International Journal on Advances in Systems and Measurements

2012 vol. 5 nr. 3&4

The *International Journal on Advances in Systems and Measurements* is published by IARIA.
ISSN: 1942-261x
journals site: http://www.iariajournals.org
contact: petre@iaria.org

International Journal on Advances in Systems and Measurements
Volume 5, Number 3 & 4, 2012

International Journal on Advances in Systems and Measurements
Volume 5, Numbers 3 & 4, 2012

CONTENTS

Weapon Detection and Classification Based on Time-Frequency Analysis of Electromagnetic Transient Images

Abdalrahman Al-Qubaa and Gui Yun Tian
School of Electrical, Electronic and Computer Engineering
Newcastle University, Newcastle upon Tyne, NE1 7RU, UK
Abd.qubaa@ncl.ac.uk, g.y.tian@ncl.ac.uk

***Abstract*—Terrorist groups, hijackers, and people hiding guns and knives are a constant and increasing threat. Concealed weapon detection has become one of the greatest challenges facing the law enforcement community today. The fact that most weapons are made from metallic materials makes electromagnetic detection methods the most prominent and preferred approach for concealed weapon detection. Each weapon has a unique electromagnetic fingerprint, determined by its size, shape and physical composition. A new detection system developed at Newcastle University that uses a walk-through metal detector with a Giant Magneto-Resistive sensor array has been utilized in this study. The system enables a two-dimensional image to be constructed from measured signals and used in later image processing. This paper addresses weapon detection using time and frequency feature extraction techniques based on this new system. The study also employs and compares two classification techniques for potential automated classification. Experimental results using guns and non-gun objects in controlled and non-controlled environments have demonstrated the potential and efficiency of the new system. The classification capabilities of the system could be developed to the point that individuals could pass through the system without the need to take off other metallic objects. The proposed techniques have the potential to produce major improvements in automatic weapon detection and classification.**

Keywords-sensor array; electromagnetic imaging; weapon detection; feature extraction; airport security.

I. INTRODUCTION

This paper, based on previous work from Al-Qubaa et al. [1], presents new results for the proposed weapon detection and classification system. There is a growing need for effective, quick and reliable security methods and techniques using new screening devices to identify weapon threats. Electromagnetic (EM) weapon detection has been used for many years, but object identification and discrimination capabilities are limited [2]. Many approaches and systems/devices have been proposed and realised for security in airports, railway stations, courts, etc. The fact that most weapons are made of metallic materials makes EM detection methods the most prominent and systems/devices built on the principle of EM induction have been prevalent for many years for the detection of suspicious metallic items carried covertly [3]. Walk-through metal detectors (WTMDs) and hand-held metal detectors (HHMDs) are commonly used as devices for detecting metallic weapons and contraband items using an EM field. Most WTMD and HHMD units use active EM techniques to detect metal objects [4][5]. Active EM means that the detector sets up a field with a source coil and this field is used to probe the environment. The applied/primary field induces eddy currents in the metal under inspection, which then generate a secondary magnetic field that can be sensed by a detector coil. The rate of decay and the spatial behaviour of the secondary field are determined by the conductivity, magnetic permeability, shape, and size of the target. Sets of measurements can then be taken and used to recover the position, the size and the shape of the objects.

Many other EM imaging techniques have been used in WTMDs. These methods include microwave [6], millimetre waves [7], terahertz waves [8], infrared imaging [9], and X-ray imaging which has been used for luggage inspections in airports [10]. All these approaches have advantages and disadvantages linked to operating range, material composition of the weapon, penetrability and attenuation factors.

Weapon detection systems currently available are primarily used to detect metal and have a high false alarm rate because they work by adjusting a threshold to discriminate between threat items and personal items, depending on the mass of the object. This leads to an increase in the false alarm level [11] [12]. Also, the human body can affect the sensitivity of the detector as when dealing with low conductivity or small materials, the human body can give a stronger signal than the material. This can cause the material to pass undetected, giving poor reliability [13]. Advanced signal processing algorithms have been used to analyse the magnetic field change generated when a person passes through a portal. Then pattern recognition and classification techniques can be used to calculate the probability that the acquired magnetic signature correlates to a weapon, or whether it is a non-weapon response [14].

Extracting distinct features from the EM signal is imperative for the proper classification of these signals. Feature extraction techniques are transforming the input image into a set of features. In other words, feature extraction is the use of a reduced representation, not a full representation, of an image to solve pattern recognition problems with sufficient accuracy. Extract or generate features from the EM signal is common method for metal detection and classification to represent the possible targets of interest. Feature extraction using Time-Frequency analysis has been used for stationary targets of backscattered signal [15]. Features are extracted from scattered field of a given candidate target from the joint time-frequency plane to obtain a single characteristic feature vector that can effectively represent the target of concern. Joint time frequency analysis was used to overcome the limitation of using the Fourier transform (FT) series to represent the EM signals which is require an infinite number of sinusoid functions [16]. The sinusoid function provides a feasible way of computing the power spectrum for EM signal, which is serves as unique fingerprint of the weapon detection response to various targets, i.e., weapons, cell phones, etc. [17].

The literature review revealed that wavelet transform (WT) are a successful method for the signal representation of time series data such as EM signals [14][18]. WT has been used to represent time series data such as ECG waveforms and mine signal detection [19]. The WT can be thought of as an extension of the classic FT except operating on a multi-resolution basis. The results obtained from [20] verify that the WT based technique produces features that are suitable for detect and identify metallic targets signal data.

After Feature extraction step, the images can be displayed for operator-assisted weapon detection or fed into a weapon detection module for automated weapon detection and classification. In [21], authors present an artificial neural network (ANN)-based scheme for metal target detection and classification. The proposed strategy involved the use of various neural networks schemes for performing feature extraction and classification tasks. It was shown that the use of an ANN in multispectral wavelengths provided a useful tool for target detection and classification. In [14], a case study on classifying metal detector signal for automated target discrimination is conducted. In this research an adaptive resonance theory networks was used and the results indicate that ANN has a vital role to improve the performance of classification. In [10], probabilistic ANN classifier used to classifies the extracted weapon candidate regions into threat and non-threat objects. The proposed framework is evaluated on a perfect database consisting of various weapons in different size, type of gun and real images and 96.48% accuracy rate has been obtained. In [22], the ANN used to differentiate between different target types such as a Glock or a starter pistol. A combination of techniques is presented that enables handguns to be effectively detected at standoff distances. Late time responses that allow threat from innocent objects (e.g., mobile phones, keys, etc.) to be distinguished from handguns. Information about the optical depth separation of the scattering corners, and the degree and shape of cross polarization allows ANN to successfully detect handguns in that research.

Support Vector Machine (SVM) has been used recently as a new machine learning methods. SVM is a concept in statistics and computer science for a set of related supervised learning methods that analyse data and recognize patterns, used for classification and regression analysis [23]. In [24], the authors revisit an attractively simple model for EM induction response of a metallic object using SVM to train and produce reliable gross characterization of objects based on the inferred tensor elements as discriminators. They focusing on gross shape and especially size to evaluate the classification success of different SVM formulations for different kinds of objects, and noticed that SVM has an inherent limitation that it takes a very long time to yield an answer in some instances. In [25], the problem of classification metallic objects using their EM induction response is solved by decomposing that response into coefficients and then using SVM and ANN to process those coefficients. The performance of each method is compared. Since it demonstrate that there is no simple relationship between sizes of objects and the overall magnitude of their coefficients, learning algorithms are necessary and useful in classifying these objects. When trained on all types of objects, both the ANN and the SVM are able to classify all objects with a good degree of accuracy. In addition, both methods show an ability to generalize for noisy test data when trained with noisy data.

A new detection system developed at Newcastle University [26] and built in a lab using an ex-service CEIA WTMD, with the addition of a Giant Magneto-Resistive (GMR) sensor array to capture the EM scattered data from any metallic objects, is used in this study.

The contributions of this paper are: firstly, improve the characterization capabilities of the new detection system through investigation of the effect of different orientations as well as the effect of concealed weapons. Secondly, investigate the feasibility of extraction features from WT and FT for weapon detection. Thirdly, automatically recognize and classify metallic threat objects by using of Support Vector Machine (SVM) and Artificial Neural Network (ANN) as classifiers.

The rest of this paper is organized as follows. Firstly, the system design and principles of operation are introduced in Section II. Next, system efficiency and test validity are reported in Section III, which is followed by details of the feature extraction methods used in Section IV. Section V explains the classification techniques used. Section VI demonstrates the classification test bed setup. Classification results are discussed in Sections VII. Finally, the conclusion is outlined in VIII.

II. SYSTEM DESIGN AND PRINCIPLES OF OPERATION

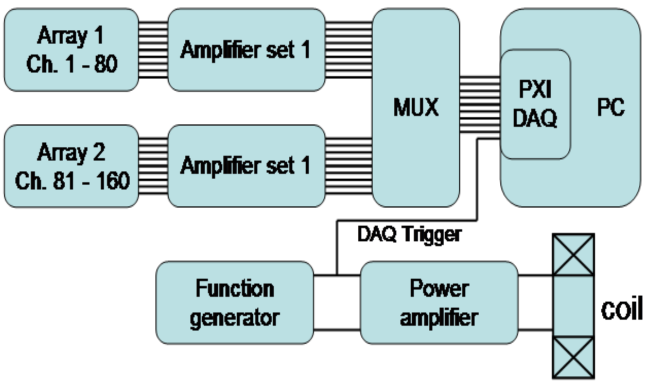
In this section, a brief description of the detection system will be given with the EM images capturing condition.

A. *System design*

The system used for the experimental tests is based around an array of NVE GMR sensors used in conjunction with the excitation coil in an ex-service CEIA (*Construction Electronics Industrial Automation)* WTMD. Figure 1 shows a block diagram of the system, the new system being converted from a typical walk through metal detector system. The signals from the sensor array are amplified using an array of signal amplifiers based on an INA111 instrumentation amplifier. Data acquisition is performed using an 80 channel PXI based National Instruments data acquisition system. The use of the PXI based system allows data to be acquired on 40 channels at a rate of 125kS/s, or 80 channels at a rate of 62.5kS/s. The channel count is further increased to a maximum of 160 by the use of multiplexer circuits. A variable excitation waveform is provided by a function generator, the signal from which is also used for data acquisition synchronisation.

The AAL002-02 GMR sensors were chosen because of their sensitivity and noise suppression compared with other common sensors such as Hall Effect models. The L in the sensor model name indicates that low hysteresis (maximum 2%) GMR material has been used to fabricate the sensor. This was chosen because it was initially intended to utilise an applied magnetic field varying from 0 to a maximum value and the lower hysteresis would minimise errors at low fields. However, after initial tests it was found that a more stable signal could be achieved by biasing the sensor response into its linear region using a DC offset superimposed on the excitation signal.

After several experimental studies the GMR sensor array was superimposed on the coil found inside the arch coil panel. The coil was positioned after taking an x-ray of the WTMD panel, and making investigation for signal measurements as shown in Figure 2.



a)

b)

Figure 1. Proposed EM system (a) Block diagram, (b) System setup in the laboratory [34].

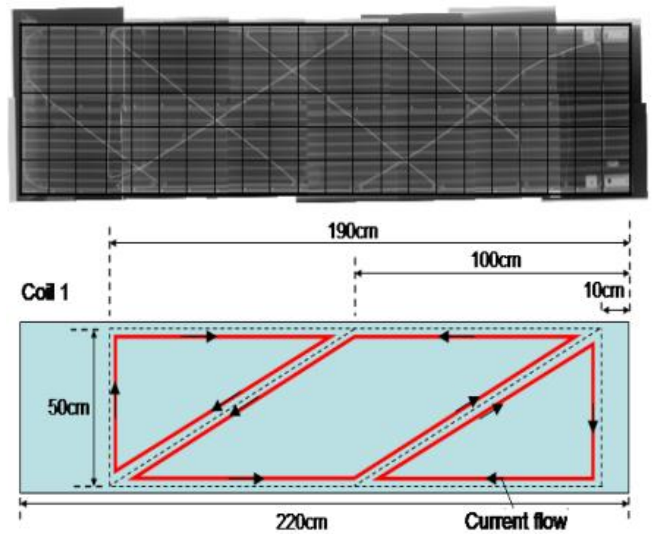


Figure 2. Coil position inside the WTMD panel.

B. *Electromagnetic field imaging*

Figure 3 illustrates the different metallic objects and their equivalent EM field images. The samples represent common threats and personal objects carried by people.

The tests were carried out using a 40 sensor array and a sample rate of 125 kHz. Thus, the temporal EM field distribution as the object moves past the array can be determined. The sensors array is aligned with the coil to pick up any distortions in the applied field due to the presence of metallic materials. Five experiments were carried out with each item and their capturing condition being summarized also in [26].

The interaction between the applied field and any sensor in the array can be captured and the pulse response from a group of sensors can be stored by moving the objects, as shown in Figure 4. If no object is present in the WTMD, the field measured by the sensor is unchanged; the presence of a metallic object causes distortion of the field, which can then be measured by the sensor.

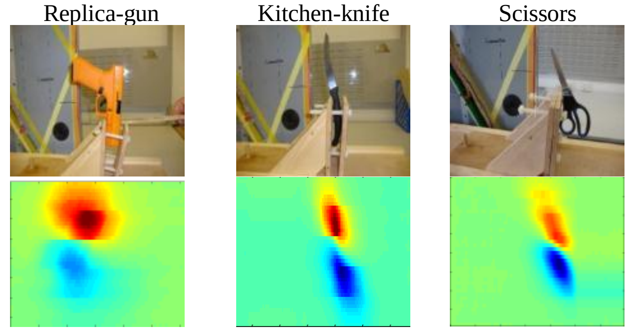


Figure 3. Samples with the equivalent EM images.

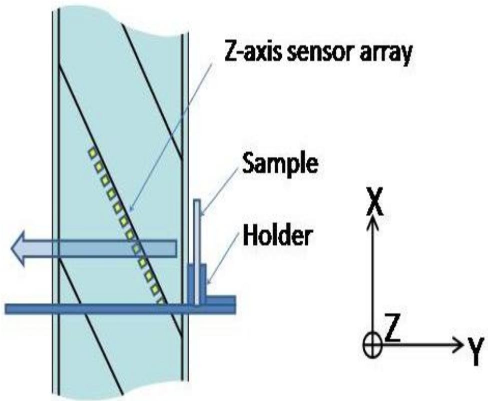


Figure 4. Object test set-up.

In the proposed system, pulsed excitation is applied to the coil. Pulsed excitation provides the opportunity to apply an interrogating field with rich frequency components in a single waveform. In the tests detailed in this paper, a pulse repetition frequency of 500Hz was used with a pulse width of 1ms and an applied current of 0.5A – 1.5A.

III. System Efficiency And Test Validity

This section explains different experimental setups used during tests to improve the ability and efficiency of the detection system. Also, repeatability is examined to check the validity and to confirm the measurements.

A. Different orientation experimental setups

To check validity of the proposed system, experiments were carried out to study the reflected signals from objects under different orientations.

Figure 6 shows the test set-up for different orientations. The objects were moved past the array dynamically and data were taken while the object was moving. Data were taken with the samples orientated in three directions.

The results of the orientation test for the kitchen knife sample are shown in Figure 6. It can be seen from the images that the feature map follows a fairly predictable evolution with the rotation of the object; in the x-direction and y-direction the object appears as a dipole distribution, with the rotation of the distribution correlating to the rotation of the object. In the z-direction, only one end of the “dipole” is presented to the array, so a uni-polar distribution is observed.

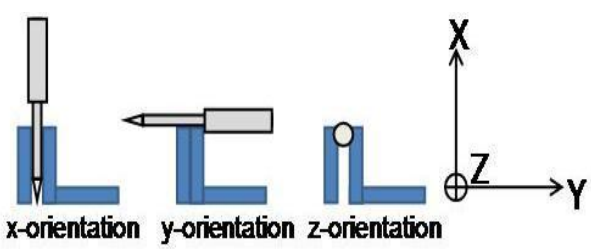


Figure 5. Orientation Test set-up

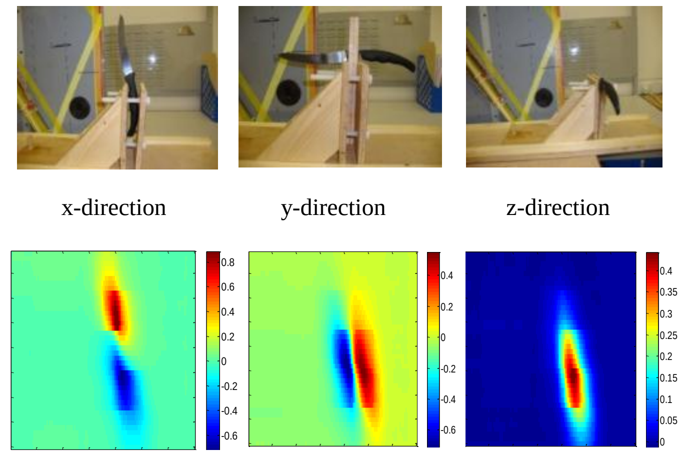


Figure 6. Different orientation results from kitchen knife sample test.

A similar trend follows for all objects; the object appears as two peaks in the feature map and as the object is rotated, this distribution is rotated from the x-direction image to the y-direction image. The z-direction, though, exhibits a clear uni-polar distribution; this is mainly due to the dimensional configuration of the sensor array.

B. Real weapon experimental setup

In order to verify the new system effectiveness using real threatening items, a series of tests using real handguns was carried out. This section reports some of the results from these tests.

Six real handguns are used (borrowed from a police station) and the specifications of these handguns are listed in Table 1. Figure 7a shows the handguns in the sample holder constructed for the tests and there reflected EM images are shown in Figure 7b. The holder was configured to ensure that the samples retained a constant and comparable distance and orientation with respect to the array during each pass through the system.

Table 1. SPECIFICATION FOR THE ALL HANDGUNS.

Sample	Description
1	Small revolver – 0.38” Smith & Wesson – Deactivated.
2	Revolver – 0.38” Enfield service revolver – Deactivated.
3	Large automatic – 9mm Glock G17 – Live.
4	Large automatic – 0.45 Colt M1911 – Replica.
5	Small revolver – Brocock Puma air pistol – Live.
6	Small automatic – 7.65mm Walther – Deactivated.

The test samples were representative of a range of handguns which would be of interest for detection. In general, it is unlikely that the deactivated weapons would give a drastically different response to the same live weapon. Although the presence of ammunition clearly increases the volume of material within the arch, it would not affect results significantly.

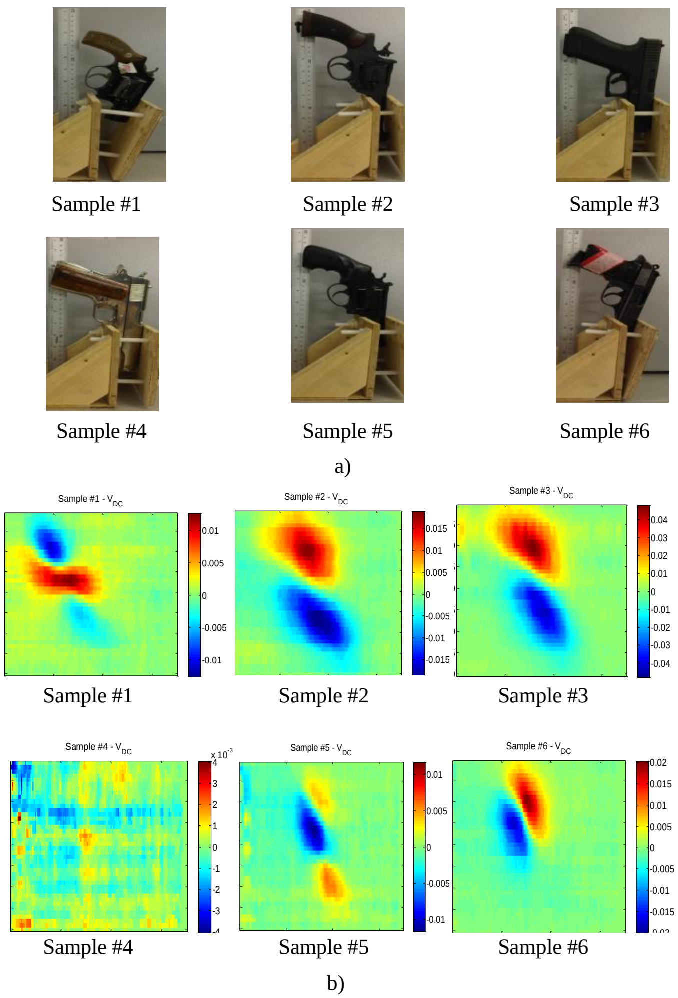


Figure 7. Handgun samples (a) In sample holder, and (b) Reflected EM image results.

The first set of tests was carried out in the same way as the previous detection tests (explained in Section II above). Results are shown in Figure 7b and are generally good, as would be expected for the controlled test setup. The one exception is sample 4, the replica gun, which is very difficult to locate in the test results. It had the lowest amplitude response and therefore the poorest signal to noise ratio.

It is notable that samples 2 and 3 gave similar results, responding to a type of dipole distribution, indicative of a ferromagnetic object made predominantly from a single type of metal. The simple form of distribution also indicates that there is very little metal in the handle of these objects, and the array saw them as a simple tube/block of metal.

Another test setup for the uncontrolled walk-through tests using real handguns is shown in Figure 8a. The handguns were carried in the inside jacket pocket of an individual walking through the arch. Test results are shown in Figure 8b. The EM images for the objects are clearly compressed along the x-axis compared to the controlled tests, due to the object moving through the arch at a greater speed; however, the actual distributions remain very similar.

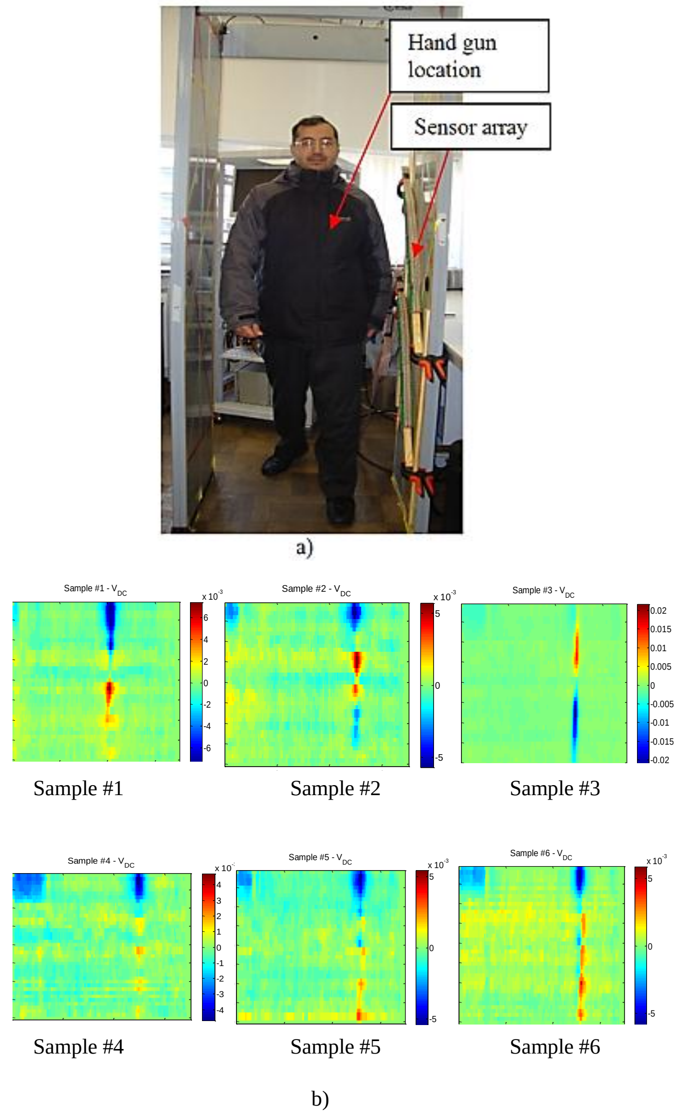


Figure 8. Walk-through test (a) Test set-up, (b) Results for all hand guns from the walk-through test.

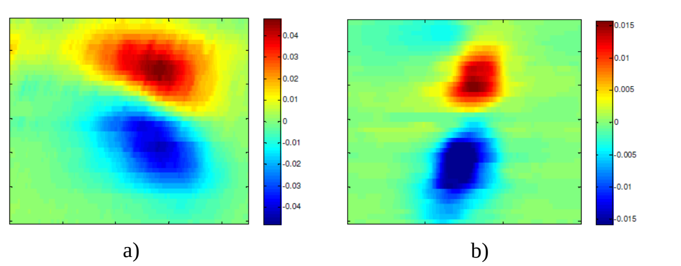


Figure 9. EM images for the sample #3 - a) Controlled, b) Uncontrolled tests.

Figure 9 shows the EM images for the controlled and walk-through tests using the handgun sample #3. The results from the walk-through test have been expanded along the horizontal axis compared to Figure 8b to aid comparison. Comparison of the plots shows that although the controlled and walk-through tests did not give identical results, the general form of the EM signatures was roughly similar. Thus, using appropriate analysis techniques it can be ascertained that the signatures are from similar, if not the same object.

C. *Measurement Stability*

In this subsection, system stability is checked by examining the repeatability of the experimental tests using a simple amplitude calculation. The overall amplitude change for each EM signal is plotted for five repetitions of the test for all handgun samples. The overall amplitude change is computed by subtracting the maximum from the minimum value of each EM signal received. The results are plotted in Figure 10 for both the controlled and walk-through test. It can be seen from Figure 10a that, the controlled test has the greatest repeatability. As well as the data trend is similar for the walk-through test as showed in Figure 10b.

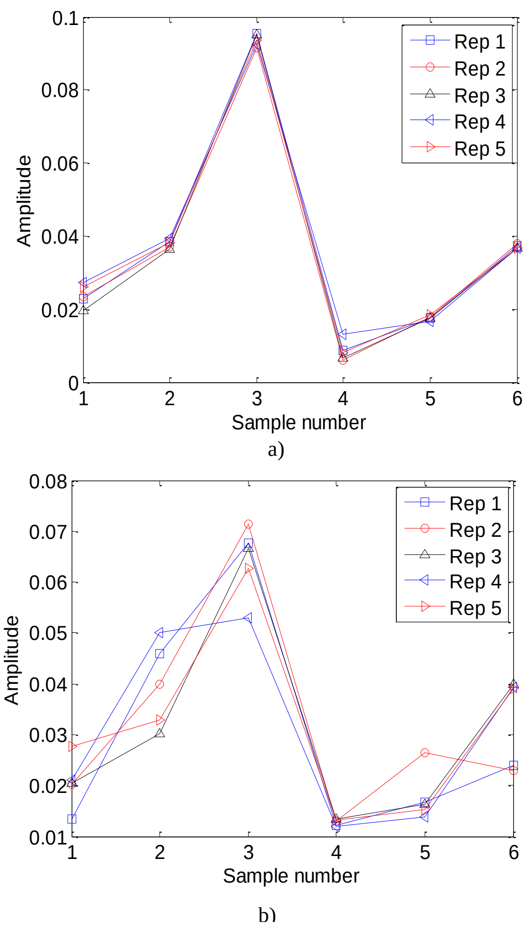


Figure 10. Overall amplitude change for (a) Controlled test, (b) Walk-through test.

IV. FEATURE EXTRACTION TECHNIQUES

Features were extracted from two different techniques (FT and WT). A brief background of FT and WT will be provided, with the motivation behind their use. In addition to detailed the feature extractions algorithm.

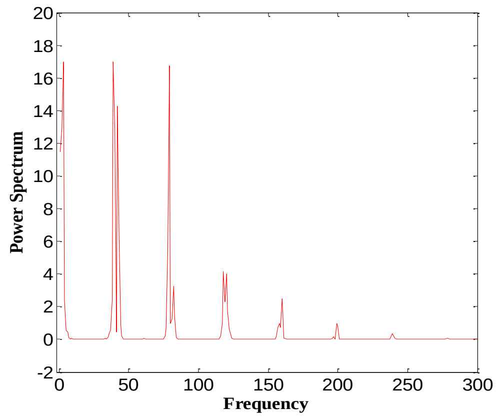
A. *Fourier Transform (FT)*

The Fourier series provides an alternative way to represent data; instead of representing the signal amplitude as a function of time, we represent the signal by how much information is contained at different frequencies. Fourier analysis is important in data acquisition as it allows one to isolate certain frequency ranges. The bridge between time and frequency representation is the FT. The signal can be decomposed as a weighted sum of sinusoid functions. This provides a feasible way of computing the power spectrum for a signal. Fast FT (FFT) is a fast algorithm of the discrete FT that represents the signals in the frequency domain. The power spectrum serves as the fingerprint of the analysed signal [16]. The absolute value will provide the total amount of information contained at a given frequency [27], and the square of the absolute value is considered the power of the signal. In this work the power spectrum (PS) of FFT for each EM image using the outcomes from the control test were utilised as a feature; each sample gave different PS, as shown in Figure 11.

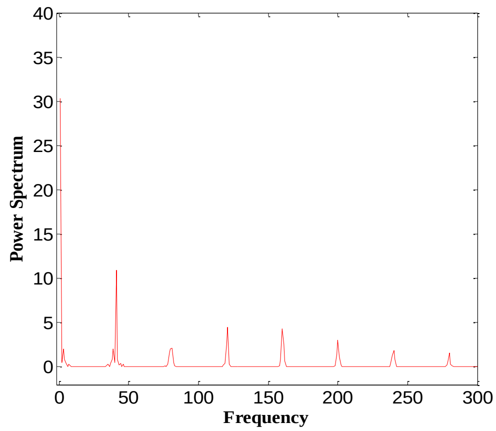
The PS results will be ($n*m$), so to reduce the data size before applied to the classifier, Principle component analysis (PCA) techniques [28] was applied and first three PCA components were selected as it represent a 99.6% of the data variance. Figure 12 shows the behaviour of the PCA feature vectors extracted from the FFT process. The test was done using the six handguns with the different other not-threat objects. It is clear from the figure that the handgun #4 gives very low response because it consists of plastic material as well as the mobile phone object gives high response because it is full charged. Figure 13 shows the flowchart of the gun classification procedure using FFT features.

B. *Wavelet Transform (WT)*

In contrast to FFT, Wavelet analysis is useful in decomposing a time series into time-frequency space simultaneously. The analysis provides information about both the amplitude of any "periodic" signals within the series, and how this amplitude varies with time. WT can be considered as an extension of the classic FFT except that it operates on a multi-resolution basis. This multi-resolution property enables a signal to be decomposed into a number of different resolutions. Each resolution represents a particular coarseness of the signal. Preservation of spatial information is another property of WT after transformation. This enables the identification of areas in the original signal that correspond to particular characteristics present in the WT data [29].



a)



b)

Figure 11. Part of the power spectrum of (a) Handgun, (b) Mobile Phone.

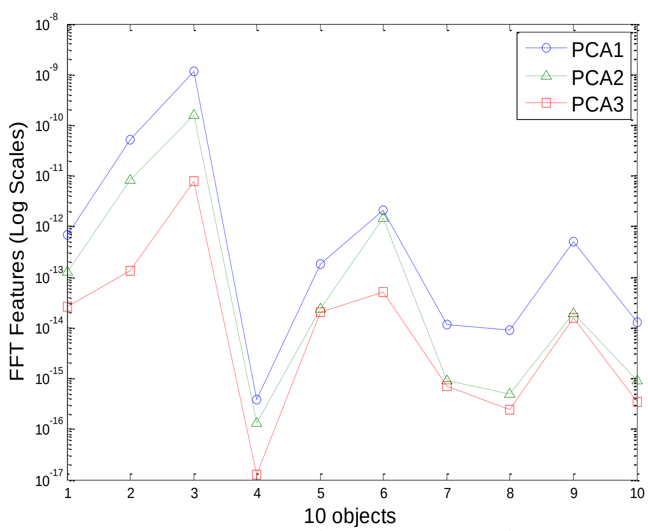


Figure 12. Feature vector extracted from the FFT process for 10 objects , #1-#6 are threat items (handgun samples) and the others (#7-#10) are not-threat items (Camera, House Key, Phone and Pen) as sorted number in the figure respectively.

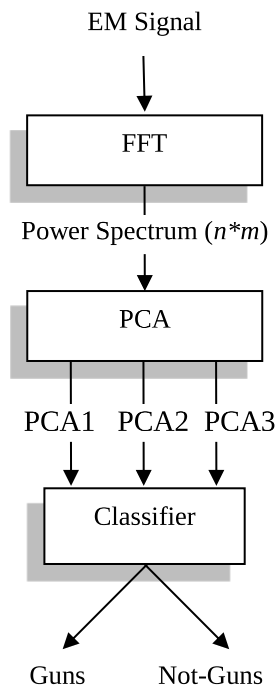


Figure 13. Flowcharts of the detailed gun classification procedure using FFT features.

The motivation behind using WT came from a preliminary analysis of the raw metal detector signal data. Literature proves that WT can be applied successfully as a method of signal representation of time series data. An example of this, is its application to electrocardiogram signals, which share a similar resemblance to metal detector signals [30][31]. Previous research also verifies that WT can be used to produce suitable features from metal detector data for target classification [14][32].

In this study, a discrete WT has been used. 'Discrete' refers to its discrete sets of dilation and translation factors, and discrete sampling of the signal. At a given scale, J, a finite number of translations were used in applying multi resolution analysis to obtain a finite number of scaling and wavelet coefficients. The signal can be represented in terms of the following coefficients (Eq. 1) [33]:

$$f(x,y) = \sum_k C_{JK}\ \varphi_{JK}(x) + \sum_{j=1}^{J}\sum_k d_{jk}\psi_{jk}(x) \qquad (1)$$

where φ_{JK} are the scaling functions, C_{Jk} are the scaling coefficients, ψ_{jk} are the mother wavelets and d_{jk} are the wavelet coefficients.. The first term in Eq. (1) gives the low resolution approximation of the signal, while the second term gives the detailed information at resolutions from the original down to the current resolution J. Daubechies order 4 has been selected from the wavelet family and three levels of wavelet decomposition have been computed [32].

Three types of statistical operation were selected from the wavelet approximation coefficients (LL) domain to be considering as a unique feature for each object, namely Entropy, standard deviation (STD) and root mean square error (RMSE). Figure 14 shows the ENT, STD and RMSE features for the three WT levels using same 10 objects in the previous test, the six handguns and the others are the not-threat objects. It can be seen from the figure that these feature give good indication to identify between handguns and the other objects. However, some of the not-threat objects are close to the handguns features, such that the entropy of the house key for instance. This leads to combine wavelet features with the FFT features to give a good classification result.

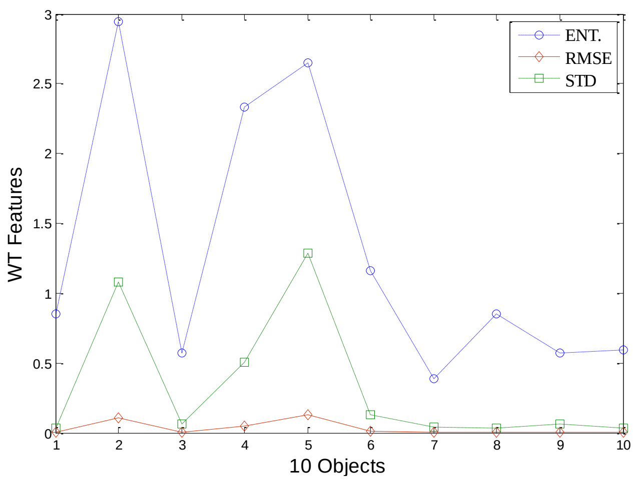


Figure 14.Wavelet features for 10 objects, #1-#6 are threat items (handgun samples) and the others (#7-#10) are not-threat items (Camera, House Key, Phone and Pen) as sorted number in the figure respectively, for one level WT analysis.

As a result, each EM image had three types of features with three levels of decompositions. Thus, a feature vector was generated consisting of nine values to be fed to the classifiers. A flowchart of the gun classification procedure using the control test and features obtained from the WT coefficient is shown in Figure 15.

V. CLASSIFICATION TECHNIQUES

In this work, two different types of classification techniques have been applied and compared to evaluate the features which extracted from the new system and to adopt the efficient classification technique for an automated process. In previous work [1], the classification of different objects was performed using cross correlation technique. In this work, the classification has been performed using ANN and SVM techniques.

A. Classification using Artificial Neural Network (ANN)

Neural networks are widely used in pattern recognition and classification since they do not need any information about the probability distribution and the priori probabilities of different classes. In this study, two feature vectors (WT and FFT) were used as an input to an ANN classifier alone and in combination with each other to find out the most suitable features for classification.

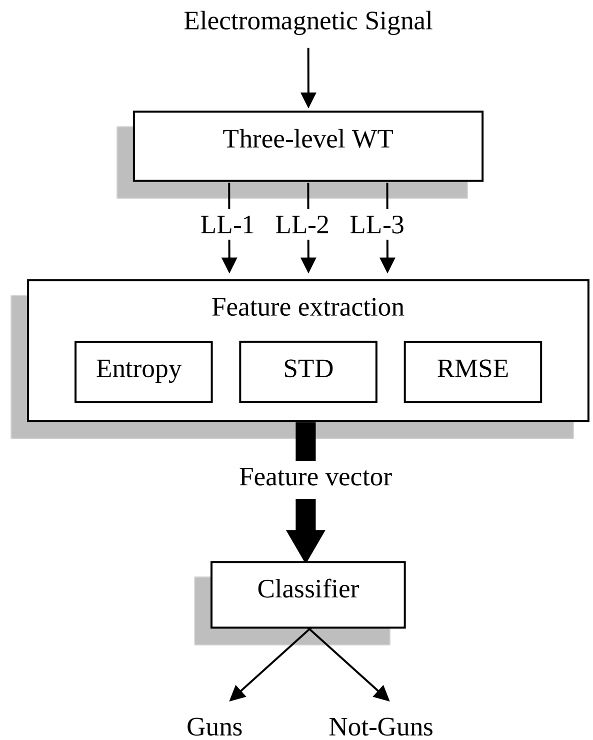


Figure 15. Flowchart of the gun classification procedure using discrete WT features.

A three layered ANN classifier was used: input, hidden and output layer. The input layer use the input feature vector element *N*, the hidden layer consisted of 2N nodes with the sigmoid activation function. The output layer consisted of one node (for our study, symbolizing whether it is a gun or not) with a linear activation function. The ANN classifier utilized a training function based on Levenberg-Marquardt optimization [34]. All ANN parameters are summarized in Table 2.

Table 2. ARTIFICIAL NEURAL NETWORK PARAMETERS.

No. of nodes in Input layer	:	Same no. of used feature vector elements (*N*).
No of nodes in Hidden layer	:	Double no. of used feature (*2N*).
No of nodes in Output layer	:	One node (Gun or not).
Transfer function	:	'*logsig*' for hidden layer, '*purelin*' for output layer
Training function	:	'*trainlm*'
Max number of Epochs	:	10000
Min performance gradient	:	1e-10

B. Classification using Support Vector Machine (SVM)

SVM is a concept in statistics and computer science for a set of related supervised learning methods that analyse data and recognize patterns. It is used for classification and regression analysis. SVM tackle classification problems by nonlinearly mapping input data into high-dimensional feature spaces, wherein a linear decision hyper plane separates the two considered classes. Different kernel functions nonlinearly maps samples into the higher dimensional space. Among them, the RBF Kernel is the most generally used. In many pattern recognition applications, SVM generalization performance is either similar or significantly better than other competing methods. In opposition to ANN, SVM have few hyper parameters to be adjusted [35].

In this work, a SVM based method was used as a second classification method to make a comparison with the results of ANN. The four feature vectors were normalised to the range [-1, +1] in each column first. The advantages of scaling are that it avoids attributes in greater numeric ranges dominating those in small numeric ranges, and it avoids numerical difficulties during calculation. In this work, the Radial Basis Function (RBF) was used as a kernel (K), as in Eq. 2 [23]:

$$K(x_i, x_j) = \exp(-\gamma \| x_i - x_j \|^2), \qquad \gamma > 0 \tag{2}$$

The RBF kernel nonlinearly maps samples into a higher dimensional space. Thus, it can handle the case when the relation between class labels and attributes is nonlinear. Furthermore, the linear kernel is a special case of RBF. Finally, the RBF kernel has fewer numerical difficulties. The LIBSVM, a library for SVM developed by Chang and Lin [23], was used in this work. After training data using SVM, the model was obtained for prediction of known objects. The model was tested using the training data to identify the classification rate.

VI. CLASSIFICATION TESTBED SETUP

Experiments were conducted with the new EM system to classify between guns and non-gun objects. The sample group consisted of twelve different objects. Six of the objects were handguns, while the others were commonly used objects that contain metallic parts. The specifications of all objects used are given in Table 3.

The handgun samples represent the most common weapons seized by the police; of particular interest is sample #5, the blank firer, which had been converted to fire live ammunition by the welding of another barrel to the existing mechanism, and the replica hand gun, sample #6, commonly used by armed robbers as a threat.

All samples constructed for the tests were controlled by a sample holder. The holder was configured to ensure that the samples retained a constant and comparable distance and orientation with respect to the array during each pass through the system. All weapon sample compositions included steel, with several incorporating other materials such as zinc alloy, aluminium, and polymers.

Table 3. OBJECTS USED IN EXPERIMENTAL TEST: (A) REAL GUNS AND (B) COMMONLY USED OBJECTS.

(a) Real Guns		(b) Other Objects	
#1	Small revolver 0.516g	**#7**	Mobile phone
#2	Small semi-automatic 0.637g	**#8**	wristwatch
#3	Medium revolver 0.937g	**#9**	House Key
#4	Medium semi-automatic 0.689g	**#10**	Screwdriver
#5	Converted blank firer 0.800g	**#11**	Scissors
#6	Replica1 1.140g	**#12**	Kitchen knife

Differences between two EM signals for two different objects were found to be close to the differences between two EM signals for the same object in two tests [26]. This conclusion was used to increase the number of samples; each object was tested five times using the new system to generate five EM samples for the same object. Hence, for the twelve objects under test, six were guns and six were non-gun objects, 60 EM samples were generated. Based on this, the classifier was trained using 48 samples for all objects (four for each object) while the remaining 12 samples were used as test samples as explained in Table 4.

Table 4. DATA SET OF THE WORK

Total No. of Images used= 60			
No. of training images = 48		No. of testing images = 12	
Guns	Non-Guns	Guns	Non-Guns
24	24	6	6

VII. CLASSIFICATION RESULTS AND DISCUSSION

In this section, an assessment of the proposed features for object classification is carried out using the two classification methods.

Firstly, each type of feature vector individually provides for the ANN and SVM as inputs. The results are shown in Table 5 and 6 respectively. It can be seen from the tables that there is a major difference between ANN and SVM, that while the overall classification rates are largest when using the ANN classifier, the SVM classifier is more sensitive and shows 0% misdetection of the guns group. Although both classification methods give a high detection rate in terms of handgun objects, both methods give a high misdetection rate for non-gun objects. This is because training on these objects is not efficient due to the diversity in their sizes, shapes and materials. The classification results could be improved using a larger database size [35].

Table 5. RESULTS OF ANN FOR EACH FEATURE VECTOR.

Feature vector	Hidden layer neurons	Objects	Correctly classified	Incorrectly classified	Classification rate
WT	18	Gun	5	1	83%
		Non-Gun	5	1	83%
			10	2	**83%**
FFT	6	Gun	5	1	83%
		Non-Gun	3	3	50%
			8	4	**67%**

Table 6. RESULTS OF SVM FOR EACH FEATURE VECTOR.

Feature vector	Objects	Correctly classified	Incorrectly classified	Classification rate
WT	Gun	6	0	100%
	Non-Gun	2	4	33%
		8	4	**67%**
FFT	Gun	6	0	100%
	Non-Gun	1	5	16%
		7	**5**	**58%**

Figure 16 shows the classification results with the average error from the ANN and SVM techniques.

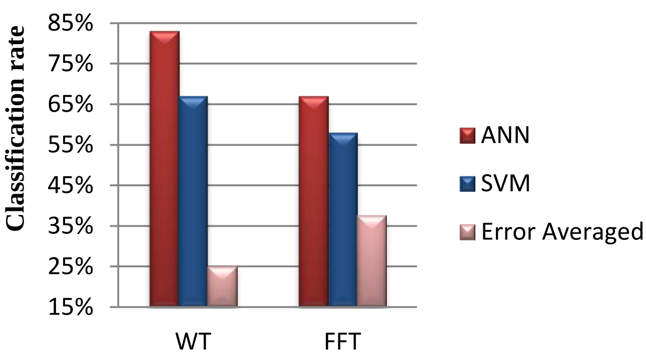


Figure 16. Classification rate results of ANN and SVM using WT and FFT features.

VIII. CONCLUSION AND FUTURE WORK

This paper has demonstrated an EM metal detector system and investigated the feasibility of object identification and classification. In comparison with conventional induction based WTMDs, the new GMR array based system has shown great potential in objects classification as the samples are made from mixed material is identified. Whereas the induction based WTMD can only discriminate between metal and non-metal object, this system has taken it a step further. Results obtained using the WT and FFT-based features have been given in detail to show the validity of the new system. The proposed features is more advanced in object characterisation as it depends on the amplitude distribution of the EM field making training possible using a database of objects; unlike traditional thresholding adopted in the induction based system, which largely depends on material volume. These features were utilized to classify 12 objects. Six were real handguns and six were different commonly used metallic items. The real handguns were examined in a controlled and uncontrolled environment such that the handguns were either controlled by a holder or carried on the inside of a jacket pocket of an individual walking through an arch. A comparison between two classification methods, ANN and SVM, showed promising results for detecting and classifying objects in security applications of EM signal. To conclude, WT features, due to their spatial frequency properties, showed almost better classification rates than FFT features, which have frequency properties. In terms of classifiers, ANN classification techniques have a better classification rate in general. However, SVM had very high sensitivity and showed very low false negative alarm (100% classification rate) in terms of handgun samples. These initial results aimed to identify possible methodologies for analysis and classify of EM signals.

For future work, the classification capabilities of the system could be developed to the point that individuals pass through the system without removing metallic objects from their person. This would be realised through "training" the system to identify threat objects by presenting the system with a wide variety of threat and non-threat objects. Further work can also utilize more EM features for accurate detection and classification of threat objects.

ACKNOWLEDGMENT

This project is funded under the Innovative Research Call in Explosives and Weapons Detection (2007), a cross-government programme sponsored by a number of government departments and agencies under the CONTEST strategy. The authors would like to thank cross-government departments for joint experimental tests.

REFERENCES

[1] A. Al-Qubaa, G. Tian, and J. Wilson, "Electromagnetic Imaging System for Weapon Detection and Classification", Proc. Fifth International Conference on Sensor Technologies and Applications, SENSORCOM 2011, August 21-27, 2011, Nice, pp. 317-321, IARIA XPS Press, ISBN: 978-1-61208-144-1.

[2] A. Agurto, Y. Li, G. Y. Tian, N. Bowring, and S. Lockwood, "A Review of Concealed Weapon Detection and Research in Perspective", Proceedings of the 2007 IEEE International Conference on Networking, Sensing and Control, London, pp. 443-448, 2007.

[3] C. A. Dionne, J. J. Schultz, R. A. Murdock, and S. A. Smith, "Detecting Buried Metallic Weapons in A Controlled Setting using a Conductivity Meter", Forensic Science International, vol. 208, pp. 18-24, 2011.

[4] CEIA Security Metal Detectors. [On-line 09.10.2012] http://www.ceia.net/security.

[5] N. G. Paulter, "Users' Guide for Hand-Held and Walk-Through Metal Detectors", NIJ Guide 600-00, NCJ 184433, Office of Science and Technology, U.S. Department of Justice, Washington, DC 20531, 2001. [On-line 09.10.2012] http://www.nij.gov/pubs-sum/184433.htm.

[6] D. M. Sheen, D. L. McMakin, and T. E. Hall, "Three-Dimensional Millimetre-Wave Imaging for Concealed Weapon Detection", IEEE transactions on microwave theory and techniques, vol. 49, pp. 1581-1592, 2001.

[7] S. Yeom, D.-S. Lee, Y. Jang, M.-K. Lee, and S.-W. Jung, "Real-Time Concealed-Object Detection and Recognition with Passive Millimeter Wave Imaging", Opt. Express, Vol. 20, pp. 9371-9381, 2012.

[8] K. B. Cooper, R. J. Dengler, N. Llombart, B. Thomas, G. Chattopadhyay, and P. H. Siegel, "THz Imaging Radar for Standoff Personnel Screening", Terahertz Science and Technology, IEEE Transactions on, vol. 1, pp. 169-182, 2011.

[9] C. Siu-Yeung and T. Nanda-Pwint, "Using Infrared Imaging Technology for Concealed Weapons Detection and Visualization", in TENCON 2010 - IEEE Region 10 Conference, pp. 228-233, 2010.

[10] H. Pourghassem, O. Sharifi-Tehrani, and M. Nejati, "A Novel Weapon Detection Algorithm in X-Ray Dual-Energy Images Based on Connected Component Analysis and Shape Features", Australian Journal of Basic and Applied Sciences, vol. 5, pp. 300-307, 2011.

[11] US Department of Justice, "Final Report - Demonstration of a Concealed Weapons Detection System Using Electromagnetic Resonances", January 2001. [On-line 09.10..2012] http://www.ncjrs.gov/pdffiles1/nij/grants/190134.pdf.

J. Ashcroft, D. J. Daniels, and S. V. Hart, "Walk-Through Metal Detectors for use in Concealed Weapon and Contraband Detection", U.S. Department of Justice, Office Justice Programs, National Institute of Justice, 2003. [On-line 22.06.2012] http://www.nij.gov/pubs-sum/193510.htm.

[12] H. M. Chen, S. Lee, R. M. Rao, M. Slamani, and P. K. Varshney, "Imaging for Concealed Weapon Detection", IEEE Signal Processing Magazine, Vol. 22, pp. 52-61, 2005.

[13] M. Tran, C. Lim, C. Abeynayake, and L. Jain, "Feature Extraction and Classification of Metal Detector Signals Using the Wavelet Transform and the Fuzzy ARTMAP Neural Network", Journal of Intelligent & Fuzzy Systems: Applications in Engineering and Technology, Vol. 21, pp.89-99, 2010.

[14] G. Turhan-Sayan, "Real Time Electromagnetic Target Classification Using a Novel Feature Extraction Technique with Pca-Based Fusion", Antennas and Propagation, IEEE Transactions on, vol. 53, pp. 766-776, 2005.

[15] D. Kotter, L. Roybal, and R. E. Polk, "Detection and Classification of Concealed Weapons using a Magnetometer-based Portal", SPIE proceeding, Sensors and Command, Control, Communications, and Intelligence (C3I) Technologies for Homeland Defence and Law Enforcement, vol. 4708, pp. 145-155, 2002.

[16] G.Y. Chen and B. Kegl, "Feature Extraction using Radon, Wavelet and Fourier Transform", Systems, Man and Cybernetics, 2007. ISIC. IEEE International Conference on , pp.1020-1025, 2007.

[17] H. Krüger and H. Ewald, "Signal Processing and Pattern Recognition for Eddy Current Sensors, used for Effective Land-Mine Detection", Autonomous and Intelligent Systems, vol. 6752, pp. 294-302, 2011.

[18] F. Zhang and M. Li, "Wavelet Analysis Method of Harmonics and Electromagnetic Interference in Coal Mines", Mining Science and Technology (China), vol. 20, pp. 576-580, 2010.

[19] M. Tran and C. Abeynayake, "Evaluation of the Continuous Wavelet Transform for Feature Extraction of Metal Detector Signals in Automated Target Detection: New Advances in Intelligent Decision Technologies", K. Nakamatsu, G. Phillips-Wren, L. Jain, and R. Howlett, Eds., ed: Springer Berlin/Heidelberg, vol. 199, pp. 245-253, 2009.

[20] M. Xi, M. R. Azimi-Sadjadi, T. Bin, A. C. Dubey, and N. H. Witherspoon, "Detection of Mines and Minelike Targets Using Principal Component and Neural-Network Methods", Neural Networks, IEEE Transactions on, vol. 9, pp. 454-463, 1998.

[21] A. A. David, B. Nicholas, D. R. Nacer, S. Matthew, G. Elizabeth, H. Stuart, and A. Ali, "A Multifaceted Active Swept Millimetre-Wave Approach to the Detection of Concealed Weapons", Millimetre Wave and Terahertz Sensors and Technology. Edited by Krapels, Keith A.; Salmon, Neil A., Proceedings of the SPIE, vol. 7117, pp. 711707-711707-10, 2008.

[22] C. C. Chang and C. J. Lin, "LIBSVM: A Library for Support Vector Machines", ACM Transactions on Intelligent Systems and Technology, vol. 2, 2011.

[23] J. Fernandez, B. Barrowes, K. O'Neill, K. Paulsen, I. Shamatava, F. Shubitidze, and K. Sun, "Evaluation of SVM Classification of Metallic Objects Based on a Magneticdipole Representation", presented at the Detection and Remediation Technologies for Mines and Minelike Targets XI, 2006.

[24] Z. Beijia, K. O'Neill, K. Jin Au, and T. M. Grzegorczyk, "Support Vector Machine and Neural Network Classification of Metallic Objects Using Coefficients of the Spheroidal MQS Response Modes", Geoscience and Remote Sensing, IEEE Transactions on, vol. 46, pp. 159-171, 2008.

[25] G. Tian, A. Al-Qubaa, and J. Wilson, "Design of an Electromagnetic Imaging System for Weapon Detection based on GMR Sensor Arrays", Sensors & Actuators A, vol. 174, pp. 75-84, 2012.

[26] Computing Fourier Series and Power Spectrum with MATLAB, [On-line 09.10.2012]. http://faculty.olin.edu/bstorey/Notes/Fourier.pdf.

[27] A. Sophian, G. Tian, D. Taylor, and J. Rudlin, "A Feature Extraction Technique Based on Principal Component Analysis for Pulsed Eddy Current NDT", NDT&E International, vol. 36, pp. 37-41, 2003.

[28] H.G.Stark, "Wavelets and Signal Processing: An Application-based Introduction", Springer, Heidelberg, 2005.

[29] S. Mahmoodabadi, A. Ahmadian, M. Abolhasan, M. Eslami, and J. Bidgoli, "ECG Feature Extraction Based on Multiresolution Wavelet Transform", 27th IEEE Annual International Conference of the Engineering in Medicine and Biology Society, pp. 3902-3905, 2005.

[30] Q. Zhao and L. Zhang, "ECG Feature Extraction and Classification Using Wavelet Transform and Support Vector Machines", International Conference on Neural Networks and Brain, ICNN&B, pp. 1089-1092, 2005.

[31] M. Tran, and C. Abeynayake, "Evaluation of the Continuous Wavelet Transform for Feature Extraction of Metal Detector Signals in Automated Target Detection", New Advances in Intelligent Decision Technologies, Springer, Heidelberg, pp. 245-253, 2009.

[32] K. Amolins, Y. Zhang, and P. Dare, "Wavelet Based Image Fusion Techniques: An Introduction, Review and Comparison", ISPRS Journal of Photogrammetry and Remote Sensing, vol. 62, pp. 249-263, 2007.

[33] H. Demuth, M. Beale, "Neural Network Toolbox for use with MATLAB", Mathworks, Inc. Natick, MA , 2002.

[34] M. Curilem, M. Chacón, G. Acuña, S. Ulloa, C. Pardo, C. Defilippi, and A. Madrid, "Comparison of Artificial Neural Networks and Support Vector Machines for Feature Selection in Electrogastrography Signal Processing", Engineering in Medicine and Biology Society, Annual International Conference of the IEEE, pp. 2774-2777, 2010.

Increase of Robustness of Production Plans using a Hybrid Optimization and Simulation Approach

Christoph Laroque
Heinz Nixdorf Institute
University of Paderborn
Paderborn, Germany
laro@hni.uni-paderborn.de

Robin Delius
Heinz Nixdorf Institute
University of Paderborn
Paderborn, Germany
robin.delius@hni.upb.de

Jan-Hendrik Fischer
University of Paderborn
Paderborn, Germany
jhfischer@gmail.com

Dennis Horstkemper
University of Paderborn
Paderborn, Germany
dhorstkemper@gmail.com

***Abstract* — We propose the use of the material flow simulation to evaluate the robustness of a production plan, which was created and optimized with no respect to unforeseen derivations. The necessary probabilities for machine failures and similar operational events on the floor can easily be integrated in the simulation model, in order to analyze, how initial plan performs in these situations. The influence of unforeseen events in daily production cannot be modeled within mathematical optimization without consuming either large amounts of computation time or requiring domain specific techniques, which increasingly decrease the maintainability of a model. We show a possible way to use simulation to evaluate and enhance a production plan. We illustrate the developed process using a real-world use-case of medium complexity and can show, that simulation is able to evaluate the robustness of a given pre-optimized production plan.**

Keywords - material flow simulation; robustness; production planning; mathematical optimization

I. Motivation

Even after overcoming the global economic crisis tremendous requirements exist within the daily operation of a production facility and its supply chain. Fluctuating demands are leading to less adequate forecast data and the need to lower capital commitment is leading to the necessarily of designing robust production planning models [1], [5], [6]. Major objective is always to be able to serve all demands in due time while causing minimal costs.

Several uncertainties exist within the production planning process. On the one hand, many unforeseen events can take place: machine failures, missing materials, changed sales demands or ill employees are only a small subset of possible examples. On the other hand, it is simply impossible to include all factors that might occur into the planning process in the first place. Therefore, planning methods are always based on different models of a production structure, which are an abstraction of reality themselves. It is the responsibility of the production planner to decide, which factors he wants to take into the account when creating the models. He always has to find a compromise between the detail level of the model (and therefore its significance) and the solvability of the optimization problem, which is created on its basis. The lot sizing and scheduling problems that are used within production planning are usually already np-complete even in their simplest form [15]. Therefore, one cannot guarantee to be able to find acceptable solutions in a timely manner while using modern operation research techniques. We show that the applying uncertain information to a simple stochastic optimization model yields either unacceptable solutions or unacceptable solution times. Thus, we have to find a way to include the aforementioned uncertainties within the production planning process without limiting its solvability significantly. Using advanced stochastic optimization techniques is out of the question: Special techniques, like the Benders Decomposition [22], are based upon the specific structure of certain optimization model. As production systems underlie a constant change however, the model needs to be able to be adapted constantly. Additionally efforts made in applying such a model for a certain production system cannot be reused, as two production systems and therefore the corresponding optimization models are seldom identical or even similar [23]. To create an approach, which allows for easy maintainability, sufficient solution quality including uncertain events and practical solution times we connect mathematical optimization models with a down streamed material flow simulation. While we always assume optimal conditions within the mathematical optimization model, we are including the uncertainties in the simulation process. This allows us to analyze whether a production plan is able to perform well creating an acceptable monetary solution under these changed conditions or not. We can improve upon scheduling decisions using rule-based machine controls within the simulation, reacting to changes in the production systems environment. In addition, we are able to create automatic or manual modifications of the plan and can evaluate these as well using additional simulations. It is easily possible to develop a more robust production plan with this toolset.

Simulations usually are used to verify the solutions of an optimization problem. However, the aim of our research is to replace parts of the optimization process with simulation methods to receive solutions with an acceptable quality on a timely manner. First, we solve a mathematical optimization problem with standard solver software like IBM ILOG CPLEX [17]. The solution generated by the optimization is represented by a mst-file, will then be converted into a machine readable production plan, which is used as an input for the simulation environment d³fact. Within this environment we can evaluate the performance of a plan against environmental influences. Additionally, we can create production plans, which are optimized for a certain scenario, meaning a distinct combination of different influences, and use several of these plans to create a new production plan within a post-processing software. This plan shows an increased performance in any scenario, which we can proof by simulating and evaluating it again. Figure 1 shows this general optimization and simulation process.

After regarding the necessary State-of-the-Art in Section II, we describe the production model and the corresponding optimization models in Section III. It is possible to include uncertainties in the planning phase within the mathematical optimization process. We discuss these methods in Section IV and analyze the corresponding solution quality and solution times. To generate a more robust production plan based upon a given near optimal plan we propose a procedure, which generates and evaluates a number of scenarios with the help of offline simulations to create a new plan. We explain the transfer of the optimization solutions into the simulation process in Section V. To cover a broad spectrum of stochastically possible scenarios; several replications of the stochastic simulation based upon the production structure are performed. This way we are able to cover a wide field of possible scenarios for machine failures and other events.

The production schedules are logged and afterwards evaluated on the base of costs and robustness. A rule-based machine control is used to reduce possible production losses when intermediate products were not assembled in due time. An additional post-processing can be used to maintain further robustness increasing actions. The effect of these actions can be evaluated using further simulations. We present these processes in Section VI. We finally evaluate the outcome of our work using a case study. Additionally, we give a conclusion (Section VII) and an outlook towards further possibilities and improvements for this approach.

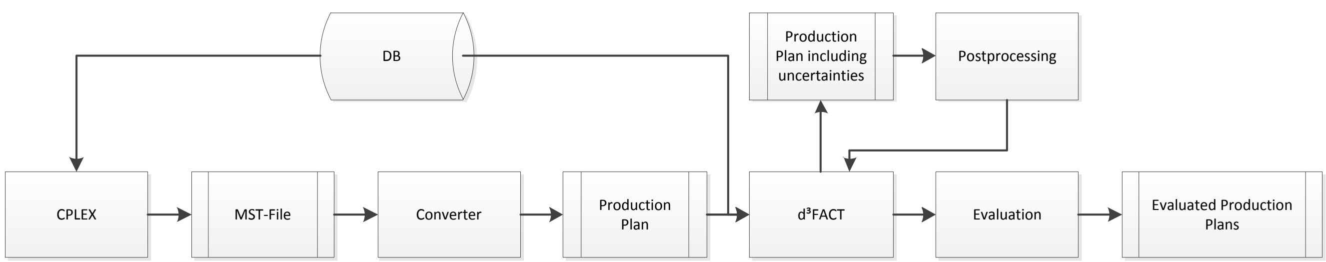


Figure 1. General Structure of presented concept

II. State of the Art

An ideal environment, free from external influences as used in most scheduling approaches is normally not given when processing a production plan. Production settings are subject to influences from human and machine failures. Additional resources and materials might not be available in due time and new demands often have to be taken into account on a short-term notice. A comprehensive overview about the execution of production plans under uncertainties is given by Aytug et al. [2]. They develop taxonomy to classify uncertainties, to be able to classify numerous facets of disturbances within operational procedures. These are characterized by four dimensions:

- Cause (e.g., machine failure)
- Context (e.g., materials have not been delivered)
- Effect (postponed starting times)
- Inclusion (reaction upon interruptions, either predictive or reactive) [2]

These aspects illustrate uncertainties within the production planning process. The effect of disturbances and interruptions depends upon the robustness of the scheduling. Schneeweiß [15] gives a basic definition of a robust plan: a plan is robust, when it is insensitive against random environmental influences. Based on this expression one cannot find any quantitative measurements however. Scholl [16] expanded upon this definition. We mainly consider two of the criteria he developed: if a plan is always valid, no matter what environmental influences may effect it, it is called “total validity robust”. One cannot assume to reach this level in practical applications though. Therefore, one is able to analyze the validity robustness in greater detail instead of using a binary value. One could analyze the amount of broken model restrictions or also weight them after their importance. Within production planning, it is especially important to stay within the machine capacities and to adhere to given deadlines. We can consider the objective function of the planning models as the result of a production planning process. Therefore, one can define the criteria of result robustness: a plan is result robust, when its result only differs in a minimal way from the original plan when random environmental influences occur. However, a good result for one scenario may often lead towards a bad result for another scenario. Additionally result and validity robustness conflict with each other: a higher validity often causes higher costs.

Simulations can fulfill two roles within robust production planning: on the one hand, one can use a simulation to simply assess and evaluate the robustness of a plan to confirm the validity of other approaches to create robust production plans. On the other hand they can be used to create robust production plans to include uncertainties.

Aytug et al [2] identified three main approaches in prior literature to create robust production plans: completely reactive procedures, robust scheduling and predictive-reactive scheduling. Completely reactive procedures only take action when disturbances in the production process already occurred. They sort and filter all jobs given to the current machine and continue with the job that appears to be the best based on this evaluation.

Robust scheduling approaches instead are creating plans, which minimize the effects of disturbances within the production procedure. Therefore, a plan for a worst-case scenario is created. Such a plan aims to be able to be processed in many different scenarios without greater difficulties. Both of these approaches share the issue, that available capacities will not be used to their full extend.

A large amount of research happens within the area of predictive-reactive scheduling. First, a plan for the whole planning horizon is created. This plan will be adapted later on. This can happen in a periodic fashion, on the occurrence of new events or in combination of both methods. In practice, these hybrid approaches are mostly used [12], [7].

Simulations are a standard tool to evaluate the robustness of production plans. This can be done based upon different target measures. Honkomp et al. [10] compare a basic deterministic simulation with multiple stochastic replications. To measure the robustness they use metrics that either compute the relation between the average objective function of the stochastic simulations and the deterministic objective function or calculate the standard deviation of the stochastic simulations towards the best deterministic objective function. Apart from cost analysis, Pfeiffer et al. [13] also consider the plan efficiency and stability. This is also done in the overview about rescheduling approaches. Usually one obtains simple efficiency measurements (e.g., delays, backlogging amounts and production times). One can also evaluate these values visually [8]. Plan changes caused by stochastic events are processed to optimize the efficiency values. However, effects of changes within the scheduling are not taken into account within these approaches. Instead of optimizing the efficiency values one might also aim to create plans that only differ minimal from the original plan. A framework to evaluate different techniques to generate robust production plans has been developed by Rasconi et al. [14].

Our work is based on well known optimization models. Typically, two different optimization models are used to create a production plan. Initially we calculate the lot sizes using a Multilevel Capacitated Lotsizing Problems (MLCLSP) based upon macro periods. Subsequently one creates a plan based upon micro periods using a Discrete Lot-Sizing and Scheduling Problem (DLSP) to determine exact production timings. As a result, the order is decided, in which the machines process their corresponding lots. We use a MLCLSP based on the formulation of Tempelmeier and Helber [9] and a DLSP based on the formulation of Drexel and Kimms [24]. These models can be combined in a hierarchical manner, relying on the formal description of hierarchical problems of Schneeweiß [35]. Several efforts have been made to include uncertain events into the formulation of these models. Additionally, special solution methods have been applied to these formulations. For example Gupta and Maranas as well as Bakir and Byrne propose a two staged model to include demand uncertainties [25][26]. Demand uncertainties are also analyzed by Chen and Chu [27]. They do however propose an adapted Lagrangian Relaxation approach to solve this problem. While demand uncertainty is a problem mostly approached in operative planning processes, more uncertainties obviously arise in tactical and strategic planning problems, which are spread about considerably longer timeframes. Three main classes of uncertainty, which were analyzed in scientific literature, are classified by Tajbakhsh et al. [28]: stochastic lead times, as they were discussed by Dolgui et al. [29] or Gurnaki and Gerchak [30], uncertainties in supply quality as considered by Radhoui et. al. [31] and uncertainties in purchasing prices. Our work however is only considering uncertainties in the operational production execution level.

All aforementioned stochastic optimization approaches to include uncertainties do not take the model life cycle, as described by Forgione [32], into account. Literature and research typically do focus on creating and implementing better and more sophistic solution methods to find better solutions to increasingly complex optimization models in a shorter timeframe. However, they do not consider the maintainability of such models, which is a key part of a models life cycle and by far exceeds its deployment time if economically used [33]. Other researchers, like Sundaram and Wolf also note, that businesses demand optimization models to be adaptable for changes within their companies. This led to the introduction of Enterprise Model management Software, which however has not been used to include uncertain planning processes as of now [34]. Nevertheless, this demand lets us to conclude, that special complex solution methods for advanced stochastic models will not receive a broad acceptance in businesses and therefore are not applicable for real world problems.

III. Production Model

To receive meaningful results we base our work on a close to reality production model with a corresponding complexity. Leaned upon a company in the supply industry of average size the model contains 21 machines with a general production structure, meaning that converging, diverging and linear substructures appear. Some of the 44 products can be produced on several machines in a parallel matter. This may possibly lead to different production and setup times as well as costs. 11 products with external consumer demands exist in total. We can classify the model into 13 different levels, which could be used to decompose the problem. Based on this assumption, a high degree of freedom exists, when a concrete production plan shall be created. Figure 2 shows the overall machine plan and material flow of the production model. We create a lot-sizing and scheduling by the combined usage of the Models MLCLSP and DLSP.

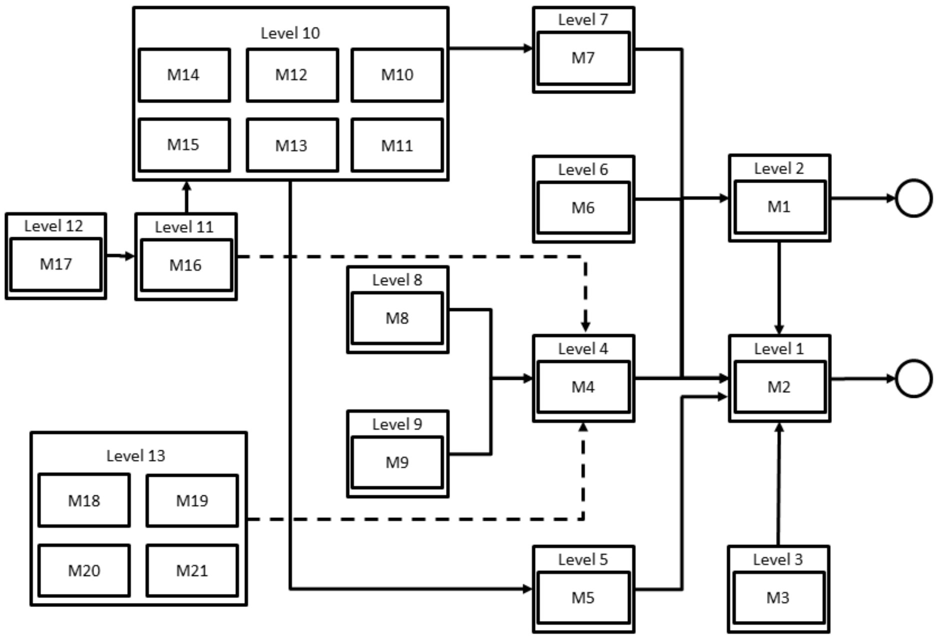


Figure 2. Machine Plan

A. *Deterministic Lotsizing Problem*

To determine the production amounts for each given period we use a MLCLSP in this paper. The basic version of the MLCLSP, as described by Tempelmeier and Helber [9] develops a cost optimal multi-period production plan based on given demands, production costs, setup costs, inventory costs and machine capacities. For this purpose the optimization problem tries to take advantage of possible synergy effects that occur when production lots for several demands are combined, creating less need for setup processes. In contrast, this might create capital commitment and inventory costs when products are created in an earlier period. Therefore, a compromise between these factors has to be found. The model considers machine capacities in particular. Each machine can only be operated for a limited amount of time per period, for example for one or several working shifts. This does force an inventory increase.

The MLCLSP is a model based on macro periods, meaning multiple actions can be done within one time period. Therefore it only determines which amounts of, which products are produced on, which machine in every given period. The model explicitly does not determine a lot scheduling. To reproduce dependencies between different products lead times are used. If a product needs another product from an earlier production level as an input, it has to be produced in an earlier period. A production of intermediate products is triggered whenever a final product is created. A bill of materials is used to determine the needed amounts.

The MLCLSP we are using contains several enhancements over the basic models used in most literature. Several additional constraints are used to comply with the complexity of real production planning. Additionally to the standard model, we allow backlogging for products that have a direct external demand. Backlogged demands do however create extraordinarily high penalty costs, as the consequences of unsatisfied demands can be as severe as the loss of a customer. Products can be manufactured on several machines in a parallel matter. We include transport lots, forcing the production of certain parts in given batch sizes, and the machine capacities are determined upon a flexible work shift model. Late and Night shifts cause additional personal costs. Also, productions on Saturdays and Sundays lead to increasing costs. We do this to reflect the increased worker salaries at this time.

The mathematical formulation of the used model is as follows:

Model MLCLSP:

Minimize $O = \sum_{j=1}^{J}\sum_{k=1}^{K}\sum_{t=1}^{T}(s_{kj} * \gamma_{ktj} + h_k * y_{kt} + p_{kt} * q_{ktj} + bo_{kt} * i_k + b_{jt} * pc_{jt})$

Under the Constraints:

$y_{k,t-1} + q_{k,t-1} - \sum_{i \in N_k} a_{ki} * q_{it} - y_{kt} + bo_{k,t+1} - bo_{kt} = d_{kt}$	∀k ∈ K,∀t ∈ T	(4.1.1)
$\sum_{k \in K_j}(tb_{kj} * q_{kt} + tr_{kj} * \gamma_{ktj}) \leq b_{jt}$	∀j ∈ J,∀t ∈ T	(4.1.2)
$q_{kt} - M * \gamma_{kt} \leq 0$	∀k ∈ K,∀t ∈ T	(4.1.3)
$q_{kt}, y_{kt}, c_{ktj} \geq 0$	∀k ∈ K , ∀t ∈ T	(4.1.4)
$\gamma_{ktj} \leq avail_{kj}$	∀k ∈ K , ∀t ∈ T, ∀j ∈ J	(4.1.5)
$\gamma_{ktj}, s_{jt}^0, s_{jt}^1 . s_{jt}^2 \in 0,1$	∀k ∈ K , ∀t ∈ T, ∀j ∈ J	(4.1.6)
$q_{ktj} = c_{ktj} * cont_k$	∀k ∈ K , ∀t ∈ T, ∀j ∈ J	(4.1.7)
$bo_{kt} \leq bomax_k$	∀k ∈ K , ∀t ∈ T	(4.1.8)
$b_{jt} = 0 + s_{jt}^0 * 480 + s_{jt}^1 * 960 + s_{jt}^2 * 1440$	∀j ∈ J,∀t ∈ T	(4.1.9)
$s_{jt}^0 + s_{jt}^1 + s_{jt}^2 = 1$	∀j ∈ J,∀t ∈ T	(4.1.10)
$\sum_{k \in K_j} \omega_{ktj} \leq 1$	∀j ∈ J,∀t ∈ T	(4.1.11)
$\omega_{ktj} \leq \gamma_{k,t-1,j}$	∀k ∈ K , ∀t ∈ T, ∀j ∈ J	(4.1.12)

Variables and constants meanings:

a_{ki}	Direct demand coefficient of products k and i
b_{jt}	Available capacity of resource j in period t
d_{kt}	Primary demand for product k in period t
pc_{jt}	Personal costs for resource j in period t
h_k	Stock expense ratio for product k
i_k	Penalty costs for backlogging of product k
J	Amount of Resources (j= 1,2,…,J)
K_j	Index set of operations performed by resource j
M	Big number
N_k	Index set of followers of product k
p_{kt}	production costs of product k in period t
q_{ktj}	Production amount of product k on resource j in period t
c_{ktj}	Amount of containers of product k processed by resource j in period t
$cont_k$	Container size/Transport lot size for product k
s_{kj}	Setup costs for product k on resource j
T	Length of planning horizon measured in periods (t=1,2,…,T)
tb_{kj}	Production time for product k on resource j
tr_{kj}	Setup time for product k on resource j
y_{kt}	Stock for product k at the end of period t
γ_{ktj}	Binary setup variable for product k on resource j in period t
bo_{kt}	Backlog variable for product k in period t
$bomax_k$	Maximal backlog amount for product k (always 0 for intermediate products)
$s_{jt}^0, s_{jt}^1, s_{jt}^2$	Binary variables used to calculate the amount of used working shifts

In the objective function the sum of setup-, stock-, production-, backlog penalty and personal costs are minimized. The following constraints enforce the creation of a valid production plan, which fulfills external demands in due time whenever possible.

Constraint 4.1.1 creates a balance between external demands on one side and production- stock and backlog amounts as well as secondary demands on the other side. Thus, it is enforced, that demands can be either satisfied by production and inventory amounts, or that appropriate backlog penalty costs are applied. To be sure that intermediate products are assembled before the final product is created, products must be created a day before the secondary demand takes place. This day of lead time is needed, as the MLCLSP does not create an exact scheduling. Machine capacities are taken into account in constraint 4.1.2. It is only possible to perform a limited amount of production and setup activities within a single period. Constraint 4.1.3 ensures that one can only produce a product on a machine when a machine was set up for that product beforehand.

Additionally, constraint 4.1.5 ensures that machines can only produce products that they can be set up for. Constraint 4.1.7 expresses that production lots always have to be a multiple of transport lots. Within constraint 4.1.8, maximum backlog amounts for each product are defined. This way we can ensure that demands for intermediate products cannot be backlogged. The constraint 4.1.9 and 4.1.10 determine the amount of working shifts used for a machine in a certain period. Constraints 4.1.11 and 4.1.12 are used to allow for a setup carry-over between different periods. This way, a machine only needs to be set up once, when a product is produced in several consecutive periods. The other constraints are used to design meaningful bounds to the variables, for example, stock amounts always have to have a positive value.

B. Deterministic Scheduling Problem

Using a DLSP one can assess a plan based upon micro periods to determine exact production timings. The solutions of the MLCLSP can be used as parameters for the DLSP. This way one can create a complete machine scheduling plan. A basic version of the DLSP can be found at Fleischmann [11] or Drexel and Kimms [24]. The production amounts within a period that have been determined using the MLCLSP can be used as external demands for the DLSP. Periods within the DLSP are chosen as the smallest meaningful unit, for example the smallest common denominator of setup- and production times. The MLCLSP includes lead times; therefore, it is not needed to take dependencies between production levels into account. Hence, we can solve the DLSP for each machine individually. This means that the solution times are rather short. We adapted the basic DLSP formulation to use a similar notation as our MLCLSP model. We did not include additional enhancements into our DLSP, as most major decisions were already done at the level of the MLCLSP. The mathematical formulation of the DLSP is as follows:

Model DLSP:

Minimize O = $\sum_{k=1}^{K} \sum_{t=1}^{T} s_k * \gamma_{kt} + h_k * y_{kt}$

Under the Constraints:

$y_{k,t-1} + q_{kt} - y_{kt} = d_{kt}$	$\forall k \in K, \forall t \in T$	(4.2.1)
$tb_k * q_{kt} + y_{kt} * tr_k = b_t * \delta_{kt}$	$\forall k \in K, \forall t \in T$	(4.2.2)
$\gamma_{kt} \geq \delta_{kt} - \delta_{k,t-1}$	$\forall k \in K, \forall t \in T$	(4.2.3)
$\sum_{k=1}^{K} \delta_{kt} \leq 1$	$\forall t \in T$	(4.2.4)
$q_{kt} \geq 0$	$\forall k \in K, \forall t \in$	(4.2.5)
$y_{kt} \geq 0$	$\forall k \in K, \forall t \in$	(4.2.6)
$\gamma_{kt}, \delta_{kt} \in (0,1)$	$\forall k \in K, \forall t \in$	(4.2.7)

Variables and constants meanings:

b_t	Available capacity in period t
d_{kt}	Primary demand for product k in period t
h_k	Stock expense ratio for product k
q_{kt}	Production amount of product k in period t
s_{kj}	Setup costs for product k
T	Length of planning horizon measured in periods (t=1,2,…,T)
tb_k	Production time for product k
tr_{kj}	Setup time for product k
y_{kt}	Stock for product k at the end of period t
γ_{kt}	Binary setup variable for a setup process of product k in period t
δ_{kt}	Binary setup variable for a setup state of product k in period t

The objective function is used to minimize the sum of setup and inventory costs. We use the DLSP to solely define a scheduling and not a lot-sizing. Therefore, it is not needed to include production and personal costs within this model.

Constraint 4.2.1 is the inventory equation, expressing that demands have to be fulfilled by production and inventory amounts. As the DLSP is solved on a per machine basis, we do not take secondary demands into account. Additionally, backlog amounts are not represented, as we do not allow backlogging within the scheduling problem. Constraint 4.2.2 ensures that a bucket is either filled by a production process or by a setup process. As the DLSP is following a “All or nothing” principle for micro periods, a period capacity either gets completely used by a process, or is not used at all. In conclusion, a setup-carryover must be possible. Constraint 4.2.3 is included for this purpose. To ensure that only one setup state can occur at any given moment, Constraint 4.2.4 exists. The other constraints are used to design meaningful bounds to the variables. For example, production and inventory amounts always have to be positive.

C. Combined Lotsizing and Scheduling

To create a complete production plan including lot-sizing and scheduling decisions, we use the models MLCLSP and DLSP combined in a hierarchical planning compound. This combination can be based on the common schema for hierarchical planning problems, which was described by Schneeweiß [35]. He considers two different planning problems as shown in Figure 3: on the top level, a planning problem is making decisions, which have to be considered by the bottom level problem. Therefore, the decisions by the top level model create a decreased room for maneuvers for the bottom level model. The models MLCLSP and DLSP can clearly be used as a top level and bottom level problem in this fashion.

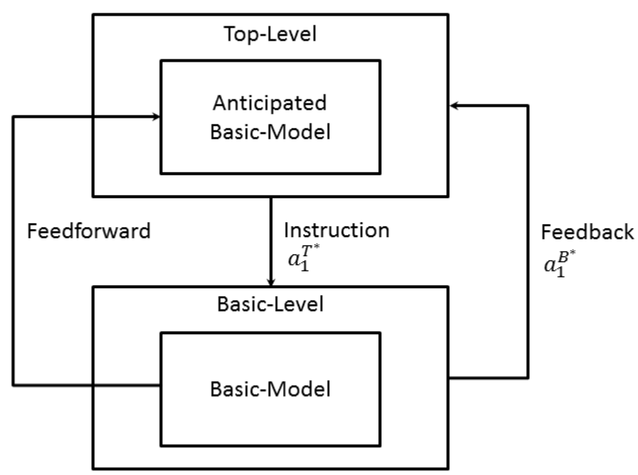


Figure 3. Hierarchical Planning Schema

The Top level tries to anticipate the decision made by the bottom level. Typically a reduced set of information about the bottom level decision problem is used. This amount of information is called "Feed forward". If all information was used and an exact anticipation of decisions of the bottom level was done, one would solve the complete bottom level problem within the top level problem. A distinction of the planning models would thus be pointless. Decisions of the top level problem can form "Instruction", which are given to the bottom level problem and have to be considered by it. This does however mean that the top level problem can make decisions that lead to situations within the bottom-level problem that cannot be solved sufficiently. In this case a manual or automatic Feedback process has to take effect, which causes a new planning process at the top level under an adapted anticipation of the bottom level models decisions.

Applied to our mathematical models, this means that the MLCLSP determines, which products shall be produced at what date. The production and setup times used by the MLCLSP are identical to the times used by the DLSP. Therefore the capacity usage can be exactly anticipated. Both models also only allow for the usage of one setup state at the same moment. Theoretically a lot determined by the MLCLSP could be produced in the form of two different lots. However, setup costs would occur twice in this case – therefore the model DLSP avoids this situation and the amount of setup processes is also correctly anticipated by the MLCLSP. Variable bounds and setup carryover constraints are also identical in function in both models. The MLCLSP therefore anticipates all decisions by the DLSP correctly, and the "Feedback" functionality is not needed when combining these models. The "Instruction" process is needed however. The variable outputs for production amounts in the MLCLSP are transferred to the input demand Parameter for the DLSP. The values can be transferred one by one, but the periods need to be adjusted, as the models operate on different time models. We assume that the demand applies at the last micro period within the DLSP that belongs to the according macro period within the MLCLSP. The amount of micro periods per macro period are determined by the amount of work shifts within the MLCLSP and therefore forms another "Instruction". The different time models also account for the need to separate the models in the first place. Including micro periods within the MLCLSP would vastly increase the decision room and prolong solution times without increasing the solution quality. The correct anticipation of most of the DLSP's decisions however allows us to exclude a lot of the information needed within the MLCLSP, leaving us with a very simple and fast to solve formulation of the DLSP.

IV. Stochastic Optimization Approach

Fuzzy parameters and uncertain information can be reproduced using stochastic methods inside the model classes we described earlier. Ideally, we already know exact probabilities for possible events in advance. Where applicable we can use appropriate prognosis methods to estimate this probabilities. Otherwise, we can only use a normal or similar distribution.

The stochastic optimization tries to find a solution that is the best for all possible combinations of parameters. Finding a solution for these models already is a np-hard problem for sharp levels of information. Finding a solution for a stochastic problem is an extremely time consuming task. Fuzzy parameters might even lead to a state explosion, meaning that an exponentially rising amounts of possible parameter combinations exist. The overwhelming amount of combinations cannot be used to create a valid solution. This situation gets even more complicated, as we use a multi-period, multilevel production structure. A problem in early periods or on a low level can lead to even more problems in later periods or levels. In many situations, one cannot find a solution that is applicable for all possible situations. Therefore, one cannot assume that that it is practical to include uncertainties in the planning process using stochastic optimization methods. Even when such a solution exists, it is unlikely that it can be found within a reasonable amount of time.

A. *Stochastic Optimization Techniques*

In literature and research, several approaches to include uncertainties into an optimization model exist. The simplest version is to create a deterministic substitute value model [36] [37]. The stochastic influences will be replaced by a deterministic value, for example by calculating the mean of the value throughout a finite set of scenarios. The expressiveness of such a model is rather limited, because edge cases with tremendous influences will not be handled at all [20]. This model class still is applicable in practice though, as it can be used to calculate minimum needed buffer times.

A more complex version of stochastic optimization models are fat solution models [38]. In this model, class one tries to find a solution that is applicable to a number of finite scenarios, and creates a solution that is ideal within this constraints [39]. The occurrence of edge cases, for example an extremely prolonged set-up time, can cause huge issues: solutions must always adhere to the worst case scenarios. Therefore, it is hard to find a valid solution for this model type. Finding good solutions is almost impossible within most mid to large size problem classes using this modeling technique.

Most sophisticated stochastic optimization approaches are based on three different methods. Multilevel stochastic models with compensation are based upon Dantzig [6]. Decisions on one level are made at an early point of time and fixed for all following levels. We consider a huge amount of possible events; therefore, we would have to model a corresponding amount of model levels. Stochastic programs with probabilistic constraints date back to Charles and Cooper [4]. Within these models, the breach of constraints is permitted for certain parameter combinations. One can only find proper solutions for this type of models when it is possible to transform the models into an equal deterministic model. Additionally, the expressive value of the model can be reduced due to the loosened constraints.

Bellman [3] introduced stochastic dynamic programming. Based upon a decision tree a backward chaining is used to conclude the ideal choice at the decision situation. All this approaches share the issue that they can only be solved efficiently, if the amount of possible scenarios can be reduced to a certain amount. However, when looking at a real production problem a seemingly infinite amount of decisions is possible.

B. Setup Of Test Cases

To test our assumption, that stochastic optimization models cannot be used to generate good solutions in a timely manner when including uncertainties, we created several simple stochastic optimization models. The production schedule execution is typically affected by unforeseen interruptions und disturbances. In our test models, cycle, and setup times are considered stochastically influenced, due to their high influence on the overall flow shop production process and their deterministic usage in the production-planning model. Material shortages, which arise from supplier unreliableness, are not taken in account and all materials are assumed of as supplied in time. However, scrap parts can also occur during the in-house production. We also include this possibility of uncertain behavior. When scrap parts are produced, the appropriate lot sizes must be expanded to compensate.

The stochastic influences are modeled with two parameters. On the one, hand the likeliness of an increased process time or an occurrence of scrap part production and on the other hand the amount of the deviation. The probability that the planned process time varies, is modeled with a uniform distribution, whereas for the duration a normal distribution is used. Ideally, one is able to use historical data to determine the probabilities for each machine individually; however, this is not possible in a hypothetical model.

First, we created a deterministic substitute value model. The variable amounts are derived from different test sets, and a ceiling modeling is used to keep the integrity of the integer inventory restriction. As expected, the resulting solution times differ only slightly from a purely deterministic optimization model. In comparison to the original plan, slightly more parts where produced in all lots and slightly more working shifts were applied.

Secondly, we created a Fat Solution model. We use this model class within our MLCLSP similarly as before: for each scenario, the setup times, production times and production amounts are altered via stochastic factors. We are unable to find any solutions using this approach. The inventory balance equation cannot be fulfilled, as the production amounts are different in each scenario. Therefore, the inventory and backlog amounts cannot be equal throughout the scenarios.

Lastly, we created a two staged stochastic optimization model with simple compensation. Inequalities of the model are accepted in some edge cases, but penalty costs apply to keep these occurrences low. Indirectly, we already include such penalties in our model, as we allow backlogging. To formulate a two staged model with simple compensation, we must simply allow different inventory and backlog amounts per scenario. Backlogging of a demand is an opportune business decision: sometimes demands simply cannot be fulfilled with the given production capacities. When a demand is backlogged until the end of the planning horizon, one even might consider ordering those parts externally through an alternative supplier or subcontractor. We can make the problem solvable by making the inventory and backlog variables scenario specific. A similar compensation cannot be done for capacity constraints though: while backlogging is an opportune business decision one cannot prolong a working day beyond the 24 hour mark. Therefore, we kept the capacity constraints in the fat solution format.

We ran the optimization process with 10 and 100 scenarios and for the different amount of machines/production levels to determine the problem sizes we can solve using this method. We cancelled the optimization runs if we either found a solution with a gap below 10% or weren't able to find such a solution within 3 hours. Tables 1 and 2 show the results of our test runs.

Obviously, we are unable to find good solutions in medium or big problem sizes, as they appear in practice. Additionally, the high penalty costs for identified solutions make it appear, that these solutions do not suffice for the creation of a sensible plan. Considering this solutions and our argument that we are not able to use much further advanced stochastic optimization methods, because we cannot guarantee their maintainability, leads us to conclude that a different approach is needed to tackle this problem. Therefore, we propose an approach combining a deterministic optimization with a down-streamed simulation.

TABLE 1. SOLUTIONS FOR 10 SCENARIOS

Model-type	Mean Value			Two Stage with Compensation		
# of Levels	Solution Time (m:s)	Solution Value (€)	GAP (%)	Solution Time (m:s)	Solution Value (€)	GAP (%)
1	00:01	809,644	1.76	00:06	850,455	3.97
2	02:01	3,923,725	1.52	03:20	4,847,523	1.05
3	00:43	3,983,490	2.62	04:44	4,910,212	1.31
4	04:11	3,958,611	1.04	06:56	5,528,856	9.93
5	105:09	5,085,347	9.34	180:00*	156,516,563	17.3
6	166:28	5,969,908	9.32	180:00*	143,278,724	87.3
7	180:00*	6,417,375	13.5	180:00*	246,840,729	21.5
8	180:00*	14,445,652	58.7	180:00*	172,457,779	85.7
9	180:00*	14,323,895	25.0	180:00*	196,765,403	81.5
10	180:00*	14,378,925	24.53	180:00*	190,311,820	76.6
11	180:00*	76,451,461	84.8	180:00*	171,081,633	72.4
12	180:00*	113,678,019	89.50	180:00*	119,017,009	54.9
13	180:00*	842,996,612	85.6	180:00*	195,472,374	71.3

TABLE 2. SOLUTIONS FOR 100 SCENARIOS

Model type	Mean Value			Two Stage with Compensation		
# of Levels	Solution Time (m:s)	Solution Value (€)	GAP (%)	Solution Time (m:s)	Solution Value (€)	GAP (%)
1	00:01	843,109	5.65	02:17	866,438	5.75
2	00:22	3,910,988	1.25	180:00*	1,768,362,646	99.72
3	00:37	3,986,675	0.72	180:00*	1,768,362,646	100
4	02:59	3,945,480	2.71	180:00*	1,768,311,953	100
5	04:11	5,085,347	10	180:00*	1,768,235,686	97.6
6	180:00*	6,414,774	14.5	180:00*	1,768,366,685	97.3
7	180:00*	6,307,645	11.0	180:00*	1,768,312,876	97.3
8	180:00*	7,134,935	15.8	180:00*	1,768,313,128	96.9
9	180:00*	15,696,427	28.4	180:00*	1,768,367,156	95.9
10	180:00*	30,413,920	62.3	180:00*	1,768,367,156	95.9
11	180:00*	49,768,107	75.4	180:00*	1,768,367,632	95.4
12	180:00*	273,806,941	95.4	180:00*	1,768,261,632	95.2
13	180:00*	158,961,462	91.9	180:00*	1,768,262,192	95.7

*: No Solution with <10% GAP could be found within a 180 minute timeframe

V. HYBRID APPROACH

To be able to simulate the results of an optimization, the solution data has to be preprocessed in order to prepare the data for the simulation model. CPLEX can export a XML-based file-format, which contains the mathematical programming solution for all variables of the problem. The Converter module reads the file line by line, whereas each line represents a variable. We mainly need two decision variables to be able to simulate the plan: the production variable q_{ktj} determines the products that are produced on a certain machine in a given time period. Additional data like production- and setup times as well as costs can be read from the database based on this production lots. Because a work shift model has been included in the mathematical optimization, every machine can have a different capacity in each period. Therefore, we also have to take the variable b_{jt} into account, which describes these capacities. We sort the lots generated by the MLCLSP in order of the results of the DLSP model to create our initial scheduling. This can be done by analyzing the occurrence of setup process variables γ_{kt}. The real scheduling and date safeguarding will be done within the simulation process. Based upon the given data we are able to calculate all needed information in a deterministic fashion. For example, we are able to calculate the stock or backlog amounts via a difference of production amounts, demands and secondary demands. Thus, we have all information needed to control the simulation procedure. These calculations are also needed to evaluate the simulation results. Therefore, it is a sensible approach to calculate these values for the original plan instead of importing every information from the mathematical model.

A. Simulation

The simulation model is implemented using the discrete event simulator d³FACT developed by our workgroup, Business Computing, esp. CIM. The extensible Java API provides a high-performance, petri-net-based material flow component [1].

The production plan information is first transferred towards the simulation logic. During the initialization of the simulation model, all machines are loading their fixed schedules for the complete planning period. It holds for each machine, which products in what amount have to be produced in each period. Furthermore, it holds the planned durations for the maintenance, production and setup processes. The lot release order is fixed and stays so, even in the case of blocked lots due, to late secondary demands. All released lots are stored in a FIFO-Queue, to be processed in their incoming order. At the beginning of each new period, all planned lots are enqueued and the production cycle starts. Prior to nearly any lot, a setup is intended for rigging the machine. If planned, a routine maintenance of the unit is performed after a given amount of work pieces.

If multiple products or machines demand the same intermediate product, a fork is needed to control the material flow. It stores and routes the tokens as needed towards the point of consumption. The built-in buffer stores the tokens until a machine starts a job and signals its demand. The fork uses a FIFO-Queue to handle the incoming requests and to minimize the mean waiting time for supply. The machine uses a strict FIFO-Queue for lots to dispatch. In this naive version, even a blocked lot with unfulfilled secondary demands waits until its demands are met. If all lots for a period are finished, the shift ends and the next jobs are dispatched in the next period.

Under certain circumstances, it is possible that in case of unmet secondary demands and fully loaded periods, lots are pushed into the following period. In this case, the moved lots are scheduled prior to the regular lots to dispatch the longer waiting jobs first. Because the planning methods calculates with one day lead-time it is easily possible that delayed lots are blocking further following demands.

B. Uncertainties in the production planning process

We include the same kind of uncertainties in our simulation as before. These are the failure of machines in the form of prolonged cycle and setup times and the production of scrap parts.

The cycle and setup times that are incorporated in the formulation of the production-planning problem, are forming the lower bound for the process execution and are modeled in the simulation.

The stochastic influences are also using the same two parameters as before. They determine the number of uncertainties that occur, and the strength of their effects.

C. Rule-based machine control

To be able to improve the production plan within the simulation we are using a rule-based machine control. We are allowing a machine to change its own scheduling plan. As a day of lead-time is included in our planning process, this should not have a negative effect on later production levels. One possible rule that we also implemented appears, when a machine is unable to produce a lot because the secondary demands cannot be met. In this case, the machine logic tries to find other lots for this period, which do not need the missing intermediate products. When such a lot exists, it is processed first while the original planned lot will be processed later. This way, we are able to ensure an even utilization of the given machine capacities. Additionally we reduce the danger of possible backlog amounts. This way we increase the validity robustness of the production plan. Another possible decision rule concerns setup carryovers. If production lots of the same product exist in successive periods, it is sensible to change the scheduling in a way, which allows this product to be produced in the end of the first period and in the beginning of the second period. Therefore, the need to setup the machines for both production lots is not applicable anymore. If one introduces a setup, carryover into the mathematical optimization highly increased solution times may occur. The discussed rule-based mechanisms however only lead towards a small increase in processing time within the simulation process. Additional rules can always be applied in a model specific fashion.

D. Evaluation

The evaluation calculates performance figures for the validity and result robustness. For measuring the validity robustness, we compare the objective value of the simulated plans with the objective value of the original plan from the mathematical optimization. A comparison of single cost values is also possible, like evaluating the influence of capital commitment costs. A plan is considered validity robust, when it does not violate any of the optimization models restrictions. The model we use does allow backlogging however. Backlogging always incurs penalty costs, which also influence the result robustness. However, one cannot assess the influence of delivery dates that could not be met, as it might lead to the loss of a customer in the extreme cases. Therefore, it is sensible to protocol every appearance of backlog amounts.

Important information considers the machine load factors. It can happen that the planned or even the maximum capacity of a machine is not sufficient to produce all lots allocated to it. These events are protocolled and evaluated separately as well. This allows for the search of admissible alternatives.

E. Post-Processing

Within the post-processing component, we are able to use additional simulation external methods to generate an improved production plan with an increased robustness based upon the simulated production plan. An increase of validity however usually creates increased costs. Therefore, we cannot assume that increased validity robustness also correlates with high result robustness.

The simplest way to increase the robustness of a plan is to extend the given capacities where possible. Our model is based on a possible three-shift production. Generally, one tries to avoid using all three shifts to avoid high personal costs during nighttime. By courtesy of the simulation we can however estimate the increase of robustness when considering the introduction of additional shifts. This allows the production planner to decide whether the additional costs are justified or not. One possible way to do this automatically is to calculate the average of the production timings after a higher number of simulations. Afterwards we can determine the average machine load factor and decide upon the amount of needed work shifts.

Another possible way to increase the robustness of the production plan is to move several lots into an earlier period, when this period contains larger capacity reserves. This process is considerably more complicated, as secondary demands also have to be fulfilled in due time. Therefor one cannot simply review available capacities for the final product. One also has to check whether available capacities for the production of all needed intermediate products exist, which often is not the case when the overall machine load factor is constantly high. Additionally an earlier production causes further inventory and capital commitment costs. Thus, this way often is not an opportune choice. In general, it lies in the responsibility of the production planner to decide, which amounts of cost increases he accepts to increase the validity robustness of his production plans. All production plans that are created within the post-processing can be simulated and evaluated again. The production planer consequently can access all information he needs to come to a corresponding decision.

VI. RESULTS

We executed several simulation runs based upon the production plan created by the deterministic mathematical optimization, including all 13 planning levels (c.f. figure 2) and using a planning horizon of 56 periods with a dynamic demand structure. We assumed a failure rate of 10% for each machine. The corresponding processes were prolonged by a standard deviation of 15% and 30%. Table 3 shows several performance indicators in a comparison of simulations with a naïve and rule-based machine control, in particular focusing delays for final products. We calculated the average values of 100 simulation runs, which took less than an hour of time, enabling us to handle real world problem sizes within a timeframe, which is applicable in practice. The rule-based machine controls objective function costs are considerably lower than the costs caused by the naïve machine control. It is noticeable that less final parts get delayed when using the rule-based machine control. Therefore, the ability to supply is increased and lower delay penalty costs occur. These also explain the lower objective cost values.

TABLE 3. COMPARISON BETWEEN DETERMINISTIC (D), RULE-BASED (RB) AND NAIVE (N)MACHINE CONTROL

Standard Deviation	Sim-Type	Objective Function	Delayed Final Products (Absolute)	Delayed Final Products (Relative)	Delay Penalty Costs	Stock Costs
	D	2.769.282,95 €	2.769.282,95 €	2.769.282,95 €	2.769.282,95 €	2.769.282,95 €
15%	RB	3.944.976,12 €	3.944.976,12 €	3.944.976,12 €	3.944.976,12 €	3.944.976,12 €
	N	4.211.949,84 €	4.211.949,84 €	4.211.949,84 €	4.211.949,84 €	4.211.949,84 €
30%	RB	4.355.206,90 €	4.355.206,90 €	4.355.206,90 €	4.355.206,90 €	4.355.206,90 €
	N	4.670.432,31 €	4.670.432,31 €	4.670.432,31 €	4.670.432,31 €	4.670.432,31 €

However, a deterministic simulation of the production plan without stochastic influences shows that no penalty costs occur. The deterministic objective function value is correspondingly low. The rule-based machine control causes an improvement in result robustness as well as validity robustness. Table 4 shows the corresponding evaluation metrics by Honkomp et al. [10].

TABLE 4. METRICS BY HONKOMP

Standard Deviation	Sim-Type	$S.D./OF_{DB}$	$\overline{OF}/OF_{DB}$
15%	Rule-Based	40,00%	1,42
	Naive	40,09%	1,52
30%	Rule-Based	42,90%	1,57
	Naive	44,40%	1,69

The first column represents the relations between the standard deviation of all objective function values of all stochastic simulation runs and the objective function value of the deterministic simulation. A lower value indicates that disturbances and environmental influences have less impact on the ability to supply. The second column represents the relations between the objective function values of stochastic and deterministic simulations. The value shows the cost increase caused by the disturbances and directly shows the result robustness. Normally, a higher robustness is gained by increased costs. However, the inclusion of penalty costs into the objective function value causes lower cost for the more robust plan.

Another reason for increased costs is personal costs. The simulation showed that more working shifts have to be introduced to be able to satisfy customer demands. The original plan was using working shift per day. The resulting plans when using either simulation method mostly used two or three shifts. The rule-based machine control delays 13% less products beyond the planned capacity restrictions, therefore needing less working shifts and causing less personal costs as well. When analyzing the problems within the production process one needs to find out where a possible bottleneck occurs. During the simulation we protocol all occurrences of backlog amounts and the connected machines, products and periods. For further analysis we can determine, which products are delayed most as shown in figure 4.

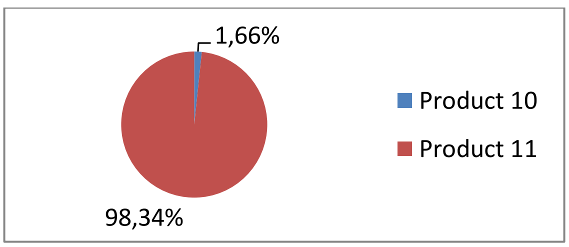


Figure 4. Delayed Final Parts according to products

Surprisingly, most delays are caused by one final part. This is an obvious sign that the production capacity for this part might not be sufficient. Alternatively, production capacities for needed intermediate products might be insufficient. This can be found out by analyzing internal delays for the intermediate products. Table 5 shows the absolute and relative internal delays for both simulation types averaged over 100 simulations. We define internal delays as the amount of intermediate products that couldn't be produced in the planned period.

The usage of the rule-based machine control also shows an improvement when considering the internal demands. Despite not leading to direct revenue losses due to unmet demands, internal delays can cause costs when changes in the production plan have to be made. These costs aren't implicitly included into our production model, but it is in the interest of the production planner to reduce these costs as well. When considering the internal delays per product we are able to find out that product 10 and product 11 are based on the same intermediate product. This product possesses several internal delays, which influence the production of the final products. We were able to find the bottleneck in our production model and can take action to reduce the impact of this issue.

TABLE 5. ANALYSIS: ACCUMULATION OF INTERNAL DELAYS

Standard Deviation	**Sim-Type**	**Internal Delays (Absolute)**	**Internal Delays (Relative)**
15%	Rule-Based	10194,38	1,81%
	Naive	11172,92	1,99%
30%	Rule-Based	16172,68	2,88%
	Naive	17266,10	3,07%

VII. CONCLUSIONS

We have shown in this paper that a material flow simulation can be used to analyze a production plan created in a mathematical optimization and to evaluate its robustness. It is easily possible to read the results of an optimization process, to transfer this data into our simulation framework. We are able to simulate the plan including probabilities for unforeseen events and fuzzy information. The results of the simulations can be used to find possible weak spots in the given plan. In several cases, we might be able to fix these weak spots through automatic post-processing or with manual changes. The effect of these changes can also be evaluated using additional simulation runs. Therefore, a production planner can decide whether he wants to implement these changes or not. Performing a large number of simulations is substantially faster than running another instance of the optimization problem. This especially holds true, if we compare runtime of our hybrid approach to the runtime of stochastic optimization methods. We were unable to find applicable solutions in real world sized problems in an acceptable timeframe using stochastic optimization methods and argued that advanced methods of decomposition for increased problem-solving efficiency are not accepted in practice due to their low maintainability. In the end, we recommend the hybrid optimization and simulation approach for practical and economic usage and expect further improvements to be made.

REFERENCES

[1] C. Laroque, R. Delius, J.-H. Fischer, and D. Horstkemper, "Increase of Robustness on Pre-optimized Production Plans Through Simulation-based Analysis and Evaluation", Proceedings of The Third International Conference on Advances in System Simulation, IARIA Xpert Publishing Services, 2011, pp. 13-20.

[2] C. Almeder, M. Preusser, and F.R. Hartl, "Simulation and optimization of supply chains: alternative or complementary approaches?", OR Spectrum 31, Springer Verlag, 2009, pp.95–119.

[3] M. Aufenanger, W. Dangelmaier, C. Laroque, and N. Rügener, "Knowledge-based Event Control For Flow-Shops Using Simulation and Rules", Proceedings of the 2008 Winter Simulation Conference,

Edited by S. J. Mason, R. R. Hill, O. Rose, T. Jefferson, J. W. Fowler, p 1952-1958, 2008.

[4] H. Aytug, M.A. Lawley, K. McKay, S. Mohan, and R.Uzsoy, “Executing production schedules in the face of uncertainties: A review and some future directions” European Journal of Operational Research 161, 2005, pp 68-110.

[5] R. Bellman, “Dynamic Programming”, Princeton University Press, Princeton, New Jersey, 1957.

[6] R. Bihlmaier, A. Koberstein, and R. Obst, “Modeling and optimizing of strategic and tactical production planning in the automotive industry under uncertainty”, OR Spectrum 31, Springer Verlag, 2009, pp. 311-336.

[7] S. Biswas and Y. Narahari “Object oriented modeling and decision support for supply chains”, Eur J Oper, Res 153, 2004, pp.704–726.

[8] A. Charnes and W.W. Cooper, “Chance-constrained programming”, Management Science 5, 1959, pp. 73-79.

[9] C. S. Chong, A. I. Sivakumar, and R. Gay, “Simulation-based scheduling for dynamic discrete manufacturing.”, Proceedings of the 2003 Winter Simulation Conference, Edited by S. Chick, P.J. Sánchez, D. Ferrin and D. J. Morrice, 2003.

[10] G. B. Dantzig, “Linear Programming under uncertainty”, Management Science 1, 1955, pp. 197-206.

[11] M. Frantzén, A. H. C. Ng, and P. Moore, “Asimulation-based scheduling system for real-time optimization and decision making support”, Robotics and Computer-Integrated Manufacturing, 2011.

[12] F. Ghezail, F. Pierreval, and S. Hajri-Gabouj, “Analysis of robustness in proactive scheduling: A graphical approach”, Computers & Industrial Engineering 58, 2010, pp. 193-198.

[13] S. Helber and H. Tempelmeier, “A heuristic for dynamic multi-item multi-level capacitated lotsizing for general product structures”, European Journal of Operational Research 75, North-Holland,1994, pp. 296-311.

[14] S. Honkomp, J. L. Mockus, and G. V. Reklaitis, “A framework for schedule evaluation with processing uncertainty“, Computers and Chemical Engineering 23, 1999, pp. 595-609.

[15] B. Fleischmann, “The discrete lot-sizing and scheduling problem”, European Journal of Operational Research 44, North-Holland, 1990, pp. 337-348.

[16] J. Maes, and L. Van Wassenhove, “Multi-Item Single-Level Capacitated Dynamic Lot-Sizing Heuristics - A General Review”, The Journal of the Operational Research Society, Vol. 39, No. 11 (1988), pp. 991-1004.

[17] A. Peiffer, B. Kádár, L. Monostori, and D. Karnok, “Simulation as one of the core technologies for digital enterprises: assessment of hybrid rescheduling methods”, International Journal of Computer Integrated Manufacturing, 21:2, 2007, pp. 206 – 214.

[18] A. Peiffer, B. Kádár, and L. Monostori, “Stability-oriented evaluation of rescheduling strategies, by using simulation“,Computers in Industry 58, 2007, pp. 630-643.

[19] R. Rasconi, A. Cesta, N. Policella, “Validating scheduling approaches against executional uncertainty”, Journal of Intelligent Manufacturing (2010), 2008, pp. 49-64.

[20] C. Schneeweiß, “Planung 2: Konzepte der Prozeß- und Modellgestaltung“, Springer, 1992.

[21] A. Scholl, „Robuste Planung und Optimierung“, Physica-Verlags, 2004.

[22] ILOG CPLEX: High-Performance Software for Mathematical Programming and Optimization, http://www.ilog.com/products/cplex, 20.12.2012

[23] F. You and I. Grossmann, “Multicut benders’ decomposition algorithm for supply chain planning under uncertainty”, Annals of Operations Research 10479, 2011.

[24] W. Dangelmaier, D. Brüggemann, B. Klöpper, and T. Rust, „OOPUS WEB - Eine flexible Plattform für die Implementierung von PPS-Tools erläutert an der Anbindung einer CSLP-orientierten Belegungsplanung in der Serienfertigung“, Beiträge zu einer Theorie der Logistik, Springer, 2008, pp. 323-347.

[25] A. Drexl and A. Kimms, “Lot sizing and scheduling - Survey and extensions”, European Journal of Operations Research, 1999, pp. 221-235.

[26] A. Gupta and C. Maranas, “Managing demand uncertainty in supply chain planning”, Computers and Chemical Engineering 27, 2003, pp. 1219–1227.

[27] M.D. Byrne and M. A. Bakir, “Stochastic linear optimisation of an mpmp production planning model”, International Journal of Production Economics 55 (1), 1998, pp. 87 – 96.

[28] H. Chen and C. Chu, “A lagrangian relaxation approach for supply chain planning with order/setup costs and capacity constraints”, Journal of Systems Science and Systems Engineering, Volume 12, Number 1 2003, pp. 98-110.

[29] M. M. Tajbakhsh, S. Zolfaghari, and C.-G. Lee, “Supply Uncertainty and Diversification: A Review”, Trends in Supply Chain Design and Management, Springer Series in Advanced Manufacturing, 2007, Part II., Part 3, p 345-368.

[30] A. Dolgui, F. Grimaud and K. Shchamialiova, “Supply Chain Management Under Uncertainties: Lot-sizing and Scheduling Rules”, Springer Series in Advanced Manufacturing, Springer, 2010, Ch. 7, pp. 181–220.

[31] H. Gurnani and Y. Gerchak, “Coordination in decentralized assembly systems with uncertain component yields”, European Journal of Operational Research 176 (3) ,2007, pp 1559 – 1576.

[32] M. Radhoui, N. Rezg, and A. Chelbi, “Integrated model of preventive maintenance, quality control and buffer sizing for unreliable and imperfect production systems”, International Journal of Production Research 47 (2), 2009, pp. 389–402.

[33] M. Radhoui, N. Rezg, and A. Chelbi, “Integrated model of preventive maintenance, quality control and buffer sizing for unreliable and imperfect production systems”, International Journal of Production Research 47 (2), 2009, pp. 389–402.

[34] A. Bharadwaj, J. Chooineh, A. Lo, and B. Shetty, “Model Management Systems: A Survey”, Annuals of Operations Research(38), 1992, pp. 17-67.

[35] D. Sundaram and E. Wolf, “Enterprise Model Management Systems”, Proceedings of EOMAS '09: Proceedings of the International Workshop on Enterprises & Organizational Modeling and Simulation, New York, 2009.

[36] C. Schneeweiß, „Elemente einer Theorie hierarchischer Planung“, OR-Spektrum 16 (2), 1994, pp. 161-168.

[37] K. Marti, “Stochastic Optimization Methods”, Springer, 2008.

[38] J. Birge and F. Louveaix, “Introduction to Stochastic Programming”, Springer, 1997.

[39] A. Madansky, “Methods of solution of linear programs udner uncertainty”, Operations Research 10 (4), 1962, pp. 463 – 471.

[40] J. Mayer, “On the numerical solution of stochastic optimization problems”, IFIP International Federation for Information processing 199 (1), 2006, pp. 193–206.

Integrating Information Systems and Scientific Computing

Claus-Peter Rückemann
Leibniz Universität Hannover,
Westfälische Wilhelms-Universität Münster (WWU),
North-German Supercomputing Alliance (HLRN), Germany
Email: ruckema@uni-muenster.de

***Abstract*—This paper presents the practical results and challenges from the real-life implementation of interactive complex systems in High End Computing environments. The successful implementation has been made possible using a new concept of higher-level data structures for dynamical applications and configuration of the resources. The discussion shows how an implementation of integrated information systems, compute and storage resources has been achieved. The implementation uses techniques ensuring to create a flexible way for communication with complex information and advanced scientific computing systems. Besides Inter-Process Communication these are mainly Object Envelopes for object handling and Compute Envelopes for computation objects. These algorithms provide means for generic data processing and flexible information exchange. Targeting mission critical environments, the interfaces can embed instruction information, validation and verification methods. The application covers challenges of collaborative implementation, legal, and security issues as well as scientific documentation and classification with these processes. The focus is on integrating information and computing systems, Distributed and High Performance Computing (HPC) resources, for use in natural sciences disciplines and scientific information systems. Implementing higher-level data structure frameworks for dynamical applications and resources configuration has led to scalable and modular integrated public / commercial information system components.**

Keywords–Integrated Systems; Information Systems; Advanced Scientific Computing; Geosciences; High Performance Computing; Computing Systems; Documentation; Classification.

I. Introduction

Based on the implementation of components for Integrated Information and Computing Systems (IICS) and resources, several aspects have been studied [1]. Besides the challenges of complex integrated systems, geoscientific and technical issues, the aspect of multi-disciplinary international projects managing and efficiently operating complex systems has been in the focus.

Regarding the feasibility of dynamical interactive applications and high end resources the goal is definitively a matter of “capability” computing. Whereas the interactive tasks should be considered to represent massively dynamical interactive requirements. Classical capability computing requirements can only be considered of secondary level importance compared to the dynamical interactive challenges of these application scenarios.

The prominent “Information System” components still completely ignore these advanced aspects and abilities for integration and computing. In High Performance Computing (HPC) supercomputers, that means computer systems at the upper performance limit of current technical capabilities for computing, are employed to solve challenging scientific problems. In consequence there is no general or common architecture and configuration for HPC resources as in the lower parts of the performance pyramid. Within the last decades a large number of implementations of information systems, computing and storage systems and other resources have been created. Nearly all of these implementations lack features for extending information systems with the various resources available. Thus, the integration can open advances using larger resources, interactive processing, and reducing time consumption for assigned tasks. Most of these applications and resources are very complex standalone systems and used that way, neglecting, that for many sophisticated use cases conjoint applications are desirable.

This paper is organised as follows. Section II presents motivation, challenges and complexity with the implementation. Section III shows the previous work continued with this research. Section IV summarises the targeted properties and challenges with IICS, distributed information, structuring information as well as unstructured information, large data volume, event multiplicity and context implementation issues. Section V presents two complex application scenarios. Sections VI and VII describe the prerequisites and basic resources configuration for the implementation. Sections VIII and IX show the components and dependencies for integrated systems and resources. Section X discusses the time and response dependence of the integrated solutions. Section XI describes the system implementation, and Section XII presents the evaluation. Sections XIII and XIV summarise the lessons learned, conclusions, and future work.

II. Motivation

With the implementation use cases for Information Systems, the suitability of Distributed and High Performance Computing resources have been studied. These use cases have focus on event triggered communication, dynamical cartography, compute and storage resources. The goal has been, to bring together the features and the experiences for creating and operating a flexible, multi-disciplinary IICS. An example that has been implemented is a spatial information system with hundreds of thousands possible ad-hoc simulations of interest, being used for geomonitoring as well as for geoprocessing. This can be implemented as a map for which any of a large number of Points of Interest (PoI) can be linked with on-demand visualisation and media generation, e.g., on-site videos on location, geology, weather. Within these interactive systems as many "next information of interest" as possible can be dynamically calculated in parallel, near real-time, in order to be of any practical use. These compute intensive tasks can, e.g., be triggered by interactive mouse movement and require to precompute the secondary information for all PoI currently seen on a screen in order to usable in time. In the following passages, we will show environmental components exactly using this implementation for many thousands of PoI. Due to the complexity of IICS, we have applied meta-instructions and signatures for algorithms and interfaces. For these cases, envelopes and Inter-Process Communication (IPC) has been used to provide a unique event and process triggered interface for event, computing, and storage access. The focus with massively interactive applications is naturally on the number of parallel requests. I/O could not been tested within the presented case study scenarios. The interactive handling of 10000 parallel application requests requires on one hand a large number of available system resources and on the other hand latencies as low as possible. For complex systems there are important properties to consider and therefore the components require some overall features, not limited to, but essentially for:

- Structuring,
- Referencing,
- Standardisation,
- Portability,
- Versioning,
- Extendability,
- Certification,
- Meta and object level,
- Ability for distributed application scenarios,
- Connectivity and interfaces.

The following sections discuss all of these aspects in detail and explain how solutions have been created.

III. Previous work

The next generation of systems necessary for providing profound means for communication and computation will have to gather methods evolved by active interdisciplinary interchange, grown with the requirements of the last decades: The information and computing disciplines need means for "in praxi" collaboration from disciplines, structural sciences, computer science, computing science, and information science. Examples are computing intensive interactive environmental monitoring and information systems or simulation supported dynamical geoinformation systems. In this manner, an efficient development and operation can be put into practice for making interactive use of systems with tenthousands of thousands of nodes, ten to hundred thousands of compute cores and hundred thousands to millions of system parts like memory bars and hard disk drives. Methodological sciences means sciences of developing methods for using resources and techniques for gathering new scientific insights. For years, "methodological sciences" or more precise "methodological techniques" have been commonly propagated to solve the problems of High End Computing. It has been commonly experienced that this is not true as there are no applicatory results regarding the essential requirements of complex and integrated high end systems. The available "methods and techniques" is not what we can use for supercomputing where every application and system architecture is different. Unfortunately up to now, this difference is implicit with common tender and operation strategy for so called "efficiency and economical" reasons.

The experiences with integrated systems have been compiled in various projects over the last years [2], [3] and legal issues and object security have been internationally discussed [4], [5]. The architecture of the framework and suitable components used, have been tested by implementing various integrated systems. The following sections show components of an integrated geoscientific information and computing systems developed in one of these case studies that can be used for environmental monitoring or feeding expert systems. None of the participating parties from industry and scientific disciplines can or will create one single application from the components discussed here. The goal is to enable the necessary operation and software stack, nevertheless, the components are modular entities. For the last years practical solutions to various requirements for communication requests have been implemented in a number of projects and case studies using various resources [6], [7], [8], [9], [10]. The most important communication facilities for IICS are:

- Communication requests with applications (example: Inter-Process Communication, IPC).
- Storage object requests (framework example: Object Envelopes, OEN).
- Compute requests (framework example: Compute Envelopes, CEN).

The idea of dynamical event management [11] and envelope-like descriptive containers has been inspired by the good experiences with the concept of Self Contained applications (SFC). Based on these components an integrated solution has been built, for use with local HPC resources supported by distributed information and compute resources. From the point of view of resources providers and integrators of HPC resources it would make very little sense to describe the application components here. Applications details have been published for several components before. For the core issues the conceptional results are by far the most important.

IV. Targeted properties and challenges

The presented approach for Integrated Systems delivers a set of features targeting essential properties regarding all aspects for the components, like with information, algorithms, and dynamical applications:

- Distributed information,
- Documentation and structuring information,
- Unstructured information,
- Large data volume,
- Event multiplicity,
- Resources integration,
- Context,
- Abstraction with context,
- Legal issues,
- Experience and qualification,
- Protection and privacy,
- Complex application scenarios.

The concept and implementation are aimed to provide a feasible long-term perspective and support a holistic level of sustainability.

A. Distributed information

The overall information is widely distributed and it is sometimes very difficult and a long lasting challenge not only to create information but even to get access to a few suitable information sources. The goal for these ambitions is an integrated knowledge base for archaeological geophysics. It will be necessary to collect data from central data centers or registers [12]. From the information point of view, an example for suitable archaeological and geophysical data that has been collected is from the North American Database of Archaeological Geophysics (NADAG) [13] and the Center for Advanced Spatial Technologies (CAST) [14] as well as with the work of the Archaeology Data Service (ADS) [15]. It must be emphasised that there is neither a standard being used for one discipline nor an international standard. All participating disciplines, services, and resources have to be prepared for challenges as big data, critical data, accessibility, longevity, and usability. The concept of this framework is designed to consider these aspects and in order to handle any objects as with the Center of Digital Antiquity [16] and with the Digital Archaeological Record (tDAR) [17], the United States' largest digital store of global archaeological data. In some cases, even concepts for active and smart cities have needed large efforts for collaboration and policies, e.g., with the Rio Operations Center, the public information management center for Rio de Janeiro, Brazil. It integrates and interconnects information from multiple government departments and public agencies in the municipality of Rio de Janeiro in order to improve city safety and responsiveness to various types of incidents, such as flash floods and landslides [18].

These efforts of using information can be supported by integration approaches based on semantics to be used for many purposes, e.g., as a guide [19], [20]. For all components presented, the main information, data, and algorithms are provided by the LX Foundation Scientific Resources [21], containing all the necessary structure and information to support any kind of implementation.

B. Structuring information

The Universal Decimal Classification (UDC) [22] is a hierarchical, multi-lingual and already widely established classification implementing faceted analysis with enumerative scheme features, allowing to build new classes by using relations and grouping. UDC is by definition multi-disciplinary. In multi-disciplinary object context [23] a faceted classification does provide advantages over enumerative concepts. Composition / decomposition and search strategies do benefit from faceted analysis. It is comprehensive and flexible extendable. A classification like UDC is necessarily complex but it has proved to be the only means being able to cope with classifying and referring to any kind of object. It is working with international functional applications, e.g., in applied sciences and medicine. Copies of referred objects can be conserved and it enables searchable relations, e.g., for comparable items regarding special object item tags. The UDC enables to use references like for object sources, may these be metadata, media, realia, dynamical information, citation as via BibTeX sources or Digital Object Identifier (DOI) [24] as well as for static sources. With interactive and dynamical use for interdisciplinary research the referenced objects must be made practically available in a generally accessible, reliable, and persistant way. A DOI-like service with appropriate infrastructure for real life object services, certification, policies and standards in Quality of Service, for reliable long-term availability object, persistency policies should be available. Therefore, for any complex application, these services must be free of costs for application users. It would not be sufficient to build knowledge machines based only on time-limited contracts with participating institutions. These requirements include the infrastructure and operation so data availability for this long-term purpose must not be depending on support from data centers providing the physical data as a "single point of storage".

C. Unstructured information

Unstructured information, the data variety, is one major complexity. For relational databases, a lot of players providing offerings in this space go through the cycle of what the needs are for structured data. As one can imagine, a lot of that work is also starting for unstructured or semi-structured data with Integrated Systems. Data access and transfer for structured data, unstructured data, and semi structured data may be different and may to a certain extend need different solutions for being effective and economic [25].

D. Large data volume

Effective handling of large data volume is increasingly important for geosciences, archaeology, and social networks in any complex context with IICS and High End Computing [5]. The large data challenge is immanent for scientific as well as for social network application scenarios.

Efficient parallel high end access on large volume data from thousands of cores is essentially depending on hardware appliances. These will be I/O appliances, internal networks (like InfiniBand architectures), and large highly parallel filesystems. This is a matter of scalability of hardware and respective software application. Splitting large volume data into many smaller files for highly parallel access can even exacerbate the problem. So with such physical constraints a general overall efficient and cost-effective solution cannot exist today.

There are scientific application scenarios where data volume cannot be reduced and where no data can be filtered and deleted and otherwise there are communication and networking implementations that may make use of filtering and data reduction. So, for complex situations we will certainly need both strategies, advancing resources and technologies development and data reduction. With the original development of dynamical components like Active Source “Kacheln” (tiles) have first been used [11] with data material, mostly based on precomputed static data, supporting the dynamic increase of the number of server resources. This precomputation is an important means in order to provide scalable services. The same aspect of Cloud scalability holds true for communication data required for application scenarios, e.g., with Meta Data Services (MDS).

Regarding High End Computing requirements, sciences can be distinguished in data intensive and compute intensive. Besides to computation with denser grids and the goals to compute for more details, the increased sensor data and mobility have increased data volume drastically, too [26]. Simple and homogeneous infrastructures for computing and I/O are widely considered best practice with most kind of research, even with ambitioned projects. Researchers and people operating technology must be aware that otherwise the infrastructure is likely to get the greatest challenge. Especially for various complex application scenarios large data volumes, compute power, and staff are needed.

Distributed resources, e.g., “Clouds”, “Nebulas”, “Superclouds”, “Skies”, are suitable for a wide number of applications [27], [28], [29] but they do have a major problem with data intensive sciences: data transfer. Not only that with large volumes this is expensive, in many case it is even technologically not feasible in any economic way.

E. Event multiplicity

Data and information velocity is to a certain extend depending from physical constraints. In many cases the end-to-end workflow from gathering data to interactively working with the information has to be considered. Table I shows the number of massively interactive communication events (n_e), number of cores (n_c), and overall response time (t_r) for integrated systems as has been implemented with this research and which is a strong aspect of motivation to use high end resources with interactive communication.

Table I
INTERACTIVE COMMUNICATION, RESOURCES, AND RESPONSE TIME AS BEING THE RESULT OF THE IMPLEMENTATION.

n_e	n_c	t_r (sec) (event multiplicity×event time)
100000	10000	10×5
10	100	1×5
1	1	1×5

This has been implemented and successfully verified for various dynamical applications using parallel compute and storage requests for geoprocessing triggered from a geospatial application. The parallel architecture of the interactive components, software and hardware, allows to precompute objects needed for interactive mode in very short time. Event multiplicity means the number of parallel events in relation to the number of available resources. Related to the event multiplicity, there are various possible cases, for example regarding cores with single core instances. The simplest one-one case, one event – one core, needs 5 seconds for the application to deliver the result. In cases with spare capacity needed, e.g., mission critical buffer capacities, these must be reserved for appropriate usage scenarios. In some cases the spare capacity will be needed per instance.

In High End Computing the multi-user performance is physically depending on hardware. If, with the above single event response time, for 10 events one hundred cores are available, there is spare capacity for several instances. In case of an event overload, which can easily be reached with integrated dynamical systems, e.g., interactive processing events bound to locations in spatial information system interfaces, the response time will get longer. The number of 100000 events is quite a realistic number per application instance. So even with 10000 cores available the interactive use will get limited response times.

F. Resources integration

For seamless operation the required resources have to be integrated into the application scenario. Availability of integrated resources can be provided by policies. The constraints, e.g., with hardware resources, energy efficiency

[30], or sustainable funding are bound to these, so being requirements that cannot be solved with case studies.

With High End Computing and Supercomputing it is the hardware defining the granularity. In general, it is bad for the overall system to splinter into hardware heterogeneity. Heterogeneity at the hardware level can result in multiple granularity as, e.g., with different architectures, separate clusters, various CPU and GPU architectures, concepts of separate physical memory, separate physical networks and so on. Whereas the theoretical overall compute power may be high, what is much more important is that the implementation and porting for these resources is most ineffective. The same holds true for the operation of heterogeneous resources. So in most cases it will be responsible to support homogeneous resources.

Batch and Scheduling has been implemented with many tools available although there is no general best solution because this will depend on the application scenario and resources available, with cluster, Grid, and Cloud [31], [32], [33], [34], [35], [36]. Parallelisation has been implemented both loosely coupled and using application triggering, OpenMP and MPI [37]. For example, components requiring wave analysis can utilise MPI or OpenMP support from additional frameworks and parallelisation support, e.g., IWAVE for the construction of regular grid finite difference and finite element methods for time-dependent partial differential equations [38]. Application performance can be analysed, e.g., with support of integrated tools [39]. A general support of High End Computing resources though is not possible with todays heterogeneous technology [40] so as there is no unifying framework, besides advanced scientific computing applications virtualisation as with the Parallel Virtual Machine (PVM), we have to live with and support several strategies. The appropriate information system resources and frameworks depend on the main focus of the application scenario. For example, in addition to geoscientific and geophysical processing, with environmental and spatial information in international context the GEOSS and GMES [41], [42], [43], [44] have been considered.

G. Context

Context is any information that can be used to characterise the situation of an entity. An entity is a person, place, or object that is considered relevant for the interaction between a user and an application, including the users and applications themselves. For long-term usage, software and hardware context cannot be certified in any suitable way today.

H. Abstraction with context

Future applications in advanced scientific computing need a much higher level of abstraction for communication and parallelisation. For many advanced scientific computing applications virtualisation as with the Parallel Virtual Machine (PVM) system is, instead of often long-winded system-specific rewriting of basic Message Passing use, widely preferred for portable and efficient use of shared and local-memory multiprocessor systems, in case of heterogeneous high end resources. Most applications can be integrated into present applications, using various interfaces, e.g., directly from Active Source or with the tkpvm interface. In addition, very complex applications supporting PVM are available, for example for seismic processing, scientific calculation and visualisation, raytracing and many more fields of advanced scientific computing.

I. Legal issues

In order to find solutions to those issues of Integrated Systems, large data volume, legal aspects of Cloud and Distributed Computing [45], as well as monitoring and forensics are actually discussed [46].

J. Experience and qualification

Experience and expertise are a most important value [47]. In High End Computing expert experience from scientific fields with application scenarios is getting increasingly significant. Non-scientific and low level skills, like parallel programming, operation or administrative experience is drastically loosing importance from the discipline point of view. Accordingly, optimisation is not feasible in most cases. The term that will describe the process best is configuration with complex and scientific applications:

- High level applications available.
- Easy to use frameworks, exploiting parallel computing resources.
- Scientists' working time is more important than optimisation. This is even more prominent if a long verification phase for algorithms and results is the consequence.
- Scientific staff is not funded for low level tasks.

K. Protection and privacy

Securing privacy protection of information for future Information and Computing Systems, e.g., with archaeological sites, is a crucial challenge [4] and requirements even increase with the increasing data volume [5], [48]. So data security has been recognised a key issue for future IICS development and components of collaboration. It needs multi-disciplinary efforts in order to implement a global framework considering privacy in such complex environments.

V. Complex application scenarios

With the implementation of use cases, especially regarding complex Information Systems, the suitability of Distributed and High Performance Computing resources supporting processing and computing of geoscientific data have been studied, e.g., with geosciences, archaeology, and environmental sciences. One of the most prominent examples for a future integration is archaeological geophysics [16], [17], [13], [14], [15].

A. Archaeology and geosciences

Currently basic analysis is being worked on for a principle solution with Archaeological IICS considering the software stack, documentation, structure, classification, and hardware. If so individually available, this should be possible without restructuring complex data all the time when migrating to different architectures and to be prepared for future resources. In case of archaeological, cultural heritage, and geoscientific information and computing systems, there is strong need for integration of different data, information, and sources with scientific computing capabilities, e.g., object information (multi-medial, photographical, textual, properties, sources, references) from natural sciences and humanities. The dynamical system components for this application scenario must enable advanced features like

- weave n-dimensional topics in time,
- use archaeological information in education,
- integrate sketch mapping,
- implement n-dimensional documentation,
- provide support by multi-disciplinary referencing and documentation,
- do discovery planning,
- and structural analysis,
- as well as multi-medial referencing.

In addition, using large data volume resources in an integrated environment like with archaeological geophysics can help to advance the state of the art in integrated systems and accelerates the pace of development for multi-disciplinary solutions and technologies as well as it increases the potential for discoveries and cognitive facilities.

B. Geoprocessing

Geoscientific data processing, "geoprocessing", means processing of geophysical and other data related to the earth or earth-like objects. Today without appropriate background in natural sciences the term is sometimes erroneously reduced to spatial data. Geophysical data processing does have completely different requirements than needed for handling of spatial data and may require a very compute intensive workflow of processes from preprocessing up to postprocessing. In many cases like Seismics or Ground Penetrating Radar although there is compute power available [49] this cannot be done automatically as geological and other information has to be analysed and evaluated. So professional experience is a most important part of the workflow that even cannot be fully automated itself.

VI. System prerequisites

System prerequisites can be considered to be part of the environment, which may be optimised for the application scenario. In almost any case with provisioning of high end resources this cannot be done by user groups. Therefore, the resources providers have been included into the process, as this is a core aspect of the "Collaboration house" framework, which has been developed for this kind of scenarios and which has been presented here (Figure 3).

For implementation and testing a suitable system architecture and hardware had been necessary. A single local system had to fulfill the following minimal criteria:

- Capacity for more than 5000 subjobs per job.
- At least one compute core available per subjob.
- Interactive batch system.
- No distributed compute and storage resources.
- Fast separate InfiniBand networks for compute and I/O.
- Highly performant parallel filesystem.
- Available for being fully configurable.

A system provided being fully configurable means especially configuration of hardware, network, operating system, middleware, scheduling, batch system. At this size this normally involves a time interval of at least three to six months.

It should be obvious that there are not many installations of some reasonable size and complexity that could be provided, configured and operated that way if in parallel to normal operation and production.

The available HPC and distributed resources at ZIV and HLRN [6], [7], [8], [9], [10]. as well as commercially provided High End Computing installations have been sufficient to fulfill all the necessary criteria.

VII. Basic resources configuration

With the systems used for the implementation some operating systems, middlewares, and communication have been available. In almost any case with high end resources in production this cannot be done by user groups. The resources providers have been included into the process.

Elementary operating system components on the resources involved are: AIX, Solaris, and various Linux distributions (SLES, Scientific Linux). Elementary middlewares, protocols, and accounting systems used for the integrated components are: Globus Toolkit, SGAS, DGAS. Unicore, SAGA, SOAP, and many others can be integrated, too. For communication and parallelisation MPI (Open-MPI [37], MPI from SGI, Intel, HP, IBM), OpenMP, MPICH, MVAPICH and other methods have been used along with IPC regarding to the type of operation and optimisation of algorithms needed. For scheduling and batch systems the resources used Moab, Torque, LoadLeveler, and SGE.

All these "tools" are only middleware, protocols, interfaces or isolated applications. They are certainly used on the system resources but they cannot integrate anything, not on the disciplines / application level, not on the services level, not on the resources level. So we want to concentrate on the important high-level issues for the further advanced view of components.

VIII. Components

Using the following concepts, we can, mostly for any system, implement:

- Application communication via IPC.
- Application triggering on events.
- Storage object requests based on envelopes.
- Compute requests based on envelopes.

For demonstration and studies flexible and open Active Source Information System components have been used for maximum transparency. This allows OO-support (object, element) on application level as well as multi-system support. Listing 1 shows a simple example for application communication with framework-internal and external applications (Inter-Process Communication, IPC).

```
catch {
 send {rasmol #1} "$what"
}
```

Listing 1. Application communication (IPC).

This is self-descriptive Tcl syntax. In this case, the IPC send is starting a molecular graphics visualisation tool and catching messages for further analysis by the components.

Listing 2 shows an example of how the communication triggering can be linked to application components.

```
text 450.0 535.0 -tags {itemtext relictrotatex} -fill yellow -text "Rotate␣x" -justify center
...
$w bind relictrotatex <Button-1> {sendAllRasMol {rotate x 10}}
$w bind relictballsandsticks <Button-1> {sendAllRasMol {spacefill 100}}
$w bind relictwhitebg <Button-1> {sendAllRasMol {set background white}}
$w bind relictzoom100 <Button-1> {sendAllRasMol {zoom 100}}
```

Listing 2. Application component triggering.

Tcl language objects like text carry tag names (relictrotatex) and dynamical events like Button events are dynamically assigned and a user defined subroutine sendAllRasMol is executed, triggering parallel visualisation. Storage object requests for distributed resources can be done via OEN. Listing 3 shows a small example of a generic OEN file.

```
<ObjectEnvelope><!-- ObjectEnvelope (OEN)-->
<Object>
<Filename>GIS_Case_Study_20090804.jpg</Filename>
<Md5sum>...</Md5sum>
<Sha1sum>...</Sha1sum>
<DateCreated>2010-08-01:221114</DateCreated>
<DateModified>2010-08-01:222029</DateModified>
<ID>...</ID><CertificateID>...</CertificateID>
<Signature>...</Signature>
<Content><ContentData>...</ContentData></Content>
</Object>
</ObjectEnvelope>
```

Listing 3. Storage object request (OEN).

OEN are containing element structures for handling and embedding data and information, like Filename and Content. An end-user public client application may be implemented via a browser plugin, based on appropriate services. With OEN instructions embedded in envelopes, for example as XML-based element structure representation, content can be handled as content-stream or as content-reference. Algorithms can respect any meta-data for objects and handle different object and file formats while staying transparent and portable. Using the content features the original documents can stay unmodified.

The way this will have to be implemented for different use cases depends on the situation, and in many cases on the size and number of data objects. However the hierarchical structured meta data is uniform and easily parsable. Further it supports signed object elements (Signature), validation and verification via Public Key Infrastructure (PKI) and is usable with sources and binaries like Active Source.

Compute requests for distributed resources are handled via CEN interfaces [50]. Listing 4 shows a generic CEN file with embedded compute instructions.

```
<ComputeEnvelope><!-- ComputeEnvelope (CEN)-->
<Instruction>
<Filename>Processing_Batch_GIS612.pbs</Filename>
<Md5sum>...</Md5sum>
<Sha1sum>...</Sha1sum>
<Sha512sum>...</Sha512sum>
<DateCreated>2010-08-01:201057</DateCreated>
<DateModified>2010-08-01:211804</DateModified>
<ID>...</ID>
<CertificateID>...</CertificateID>
<Signature>...</Signature>
<Content><DataReference>https://doi...</DataReference></Content>
<Script><Pbs>
<Shell>#!/bin/bash</Shell>
<JobName>#PBS -N myjob</JobName>
<Oe>#PBS -j oe</Oe>
<Walltime>#PBS -l walltime=00:10:00</Walltime>
<NodesPpn>#PBS -l nodes=8:ppn=4</NodesPpn>
<Feature>#PBS -l feature=ice</Feature>
<Partition>#PBS -l partition=hannover</Partition>
<Accesspolicy>#PBS -l naccesspolicy=singlejob</Accesspolicy>
<Module>module load mpt</Module>
<Cd>cd $PBS_O_WORKDIR</Cd>
<Np>np=$(cat $PBS_NODEFILE | wc -l)</Np>
<Exec>mpiexec_mpt -np $np ./dyna.out 2>&1</Exec>
</Pbs></Script>
</Instruction>
</ComputeEnvelope>
```

Listing 4. Compute request (CEN).

Content can be handled as content-stream or as content-reference (Content, ContentReference). Compute instruction sets are self-descriptive and can be pre-configured to the local compute environment. In this case, standard PBS batch instructions like walltime and nodes are used. The way this will have to be implemented for different use cases depends on the situation, and in many cases on the size and number of data objects.

An important benefit of content-reference with high performant distributed or multicore resources is that references can be processed in parallel on these architectures. The number of physical parallel resources and the transfer capacities inside the network are limiting factors.

IX. Integrated Systems with Coupled Resources

Figure 1 shows the applied integration of the information and communication systems with coupled computation resources, namely compute resources and storage resources. For integrating the features of information and communication systems with powerful compute resources and storage, it has been necessary to implement interfaces and software applications being able to efficiently use the benefits of High End Computing resources.

Following the results of the long-term case studies [51] three columns namely disciplines (as geosciences), services (as middleware and compute services), resources (computing and storage) had to be figured out for this scenario.

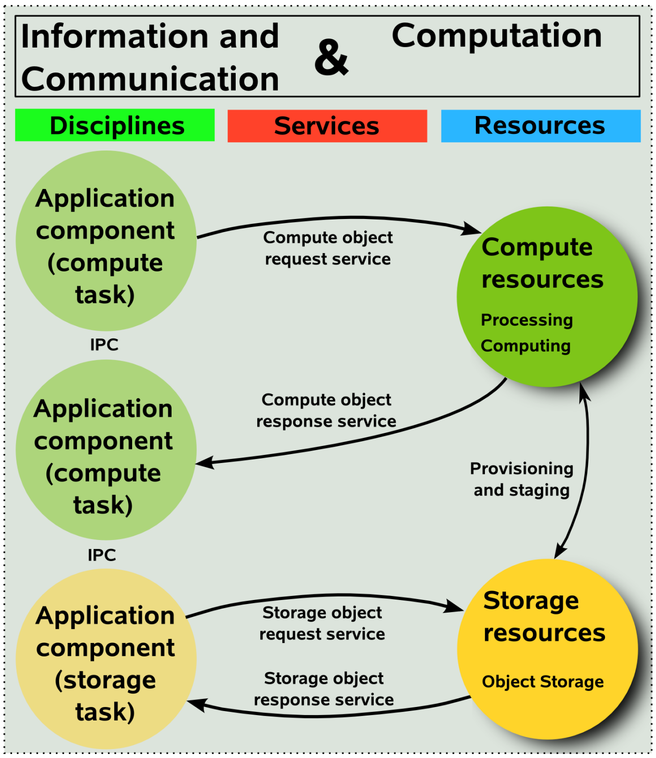


Figure 1. Integrated systems: disciplines integrating resources.

The discipline column shows application components with the state for a compute task and an application component with state for a storage task. Local tasks, ordinary communication between the applications without the need for external computing power, can as usual be done using a local service, for example using Inter-Process Communication (IPC).

Using services, requests can be sent to the configured compute object request service for compute intensive tasks. Results delivered from the computation are delivered for the compute object response service, giving the desired information back the one of the application components. Compute Envelopes (CEN) can be used for exchange of the compute requests. The resources column does provide compute resources for processing and computing as well as storage resources for object storage. Commonly, these resources are separated for backend use with high end applications customised on the compute resources. Application components may trigger storage tasks using a storage object request service. Data objects are handled by the service and delivered to the storage resources. Request for retrieval from the storage are handled by the storage object response service. Object Envelopes (OEN) can be used for exchange of the object requests. For enabling overall scalable integrated systems, mostly for large data volume, the computing and storage resources can communicate for using stored data from within compute tasks and for provisioning and staging of data. These services are so far using a loosely coupled parallel computing, parallelised on the application component level. Each single task can itself contain scalable and loosely to highly parallel computing jobs running on the available compute resources. MPI and OpenMP can be used here. The CEN Envelopes are used to transfer the tasks and their description. The user has to ensure that with using the resources the interactivity and latencies of the integrated system still result in appropriate and usable comprehensive system. Among the compute and storage resources a provisioning and staging mechanism for data and resources requests and responses can be used. Therefore, triggering of computing for storage operations and triggering of storage operations for computing are available.

X. Time Dependence

The same reason why opening large resources for information system purposes is desirable, there is still a dependence on time consumption for interactive and batch processing. Table II shows the characteristic tasks and times that have been considered practical [51] with the current information system applications, for example with environmental monitoring and information system components.

Table II
Tasks and times regarding the overall integrated systems.

Task	*Compute / Storage Times*
In-time events requests	1–3 seconds
Interactive requests	up to 3 minutes
Data processing	1–24 hours
Processing data transfer	n days
Object storage interval	n weeks
Object archive interval	n years

The different tasks afford appropriate policies for interactive and batch use inside the resources network. Besides that, the user and the developer of the application components can use the computing and storage interfaces in order to extend the application facilities using these non-local resources.

Nevertheless, for configuring the system and for implementing new operations the decisions have to be made, which type of implementation would be more suitable.

Interactive request are mostly not acceptable when response times are longer than a few minutes, even for specialised information systems. HPC systems have shown a good performance for parallelisation of interactive subjobs, being in the range of minutes. Whereas distributed resources are much less scalable and provide less performance due to smaller and mostly different resource architecture types and non-exclusive use. Compute times for 1 to 24 hours will force to decide about the field of operation of the system application, when assigning the tasks and events. For example, those compute resources doing computation on large jobs are the computational bottleneck for interactive use. On the other hand, for information system purposes, for example needing visual updates within longer intervals, like for special monitoring purposes for environmental, weather, volcano or catastrophes monitoring and using remote sensing, this scenario is very appropriate. Storage resources and object management can reduce the upload and staging times for objects that can be used more than once. Service providers are confronted with the fact that highly performant storage with reliable and long time interval archiving facilities will be needed at a reasonable price.

XI. Implemented system

The system implemented integrates the component features described from the projects and case studies. Figure 2 shows the implementation of the integrated systems and resources. As Resources Oriented Architectures (ROA) and Services Oriented Architectures (SOA) in themselves are not sufficient for a sustainable long-term development, an important aspect here are the disciplines and their application scenarios with a definable but loosely coupling of services and resources. The components were taken from the GEXI case studies and the well known actmap components [51], [52]. These components handle information like spatial and remote sensing data, can be used for dynamical cartography and trigger events, provide IPC and network communication, and integrate elements from remote compute and storage resources as available with existing compute resources [8], [9], [10]. Processing and computing tasks can for example consist of raytracing, seismic stacking, image transformation and calculation, pattern recognition, database requests, and post processing. The modularisation for development and operation of advanced HPC and application resources and services can improve the multi-disciplinary cooperation. The complexity of operation and usage policies is reduced.

A. Application components

The integrated system is built in three main columns, application components in use with scientific tasks for various disciplines, meaning the conventional scientific desktop and information system environment, services, and resources. These columns are well understood from the "Collaboration house" framework (Figure 3). As well as analysing and separating the essential layers for building complex integrated systems, it is essential that these allow a holistic view on the overall system, for operation, development, and strategies level. The collaboration framework developed [53] and studied for integrated information and computing is the "Collaboration house" [54], [55] (with Grid Computing and GIS initially known as Grid-GIS house and GIS house).

The present framework is considering the respective state of the art resource provisioning for establishing IICS based on multi-disciplinary international collaboration (Figure 3). Implementations in general are complex and can only be handled with a modular architecture being able to separate tasks and responsibilities for resources (HPC, HEC with Grid, Cloud, Sky Computing), services, and disciplines. Thus, the framework supports multiple tenants and groups as well as sole tenants. In opposite to the conventional isolated usage scheme, interaction and communication is not restricted to happen inside the disciplines and resources columns only. Non-isolated usage and provisioning can speed up the development of new components and the modification of existing components in complex environments. The workflows with the application scenarios (Figure 2) are:

a) Application communication.
b) Storage task.
c) Compute task.

These tasks can consists of a request, triggered by some event, and a response, when the resources operation is finished. The response can contain data with the status or not, in case that for example an object has been stored on the resources. Based on this algorithm, task definition can be reasonably portable, transparent, extendable, flexible, and scalable.

B. Application communication

a) Request: The internal and framework-external application is triggered from within the framework components (rasmol is used in the example). From within an actmap component a task to an application component is triggered. IPC calls are used with data and information defined for the event.

Response: A framework-external application is started (rasmol locally on the desktop). The external application can further be triggered from the applications available.

C. Storage task

b) Request: From within an actmap component a storage task is triggered. The stored OEN definition is used to transmit the task to the services. The services do the validation, configuration checks, create the data instructions and initiate the execution of the object request and processing for the resources.

Response: The processing output is transmitted to the services for element creation and the element (in this example a photo image) is integrated into the actmap component.

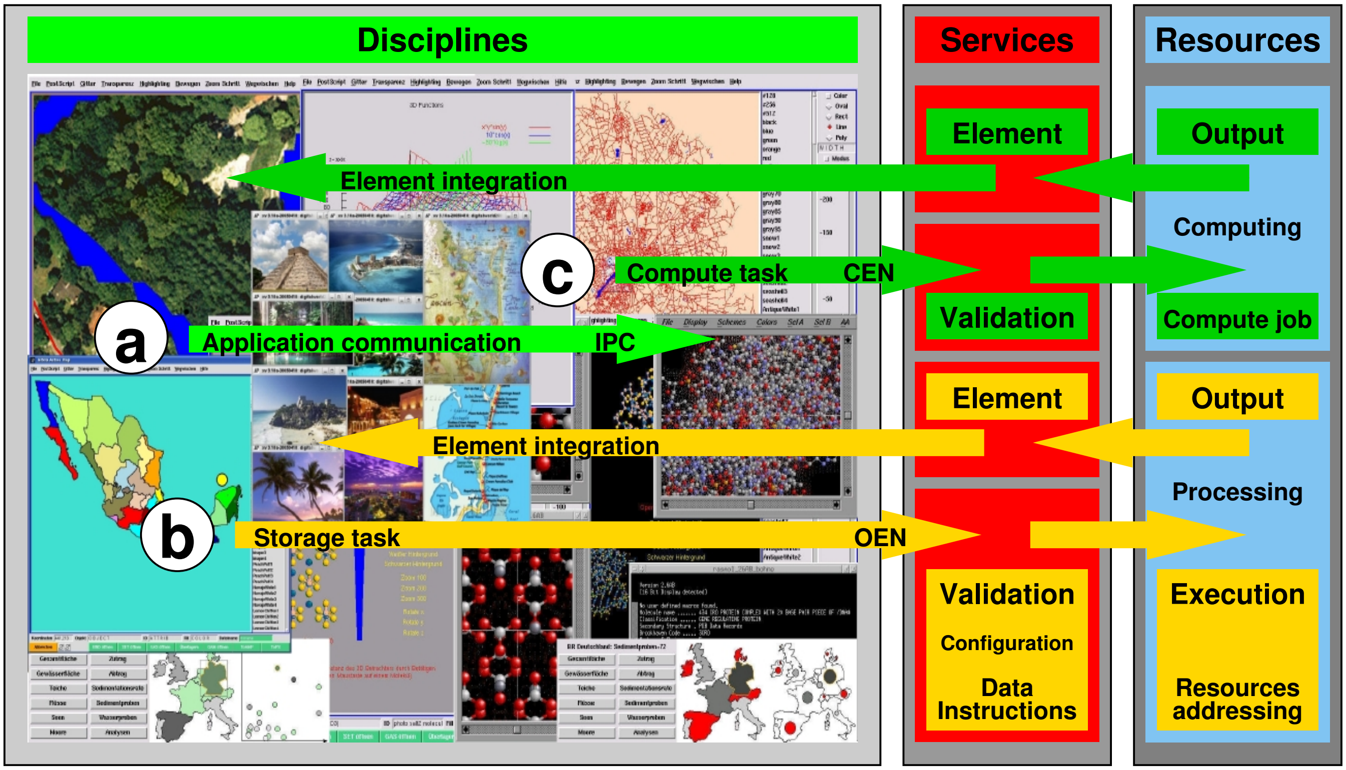


Figure 2. Implementation of the different tasks with integrated systems and resources for advanced scientific computing, utilising the disciplines, services, and resources columns. The disciplines column is triggering the different tasks from the application components (screenshots showing examples). Application communication (local), compute tasks (remote), and storage tasks (remote) are using different resources. If remote high end resources are available these can be used without additional effort. If these are not available then locally available resources can be configured.

D. Compute task

c) Request: From within an actmap component a compute task is triggered. The stored CEN definition is used to transmit the task to the services. The services do the validation (configuration checks, create the compute instructions and initiate the execution of the compute request and compute job for the resources.

Response: The processing output is transmitted to the services for element creation and the element (in this example a remote sensing image and vector object) is integrated into the actmap component.

XII. Evaluation

The target has been to integrate application communication, computing, and storage resources for handling computing requests and content for distributed storage within one system architecture. The technical details of the components have been discussed in several publications and used in applications publically available. The case study has demonstrated that existing information systems and resources can be easily integrated using envelope interfaces in order to achieve a flexible computing and storage access. As the goal has been to demonstrate the principle and for the modular system components used and due to the previous experiences, the services necessary for integration afforded minimal scripting work. Modern information and computing systems object management is a major challenge for software and hardware infrastructure. Resulting from the case studies with information systems and compute resources, signed objects embedded in OEN can provide a flexible solution. The primary benefits shown from the case studies of this implementation are:

- Build a defined interface between dedicated information system components and computing system components.
- Uniform algorithm for using environment components.
- Integration of information and computing systems.
- Speed-up the development of new components and the modification of existing components in complex environments.
- Portable, transparent, extendable, flexible, and scalable.
- Hierarchical structured meta data, easily parsable.
- OO-support (object, element) on application level.
- Multi-system support.
- Support for signed object elements, validation and verification via PKI.
- Usable with sources and binaries like Active Source.
- Portable algorithms in between different object and file formats, respecting meta-data for objects.

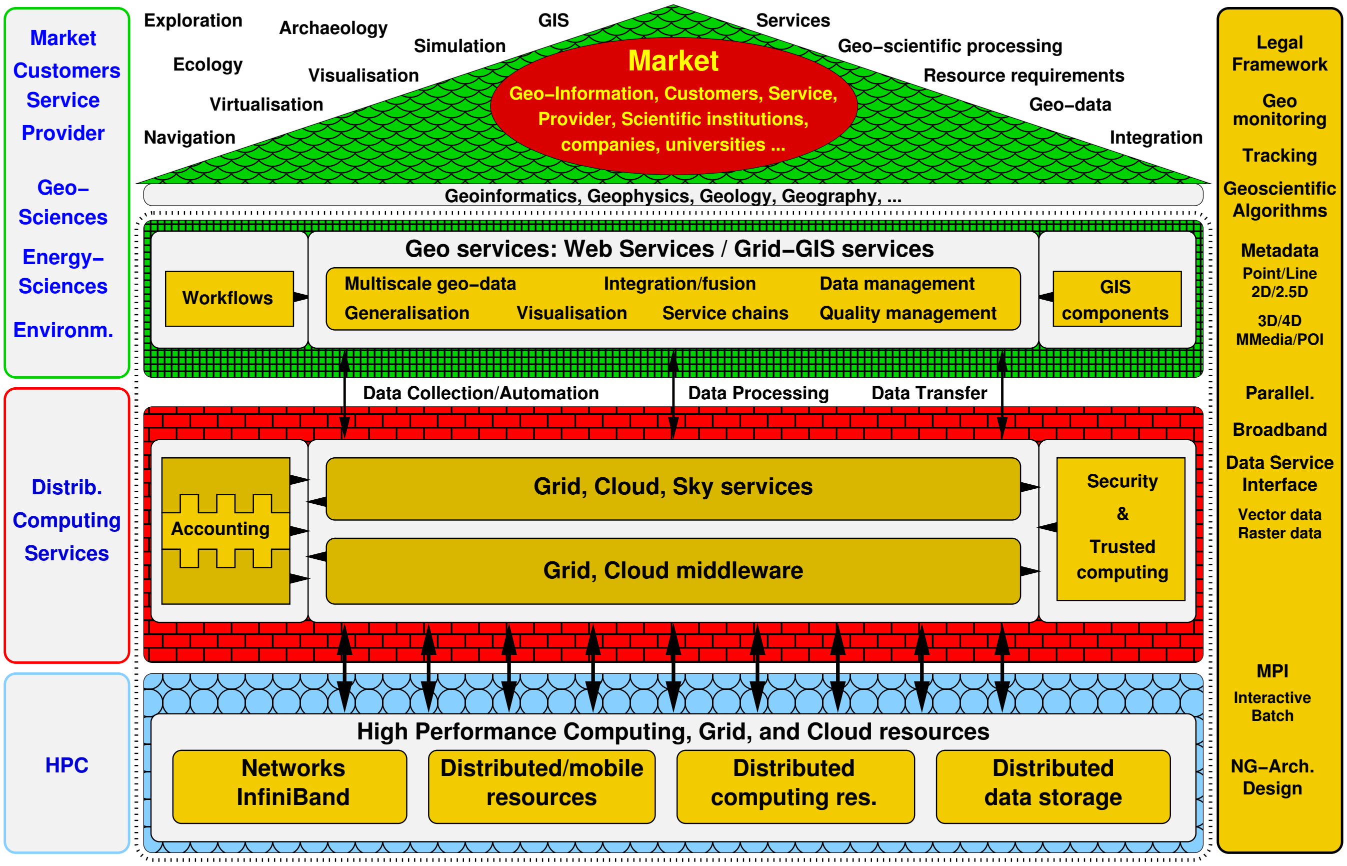


Figure 3. Framework for integrating information systems and scientific computing: This framework has been successfully used with the efficient implementation of all the components presented with this research. Application components have been developed within disciplines (top). System prerequisites and resources configuration have been supported by the resources providers (HPC and distributed computing services) operating the resources, compute and storage. The rightmost column shows examples for the features that have been implemented.

- Original documents and sources can stay unmodified.
- The solution is most transparent, extendable, flexible, and scalable, for security aspects and modularisation.
- Handling of cooperation and operation policies is less complex [56].
- Guaranteed data integrity and authentication derived from the cryptographic strength of current asymmetric algorithms and digital signature processes.
- Flexible meta data association for any object and data type, including check sums and time stamps.

Main drawbacks are:

- Additional complexity due to additional resources and system environment features like batch scripting (Condor [32], Moab / Torque [57]) and using verification/PKI.
- Complexity of parsing and configuration.
- Additional software clients might come handy to handle resources and generate, store and manage associated data and certificates.

The context is an important aspect, though it cannot be called "drawback" here. With closed products, e.g., when memory requirements are not transparent, it is difficult for users to specify their needs. Anyhow, testing is in many cases not the answer in productive environments. Separate measures have to be taken to otherwise minimise possible problems and ease the use of resources in productive operation.

Even in the face of the drawbacks, for information systems making standardised use of large numbers of accesses via the means of interfaces, the envelopes can provide efficient management and access, as programming interfaces can.

XIII. Lessons Learned

The integration of Information Systems and scientific computing has been successfully implemented for various case studies. By using information system components and external resources it has been possible to provide a very flexible and extensible solution for complex application scenarios. OEN and CEN, based on generic envelopes, are very flexible and extensible for creating portable, secure objects handling and processing components with IICS. This way context for complex application scenarios can be addressed regarding legal issues, operation, security, and privacy,

A comprehensive review of the most up-to-date results within the GEXI [51] project has been necessary. The most prominent aspects of dynamical integrated systems regarding information systems and disciplines, resources, and computing have been studied. All these issues of implementation and usage have been implemented considering the respective state of the art [55]. The efficiency of the solution is depending on the environment and system architecture provided, so resources should be accessible and configurable in a most user-friendly way.

Case studies including structuring information as with LX [58], with the use of classification like UDC, distributed information, and unstructured information, including large data volume have show that IICS are scalable to a high extend due to their flexible means of configuration. As an advanced target is to cope with the challenges of structuring and classification of information in complex systems, the LX resources have proved a valuable means providing structured information along with UDC having shown excellent applicability for multi-disciplinary classification of any object to be documented. For multi-disciplinary context as well as for special disciplines, knowledge resources integrated with UDC can provide an excellent means of guidance for using knowledge resources and building manual or automated decision making and processing workflows.

The case studies demonstrated that very different kinds of object data structures and instruction sets may be handled with the envelopes, in embedded or referenced use. Meta data, signatures, check sums, and instruction information can be used and customised in various ways for flexibly implementing information and computing system components. Support for transfer and staging of data in many aspects further depends on system configuration and resources as for example physical bottlenecks cannot be eliminated by any kind of software means.

For future IICS an interface layer between user configuration and system configuration would be very helpful. From system side in the future we need least operation-invasive functioning operating system resources limits, e.g., for memory and a flexible limits management. Homogeneous compute and storage resources and strong standardisation efforts for the implementation could support the use of high end resources regarding economic and efficient operation and use.

On the side of system resources integration, for the next generation of system resources we need I/O thresholds being defined and under control of the operating and management system. Memory management and limits are essential for providing the necessary high availability solutions for the resources providers. Regarding many challenges, service providers try to reduce planning and development requirements for their business, e.g., with unstructured data it is possible to avoid file system limitations by using scale-out Network Attached Storage (NAS) or object storage systems [59] while moving problems to other critical challenges like scalability issues, which tend to be moved to different facilities in the scenario. It is possible to proceed this way for business reasons but research in overall solutions does not carry advantages from this.

Benefits of the developments shown will be a convergence of technologies for integrated intelligent system components, as with Multi-Agent Systems (MAS) [60], advanced criticality management, and an ease of use for the overall ecosystem of use and development, services, and operation.

XIV. Conclusion and future work

It has been demonstrated that IICS can be successfully built from information system and advanced scientific computing components, employing a flexible and portable collaboration and envelopes framework.

All the enlisted challenges have been addressed with the implementation, using the collaboration framework, dynamical components, and structured and classified information. For those cases where it is not possible to provide general solutions for all these issues, the infrastructure concept is provided for building complex efficient systems. In some case like physical constraints, we even have to see advanced technology being developed besides any optimised algorithms.

For this implementation, Object Envelopes, Compute Envelopes, and IPC have been used. The application strongly benefits from an integration of supercomputing resources with the information system components. This is especially important for application scenarios with high event multiplicity. For the dynamical interactive application scenarios there are no implicit requirements for application checkpointing features. Other applications running in batch on a large number of cores for long runtimes should have checkpointing features in order to be candidates for High Performance Computing resources. Mission critical application scenarios need to define the multiplicity and the resources. This can be handled transparently by a dynamical event management as with Active Source.

For the case study Active Source components and Distributed and High Performance Computing resources provided the information system and computing environment. Objects necessary with the information system components have been documented and classified in order to support the automation of processes. With IICS the following main results have been achieved.

Local and inter-application communication can be done using IPC. Object Envelopes can be natively used for handling objects and implementing validation and verification methods for communication. Compute Envelopes can be used in order to define information system computation objects and embed instruction information. These algorithms provide means for generic data processing and flexible information exchange.

It has been shown that the collaboration concept is least invasive to the information system side as well as to the resources used whereby being very modular and scalable. The services in between can hold most of the complexity and standardisation issues and even handle products that are meant to be commercially used or licensed.

In the future, we will have to integrate the features for computing, storage, envelopes, structuring, and classification into a global framework for communication purposes and defining standardised interfaces. UDC will be integrated for classification and reference for objects from multi-disciplinary context for creating Archaeological and Universal IICS [61].

The state of the art implementation has demonstrated a flexible basic approach in order to begin to pave the way and show the next aspects to go on with for future IICS [61] for multi-disciplinary applications.

Acknowledgements

We are grateful to all national and international academic and industry partners in the GEXI cooperations for the innovative constructive work and to the scientific colleagues at the Leibniz Universität Hannover, the Institute for Legal Informatics (IRI), and the Westfälische Wilhelms-Universität (WWU), sharing their experiences working on resources on ZIV, HLRN, and D-Grid and for participating in fruitful case studies as well as the participants of the EULISP Programme for prolific discussion of scientific, legal, and technical aspects over the last years. My thanks go to the committee members and participants of the INFOCOMP conferences in Barcelona and Venice as well as the DigitalWorld conferences for inspiring and fruitful discussions on multi-disciplinary scientific documentation, classification, knowledge discovery, management, and application.

References

[1] C.-P. Rückemann, "Implementation of Integrated Systems and Resources for Information and Computing," in *Proceedings of the International Conference on Advanced Communications and Computation (INFOCOMP 2011), October 23–29, 2011, Barcelona, Spain*. XPS, Xpert Publishing Services, 2011, pp. 1–7, Rückemann, C.-P. and Christmann, W. and Saini, S. and Pankowska, M. (eds.), ISBN-13: 978-1-61208-009-3 (CDROM), ISBN-13: 978-1-61208-161-8, URL: http://www.thinkmind.org/download.php?articleid=infocomp_2011_1_10_10002 [accessed: 2012-05-27].

[2] C.-P. Rückemann, "Using Parallel MultiCore and HPC Systems for Dynamical Visualisation," in *Proceedings, GEOWS 2009, Cancun, Mexico*. IEEE CSP, IEEE Xplore, 2009, pp. 13–18, ISBN: 978-0-7695-3527-2, URL: http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=4782685&isnumber=4782675 [accessed: 2011-02-20].

[3] C.-P. Rückemann, "Envelope Interfaces for Geoscientific Processing with High Performance Computing and Information Systems." XPS, 2011, pp. 23–28, ISBN: 978-1-61208-003-1, URL: http://www.thinkmind.org/download.php?articleid=geoprocessing_2011_2_10_30030 [accessed: 2011-03-20].

[4] *ICDS-GEOProcessing-CYBERLAWS Joint International Panel, Panel on Privacy Invasion and Protection in Digital Society at DigitalWorld 2011, February 26, 2011, Gosier, Guadeloupe, France, The International Conference on Digital Society (ICDS 2011), The International Conference on Advanced Geographic Information Systems, Applications, and Services (GEOProcessing 2011), The International Conference on Technical and Legal Aspects of the e-Society (CYBERLAWS 2011), February 23–28, 2011, Gosier, Guadeloupe, France / DigitalWorld 2011, [Lecture]*, 2011, URL: http://www.iaria.org/conferences2011/filesICDS11/ICDS_2011_PANEL.pdf [accessed: 2012-04-03].

[5] *GEOProcessing International Panel, Panel on Challenges in Handling Large Data Volume for GEO Processing, January 31, 2012, Valencia, Spain International Panel GEOProcessing, The International Conference on Advanced Geographic Information Systems, Applications, and Services (GEOProcessing 2012), Polytechnic University of Valencia, January 30 – February 4, 2012, Valencia, Spain / DigitalWorld 2012, [Lecture]*, 2012, URL: http://www.iaria.org/conferences2012/filesGEOProcessing12/GEO_2012_PANEL.pdf [accessed: 2012-03-04], URL: http://www.iaria.org/conferences2012/ProgramGEOProcessing12.html (Programme) [accessed: 2012-03-04].

[6] "D-Grid, The German Grid Initiative," 2008, URL: http://www.d-grid.de [accessed: 2012-12-09].

[7] ZIVGrid, "ZIV der WWU Münster – ZIVGrid," 2008, URL: http://www.uni-muenster.de/ZIV/Server/ZIVGrid/ [accessed: 2008-12-23].

[8] "ZIVHPC, HPC Computing Resources," 2011, URL: https://www.uni-muenster.de/ZIV/Technik/ZIVHPC/index.html [accessed: 2011-07-10].

[9] "HLRN, North-German Supercomputing Alliance (Norddeutscher Verbund für Hoch- und Höchstleistungsrechnen)," 2011, URL: http://www.hlrn.de [accessed: 2011-07-10].

[10] "ZIVSMP, SMP Computing Resources," 2011, URL: https://www.uni-muenster.de/ZIV/Technik/ZIVHPC/ZIVSMP.html [accessed: 2011-07-10].

[11] C.-P. Rückemann, "Beitrag zur Realisierung portabler Komponenten für Geoinformationssysteme. Ein Konzept zur ereignisgesteuerten und dynamischen Visualisierung und Aufbereitung geowissenschaftlicher Daten," Diss., Westfälische Wilhelms-Universität, Münster, Deutschland, 2001, 161 (xxii+139) S., OPAC, OCLC: 50979238, URL: http://www.user.uni-hannover.de/cpr/x/publ/2001/dissertation/wwwmath.uni-muenster.de/cs/u/ruckema/x/dis/download/dis3acro.pdf [accessed: 2012-01-15], URL: http://wwwmath.uni-muenster.de/cs/u/ruckema/x/dis/download/dis3acro.pdf [accessed: 2012-01-15].

[12] N. P. Service, "National Register of Historic Places Official Website, Part of the National Park Service (NPS)," 2012, NPS, URL: http://www.nps.gov/nr [accessed: 2012-03-18].

[13] "North American Database of Archaeological Geophysics

(NADAG)," 2012, University of Arkansas, URL: http://www.cast.uark.edu/nadag/ [accessed: 2012-04-08].

[14] "Center for Advanced Spatial Technologies (CAST)," 2012, University of Arkansas, URL: http://www.cast.uark.edu/ [accessed: 2012-04-08].

[15] "Archaeology Data Service (ADS)," 2012, URL: http://archaeologydataservice.ac.uk/ [accessed: 2012-04-08].

[16] "Center for Digital Antiquity," 2012, Arizona State University, USA, URL: http://www.digitalantiquity.org/ [accessed: 2012-01-08].

[17] "The Digital Archaeological Record (tDAR)," 2012, URL: http://www.tdar.org [accessed: 2012-01-08].

[18] IBM, "City Government and IBM Close Partnership to Make Rio de Janeiro a Smarter City," *IBM News room - 2010-12-27, USA*, 2012, URL: http://www-03.ibm.com/press/us/en/pressrelease/33303.wss [accessed: 2012-03-18].

[19] D. Cacciagrano, E. Merelli, L. Vito, A. Sergiacomi, and S. Carota, "Semantics on the Cloud: Toward an Ubiquitous Business Intelligence 2.0 ERP Desktop," in *Proceedings of The Sixth International Conference on Advances in Semantic Processing (SEMAPRO 2012), September 23–28, 2012, Barcelona, Spain*. XPS, Xpert Publishing Services, 2012, Cacciagrano, D. R., Dini, P. (eds.), pages 42–47, ISBN-13: 978-1-61208-240-0, URL: http://www.thinkmind.org/download.php?articleid=semapro_2012_2_40_50119 [accessed: 2012-12-01].

[20] E. Bartocci, D. Cacciagrano, M. R. Di Berardini, E. Merelli, and L. Vito, "UBioLab: a web-LABoratory for Ubiquitous in-silico experiments," *Journal of Integrative Bioinformatics*, vol. 9, no. 1, 2012, pages 1–20, ISSN: 1613-4516, DOI: 10.2390/biecoll-jib-2012-192 [accessed: 2012-12-01].

[21] "LX-Project," 2012, URL: http://www.user.uni-hannover.de/cpr/x/rprojs/en/#LX (Information) [accessed: 2012-02-26].

[22] "Universal Decimal Classification Consortium (UDCC)," 2012, URL: http://www.udcc.org (Information) [accessed: 2012-02-19].

[23] C.-P. Rückemann, *Current Research on Multidisciplinary Information and Communication Systems*. Publisher of the University of Economics in Katowice, Poland, 2012, pp. 146–168, in: Dyduch, W. and Pankowska, M. (eds.), Internet Communication Management, International Week, URL: http://www.intweek.ue.katowice.pl/, Monograph, 192 pages, ISBN: 978-83-7246-776-8 (softcover), (Peer Review).

[24] "Digital Object Identifier (DOI) System, The International DOI Foundation (IDF)," 2012, URL: http://www.doi.org (Information) [accessed: 2012-02-19].

[25] N. Hemsoth, "IBM Big Data VP Surveys Landscape," *Datanami*, 2012, March 09, 2012, URL: http://www.datanami.com/datanami/2012-03-09/ibm_big_data_vp_surveys_landscape.html [accessed: 2012-03-18].

[26] N. Hemsoth, "A Floating Solution for Data-Intensive Science," *Datanami*, 2012, March 12, 2012, URL: http://www.datanami.com/datanami/2012-03-12/a_floating_solution_for_data-intensive_science.html [accessed: 2012-03-18].

[27] "Hpc cloud targets oil and gas space," HPC in the Cloud, November 21, 2012, 2012, URL: http://www.hpcinthecloud.com/hpccloud/2012-11-21/hpc_cloud_targets_oil_and_gas_space.html [accessed: 2012-12-01].

[28] "Cyrusone, dell, r systems deliver oil & gas cloud-based solution."

[29] T. Trader, "Gartner Lays Out Five-Part Cloud Strategy," *HPC in the Cloud*, 2012, April 03, 2012, URL: http://www.hpcinthecloud.com/ct/uz5657276Biz13083207 [accessed: 2012-04-08].

[30] EU, "European Code of Conduct for ICT, EU Code of Conduct for Data Centres," 2012, European Commission Joint Research Center, URL: http://re.jrc.ec.europa.eu/energyefficiency/html/standby_initiative_data_centers.htm [accessed: 2012-03-18].

[31] "IBM Tivoli Workload Scheduler LoadLeveler," 2005, URL: http://www-03.ibm.com/systems/software/loadleveler/ [accessed: 2010-10-10].

[32] "Condor, High Throughput Computing," 2011, URL: http://www.cs.wisc.edu/condor/ [accessed: 2010-12-26].

[33] "Sun Grid Engine," 2010, URL: http://gridengine.sunsource.net/ [accessed: 2010-10-10].

[34] "Moab Workload Manager Administrator's Guide," 2010, URL: http://www.clusterresources.com/products/mwm/ [accessed: 2009-11-16], URL: http://www.adaptivecomputing.com/resources/docs/mwm/ [accessed: 2012-12-09].

[35] "Torque Admin Manual," 2010, URL: http://www.clusterresources.com/torquedocs21/ [accessed: 2010-10-10].

[36] T. Trader, "Adaptive Computing Revs Up Moab," *HPC in the Cloud*, 2012, March 20, 2012, URL: http://www.hpcinthecloud.com/hpccloud/2012-03-20/adaptive_computing_revs_up_moab.html [accessed: 2012-03-25].

[37] "Open-MPI," 2011, URL: http://www.open-mpi.org [accessed: 2011-07-12].

[38] "IWAVE," 2012, URL: http://www.trip.caam.rice.edu/software/iwave/doc/html/index.html [accessed: 2012-03-18].

[39] "HPCToolkit," 2012, URL: http://hpctoolkit.org [accessed: 2012-03-18].

[40] M. Wolfe, "The Heterogeneous Programming Jungle," *HPCwire*, 2012, March 19, 2012, URL: http://www.hpcwire.com/hpcwire/2012-03-19/the_heterogeneous_programming_jungle.html [accessed: 2012-03-25].

[41] "Group on Earth Observations (GEO)," 2012, URL: http://www.earthobservations.org [accessed: 2012-03-25].

[42] "Global Earth Observation System of Systems (GEOSS)," 2012, URL: http://www.earthobservations.org/geoss.shtml [accessed: 2012-03-25].

[43] "Shared Environmental Information System (SEIS)," 2012, URL: http://ec.europa.eu/environment/seis/ [accessed: 2012-03-25].

[44] "Global Monitoring for the Environment and Security (GMES)," 2012, URL: http://www.gmes.info [accessed: 2012-03-25].

[45] D. Snead, "Cloud Computing: Legal Traps to Look Out For," *Tutorial, January 30, 2012, Valencia, Spain International Panel GEOProcessing, The International Conference on Advanced Geographic Information Systems,*

Applications, and Services (GEOProcessing 2012), Polytechnic University of Valencia, January 30 – February 4, 2012, Valencia, Spain / DigitalWorld 2012, [Lecture], 2012, URL: http://www.iaria.org/conferences2012/ProgramGEOProcessing12.html [accessed: 2012-02-26].

[46] S. Naqvi, "Digital Investigations and Forensic Analysis - Practices and Technologies," *Tutorial, January 30, 2012, Valencia, Spain International Panel GEOProcessing, The International Conference on Advanced Geographic Information Systems, Applications, and Services (GEOProcessing 2012), Polytechnic University of Valencia, January 30 – February 4, 2012, Valencia, Spain / DigitalWorld 2012, [Lecture]*, 2012, URL: http://www.iaria.org/conferences2012/ProgramGEOProcessing12.html [accessed: 2012-02-26].

[47] D. Malzahn, "The Electronic Silverback, Absorb Knowledge Loss in Industry by Social Network Approach," in *Proceedings International Conference on Information, Process, and Knowledge Management (eKNOW 2012) Polytechnic University of Valencia (UPV), January 30 – February 4, 2012, Valencia, Spain / DigitalWorld 2012*. XPS, Xpert Publishing Services, 2012, pp. 47–52, ISBN 978-1-61208-181-6.

[48] C.-P. Rückemann, "Securing Privacy for Future Information and Computing Systems," *Privacy Invasion and Protection in Digital Society at DigitalWorld 2011, February 26, 2011, ICDS-GEOProcessing-CYBERLAWS Panel, The International Conference on Digital Society (ICDS), The International Conference on Advanced Geographic Information Systems, Applications, and Services (GEOProcessing), The International Conference on Technical and Legal Aspects of the e-Society (CYBERLAWS), February 23–28, 2011, Gosier, Guadeloupe, France*, 2011, URL: http://www.iaria.org/conferences2011/filesICDS11/ICDS_2011_PANEL.pdf [accessed: 2012-03-25].

[49] HPCwire, "HPCwire: GRC and Super Micro Partner on Geophysical Processing Cluster," *HPCwire*, 2012, March 22, 2012, URL: http://www.hpcwire.com/hpcwire/2012-03-22/grc_and_super_micro_partner_on_geophysical_processing_cluster.html [accessed: 2012-03-25].

[50] C.-P. Rückemann and B. Gersbeck-Schierholz, "Object Security and Verification for Integrated Information and Computing Systems," in *Proceedings, CYBERLAWS 2011, Gosier, Guadeloupe, France*. XPS, 2011, pp. 1–6, ISBN: 978-1-61208-003-1, URL: http://www.thinkmind.org/download.php?articleid=cyberlaws_2011_1_10_70008 [accessed: 2012-11-25].

[51] "Geo Exploration and Information (GEXI)," 1996, 1999, 2010, 2012, URL: http://www.user.uni-hannover.de/cpr/x/rprojs/en/index.html#GEXI (Information) [accessed: 2012-12-06].

[52] "Applications with Active Map Software, Screenshots," 2005, URL: http://wwwmath.uni-muenster.de/cs/u/ruckema/x/sciframe/en/screenshots.html [accessed: 2011-02-20].

[53] C.-P. Rückemann, "Legal Issues Regarding Distributed and High Performance Computing in Geosciences and Exploration," in *Proceedings of the Int. Conf. on Digital Society (ICDS 2010), The Int. Conf. on Technical and Legal Aspects of the e-Society (CYBERLAWS 2010), February 10–16, 2010, St. Maarten, Netherlands Antilles*. IEEE Computer Society Press, IEEE Xplore Digital Library, 2010, pp. 339–344, ISBN: 978-0-7695-3953-9, URL: http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=5432414 [accessed: 2011-02-20].

[54] C.-P. Rückemann, "Dynamical Parallel Applications on Distributed and HPC Systems," *International Journal on Advances in Software*, vol. 2, no. 2, 2009, ISSN: 1942-2628, URL: http://www.iariajournals.org/software/ [accessed: 2012-11-25].

[55] C.-P. Rückemann, *Queueing Aspects of Integrated Information and Computing Systems in Geosciences and Natural Sciences*. InTech, 2011, pp. 1–26, Chapter 1, in: Advances in Data, Methods, Models and Their Applications in Geoscience, 336 pages, ISBN-13: 978-953-307-737-6 (hardcover), DOI: 10.5772/29337 [accessed: 2012-12-01], (Peer Review), URL: http://www.intechopen.com/download/pdf/pdfs_id/25377 [accessed: 2012-12-02], URL: http://www.intechopen.com/articles/show/title/queueing-aspects-of-integrated-information-and-computing-systems-in-geosciences-and-natural-sciences [accessed: 2012-12-02] URL: http://dx.doi.org/10.5772/29337 [accessed: 2012-12-02].

[56] "EULISP Lecture Notes, European Legal Informatics Study Programme, Institute for Legal Informatics, Leibniz Universität Hannover (IRI / LUH)," 2011, URL: http://www.eulisp.de [accessed: 2012-11-25].

[57] "Moab Admin Manual, Moab Users Guide," 2011, URL: http://www.clusterresources.com/products/mwm/moabdocs/index.shtml [accessed: 2011-02-20].

[58] C.-P. Rückemann, *Integrated Information and Computing Systems for Advanced Cognition with Natural Sciences*. Premier Reference Source, Information Science Reference, IGI Global, 701 E. Chocolate Avenue, Suite 200, Hershey PA 17033-1240, USA, 2012, pages 1–26, in: Rückemann, C.-P. (ed.), Integrated Information and Computing Systems for Natural, Spatial, and Social Sciences, 543 (xxiv + 519) pages, ISBN: 978-1-4666-2190-9, EISBN: 978-1-4666-2191-6, LCCN: 2012019278, DOI: 10.4018/978-1-4666-2190-9 [accessed: 2012-12-01].

[59] P. Speciale, "Big Unstructured Data and the Case for Optimized Object Storage," *Datanami*, 2012, March 20, 2012, URL: http://www.datanami.com/datanami/2012-03-20/big_unstructured_data_and_the_case_for_optimized_object_storage.html [accessed: 2012-03-25].

[60] U. Inden, S. Tieck, and C.-P. Rückemann, "Rotation-oriented Collaborative Air Traffic Management," in *Proceedings of the International Conference on Advanced Communications and Computation (INFOCOMP 2011), October 23–29, 2011, Barcelona, Spain*. XPS, 2011, pp. 25–30, iSBN-13: 978-1-61208-009-3, URL: http://www.thinkmind.org/download.php?articleid=infocomp_2011_2_10_10089 [accessed: 2012-05-27].

[61] C.-P. Rückemann, "Enabling Dynamical Use of Integrated Systems and Scientific Supercomputing Resources for Archaeological Information Systems," in *Proceedings of the International Conference on Advanced Communications and Computation (INFOCOMP 2012), October 21–26, 2012, Venice, Italy*. XPS, Xpert Publishing Services, 2012, Rückemann, C.-P. and Dini, P. and Hommel, W. and Pankowska, M. and Schubert, L. (eds.), pages 36–41, ISBN-13: 978-1-61208-226-4, URL: http://www.thinkmind.org/download.php?articleid=infocomp_2012_3_10_10012 [accessed: 2012-12-01], URL: http://www.thinkmind.org/index.php?view=article&articleid=infocomp_2012_3_10_10012 [accessed: 2012-12-01].

Recent Development and Applications of SUMO – Simulation of Urban MObility

Daniel Krajzewicz, Jakob Erdmann, Michael Behrisch, and Laura Bieker
Institute of Transportation Systems
German Aerospace Center
Rutherfordstr. 2, 12489 Berlin, Germany
e-mail: Jakob.Erdmann@dlr.de, Daniel.Krajzewicz@dlr.de, Michael.Behrisch@dlr.de, Laura.Bieker@dlr.de

Abstract—**SUMO is an open source traffic simulation package including the simulation application itself as well as supporting tools, mainly for network import and demand modeling. SUMO helps to investigate a large variety of research topics, mainly in the context of traffic management and vehicular communications. We describe the current state of the package, its major applications, both by research topic and by example, as well as future developments and extensions.**

Keywords-microscopic traffic simulation; traffic management; open source; software

I. INTRODUCTION

SUMO ("Simulation of Urban MObility") [1][2] is a microscopic, inter- and multi-modal, space-continuous and time-discrete traffic flow simulation platform. The implementation of SUMO started in 2001, with a first open source release in 2002. There were two reasons for making the work available as open source under the gnu public license (GPL). The first was the wish to support the traffic simulation community with a free tool into which own algorithms can be implemented. Many other open source traffic simulations were available, but being implemented within a student thesis, they got unsupported afterwards. A major drawback – besides reinvention of the wheel – is the almost non-existing comparability of the implemented models or algorithms, and a common simulation platform is assumed to be of benefit here. The second reason for making the simulation open source was the wish to gain support from other institutions.

Within the past ten years, SUMO has evolved into a full featured suite of traffic modeling utilities including a road network importer capable of reading different source formats, demand generation and routing utilities, which use a high variety of input sources (origin destination matrices, traffic counts, etc.), a high performance simulation usable for single junctions as well as whole cities including a "remote control" interface (TraCI, see Section II. D.) to adapt the simulation online and a large number of additional tools and scripts. The major part of the development is undertaken by the Institute of Transportation Systems at the German Aerospace Center (Deutsches Zentrum für Luft- und Raumfahrt, DLR). External parties supported different extensions to the simulation suite.

In this paper, we will survey some of the recent developments and future prospects of SUMO. We start with an overview of the applications in the suite, showing how they help in preparing and performing a traffic simulation. Then, major research topics, which can be addressed using SUMO are presented. We then outline the usage of SUMO within some recent research projects. Finally, we present recent extensions and discuss current development topics.

II. THE SUMO SUITE

SUMO is not only a traffic simulation, but rather a suite of applications, which help to prepare and to perform the simulation of a traffic scenario. As the simulation application "*sumo*", which is included in the suite, uses own formats for road networks and traffic demand, both have to be imported or generated from existing sources of different kind. Having the simulation of large-scale areas as the major application for *sumo* in mind, much effort has been put into the design and implementation of heuristics which determine missing, but needed attributes.

In the following, the applications included in the suite are presented, dividing them by their purpose: network generation, demand generation, and simulation.

A. Road Network Generation

SUMO road networks represent real-world networks as graphs, where nodes are intersections, and roads are represented by edges. Intersections consist of a position, a shape, and right-of-way rules, which may be overwritten by a traffic light. Edges are unidirectional connections between two nodes and contain a fixed number of lanes. A lane contains geometry, the information about vehicle classes allowed on it, and the maximum allowed speed. Therefore, changes in the number of lanes along a road are represented using multiple edges. Such a view on road networks is common; though some other approaches, such as Vissim's [3] network format or the OpenDRIVE [4] format, exist. Besides this basic view on a road network, SUMO road networks include traffic light plans, and connections between lanes across an intersections describing which lanes can be used to reach a subsequent lane.

SUMO road networks can be either generated using an application named "*netgenerate*" or by importing a digital road map using "*netconvert*". *netgenerate* builds three different kinds of abstract road networks: "manhattan"-like grid networks, circular "spider-net" networks, and random networks. Each of the generation algorithms has a set of options, which allow adjusting the network's properties. Figure 1 shows examples of the generated networks.

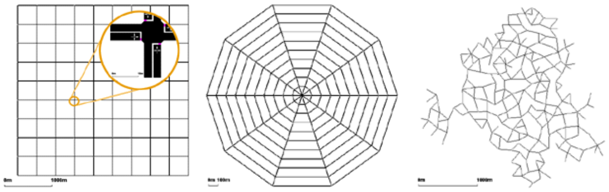


Figure 1. Examples of abstract road networks as built using "netgenerate"; from left to right: grid ("manhattan", spider, and random network)

The road network importer *netconvert* converts networks from other traffic simulators such as VISUM [5], Vissim, or MATSim [6]. It also reads other common digital road network formats, such as shapefiles or OpenStreetMap [7]. Besides these formats, *netconvert* is also capable to read less known formats, such as OpenDRIVE or the RoboCup network format. Figures 2 and 3 show the capabilities to import road networks from OpenStreetMap by example, comparing the original rendering on OpenStreetMap's web pages against SUMO rendering of the imported network.

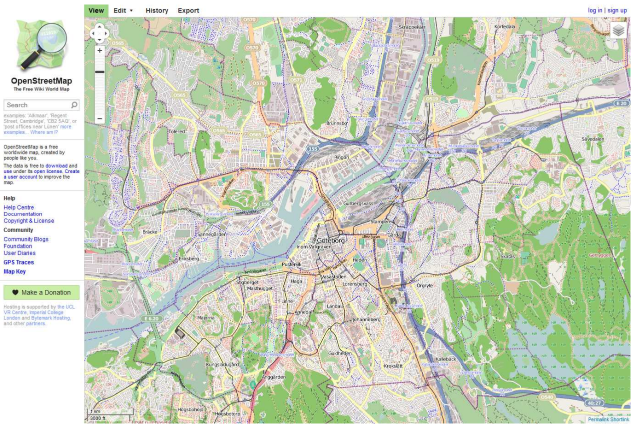


Figure 2. Original OpenStreetMap network of Gothenborg.

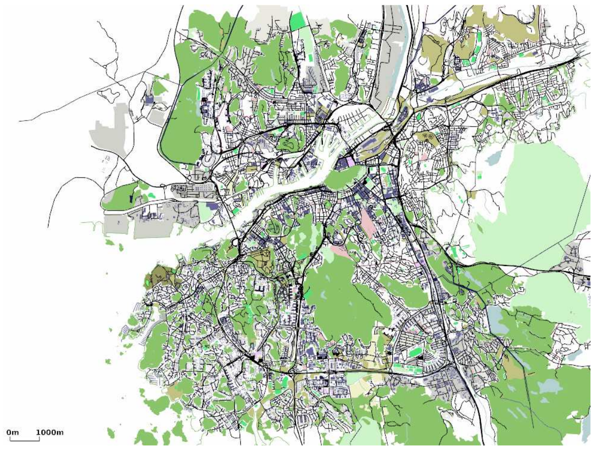


Figure 3. Gothenborg network imported into SUMO.

Additionally, *netconvert* reads a native, SUMO-specific, XML-representation of a road network graph referred to as "plain" XML, which allows the highest degree of control for describing a road network for SUMO. This XML representation is broken into five file types, each for description of nodes, edges, optionally edge types, connections, and (fixed) traffic light plans. Edge types name sets of default edge attributes, which can be referenced by the later loaded edges. Nodes describe the intersections, edges the road segments. Connections describe which lanes incoming into an intersection are connected to which outgoing lanes. The simulation network created by *netconvert* contains heuristically computed values wherever the inputs are incomplete as well as derived values such as the exact geometry at junctions. It is also possible to convert a simulation network back into the "plain" format. Multiple input formats can be loaded at the same time and are automatically merged. Since the "plain" format allows specifying the removal of network elements and the adaption of single edge and lane parameters, it can be used for a wide range of network modifications. To support such modifications SUMO additionally provides the python tool *netdiff.py*, which computes the (human-readable) difference D between two networks A and B. Loading A and D with *netconvert* reproduces B.

Most of the available digital road networks are originally meant to be used for routing (navigation) purposes. As such, they often lack the grade of detail needed by microscopic road traffic simulations: the number of lanes, especially in front of intersections, information about which lanes approach which consecutive ones, traffic light positions and plans, etc., are missing. Sharing the same library for preparing generated/imported road networks, see Figure 4, both, *netgenerate* and *netconvert*, try to determine missing values using heuristics. A coarse overview on this preparation process can be found in [8]. However, most of the algorithms described in [8] have been reworked since its publication. Additional, optional heuristics guess locations of highway on- and off-ramps, roundabouts, traffic lights, etc.

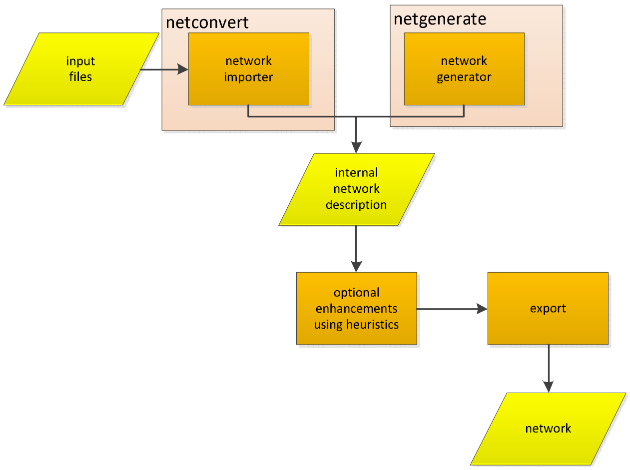


Figure 4. Common network preparation procedure in *netconvert* and *netgenerate*.

Even with the given functionality, it should be stated that preparing a real-world network for a microscopic simulation is still a time-consuming task, as the real-world topology of more complicated intersections often has to be improved manually. A new tool named "*netedit*" allows editing road networks graphically. This is in many cases simpler and faster than preparing XML input files. It also combines the otherwise separate steps of network generation and inspection with *netconvert* and the simulation GUI. *netedit* is not yet available for public use.

B. *Vehicles and Routes*

SUMO is a purely microscopic traffic simulation. Each vehicle is given explicitly, defined at least by a unique identifier, the departure time, and the vehicle's route through the network. By "route" we mean the complete list of connected edges between a vehicle's start and destination. If needed, each vehicle can be described in a finer detail using departure and arrival properties, such as the lane to use, the velocity, or the exact position on an edge. Each vehicle can get a type assigned, which describes the vehicle's physical properties and the variables of the used movement model. Each vehicle can also be assigned to one of the available pollutant or noise emission classes. Additional variables allow the definition of the vehicle's appearance within the simulation's graphical user interface.

A simulation scenario of a large city easily covers one million vehicles and their routes. Even for small areas, it is hardly possible to define the traffic demand manually. The SUMO suite includes some applications, which utilize different sources of information for setting up a demand.

For large-scale scenarios usually so-called "origin/ destination matrices" (O/D matrices) are used. They describe the movement between so-called traffic analysis zones (TAZ) in vehicle numbers per time. For use in SUMO these matrices must be disaggregated into individual vehicle trips with depart times spread across the described time span. Unfortunately, often, a single matrix is given for a single day, which is too imprecise for a microscopic traffic simulation since flows between two TAZ strongly vary over the duration of a day. For example, people are moving into the inner-city centers to get to work in the morning, and leave the inner-city area in the afternoon or evening. Such direction changes cannot be retrieved from an aggregated 24h matrix. Much more useful but only sometimes available are matrices with a scale of 1h. The SUMO suite includes "*od2trips*", an application for converting O/D matrices to single vehicle trips. An hourly load curve can be given as additional input for splitting the daily flows into more realistic hourly slices. Besides disaggregating the matrix, the application also optionally assigns an edge of the road network as depart/arrival position, respectively. The mapping from traffic assignment zones to edges must be supplied as another input.

The resulting trips obtained from *od2trips* consist of a start and an end road together with a departure time. However, the simulation requires the complete list of edges to pass. Such routes are usually calculated by performing a dynamic user assignment (DUA). This is an iterative process employing a routing procedure such as shortest path calculation under different cost functions. Details on the models used in SUMO can be found in Section III.B.

SUMO includes two further route computation applications. The first, "*jtrrouter*", uses definitions of turn percentages at intersection for computing routes through the network. Such an approach can be used to set up the demand within a part of a city's road network consisting of few nodes. The second, "*dfrouter*", computes routes by using information from inductive loop or other cross-section detectors. This approach is quite successful when applied to highway scenarios where the road network does not contain rings and the highway entries and exits are completely covered by detectors. It fails on inner-city networks with rings or if the coverage with detectors is low.

It should be noted, that, while digital representations of real-world road networks became available in good quality in recent years, almost no sources for traffic demand are freely available. Within most of our (DLR's) projects, a road administration authority was responsible for supporting the demand information, either in form of O/D-matrices or at least by supplying traffic counts, which were used to calibrate a model built on rough assumptions.

Two tools enclosed in the SUMO package try to solve this problem by modeling the mobility wishes of a described population. "*SUMO Traffic Modeler*" by Leontios G. Papaleontiou [9] offers a graphical user interface allowing the user to set up demand sources and sinks graphically. "*activitygen*" written by Piotr Woznica and Walter Bamberger from TU Munich has almost the same capabilities, but has no user interface. Both tools are included in the suite and both use own models for creating mobility wishes for an investigated area, requiring different data. They are both under evaluation, currently.

Figure 5 summarizes the possibilities to set up a demand for a traffic simulation using tools included in the SUMO package.

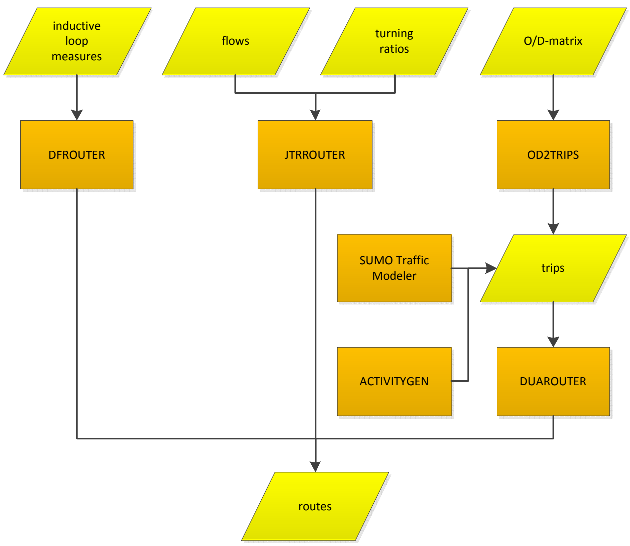


Figure 5. Supported methods for demand generation.

C. Simulation

The application "*sumo*" performs a time-discrete simulation. The default step length is 1s, but may be chosen to be lower, down to 1ms. Internally, time is represented in microseconds, stored as integer values. The maximum duration of a scenario is so bound to 49 days. The simulation model is space-continuous and internally, each vehicle's position is described by the lane the vehicle is on and the distance from the beginning of this lane. When moving through the network, each vehicle's speed is computed using a so-called car-following model. Car-following models usually compute an investigated vehicle's (ego) speed by looking at this vehicle's speed, its distance to the leading vehicle (leader), and the leader's speed. SUMO uses an extension of the stochastic car-following model developed by Stefan Krauß [10] per default. Krauß' model was chosen due to its simplicity and its high execution speed.

The model by Krauß has proved to be valid within a set of performed car-following model comparisons [11][12][13]. Nonetheless, it has some shortcomings, among them its conservative gap size, yielding in a too low gap acceptance during lane changing, and the fact that the model does not scale well when the time step length is changed. To deal with these issues, an application programmer interface (API) for implementing other car-following models was added to *sumo*. Currently, among others, the following models are included: the intelligent driver model (IDM) [14], Kerner's three-phase model [15], and the Wiedemann model [16]. It must be stated, though, that different problems were encountered when using these models in complex road networks, probably due to undefined side-constraints and/or assumptions posed by the simulation framework. For this reason, the usage of different car-following models should be stated to be experimental only, at the current time. Being a traffic flow simulation, there are only limited possibilities to reflect individual driver behavior; it is however possible to give each vehicle its own set of parameters (ranging from vehicle length to model parameters like preferred headway time) and even to let different models run together. The computation of lane changing is done using a model developed during the implementation of SUMO [17].

Two versions of the traffic simulation exist. The application "*sumo*" is a pure command line application for efficient batch simulation. The application "*sumo-gui*" offers a graphical user interface (GUI) rendering the simulation network and vehicles using openGL. The visualization can be customized in many ways, i.e., to visualize speeds, waiting times and to track individual vehicles. Additional graphical elements – points-of-interest (POIs), polygons, and image decals – allow to improve a scenario's visual appearance. The GUI also offers some possibilities to interact with the scenario, e.g. by switching between prepared traffic signal programs, changing reroute following grades, etc. Figure 6 shows a single intersection simulated in *sumo-gui*. *sumo-gui* offers all features the command line version *sumo* supports.

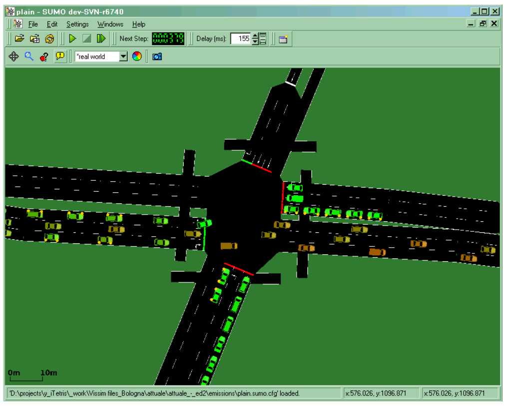


Figure 6. Screenshot of the graphical user interface coloring vehicles by their CO_2 emission.

SUMO allows generating various outputs for each simulation run. These range from simulated inductive loops to single vehicle positions written in each time steps for all vehicles and up to complex values such as information about each vehicle's trip or aggregated measures for all streets and/or lanes. Besides conventional traffic measures, SUMO was extended by a noise emission and a pollutant emission / fuel consumption model, see also Section V.A. All output files generated by SUMO are in XML-format.

D. On-Line Interaction

In 2006, the simulation was extended by the possibility to interact with an external application via a socket connection. This API, called "TraCI" for "Traffic Control Interface" was implemented by Axel Wegener and his colleagues at the University of Lübeck [18], and was made available as a part of SUMO's official release. Within the iTETRIS project, see Section IV.B, this API was reworked, integrating it closer into SUMO's architecture.

To enable on-line interaction, SUMO has to be started with an additional option, which obtains the port number to listen to. After the simulation has been loaded, SUMO starts to listen on this port for an incoming connection. After being connected, the client is responsible for triggering simulation steps in SUMO as well as for closing down the connection what also forces the simulation to quit. The client can access values from almost all simulation artifacts, such as intersections, edges, lanes, traffic lights, inductive loops, and of course vehicles. The client may also change values, for example instantiate a new traffic light program, change a vehicle's velocity or force it to change a lane. This allows complex interaction such as online synchronization of traffic lights or modeling special behavior of individual vehicles.

While DLR uses mainly a client-library written in Python when interacting with the simulation, the client can be written in any programming language as long as TCP sockets are supported. A Python API as well as a freely available Java API [19] are included with SUMO and support for other programming languages may follow.

III. Research Topics

In the following, the major research topics addressed using SUMO are presented. The list is mainly based on observations of published papers which cite SUMO.

A. Vehicular Communication

The probably most popular application for the SUMO suite is modeling traffic within research on V2X – vehicle-to-vehicle and vehicle-to-infrastructure – communication. In this context, SUMO is often used for generating so-called "trace files", which describe the movement of communication nodes by converting the output of a SUMO simulation into a format the used communication simulator can read. Such a post-processing procedure allows feeding a communication simulator with realistic vehicle behavior, but fails on simulating the effects of in-vehicle applications that change the vehicles' behavior. To investigate these effects, a combined simulation of both, traffic and communication is necessary [20]. For such research, SUMO is usually coupled to an external communication simulation, such as ns2 or ns3 [21] using TraCI. For obtaining a functioning environment for the simulation of vehicular communications, a further module that contains the model of the V2X application to simulate is needed. Additionally, synchronization and message exchange mechanisms have to be involved.

TraNS [22] was a very popular middleware for V2X simulation realizing these needs. It was build upon SUMO and ns2. Here, TraNS' extensions to ns2 were responsible for synchronizing the simulators and the application had also to be modeled within ns2. TraNS was the major reason for making TraCI open source. After the end of the projects the original TraNS authors were working on, TraNS was no longer maintained. Since the TraCI API was changed after the last TraNS release, TraNS only works with an outdated version of SUMO.

A modern replacement for TraNS was implemented within the iTETRIS project [23]. The iTETRIS system couples SUMO and ns2's successor ns3. ns3 was chosen because ns2 was found to be unstable when working with a large number of vehicles. Within the iTETRIS system, the "iTETRIS Control System", an application written in c++ is responsible for starting and synchronizing the involved simulators. The V2X applications are modeled as separate, language-agnostic programs. This clear distribution of responsibilities allows to implement own applications conveniently in the user's favorite programming language.

The Veins framework [20] couples SUMO and OMNET++ [24], a further communication simulator. A further, very flexible approach for coupling SUMO with other applications is the VSimRTI middleware developed by Fraunhofer Fokus [25]. Its HLA-inspired architecture not only allows the interaction between SUMO and other communication simulators. It is also able to connect SUMO and Vissim, a commercial traffic simulation package. In [25], a system is described where SUMO was used to model large-scale areas coarsely, while Vissim was used for a fine-grained simulation of traffic intersections.

Many vehicular communication applications target at increasing traffic safety. It should be stated, that up to now, microscopic traffic flow models are not capable of modelling real collisions and thus derive safety-related measures indirectly, for instance by detecting full braking. SUMO's strength lies in simulation of V2X applications that aim at improving traffic efficiency. Additionally, evaluating concepts for forwarding messages to their defined destination ("message routing") can be done using SUMO, see, for example, [26] or [27].

B. Route Choice and Dynamic Navigation

The assignment of proper routes to a complete demand or a subset of vehicles is investigated both, on a theoretical base as well as within the development of new real-world applications. On the theoretical level, the interest lies in a proper modeling of how traffic participants choose a route – a path through the given road network – to their desired destination. As the duration to pass an edge of the road graph highly depends on the numbers of participants using this edge, the computation of routes through the network under load is a crucial step in preparing large-scale traffic simulations. Due to its fast execution speed, SUMO allows to investigate algorithms for this "user assignment" or "traffic assignment" process on a microscopic scale. Usually, such algorithms are investigated using macroscopic traffic flow models, or even using coarser road capacity models, which ignore effects such as dissolving road congestions.

The SUMO suite supports such investigations using the *duarouter* application. Two algorithms for computing a user assignment are implemented, c-logit [28] and Gawron's [29] dynamic user assignment algorithm. Both are iterative and therefore time consuming. Possibilities to reduce the duration to compute an assignment were evaluated and are reported in [30]. A further possibility to reduce the computational effort is given in [31]. Here, vehicles are routed only once, directly by the simulation and the route choice is done based on a continuous adaptation of the edge weights during the simulation.

Practical applications for route choice mechanisms arise with the increasing intelligence of navigation systems. Modern navigation systems as Tom Tom's IQ routes ([32]) use on-line traffic information to support the user with a fastest route through the network regarding the current situation on the streets. One research topic here is to develop new traffic surveillance methods, where vehicular communication is one possibility. With the increased penetration rate of vehicles equipped with a navigation device, further questions arise: what happens if all vehicles get the same information? Will they all use the same route and generate new congestions? These questions are not only relevant for drivers, but also for local authorities as navigation devices may invalidate concepts for keeping certain areas calm by routing vehicles through these areas. SUMO allows addressing these topics, see, e.g., [33].

C. Traffic Light Algorithms

The evaluation of developed traffic light programs or algorithms for making traffic lights adaptable to current traffic situation is one of the main applications for microscopic traffic flow simulations. As SUMO's network model is relatively coarse compared to commercial applications such as Vissim, SUMO is usually not used by traffic engineers for evaluating real-life intersections. Still, SUMO's fast execution time and its open TraCI API for interaction with external applications make it a good candidate for evaluating new traffic control algorithms in abstract scenarios.

The first investigation in traffic lights was performed within the project "OIS" [34] where a traffic light control algorithm, which used queue lengths determined by image processing should have been evaluated. As a real-world deployment of the OIS system was not possible due to legal constraints, the evaluation had to be done using a simulation. The simulation was prepared by implementing a real-world scenario, including real-world traffic light programs. The simulation application itself was extended by a simulated sensor, which allowed retrieving queue lengths in front of the intersection similar to the real image processing system. The traffic light control was also implemented directly into the simulation. At the end, the obtained simulation of OIS-based traffic control was compared against the real-world traffic lights, [34] shows the results.

In ORINOKO, a German project on traffic management, the focus was put on improving the weekly switch plans within the fair trade center area of the city of Nürnberg. Here, the initial and the new algorithm for performing the switch procedure between two programs were implemented and evaluated. Additionally, the best switching times were computed by a brute-force iteration over the complete simulated day and the available switch plans.

By distinguishing different vehicle types, SUMO also allowed to simulate a V2X-based emergency vehicle prioritization at intersections [35]. Other approaches for traffic light control were also investigated and reported by other parties, see, e.g., [36], or [37].

As mentioned before, the first investigations were performed by implementing the traffic light algorithms to evaluate directly into the simulation's core. Over the years, this approach was found to be hard to maintain. Using TraCI seems to be a more sustainable procedure currently.

D. *Evaluation of Traffic Surveillance Systems*

Simulation-based evaluation of surveillance systems mainly targets on predicting whether and to what degree the developed surveillance technology is capable to fulfill the posed needs at an assumed rate of recognized and/or equipped vehicles. Such investigations usually compare the output of the surveillance system, fed with values from the simulation to the according output of the simulation. An example will be given later, in Section IV.A. on the project "TrafficOnline".

A direct evaluation of traffic surveillance systems' hardware, for example image processing of screenshots of the simulated area, is uncommon, as the simulation models of vehicles and the environment are too coarse for being a meaningful input to such systems. Nonetheless, the simulation can be used to compute vehicle trajectories, which can be enhanced to match the inputs needed by the evaluated system afterwards. An example of such an investigation is the evaluation of hyperspectral sensors reported in [38].

Besides evaluating developed surveillance systems, possibilities to incorporate traffic measurements of various kinds into a simulation are evaluated, see for example Section IV.C. on "VABENE".

IV. Recent and Current Projects

SUMO was used in past research projects performed by the DLR and other parties. In the following, some of the recently performed projects are described.

A. *TrafficOnline*

Within the TrafficOnline project, a system for determining travel times using GSM telephony data was designed, implemented, and evaluated. SUMO was used to validate this system's functionality and robustness. In the following, we focus on the simulation's part only, neither describing the TrafficOnline system itself, nor the evaluation results.

The outline for using the simulation was as follows. Real-world scenarios were set up in the simulation. When being executed, the simulation was responsible for writing per-edge travel time information as well as simulated telephony behavior values. The TrafficOnline system itself obtained the latter, only, and computed travel times in the underlying road network. These were then compared to the travel times computed by the simulation. The overall procedure is shown in Figure 7.

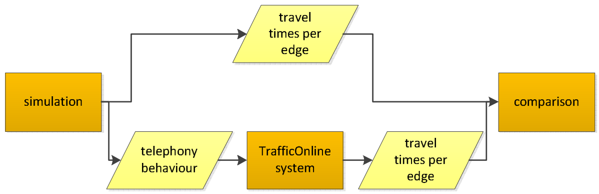


Figure 7. Overall process of TrafficOnline validation.

The evaluation was performed using scenarios located in and around Berlin, Germany, which covered urban and highway situations. The road networks were imported from a NavTeq database. Manual corrections were necessary due to the limits of digital road networks described earlier in Section II.A.

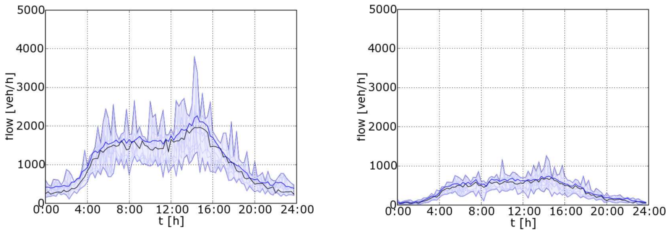


Figure 8. Validation of the traffic flows in TrafficOnline.

Measurements from inductive loops were used for traffic modeling. Figure 8 shows two examples of validating the traffic simulation by comparing simulated (black) and real-world inductive loop measures (blue, where dark blue indicates the average value). For validating the robustness of the TrafficOnline system, scenario variations have been implemented, by adding fast rail train lines running parallel to a highway, or by implementing additional bus lanes, for example. Additionally, scenario variations have been built by scaling the simulated demand by +/- 20%.

For validating the TrafficOnline system, a model of telephony behavior was implemented, first. The telephony model covered the probability to start a call and a started call's duration, both retrieved from real-world data. For an adequate simulation of GSM functionality, the real-world GSM cell topology was put onto the modeled road networks. It should be noted, that dynamic properties of the GSM network, such as cell size variations, or delays on passing a cell border, have not been considered. Figure 9 shows the results of validating the simulated telephone call number (black) against the numbers found in real-world data (green, dark green showing the average call number) over a day for two selected GSM cells.

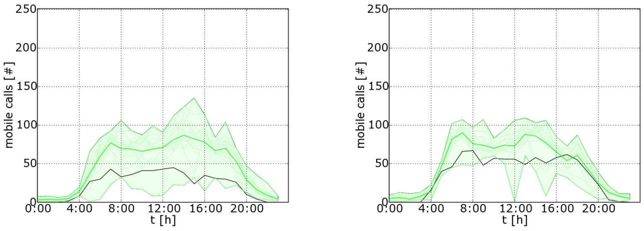


Figure 9. Validation of the telephony behavior in TrafficOnline.

B. *iTETRIS*

The interest in V2X communication is increasing but the deployment of this technology is still expensive, and ad-hoc implementations of new traffic control systems in the real world may even be dangerous. For research studies where the benefits of a system are measured before it is deployed, a simulation framework, which simulates the interaction between vehicles and infrastructure is needed, as described in III.A. The aim of the iTETRIS project was to develop such a framework, coupling the communication simulator ns3 and SUMO using an open source system called "iCS" – iTETRIS Control System – which had to be developed within the project. In contrary to other, outdated solutions such as TraNS, iTETRIS was meant to deliver a sustainable product, supported and continued to be developed after the project's end.

Besides implementing the V2X simulation system itself, which was already presented in Section III.A., the work within iTETRIS included a large variety of preparation tasks and – after completing the iCS implementation – the evaluation of traffic management applications as well as of message routing protocols.

The preparations mainly included the investigation of real world traffic problems and their modelling in a simulation environment. The city of Bologna, who was a project partner in iTETRIS, supported traffic simulation scenarios covering different parts of the town, mainly as inputs for the simulations Vissim and VISUM, both commercial products of PTV AG. These scenarios were converted into the SUMO-format using the tools from the SUMO package. Besides the road networks and the demand for the peak hour between 8:00am and 9:00am, they included partial definitions of the traffic lights, public transport, and other infrastructure information.

One of the project's outputs is a set of in-depth descriptions of V2X-based traffic management applications, including different attempts for traffic surveillance, navigation, and traffic light control. In the following, one of these applications, the *bus lane management*, is described, showing the complete application design process, starting at problem recognition, moving over the design of a management application that tries to solve it, and ending at its evaluation using the simulation system. A more detailed report on this application is [39].

Public transport plays an important role within the city of Bologna, and the authorities are trying to keep it attractive by giving lanes, and even streets free to public busses only. On the other hand, the city is confronted with event traffic – e.g., visitors of football matches, or the fair trade centre – coming in the form of additional private passenger cars. One idea developed in iTETRIS was to open bus lanes for private traffic in the case of additional demand due to such an event. The application was meant to include two sub-systems. The first one was responsible for determining the state on the roads. The second one used this information to decide whether bus lanes shall be opened for passenger cars and should inform equipped vehicles about giving bus lanes for usage.

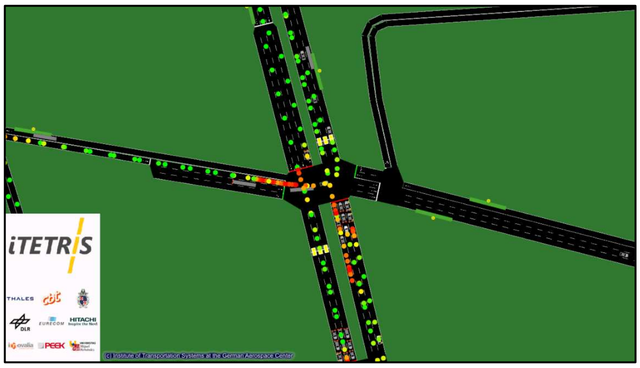


Figure 10. Speed information collection by RSUs. Each dot represents one data point, the color represents the speed (green means fast, red slow).

In order to use standardised techniques, traffic surveillance was implemented by collecting and averaging the speed information contained in the CAMs (cooperative awareness messages) at road side units (RSUs) placed at major intersections (see Figure 10). As soon as the average speed falls below a threshold, the application, assuming a high traffic amount, gives bus lanes free for passenger cars. The RSU sends then the information about free bus lanes to vehicles in range.

The evaluations show that the average speed was usable as an indicator for an increase of traffic demand. Though, as the usage of this measure is rather uncommon, further investigations and validations should be performed. When coming to measure the benefits of using bus-lanes for private vehicles, the application did not prove its benefits at all. At higher penetration rates, the average travel time of all distinguished transport modes – busses, vehicles not equipped with V2X devices, equipped vehicles, as well as rerouted vehicles – climbs above the respective average travel times without the application. The main reason is that vehicles, which use bus lanes tend to either decelerate the busses or are blocked by busses.

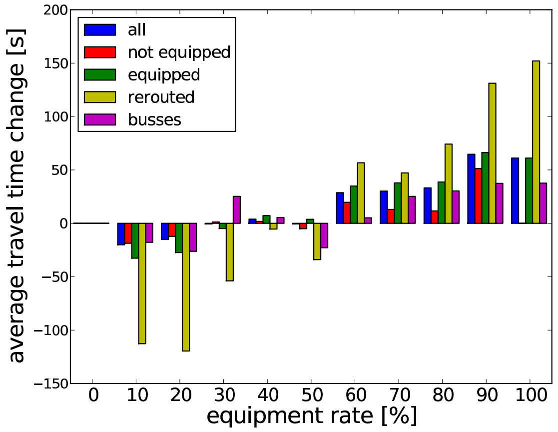


Figure 11. Average travel times changes per vehicle class over equipment rates.

The results show, that a naive implementation of the application does not take into account traffic behaviour and degrades with increasing penetration rate. This effect was observed in studies on other V2X-based traffic management applications as well. It also shows that proper design and a fine-grained evaluation of developed applications are needed.

C. VABENE

Big events or catastrophes may cause traffic jams and problems to the transport systems, causing additional danger for the people who live in the area. Public authorities are responsible for taking preparatory actions to prevent the worst case. The objective of VABENE is to implement a system that supports public authorities to decide which action should be taken. This system is the successor of demonstrators used during the pope's visit in Germany in 2005 and during the FIFA World Cup in 2006.

One focus of VABENE lies on simulating the traffic of large cities. The system shows the current traffic state of the whole traffic network, helping the traffic manager to realize when a critical traffic state will be reached. To simulate the traffic of a large region such as Munich and the area around Munich at multiple real-time speed, a mesoscopic traffic model was implemented into SUMO. This model has not yet been released to the public and is available for internal proposes only.

Similar to the *TrafficOnline* Project (Section A), the road networks were imported from a NavTeq database and adapted manually where needed. The basic traffic demand was computed from O/D-matrices supplied by traffic authorities.

The simulation is restarted every 10 minutes, loads a previously saved state of the road network and computes the state for half an hour ahead. While running, the simulation state is calibrated using traffic measurements from various sources such as inductive loops, floating car data and (if available) an airborne traffic surveillance system. This calibration is performed by comparing simulated vehicle counts with measured vehicle counts at all network edges for which measurements are currently available. Depending on this comparison, vehicles are removed prematurely from the simulation or new vehicles are inserted. Also, the maximum speed for each edge is set to the average measured speed.

A crucial part of this calibration procedure is the selection of a route for inserted vehicles. This is accomplished by building a probability distribution of possible routes for each network edge out of the basic traffic demand and then sampling from this distribution.

The accuracy of the traffic prediction depends crucially on the accuracy of the basic traffic demand. To lessen this problem we are currently investigating the use of historical traffic measurements to calibrate the simulation wherever current measurements are not yet available. However, this approach carries the danger of masking unusual traffic developments, which might already be foreseeable from the latest measurements.

Both, the current traffic state as well as the prediction of the future state is presented to the authorities in a browser-based management interface. The management interface allows to investigate the sources of collected information, including inductive loops, airborne and conventional images, as well as to monitor routes or evaluate the network's current accessibility, see Figure 12.

Figure 12. Screenshot of the “EmerT” portal used in VABENE showing travel time isochrones.

D. CityMobil

Microscopic traffic simulations also allow the evaluation of large scale effects of changes in vehicle or driver behavior such as the introduction of automated vehicles or electromobility. The former was examined with the help of

SUMO in the EU project CityMobil where different scenarios of (partly) automated cars or personal rapid transit were set up on different scales, from a parking area up to whole cities.

On a small scale, the benefits of an autonomous bus system were evaluated. In this scenario, busses are informed about waiting passengers and adapt their routes to this demand. On a large scale, the influence of platooning vehicles was investigated, using the model of a middle-sized city of 100.000 inhabitants. Both simulations showed positive effects of transport automation.

V. Recent Extensions

A. Emission and Noise Modeling

Within the iTETRIS project, SUMO was extended by a model for noise emission and a model for pollutant emission and fuel consumption. This was required within the project for evaluating the ecological influences of the developed V2X applications.

Both models are based on existing descriptions. 7 models for noise emission and 15 pollutant emission / fuel consumption models were evaluated, first. The parameter they need and their output were put against values available within the simulation and against the wanted output, respectively. Finally, HARMONOISE [40] was chosen as noise emission model. Pollutant emission and fuel consumption is implemented using a continuous model derived from values stored in the HBEFA database [41].

The pollutant emission model's implementation within SUMO allows to collect the emissions and fuel consumption of a vehicle over the vehicle's complete ride and to write these values into a file. It is also possible to write collected emissions for lanes or edges for defined, variable aggregation time intervals. The only available noise output collects the noise emitted on lanes or edges within pre-defined time intervals, a per-vehicle noise collecting output is not available. Additionally, it is possible to retrieve the noise, emitted pollutants, and fuel consumption of a vehicle in each time step via TraCI, as well as to retrieve collected emissions, consumption, and noise level for a lane or a road.

Besides measuring the level of emissions or noise for certain scenarios, the emission computation was also used for investigating new concepts of vehicle routing and dependencies between the traffic light signal plans and emissions [42].

B. Person-based Intermodal Traffic Simulation

A rising relevance of intermodal traffic can be expected due to ongoing urbanization and increasing environmental concerns. To accommodate this trend SUMO was extended by capabilities for simulating intermodal traffic. We give a brief account of the newly added concepts.

The conceptual center of intermodal traffic is the individual person. This person needs to undertake a series of trips where each may be taken with a different mode of transport such as personal car, public bus or walking. Trips may include traffic related delays, such as waiting in a jam, waiting for a bus or waiting to pick up an additional passenger. It is important to note that earlier delays influence later trips of a simulated person. The above concept is reflected in an extension of the SUMO route input. One can now specify a person as a list of rides, stops and walks. A ride can stand for any vehicular transportation, both private and public. It is specified by giving a starting edge, ending edge and a specification of the allowed vehicles. Stops correspond to non-traffic related activities such as working or shopping. A walk models a trip taken by foot but it can also stand for other modes of transport that do not interfere with road traffic. Another extension concerns the vehicles. In addition to their route, a list of stops and a line attribute can be assigned. Each stop has a position, and a trigger which may be either a fixed time, a waiting time or the id of a person for which the vehicle must wait. The line attribute can be used to group multiple vehicles as a public transport route.

These few extensions are sufficient to express the above mentioned person trips. They are being used within the TAPAS [43][44] project to simulate intermodal traffic for the city of Berlin. Preliminary benchmarks have shown that the simulation performance is hardly affected by the overhead of managing persons. In the future the following issues will be addressed:

- Online rerouting of persons. At the moment routing across trips must be undertaken before the start of the simulation. It is therefore not possible to compensate a missed bus by walking instead of waiting for the next bus.
- Smart integration of bicycles. Depending on road infrastructure bicycle traffic may or may not interact with road traffic.
- Import modules for importing public time tables.

VI. Current Development

As shown, the suite covers a large variety of functionalities, and most of them are still under research. In the past, applications from the SUMO suite were adapted to currently investigated projects' needs, while trying to keep the already given functionality work. This development context will be kept for the next future, and major changes in functionality are assumed to be grounded on the investigated research questions. Nonetheless, a set of "strategic" work topics exist and will be presented in the following sub-sections. They mainly target on increasing the simulation's validity as well as the number of situations the simulation is able to replicate, and on establishing the simulation as a major tool for evaluation of academic models and algorithms for both traffic simulation as well as for evaluation of traffic management applications.

A. Car-Following and Lane-Change API

One of the initial tasks SUMO was developed for was the comparison of traffic flow models, mainly microscopic car-following and lane-changing models. This wish requires a clean implementation of the models to evaluate. Within the iTETRIS project, first steps towards using other models than the used Krauß extension for computing the vehicles' longitudinal movement were taken by implementing an API for embedding other car-following models. Some initial implementations of other models exist, though not all of them are able to deal correctly with multi-lane urban traffic. What is already possible to do with car-following models is also meant to be implemented for lane-change models.

B. *Model Improvements*

While evaluation of academic driver behavior models is one of the aimed research topics, most models are concentrating to describe a certain behavior, e.g., spontaneous jams, making them inappropriate to be used within complex scenarios which contain a large variety of situations. In conclusion, next steps of SUMO development will go beyond established car-following models. Instead, an own model will be developed, aiming on its variability mainly.

C. *Interoperability*

SUMO is not the only available open source traffic simulation platform. Some other simulators, such as MATsim, offer their own set of tools for demand generation, traffic assignment etc. It is planned to make these tools being usable in combination with SUMO by increasing SUMO's capabilities to exchange data. Besides connecting with other traffic simulation packages, SUMO is extended for being capable to interact with driving or world simulators. Within the DLR project "SimWorld Urban", SUMO is connected to the DLR driver simulator, allowing to perform simulator test drives through a full-sized and populated city area.

VII. SUMMARY

We have presented a coarse overview of the microscopic traffic simulation package SUMO, presenting the included applications along with some common use cases, and the next development steps. The number of projects and the different scales (from single junction traffic light control to whole city simulation) present the capabilities of the simulation suite. Together with its import tools for networks and demand and recently added features such as emission modeling and the powerful TraCI interface, SUMO aims to stay one of the most popular simulation platforms not only in the field of vehicular communication. We kindly invite the reader to participate in the ongoing development and implement his or her own algorithms and models. Further information can be obtained via the project's web site [2].

REFERENCES

[1] M. Behrisch, L. Bieker, J. Erdmann, and D. Krajzewicz, "SUMO - Simulation of Urban MObility: An Overview," in SIMUL 2011, The Third International Conference on Advances in System Simulation, 2011.

[2] DLR and contributors, SUMO Homepage [Online], http://sumo.sourceforge.net/, accessed July 03, 2012.

[3] PTV AG, Vissim homepage [Online], http://www.ptv-vision.com/en-uk/products/vision-traffic-suite/ptv-vissim/overview/, accessed July 03, 2012.

[4] OpenDRIVE consortium, OpenDRIVE homepage [Online], http://www.opendrive.org/, accessed July 03, 2012.

[5] PTV AG, VISUM homepage [Online], http://www.ptv-vision.com/de/produkte/vision-traffic-suite/ptv-visum/, accessed July 03, 2012.

[6] MATSim homepage [Online], http://www.matsim.org/, accessed July 03, 2012.

[7] OpenStreetMap homepage [Online], http://www.openstreetmap.org/, accessed July 03, 2012.

[8] D. Krajzewicz, G. Hertkorn, J. Ringel, and P. Wagner, "Preparation of Digital Maps for Traffic Simulation; Part 1: Approach and Algorithms," in Proceedings of the 3rd Industrial Simulation Conference 2005, pp. 285–290. EUROSIS-ETI. 3rd Industrial Simulation Conference 2005, Berlin (Germany). ISBN 90-77381-18-X.

[9] L. G. Papaleondiou and M. D. Dikaiakos, "TrafficModeler: A Graphical Tool for Programming Microscopic Traffic Simulators through High-Level Abstractions," in Proceedings of the 69th IEEE Vehicular Technology Conference, VTC Spring 2009, Spain, 2009.

[10] S. Krauß, "Microscopic Modeling of Traffic Flow: Investigation of Collision Free Vehicle Dynamics," PhD thesis, 1998.

[11] E. Brockfeld, R. Kühne, and P. Wagner, "Calibration and Validation of Microscopic Traffic Flow Models," in Transportation Research Board [ed.]: TRB 2004 Annual Meeting, pp. 62–70, TRB Annual Meeting, Washington, DC (USA) 2004.

[12] E. Brockfeld and P. Wagner, "Testing and Benchmarking of Microscopic Traffic Flow Models," in Proceedings of the 10th World Conference on Transport Research, pp. 775-776, WCTR04 - 10th World Conference on Transport Research, Istanbul (Turkey) 2004.

[13] E. Brockfeld, R. Kühne, and P. Wagner, "Calibration and Validation of Microscopic Traffic Flow Models," Transportation Research Records, 1934, pp. 179–187, 2005.

[14] M. Treiber and D. Helbing, "Realistische Mikrosimulation von Strassenverkehr mit einem einfachen Modell," Symposium Simulationstechnik (ASIM), 2002.

[15] B. Kerner, S. Klenov, and A. Brakemeier, "Testbed for wireless vehicle communication: A simulation approach based on three-phase traffic theory," in Proceedings of the IEEE Intelligent Vehicles Symposium (IV'08), pp. 180–185, 2008.

[16] R. Wiedemann, „Simulation des Straßenverkehrsflußes," in Heft 8 der Schriftenreihe des IfV, Institut für Verkehrswesen, Universität Karlsruhe, 1974.

[17] D. Krajzewicz, "Traffic Simulation with SUMO - Simulation of Urban Mobility," in J. Barceló, "Fundamentals of Traffic Simulation," International Series in Operations Research and Management Science. Springer, pp. 269–294, ISBN 978-1-4419-6141-9. ISSN 0884-8289, 2010.

[18] A. Wegener, M. Piórkowski, M. Raya, H. Hellbrück, S. Fischer, and J.-P. Hubaux, "TraCI: An Interface for Coupling Road Traffic and Network Simulators," in Proceedings of the 11th communications and networking simulation symposium, 2008.

[19] Politecnico di Torino, TraCI4J Homepage [Online], http://sourceforge.net/apps/mediawiki/traci4j/index.php?title=Main_Page, accessed July 09, 2012.

[20] C. Sommer, Z. Yao, R. German, and F. Dressler, "On the need for bidirectional coupling of road traffic microsimulation and network simulation," in Proceedings of the 1st ACM SIGMOBILE workshop on Mobility models, pp. 41–48, 2008.

[21] ns3 Homepage [Online], http://www.nsnam.org/, accessed January 26, 2011.

[22] M. Piórkowski, M. Raya, A. Lugo, P. Papadimitratos, M. Grossglauser, and J.-P. Hubaux, "TraNS: Realistic Joint Traffic and Network Simulator for VANETs," ACM SIGMOBILE Mobile Computing and Communications Review, pp. 31-33, 2008.

[23] iTETRIS Homepage [Online], http://www.ict-itetris.eu/10-10-10-community/, accessed July 09, 2012.

[24] OMNET++ Homepage [Online], http://www.omnetpp.org/, accessed July 09, 2012.

[25] D. Rieck, B. Schuenemann, I. Radusch, and C. Meinel, "Efficient Traffic Simulator Coupling in a Distributed V2X Simulation Environment," in SIMUTools '10: Proceedings of the 3rd International ICST Conference on Simulation Tools and Techniques, Torremolinos, Malaga, Spain, 2010. ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering), ICST, Brussels, Belgium, pp. 1-9, ISBN: 978-963-9799-87-5.

[26] D. Borsetti and J. Gozalvez, "Infrastructure-Assisted Geo-Routing for Cooperative Vehicular Networks," in Proceedings of the 2nd IEEE (*) Vehicular Networking Conference (VNC 2010), 2010, New Jersey (USA).

[27] M. A. Leal, M. Röckl, B. Kloiber, F. de Ponte-Müller, and T. Strang, "Information-Centric Opportunistic Data Dissemination in Vehicular Ad Hoc Networks," in International IEEE Conference on Intelligent Transportation Systems (ITSC), 2010, Madeira Island (Portugal).

[28] E. Cascetta, A. Nuzzolo, F. Russo, and A. Vitetta, "A modified logit route choice model overcoming path overlapping problems," in Transportation and traffic theory: Proceedings of the 13th International Symposium on Transportation and Traffic Theory. Pergamon Press, Lyon, France, 1996.

[29] C. Gawron, "Simulation-based traffic assignment – computing user equilibria in large street networks," Ph.D. Dissertation, University of Köln, Germany, 1998

[30] M. Behrisch, D. Krajzewicz, and Y.-P. Wang, "Comparing performance and quality of traffic assignment techniques for microscopic road traffic simulations," in Proceedings of DTA2008. DTA2008 International Symposium on Dynamic Traffic Assignment, Leuven (Belgien), 2008.

[31] M. Behrisch, D. Krajzewicz, P. Wagner, and Y.-P. Wang, "Comparison of Methods for Increasing the Performance of a DUA Computation," in Proceedings of DTA2008. DTA2008 International Symposium on Dynamic Traffic Assignment, Leuven (Belgien), 2008.

[32] R.-P. Schäfer, "IQ routes and HD traffic: technology insights about tomtom's time-dynamic navigation concept," in Proceedings of the the 7th joint meeting of the European software engineering conference and the ACM SIGSOFT symposium on The foundations of software engineering (ESEC/FSE '09). ACM, New York, NY, USA, 171-172. DOI=10.1145/1595696.1595698, 2009.

[33] D. Krajzewicz, D. Teta Boyom, and P. Wagner, "Evaluation of the Performance of city-wide, autonomous Route Choice based on Vehicle-to-vehicle-Communictaion," TRB 2008 (87. Annual Meeting), 2008, Washington DC, USA.

[34] D. Krajzewicz, E. Brockfeld, J. Mikat, J. Ringel, C. Rössel, W. Tuchscheerer, P. Wagner, and R. Woesler, "Simulation of modern Traffic Lights Control Systems using the open source Traffic Simulation SUMO," in Proceedings of the 3rd Industrial Simulation Conference 2005, pp. 299–302, EUROSIS-ETI, 3rd Industrial Simulation Conference 2005, Berlin (Germany). ISBN 90-77381-18-X.2005.

[35] L. Bieker, "Emergency Vehicle prioritization using Vehicle-to-Infrastructure Communication," Young Researchers Seminar 2011 (YRS2011), 2011, Copenhagen, Denmark.

[36] D. Greenwood, B. Burdiliak, I. Trencansky, H. Armbruster, and C. Dannegger, "GreenWave distributed traffic intersection control," in Proceedings of The 8th International Conference on Autonomous Agents and Multiagent Systems - Volume 2, pp. 1413–1414, 2009.

[37] O. H. Minoarivelo, "Application of Markov Decision Processes to the Control of a Traffic Intersection," postgraduate diploma, University of Barcelona, 2009

[38] J. Kerekes, M. Presnar, K. Fourspring, Z. Ninkov, D. Pogorzala, A. Raisanen, A. Rice, J. Vasquez, J. Patel, R. MacIntyre, and S. Brown, "Sensor Modeling and Demonstration of a Multi-object Spectrometer for Performance-driven Sensing," in Proceedings of Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery XV, SPIE Vol. 7334, Defense and Security Symposium, Orlando, Florida, 2009, DOI: 10.1117/12.819265.

[39] L. Bieker and D. Krajzewicz, "Evaluation of opening Bus Lanes for private Traffic triggered via V2X Communication," (FISTS 2011), 2011, Vienna, Austria.

[40] R. Nota, R. Barelds, and D. van Maercke, "Harmonoise WP 3 Engineering method for road traffic and railway noise after validation and fine-tuning," Technical Report Deliverable 18, HARMONOISE, 2005.

[41] INFRAS. HBEFA web site [Online], http://www.hbefa.net/, accessed July 09, 2012.

[42] D. Krajzewicz, L. Bieker, E. Brockfeld, R. Nippold, and J. Ringel, "Ökologische Einflüsse ausgewählter Verkehrsmanagementansätze," In Heureka '11, 2011, Stuttgart, Germany.

[43] R. Cyganski and A. Justen, "Maßnahmensensitive Nachfragemodellierung in mikroskopischen Personenverkehrs-modellen," Deutsche Verkehrswissenschaftliche Gesellschaft, Schriftenreihe B, 2007.

[44] G. Hertkorn and P. Wagner, "Travel demand modelling based on time use data," in 10th International conference on Travel Behaviour Research, August 2004.

Hybrid WDM–XDM PON Architectures for Future Proof Access Networks

Rodolfo Alvizu, Alfredo Arcia, Maybemar Hernández, Mónica Huerta
Departamento de Electrónica y Circuitos
Universidad Simón Bolívar (USB)
Sartenejas, Venezuela
ralvizu@usb.ve, arcialfredo@hotmail.com, maybemarh@usb.ve, mhuerta@usb.ve

Idelfonso Tafur Monroy
Department of Photonics Engineering
Technical University of Denmark (DTU)
Lyngby, Denmark
idtm@fotonik.dtu.dk

Abstract— **The design of future proof optical access networks is a challenging and hot research topic. In this work we summarized the requirements for future proof networks and the expected evolution of optical access networks. The current network infrastructure is composed by too many domains where the electronic processing bottleneck must be relieved. The evolution and penetration of Optical Access Networks threatens to create a higher electronic bottleneck at Metro–Access interfaces. In this paper we evaluated an enhanced version of time–wavelength access architecture with optical bypass and traffic self–aggregation based on colorless customer terminals. The purpose of this hybrid Passive Optical Network with Wavelength Division Multiplexing and Time Division Multiplexing architecture is to allow for a transparent Metro–Access interface with low latency and reduced power consumption. The time–wavelength access architecture is based on the nonuniform traffic profile in networks to introduce optical bypass. For comparison a Passive Optical Network with Wavelength Division Multiplexing architecture was used as reference. Simulation results obtained by a developed model in OPNET Modeler showed that the Passive Optical Network with Wavelength Division Multiplexing tends to present electronic bottleneck issues, even though delays and loss rates were not very different in both architectures. These bottleneck issues can be avoided by the introduction of optical bypass and traffic self–aggregation. Optical bypass allows nodes to avoid electronic processing; therefore, delay and power consumption can be reduced. The proposed architecture is an interesting useful approach for Metro–Access integration and future access networks. This paper represents a baseline for future design of hybrid Passive Optical Network with Wavelength Division Multiplexing architectures and optical bypass for traffic self–aggregation. Such hybrid schemes can be based on multiplexing technologies like Orthogonal Frequency Division Multiplexing.**

Keywords– Future access network; optical bypass; Self–Aggregation; Hybrid WDM–PON; Metro–Access.

I. INTRODUCTION

This paper, based on the work showed in [1], presents new results, current trends, and the general scenario for future proof optical access networks. The world has been witnessing an explosion of high bandwidth consuming applications, and an ever rising bandwidth demand by users. According to Cisco's Visual Networking Index, global IP (Internet Protocol) traffic will grow fourfold from 2010 to 2015 [2]. Globally, Internet video traffic is expected to grow from 40% of all consumer Internet traffic in 2010 to 61% in 2015. New solutions must allow continuous growth of access networks.

Currently implemented networks are composed by several domains like access, metro, and core. Within these layers there are too many IP routers and nodes that are costly in energy and packet delay due to the Optical– Electronic–Optical (O–E–O) conversions. Multilayer network compatible with Giga bits per second (Gbps) access rates are being aggressively developed, able to endure high scalability and flexibility, guaranteed end to end performance and survivability, energy efficiency and lower cost of ownership [3]. Due to the massive rollouts required at access networks, this domain presents additional technologic and economic challenges. Those additional challenges have increased the bottleneck at access networks to deliver enough bandwidth to end users. The introduction of optical fiber to the customer sites has been accepted as a solution to relieve the access bandwidth bottleneck and to cope with the ever increasing users' bandwidth demand [3][4].

To successfully deliver enough bandwidth to end users, the bottleneck at access networks must be released. As consequence, the introduction of optical fiber to the home (FTTH) or fiber to the customer sites (FTTx; stands for fiber to the x, where x stands for Curb, Building, Premises, etc.) has been accepted as a solution to relieve the access bandwidth bottleneck and to cope with the ever increasing users' bandwidth demand [5][6][7][8].

The FTTH Council announced a continuous global growth of all fiber networks [9]. The average broadband access network speed grew 97% from 2009 to 2010 [2]. Such a growth has been allowed by deployments of PONs (Passive Optical Networks) based FTTx. The total number of FTTH subscribers was about 38 million at the end of 2011 and is expected to reach about 90 million at the end of 2015 [10].

PON is a point–to–multipoint optical network, which connects an optical line terminal (OLT) at the carrier's Central Offices (COs) with several optical network units (ONUs) at the customer sites. This is done through one or more 1:N optical splitters. The success of PON relies on its high bandwidth, infrastructure cost sharing and its simple maintenance and operation, which results from the absence of electronic active components between the OLT and the

ONUs. As a consequence of the point to multi–point nature, PON's upstream channel requires a multiple access technology. Today's standards and deployments of FTTx are based on Time Division Multiple access PON (TDM–PON). TDM–PON uses a single wavelength for downstream (CO to users) and upstream (users to CO). Upstream and downstream channels are multiplexed in a single fiber through Coarse WDM (CWDM) technology standardized according to ITU (International Telecommunication Union) G.694.2 (CWDM spectral grid). TDM PON keeps the cost of access networks down, by sharing among all users the bandwidth available in a single wavelength. There are two main TDM–PON standards used for mass rollouts [9][11]:

- Ethernet PON (EPON) technology: specified by the IEEE (Institute of Electrical and Electronics Engineers) as the 802.3ah standard, which is widely deployed in United States of America and in Europe.
- Gigabit PON (GPON) technology: specified by the ITU–T G.984 standard, which is broadly deployed in Japan and South Korea.

However, TDM–PON cannot cope with future access networks' requirements regarding to aggregated bandwidth, reach and power budget [11]. To cope with this requirements, it has been widely accepted that the next step of evolution for PON architectures is the introduction of wavelength division multiple access in PON (WDM–PON) [6][10][11][12]. The WDM–PON approach assigns an individual wavelength to each ONU. This strategy allows the use of higher bandwidth for each ONU, longer reach, better scalability towards higher users' concentration, and provides transparent bit rate channels [11][12].

Hybrid schemes of WDM–PON are a current hot topic in the development of future proof networks. A combination of TDM and WDM over PON turns out into a hybrid optical network known as TDM–WDM–PON. In this scheme, a set of wavelengths is shared over time by different ONUs instead of being dedicated as in WDM–PON. TDM–WDM–PON improves the network resource's efficiency usage [13][14][15].

Another WDM–PON hybrid scheme arises from the possibility of transmitting subcarriers over a dedicated wavelength; this is the principle of Orthogonal Frequency Division Multiplexing with WDM–PON (OFDM –WDM–PON). This approach offers great capacity and applicability for the future passive optical networks [16].

A Combination of the Time Orthogonal Code Division Multiplexing (OCDM) and WDM over passive optical networks turns out into a hybrid optical network known as OCDM–WDM–PON. With OCDMA each ONU have a code word to differentiate their optical transmissions. It provides data transparency and security to end users, and it's an interesting access technique for future optical access networks [8][17].

The continuous growth of the users bandwidth demand is leading to more than 100 Gbps optical access systems [18]. This scenario implies that COs, supporting higher concentration of customers (with split ratio extended far beyond 1:64), will have to aggregate traffic in volumes reaching Tera bits per second (Tbps). Thus, there will be a higher congestion for management of the increasing bandwidth demand at the Metro–Access interface, and higher requirements of future applications on bandwidth guarantee and low latency.

In this paper, we present an accurate performance assessment of an hybrid WDM–PON architecture originally proposed in [19]. This architecture introduces a time–wavelength (t–λ) routing for an on–the–fly (optical bypass) routed, self–aggregating Metro–Access interface. The t–λ routing architecture, based on nonuniform traffic distribution in access networks, introduces optical bypass toward the most requested destinations. Its goal is to relieve the electronic bottleneck and simplify the Metro–Access interface.

Wieckowski et al. [19] have outlined the advantages of the t–λ routing architecture over WDM–PON. However, there is a situation in their simulation model that produces delay variations and packet reordering problems. The present enhanced version of t–λ routing architecture avoids those problems, thus allowing traffic to cope with the strict requirements of delay sensible traffic.

Simulation results obtained show that by using this enhanced t– λ routing architecture, the COs become congestion–free, a reduction of the network's power consumption is achieved, and the network is able to address different requirements of the traffic. Therefore, the present work proves that the use of optical bypass and traffic self–aggregation is a very attractive approach for future proof access networks.

The paper is organized as follows: Section II describes current Network domains. Section III summarizes the requirements for future optical networks. Section IV explains the expected evolution of optical access networks. Section V describes the $t-\lambda$ routing optical access network architecture. Section VI presents the related work of this paper. Section VII describes the simulation model developed and presents the simulation results. Section VIII concludes the paper and exposes future works.

II. TELECOM NETWORK DOMAINS

In general, Telecom Networks are divided into three major networks domains: core, metro and access as depicted in Figure 1. Core networks represent the backbone infrastructure covering nationwide and global distances. It aggregates and distributes traffic from large cities (as network nodes), countries and even continents. The core domain is usually based on a mesh interconnection pattern and carries huge amounts of traffic collected through the interfaces with the metro networks. Links in the core network could have a reach of a few hundreds to a few thousands of kilometers. In the backbone network, optical technologies are widely deployed based on IP over SONET (Synchronous Optical Network) / SDH (Synchronous Digital Hierarchy), IP over SONET/SDH over WDM (Wavelength–Division Multiplexing) and more recently IP over WDM [20].

The metro network domain is the part of a telecom infrastructure that covers metropolitan areas. This domain interconnects several nodes for aggregation of business and

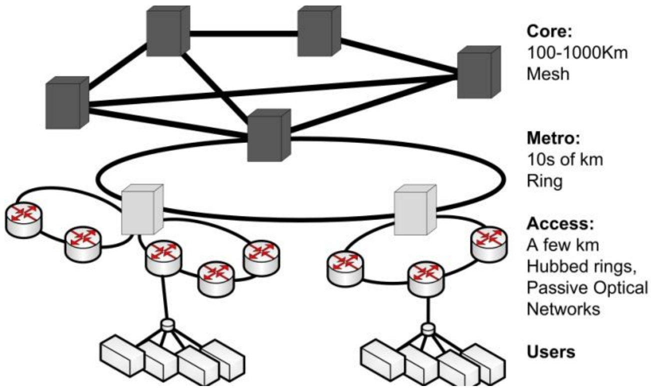


Figure 1. Telecom Network Domains [20].

residential subscribers' traffic. The metro represent the connection between the core domain and the access domain, allowing for Internet connectivity. Links in the metro could have a reach of a few tens to a few hundreds of kilometers. Metro domain is mostly based on SONET/SDH optical ring networks. Other commonly used metro networks are Metro Ethernet [21] and Metro WDM ring networks [22].

The access network connects business and residential customers to the rest of the network through their service provider Metro network. It aggregates and distributes traffic from several end user localities, covering areas of few kilometers. The Access is still dominated by several versions of Digital Subscriber Line (xDSL), but there is an increasing deployment of FTTH or FTTx, based on different flavors of PONs. Optical access networks are usually based on treelike topologies [11].

A. *Traffic Profile between network domains*

In general the Telecom network domains are interconnected by one or a few interfacing nodes (gateways), e.g., there is typically only one interconnection node between the Access and the Metro network, and between the Metro and the Core network. Due to our globalized society, it is well known by the academy that there is more traffic crossing through different network domains (remote or transit traffic), than traffic staying within the same network domain (local traffic) [23][24].

It has been observed that in a multiple interfacing scenario, traffic demands at IP routing nodes are not evenly distributed among all destinations. About four to five major destinations comprise the 80–95% of the outgoing traffic in the routers, and 50–70% of the traffic goes to one major destination [23]. This behavior can lead to insufficient capacity and traffic bottlenecks at specific points of the network. Thus, such traffic behavior must be considered in the design of new generation networks.

B. *Metro–Access Interface*

The fast penetration and the evolution of optical access networks will produce higher congestion at COs and Metro–Access interfacing nodes. Consequently, there will be a large pressure to deal with the increasing Quality of Service (QoS) demanded by future applications. Thus, future–proof, cost and energy efficient Metro–Access interfaces have been extensively investigated [19][25][26].

III. REQUIREMENTS FOR THE OPTICAL NETWORKS OF THE FUTURE

On the nearest future, the optical networks should support a major number of users and better traffic aggregation while keeping low power consumption. In this section, general future optical network's requirements are first presented. Later the specific requirements for access networks are highlighted.

A. *Future Optical Network's Requirements*

Currently implemented networks are composed by too many domains like access, metro, backhaul, outer core and inner core [3]. Within these layers there are many IP routers and nodes that are costly in energy and packet delay due to the O–E–O conversions and interfacing tasks.

It will soon be necessary the design of a multilayer network compatible with Gbps access rates, able to support high scalability and flexibility, guaranteed end to end performance and survivability, energy efficiency, and lower cost of ownership [3].

At the nodes level the legacy of O–E–O conversions have been changed for the idea of all optical networks (AON). Thus networks may move more functions to the optical domain to take advantage of the scalability of optics as the bit rate increases.

The Optical Network of the future should be remotely reconfigurable. It must be able to turn on and turn off wavelengths based on demand, and to change the optically routed paths without losing or delaying the traffic in the network. The solution to this requirement could be the introduction of Reconfigurable Optical Add Drop Multiplexer (ROADM), with All Optical Switches (AOS) [27]. Future optical network's requirements are summarized below in terms of network capacity, digital transmitter and receiver and the fiber design.

1) *Network Capacity*

There are three approaches for increasing the network capacity: increase the number of wavelengths supported by the fiber, increase the bit rate of every wavelength (spectral efficiency), and reduce the amount of signal distortion accumulated per distance unit. These approaches will be necessary in the nearest future, and they all require advanced modulations formats [27]. Modulations schemes like Quadrature Amplitude Modulation (QAM), Polarization Multiplexing and Orthogonal Frequency Division Multiplexing (OFDM) are currently being investigated. Phase Modulation Quadrature Phase Shift Keying (PM–QPSK) with coherent receivers is the main choice in the industry [28].

a) *Increasing the bit Rate per Wavelength.*

One method to increase capacity is the use of larger signal constellations like Dual Polarization Quadrature Amplitude Modulation (DP–MQAM) or DP–MQPSK. Using a modulation format with low Optical Signal to Noise Ratio (OSNR) requirement increases the achievable capacity.

However, data rate cannot be increased indefinitely by higher–order modulation because of the nonlinear Shannon's limit [29].

A challenge for increasing the bit rate of the wavelengths is the cost. As consequence of the development of high speed electronics, with each fourfold of the bit rate the cost of the associated transponder increments by a 2 or 2.5 factor [27].

b) Increasing the number of channels.

Instead of increasing the bit rate, a successful approach is to increase the number of wavelengths, reducing the inter channel spacing and possibly the bit rate.

Optical Orthogonal Frequency Division Multiplexing (O–OFMD) is a multiplexing scheme for high speed transmission. The high data rates are accomplished by parallel transmission of partially overlapped subcarriers at a lower data rate. The principle of this method is based on the orthogonality of the subcarriers frequencies, which allows overlapping the spectrum in an efficient way and eliminating the guard frequencies that were used in Optical Frequency Division Multiplexing (O–FDM) [30].

O–OFDM has the best potential of all the transmission technologies currently being investigated. The highest spectral efficiency and highest capacity reported for a WDM system in a single–core single–mode fiber was achieved using O–OFDM [16][31].

c) Nonlinearity Compensation.

The fiber capacity is limited by the nonlinearities. In the absence of noise, a single–channel signal is limited by self–phase modulation (SPM). Whereas Wavelength Division Multiplexing (WDM) systems are limited by cross–phase modulation (XPM) and four–wave mixing (FWM) [29].

The Nonlinear Schrodinger Equation (NLSE) is deterministic; it means that SPM, XPM and FWM could be compensated with digital signal processors (DSPs). As the DSP's capability improves, and systems seek to achieve highest capacity, nonlinearity compensation could become essential [29].

For the optical networks of the future, increasing spectral efficiency is the most important step as long as the reach in not sacrificed to the point of increasing regeneration costs [32]. Adding amplifiers seems reasonable, assuming Raman amplification is needed anyway, as long as the ROADM functionality and line stability are not sacrificed [32].

Reducing channel spacing appears to be the best long term solution for scalability assuming coherent detection systems as long as the performance is not sacrificed [32].

2) Digital Transmitter and Receiver

Next generation systems will continue with the trend that enabled the technologies of 100 Gbps. DSP will play an important role in the transmitter and receivers of the future optical networks, where advanced algorithms will be used to compensate fiber impairments [29].

a) Digital Coherent Receivers.

The advantage of digital coherent receivers stems on the ability to manipulate the electric field in the two signal polarizations [29]. Coherent receivers in conjunction with analog to digital converters ADC above Nyquist Rate, permit to manipulate the information contained on the digitalized waveform. Therefore, any information manipulated by hardware in the analog domain could be achieved in the digital domain by advanced software. One of the many benefits of coherent receivers is the correction of Chromatic Dispersion (CD) [29].

An additional benefit of digital coherent receivers is that it facilitates the channel demultiplexing. In traditional WDM systems, the channel of interest is selected by optical filters. In consequence, guard bands are required in WDM reducing the spectral efficiency [29]. With a digital coherent receiver, digital filters could be used for increasing the spectral efficiency.

Last but not least, a digital coherent receiver enables novel modulation formats e.g., OFDM [29].

b) Digital Transmitters.

DSP can also be used on the transmitter to generate the data above the Nyquist Rate, and Digital to Analog Converters (DAC) to acquire the signal to be sent to the receptor.

A software–defined optical system with DSP at both transmitter and receiver enables the most agile platform. It allows channel impairments to be compensated by powerful DSP algorithms, and enables new signal multiplexing and demultiplexing paradigms. In addition, the channel data rate, modulation format, and coding scheme can all be programmed by network management in response to channel conditions changes. Therefore, the network can become more flexible and tunable, allowing to route optical signals through several distances, fiber types, and different number of ROADMs in the path. Such flexibility is achieved by sensing the channel quality and adjusting the modulation format and coding scheme to provide reliable end–to–end connection at the highest data rate possible [29].

3) Fiber Design

The primary interest in optical networks is to increase system capacity. Even without nonlinearity, achievable capacity scales only logarithmically with power. A power–efficient method is to transmit information over parallel channels [29].

Future systems could require space–division multiplexing (SDM). The simplest SDM method is to use multiple fibers. This requires parallel transmitters, fibers, amplifiers, and receivers. System complexity will scale approximately linearly with capacity, so cost reduction per bit will only be achieved by minimizing the cost of inline amplifiers and transponders [29].

An alternative strategy is to have SDM within a single strand of fiber. Two schemes have been proposed: multicore fiber (MCF), and multimode fiber (MMF) [29].

B. Future Optical Access Network's Requirements

Due to the proximity of access networks with the end users, the access network presents additional technological and economic challenges. In consequence, those additional challenges had increased the bottleneck at access networks to deliver enough bandwidth to end users. The introduction of optical fiber to the customer sites has been accepted as the solution to relieve the access bandwidth bottleneck and to cope with the ever increasing users' bandwidth demand [4][33].

Requirements of network capacity, adaptation, scalability, energy efficiency, data integrity and other aspects will be summarized below.

1) Network Capacity

The future optical access network must be faster with increased bandwidth compared to current standards (e.g., EPON and GPON). Some authors expect that the minimum bit rate required in future PON architectures for downstream will be around 10Gbps, while the upstream requirement will be around 2.5 Gbps [34].

If ONUs shares the same channel for downstream and upstream, the bandwidth allocation will be dynamic. The dynamic bandwidth allocation allows operators to manage several kinds of traffic and different transmission bit rates upon requests. There are two kinds of bandwidth allocation mechanisms used in PON: status–reporting mechanism and non–status–reporting mechanism. In non–status–reporting, the CO continuously allocates a small amount of extra bandwidth to each ONU. The status–reporting mechanism and dynamic bandwidth allocation (DBA) are based on the Multi–Point Control Protocol (MPCP) specified in the IEEE 802.3ah standard [35].

Hybrid WDM–PON architectures such as TDM–WDM–PON, OCDM–WDM–PON and OFDM–WDM –PON allow to achieve higher transmission bit rate and to have more disposition of the bandwidth.

2) Adaption and Scalability

The massive deployment of fiber is limited by infrastructure investment. Hence, it is required to ensure future adaptation and scalability of the inversion. A major challenge for service providers is to keep simple operating procedures and ensure convergence of different networks. A migration to packet traffic transport platforms and the tendency towards a common access architecture facilitate the acquisition of new access technologies [36] .

The system must be flexible with the ability to be incorporated in sequential and modular way. Satisfy the needs of the operator in terms of network administration and implementation costs. The network must allow different types of user (e.g., business user, residential users, etc.) where each user may have different requirements and be subject to individual claims [36].

3) Costs and Energy Eficiency

The cost is the current limiting factor for massive deployments of optical access networks. There are different costs to consider in the networks: implementation, maintenance, infrastructure, network improvement and environmental impact. Energy efficiency has become a very important aspect for network design, due to increased operating costs related to energy consumption, increasing awareness of global warming and climate change. Due to its low power consumption, PON is considered as a green technology [9].

4) Integrity and Other Aspects

Redundancy includes automatic reconnection across redundant network elements, and should minimize the impact in case of failure. The system must be able to provide mechanisms for fault detection as monitoring and diagnosis for proper management. The network must also be able to locate and remotely provide a solution at the full extent of the network [36].

The systems must support heterogeneous access networks. The convergence of networks offers the possibility to optimize the total costs and to provide access technology solutions. In this context, fixed access backhauling and mobile backhauling must be considered [36].

Network operators are looking forward to simplify the network structure and reduce the number of access sites. Node consolidation would satisfy this requirement improving the overall cost efficiency of the network. Node consolidation is possible by increasing the reach (@100 Km) and the number of users supported by each access site [36].

Future access networks must provide data security and integrity to the customers. The terminal equipment should be simpler in terms of administration and configuration as possible (i.e., plug and play) [36].

IV. Evolution of Optical Access Networks

In this section the evolution of optical access networks towards a future proof solution is presented.

A. TDM–PON

Time Division Multiplexing (TDM) is a multiplexing scheme that allows N users to share in time the bandwidth offered by the provider. In optical communications the shared resource is an assigned wavelength [13]. Currently, there are a few implemented standards of TDM based PON. However, the most employed are EPON (IEEE 802.3ah), principally adopted in Asia, and GPON (ITU–T G.984) in Europe and North America [13].

EPON transports Ethernet frames in PON, this standard combines the low costs of the Ethernet devices with the passive optical components of PON. IEEE 802.3 (Ethernet in the First Mile) in 2004 established the standard for symmetric traffic at 1 Gbps, for a 10 to 20 km link, with 16 ONUs per each OLT.

GPON transmits TDM data frames, at a data rate of 2.5 Gbps for the downlink and 1.25 Gbps for the uplink.

With the higher bandwidth requirements, it was necessary to implement a 10 Gbps standard (10 GE–PON). However, as the number of users and the bandwidth requirements increase, the TDM–PON architecture becomes insufficient.

B. WDM–PON

A straightforward and widely accepted upgrade for TDM–PON is the introduction of wavelength division multiple access in PON [6][10][11][12]. The WDM–PON approach assigns an individual wavelength to each ONU. This strategy allows the use of higher bandwidth for each ONU, longer reach, better scalability towards higher users' concentration, and provides transparent bit rate channels

[11][12]. WDM–PON provides security and integrity to end users, and it facilitates network upgradeability for enhanced future–proofing.

The high cost of WDM–PON has made it less attractive for implementation. To date, the only commercially available WDM–PON system has a cost roughly twice of EPON or GPON [10].

The GigaWaM Project aims to develop essential optical subsystem components required for a future–proof WDM–PON broadband access system providing each end user with up to 1 Gbps bidirectional data bandwidth. A main goal of GigaWaM Project is to achieve a target cost below that of a typical TDM–PON implementing 70 Mbps per user [10].

C. Hybrid WDM–PON

In this subsection, three hybrid access technologies that provide improved capacity, scalability, and spectral efficiency are presented. Such technologies are envisioned to meet the requirements of future services and applications.

Even though high speed digital signal processors are in continuous evolution, the photonic circuits still an immature technology. Therefore, practical implementation of Hybrid WDM–PON architectures represents a major challenge due to the complexity and costs in both photonics and high speed electronics hardware [8].

1) TDM–WDM–PON

A Combination of TDM and WDM over PON turns out into a hybrid optical network known as TDM–WDM–PON. In this scheme a set of wavelengths are shared over time by different ONUs instead of being dedicated as in WDM–PON. This architecture improves the efficiency of the network resources [13].

The design and implementation of this architecture present new challenges, because the network should be dynamic, in the sense that each ONU must tune into the different wavelengths of the OLT. Also, the transmission and reception of the TDM packets shall be configurable.

An architecture based in TDM–WDM–PON is proposed in [13] using new modulation schemes. On the downlink, Phase Shift Keying (PSK) is used. On the Uplink the signals are remodulated with On–Off Keying (OOK) using a delay Interferometer (DI), and an NxN cyclic Arrayed Wave guide Grating (AWG). In this scheme, the interference between both, uplink and downlink is decreased. Also, only one instead of two dedicated DI are required on each ONU. The architecture was demonstrated at a data rate of 10 Gbps on the downlink and 1.25 Gbps on the Uplink.

High capacity and long reach architectures have been demonstrated [14][15]. A TDM–WDM–PON with capacity of 8192 ONUs, symmetric traffic of 320 Gbps and links of 135.1 km length was demonstrated [14]. The configuration of the network consisted of 32 wavelengths in the downlink and another 32 in the uplink, with 50 GHz spacing. The transmission was in burts at 10 Gbps each. The total capacity achieved was 256 ONUs per wavelength. A similar architecture with capacity of 16384and 100 km reach is demonstrated in [15].

2) OFDM–WDM–PON

OFDM is a multiplexing scheme that offers a major spectral efficiency due to the use of orthogonal frequencies for data transmission. It has been demonstrated that N carriers could be transmitted over the same frequency band without interfering each other [16]. With respect to other multiplexing schemes like Single Carrier Multiplexing (SCM) or Frequency Division Multiplexing (FDM), OFDM–WDM–PON is more efficient, because it eliminates the guard bands between the different channels.

An OFDM channel offers many possibilities for modulation formats, such as amplitude, phase, polarization and intensity. For this reason, depending on the type of the modulation format selected, it is possible to obtain different reception schemes. Among these architectures are: amplitude modulation Optical OFDM (O–OFDM), optical modulation with Coherent detection (CO–OFDM), and All Optical OFDM (AO–OFDM).

OFDM is an option that may offer a bigger capacity for the optical fiber systems. There are multiple variations for the implementation of optical OFDM. In the simplest form, it is possible the transmission of multiple dedicated optical subcarriers over the passive optical network, this is known as OFDM–PON. A more complex scheme is to provide multiple access by sharing the subcarriers as needed. Therefore, subcarriers are allocated to different users in different slots of time; this is the case of TDM–OFDMA–PON. Last but not least important, there is the possibility of transmitting the subcarriers over a dedicated wavelength; this is the principle of OFDMA–WDM–PON. This option offers the biggest capacity and applicability for the future optical networks [16].

Record of 1.92 Tbps (40λ × 48Gbps/λ) coherent DWDM –OFDMA–PON without high–speed ONU–side was demonstrated over 100km straight Standard Single Mode Fiber (SSMF) with a 1:64 passive split [37]. Novel optical–domain OFDMA sub–band selection, coherent detection, and simple RF components are exploited. As an experimental verification of a next–generation optical platform capable of delivering 1 Gbps to 1000 users over 100km, this is a promising architecture for future optical Metro– Access consolidation.

The proposed DWDM–OFDMA–PON architecture consists of optical carriers separated by a 50 GHz band for both the uplink and the downlink, without the use of high speed ONUs in the downlink, because of the coherent detection. The optical channels contain four subcarriers at 12 Gbps data rate each. The result is 48vGbps–λ speed. For the DS/US 40 wavelengths on the 1532.29–1563.45nm range are used [37].

Symmetric 1.2 Tbps DWDM–OFDMA–PON is demonstrated in [31] over a SSMF with a 1:32 passive splitter. The proposed system supports 800 ONUs with reduced complexity and it could be a solution for Metro–Access consolidation.

The proposed architectures in [16][37] are resumed in [38], where the future terabit optical networks based in DWDM–OFDMA–PON are shown. N subcarriers are generated and then modulated by an optical continuous wave. All the wavelengths are transmitted with an AWG, and an Optical Single Side Band (OSSB) signal is produced with an interleaver (IL). The transmission is made through a SSMF. After X kilometers a local exchange is made, where the signals are amplified and rerouted through an AWG, and DS/US pairs are retransmitted to the ONUs. Once the signal arrive to the destination each ONU tune in to the subcarrier of interest. On the uplink the subcarrier is mapped back to the band from where it was selected, and then is transmitted in the wavelength of US. The signals from different ONUs are combined and received coherently by the OLT. This architecture supports λ x N without dispersion compensation.

A symmetric transmission on baseband and 4–Amplitude Shift Keying (ASK) Fast OFDM (F–OFDM) was realized for colorless ONUs [37]. The difference between this scheme and the one showed in [38] is that the DS wavelength is reused for the US transmission, in consequence the ONU is colorless. The achieved capacity was 160 Gbps (16λx10Gb/s) in symmetric traffic.

All the presented architectures show a trend to metro access integration, since the main goal is to achieve the highest transmission speed and a long reach.

3) OCDM–WDM–PON

A Combination of the OCDM and WDM over passive optical networks turns out into a hybrid optical network known as OCDM–WDM–PON.

With OCDMA each ONU has a code word to differentiate their optical transmissions. At the ONUs an optical operation encodes each bit before transmission. At the receiver, an inverse decoding operation must be performed to retrieve data. OCDMA has a scheme similar to WDMA, offering point to point virtual links over one physical point to multipoint architecture, providing data transparency and security to end users. OCDMA supports a higher number of users than TDMA asynchronously, eliminating the processing and network overhead required for synchronization of TDMA. Thus OCDMA is presented as an interesting access technique for the optical access network of the future [8][17].

OCDMA–WDM–PON has unique features as full asynchronous transmission, low–latency access, demand –driven capacity, and data integrity. Using OCDMA–WDM–PON it has been shown experimental speeds higher than standard 10G–PON [39]. The authors in [39] included a multiport device encoder and decoder with the ability to generate and process calculations with multiple codes on multiple wavelengths. The multiport allows multiple users to connect and thus the cost of infrastructure can be shared between multiple users. The implementation of an 8x8 multi–port using 4 wavelengths, each at 40 Gbps, permits to achieve a maximum transmission speed of 2.56 Tbps (40Gbps x 8λ x 4 OC x 2 PM), over a 50 km fiber.

Mechanisms such as VPN (Virtual Private Network) in OCDMA system can be generated with Walsh coding. Some studies has proposed the creation of VPN using keys to feed each of the PRBS (Pseudo Random Bit Sequence) producing non–orthogonal sequences with larger spacing, making it more attack resistant [40].

The OCDM–WDM–PONs suffer from a lack of transitional models that take into account legacy systems [7].

V. Time–Wavelength ($t-\lambda$) Routing Optical Access network Architecture

In order to design a future proof optical Access Network, the nonuniform traffic profile represents an interesting useful behavior to be exploited by architectures with optical bypass. Optical bypass is a technique to avoid the electronic routing tasks like: optic–electronic conversions (O–E), error correction algorithms, electronic buffering, electronic processing, and electronic–optic conversions (E–O). Therefore, Optical bypass will reduce power consumption and node complexity.

The t–λ routing optical access architecture exploits the traffic profile of Metro–Access networks to introduce optical bypass and passive self–aggregation of Metro–Access traffic [19]. The goal of the architecture is to use the nonuniform traffic distribution of access networks to select the major traffic portion to transmit it through all optical channels (on–the–fly routed). In consequence, the t–λ routing optical access architecture reduces the electronic traffic aggregation and routing tasks.

The selection of the traffic portion, which will be optically bypassed relies on the analysis and evaluation of a multiple gateway traffic–distribution [23]. Based on this analysis up to 70% of the traffic generated in the access networks is destined to the Metro–Access interfacing node (called major destination). In consequence, the t–λ routing approach is designed to perform optical bypass and traffic self– aggregation at COs of the major destination traffic (up to 70% of all traffic generated by the users). The minor destination traffic (local traffic) is sent through electronically routed channels using common stop–and–forward policies.

The underlying idea of the t–λ routing approach is presented in Figure 1 using a PON to connect N=3 ONUs.

The architecture arranges ONUs in groups that are equal to the number of wavelength in the t–λ frame (3 wavelengths in the case depicted in Figure 2). For each PON the COs assign a specific wavelength/channel to perform the optical bypass (passive channel) for upstream major traffic. To avoid wavelength interference at the COs network, the wavelength assignment must ensure different passive wavelengths for each PON sharing links in their path towards the major destination.

Based on the requirements of future access networks, the t–λ routing architecture employs colorless reflective ONUs. Reflective ONUs allow providers to have centralized control over the access network at the COs. The COs send Continuous Wave (CW) seed light to ONUs based on the t–λ frame. Each ONU sorts packets at the customer sites on the basis of their destination. Major traffic is introduced in a buffer associated with the passive wavelength (passive channel). Minor destination traffic is arranged in other buffers associated with electronically routed channels.

Colorless reflective semiconductor optical amplifier modulators (RSOA) are used to transmit data by modulating

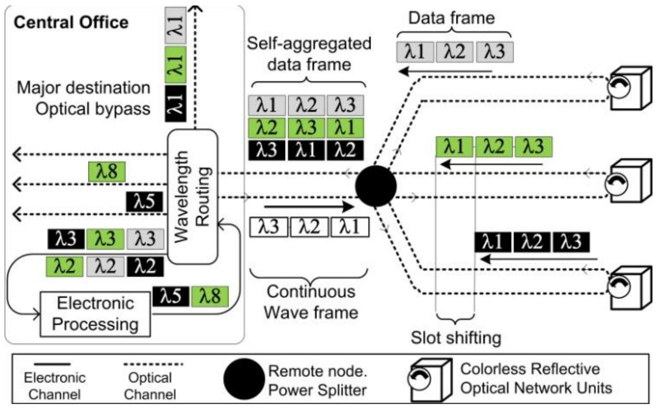


Figure 2. Underlying idea of the time–wavelength assignment and Central Office (CO) schematic diagram, for time–wavelength routed architecture to provide on–the–fly routing to Major destinations. CW: Continuous Wave.

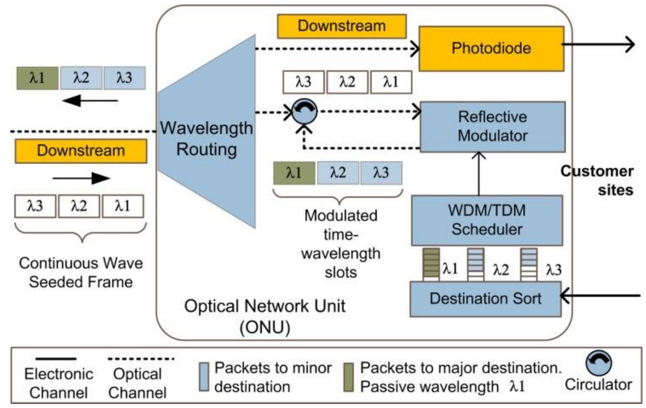


Figure 3. Basic Reflective Optical Network Unit (ONU) scheme. Downstream and Upstream channels coming from the Central Office (CO) are separated by a wavelength filter. The seeded upstream Continuous Wave (CW) is modulated and reflected back to the CO by the colorless reflective semiconductor optical amplifier (RSOA). Incoming customer traffic is sorted by destination in specific buffers.

the CW seeded by COs [25][41]. Major destination traffic only modulates the predefined passive wavelength.

A collision free optical self–aggregation will be achieved by assuring that the upstream signals arrive at the remote node (splitting and combining point) properly adjusted in time based on the t–λ frame (see Figure 1). In consequence, each wavelength at the CO will contain self–aggregated data from several ONUs. At COs the self–aggregated major destination traffic is optically routed on–the–fly (optical bypass). The rest of the wavelengths are sent to local processing units for destination inspection, routing and forwarding (electronically routed channels). As shown in Figure 1, the CO assigns λ1 as passive channel. Hence λ1 is self–aggregated and on–the–fly routed towards the major destination.

VI. Related Work

In a former performance evaluation, the authors have shown the advantages of t–λ routing architecture over a conventional WDM–PON. It was therefore proposed as solution for a transparent and self–aggregating Metro–Access integration architecture [19]. In the former evaluation the architectures were assessed by means of dedicated discrete step–based simulation models, with a Poisson process for traffic arrivals and packet Loss Rate (LR) as the performance metric [19]. In the present work, an event–based simulation model has been developed. The developed model introduces some enhancements in relation with the previous architecture.

A. Excess Traffic in passive channels

In the simulation model developed in the prior evaluation, the excess traffic in passive channels was distributed to electronically routed channels. Excess traffic in passive channels constitutes major traffic (traffic destined to a major destination) arriving at ONUs, which produces overflows of buffers associated with the passive channels (passive buffers). Therefore, upon passive channel overflows the major excess traffic is electronically routed towards the major destination. This approach allows to use passive channels resources at the maximum and to distribute major excess traffic trough the electronically routed channels.

LR was the only performance metric used in the former evaluation, and advantages of the t–λ routing architecture over WDM–PON have been shown [19]. But customers are demanding applications with requirements beyond bandwidth and loss rate. According to Cisco's Visual Networking Index, globally Internet video traffic will grow from 40% of all consumer's Internet traffic in 2010 to 61% in 2015 [2]. Inelastic traffic applications especially video–related traffic hava been flooding the networks, and future optical access networks must address the requirements of low delay and jitter.

There is a negative behavior in the t–λ routing architecture if the excess traffic in passive channels is distributed through electronically routed channels. In general, using this approach the packets will arrive at major destination with high delay variations. Such delay variations lead to constant packet reordering at the major destination. Packet reordering can be compensated with buffering at the price of an additional delay. In future optical access networks (managing traffic volumes of Tera bits per second), the buffer size for packet reordering could be prohibitive. We analyze two types of Internet traffic for discussing the traffic behavior and the consequences of this approach:

1) Inelastic traffic (voice, video–related and time sensitive traffic)

The packet reordering compensation adds additional delay. Thus, inelastic traffic will be most likely to be discarded, because even when the destination is reached, it could not meet the low delay and jitter requirements. Additionally, this traffic (most likely to be discarded) will compete inside the t–λ routing architecture for resources with minor destination traffic.

2) Elastic traffic (web, mail)

TCP (Transport Control Protocol), the standard Internet transport protocol for non delay sensible traffic, has been proven to be packet reordering–sensible. TCP produces unnecessary traffic retransmission under constant packet reordering, and can lead to bursty traffic behavior, thus increasing the network overload [42].

Based on the negative behavior in elastic and inelastic traffic produced by the use of the approach presented in [19], in this work we use a different approach for excess major traffic.

The t–λ routing architecture has been proposed as a solution for a transparent and self aggregating Metro–Access integration. Considering that inelastic traffic will flood future networks, we developed an enhanced t–λ routing architecture model. In the enhanced version, the excess traffic in passive channels is dropped at the ONUs. In consequence, the delay variations and the packet reordering problems are avoided.

By dropping the excess traffic at ONUs the electronic routing tasks are reduced, avoiding the electronic bottleneck at COs (i.e., it moves the network bottleneck of major traffic from the COs to the ONUs). The considered approach in the proposed enhanced t–λ routing architecture is that if some packets should be discarded by a network bottleneck, it is better that happen as soon as possible.

B. Routing algorithm

Routing in the network model implemented in [19], is based on the Dijkstra algorithm. For each network node, all possible paths to other nodes are calculated. Next, the shortest paths to all achievable destinations are selected. The shortest path in the Dijkstra algorithm is the smallest weight path, so it potentially introduces the smallest delay and losses. The unique names of destinations with the "best" next hops are written into the routing table. Thus, the Dijkstra algorithm was implemented to consider only one shortest path even if there is more than one equal weight shortest path to any particular destination.

The proposed enhanced routing approach considers all different equal weight shortest paths and evenly shares the traffic between them, performing a network load balancing. Load balancing improves network performance by reducing losses caused by buffer overflows and by reducing the average packet end–to–end delay due to the reduction of the time that packets spent in the buffers.

In the developed model, the nodes manage routing tables based on the OPNET Modeler DJK package. The DJK package is based on a variation of Dijkstra algorithm to generate the set of shortest paths for each node [43].

VII. Performance Assessment of WDM–PON vs. $t-\lambda$ Routing Optical Access network

In the present work, the enhanced t–λ routing architecture and a conventional WDM–PON architecture (as reference) have been evaluated by means of simulation models developed in OPNET Modeler. OPNET Modeler is an event–based state–of–the–art network system modeling and simulation environment [43].

A. Performance metrics

In the related work, the authors had only considered LR to assess the architectures. Nevertheless, the delay metrics had gained more relevance to offer today services. Globally, Internet video traffic is expected to grow from 40% of all consumer Internet traffic in 2010 to 61% in 2015 [2]. In the present work, the developed simulation model in OPNET Modeler introduces *LR*, end–to–end delay and average buffer delay as performance metrics.

1) Loss Rate LR

It is the relation between lost packets and sent packets (1). When a specific node, is congested the packet loss is unavoidable. Based on *LR* we determined if the t–λ routing architecture can effectively avoid the packet loss due to the electronic routing bottleneck (the cause of network congestion).

$$LR = \frac{Packets_{lost}}{Packets_{sent}} \quad (1)$$

2) End–to–End delay EED

It is the time difference between the time instant when the first bit of a packet arrives at an ONU and the time instant when the last bit of a packet is received at its destination. Packet EED is calculated as the summation of buffering delay, processing delay and transmission delay. The EED is an important metric for the performance assessment of optical bypass in the t–λ routing architecture, and the benefits of avoiding electronic routing at COs.

3) Average Buffer Delay

Defines the average time experienced by each packet into a specific buffer. The buffer delay is the time difference between the time instant when the first bit of a packet arrives at a specific buffer and the time instant when the last bit of a packet gets out of the buffer. It is calculated based on (2).

$$\overline{Buffer\_Delay_i} = \frac{1}{N_i}\sum_{n=1}^{N_i} Buffer\_Delay_n \quad (2)$$

Where $\overline{Buffer\_Delay_i}$ is the average delay experienced by the packets that went out in the time window i. N_i is the number of packets that went out of the buffer in the time window i. $Buffer\_Delay_n$ is the delay experienced by the n^{th} packet in the i^{th} time window.

The variations in the EED experienced by the packets are consequence of medium access, electronic routing and aggregation tasks (processing and buffering delay). Thus buffer delay is an interesting useful metric, which allows determining bottlenecks and congestion points in the evaluated networks.

B. Traffic model

At ONU nodes a traffic generator models incoming traffic. The traffic is generated based on probabilistic distributions for: packet inter–arrival times, packet length and packet destination.

The traffic model used in [19] was Poisson, with fixed length. In the present work a more accurate traffic model was introduced for a more accurate performance assessment.

At ONU nodes a traffic generator can generates incoming traffic with several short range dependence distribution (e.g., Poisson) and long range dependence distribution (e.g., Fractal Point Process), using a variation of the OPNET simple source and the raw packet generator respectively [43]. It is well known that Internet traffic presents self–similar behavior, which reproduces the burstiness of Internet traffic [44]. Hence, for the performance evaluation the traffic generator was configured to generate incoming traffic with self–similar inter–arrivals times.

C. Validation of t–λ Routing Architecture Model

For validation purposes the simple topology presented in Figure 4 was used.

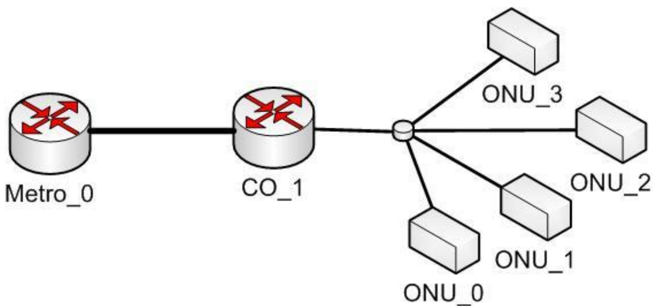


Figure 4. Validation scenario. Four Optical Network Units (ONU) connected through a Passive Optical Network to a Central Office (CO_1), the CO_1 is connected to the Metro–Access interfacing node (Metro_0).

The validation scenario consisted in a PON based on the t–λ routing architecture connected to $N = 4$ ONUs and the CO_1 was connected to the Metro–Access interfacing node (Metro_0).

1) Scenario Configuration

The t–λ frame period was set to $T_{frame} = 125\ \mu\ seconds$ long. Therefore, each t–λ slot time was set to $T_{Slot} = \frac{T_{frame}}{4} seconds$. There were only two possible destinations in the network: CO_1 and Metro_0. Optical bypass was performed for traffic destined to Metro_0 (through the t–λ passive channel). ONU–CO_1 links had four WDM channels to accomplish the t–λ frame, and a transmission rate of $R = 155.520\ Mbps$ (OC3). The CO_1–Metro_0 link was configured with two WDM channels at $R = 155.520\ Mbps$. CO_1 and Metro_0 processing capacity were in line with the transmission links rate $C = 155.520\ Mbps$. The buffer size was set to introduce a maximum buffer delay of 10 $msec$ based on the following equation:

$$Buffer_{length} = R \cdot Max\_buffer\_delay \quad (3)$$

The traffic inter–arrival time was set to present self–similar properties generated with a Power ON–Power OFF model with Hurst parameter H= 0.8. Packet length was exponentially distributed with a mean of $1024\ bit$. Simulation time was set to $100\ sec$.

2) Theoretic Analisys

Each ONU had $N = 4$ buffers associated with each t–λ channel. Because of the time division multiplexing, in average each ONU experienced a transmission rate of $R_i = \frac{R}{N}\ bps$.

The t–λ routing architecture was proposed to introduce the optical bypass of only one passive channel per each CO [19]. In consequence, for each ONU the optical bypassed traffic (destined to Metro_0) experienced in average a transmission rate of $R_i = 0.25 \cdot R\ bps$. While the rest of the traffic (destined to CO_1) experienced in average a transmission rate of $3R_i\ bps$.

Even though the CO_1 – Metro_0 link had 2 WDM channels at $R\ bps$, the t–λ routing architecture used only one WDM channel for traffic destined to Metro_0. Channel sub utilization was expected due to the lack of packet segmentation to fill up the t–λ slots in the simulation model.

The developed model considered error free optical links. Therefore, as CO_1 and Metro_0 processing capacity were in line with the transmission rate, losses were only produced by buffer overflow.

There was a major bottleneck at the passive channel of the t–λ routing architecture. When the traffic destined to Metro_0 exceeds arrivals at $R_i = 0.25 \cdot R\ bps$ the buffer associated with the passive channel overflowed and in consequence, packets were lost.

3) Simulation Results

Figure 5 depicts the results of Loss Rate (*LR*) vs. Offered Traffic (*A*) for four different traffic distributions. The distributions ranges from 25% of *A* destined to Metro_0 (75% destined to CO_1) to 100% of *A* destined to Metro_0.

a) Case 1: 25% of A destined to Metro_0 and 75% of A destined to CO_1

In this case, the t–λ routing architecture perceived a uniform distribution of *A* in the four t–λ channels. When $A < 80\%$ there were no buffers overflow because there is none overloaded channel. For $A > 80\%$ the four channels started to lost packets due to buffers overflow. When the network was fully loaded ($A = 100\%$) the four channels were overloaded and the overall *LR* was 20%. This 20% of *LR* was consequence of the channel sub utilization. Therefore, there was a 5% of capacity in each channel being lost due to the channel sub utilization.

b) Case 2: 50% of A destined to Metro_0 and 50% of A destined to CO_1

The 50% of *A* destined to CO_1 was transmitted by three not overloaded channels. Those three channels handled without overloading up to 60% of *A* (20% per each channel). For $A > 40\%$ the passive channel started to get overloaded. For $A = 100\%$ there was 30% of *LR*, due to only 20% of *A* was transmitted through the passive channel.

c) Case 3: 75% of A destined to Metro_0 and 25% of A destined to CO_1

In this case, 25% of *A* was transmitted loss–free towards CO_1 through 3 t–λ channels. When $A = 100\%$, from the 75% of *A* destined to Metro_0 the passive channel was only

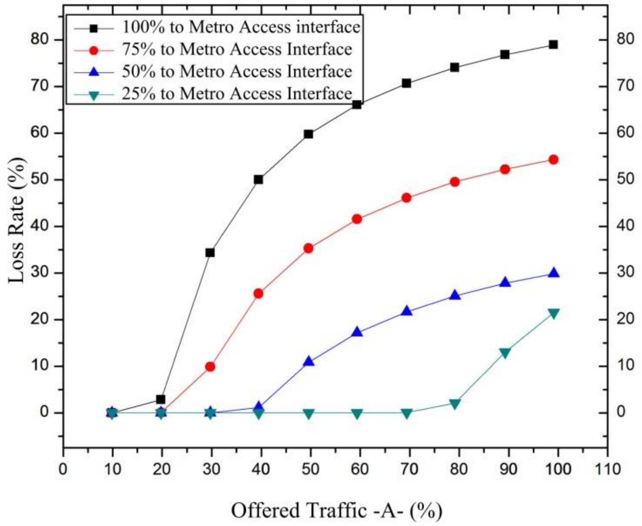


Figure 5. Loss Rate (LR) vs. Offered Traffic (A) for the t–λ Routing Architecture. The four curves represents diferent traffic distributions.

able to transmit up to 20% of A. In consequence, for $A = 100\%$ LR was expected to be 55%.

d) Case 4: 100% of A destined to Metro_0

For this case, when $A > 20\%$ the LR started to increase because the passive channel was overloaded. For $A = 100\%$ a 80% of LR was expected.

Figure 6 presents the curve of Carried Traffic (A_c) vs. Offered Traffic (A) for link ONU_0 – CO_1 using the first case of traffic distribution (25% of A destined to Metro_0 and 75% of A destined to CO_1). In this case the Offered Traffic was uniformly distributed through the four t–λ channels. For $A < 80\%$ the curve obtained is a straight line where $A_c = A$. When $A \geq 80\%$ the four channels started to get overloaded and their associated buffers started to overflow. Therefore, the Carried Traffic A_c had a limit of 80%. These results were consistent with the analysis presented with Figure 5.

Figure 7 depicts the average Buffer Delay for the associated passive channel of ONU_0 using the first case of traffic distribution (25% of A destined to Metro_0 and 75% of A destined to CO_1). There are six curves corresponding to $A = [100\%, 90\%, 80\%, 70\%, 60\%, 10\%]$.

To evaluate the results of Figure 7, it has to be remember that each channel perceived in average a transmission rate of $R_i = 0.25 \cdot R\ bps$, but 5% of the channel capacity was lost by the channel sub utilization. Therefore, the buffer of each channel must be overflowed for $A > 80\%$.

In Figure 7, for $A \leq 80\%$ there is a proportional increase of the average Buffer Delay related to A. For $A = 80\%$ the mean value of the average buffer delay was $1.17\ msec$. For $A > 80\%$ the mean value of the average buffer delay was $11.58\ msec$; showing that the buffer was overflowed.

The maximum buffer delay was set to $10\ msec$. However such maximum delay was defined by (3), where no channel sub utilization was considered. However, the channel sub utilization made packets to stay more than $10\ msec$ at the buffers.

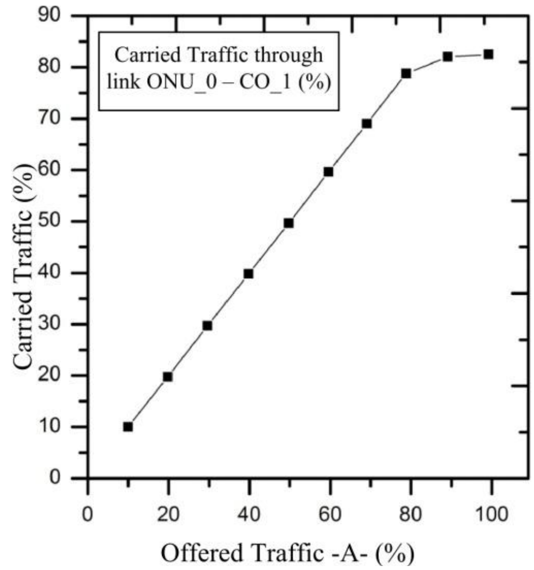


Figure 6. Carried Traffic (A_c) vs. Offered Traffic (A) for link ONU_0 – CO_1 using the first case of traffic distribution (25% of A destined to Metro_0 and 75% of A destined to CO_1).

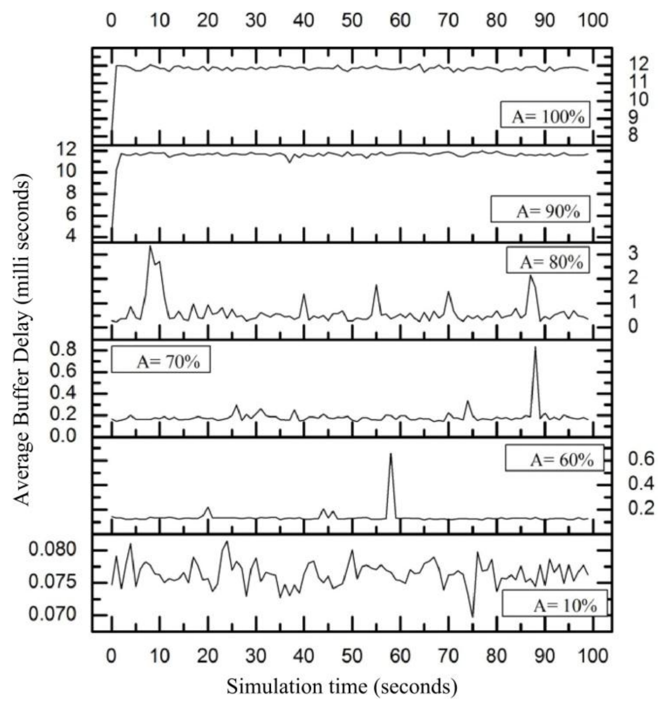


Figure 7. Average Buffer Delay for the associated passive channel of ONU_0 using the first case of traffic distribution (25% of A destined to Metro_0 and 75% of A destined to CO_1). Results are presented for five values of Offered Traffic $A = [100\%, 90\%, 80\%, 70\%, 60\%, 10\%]$.

D. WDM–PON Model Validation

For validation of developed WDM–PON model, the simple topology presented in Figure 4 was used with the same configuration parameters. Thus, allowing for model validation and performance comparison between both architectures.

1) Theoretic Analisys

In WDM–PON, a WDM channel was dedicated to each ONU, i.e., Offered Traffic perceived in average a transmission rate of $R\ bps$. Therefore, there must be no buffer overflow at ONUs even when $A = 100\%$.

The processing capacity of the nodes was in line with the transmission rate. In consequence, CO_1 was capable of processing all arriving packets, presenting a maximum for $N = 4$ ONUs of $4R\ bps$.

There must be a bottleneck at the CO_1 – Metro_0 link because there were only 2 WDM channels at $R\ bps$. Thus, at Metro_0 can arrive traffic at a maximum of $2R\ bps$. When traffic destined to Metro_0 arrived at CO_1 at more than $2R\ bps$ buffers started to overflow and the loss rate started to increase.

2) Simulation Results and Analysis

Figure 8 shows gathered results of Loss Rate (*LR*) vs. Offered Traffic (*A*). Results are presented for three cases of traffic distribution.

a) 50% of A destined to Metro_0 and 50% of A destined to CO_1

The curve associated with this case shows that there was no packet loss. As it was expected only when at CO_1 arrives traffic destined to Metro_0 at more than $2R\ bps$ there will be packet loss.

b) 75% of A destined to Metro_0 and 25% of A destined to CO_1

In this case, the bottleneck imposed by the 2 WDM channels started to produce buffer overflows for $A \geq 70\%$. For $A \geq 70\%$ the ONUs were generating traffic destined to Metro_0 at more than $2R\ bps$. For $A = 100\%$ there was a *LR* of 25% because 75% of the offered load requested the CO_1 – Metro_0 link and it only supported 50% of the offered load ($2R\ bps$).

c) 100% of A destined to Metro_0

It can be seen that for $A > 50\%$ the *LR* curve started to grow, as was expected. When 100% of generated traffic was destined to Metro_0, the 2 WDM channels of CO_1 – Metro_0 link only supported up to 50% of the load without packet loss.

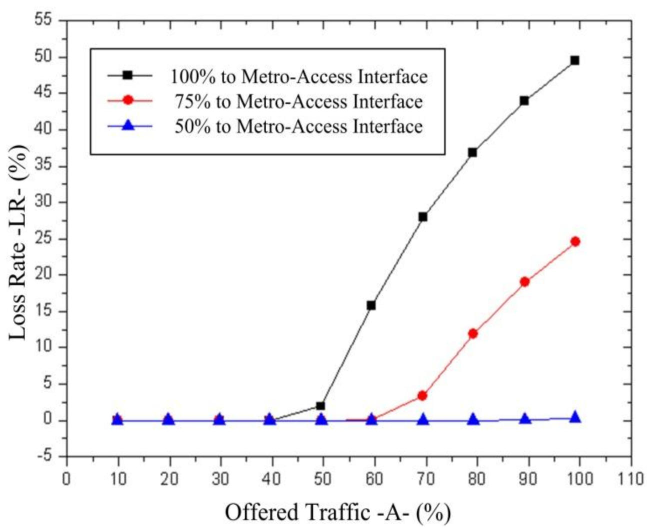


Figure 8. Loss Rate (LR) vs. Offered Traffic (A) for the WDM–PON Architecture. The three curves represents diferent traffic distributions.

E. Performance Assessment

Figure 9 presents the simple network topology used in the simulation experiments. This topology was selected in order to establish some comparison with the former evaluation [19]. It consisted of five COs with four ONUs connected to each CO and a Metro–Access Interfacing Node (MN) as major destination. The COs were connected to each other in a ring arrangement. Just one CO was connected to the MN.

1) Scenario Configuration

The set up of the performance assessment scenario was as follow. Transmission Rate was set at $R = 125\ Mbps$ (one magnitude order below EPON standard rates). The processing rate of the nodes was set to be on line with the transmission rate. Buffer sizes were assigned to limit the maximum buffer delay at $1\ msec$ in relation with the transmission rate; based on design considerations assumed in [45].

The applied traffic model had self–similar arrivals processes with Hurst parameter H ~ 0,74; based on empirical traffic evaluations [44]. The traffic distribution was the same as used in the former evaluation. It was taken from a multiple gateway traffic assessment, where 70% of A was destined to the major destination (i.e., the MN) and the rest was equally distributed among the minor destinations (i.e., the five COs) [19][23].

The t– λ frame period was set to $T_{frame} = 125\ \mu\,sec$; compatible with the Synchronous Digital Hierarchy (SDH).

Each frame is composed by four t–λ slots (same number as ONUs connected to each CO). The time slots were assigned to specific wavelengths, and the wavelengths were associated with the MN as the major destination and with the COs as minor destinations.

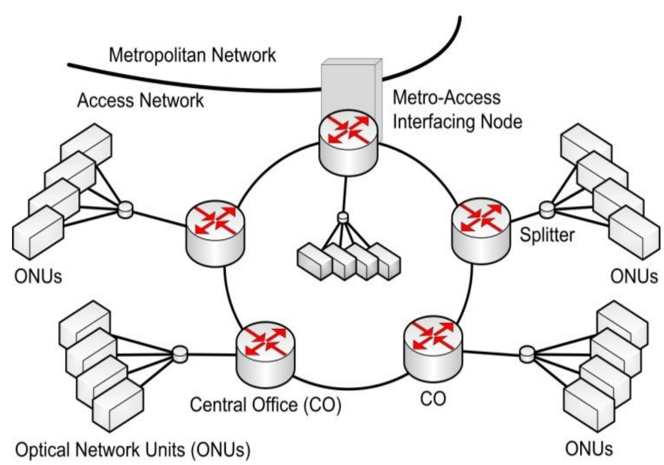


Figure 9. Simple Access network topology used for the performance evaluation, based on a ring interconection of COs. The Metro–Access interfacing Node (MN) represents the major destination (for the performance assessment up to 70% of traffic was destined to MN).

In conventional WDM–PON, each ONU had its own dedicated wavelength, which is terminated at the CO, i.e., no passive channels provided.

2) Simulation Results and Analysis

Performance assessment was carried out using a worst case electronic bottleneck scenario; i.e., up to 70% of the offered traffic was destined to the major destination (MN). As there were four ONUs per each CO (see Figure 9), the t–λ frame was composed by four time shared wavelengths. Only one wavelength was optically bypassed towards the major destination per each CO (i.e., one on–the–fly routed path established from CO to MN). In this way, each ONU perceived up to 25% of the transmission rate to send traffic through the passive channel, producing fast overload of the passive channels.

Figure 10 shows the simulation results for *LR* and *EED* vs. A for the enhanced t–λ routing architecture and the conventional WDM–PON as reference. As can be observed in Figure 10, the conventional WDM–PON has a superior performance, based on *LR*, because of the expected fast overload of passive channels in the enhanced t–λ routing architecture. However the *LR* in WDM–PON tends to worse when A increases (higher degree of congestion in the network) as a consequence of the COs' electronic bottleneck.

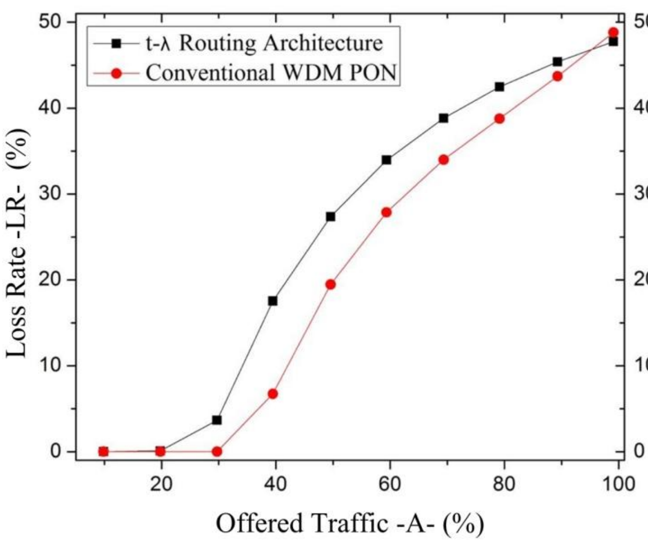


Figure 10. Loss Rate (*LR*) vs. Offered Traffic (*A*); for 70% of *A* destined to Metro Node (MN).

Figure 11 presents the *EED* experienced by the enhanced t–λ routing architecture, based on three curves: electronically routed paths, on–the–fly routed paths (optically bypassed), and overall paths. Only one WDM–PON EED curve is shown, because all packets were electronically routed in WDM–PON.

Figure 11 shows that the electronically routed paths of the enhanced t–λ routing architecture presented the lowest *EED* $\forall A$, because minor destination traffic (local traffic) perceived congestion free COs. The on–the–fly routed paths *EED* curve suggests that the ONUs passive buffers were overloaded for $A \geq 40\%$. In Figure 11 the enhanced t–λ overall EED curve indicates that when the passive buffers were overloaded ($A \geq 40\%$) there was an increasing portion of major traffic being lost, as is clearly showed in Figure 10.

The WDM–PON *LR* and *EED* curves depicted the WDM electronic bottleneck problem. Even though we have moved the bottleneck from COs to the ONUs in the enhanced t–λ architecture, producing a fast overload of the passive channels; the WDM–PON performance tends to worse when the network is highly loaded ($A \geq 70\%$). Figure 11 shows that the WDM–PON COs tend to get congested when the network load is increased. For $A \geq 70\%$ the WDM−PON *EED* becomes worse than the *EED* experienced by the enhanced t–λ architecture.

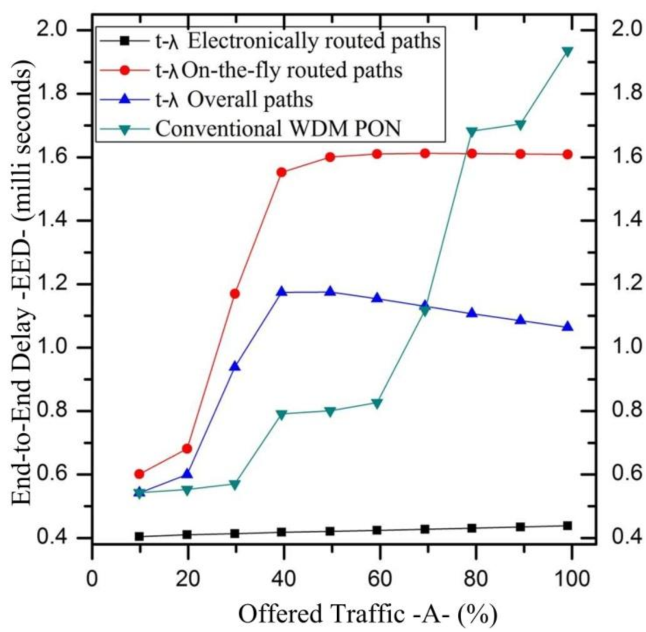


Figure 11. End–to–End Delay (*EED*) vs. Offered Traffic (*A*); for 70% of *A* destined to Metro Node (MN).

VIII. CONCLUSION AND FUTURE WORK

In a future scenario with higher users' bandwidth demand for video–related traffic (video–related representing 61% of all customer traffic for 2015) and faster networks, the access network must successfully deliver the demanded bandwidth and cope with the strict traffic requirements on low delay and jitter.

A former t–λ routing architecture model feature was found, which produces constant packet reordering and performance problems. To avoid such packet reordering and performance problems, we have proposed an enhanced version of the t–λ routing architecture, which effectively avoids the reordering packet problems. The proposed scheme was evaluated against a traditional WDM–PON architecture using developed simulation models in the OPNET Modeler tool.

Although EED on WDM–PON and t–λ routing are not very different, our simulation results showed that the WDM–PON leads to congestion at COs in presence of nonuniform access traffic distribution, as a consequence of the electronic bottleneck. In spite of the fast overload of the passive channel at ONUs, the use of the proposed enhanced t– λ routing architecture, allows the COs to remain congestion free. In consequence, the introduction of optically bypassed channels leads to congestion avoidance, thus the network can more efficiently support different traffic requirements while reducing the power consumption.

In a future proof optical access scenario, each CO must manage much more than 64 ONUs. However, the t–λ routing architecture cannot meet this requirement because it presents a limitation in the number of ONUs managed by the CO.

The t–λ frame is composed by the same number of wavelengths as ONUs connected to the CO. Because of the time division multiplexing, in average each ONU experience a transmission rate $R_i = \frac{R}{N}\,bps$. The architecture was proposed to introduce the optical bypass of only one passive channel per CO [19]. In consequence, for each ONU the optical bypassed traffic experience in average a transmission rate of $R_i = \frac{R}{N}\,bps$, while the rest of the traffic experience in average a transmission rate of $(N-1)R_i bps$. Therefore, as the number of ONUs (N) increases, the average transmission rate of the passive channel perceived by each ONU decreases. In the case of a CO with 128 ONUs the average transmission rate of the passive channel perceived by each ONU will be $R_i = \frac{R}{128}\,bps$. In such scenario, the performance of the t–λ routing architecture will be very poor, because there will be one or just few channels associated with the major destination (up to 80% of offered load destined to major destination) increasing the bottleneck at the passive channel.

Hybrid Wavelength Division Multiplexing-Passive Optical Networks (XDM–WDM–PON) seems to be the solution to tackle the requirements for optical access networks of the future. The introduction of optical bypass (on–the–fly routed) channels based on the nonuniform access traffic distribution to achieve a transparent, low latency and low power consumption Metro–Access interface could represent an interesting useful approach to consider in XDM–WDM–PON architectures. Based on our simulation results, we propose to combine the best of the t–λ routing architecture and XDM–WDM–PON. A more effective combination could be by means of hybrid OFDM –WDM–PON or OCDM–WDM–PON, in substitution of the time division multiplexing. Such hybrid WDM–PON must be able to provide transparent ONU–CO connections and transparent Metro–Access optical bypass routed paths (releasing electronic bottleneck) without restrictions to increase the number of ONUs per CO.

It would be convenient to conduct some additional experiments introducing inelastic and elastic traffic differentiation. Using traffic differentiation can assure that only inelastic traffic will be sent by the optical bypassed channels, whereas elastic traffic could be sent through electronically routed channels.

ACKNOWLEDGMENT

This work has been supported by the Applied Networks and Telematics Group at Simon Bolivar University (GRETA–USB), Venezuela, and the Metro Access and Short Range Systems Group at Technical University of Denmark, Denmark.

REFERENCES

[1] R. Alvizu, M. Huerta, R. Gonzales, R. Martinez, L. Rodríguez, and I. Tafur, "On-the-Fly Routing and Traffic Self-Aggregation Towards Metro-Access Integration," in *INFOCOMP 2011, The First International Conference on Advanced Communications and Computation*, Barcelona, Spain, 2011.

[2] *Cisco Visual Networking Index: Forecast and Methodology, 2010-2015. http://www.cisco.com . 12.12.2011.*

[3] A. Di Giglio, *et al.*, "STRONGEST: Challenges for network architectures," in *Future Network and Mobile Summit, 2010*, 2010, pp. 1-8.

[4] J. i. Kani, "Enabling Technologies for Future Scalable and Flexible WDM-PON and WDM/TDM-PON Systems," IEEE J. Sel. Topics Quantum Electron., vol. 16, pp. 1290-1297, 2010.

[5] P. E. Green, "Fiber to the home: the next big broadband thing," IEEE Commun. Mag., vol. 42, pp. 100-106, 2004.

[6] L. G. Kazovsky, W.-T. Shaw, D. Gutierrez, N. Cheng, and S.-W. Wong, "Next-Generation Optical Access Networks," J. Lightw. Technol., vol. 25, pp. 3428-3442, 2007.

[7] X. Shilin, Z. Zhao, B. Meihua, Z. Min, and C. He, "Evolution scenarios for passive optical access networks," in *International Conference on Information Communications and Signal Processing (ICICS) 2011 8th*, 2011, pp. 1-5.

[8] E. Wong, "Next-Generation Broadband Access Networks and Technologies," J. Lightw. Technol., vol. 30, pp. 597-608, 2012.

[9] *Fiber-to-the-Home (FTTH) Council website. http://www.ftthcouncil.org/. 12.12.2012.*

[10] K. Prince, *et al.*, "GigaWaM−Next-Generation WDM-PON Enabling Gigabit Per-User Data Bandwidth," J. Lightw. Technol., vol. 30, pp. 1444-1454, 2012.

[11] K. Grobe and J. P. Elbers, "PON in adolescence: from TDMA to WDM-PON," IEEE Commun. Mag., vol. 46, pp. 26-34, 2008.

[12] C. Gee-Kung, *et al.*, "Key Technologies of WDM-PON for Future Converged Optical Broadband Access Networks [Invited]," IEEE J. Opt. Commun. Netw., vol. 1, pp. C35-C50, 2009.

[13] Z. Luying, C. Xiaofei, Y. Yong-Kee, and N. Lek heng, "Hybrid WDM-TDM PON architectures and DWBA algorithms," in *Communications and Networking in China (CHINACOM), 2010 5th International ICST Conference on*, 2010, pp. 1-6.

[14] P. Ossieur, et al., "A 135-km 8192-Split Carrier Distributed DWDM-TDMA PON With 2 x 32 x 10 Gb/s Capacity," J. Lightw. Technol., vol. 29, pp. 463-474, 2011.

[15] P. Ossieur, et al., "Demonstration of a 32 x 512 Split, 100 km Reach, 2x 32 x 10 Gb/s Hybrid DWDM-TDMA PON Using Tunable External Cavity Lasers in the ONUs," J. Lightw. Technol., vol. 29, pp. 3705-3718, 2011.

[16] N. Cvijetic, "OFDM for Next-Generation Optical Access Networks," J. Lightw. Technol., vol. 30, pp. 384-398, 2012.

[17] K. Fouli and M. Maier, "OCDMA and Optical Coding: Principles, Applications, and Challenges [Topics in Optical Communications]," IEEE Commun. Mag., vol. 45, pp. 27-34, 2007.

[18] J. i. Kani, *et al.*, "Next-generation PON-part I: Technology roadmap and general requirements," IEEE Commun. Mag., vol. 47, pp. 43-49, 2009.

[19] M. Wieckowski, A. V. Osadchiy, J. P. Turkiewicz, and I. T. Monroy, "Performance assessment of flexible time-wavelength routing for a self-aggregating transparent Metro-access interface," in *35th European Conference on Optical Communication (ECOC)* 2009, pp. 1-2.

[20] B. Mukherjee, *Optical WDM Networks*: Springer, 2006.
[21] *Metro Ethernet services - a technical overview. http://metroethernetforum.org/PDF_Documents/metro-ethernet-services.pdf. 12.12.2012*.
[22] F. Davik, M. Yilmaz, S. Gjessing, and N. Uzun, "IEEE 802.17 resilient packet ring tutorial," IEEE Commun. Mag., vol. 42, pp. 112-118, 2004.
[23] P.-L. Tsai and C.-L. Lei, "Analysis and evaluation of a multiple gateway traffic-distribution scheme for gateway clusters," Comput. Commun., vol. 29, pp. 3170-3181, 2006.
[24] E. Van Breusegern, *et al.*, "Overspill routing in optical networks: a true hybrid optical network design," IEEE J. Sel. Areas Commun., vol. 24, pp. 13-25, 2006.
[25] A. V. Osadchiy, J. B. Jensen, P. Jeppesen, and I. T. Monroy, "Colorless receiver enabling crossconnect based metro-access interfacing nodes for optically labelled DQPSK payload signals," in *21st Annual Meeting of the IEEE Lasers and Electro-Optics Society (LEOS)*, 2008, pp. 612-613.
[26] S. Chatzi, J. A. Lazaro, J. Prat, and I. Tomkos, "A quantitative techno-economic comparison of current and next generation metro/access converged optical networks," in *36th European Conference and Exhibition on Optical Communication (ECOC)*, 2010, pp. 1-3.
[27] J. Berthold, A. A. M. Saleh, L. Blair, and J. M. Simmons, "Optical Networking: Past, Present, and Future," J. Lightw. Technol., vol. 26, pp. 1104-1118, 2008.
[28] C. Zhang, W. Zhao, and X. Yi, "100Gb/s technology and future development in China," in *OptoeElectronics and Communications Conference (OECC), 2011 16th*, 2011, pp. 169-171.
[29] E. Ip, *et al.*, "100G and Beyond Transmission Technologies for Evolving Optical Networks and Relevant Physical-Layer Issues," 100G and Beyond Transmission Technologies for Evolving Optical Networks and Relevant Physical-Layer Issues, vol. 100, pp. 1065-1078, 2012.
[30] R. Van Nee and R. Prasad, *OFDM for Wireless Multimedia Communications*: Artech House, 2000.
[31] N. Cvijetic, M.-F. Huang, E. Ip, Y.-K. Huang, D. Qian, and T. Wang, "1.2 Tb/s Symmetric WDM-OFDMA-PON over 90km Straight SSMF and 1:32 Passive Split with Digitally-Selective ONUs and Coherent Receiver OLT," in *Optical Fiber Communication Conference and Exposition (OFC/NFOEC), and the National Fiber Optic Engineers Conference* 2011, pp. 1-3.
[32] G. Wellbrock, "Preparing for the future," in *36th European Conference and Exhibition on Optical Communication (ECOC)*, 2010, pp. 1-10.
[33] A. Di Giglio, *et al.*, "STRONGEST: Challenges for network architectures," in *Future Network and Mobile Summit*, 2010, pp. 1-8.
[34] F. J. Effenberger, J. i. Kani, and Y. Maeda, "Standardization trends and prospective views on the next generation of broadband optical access systems," IEEE J. Sel. Areas Commun., vol. 28, pp. 773-780, 2010.
[35] S. Huan, K. Byoung-Whi, and B. Mukherjee, "Long-reach optical access networks: A survey of research challenges, demonstrations, and bandwidth assignment mechanisms," IEEE Communications Surveys & Tutorials, vol. 12, pp. 112-123, 2010.
[36] P. Chanclou, A. Cui, F. Geilhardt, H. Nakamura, and D. Nesset, "Network operator requirements for the next generation of optical access network," IEEE Netw., vol. 26, pp. 8-14, 2012.
[37] N. Cvijetic, et al., "1.92Tb/s coherent DWDM-OFDMA-PON with no high-speed ONU-side electronics over 100km SSMF and 1:64 passive split," Opt. Express, vol. 19, pp. 24540-24545, 2011..
[38] N. Cvijetic, et al., "Terabit Optical Access Networks Based on WDM-OFDMA-PON," J. Lightw. Technol., vol. 30, pp. 493-503, 2012.
[39] N. Kataoka, N. Wada, G. Cincotti, and K. Kitayama, "2.56 Tbps (40-Gbps x 8-wavelength x 4-OC x 2-POL) asynchronous WDM-OCDMA-PON using a multi-port encoder/decoder," in 37th European Conference and Exhibition on Optical Communication (ECOC), 2011, pp. 1-3.
[40] A. A. Ortega, V. A. Bettachini, J. I. Alvarez-Hamelin, and D. F. Grosz, "Point-to-point and Point-to-multipoint CDMA Access Network with Enhanced Security," 2011, p. ATuB6.
[41] F. Payoux, P. Chanclou, and N. Genay, "WDM-PON with colorless ONUs," in *Optical Fiber Communication and the National Fiber Optic Engineers Conference, 2007. OFC/NFOEC 2007. Conference on*, 2007, pp. 1-3.
[42] L. Ka-Cheong, V. O. K. Li, and Y. Daiqin, "An Overview of Packet Reordering in Transmission Control Protocol (TCP): Problems, Solutions, and Challenges," IEEE Trans. Parallel Distrib. Syst., vol. 18, pp. 522-535, 2007.
[43] *OPNET Technologies Inc. http://www.opnet.com/. 12.12.2012*
[44] W. E. Leland, M. S. Taqqu, W. Willinger, and D. V. Wilson, "On the self-similar nature of Ethernet traffic (extended version)," IEEE/ACM Trans. Netw., vol. 2, pp. 1-15, 1994.
[45] B. Skubic, C. Jiajia, J. Ahmed, L. Wosinska, and B. Mukherjee, "A comparison of dynamic bandwidth allocation for EPON, GPON, and next-generation TDM PON," IEEE Commun. Mag, vol. 47, pp. S40-S48, 2009.

Modeling and Simulation of Bacterial Self-Organization in Circular Container Along Contact Line as Detected by Bioluminescence Imaging

Romas Baronas, Žilvinas Ledas
Faculty of Mathematics and Informatics
Vilnius University
Naugarduko 24, LT-03225 Vilnius, Lithuania
Emails: romas.baronas@mif.vu.lt, zilvinas.ledas@mif.vu.lt

Remigijus Šimkus
Institute of Biochemistry
Vilnius University
Mokslininku 12, LT-08662 Vilnius, Lithuania
Email: remigijus.simkus@bchi.vu.lt

Abstract*—Mathematical modeling and numerical simulation of quasi-one dimensional spatiotemporal pattern formation along the three phase contact line in the fluid cultures of lux-gene engineered *Escherichia coli* is investigated in this paper. The numerical simulation is based on a one-dimensional-in-space mathematical model of a bacterial self-organization as detected by quasi-one-dimensional bioluminescence imaging. The pattern formation in a luminous *E. coli* colony was mathematically modeled by the nonlinear reaction-diffusion-chemotaxis equations. The numerical simulation was carried out using the finite difference technique. Regular oscillations as well as chaotic fluctuations similar to the experimental ones were computationally simulated. The influence of the signal-dependent as well as density-dependent chemotactic sensitivity, the non-local sampling and the diffusion nonlinearity on the pattern formation was investigated. The computational simulations showed that a constant chemotactic sensitivity, a local sampling and a linear diffusion can be applied for modeling the formation of the bioluminescence patterns in a colony of luminous *E. coli.

***Keywords*-chemotaxis; reaction-diffusion; pattern formation; simulation; whole-cell biosensor.**

I. Introduction

This paper is an extension of work originally reported in The Third International Conference on Advances in System Simulation [1].

Various microorganisms respond to certain chemicals found in their environment by migrating towards higher (chemoattraction) or lower (chemorepulsion) concentrations of the substance. The directed movement of microorganisms in response to chemical gradients is called chemotaxis [2]. Chemotaxis plays a crucial role in a wide range of biological phenomena, e.g., within the embryo, the chemotaxis affects avian gastrulation and patterning of the nervous system [3]. Often, microorganisms not only move up chemical gradients towards a chemoattractant, but they are also able to produce more of the chemoattractant. This is the effect that produces the aggregation of the motile microorganisms into local clusters with high density and hence results in a pattern formation [4].

Although the chemotaxis has been observed in many bacterial species, *Escherichia coli* is one of the mostly studied examples. *E. coli* responds to the chemical stimulus by alternating the rotational direction of their flagella [2], [3].

Various mathematical models based on the Patlak-Keller-Segel model have been successfully used as important tools to study the mechanisms of the chemotaxis [5]. An excellent review on the mathematical modeling of the chemotaxis has been presented by Hillen and Painter [6].

Bacterial species including *E. coli* have been observed to form various patterns under different environmental conditions [4], [7], [8]. Bacterial populations are capable of self-organization into states exhibiting strong inhomogeneities in density [9], [10]. Recently, the spatiotemporal patterns in the fluid cultures of *E. coli* have been observed by employing lux-gene engineered cells and a bioluminescence imaging technique [11], [12]. However, the mechanisms governing the formation of bioluminescence patterns still remain unclear.

Over the last two decades, lux-gene engineered bacteria have been successfully used to develop whole cell-based biosensors [13]. A whole-cell biosensor is an analyte probe consisting of a biological element, such as a genetically engineered bacteria, integrated with an electronic component to yield a measurable signal [14]. Whole-cell biosensors have been successfully used for the detection of environmental pollutant bioavailability, various stressors, including dioxins, endocrine-disrupting chemicals, and ionizing radiation [15]. To solve the problems currently limiting the practical use of whole-cell biosensors, the bacterial self-organization within the biosensors have to be comprehensively investigated.

In this paper, the bacterial self-organization in a small circular container near the three phase contact line is investigated [11], [12]. A computational model for efficient simulating the formation of the spatiotemporal patterns experimentally detected by quasi-one-dimensional bioluminescence imaging in the fluid cultures of *E. coli* has recently been developed [16], [17]. The pattern formation in a luminous *E. coli* colony was modeled by the nonlinear reaction-diffusion-chemotaxis equations. The mathematical model was formulated in a one-dimensional space. Several different model variations were analyzed, and a minimal model was obtained for simulating the formation of the bioluminescence patterns representing the self-organization of luminous *E. coli*.

The aim of this work was to improve the already existing computational model by introducing the nonlinear diffusion of cells, the non-local sampling and several kinds of the chemotactic sensitivity [6]. By extending the model in this way, the improvements of the patterns simulated using extended model were expected. In this paper, the pattern formation is computationally investigated assuming two kinds of the chemotactic sensitivity, the signal-dependent sensitivity and the density-dependent sensitivity. The non-local sampling and the nonlinear diffusion are investigated individually and collectively. The numerical simulation at transient conditions was carried out using the finite difference technique [18]. The computational model was validated by experimental data. Regular oscillations as well as chaotic fluctuations similar to experimental ones were computationally simulated. By varying the input parameters the output results were analyzed with a special emphasis on the influence of the model parameters on the spatiotemporal pattern formation in the luminous *E. coli* colony.

The rest of the paper is organized as follows. Section II provides a state of the art on the mathematical modeling of bacterial self-organization. Section III describes the mathematical model of the bacterial self-organization in a circular container. The computational modeling of a physical experiment is discussed in Section IV. Section V is devoted to the results of the numerical simulation where the effects of different chemotactic sensitivity functions, the non-local gradient and the diffusion nonlinearity are investigated. Finally, the main conclusions are summarized in Section VI.

II. Models of Bacterial Self-Organization

Different mathematical models based on advection-reaction-diffusion equations have already been developed for computational modeling the pattern formation in bacterial colonies [7], [8], [19], [20], [21]. The system of coupled nonlinear partial differential equations introduced by Keller and Segel are still among the most widely used [5], [6].

According to the Keller and Segel approach, the main biological processes can be described by a system of two conservation equations ($x \in \Omega,\ t > 0$),

$$\begin{aligned} \frac{\partial n}{\partial t} &= \nabla \left(D_n \nabla n - h(n,c) n \nabla c \right) + f(n,c), \\ \frac{\partial c}{\partial t} &= \nabla \left(D_c \nabla c \right) + g_p(n,c) n - g_d(n,c) c, \end{aligned} \tag{1}$$

where x and t stand for space and time, $n(x,t)$ is the cell density, $c(x,t)$ is the chemoattractant concentration, $D_n(n)$ and D_c are the diffusion coefficients, $f(n,c)$ stands for cell growth and death, $h(n,c)$ stands for the chemotactic sensitivity, g_p and g_d describe the production and degradation of the chemoattractant [5], [21].

Both diffusion coefficients, D_n and D_c, are usually assumed to be constant. However, the nonlinear cell diffusion depending on the chemoattractant concentration or/and the cell density is also considered [6]. In this work, we consider the nonlinear diffusion of the form

$$D_n(n) = D_n \left(\frac{n}{n_0} \right)^m, \tag{2}$$

where and below n_0 is the maximal density ("carrying capacity") of the cell population ($n < n_0$) [22]. At $m < 0$ the rate of diffusion increases with increasing the cell density, while at $m > 0$ the rate decreases with increasing the cell density. Accepting $m = 0$ leads to a constant rate of the cell diffusion. Since the proper form of the diffusion coefficient D_n to be used for the simulation of the spatiotemporal pattern formation in the fluid cultures of lux-gene engineered *E. coli* is unknown, the simulation was performed at different vales of m.

The cell growth $f(n,c)$ is usually assumed to be a logistic function,

$$f(n,c) = k_1 n \left(1 - \frac{n}{n_0} \right), \tag{3}$$

where k_1 is the constant growth rate of the cell population [7].

Various chemoattractant production functions have been used in chemotactic models [6]. Usually, a saturating function of the cell density is used indicating that, as the cell density increases, the chemoattractant production decreases. The Michaelis-Menten function is widely used to express the production rate g_p [5], [20], [23],

$$g_p(n,c) = \frac{k_2}{k_3 + n}. \tag{4}$$

The degradation or consumption g_d of the chemoattractant is typically constant,

$$g_d(n,c) = k_4. \tag{5}$$

Values of k_2, k_3 and k_4 are not exactly known yet [21].

The function $h(n,c)$ controls the chemotactic response of the cells to the chemoattractant. The signal-dependent sensitivity and the density-dependent sensitivity are two main kinds of the chemotactic sensitivity $h(n,c)$ [6]. In order to reproduce the experimentally observed bands Keller and Segel introduced a chemotactic (signal-dependent) sensitivity of the following form [24]:

$$h(n,c) = \frac{k_5}{c}. \tag{6}$$

Since the bacterial current flow declines at low chemical concentrations and saturates at high concentrations, Lapidus and Schiller derived the "receptor" chemotactic (signal-dependent) sensitivity for *E. coli* [19],

$$h(n,c) = \frac{k_6}{(k_7+c)^2}. \quad (7)$$

Assuming that cells carry a certain finite volume, a density-dependent chemotactic sensitivity function as well as volume-filling model were derived by Hillen and Painter [25],

$$h(n,c) = k_8\left(1-\frac{n}{n_0}\right). \quad (8)$$

Another form for the density-dependent chemotactic sensitivity has been introduced by Velazquez [26],

$$h(n,c) = \frac{k_9}{k_{10}+n}. \quad (9)$$

In the simplest form, the chemotactic sensitivity is assumed to be independent of the chemoattractant concentration c as well as the cell density n, i.e., $h(n,c)$ is constant, $h(n,c) = k_8$. Since the proper form of the chemotactic sensitivity function $h(n,c)$ to be used for the simulation of the spatiotemporal pattern formation in the fluid cultures of lux-gene engineered *E. coli* remains unknown, all these four forms of the function $h(n,c)$ were used to find out the most useful form.

E. coli is able to detect a gradient by sampling the chemoattractant concentration over the time and adjusting their movement accordingly. As a result, the signal detected by the cell is non-local and the non-local gradient can be used to model this behaviour [27], [28],

$$\overset{\circ}{\nabla}_\rho\, c(x,t) = \frac{n}{|S^{n-1}|\,\rho}\int_{S^{n-1}} \sigma c(x+\rho\sigma,t)d\sigma, \quad (10)$$

where S^{n-1} denotes the $(n-1)$-dimentional unit sphere in $\mathbb{R}^n$ and ρ is the sampling radius. When $\rho \to 0$, this model collapses to the ordinary model with local sampling.

It was recently shown that the Keller-Segel approach can be applied to the simulation of the formation of the spatiotemporal patterns experimentally detected by bioluminescence imaging in the fluid cultures of *E. coli* [16], [17]. This work aims to improve the already existing computational model by introducing the nonlinear diffusion (2) of cells, the non-local sampling (10) and different kinds of the chemotactic sensitivity (6)-(9). The improvement of the patterns simulated using the extended model was expected.

III. Model for Luminous *E. Coli*

When modeling the self-organization of luminous *E. Coli* in a circular container along the three phase contact line [11], [12], the mathematical model can be defined in one spatial dimension - on the circumference of the vessel [16], [17].

A. Governing Equations

Replacing f, g_p, g_d, D_n and ∇c with the concrete expressions above, the governing equations (1) reduce to a cell kinetics model with the nonlinear signal kinetics, the nonlinear cell diffusion, the nonlinear chemotactic sensitivity and the non-local sampling,

$$\begin{aligned} \frac{\partial n}{\partial t} &= D_n \nabla\left(\left(\frac{n}{n_0}\right)^m \nabla n\right) - \\ &\quad - \nabla\left(h(n,c) n \overset{\circ}{\nabla}_\rho\, c\right) + k_1 n\left(1-\frac{n}{n_0}\right), \quad (11) \\ \frac{\partial c}{\partial t} &= D_c \Delta c + \frac{k_2 n}{k_3+n} - k_4 c, \quad x \in (0,l), \quad t>0, \end{aligned}$$

where Δ is the Laplace operator formulated in the one-dimensional Cartesian coordinate system, and l is the length of the contact line, i.e., the circumference of the vessel. Assuming R as the vessel radius, $l = 2\pi R$, $x \in (0, 2\pi R)$.

B. Initial and Boundary Conditions

A non-uniform initial distribution of cells and zero concentration of the chemoattractant are assumed,

$$\begin{aligned} &n(x,0) = n_{0x}(x), \\ &c(x,0) = 0, \quad x \in [0,l], \end{aligned} \quad (12)$$

where $n_{0x}(x)$ stands for the initial ($t=0$) spatially-varying cell density.

For the bacterial simulation on a continuous circle of the length l of the circumference, the following periodicity conditions are applied as the boundary (matching) conditions ($t>0$):

$$\begin{aligned} n(0,t) = n(l,t), \quad \left.\frac{\partial n}{\partial x}\right|_{x=0} = \left.\frac{\partial n}{\partial x}\right|_{x=l}, \\ c(0,t) = c(l,t), \quad \left.\frac{\partial c}{\partial x}\right|_{x=0} = \left.\frac{\partial c}{\partial x}\right|_{x=l}. \end{aligned} \quad (13)$$

C. Dimensionless Model

In order to define the main governing parameters of the mathematical model (11)-(13), a dimensionless mathematical model has been derived by introducing the following dimensionless parameters [4], [6], [23]:

$$\begin{aligned} &u = \frac{n}{n_0}, \quad v = \frac{k_3 k_4 c}{k_2 n_0}, \\ &t^* = \frac{k_4 t}{s}, \quad x^* = \sqrt{\frac{k_4}{D_c s}}x, \\ &D = \frac{D_n}{D_c}, \quad r = \frac{k_1}{k_4}, \quad \phi = \frac{n_0}{k_3}, \quad \rho^* = \frac{\rho}{l}, \\ &\chi(u,v) = \frac{k_2 n_0}{k_3 k_4 D_c} h(n_0 u, k_2 n_0 c/(k_3 k_4)). \end{aligned} \quad (14)$$

Dropping the asterisks, the dimensionless governing equations then become ($t>0$)

$$\frac{\partial u}{\partial t} = \frac{\partial}{\partial x}\left(Du^m \frac{\partial u}{\partial x}\right) - \frac{\partial}{\partial x}\left(\chi(u,v)u \overset{\circ}{\nabla}_\rho v\right) + \\ + sru(1-u), \tag{15}$$

$$\frac{\partial v}{\partial t} = \frac{\partial^2 v}{\partial x^2} + s\left(\frac{u}{1+\phi u} - v\right), \quad x \in (0,1),$$

where x and t stand for the dimensionless space and time, respectively, u is the dimensionless cell density, v is the dimensionless chemoattractant concentration, r is the dimensionless growth rate of the cell population, ϕ stands for saturating of the signal production, $\chi(u, v)$ is the dimensionless chemotactic sensitivity, and s stands for the spatial and temporal scale.

Assuming the one-dimensional Cartesian coordinate system the non-local gradient can be described as

$$\overset{\circ}{\nabla}_\rho v(x,t) = \frac{v(x+\rho,t) - v(x-\rho,t)}{2\rho}. \tag{16}$$

For the dimensionless simulation of the spatiotemporal pattern formation in a luminous *E. coli* colony, four forms of the chemotactic sensitivity function $\chi(u, v)$ were used to find out the best fitting pattern for the experimental data [11], [12], [16],

$$\chi(u,v) = \frac{\chi_0}{(1+\alpha v)^2}, \tag{17a}$$

$$\chi(u,v) = \chi_0 \frac{1+\beta}{v+\beta}, \tag{17b}$$

$$\chi(u,v) = \chi_0\left(1 - \frac{u}{\gamma}\right), \tag{17c}$$

$$\chi(u,v) = \frac{\chi_0}{1+\epsilon u}. \tag{17d}$$

The first two forms (17a) and (17b) of the function $\chi(u, v)$ correspond to the signal-dependent sensitivity, while the other two (17c) and (17d) - for the density-dependent sensitivity [6]. Accepting $\alpha = 0$, $\beta \to \infty$, $\gamma \to \infty$ or $\epsilon = 0$ leads to a constant form of the chemotactic sensitivity, $\chi(u, v) = \chi_0$.

The initial conditions (12) take the following dimensionless form:

$$u(x,0) = 1 + \varepsilon(x), \\ v(x,0) = 0, \quad x \in [0,1], \tag{18}$$

where $\varepsilon(x)$ is a random spatial perturbation.

The boundary conditions (13) transform to the following dimensionless equations ($t > 0$):

$$u(0,t) = u(1,t), \quad \frac{\partial u}{\partial x}\Big|_{x=0} = \frac{\partial u}{\partial x}\Big|_{x=1}, \\ v(0,t) = c(1,t), \quad \frac{\partial v}{\partial x}\Big|_{x=0} = \frac{\partial v}{\partial x}\Big|_{x=1}. \tag{19}$$

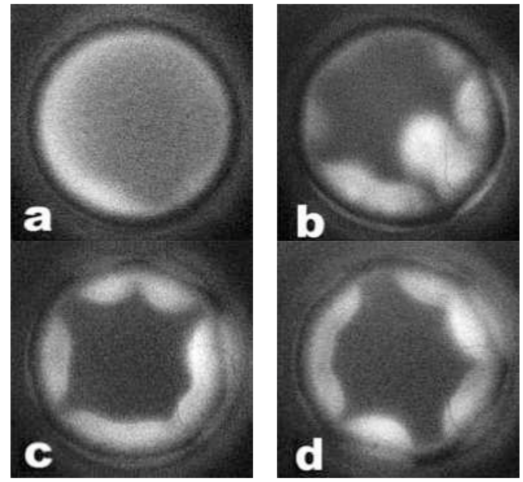


Figure 1. Top view bioluminescence images of the bacterial cultures in the cylindrical glass vessel. The images were captured at 5 (a), 20 (b), 40 (c), 60 (d) min [12].

According to the classification of chemotaxis models, the dimensionless model of the pattern formation is a combination of the signal-dependent sensitivity (M2), the density-dependent sensitivity (M3), the non-local sampling (M4), the nonlinear diffusion (M5), the saturating signal production (M6) and the cell kinetics (M8) models [6]. At certain values of the model parameters the dimensionless model (15), (18) and (19) reduces to the minimal model (M1) for the chemotaxis [6].

IV. Numerical Simulation

The mathematical model (11)-(13), as well as the corresponding dimensionless model (15), (18), (19), has been defined as an initial boundary value problem based on a system of nonlinear partial differential equations. No analytical solution is possible because of the nonlinearity of the governing equations of the model [4]. Hence the bacterial self-organization was simulated numerically.

The numerical simulation was carried out using the finite difference technique [18]. To find a numerical solution of the problem a uniform discrete grid with 760 points and the dimensionless step size $1/760$ (dimensionless units) in the space direction was introduced, $760 \times 1/760 = 1$. A constant dimensionless step size 2.5×10^{-7} was also used in the time direction. An explicit finite difference scheme has been built as a result of the difference approximation [17], [18], [29], [30]. The digital simulator has been programmed by the authors in Free Pascal language [31].

The computational model was applied to the simulation of bioluminescence patterns observed in a small circular containers made of glass [12], [16]. Figure 1 shows typical top view bioluminescence images of bacterial cultures illustrating an accumulation of luminous bacteria near the

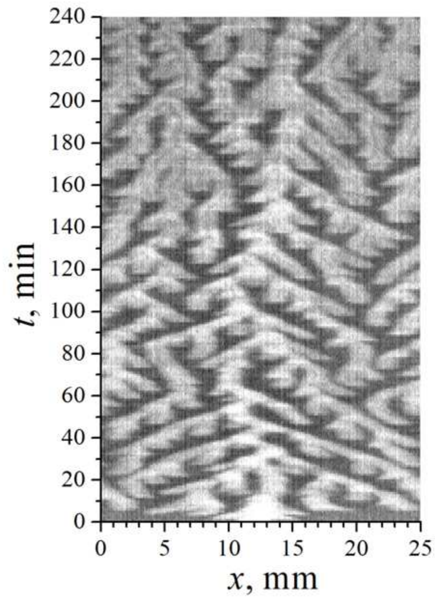


Figure 2. Space-time plot of bioluminescence measured along the contact line of the cylindrical vessel [12], [16].

contact line. The images were captured at different times of the population evolution.

In general, the dynamic processes in unstirred cultures are rather complicated and need to be modeled in three dimensional space [2], [11], [12]. Since luminous cells concentrate near the contact line, the three-dimensional processes were simulated in one dimension (quasi-one dimensional rings in Figure 1). Figure 2 shows the corresponding space-time plot of quasi-one-dimensional bioluminescence intensity.

By varying the model parameters the simulation results were analyzed with a special emphasis to achieving a spatiotemporal pattern similar to the experimentally obtained pattern shown in Figure 2. Figure 3 shows the results of the informal pattern fitting, where Figures 3a and 3b present the simulated space-time plots of the dimensionless cell density u and the chemoattractant concentration v, respectively. The corresponding values $\bar{u}$ and $\bar{v}$ averaged on the circumference of the vessel are depicted in Figure 3c,

$$\bar{u}(t) = \int_0^1 u(x,t)\,dx, \qquad \bar{v}(t) = \int_0^1 v(x,t)\,dx\,. \tag{20}$$

Regular oscillations as well as chaotic fluctuations similar to the experimental ones were computationally simulated. Accepting the constant form of the chemotactic sensitivity ($\chi(u,v) = \chi_0$) and the simple gradient, the dynamics of the bacterial population was simulated at the following values of the model parameters [16]:

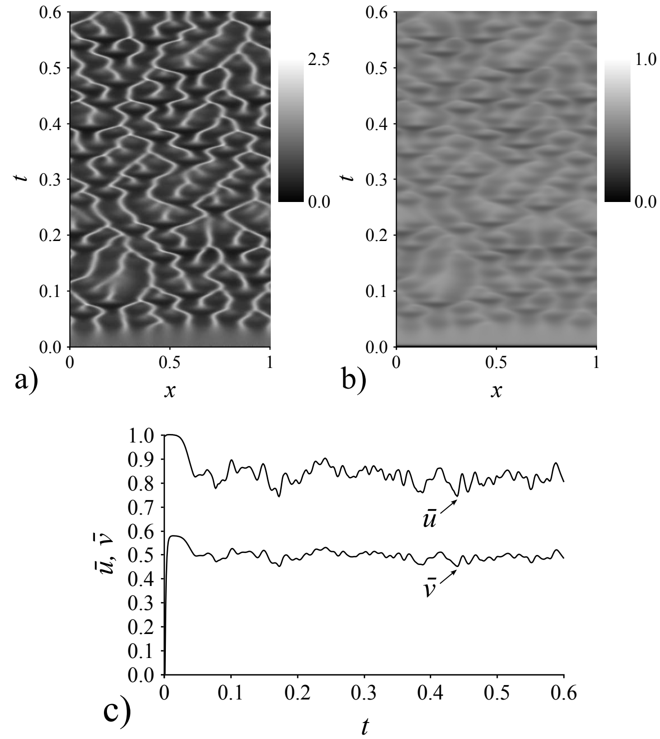


Figure 3. Simulated space-time plots of the dimensionless cell density u (a) as well as the chemoattractant concentration v (b) and the dynamics of the corresponding averaged values $\bar{u}$ and $\bar{v}$ (c). Values of the parameters are as defined in (21).

$$D = 0.1, \quad \chi_0 = 6.2, \quad \rho = 0, \quad r = 1, \qquad \phi = 0.73, \quad s = 625, \quad m = 0. \tag{21}$$

A spatially-varying random perturbation $\varepsilon(x)$ of the dimensionless cell density u of 10% was applied for the initial distribution of bacteria near the three phase contact line when simulating the spatiotemporal patterns.

Due to a relatively great number of model parameters, there is no guarantee that the values (21) mostly approach the pattern shown in Figure 2. Similar patterns were achieved at different values of the model parameters. The linearization and the stability analysis of homogenous solutions of the Keller-Segel model showed similar effects [32], [33]. An increase in one parameter can be often compensated by decreasing or increasing another one. Because of this, it is important to investigate the influence of the model parameters on the pattern formation and to develop a mathematical model containing a minimal number of parameters and ensuring a qualitative analysis of bacterial pattern formation in a liquid medium [6], [10], [17], [21].

V. Results and Discussion

By varying the input parameters the output results were analyzed with a special emphasis on the influence of the

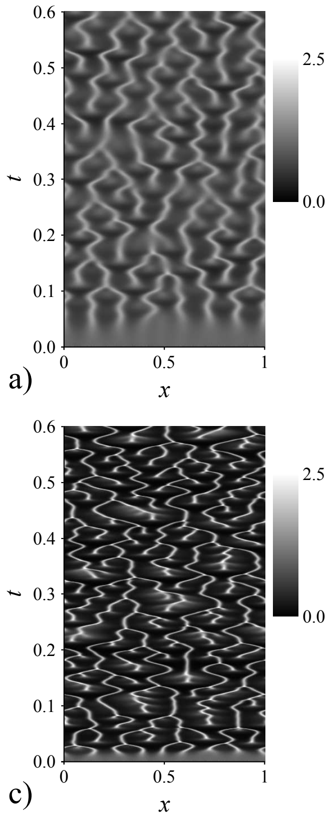


Figure 4. Spatiotemporal plots of the dimensionless cell density u for two forms of the signal-dependent chemotactic sensitivity $\chi(u, v)$: (17a) ($\alpha = 0.05$) (a), ($\alpha = 0.07$) (b) and (17b) ($\beta = 2$) (c), ($\beta = 10$) (d). Values of the other parameters are as defined in (21).

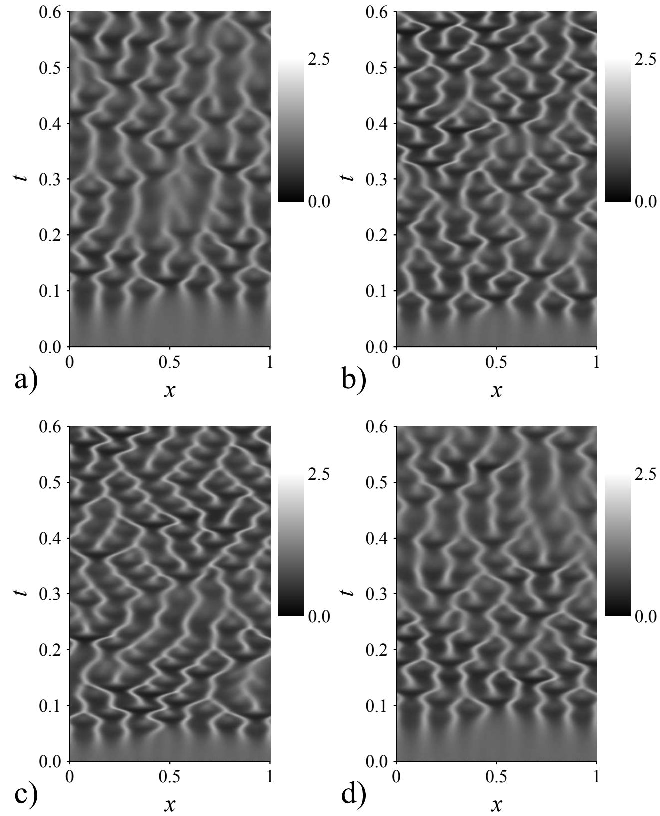


Figure 5. Spatiotemporal plots of the dimensionless cell density u for two forms of the density-dependent chemotactic sensitivity $\chi(u, v)$: (17c) ($\gamma = 10$) (a), ($\gamma = 15$) (b) and (17d) ($\epsilon = 0.05$) (c), ($\epsilon = 0.1$) (d). Values of the other parameters are as defined in (21).

chemotactic sensitivity, the non-local gradient and the diffusion nonlinearity on the spatiotemporal pattern formation in the luminous *E. coli* colony. Figure 3a shows the spatiotemporal pattern for the constant form of the chemotactic sensitivity ($\chi(u, v) = \chi_0$) applying the simple gradient ($\rho = 0$) and the linear diffusion ($m = 0$).

The effects of the different chemotactic sensitivity functions were investigated assuming the linear diffusion ($m = 0$) and the simple gradient ($\rho \to 0$). The non-local gradient and the nonlinear diffusion was analyzed separately and together assuming the constant chemotactic sensitivity.

A. The Effect of the Signal-Dependent Sensitivity

The signal-dependent sensitivity was computationally modeled by two forms of the chemotactic sensitivity function $\chi(u, v)$: (17a) and (17b). The spatiotemporal patterns of the dimensionless cell density u were simulated at very different values of α and β. Figure 4 shows the effect of the signal-dependence of the chemotactic sensitivity on the pattern formation.

Accepting $\alpha = 0$ or $\beta \to \infty$ leads to a signal-independence, i.e., a constant form, of the chemotactic sensitivity, $\chi(u, v) = \chi_0$. Results of the multiple simulations showed that the simulated patterns distinguish from the experimental one (Figure 2) when increasing α-parameter (Figures 4a and 4b) or decreasing β-parameter (Figures 4c and 4d). Because of this, there is no practical reason for application of a non-constant form of the signal-dependent sensitivity to modeling the formation of the bioluminescence patterns in a colony of luminous *E. coli.* Consequently, the signal-dependence of the chemotactic sensitivity can be ignored when modeling the pattern formation in the luminous *E. coli* colony.

B. The Effect of the Density-Dependent Sensitivity

Two forms, (17c) and (17d), of the chemotactic sensitivity function $\chi(u, v)$ were employed for computational modeling of the density-dependent chemotactic sensitivity. The spatiotemporal patterns of the cell density u were simulated at various values of γ and ϵ. Figure 5 shows how the density-dependence affects the pattern formation.

Accepting $\gamma \to \infty$ or $\epsilon = 0$ leads to a density-independence, i.e., a constant form, of the chemotactic sensitivity, $\chi(u, v) = \chi_0$. Multiple simulation showed that the simulated patterns distinguish from the experimental one (Figure 2) when decreasing γ-parameter (Figures 5a and 5b)

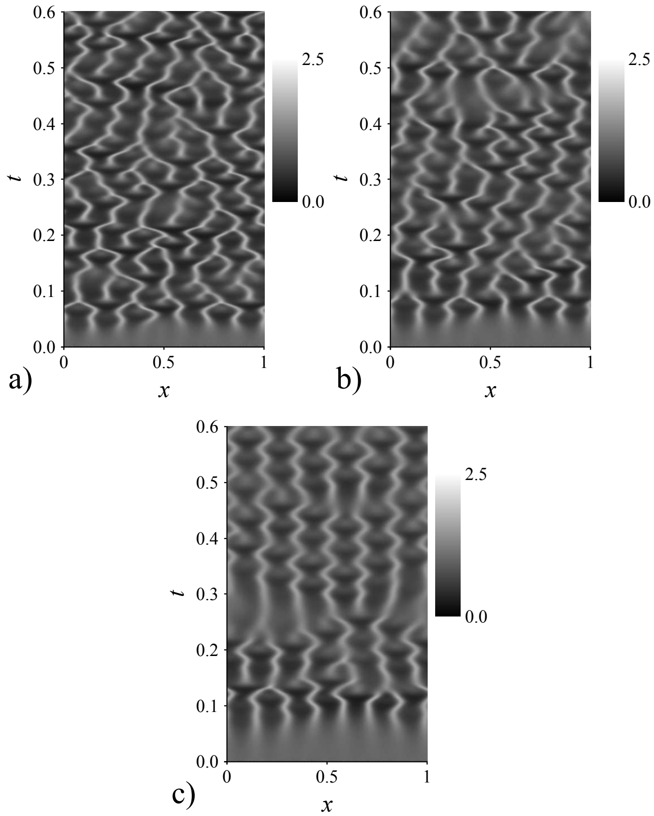


Figure 6. Spatiotemporal plots of the dimensionless cell density u when using the non-local sampling ($\rho = 0.008$) (a), ($\rho = 0.012$) (b), ($\rho = 0.016$) (c). Values of the other parameters are as defined in (21).

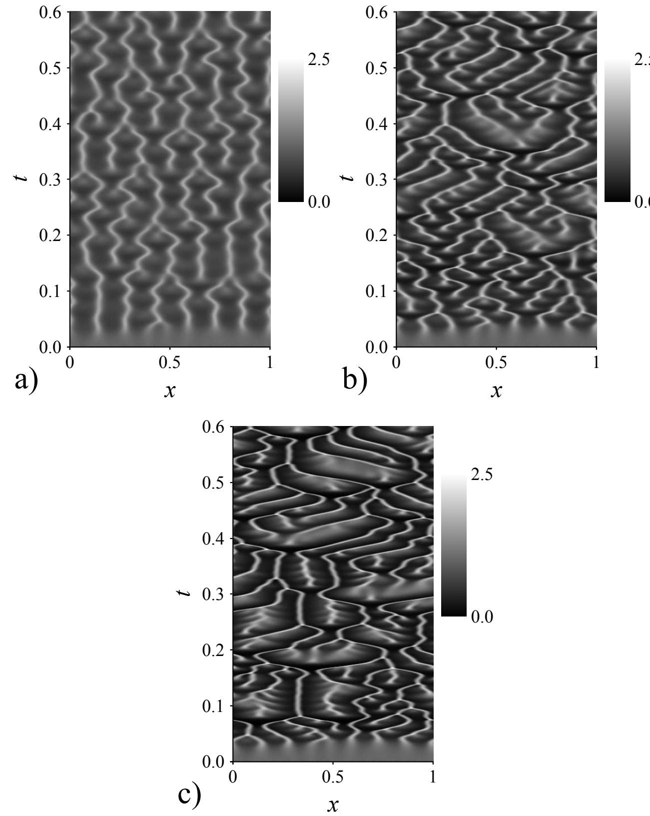


Figure 7. Spatiotemporal plots of the dimensionless cell density u when using the nonlinear diffusion ($m = -0.6$) (a), ($m = 0.2$) (b), ($m = 0.6$) (c). Values of the other parameters are as defined in (21).

or increasing ϵ-parameter (Figures 5c and 5d). Because of this, similarly to the signal-dependent chemotactic sensitivity, there is no practical reason for application of a non-constant form also of the density-dependent sensitivity when modeling the pattern formation in a colony of luminous *E. coli.*

A simple constant form ($\chi(u, v) = \chi_0$) of the chemotactic sensitivity can be successfully applied to modeling the formation of the bioluminescence patterns in a colony of luminous *E. coli.* Oscillations and fluctuations similar to experimental ones can be computationally simulated ignoring the signal-dependence as well as the density-dependence of the chemotactic sensitivity.

C. The Effect of the Non-Local Sampling

The non-local sampling was modeled by using non-local gradient (16). The constant chemotactic sensitivity ($\chi(u, v) = \chi_0$) was used in these simulations. The spatiotemporal patterns of the dimensionless cell density u were simulated at various values of the effective sampling radius ρ. Figure 6 shows how the non-local sampling affects the pattern formation in the luminous *E. coli* colony.

Accepting $\rho = 0$ leads to a model with the local sampling and the simple gradient, operator $\overset{\circ}{\nabla}_\rho$ approaches ∇. The computational results showed that the simulated patterns get dissimilar from the experimental one (Figure 2) when increasing ρ-parameter (Figure 6). As it can be seen from Figure 6c, merging of the different "branches" in the pattern is almost gone and this merging behaviour is essential to get patterns similar to experimental ones. Because of this, there is no practical reason for application of applying the non-local gradient to modeling the formation of the patterns in a colony of luminous *E. coli.*

D. The Effect of the Nonlinear Diffusion

The nonlinear diffusion was modeled by using the following form of the diffusion function: $D(u) = Du^m$ [22]. The chemotactic sensitivity was assumed to be constant ($\chi(u, v) = \chi_0$) in these simulations. The spatiotemporal patterns of the dimensionless cell density u were simulated at various values of m-parameter. Figure 7 shows the effect of the nonlinearity of the diffusion.

Accepting $m = 0$ leads to a model with the linear diffusion. Results of the simulations at different m values show that patterns tend to drift away from the experimental one (Figure 2) when increasing ($m \to \infty$) (Figure 7c) or decreasing ($m \to -\infty$) (Figure 7a) the m-parameter. The pattern shown in Figure 7a contains less mergers of different "branches" (as a result of $m \ll 0$). Figure 7 exhibits the "branch" movements that are distorted compared to the experimentally observed ones (as a result of $m \gg 0$). Therefore, there is no need to use the nonlinear diffusion for modeling the pattern formation in a colony of luminous *E. coli.*

E. The Effect of the Non-Local Sampling With the Nonlinear Diffusion

From the simulations with the non-local gradient (the non-local sampling) and the nonlinear diffusion it was seen that the increasing the non-local gradient parameter ρ has visually opposite effect to the increasing nonlinear diffusion parameter m (Figure 6c versus Figure 7c). As a result, additional numerical experiments were carried out to determine how the non-local sampling combined with the nonlinear diffusion affects the pattern formation. Various combinations of ρ- and m-parameter values were used to simulate the spatiotemporal patterns of the dimensionless cell density u along the three phase contact line of the cylindrical vessel. Figure 8 shows the effects of the non-local sampling and the diffusion nonlinearity.

From Figure 8 it can be seen that the simulated patterns (especially Figure 8a) are more similar to the experimentally observed one (Figure 2) than those shown in Figures 6c and 7c. When analyzing the most distorted case (Figure 6c), one can see that the merging behaviour can be regained by using the nonlinear diffusion (Figures 8c and 8d), but the result is not quite similar to the desired one. However, if the nonlinear diffusion is added to the case shown in Figure 6b, the results (Figures 8a and 8b) become much better than those obtained considering the non-local sampling and the diffusion nonlinearity separately. This means that when increasing ρ, one should consider increasing m, respectively. On the other hand, the comparison of Figure 8a with Figure 3a does not confirm that the model with the non-local sampling and the nonlinear diffusion is capable to produce a result that better matches experimentally observed one. Because of this, there is no practical need for applying the non-local sampling as well as the nonlinear diffusion for the computational modeling of the pattern formation in a colony of luminous *E. coli.*

F. A minimal model

In the previous sections it was shown that the pattern formation along the contact line in a cellular population can be modeled at the following values of the model parameters: $m = 0$, $\alpha = 0$, $\beta \to \infty$, $\gamma \to \infty$, $\epsilon = 0$. The simulated patterns at these values tend to have the desired properties similar to the experimental ones (Figure 2) - emergence and merging of the strands are present and regular. Accepting

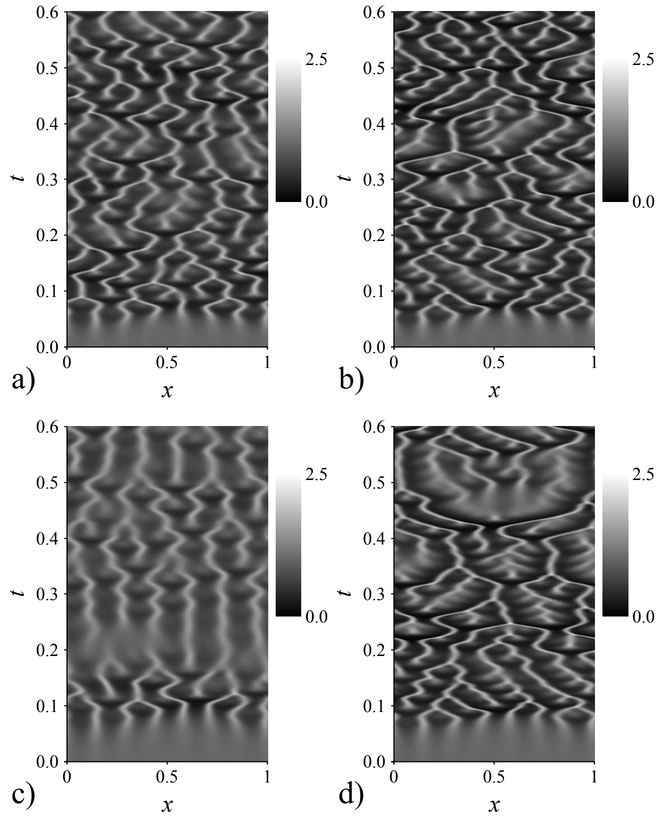


Figure 8. Spatiotemporal plots of the dimensionless cell density u when using the non-local sampling and the nonlinear diffusion ($\rho = 0.012$, $m = 0.2$) (a), ($\rho = 0.012$, $m = 0.4$) (b), ($\rho = 0.016$, $m = 0.2$) (c), ($\rho = 0.016$, $m = 0.6$) (d). Values of the other parameters are as defined in (21).

theses values leads to a reduction of the governing equations (15) to the following form:

$$
\begin{aligned}
&\frac{\partial u}{\partial t} = D\frac{\partial^2 u}{\partial x^2} - \chi_0 \frac{\partial}{\partial x}\left(u\frac{\partial v}{\partial x}\right) + sru(1-u), \\
&\frac{\partial v}{\partial t} = \frac{\partial^2 v}{\partial x^2} + s\left(\frac{u}{1+\phi u} - v\right), \\
&x \in (0,1), \quad t > 0.
\end{aligned}
\tag{22}
$$

The governing equations (22), the initial (18) and the boundary (19) conditions form together a minimal mathematical model suitable for simulating the pattern formation in a colony of luminous *E. coli.*

According to the classification of the chemotaxis models introduced by Hillen and Painter [6], the minimal model (22) is a combination of two models: the nonlinear signal kinetics model M6 and the cell kinetics model M8. This combination of the models has comprehensively been analyzed by Maini and others [4], [20], [23].

The governing equations (22) contain five parameters, D, χ_0, r, ϕ and s. The diffusion parameter D is necessary because of an inequality of the dimensional diffusion coefficients D_n and D_c [4], [21]. The model parameter s is required to support the spatial and temporal scale for simulating systems and processes of the interest. The essential parameter χ_0 controls the chemotactic response of the cells to the concentrations of the attractant and allows to reproduce the experimentally observed bands. Earlier, it was shown that r and ϕ are also essential for modeling the pattern formation in a colony of luminous *E. coli* [17].

VI. Conclusions

The quasi-one dimensional spatiotemporal pattern formation along the three phase contact line in the fluid cultures of lux-gene engineered *Escherichia coli* can be simulated and studied on the basis of the Patlak-Keller-Segel model.

The mathematical model (11)-(13) and the corresponding dimensionless model (15), (18), (19) of the bacterial self-organization in a circular container as detected by bioluminescence imaging may be successfully used to investigate the pattern formation in a colony of luminous *E. coli*.

A constant function ($\chi(u, v)$ as well as $h(n, c)$) of the chemotactic sensitivity can be used for modeling the formation of the bioluminescence patterns in a colony of luminous *E. coli* (Figures 4 and 5). Oscillations and fluctuations similar to experimental ones (Figure 2) can be computationally simulated ignoring the signal-dependence as well as the density-dependence of the chemotactic sensitivity (Figure 3a).

The local sampling and the linear diffusion can be successfully applied to modeling the formation of the bioluminescence patterns in a colony of luminous *E. coli*. The influence of the non-local gradient to the pattern formation can be partially compensated with the nonlinear diffusion (Figures 6, 7 and 8). However, the non-local sampling and the nonlinear diffusion do not yield in patterns more similar to the experimentally observed ones when compared to the patterns obtained by the corresponding model with the local sampling and the linear diffusion.

The more precise and sophisticated two- and three-dimensional computational models implying the formation of structures observed on bioluminescence images are now under development and testing.

Acknowledgment

The research by R. Baronas is funded by the European Social Fund under Measure VP1-3.1-ŠMM-07-K ”Support to Research of Scientists and Other Researchers (Global Grant)”, Project ”Developing computational techniques, algorithms and tools for efficient simulation and optimization of biosensors of complex geometry”.

References

[1] R. Baronas, “Simulation of bacterial self-organization in circular container along contact line as detected by bioluminescence imaging,” in *SIMUL 2011: The Third International Conference on Advances in System Simulation*, Barcelona, Spain, Oct. 2011, pp. 156–161.

[2] M. Eisenbach, *Chemotaxis*. London: Imperial College Press, 2004.

[3] T. C. Williams, *Chemotaxis: Types, Clinical Significance, and Mathematical Models*. New York: Nova Science, 2011.

[4] J. D. Murray, *Mathematical Biology: II. Spatial Models and Biomedical Applications*, 3rd ed. Berlin: Springer, 2003.

[5] E. F. Keller and L. A. Segel, “Model for chemotaxis,” *J. Theor. Biol.*, vol. 30, no. 2, pp. 225–234, 1971.

[6] T. Hillen and K. J. Painter, “A user’s guide to PDE models for chemotaxis,” *J. Math. Biol.*, vol. 58, no. 1-2, pp. 183–217, 2009.

[7] E. O. Budrene and H. C. Berg, “Dynamics of formation of symmetrical patterns by chemotactic bacteria,” *Nature*, vol. 376, no. 6535, pp. 49–53, 1995.

[8] M. P. Brenner, L. S. Levitov, and E. O. Budrene, “Physical mechanisms for chemotactic pattern formation by bacteria,” *Biophys. J.*, vol. 74, no. 4, pp. 1677–1693, 1998.

[9] S. Sasaki *et al.*, “Spatio-temporal control of bacterial-suspension luminescence using a PDMS cell,” *J. Chem. Engineer. Japan*, vol. 43, no. 11, pp. 960–965, 2010.

[10] K. J. Painter and T. Hillen, “Spatio-temporal chaos in a chemotactic model,” *Physica D*, vol. 240, no. 4-5, pp. 363–375, 2011.

[11] R. Šimkus, “Bioluminescent monitoring of turbulent bioconvection,” *Luminescence*, vol. 21, no. 2, pp. 77–80, 2006.

[12] R. Šimkus, V. Kirejev, R. Meškienė, and R. Meškys, “Torus generated by *Escherichia coli*,” *Exp. Fluids*, vol. 46, no. 2, pp. 365–369, 2009.

[13] S. Daunert *et al.*, “Genetically engineered whole-cell sensing systems: coupling biological recognition with reporter genes,” *Chem. Rev.*, vol. 100, no. 7, pp. 2705–2738, 2000.

[14] Y. Lei, W. Chen, and A. Mulchandani, “Microbial biosensors,” *Anal. Chim. Acta*, vol. 568, no. 1-2, pp. 200–210, 2006.

[15] M. B. Gu, R. J. Mitchell, and B. C. Kim, “Whole-cell-based biosensors for environmental biomonitoring and application,” *Adv. Biochem. Eng. Biotechnol.*, vol. 87, pp. 269–305, 2004.

[16] R. Šimkus and R. Baronas, “Metabolic self-organization of bioluminescent *Escherichia coli*,” *Luminescence*, vol. 26, no. 6, pp. 716–721, 2011.

[17] R. Baronas and R. Šimkus, “Modeling the bacterial self-organization in a circular container along the contact line as detected by bioluminescence imaging,” *Nonlinear Anal. Model. Control*, vol. 16, no. 3, pp. 270–282, 2011.

[18] A. A. Samarskii, *The Theory of Difference Schemes*. New York-Basel: Marcel Dekker, 2001.

[19] I. R. Lapidus and R. Schiller, “Model for the chemotactic response of a bacterial population,” *Biophys. J.*, vol. 16, no. 7, pp. 779–789, 1976.

[20] P. K. Maini, M. R. Myerscough, K. H. Winters, and J. D. Murray, "Bifurcating spatially heterogeneous solutions in a chemotaxis model for biological pattern generation," *Bull. Math. Biol.*, vol. 53, no. 5, pp. 701–719, 1991.

[21] R. Tyson, S. R. Lubkin, and J. D. Murray, "Model and analysis of chemotactic bacterial patterns in a liquid medium," *J. Math. Biol.*, vol. 38, no. 4, pp. 359–375, 1999.

[22] R. Kowalczyk, "Preventing blow-up in a chemotaxis model," *J. Math. Anal. Appl.*, vol. 305, no. 2, pp. 566–588, 2005.

[23] M. R. Myerscough, P. K. Maini, and K. J. Painter, "Pattern formation in a generalized chemotactic model," *Bull. Math. Biol.*, vol. 60, no. 1, pp. 1–26, 1998.

[24] E. F. Keller and L. A. Segel, "Travelling bands of chemotactic bacteria: A theoretical analysis," *J. Theor. Biol.*, vol. 30, no. 2, pp. 235–248, 1971.

[25] K. Painter and T. Hillen, "Volume-filling and quorum-sensing in models for chemosensitive movement," *Can. Appl. Math. Quart.*, vol. 10, no. 4, pp. 501–543, 2002.

[26] J. J. L. Velazquez, "Point dynamics for a singular limit of the Keller-Segel model. I. Motion of the concentration regions," *SIAM J. Appl. Math.*, vol. 64, no. 4, pp. 1198–1223, 2004.

[27] H. G. Othmer and T. Hillen, "The diffusion limit of transport equations II: Chemotaxis equations," *SIAM J. Appl. Math.*, vol. 62, no. 4, pp. 1122–1250, 2002.

[28] T. Hillen, K. Painter, and C. Schmeiser, "Global existence for chemotaxis with finite sampling radius," *Discr. Cont. Dyn. Syst. B*, vol. 7, no. 1, pp. 125–144, 2007.

[29] R. Baronas, F. Ivanauskas, and J. Kulys, *Mathematical Modeling of Biosensors: An Introduction for Chemists and Mathematicians*, ser. Springer Series on Chemical Sensors and Biosensors, G. Urban, Ed. Dordrecht: Springer, 2010.

[30] R. Čiegis and A. Bugajev, "Numerical approximation of one model of bacterial self-organization," *Nonlinear Anal. Model. Control*, vol. 17, no. 3, pp. 253–270, 2012.

[31] S. Leestma and L. Nyhoff, *Pascal Programming and Problem Solving*, 4th ed. New York: Prentice Hall, 1993.

[32] R. Tyson, L. Stern, and R. J. LeVeque, "Fractional step methods applied to a chemotaxis model," *J. Math. Biol.*, vol. 41, no. 5, pp. 455–475, 2000.

[33] P. Romanczuk, U. Erdmann, H. Engel, and L. Schimansky-Geier, "Beyond the Keller-Segel model. Microscopic modeling of bacterial colonies with chemotaxis," *Eur. Phys. J. Special Topics*, vol. 157, no. 1, pp. 61–77, 2007.

Empty Container Management in Mult-port System with Inventory-based Control

Loo Hay Lee, Ek Peng Chew, and Yi Luo
Department of Industrial and Systems Engineering
National University of Singapore
Singapore
iseleelh@nus.edu.sg, isecep@nus.edu.sg, luoyi@nus.edu.sg

Abstract – One of challenges that shipping companies face today is to effectively manage empty containers in order to meet customer demands and to reduce operation cost. This paper considers the joint empty container repositioning and container fleet sizing problem in a multi-port system with inventory-based control mechanism. A single-level threshold policy with repositioning rule in terms of minimizing the repositioning cost is proposed to manage empty containers. The objective is to optimize the fleet size and the parameters of the policy to minimize the expected total cost per period. Taking advantage of an interesting property of the problem, this paper proposes two approaches, a non-linear programming and a gradient search approach to solve the problem. From a methodological perspective, this paper deploys infinitesimal perturbation analysis method to improve computational efficiency. A numerical study demonstrates the effectiveness of the proposed policy and provides some managerial insights for shipping companies in managing empty containers.

Keywords - empty container management; repositioning; inventory control; simulation; infinitsimal perturbation analysis

I. INTRODUCTION

Growth in maritime transportation industries has been stimulated by the increase in international merchandise trade as a result of globalization in the last few decades. In particular, the containerization of cargo transportation has been the fastest growing sector of the maritime industries. Containerized cargos have grown at an annual average rate of 9.5% over the period 1987 through to 2006, which is exceeding the average maritime trade growth rate 4.1% over the same period [2]. In 2010, it estimated that container trade volumes reached 140 million twenty-foot equivalent units (TEUs) [3]. The growth of containerized shipping has presented challenges inevitably, in particular to the management of empty containers (ECs) arising from the highly imbalanced trade between continents. For example, in 2010, the annual container flow from Asia to Europe was 13.5 million TEUs and 5.6 million TEUs in the reverse direction, resulting in container flow imbalance of 7.9 million TEUs [3]. To manage the imbalance of container flows, repositioning ECs is an essential approach used by shipping companies. It is reported that empty container movements have accounted for at least 20% of the global port handling activity ever since 1998 [2]. Song et al. [4] estimated that the cost of repositioning ECs around the globe exceeded US$15 billions, which was 27% of the total world fleet running cost based on the data in 2002. If the cost of repositioning ECs can be reduced, the shipping companies can increase profit and improve competitiveness. Therefore, how to effectively and efficiently operate ECs is a very important issue for shipping companies and it is known as the problem of empty container repositioning (ECR).

Research on ECR has increased quite substantially in recent years. Much of work has adopted deterministic programming approach, which uses dynamic network programming formulations (see, e.g., [5][6][7][8]). The stochastic factors of the problem, such as future customer demands and returned containers, have attracted much attention since 1990s. In [9], Cheung and Chen proposed a two-stage stochastic network model to determine the ECR decisions. In [10], Lam et al. formulated the ECR problem as a dynamic stochastic programming. In [11], Long et al. applied a two-stage stochastic model to incorporate uncertainties in ECR problem with random demand, supply and ship space capacity. The mathematical models often successfully capture the dynamic and stochastic nature of the problem, while give rise to some concerns, such as requirement of a pre-specified planning horizon, sensitivity of the decisions to data accuracy and variability, and implementation of the decisions in the stochastic systems [12][13].

Another interesting development is to explore inventory-based control policies for managing ECs. Characterized by a set of rules and a set of parameters, such policies utilize the feedback of inventory information to manage ECs. Once the rules and parameters are designed in advance, ECs can be repositioned by following these simple rules. One of the advantages of using inventory-based control policies is that it is easy to operate and easy to understand, while producing near-optimal or even optimal solutions [14].

In this study, we consider the joint ECR and container fleet sizing problem in a multi-port system with inventory-based control mechanism. The system comprises a set of ports connected to each other and a fleet of own containers are used to meet the stochastic customer demands. A single-level threshold policy with repositioning rule in terms of minimizing the repositioning cost is proposed to manage ECs with periodical review. Although in general such policy does not guarantee optimality, its simplicity in implementation makes them attractive in practice and this fact motivates us to study its application in the ECR problem. The objective of this study is to optimize the fleet size and the parameters of the policy in terms of minimizing the expected total cost per period, including repositioning cost and holding and leasing cost.

The remainder of the paper is organized as follows: related work is presented in Section II. Then, Section III presents the mathematical formulation of the ECR problem and discusses the methods to solve the problem, followed by the description of the infinitesimal perturbation analysis (IPA) based gradient algorithm in Section IV. Section V illustrates the numerical studies. Finally, the work in this study is concluded and several issues for future research are discussed.

II. Related Work

Inventory-based control policies for ECR problem have recently received increasing attentions. Studies in [15][16][17][18][19] focused on the examination of the structural properties of the optimal inventory-based repositioning policies and demonstrated that the optimal repositioning policies were of threshold type in some situations such as one-port and two-port systems. Once the parameters and rules of such policies are designed in advance, they are easy to operate and easy to understand from a managerial perspective. Several researchers further considered the implementation of the threshold-type control policies in more general systems. Song and Carter [14] addressed the ECR problem in a hub-and-spoke system with the assumption that ECs can be only repositioned between hub and spokes. Song and Dong [20] considered the implementation of a two-level threshold policy in a cycle shipping route system, and extended to optimize the fleet size and the parameters of the policy by using a simulation-based method with genetic algorithm in a typical liner shipping system [13]. Song and Dong [21] studied the repositioning policy for ECs with flexible destination ports in a liner shipping system. Yun et al. [22] considered the ECR problem in the inland system and optimized the parameters of the policy by applying a simulation optimization tool.

From the literature we find that the current studies are inadequate in addressing the implementation of the inventory-based control policies in general maritime systems. To the best of our knowledge, few of the studies have considered the implementation of such policies in multi-port system with direct empty container flows between each pair of ports, to which this study attempts to contribute. Li et al. [23] considered the ECR problem in a multi-port system, and proposed a heuristic algorithm based on a two-level threshold policy to allocate ECs among multiple ports. However, it could be computationally expensive when the numbers of ports and fleet size are very large. Besides, the fleet sizing problem is also not fully studied in such system. Moreover, most policies in literature apply simple rule, such as linear rationing rule to allocate ECs. The repositioning rule in terms of minimizing the repositioning cost has not been considered yet.

This paper is an extended version of [1]. It provides a detailed description of the IPA-based gradient technique, as well as a previously omitted mathematical proof on the important property in [1]. Moreover, the gradient technique is extended to optimize the fleet size.

III. Problem Formulation

A multi-port system, which consists of ports connected to each other, is considered in this study. A fleet of own ECs meets exogenous customer demands, which are defined as the requirements for transforming ECs to laden containers and then transporting these laden containers from original ports to destination ports. A single-level threshold policy with periodical review is adopted to manage ECs. At the beginning of a period, the ECR decisions are made for each port, involving whether to reposition ECs, to or from which ports, and in what quantity. Then, when the customer demands occur in the period, those ECs that are currently stored at the port and those ECs that are repositioned to the port in this period can be used to satisfy customer demands. If it is not enough, additional ECs will be leased from vendors.

Several assumptions are made as follows:

- Only one type of container, i.e., TEU is considered.
- The customer demands must be satisfied in each period; and customer demands for each pair of ports in each period follow independent normal distributions.
- Short-term lasing is considered and the quantity of the leased ECs is always available in each port at any time.
- The leased ECs are not distinguished from owned ECs, i.e., the shipping company can return own ECs to vendors when it has sufficient ECs available.
- The travel time for each pair of ports is less than one period length.
- When the repositioned ECs arrive at destination ports, they will become available immediately; and when laden containers arrive at destination ports, they will become empty and be available at the beginning of next period.
- The cost of repositioning an EC from p to port m is the summation of the handling cost of an EC at port p, the handling cost of an EC at port m, and the transportation cost of an EC from p to port m.

The notations used in this paper are presented in Table I. In every period t, the ECR decisions are firstly made at the beginning of this period. Then, the inventory position can be obtain by

$$y_{p,t} = x_{p,t} + a_{p,t}, \ \forall p \in P \tag{1}$$

After customer demands are realized and the laden containers become available, the beginning on-hand inventory of the next period can be updated by

$$x_{p,t+1} = y_{p,t} + \varphi_{p,t}, \ \forall p \in P \tag{2}$$

It should be noted that $x_{p,t}$ can be negative. This is due to the fact that customer demands are random and beyond control. If $x_{p,t}$ is negative, it represents the number of containers that are leased from port p and stored at other ports. In this

TABLE I. LIST OF NOTATIONS

Symbol	Description
N	the fleet size, which is the number of owned ECs
P	the set of ports
t	the discrete time decision period
P_t^S	the surplus port subset in period t
P_t^D	the deficit port subset in period t
$x_{p,t}$	the beginning on-hand inventory of port p in period t
$y_{p,t}$	the inventory position of port p in period t after making the ECR decisions
$z_{p,m,t}$	the number of ECs repositioned from port p to port m in period t
$\varepsilon_{p,m,t}$	the random customer demand from port p to port m in period t
$u_{p,t}^S$	the number of estimated EC supply of surplus port p in period t
$u_{p,t}^D$	the number of estimated EC demand of deficit port p in period t
$\mathbf{x}_t$	$=[x_{p,t}]_{p\in P}$ the vector of the beginning on-hand inventory in period t
$\mathbf{y}_t$	$=[y_{p,t}]_{p\in P}$ the vector of the inventory position in period t
$\mathbf{Z}_t$	$=(z_{p,m,t})_{p\in P,m\in P,p\neq m}$ the array of repositioned quantities for all ports
ω_t	$=(\varepsilon_{p,m,t})_{p\in P,m\in P,p\neq m}$ the stochastic customer demands in period t
$C_{p,m}^R$	the cost of repositioning an EC from port p to port m, $\forall p\neq m$
C_p^H	the cost of holding an EC at port p per period
C_p^L	the cost of leasing an EC at port p per period
γ_p	the threshold of port p
$\boldsymbol{\gamma}$	$=[\gamma_p]_{p\in P}$ vector of the thresholds
$\eta_{p,t}^O$	$=\sum_{m\in P,m\neq p}\varepsilon_{p,m,t}$ the number of total exported laden containers of port p in period t
$a_{p,t}$	$=\sum_{l\in P,l\neq p}z_{l,p,t}-\sum_{m\in P,m\neq p}z_{p,m,t}$ the net actual imported EC of port p in period t
$F_p(.)$	the cumulative distribution function for $\eta_{p,t}^O$
$\varphi_{p,t}$	$=\sum_{l\in P,l\neq p}\varepsilon_{l,p,t}-\eta_{p,t}^O$ the net customer demand of port p in period t
$E_{p,t}$	the set of ports whose net actual imported ECs is changed by perturbing the estimated supply or demand of port p in period t, $e_{p,t}\in E_{p,t}$ and $e_{p,t}\neq p$
$Q_{p,t}$	the set of ports whose beginning on-hand inventory in period t is affected by perturbing threshold of port p, $q_{p,t}\in Q_{p,t}$ and $q_{p,t}\neq p$
$\pi_{p,t}$	the corresponding dual variable for port p constraint in the transportation model in period t
q_t^N	the port whose beginning on-hand inventory in period t is affected by perturbing the fleet size
$I\{.\}$	a indicator function, which takes 1 if the condition in the brace is true and otherwise 0

situation there are no ECs stored at port p; otherwise, they will be returned to the vendor to reduce the number of leased containers according to the assumption. If $x_{p,t}$ is positive, it represents the number of ECs that are available at port p, which implies that there are no container leased out from vendor at this port. Note that there are N owned ECs in the system, we have

$$N=\sum_{p\in P}x_{p,t}\quad\forall t \tag{3}$$

Let $J(\mathbf{x}_t,\boldsymbol{\gamma},\omega_t)$ be the total cost in period t. It can be defined as:

$$J(\mathbf{x}_t,\boldsymbol{\gamma},\omega_t)=H(\mathbf{Z}_t)+G(\mathbf{y}_t,\omega_t) \tag{4}$$

where the value of $\mathbf{Z}_t$ is determined by the beginning on-hand inventory $\mathbf{x}_t$ and the policy $\boldsymbol{\gamma}$; $H(\mathbf{Z}_t)$ and $G(\mathbf{y}_t,\omega_t)$ are the EC repositioning cost and the EC holding and leasing cost in period t, respectively. They are defined as follows

$$H(\mathbf{Z}_t)=\sum_{p\in P}\sum_{m\in P,m\neq p}C_{p,m}^R\cdot z_{p,m,t} \tag{5}$$

$$\begin{aligned}G(\mathbf{y}_t,\omega_t)&=\sum_{p\in P}g(y_{p,t},\eta_{p,t}^O)\\&=\sum_{p\in P}\left(C_p^H(y_{p,t}-\eta_{p,t}^O)^+ + C_p^L(\eta_{p,t}^O-y_{p,t})^+\right)\end{aligned} \tag{6}$$

where $g(y_{p,t},\eta_{p,t}^O)$ represents the EC holding and leasing cost of port p in period t ; $x^+=\max(0,x)$. More specifically, the EC repositioning cost $H(\mathbf{Z}_t)$ refers to the cost of repositioning ECs between multiple ports. The EC holding and leasing cost $G(\mathbf{y}_t,\omega_t)$ is the cost incurred when ECs are stored at some ports and additional ECs are leased from vendors at the other ports.

Next, a single-level threshold policy is developed to make the ECR decisions $\mathbf{Z}_t$ at the beginning of period t.

A. A Single-level Threshold Policy

Note that when the ECR decisions are made in a period, the customer demands in this period have not been realized yet. Hence, we try to maintain the inventory position at a target threshold value $\boldsymbol{\gamma}$. More specifically, γ_p is the target threshold of port p. In period t, if $x_{p,t}>\gamma_p$, then port p is a surplus port and the quantity excess of γ_p should be repositioned out to other ports that may need ECs to try to bring the inventory position down to γ_p; if $x_{p,t}<\gamma_p$, then it is a deficit port and ECs should be repositioned in from other ports that may supply ECs to try to bring the inventory position up to γ_p; if $x_{p,t}=\gamma_p$, then it is a balanced port and nothing is done.

From the policy, therefore, if there are no surplus or deficit ports in period t, no ECs should be repositioned. Otherwise, ECs should be repositioned from surplus ports to deficit ports in the right quantity at the least movement. Without loss of generality, the ECR decisions in period t are considered. The two subsets, i.e., surplus port subset and deficit port subset can be obtained as $P_t^S=\{i:x_{i,t}>\gamma_i\}$ and $P_t^D=\{j:x_{j,t}<\gamma_j\}$, respectively. For a surplus port, its number of excess ECs, namely the number of estimated EC supply is calculated by (7); and for a deficit port, its number of estimated EC demand by (8).

$$u_{i,t}^S=x_{i,t}-\gamma_i,\ \ \forall i\in P_t^S \tag{7}$$

$$u_{j,t}^D=\gamma_j-x_{j,t},\ \ \forall j\in P_t^D \tag{8}$$

If either P_t^S or P_t^D are empty, then $\mathbf{Z}_t=0$. Otherwise, the value of $z_{l,m,t},\forall l\in(P-P_t^S),m\in(P-P_t^D),l\neq m$ should be equal to zero, and the value of $z_{i,j,t},\forall i\in P_t^S,j\in P_t^D$ are determined by solving a transportation model.

Note that when P_t^S and P_t^D are nonempty, the total number of estimated EC supplies $\sum_{i\in P_t^S} u_{i,t}^S$ could be not equal to the total number of estimated EC demands $\sum_{j\in P_t^D} u_{j,t}^D$. Thus, we propose to move as many excess ECs of surplus ports as possible to the deficit ports to satisfy their demands. Hence, when $\sum_{i\in P_t^S} u_{i,t}^S \geq \sum_{j\in P_t^D} u_{j,t}^D$, a transportation model is formulated as follows to determine the repositioned quantities in each period:

$$\min_{z_{i,j,t}} \sum_{i\in P_t^S} \sum_{j\in P_t^D} C_{i,j}^R \cdot z_{i,j,t} \tag{9}$$

$$\text{s.t. } \sum_{j\in P_t^D} z_{i,j,t} \leq u_{i,t}^S, \ \ \forall i \in P_t^S \tag{10}$$

$$\sum_{i\in P_t^S} z_{i,j,t} \geq u_{j,t}^D, \ \ \forall j \in P_t^D \tag{11}$$

$$z_{i,j,t} \geq 0, \ \ \forall i \in P_t^S, j \in P_t^D \tag{12}$$

Constraints (10) ensure that the repositioned out ECs of a surplus port should not exceed its estimated EC supply; constraints (11) ensure that the EC demand of a deficit port can be fully satisfied; constraints (12) are the non-negative quantity constraints. When $\sum_{i\in P_t^S} u_{i,t}^S < \sum_{j\in P_t^D} u_{j,t}^D$, similar model is built by substituting $\sum_{j\in P_t^D} z_{i,j.t} \geq u_{i,t}^S, \forall i \in P_t^S$ and $\sum_{i\in P_t^S} z_{i,j.t} \leq u_{j,t}^D, \forall j \in P_t^D$ for (10) and (11), respectively.

B. The Optimization Problem

Let $J(N,\boldsymbol{\gamma})$ be the expected total cost per period with the fleet size N and policy $\boldsymbol{\gamma}$. The problem, which is to optimize the fleet size and the parameters of the policy in terms of minimizing the expected total cost per period, can be formulated as

$$\min_{N,\gamma} J(N,\gamma) \tag{13}$$

subject to the inventory dynamic equations (1)~(3) and the single-level threshold policy. As in many papers on empty container movement (see, e.g., [9] [11]), the variables that relate to the flow of ECs are considered as continuous variables in this study. That is, the fleet size and the parameters of the policy are considered as real numbers in this study. It is fine when the values of these variables are large.

In general, it is difficult to solve problem (13), since there is no closed-form formulation for the computation of $J(N,\boldsymbol{\gamma})$ involving the repositioned empty container quantities determined by transportation models. However, when the transportation model is balanced in period t, i.e., the total the total number of estimated EC supplies, namely $\sum_{i\in P_t^S} u_{i,t}^S$ is equal to the total number of estimated EC demands, namely $\sum_{j\in P_t^D} u_{j,t}^D$, the excess ECs in all surplus ports can be fully repositioned out to satisfy the demands of all deficit ports. Hence, the inventory position of each port can be kept at its target threshold level in this period, i.e., $y_{p,t} = \gamma_p$. Further, the EC repositioning cost in next period will be only related to the customer demands in this period. From (3), (7) and (8), we obtain that $\sum_{i\in P_t^S} u_{i,t}^S = \sum_{j\in P_t^D} u_{j,t}^D \ \forall t$ if and only if $N = \sum_{p\in P} \gamma_p$. Hence, an important property of the problem is presented as follows:

Property I: When the fleet size is equal to the sum of thresholds, the inventory position of a port can be always maintained at its target threshold value and then the EC repositioning cost in a period is only dependent on the customer demands.

Let Scenario-I (Scenario-II) be the scenario in which $N = \sum_{p\in P} \gamma_p \ (\neq \sum_{p\in P} \gamma_p)$. Taking advantage of this property, we propose two approaches to solve problem (13) under both scenarios, respectively.

1) Scenario-I

Consider the problem under Scenario-I. From Property I, it implies that $y_{p,t} \equiv \gamma_p \ \forall p \in P, t$, and only the EC holding and leasing cost in a period is related to values of N and $\boldsymbol{\gamma}$. Hence, the optimal solution which minimizes $J(N,\boldsymbol{\gamma})$ should be equivalent to the optimal solution which minimizes the expected EC holding and leasing cost per period. From (6), since that holding and leasing cost function for each port is independent, the problem (13) under Scenario-I can be simplified to an non-linear programming (NLP) problem as follows:

$$\min_{N,\gamma} \ \sum_{p\in P} E\left(\left(C_p^H \cdot (\gamma_p - \eta_p^O)^+ + C_p^L \cdot (\eta_p^O - \gamma_p)^+\right)\right)$$

$$\text{s.t. } N = \sum_{p\in P} \gamma_p$$

where the subscript t in the notation of $\eta_{p,t}^O$ is dropped since customer demands in each period are independent and identically distributed. The NLP problem can be further decomposed into P independent newsvendor problems with the optimal solutions as follows:

$$\gamma_p^* = F_p^{-1}\left(C_p^L \big/ \left(C_p^L + C_p^H\right)\right) \tag{14}$$

$$N^* = \sum_{p\in P} \gamma_p^* \tag{15}$$

where $F_p^{-1}(.)$ is the inverse function of $F_p(.)$.

2) Scenario-II

The problem under Scenario-II is more complex than that under Scenario-I, since the EC repositioning cost is also affected by the values of N and $\boldsymbol{\gamma}$. Consider that the minimum expected holding and leasing cost of problem (13) can only be achieved under Scenario-I due to the convexity of the holding and leasing cost function. The problem under Scenario-II is worth to study if and only if the minimum expected EC repositioning cost under this scenario could be less than that under Scenario-I.

Consider there is no closed-form formulation for the expected EC repositioning cost. Without loss of generality, the EC repositioning costs in a period under both scenarios are mathematically compared with same customer demands. More specifically, we have Scenario-I with parameters $(N, \boldsymbol{\gamma}^{\mathrm{I}})$ and Scenario-II with parameters $(N, \boldsymbol{\gamma}^{\mathrm{II}})$ where $N = \sum_{p \in P} \gamma_p^{\mathrm{I}} \neq \sum_{p \in P} \gamma_p^{\mathrm{II}}$. Let other various quantities in both scenarios be distinguished by displaying the arguments $.^{\mathrm{II}}$ and $.^{\mathrm{I}}$, respectively. We have next proposition.

Proposition 3.1: For all t, we have

$$H\left(\mathbf{x}_{t+1}^{\mathrm{II}}, \boldsymbol{\gamma}^{\mathrm{II}}\right)\Big|_{\omega_t} \leq H\left(\mathbf{x}_{t+1}^{\mathrm{I}}, \boldsymbol{\gamma}^{\mathrm{I}}\right)\Big|_{\omega_t}$$

with same customer demand ω_t.

Proof: The proof is given in Appendix A. ■

From this proposition, it implies that the expected EC repositioning cost per period under Scenario-II could be less than that under Scenario-I. Therefore, it is worth to study the problem under Scenario-II, and the simulation technique is adopted to estimate $J(N, \boldsymbol{\gamma})$ with given values of N and $\boldsymbol{\gamma}$ as follows:

$$J(N, \boldsymbol{\gamma}) \approx \frac{1}{T} \sum_{t=1}^{T} J(\mathbf{x}_t, \boldsymbol{\gamma}, \omega_t) = \frac{1}{T} \sum_{t=1}^{T} \left(H(\mathbf{x}_t, \boldsymbol{\gamma}) + G(\mathbf{y}_t, \omega_t) \right) \quad (16)$$

where T is the amount of the simulation periods. A gradient search method is developed to solve the problem in next section.

IV. IPA-BASED GRADIENT TECHNIQUE

IPA [24] is able to estimate the gradient of the objective function from one single simulation run, thus reducing the computational time. Moreover, it has been shown that variance of IPA estimator is lower, compared with many other gradient estimators [25]. The main idea behind IPA is that: if a decision parameter of a system is perturbed by an infinitesimal amount, the derivation of the system performance to that parameter can be estimated by tracing its pattern of propagation through the system. This will be a function of the fraction of the propagations that die before having a significant effect on the response of interest [26]. Once the performance derivatives are obtained, they can be imbedded in optimization algorithms to optimize the interest parameters. Successful applications of IPA have been reported in inventory modeling [27], stochastic flow models [28][29], persistent monitoring problems [30], budget allocation for effective data collection [31], multi-location transshipment problem with capacitated production [32]. In this study, an IPA-based gradient technique is proposed to search the optimal solution under Scenario-II. The overall gradient technique is briefly described in Figure 1.

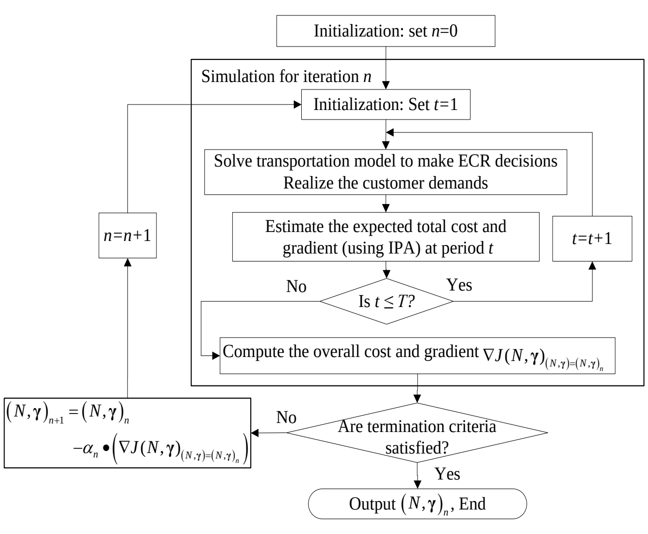


Figure 1. The flow of the IPA-based gradient technique

For a given $(N, \boldsymbol{\gamma})$ at iteration n, simulation is run to estimate the expected cost $J(N, \boldsymbol{\gamma})$ in (16). At the same time, the gradient vector of the expected cost $\nabla J(N, \boldsymbol{\gamma})_{(N,\boldsymbol{\gamma})=(N,\boldsymbol{\gamma})_n}$ can be estimated. Briefly, the simulation is run by firstly making the ECR decision in a given period by solving transportation models and then the customer demands are realized. At the same time, the cost and the gradient for this period can also be computed. After all periods are run, the overall cost and gradient can be computed from the individual values obtained in each period. Based on the gradient information, the parameters can be updated by using the steepest descent algorithm, i.e., $(N, \boldsymbol{\gamma})_{n+1} = (N, \boldsymbol{\gamma})_n - \alpha_n \left(\nabla J(N, \boldsymbol{\gamma})_{(N,\boldsymbol{\gamma})=(N,\boldsymbol{\gamma})_n} \right)$. α_n is the step size at the nth iteration. After $(N, \boldsymbol{\gamma})$ is updated, the process for the new iteration is repeated until the stopping criterion is met.

To estimate the gradient of expected cost, we take a partial derivation of (16) with respect to the fleet size and the threshold of port i, respectively and have

$$\begin{aligned} \partial J(N, \boldsymbol{\gamma})/\partial N &\approx 1/T \cdot \sum_{t=1}^{T} \left(\partial J(\mathbf{x}_t, \boldsymbol{\gamma}, \omega_t)/\partial N \right) \\ &\approx \frac{1}{T} \sum_{t=1}^{T} \left(\frac{\partial H(\mathbf{Z}_t)}{\partial N} + \sum_{p \in P} \frac{\partial E\left(g(y_{p,t}, \eta_{p,t}^{O})\right)}{\partial y_{p,t}} \cdot \frac{\partial y_{p,t}}{\partial N} \right) \end{aligned} \quad (17)$$

$$\begin{aligned} \partial J(N, \boldsymbol{\gamma})/\partial \gamma_i &\approx 1/T \cdot \sum_{t=1}^{T} \left(\partial J(\mathbf{x}_t, \boldsymbol{\gamma}, \omega_t)/\partial \gamma_i \right) \\ &\approx \frac{1}{T} \sum_{t=1}^{T} \left(\frac{\partial H(\mathbf{Z}_t)}{\partial \gamma_i} + \sum_{p \in P} \frac{\partial E\left(g(y_{p,t}, \eta_{p,t}^{O})\right)}{\partial y_{p,t}} \cdot \frac{\partial y_{p,t}}{\partial \gamma_i} \right) \end{aligned} \quad (18)$$

Here, for the EC holding and leasing cost function, the expected cost function is utilized to estimate the gradient instead of sample cost function since we are able to get the explicit function to evaluate the average gradient. $\partial H(\mathbf{Z}_t)/\partial N$ or $\partial H(\mathbf{Z}_t)/\partial \gamma_i$, which measures the impact of the transportation cost in period t when the fleet size or threshold of port i is changed, can be found using the dual model information from the transportation model. $\partial g(y_{p,t}, \eta_{p,t}^{O})/\partial y_{p,t}$, which measures the impact of the holding and leasing cost function of port p in period t when the inventory position is changed, can be easily found taking the derivation of (6) with respect to the inventory position level. $\partial y_{p,t}/\partial N$ or $\partial y_{p,t}/\partial \gamma_i$, which measures the impact of the inventory position level of port p in period t when the fleet size or threshold of port i change, can be estimated using the IPA technique.

Next, the gradient of expected total cost with respect to the threshold is analyzed.

A. Gradient with Respect to the Threshold

In this section, the gradient of expected total cost with respect to the threshold is studied with given fleet size N.

The nominal path is defined as the sample path generated by the simulation model with parameter $\boldsymbol{\gamma}$ and the perturbed path as the sample path generated using the same model and same random seeds, but with parameter $(\boldsymbol{\gamma})'$, where $(\boldsymbol{\gamma})' = \boldsymbol{\gamma} + \Delta\boldsymbol{\gamma}$. Without loss of generality, only the threshold of port i is perturbed and the thresholds of the other ports are unchanged, i.e., $(\gamma_i)' = \gamma_i + \Delta\gamma_i$ and $(\gamma_p)' = \gamma_p$ for other $p \in (P - \{i\})$, where the value of $\Delta\gamma_i$ is sufficiently small. By "sufficiently" small, we mean such that the surplus port and deficit port subsets are same in the both paths in every period. Oftentimes, the changes in various quantities will be presented by displaying with argument Δ. For example, $\Delta\mathbf{x}_t$ represents the change in the beginning on-hand inventory in period t.

The threshold of port i is perturbed with same quantities in all periods and the representative perturbation flow in period t is shown in Figure 2. The perturbation of $\Delta\mathbf{x}_t$ and the perturbation of $\Delta\gamma_i$ will affect the estimated EC supply or demand, i.e., $u_{p,t}^S$ or $u_{p,t}^D$ of some ports, which are the right-hand side (RHS) of the constraints in the transportation model. Hence, the perturbation of $\Delta u_{p,t}^S$ or $\Delta u_{p,t}^D$ could change the transportation cost and the net actual imported ECs, i.e., $a_{p,t}$ of some ports. The perturbations of $\Delta a_{p,t}$ and $\Delta\mathbf{x}_t$ will affect the perturbation on the EC inventory position $y_{p,t}$ of some ports. Further, perturbation of $\Delta y_{p,t}$ will affect the total holding and leasing cost in this period and the beginning on-hand inventory of next period.

From the definition of $Q_{i,t}$, we have $Q_{i,t} = \{p: \Delta x_{p,t} \neq 0\}$. Thus, depending on the status of $Q_{i,t}$, only two types of scenarios are possible – one with $Q_{i,t} = \emptyset$, the other with

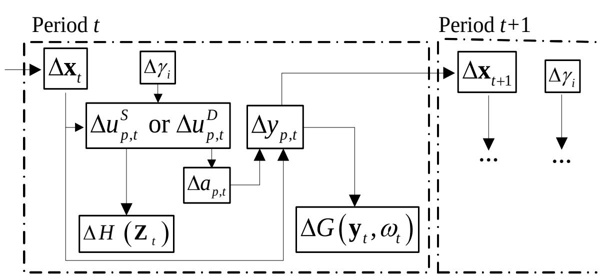


Figure 2. The Perturbation Flow

$Q_{i,t} \neq \emptyset$. Since the value of $\mathbf{x}_1$, i.e., the initial on-hand inventory, is given, it implies that $\Delta\mathbf{x}_1 = 0$ and $Q_{i,1} = \emptyset$. If the perturbation in the first period is propagated to the second period, it could lead $Q_{i,2} \neq \emptyset$. Hence, the perturbations in the first two periods are investigated in order to conclude the general formulations for the perturbation terms in (18).

1) Perturbation with Respect to Threshold in the First Period

The perturbations in period $t = 1$ are traced following the flow in Figure 2.

Since $\Delta\mathbf{x}_t = 0$, we have $\Delta u_{i,t}^S = -\Delta\gamma_i$ or $\Delta u_{i,t}^D = \Delta\gamma_i$, and $\Delta u_{p,t}^S$ or $\Delta u_{p,t}^D = 0$ for other $p \in P$ from (7) and (8). It implies that only the RHS of port i constraint is changed.

From the transportation model, the perturbation of $\Delta u_{i,t}^S$ or $\Delta u_{i,t}^D$ will affect the perturbation of $\Delta a_{p,t}$, depending on the status of port i constraint at the optimal solution. Only two scenarios are possible:

Scenario 1: The port i constraint is not binding. The perturbation of $\Delta u_{i,t}^S$ or $\Delta u_{i,t}^D$ will not affect the optimal repositioned quantities. Hence, it implies that $\Delta a_{p,t} = 0$ for all $p \in P$.

Scenario 2: The port i constraint is binding. The perturbation could affect the optimal repositioned quantities of some ports. However, only for a pair of ports, i.e., i and k, their net actual imported ECs will be changed (explained in Appendix B); while for the other port, no changes. It implies that $\Delta a_{p,t} \neq 0$ for any $p = i, k$, and $\Delta a_{p,t} = 0$ for other $p \in (P - \{i, k\})$.

From the definition of $E_{i,t}$, we get that $E_{i,t} = \{p: \Delta a_{p,t} \neq 0\}$. Thus, if the port i constraint is not binding, we have $E_{i,t} = \phi$. Otherwise, we have $E_{i,t} = \{i, e_{i,t}\}$ with $e_{i,t} = k$. It implies that $e_{i,t}$ is unique in period t. To find $E_{i,t}$, a modified stepping stone (MSS) approach (elaborated in Appendix B) is proposed.

Without investigating the details on the perturbation of $\Delta a_{p,t}$, we next consider the perturbation on $y_{p,t}$. From (1), we have $\Delta y_{p,t} = \Delta a_{p,t}, \forall p \in P$ since $\Delta x_{p,t} = 0,\ \forall p \in P$. Hence, we can get that:

- If $E_{i,t} = \phi$, $\Delta y_{p,t} = 0$ for all $p \in P$.
- If $E_{i,t} \neq \phi$, $\Delta y_{p,t} \neq 0$ for any $p \in E_{i,t}$, and $\Delta y_{p,t} = 0$ for other $p \in (P - E_{i,t})$. From the transportation model, for the port with binding constraint, its inventory position will be equal to its threshold. It implies that $\Delta y_{i,t} = \Delta\gamma_i$. Further, we have $\Delta y_{e_{i,t},t} = -\Delta\gamma_i$ since $\sum_{p \in P} \Delta y_{p,t} = \Delta N = 0$ from (1) and (3).

We continue to study how the perturbations are carried forward to next period. From (2), we have $\Delta x_{p,t+1} = \Delta y_{p,t}, \forall p \in P$. It implies that the perturbation on $x_{p,t+1}$ will be fully from the perturbation on $y_{p,t}$, and then we have $Q_{i,t+1} = \{p: \Delta y_{p,t} \neq 0\}$. From the analysis on the perturbation of $\Delta y_{p,t}$ in the case of $Q_{i,t} = \emptyset$, we have

$$Q_{i,t+1} = \begin{cases} \emptyset, & \text{if } E_{i,t} = \emptyset \\ \{i, q_{i,t+1}\}, & \text{otherwise} \end{cases} \quad (19)$$

where $q_{i,t+1} = e_{i,t}$ when $E_{i,t} \neq \emptyset$. And when $Q_{i,t+1} \neq \emptyset$, we have $\Delta x_{i,t+1} = \Delta\gamma_i$, $\Delta x_{q_{i,t+1},t+1} = -\Delta\gamma_i$, and $\Delta x_{p,t+1} = 0$ for other $p \in (P - Q_{i,t+1})$. It indicates that if there are perturbations propagated to the next period, only for a pair of ports, i and $q_{i,t+1}$, their beginning on-hand inventories will be changed by $\Delta\gamma_i$ and $-\Delta\gamma_i$, respectively; while for the other ports, no changes.

From the analysis on the perturbations of the relevant variables, we have next lemma.

Lemma 4.1: In the period t=1, we have

1) $\partial H(\mathbf{Z}_t)/\partial\gamma_i = (-1)^{I(i\in P_t^S)} \cdot \pi_{i,t}$

2) $U_{p,t} = \begin{cases} -C_p^L, & \text{if } y_{p,t} < 0 \\ \left(C_p^H + C_p^L\right) \cdot F_p(y_{p,t}) - C_p^L, & \text{if } y_{p,t} \geq 0 \end{cases}, \forall p \in P$

3) $\frac{\partial y_{p,t}}{\partial\gamma_i} = \begin{cases} 1, & \text{if } I(Q_{i,t+1} \neq \emptyset, p = i) = 1 \\ -1, & \text{if } I(Q_{i,t+1} \neq \emptyset, p = q_{i,t+1}) = 1 \\ 0, & \text{otherwise} \end{cases}, \forall p \in P$

where $U_{p,t} = \partial E\left(g(y_{p,t}, \eta_{p,t}^O)\right)/\partial y_{p,t}$.

Proof: Since only the RHS of port i 's constraint is changed in the first period with $\Delta u_{i,t}^S = -\Delta\gamma_i$ or $\Delta u_{i,t}^D = \Delta\gamma_i$, the perturbation on the optimal transportation cost value in the first period is estimated by $\Delta H(\mathbf{Z}_t) = (-1)^{I(i\in P_t^S)} \cdot \Delta\gamma_i \cdot \pi_{i,t}$ from the sensitivity analysis about RHS of linear programming. By hypothesis the value of $\Delta\gamma_i$ is sufficiently small. The change in the RHS value is within its allowable range and the status of all constraints remains unchanged in the perturbed path. In our problem with real variables, the probability of having degenerate optimal solutions or multiple optimal solutions is close to 0. Hence, $\frac{\partial H(\mathbf{Z}_t)}{\partial\gamma_i} = \lim_{\Delta\gamma_i\to 0}\frac{\Delta H(\mathbf{Z}_t)}{\Delta\gamma_i} = (-1)^{I(i\in P_t^S)} \cdot \pi_{i,t}$.

For assertion 2), by taking the first derivation of the holding and leasing cost function of port p in (6) with respect to the inventory position, it is easy to have the equation.

For assertion 3), recall that the perturbation on $y_{p,t}$ is equal to the perturbation on $x_{p,t+1}$ for all $p \in P$, i.e., $\Delta y_{p,t} = \Delta x_{p,t+1}$. We have $\frac{\partial y_{p,t}}{\partial\gamma_i} = \lim_{\Delta\gamma_i\to 0}\frac{\Delta y_{p,t}}{\Delta\gamma_i} = \lim_{\Delta\gamma_i\to 0}\frac{\Delta x_{p,t+1}}{\Delta\gamma_i}, \forall p \in P$. Hence, if $Q_{i,t+1} = \emptyset$, we have $\partial y_{p,t}/\partial\gamma_i = 0$ for all $p \in P$. Otherwise, we have $\partial y_{i,t}/\partial\gamma_i = 1$, $\partial y_{q_{i,t+1},t}/\partial\gamma_i = -1$, and $\partial y_{p,t}/\partial\gamma_i = 0$ for other $p \in (P - Q_{i,t+1})$. ■

2) Perturbation with Respect to Threshold in the Second Period

If $Q_{i,2} = \emptyset$, the analysis about the perturbation in the second period will be similar with that in the first period. Otherwise, the other case with $Q_{i,t} = \{i, q_{i,t}\}$ for t=2 will be studied.

Similarly, the perturbations on relevant variables are analyzed. We can have $\Delta u_{q_{i,t},t}^S = -\Delta\gamma_i$ or $\Delta u_{q_{i,t},t}^D = \Delta\gamma_i$ and $\Delta u_{p,t}^S$ or $\Delta u_{p,t}^D = 0$ for other $p \in (P - \{q_{i,t}\})$. It implies that only the RHS of port $q_{i,t}$ constraint is changed.

Hence, if the port $q_{i,t}$ constraint is not binding, we have $E_{q_{i,t},t} = \emptyset$ with $\Delta a_{p,t} = 0$ for all $p \in P$. Otherwise, we have $E_{q_{i,t},t} = \{q_{i,t}, e_{q_{i,t},t}\}$ with $\Delta a_{p,t} \neq 0$ for any $p \in E_{q_{i,t},t}$, and $\Delta a_{p,t} = 0$ for other $p \in (P - E_{q_{i,t},t})$.

Differently, the perturbation on $y_{p,t}$ could be from the both perturbations of $\Delta\mathbf{x}_t$ and $\Delta a_{p,t}$ in this period since $\Delta\mathbf{x}_t \neq 0$. Hence, from (1), i.e., $y_{p,t} = x_{p,t} + a_{p,t}$, we obtain:

- If $E_{q_{i,t},t} = \phi$, we have $\Delta y_{i,t} = \Delta\gamma_i$, $\Delta y_{q_{i,t},t} = -\Delta\gamma_i$, and $\Delta y_{p,t} = 0$ for other $p \in (P - \{i, q_{i,t}\})$.
- If $E_{q_{i,t},t} \neq \phi$, we have $\Delta y_{p,t}$ could be $\neq 0$ for any $p = i, e_{q_{i,t},t}$, and $\Delta y_{p,t} = 0$ for other $p \in \left(P - \{i, e_{q_{i,t},t}\}\right)$. Depending on whether $e_{q_{i,t},t} = i$, we get that:

✧If $e_{q_{i,t},t} \neq i$, then it implies that $\Delta y_{i,t} = \Delta x_{i,t} = \Delta\gamma_i$, and $\Delta y_{e_{q_{i,t},t},t} = -\Delta\gamma_i$.

✧If $e_{q_{i,t},t} = i$, then it implies that $\Delta y_{i,t} = 0$.

From the analysis about the perturbation of $\Delta y_{p,t}$ in the case of $Q_{i,t} \neq \emptyset$, we have:

$$Q_{i,t+1} = \begin{cases} \emptyset, & \text{if } E_{q_{i,t},t} \neq \emptyset \text{ and } e_{q_{i,t},t} = i \\ \{i, q_{i,t+1}\}, & \text{otherwise} \end{cases} \quad (20)$$

where $q_{i,t+1}$ is $= q_{i,t}$ when $E_{q_{i,t},t} = \emptyset$, or $= e_{q_{i,t},t}$ when $E_{q_{i,t},t} \neq \emptyset, e_{q_{i,t},t} \neq i$. When $Q_{i,t+1} \neq \emptyset$, the values of the perturbations of $\Delta x_{i,t+1}$ and $\Delta x_{q_{i,t+1},t+1}$ are same with that in the first period.

3) Total Perturbation with Respect to Threshold

Based on the analysis about $Q_{i,t}$ for $t = 1, 2$, it can be concluded that in any period t, $Q_{i,t}$ will be either empty or consists of a pair of ports, i.e., $\{i, q_{i,t}\}$. It implies that the analysis about the perturbations in period $t > 2$ by the perturbation of $\Delta\gamma_i$ will be similar with that in either of the first two periods.

From lemma 4.1 and (18), in a general form, we have

$$\frac{\partial J(N,\gamma)}{\partial\gamma_i} \approx \frac{1}{T}\sum_{t=1}^{T}\begin{pmatrix} I(Q_{i,t} = \emptyset) \bullet (-1)^{I(i\in P_t^S)} \bullet \pi_{i,t} \\ +I(Q_{i,t} \neq \emptyset) \bullet (-1)^{I(q_{i,t}\in P_t^S)} \bullet \pi_{q_{i,t},t} \\ +I(Q_{i,t+1} \neq \emptyset) \bullet \left(U_{i,t} - U_{q_{i,t+1},t}\right) \end{pmatrix} \forall i \in P \quad (21)$$

where the first two terms on the RHS of (21) present the perturbations on the EC repositioning cost under the two conditions, i.e., $Q_{i,t} = \emptyset$ and $Q_{i,t} \neq \emptyset$, respectively; the third term presents the perturbation on the EC holding and leasing cost when $Q_{i,t+1} \neq \emptyset$; For $t = 1$, $Q_{i,t} = \emptyset$, and for $t > 1$, $Q_{i,t}$ can be obtained from either (19) and (20), depending on the status of $Q_{i,t-1}$; the values of $U_{i,t}$ and $U_{q_{i,t+1},t}$ can be calculated from assertion 2) of lemma 4.1.

B. Gradient with Respect to the Fleet Size

In this section, the gradient of expected total cost with respect to the fleet size is studied with given policy $\boldsymbol{\gamma}$.

Similarly, the nominal path is defined as the sample path generated by the simulation model with parameter N and the perturbed path as the sample path generated using the same model and same random seeds, but with parameter $(N)'$, where $(N)' = N + \Delta N$ and ΔN is sufficiently small. Since $\sum_{p \in P} \Delta x_{p,t} = \Delta N$ from (3), we can perturb $x_{i,1}$ by ΔN to investigate the perturbations on relevant variables . That is, we set $q_t^N = i$ with $\Delta x_{q_t^N,t} = \Delta N$ and $\Delta x_{p,t} = 0$ for other $p \in (P - \{q_t^N\})$ in the period $t = 1$.

The perturbation flow for this problem is similar with that shown in Figure 2, but the unique difference is that there is no perturbation from the threshold. In other word, the perturbations on various variables are only from the perturbation on the beginning on-hand inventory.

1) Perturbation with Respect to the Fleet Size in the First Period

In the period $t = 1$, we have $\Delta u^S_{q_t^N,t} = \Delta N$ or $\Delta u^D_{q_t^N,t} = -\Delta N$ and $\Delta u^S_{p,t}$ or $\Delta u^D_{p,t} = 0$ for other $p \in (P - \{q_t^N\})$. It implies that only the RHS of port q_t^N constraint is changed. Hence, we have

- If the port q_t^N constraint is not binding, we have $E_{q_t^N,t} = \emptyset$. And then $\Delta y_{q_t^N,t} = \Delta N$ and $\Delta y_{p,t} = 0$ for other $p \in (P - \{q_t^N\})$.
- Otherwise, we have $E_{q_t^N,t} = \left\{ e_{q_t^N,t} \right\}$. And then $\Delta y_{e_{q_t^N,t},t} = \Delta N$ and $\Delta y_{p,t} = 0$ for other $p \in \left(P - E_{q_t^N,t}\right)$.

Similarly, we have

$$q_{t+1}^N = \begin{cases} q_t^N, & \text{if } E_{q_t^N,t} = \emptyset \\ e_{q_t^N,t}, & \text{otherwise} \end{cases} \tag{22}$$

where $\Delta x_{q_{t+1}^N,t+1} = \Delta N$ and $\Delta x_{p,t+1} = 0$ for other $p \in \left(P - E_{q_t^N,t}\right)$. It implies that in each period, there is a unique port whose beginning on-hand inventory will be affected by ΔN. And we can have next lemma.

Lemma 4.2: In any period t, we have

$$\frac{\partial y_{p,t}}{\partial N} = \begin{cases} 1, \text{if } p = q_{t+1}^N \\ 0, \text{othewise} \end{cases}, \forall p \in P$$

Proof: The proof is similar with that in the assertion 3) of the lemma 4.1. ∎

2) Total Perturbation with Respect to the Fleet Size

From lemma 4.2 and (17), in a general from, we have:

$$\partial J(N, \gamma)/\partial N \approx 1/T \sum_{t=1}^{T} \left((-1)^{I(q_t^N \in P_t^S)+1} \bullet \pi_{q_t^N,t} + U_{q_{t+1}^N,t} \right) \tag{23}$$

where the first term on the right hand side of (23) presents the perturbation on the EC repositioning cost; the second term presents the perturbation on the holding and leasing cost; for $t = 1$, $q_t^N = i$, and for $t > 1$, q_t^N can be obtained from (22); the value of $U_{q_t^N,t}$ can be calculated from the assertion 2) of lemma 4.1.

V. Numerical Results

In this section, we aim to evaluate the performance of the proposed single-level threshold policy (STP) and provide some insights about the EC management for shipping companies.

While the proposed policy can be applied in a multi-port system with an arbitrary number of ports, three problems that differ in the number of ports are considered: problem 1 has 6 ports, problem 2 has 9 ports and problem 3 has 12 ports, which represent small, moderate and large systems, respectively. Besides, since the trade imbalance could be the most important factor affecting the performance of the proposed policy, three kinds of trade imbalance patterns are designed, which include balanced trade pattern, moderately imbalanced trade pattern and severely imbalanced trade pattern. Therefore, a total of 9 cases are studied.

For each problem, the cost parameters and the average customer demands from port p to port m in a period, denoted by $\mu_{p,m}$, are randomly generated. In the balanced trade pattern, we set $\mu_{p,m} = \mu_{m,p}\ \forall p \neq m$ to balance the customer demands between any pair of ports. In the moderately (severely) imbalanced trade pattern, we double (treble) the values of one port's or several ports' exported laden containers and keep other values remain the same as those in the balanced trade pattern. The customer demands, i.e., $\varepsilon_{p,m,t}\ \forall p \neq m$ are assumed to follow normal distribution with mean $\mu_{p,m}$ and standard variance 0.2 $\mu_{p,m}$, and be left-truncated at zero. For example, for problem 1, the holding cost parameter is uniformly generated from the interval (0, 5), the repositioning cost parameter is uniformly taken from the interval (5, 10), and the leasing cost parameter is uniformly generated from the interval (10, 30). This reflects the general view that the repositioning cost parameter is greater than the holding cost parameter, while much less than the leasing cost parameter. The parameter $\mu_{p,m}$ in the balanced trade pattern is uniformly generated from the interval (0, 200). In Appendix C, the cost parameters and the parameters of average customer demands in different trade imbalance patterns of problem 1 are presented.

For a multi-port system with STP, the optimal values of the fleet size and the thresholds can be obtained through sequentially solving the problem (13) under Scenario-I and Scenario-II. Under Scenario-I, the NLP is solved by Matlab (version 7.0.1). Under Scenario-II, the IPA-gradient based algorithm is coded in Visual C++ 5.0. All the numerical studies are tested on and Intel Duo Processor E6750 2.67GHz CPU with 4.00 GB RAM under the Microsoft Vista Operation System. Based on preliminary experiments, we set the simulation period $T = 10100$with 100 warm-up periods. The termination criteria for searching are that the maximum iteration, namely n_{max} is achieve or the total cost per period in current iteration is significantly greater than that in last iteration. We set $n_{max} = 1000$.

A. Policy Performance Evaluation

To evaluate the performance of the proposed STP, a match back policy (MBP) is introduced for comparison. Such policy is widely accepted and applied in practice and its principle is to balance the container flow in each pair of ports. In other words, ECs to be repositioned from port p to port m in period $t + 1$ should try to match the difference between the total number of laden containers exported from port m to port p and the total number of laden containers exported from port p to port m in period t. Mathematically, we have

$$\left(z_{p,m,t+1}\right)_{MBP} = \left(\varepsilon_{m,p,t} - \varepsilon_{p,m,t}\right)^{+} \quad \forall t \qquad (24)$$

When MBP is adopted, the repositioning cost is independent from the fleet size. Hence, the fleet size which minimizes the expected holding and leasing cost will be the optimal fleet size minimizing the expected total cost. We have $(N^*)_{MBP} = \sum_{p \in P} F_p^{-1}\left(C_p^L / \left(C_p^L + C_p^H\right)\right)$.

To facilitate the comparison of STP and MBP, the percentage of expected total cost reduction achieved by STP from MBP is given in Figure 3. We use a doublet, i.e., (number of ports, trade pattern), to present a particular case. For example, (6, B), (6, M-IB), and (6, S-IB) mean the cases for problem 1 with 6 ports in the balanced, moderately imbalanced and severely imbalanced trade patterns, respectively.

There are three main observations from the results. First, STP outperforms MBP in all cases. As expected, the reduction of total cost by STP is major from the reduction of the repositioning cost. Second, as the system becomes larger, STP can reduce more cost. Hence, it is important for shipping companies to use better method in repositioning ECs, instead of resorting to simple way such as the MBP, especially in the complex system. Another observation is that for a problem, it seems that in the imbalanced trade patterns, the advantage of using STP over MBP seems not as great as that in the balanced trade pattern. However, STP still can reduce cost from MBP in the imbalanced trade pattern, such as over 30% cost reduction for the problem with 12 ports.

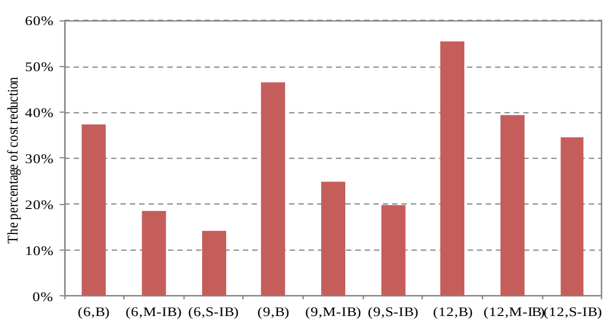


Figure 3. Percentages of expected total cost per period reduction achieved by STP from MBP

B. Policy Performance Sensitivity to the Fleet Size

Considering the fact that container fleet is often fixed by shipping companies in practice, we further investigate the policy performance under both policies with given fleet size. If the fleet size is given, the optimal thresholds under STP can be found by the two proposed approaches with little adjustment. Hence, let N^* be the optimal fleet size under STP. The fleet size is varied from $0.7N^*$ to $1.3N^*$ in all 7 cases to investigate the affect of the fleet size on the expected total cost.

Figure 4 shows the expected total cost per period under both policies in case (6, B) with different fleet sizes. First, it is also observed that STP outperforms MBP in all different fleet size cases. It reveals that the expected total cost per period savings achieved by STP over MBP are of the order of 13.18%~37.72%. One possible explanation is that the proposed policy makes the ECR decisions in terms of minimizing the repositioning cost. Besides, the trend of the diamond line shows that, the cost saving achieved by STP from MBP increases gradually when the fleet size increases, and as the fleet size increases further the saving decreases. It is also possible to have a small cost saving percentage when the fleet size is too little. The reason is that when the fleet size is severely insufficient to satisfy the customer demands, a large number of ECs are leased and few requirements for repositioning ECs. Third, the results also show that the minimum expected total cost per period appears to be convex with respect to the fleet size under both policies. It reflects the intuition that the optimal fleet size is the trade-off between the repositioning cost and the holding and leasing cost.

Similarly observations can be also found in other cases.

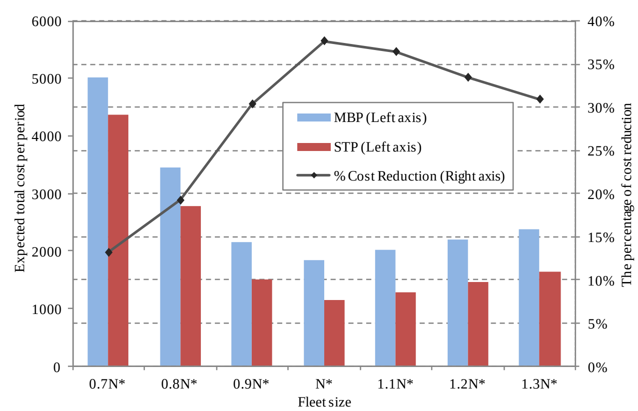


Figure 4. Comparison of STP and MBP in case (6, B)

C. *Sensitivity Analysis of the Thresholds*

Since the single-level threshold policy is applied to manage ECs in the multi-port system, the thresholds of the policy will significantly affect the performance of the proposed policy. Many factors may impact the thresholds of the policy. The most significant factors are the fleet size and the parameters of leasing cost and holding cost. Next, based on the case (6, B), we explore the sensitivity of the thresholds.

As shown in Figure 5, the optimal threshold values generally increase as the fleet size increases. More interesting, as the fleet size increases from $1.1N^*$ to $1.3N^*$, ports 1~5 have very small increments on their thresholds, while port 6 has large increment on its threshold. This reflects the fact that as the fleet size is much greater than its optimal value, the low holding cost at port 6 works in favor of keeping a large number of ECs at this port. While as the fleet size decrease from $0.9N^*$ to $0.7N^*$, port 3 has the largest decrement on its threshold. A possible explanation is that as owned ECs are insufficient at all ports, many leasing containers have to be used to satisfy the customer demands. And the low leasing cost at port 3 supports it to keep low inventory level of ECs and lease a large number of ECs.

Next, the sensitivity of the thresholds to the holding and leasing cost parameters is considered. From (21), we obtain that $\partial\left(\frac{\partial(J(N,\boldsymbol{\gamma}))}{\partial\gamma_i}\right)/\partial C_i^L \le 0$ and $\partial\left(\frac{\partial(J(N,\boldsymbol{\gamma}))}{\partial\gamma_i}\right)/\partial C_{q_{i,t}}^L \ge 0$. It implies that $\frac{\partial(J(N,\boldsymbol{\gamma}))}{\partial\gamma_i}$ will decrease (increase) as C_i^L ($C_{q_{i,t}}^L$) increases. Thus, when the leasing cost of port i increases, the threshold for this port will increase while the thresholds for some other ports decrease. The similar property for the holding cost parameter can be derived. That is, as the holding cost of port i increases, the threshold of this port will decrease while the thresholds for some other ports increase.

Focusing on the case (6, B) with optimal fleet size under STP, we consider two more cases, i.e., cases A and B, in which the holding and leasing cost parameters of port 6 are increased by 2 times, respectively. Figure 6 shows the results about thresholds changes by cases A and B from the original case. It can be observed that when the holding (leasing) cost

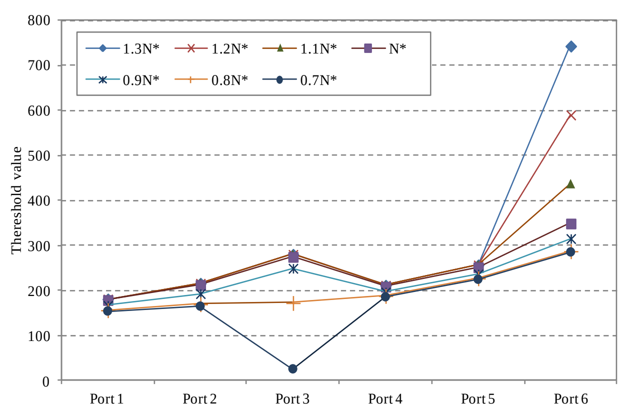


Figure 5. The optimal threshold value for case (6, B) under STP

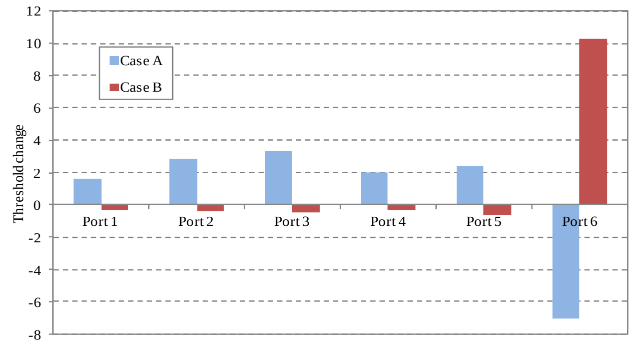


Figure 6. Optimal thresholds changes in cases A and B

parameter of port 6 increases by 2 times, the threshold of port 6 decreases (increases), but the thresholds of other ports increase (decrease). This testifies above phenomenon. Hence, the results reflect the fact that a higher leasing cost at a port works in favor of keeping a large number of ECs while a higher holding cost encourages repositioning out more ECs in a surplus port or repositioning in less ECs in a deficit port.

VI. CONCLUSION AND FUTURE WORK

In this paper, the joint ECR and container fleet sizing problem in a multi-port system is addressed. A single-level threshold policy is developed to reposition ECs periodically by taking into account customer demand uncertainty and dynamic operations. Two approaches, a NLP and IPA-based gradient technique are developed to solve the problem optimizing the fleet size and the parameters of thresholds under Scenario-I and Scenario-II, respectively. The numerical results provide insight that by repositioning the ECs intelligently, the total cost can be significantly reduced.

The main contributions of the study are as follows: (a) a single-level threshold policy with a repositioning rule in terms of minimizing repositioning cost was developed for repositioning ECs in a multi-port system. To the best of our knowledge, few works considered the repositioning rule related to the repositioning cost; (b) the optimal values of the fleet size and the thresholds of the policy can be obtained. These values could be used as reference points for shipping companies to make strategic decisions; (c) by developing the method to solve the difficult EC management problem, i.e., using IPA to estimate gradient, it is innovative and provides a potential methodology contribution in this field.

There are several limitations in the paper. Firstly, ECs were assumed to be dispatched in the right one period. Further research is needed to relax the one-period assumption and consider the problem with different dimensions for the repositioning time. The main challenge is to track the perturbations along the sample path. The second limitation is that the simulation time of the proposed algorithm could be further reduced. Although the proposed algorithm can be utilized to solve the problem with an arbitrary number of ports, as the number of ports increase, more simulation time is needed to search the optimal solution. For example, given a starting point, the proposed algorithm under Scenario-II takes about 1 hour to get the minimum expected total cost for the small size system, and more time for a larger system. Thus, improving the effectiveness of the proposed algorithm should be an interesting area for future research. We could consider combining the gradient search approach with the evolutionary approach to take advantage of both approaches. In addition, a comparison of the proposed algorithms with other meta-heuristics could be another possible research direction.

APPENDIX A

Note that when ECs are repositioned in a period, the repositioning cost will be the minimum objective values of the transportation models. Hence, the constraints of ports under both scenarios are investigated and some lemmas will be presented, which will be used in proving proposition 2.1.

From property I, we can get that under Scenario-I, $\mathbf{y}_t^{\mathrm{I}} = \boldsymbol{\gamma}^{\mathrm{I}}, \forall t$. Similarly, under Scenario-II, we can have $\mathbf{y}_t^{\mathrm{II}} \geq \boldsymbol{\gamma}^{\mathrm{II}}$ $(\leq \boldsymbol{\gamma}^{\mathrm{II}})$ if N is $> \sum_{p\in P}\gamma_p^{\mathrm{II}}$ $(< \sum_{p\in P}\gamma_p^{\mathrm{II}})$. It can be explained as follows. When $N$is $> \sum_{p\in P}\gamma_p^{\mathrm{II}}$ $(< \sum_{p\in P}\gamma_p^{\mathrm{II}})$, the transportation model in each period should be unbalanced, i.e., the total number of estimated EC supplies, namely $\sum_{i\in P_t^S} u_{i,t}^S$ is greater (less) than the total number of estimated EC demands, namely $\sum_{j\in P_t^D} u_{j,t}^D$. This unbalanced situation, such as $\sum_{i\in P_t^S} u_{i,t}^S$ is $> \sum_{j\in P_t^D} u_{j,t}^D$, means that all EC demands at the deficit ports can be satisfied and some excess ECs still stay at some surplus ports. Hence, the inventory position of each port in each period could be not less (greater) than its target threshold level when N is $> \sum_{p\in P}\gamma_p^{\mathrm{II}}$ $(< \sum_{p\in P}\gamma_p^{\mathrm{II}})$. The surplus port and deficit port subsets under both scenarios are investigated firstly.

Lemma A.1: For all period t, we have $P_{t+1}^{S,\mathrm{I}} \subseteq P_{t+1}^{S,\mathrm{II}}$ and $P_{t+1}^{D,\mathrm{II}} \subseteq P_{t+1}^{D,\mathrm{I}}$ if $N > \sum_{p\in P}\gamma_p^{\mathrm{II}}$; $P_{t+1}^{S,\mathrm{II}} \subseteq P_{t+1}^{S,\mathrm{I}}$ and $P_{t+1}^{D,\mathrm{I}} \subseteq P_{t+1}^{D,\mathrm{II}}$, if $N < \sum_{p\in P}\gamma_p^{\mathrm{II}}$ with same customer demand ω_t.

Proof: With the help of (2), we have $x_{p,t+1}^{\mathrm{I}} = \gamma_p^{\mathrm{I}} + \varphi_{p,t}, \forall p \in P$. It implies that $\varphi_{i,t} > 0, \forall i \in P_{t+1}^{S,\mathrm{I}}$ and $\varphi_{j,t} < 0, \forall j \in P_{t+1}^{D,\mathrm{I}}$.

Note that both scenarios are investigated with same ω_t. We have $x_{i,t+1}^{\mathrm{II}} > y_{i,t}^{\mathrm{II}} \geq \gamma_i^{\mathrm{II}}, \forall i \in P_{t+1}^{S,\mathrm{I}}$ if $N > \sum_{p\in P}\gamma_p^{\mathrm{II}}$, and $x_{j,t+1}^{\mathrm{II}} < y_{j,t}^{\mathrm{II}} \leq \gamma_j^{\mathrm{II}}, \forall j \in P_{t+1}^{D,\mathrm{I}}$ if $N < \sum_{p\in P}\gamma_p^{\mathrm{II}}$. It implies that $P_{t+1}^{S,\mathrm{I}} \subseteq P_{t+1}^{S,\mathrm{II}}$ if $N > \sum_{p\in P}\gamma_p^{\mathrm{II}}$, and $P_{t+1}^{D,\mathrm{I}} \subseteq P_{t+1}^{D,\mathrm{II}}$ if $N < \sum_{p\in P}\gamma_p^{\mathrm{II}}$.

Note that $P_t^S \cup P_t^D = P$ for any scenario. We have $P_{t+1}^{D,\mathrm{II}} \subseteq P_{t+1}^{D,\mathrm{I}}$ if $N > \sum_{p\in P}\gamma_p^{\mathrm{II}}$ and $P_{t+1}^{S,\mathrm{II}} \subseteq P_{t+1}^{S,\mathrm{I}}$ when $N < \sum_{p\in P}\gamma_p^{\mathrm{II}}$. ■

From lemma A.1, there are only two possible cases about the port subsets under both scenarios. One is that there are same port subsets. The other is that there are greater (smaller) surplus port subset and smaller (greater) deficit port subset under Scenario-II than that under Scenario-I when the fleet size is greater (less) than the sum of thresholds. Furthermore, for the port that is common to surplus or deficit port subsets under both scenarios, its quantity of excess (deficit) ECs is investigated.

Lemma A.2: For all period t, we have

1) If $N > \sum_{p\in P}\gamma_p^{\mathrm{II}}$, for $i \in P_{t+1}^{S,\mathrm{I}}$ and $j \in P_{t+1}^{D,\mathrm{II}}$ $(P_{t+1}^{D,\mathrm{II}} \neq \emptyset)$, we have $u_{i,t+1}^{S,\mathrm{II}} \geq u_{i,t+1}^{S,\mathrm{I}}$ and $u_{j,t+1}^{D,\mathrm{II}} \leq u_{j,t+1}^{D,\mathrm{I}}$.

2) If $N < \sum_{p\in P}\gamma_p^{\mathrm{II}}$, for $i \in P_{t+1}^{S,\mathrm{II}}$ $(P_{t+1}^{S,\mathrm{II}} \neq \emptyset)$ and $j \in P_{t+1}^{D,\mathrm{I}}$, we have $u_{i,t+1}^{S,\mathrm{II}} \leq u_{i,t+1}^{S,\mathrm{I}}$ and $u_{j,t+1}^{D,\mathrm{II}} \geq u_{j,t+1}^{D,\mathrm{I}}$.

with same customer demand ω_t.

Proof: From (2), (7) and (8), we get that $u_{i,t+1}^S = y_{i,t} + \varphi_{i,t} - \gamma_i$ and $u_{j,t+1}^D = \gamma_j - y_{j,t} - \varphi_{j,t}, \forall i \in P_{t+1}^S, j \in P_{t+1}^D$.

Hence, with help of lemmas A.1, when $N > \sum_{p\in P}\gamma_p^{\mathrm{II}}$, we have $u_{i,t+1}^{S,\mathrm{II}} \geq \varphi_{i,t} = u_{i,t+1}^{S,\mathrm{I}}, \forall i \in P_{t+1}^{S,\mathrm{I}}$, and if $P_{t+1}^{D,\mathrm{II}}$ is non-empty, $u_{j,t+1}^{D,\mathrm{II}} \leq -\varphi_{j,t} = u_{j,t+1}^{D,\mathrm{I}}, \forall j \in P_{t+1}^{D,\mathrm{II}}$.

Similarly, when $N < \sum_{p\in P}\gamma_p^{\mathrm{II}}$, we have $u_{i,t+1}^{S,\mathrm{II}} \leq u_{i,t+1}^{S,\mathrm{I}}, \forall i \in P_{t+1}^{S,\mathrm{II}}$ if $P_{t+1}^{S,\mathrm{II}} \neq \emptyset$, and $u_{j,t+1}^{D,\mathrm{II}} \geq u_{j,t+1}^{D,\mathrm{I}}, \forall j \in P_{t+1}^{D,\mathrm{I}}$. ■

From lemma A.2, it implies that for example, when the fleet size is greater than the sum of thresholds, for the common surplus port under both scenarios, its amount of excess ECs under Scenario-II is not less than that under Scenario-I.

We return now to the proof of proposition 2.1.

Proof: Since $P_t^{S,\mathrm{I}}$ and $P_t^{D,\mathrm{I}}$ are both nonempty, from the transportation model, we have $H(\mathbf{x}_{t+1}^{\mathrm{I}}, \boldsymbol{\gamma}^{\mathrm{I}})|_{\omega_t} > 0$. Hence, when either $P_{t+1}^{D,\mathrm{II}}$ or $P_{t+1}^{S,\mathrm{II}}$ is empty, we can get that $H(\mathbf{x}_{t+1}^{\mathrm{II}}, \boldsymbol{\gamma}^{\mathrm{II}})|_{\omega_t} = 0 < H(\mathbf{x}_{t+1}^{\mathrm{I}}, \boldsymbol{\gamma}^{\mathrm{I}})|_{\omega_t}$.

When $P_{t+1}^{D,\mathrm{II}}$ is nonempty, in the case of $N > \sum_{p\in P}\gamma_p^{\mathrm{II}}$, the constraints of the transportation models in period $t+1$ under both scenarios should be (10) and (11), which are as follows:

$$\sum_{j\in P_{t+1}^D} z_{i,j,t+1} \leq u_{i,t+1}^S, \quad \forall i \in P_{t+1}^S$$

$$-\sum_{i\in P_{t+1}^S} z_{i,j,t+1} \leq -u_{j,t+1}^D, \quad \forall j \in P_{t+1}^D.$$

Based on the sensitivity analysis about RHS of linear programming, we get that increasing the RHS coefficient value of above constraint could reduce the minimum objective value. Depending on the port subsets under both scenarios, from lemmas A.1 and A.2, there are only two possible cases:

- Case 1: $P_{t+1}^{S,\mathrm{II}} = P_{t+1}^{S,\mathrm{I}}$ and $P_{t+1}^{D,\mathrm{II}} = P_{t+1}^{D,\mathrm{I}}$. It implies that the transportation models under both scenarios have similar constraints for all ports $i \in P_{t+1}^{S,\mathrm{I}}, j \in P_{t+1}^{D,\mathrm{I}}$. Hence, based on the model under the scenario-I, we increase the RHS values from $u_{i,t+1}^{S,\mathrm{I}}$ and $-u_{j,t+1}^{D,\mathrm{I}}$ up to $u_{i,t+1}^{S,\mathrm{II}}$ and $-u_{j,t+1}^{D,\mathrm{II}}$, respectively. Then, the new model will be equivalent to the model under scenario-II, and we can have $H(\mathbf{x}_{t+1}^{\mathrm{II}}, \boldsymbol{\gamma}^{\mathrm{II}})|_{\omega_t} \leq H(\mathbf{x}_{t+1}^{\mathrm{I}}, \boldsymbol{\gamma}^{\mathrm{I}})|_{\omega_t}$.

• Case 2: $P_{t+1}^{S,\mathrm{II}} \supset P_{t+1}^{S,\mathrm{I}}$ and $P_{t+1}^{D,\mathrm{II}} \subset P_{t+1}^{D,\mathrm{I}}$. It implies that the transportation models under both scenarios have similar constraints for some ports $i \in P_{t+1}^{S,\mathrm{I}}, j \in P_{t+1}^{D,\mathrm{II}}$. Similarly, based on the model under the scenario-I (scenario-II), we increase (decrease) the RHS values of ports that are $\in \left(P_{t+1}^{D,\mathrm{I}} - P_{t+1}^{D,\mathrm{II}}\right)$ $\left(\left(P_{t+1}^{S,\mathrm{II}} - P_{t+1}^{S,\mathrm{I}}\right)\right)$ up (down) to zero. Then, the minimum objective cost of the new model-I (model-II), denoted by C' (C''), should be not greater (less) than that of the original model, i.e., $C' \leq H(\mathbf{x}_{t+1}^{\mathrm{I}}, \boldsymbol{\gamma}^{\mathrm{I}})|_{\omega_t}$ ($C'' \geq H(\mathbf{x}_{t+1}^{\mathrm{II}}, \boldsymbol{\gamma}^{\mathrm{II}})|_{\omega_t}$). For the new model-I and model-II, they have similar constraints for ports $i \in P_{t+1}^{S,\mathrm{I}}, j \in P_{t+1}^{D,\mathrm{II}}$. Hence, as that in case 1, we have $C'' \leq C'$. Finally, we get that $H(\mathbf{x}_{t+1}^{\mathrm{II}}, \boldsymbol{\gamma}^{\mathrm{II}})|_{\omega_t} \leq C'' \leq C' \leq H(\mathbf{x}_{t+1}^{\mathrm{I}}, \boldsymbol{\gamma}^{\mathrm{I}})|_{\omega_t}$.

Similarly, when $P_{t+1}^{S,\mathrm{II}} \neq \emptyset$ in the case of $\sum_{p \in P} \gamma_p^{\mathrm{II}} > N$, we can get that $H(\mathbf{x}_{t+1}^{\mathrm{II}}, \boldsymbol{\gamma}^{\mathrm{II}})|_{\omega_t} \leq H(\mathbf{x}_{t+1}^{\mathrm{I}}, \boldsymbol{\gamma}^{\mathrm{I}})|_{\omega_t}$. ■

APPENDIX B

A numerical example is firstly presented to show that how the perturbation in the supply or demand of one binding port, such as i, will change the net actual imported ECs of a pair of ports. Then, a modified stepping stone (MSS) method is proposed to find $E_{i,t}$.

Consider a numerical example with three surplus ports and two deficit ports. ECs are repositioned from the surplus ports to the deficit ports. Assume the total number of EC supplies is greater than the total number of EC demands. And the transportation tableau is presented in Figure 7. Note that one dummy node is created in the transportation tableau to ensure that the total number of demands is equal to the total number of supplies. For illustration, the value of cell (m, l) represents the number of ECs repositioned from the m^{th} surplus port to the l^{th} deficit port. Two sub-examples are investigated.

Surplus port	Deficit port			
	1	2	3(dummy)	Change of supply
1	•			
2	•		•	
3		•	•	
Change of demand				

Figure 7. The transportation tableau
● Basic cell

Sub-example 1: Perturb the number of EC supply of the first surplus port by $\Delta\gamma$

Consider the balance of the transportation tableau. Perturbing the number of EC supply of the first surplus port by $\Delta\gamma$ implies that the number of EC demand of the dummy deficit port will be perturbed by $\Delta\gamma$. However, since cell $(1, 3)$ is a non-basic cell (see Figure 7), its value should be kept as zero and cannot be changed by the perturbation. To track the changes on the basic variables, a loop is developed, which begins at cell $(1, 3)$ and is back to this cell (see Figure 8).

Surplus port	Deficit port			
	1	2	3(dummy)	Change of supply
1	$\Delta\gamma$ •		◎	$\Delta\gamma$
2	$-\Delta\gamma$ •		• $\Delta\gamma$	
3		•	•	
Change of demand			$\Delta\gamma$	

Figure 8. Perturb the number of EC supply of the first surplus port by $\Delta\boldsymbol{\gamma}$
◎ Beginning cell

The loop in Figure 8 consists of successive horizontal and vertical segments whose end nodes must be basic variables, except for the two segments starting and ending at the non-basic variable.

More specifically, the changes on the basic variables can be obtained as follows: first increase the value of cell $(1, 1)$ by $\Delta\gamma$; then go around the loop, alternately decrease and then increase basic variables in the loop by $\Delta\gamma$, i.e. decrease the value of cell $(2, 1)$ and increase the value of cell $(2, 3)$ by $\Delta\gamma$. It is observed that the total numbers of repositioning out ECs at the first and the second surplus ports will be increased and decreased by $\Delta\gamma$, respectively.

Sub-example 2: Perturb the number of EC demand of the first deficit port by $\Delta\gamma$

Perturbing the number of EC demand of a deficit port (not the dummy deficit port) by $\Delta\gamma$ will result in the total number of demands from deficit ports (excluding the dummy deficit port) increasing by $\Delta\gamma$. To maintain the balance of the transportation tableau, a hypothetic surplus port is introduced and its supply is assumed to increase by $\Delta\gamma$. Besides, to make the transportation tableau non-degeneracy, the cell (the hypothetic surplus port, the dummy deficit port), i.e., cell (4, 3) in Figure 9, is set as a basic cell.

To track the changes on the basic variables by the perturbation of $\Delta\gamma$ on the number of EC demand of the first deficit port, similar loop is developed in Figure 9. It is observed that the total number of repositioning in ECs at the first deficit port and the total number of repositioning out ECs at the second surplus port will be both increased by $\Delta\gamma$.

Just as the loop using in the typical stepping stone method, the loop in either Figure 8 or Figure 9 is unique. The uniqueness of the loop guarantees that there will be only a pair of ports, whose net actual imported ECs will be affected by perturbing the supply or demand of one binding port.

Surplus port	Deficit port			
	1	2	3(dummy)	Change of supply
1	•			
2	$\Delta\gamma$ •		• $-\Delta\gamma$	
3		•	•	
4 (hypothetic)	◎		• $\Delta\gamma$	$\Delta\gamma$
Change of demand	$\Delta\gamma$			

Figure 9. Perturb the number of EC demand of the first deficit port by $\Delta\boldsymbol{\gamma}$

TABLE II. LIST OF ADDITIONAL NOTATIONS

	Description
M_t,	the number of surplus ports in period t
L_t	the number of deficit ports in period t
$I_{i,t}^S$	the index of the surplus port i in the transportation tableau of period t
$I_{j,t}^D$	the index of the deficit port j in the transportation tableau of period t

Similar results can be obtained under the other case in which the total number of EC supplies is less than the total number of EC demands.

Some additional notations are defined Table II.

When the number of EC supply or demand of port i in period t is perturbed, a MSS method is proposed to find $E_{i,t}$ with the rules as follows.

MSS method to find $E_{i,t}$

Step 1. Build the transportation tableau based on the transportation model solutions in period t. Create a dummy deficit node, $L_t + 1$ and a dummy surplus node, $M_t + 1$. And arbitrarily insert the value ε into cell($M_t + 1$, $L_t + 1$).

Step 2. If port i is a surplus port in period t, i.e. $i \in P_t^S$, select cell ($I_{i,t}^S$, $L_t + 1$); if port i is a deficit port in period t, i.e. $i \in P_t^D$, select cell ($M_t + 1, I_{i,t}^D,$).

Step 3. If the selected cell is a basic cell, no port's total repositioned empties quantity will be affected, set $E_{i,t} = \emptyset$ and stop. Otherwise, go to Step 4.

Step 4. Beginning at the selected cell, trace a loop back to the cell, turning corners only on basic cells. Only successive horizontal & vertical moves are allowed.

Step 5. If the total EC supply is more than the total EC demand, record the basic cell $(v, L_t + 1)$, $v \neq (M_t + 1)$ in this loop and find port k with $I_{k,t}^S = v$. Otherwise, record the basic cell ($M_t + 1, v$) , $v \neq (L_t + 1)$ in this loop and find port k with $I_{k,t}^D = v$. Then, $E_{i,t} = \{i, e_{i,t}\}$ with $e_{i,t} = k$.

APPENDIX C

The parameters of problem 1 are presented Table III and Table IV.

TABLE III. THE COST PARAMETERS FOR PROBLEM 1

Original port p	Destination port m	$C_{p,m}^R$	C_p^H	C_p^L
1	2	8.421	4.396	16.537
	3	9.917		
2	1	6.761	2.374	12.395
	3	5.848		
3	1	9.027	1.900	24.599
	2	7.053		

TABLE IV. THE PARAMETERS OF AVERAGE CUSTOMER LADEN DEMAND IN DIFFERENT TRADE IMBALANCE PATTERNS FOR PROBLEM 1

Original port p	Destination port m	$\mu_{p,m}$		
		Balanced	Moderately imbalanced	Severely imbalanced
1	2	95.950	191.900	287.850
	3	126.385	252.771	379.155
2	1	95.950	95.950	95.950
	3	176.833	176.833	176.833
3	1	126.385	126.385	126.385
	2	176.833	176.833	176.833

REFERENCES

[1] L. H. Lee, E. P. Chew, and Y. Luo, "Inventory-Based Empty Container Repositioning in a Multi-Port System," in *INFOCOMP 2011, The First International Conference on Advanced Communications and Computation*, 2011, pp. 80-85.

[2] Economic and social commission for Asia and the Pacific. (2007). *Regional Shipping and Port Development: Container Traffic Forecast 2007 Update* [Online; accessed 18-Dec-2012]. Available: http://www.unescap.org/ttdw/publications/tis_pubs/pub_2484/pub_2484_fulltext.pdf

[3] United Nations Conference on Trade and Development. (2011). *Review of maritime transport* [Online;accessed 18-Dec-2012]. Available: http://unctad.org/en/Docs/rmt2011_en.pdf

[4] D. P. Song et al., "On cost-efficiency of the global container shipping network," *Maritime Policy & Management*, vol. 32, pp. 15-30, 2005.

[5] T. G. Crainic, P. Dejax, and L. Delorme, "Models for multimode multicommodity location problems with interdepot balancing requirements," *Annals of Operations Research*, vol. 18, pp. 277-302, 1989.

[6] B. Bourbeau, T. G. Crainic, and B. Gendron, "Branch-and-bound parallelization strategies applied to a depot location and container fleet management problem," *Parallel Computing*, vol. 26, pp. 27-46, 2000.

[7] I. K. Moon, A. D. D. Ngoc, and Y. S. Hur, "Positioning empty containers among multiple ports with leasing and purchasing considerations," *OR Spectrum*, vol. 32, pp. 765-786, 2010.

[8] Q. Meng and S. Wang, "Liner shipping service network design with empty container repositioning," *Transportation Research Part E: Logistics and Transportation Review*, vol. 47, pp. 695-708, 2011.

[9] R. K. Cheung and C. Y. Chen, "A two-stage stochastic network model and solution methods for the dynamic empty container allocation problem," *Transportation Science*, vol. 32, pp. 142-162, 1998.

[10] S. W. Lam, L. H. Lee, and L. C. Tang, "An approximate dynamic programming approach for the empty container allocation problem," *Transportation Research Part C*, vol. 15, pp. 265-277, 2007.

[11] Y. Long, L. H. Lee, and E. P. Chew, "The sample average approximation method for empty container

repositioning with uncertainties," *European Journal of Operational Research*, vol. 222, pp. 65-75, 2012.

[12] S. T. Choong, M. H. Cole, and E. Kutanoglu, "Empty container management for intermodal transportation networks," *Transportation Research Part E*, vol. 38, pp. 423-438, 2002.

[13] J. X. Dong and D. P. Song, "Container fleet sizing and empty repositioning in liner shipping systems," *Transportation Research Part E*, vol. 45, pp. 860-877, 2009.

[14] D. P. Song and J. Carter, "Optimal empty vehicle redistribution for hub-and-spoke transportation systems," *Naval Research Logistics*, vol. 55, pp. 156 - 171, 2008.

[15] J. A. Li, K. Liu, S. C. H. Leung, and K. K. Lai, "Empty container management in a port with long-run average criterion," *Mathematical and Computer Modelling*, vol. 40, pp. 85-100, 2004.

[16] D. P. Song and Q. Zhang, "A Fluid Flow Model for Empty Container Repositioning Policy with a Single Port and Stochastic Demand," *SIAM Journal on Control and Optimization*, vol. 48, pp. 3623-3642, 2010.

[17] D. P. Song, "Characterizing optimal empty container reposition policy in periodic-review shuttle service systems," *Journal of the Operational Research Society*, vol. 58, pp. 122-133, 2007.

[18] D. P. Song and C. F. Earl, "Optimal empty vehicle repositioning and fleet-sizing for two-depot service systems," *European Journal of Operational Research*, vol. 185, pp. 760-777, 2008.

[19] D. Song, J. Dong, and M. Roe, "Optimal container dispatching policy and its structure in a shuttle service with finite capacity and random demands," *International Journal of Shipping and Transport Logistics*, vol. 2, pp. 44-58, 2010.

[20] D. P. Song and J. X. Dong, "Empty Container Management in Cyclic Shipping Routes," *Maritime Economics & Logistics*, vol. 10, pp. 335-361, 2008.

[21] D. P. Song and J. X. Dong, "Effectiveness of an empty container repositioning policy with flexible destination ports," *Transport Policy*, vol. 18, pp. 92-101, 2010.

[22] Y. W. Yun, M. Y. Lee, and S. Y. Choi, "Optimal inventory control of empty containers in inland transportation system," *International Journal of Production Economics*, vol. 133, pp. 451-457, 2011.

[23] J. A. Li, S. C. H. Leung, Y. Wu, and K. Liu, "Allocation of empty containers between multi-ports," *European Journal of Operational Research*, vol. 182, pp. 400-412, Oct. 2007.

[24] Y. Ho, M. Eyler, and T. Chien, "A gradient technique for general buffer storage design in a production line," *International Journal of Production Research*, vol. 17, pp. 557-580, 1979.

[25] R. Suri, "Perturbation analysis: the state of the art and research issuesexplained via the GI/G/1 queue," *Proceedings of the IEEE*, vol. 77, pp. 114-137, 1989.

[26] F. Azadivar, "Simulation optimization methodologies," in *Proceedings of the 31st conference on Winter simulation*, 1999, pp. 93-100.

[27] M. C. Fu, "Sample path derivatives for (s, S) inventory systems," *Operations Research*, vol. 42, pp. 351-364, 1994.

[28] Y. Wardi, R. Adams, and B. Melamed, "A unified approach to infinitesimal perturbation analysis in stochastic flow models: the single-stage case," *Automatic Control, IEEE Transactions on*, vol. 55, pp. 89-103, 2010.

[29] M. Chen, J. Q. Hu, and M. C. Fu, "Perturbation analysis of a dynamic priority call center," *Automatic Control, IEEE Transactions on*, vol. 55, pp. 1191-1196, 2010.

[30] C. G. Cassandras, X. C. Ding, and X. Lin, "An optimal control approach for the persistent monitoring problem," in *Decision and Control and European Control Conference (CDC-ECC), 2011 50th IEEE Conference on*, 2011, pp. 2907-2912.

[31] W. P. Wong, W. Jaruphongsa, and L. H. Lee, "Budget allocation for effective data collection in predicting an accurate DEA efficiency score," *Automatic Control, IEEE Transactions on*, vol. 56, pp. 1235-1246, 2011.

[32] D. Özdemir, E. Yücesan, and Y. T. Herer, "Production, Manufacturing and Logistics Multi-Location Transshipment Problem With Capacitated Production," *European Journal of Operational Research*, In press.

Visual Customer Behavior Analysis at the Point of Sale

Johannes Kröckel, Freimut Bodendorf
Department of Information Systems
University of Erlangen-Nuremberg
Lange Gasse 20, Nuremberg, Germany
{johannes.kroeckel, freimut.bodendorf}@wiso.uni-erlangen.de

***Abstract*— In times of low-cost broadband and mobile internet flat rates advantages such as continuous availability and a large range of products enable online retailers to mature. Since competition grows, web shop owners have to adapt to their customers to remain competitive. Therefore, they record and analyze data about all kinds of customer activities on the web shop to derive optimization potentials and to offer customized services. Stationary retailers, which are regularly in direct competition to web shops lack these possibilities. They are limited to data derived from enterprise resource planning systems, checkout systems or loyalty cards. Hence, the behavior of customers at the point of sale cannot be considered, yet. To address this lack of information an approach is presented that applies video and infrared cameras to record and analyze customer movements and activities at the point of sale. The approach aims at extracting valuable information for the management and sales staff of stationary retailers. Based on the information, services are developed to support decisions regarding the store structure, product range and customer approach.**

Customer tracking and tracing, Customer behavior, Retailer Support.

I. INTRODUCTION

Low prices, short delivery periods and 24/7 availability are only three reasons that help web shops to extend their customer base at the expense of stationary retailers. Besides, there are few reasons left for customers to buy goods like books or music in stationary stores since a physical experience is not required. Consequently, stationary shop operators have to come up with sophisticated, individual approaches for attracting and retaining customers. For this, knowledge about the customers as well as their on-site buying behavior is required [1].

While click paths, bounce rates and page impressions as well as time spent on websites are common key figures for internet shops stationary retailers lack any sources of information about their customers' behavior [2][3][4]. Data recorded from electronic checkout counters or merchandise planning and control systems are not sufficient to reveal individual customer behavior.

Companies like Envirosell or Shopperception try to overcome this information shortage by data gathered from manually conducted observations [5]. However, these strategies are designed for a limited time only. Continuous observation over a longer period of time like weeks or months would be very expensive due to the required human resources. Besides that, an objective documentation of results by the observing persons cannot be guaranteed. To enable quick reaction to contemporary customer behavior a continuous automated monitoring similar to the one for web shops is needed.

Movements and activities can be considered as real world equivalents of clicks. Movements describe how customers walk through the shopping environment and provide useful data on their speed, regions of interest and behavior towards other people in the surrounding area. Activities in terms of interactions with products enable to gain information about viewed or purchased products and therefore about customers' buying behavior.

The extraction of movement and activity data using various kinds of sensors is strongly discussed in fields like computer vision and data mining (see Section II). However, data is rarely used for gaining information for retail managers and sales personnel. Besides that, tracking of movements and activities is mainly conducted with expensive high tech equipment, which enables scientific applications but rarely allows feasible solutions for real world applications. Although companies like Visapix and Vitracom offer tracking systems for sensing and analyzing position data their software solutions require specific hardware components. The additional hardware expenses often exceed retailer's budgets. Moreover, these software solutions provide limited data analysis. Therefore, a cost-efficient and practical approach is required.

For the proposed concept existing algorithms are combined and work with low-cost sensors. The extracted data is analyzed in a way to allow retailers to improve their retail environment. Thus, they can manage better to retain existing and attract new customers.

To gain this information, first, raw data about the movements and activities need to be recorded. Therefore, Section III.A describes the extraction of walking paths using surveillance cameras and methods from the field of video mining. Then, Section III.B gives an overview of capturing and extracting activities using the Microsoft Kinect Controller and data mining methods.

Analyses of the derived raw data are described in the succeeding sections IV and V. In Section IV data are processed to gain an overview of the comprehensive behavior of customers at the point of sale. Subsequently, in Section V the behavior of individual customers is considered.

II. Related Work

In recent years, a variety of sensor-based solutions has been developed to record context information [6][7][8]. Especially the extraction and analysis of location data is a frequently discussed topic. The approaches mainly use gathered location data in the field of ubiquitous and mobile computing [9][10][11]. Therefore, the authors apply cell phone compatible positioning technologies like GSM or GPS for outdoor location tracking. The presented approach requires position determination inside a store. That means, it has to be more accurate than GSM and, in contrast to GPS, available indoors [12]. Indoor position localization is among others achieved using technologies such as RFID, Bluetooth and WIFI [13]. However, these technologies lack of accurate results in indoor environments (minimum deviation ~1.0-2.0m). For example, that makes it impossible to capture product group related behavior. Beyond that, transmitters or receivers need to be carried around by the persons to be tracked, which might influence their behavior. Therefore, the presented approach applies surveillance cameras, infrared cameras and algorithms from the fields of video mining to extract movement and activity data in a retail environment without bothering customers.

The extraction of position data by using image sensors is among others described by Wang et al. [11], Gavrila [14] and Perl [15]. However, none of the mentioned approaches considers the conditions at the point of sale. Besides, the use of the gathered data is discussed very little.

The analysis of movement data is among others described by Andrienko et al. [16] as well as Ashbrook and Starner [9]. In their works waypoints are assigned to well-known points of interest such as buildings or places. Based on the aggregation further movements are predicted. Gutjahr [10] extended this approach to include other sources for position data. Because the structure of a shopping environment changes continuously, it makes little sense to highlight static objects as characteristic points of interest. Rather, useful information for retailers is desired that can be extracted without further knowledge of the environment.

Shortly after its introduction the Microsoft Kinect, which was originally conceived as a controller for the Microsoft Xbox 360 game console has been used in various fields of application and especially for research purposes. The easy to use gesture recognition was applied for different purposes such as innovative user interfaces or robot steering [17][18][19]. However, using the Kinect for point of sale data collection is not considered yet.

Furthermore, the combined analysis of both movement and activity data to reveal customer behavior information for retailers has not been discussed yet.

III. Data Collection

The section comprises a brief overview of the approaches being used to extract movement and activity data from raw footage. Section III.A presents two methods using image data captured from an aerial and a lateral perspective. Subsequently, the methods are weighed against each other. Section III.B addresses the extraction of activity data based on infrared sensors implemented by the Microsoft Kinect Controller.

A. Movement Data

The overall concept presented in this work is inspired by Fillbrandt [20]. In his doctoral thesis he introduces an approach for a modular single or multi camera system tracking human movements in a well-known environment. Therefore, persons are detected on images by a set of computer vision algorithms. Afterwards, location estimation is executed. Finally, the single location data of a person are connected resulting in a trajectory.

In this work two approaches are presented to capture movement data. The lateral approach tracks customers by using cameras with a lateral point of view. The aerial approach applies cameras mounted on the ceiling.

1) Lateral Tracking

The lateral approach for customer tracking uses cameras mounted at the upper end of a corridor, which enables the observation of an entire corridor area. Footage is recorded by network cameras that enable real-time applications as well as subsequent analyses.

For person detection the histogram of oriented gradients algorithm proposed by Dalal and Triggs [21] is applied. The approach is suitable for the detection of people on images. First, the images are converted into gray scale. After that, they are transformed into gradient maps. Then, small pixel areas are analyzed regarding their one-dimensional gradient direction. The gradient maps of a variety of images containing and lacking persons are used to train a support vector machine to extract distinctive features [22] (see Figure 1a).

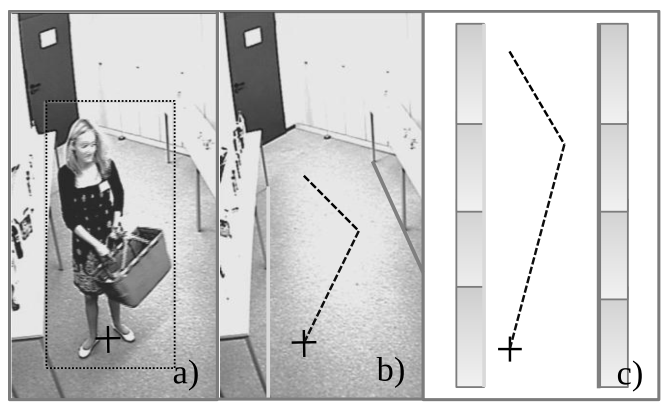


Figure 1. Lateral person tracking

Due to the wide angle distortion effect of the camera's lens especially the locations of persons being further away from the camera are perspectively distorted. That means they cannot be used for true to scale calculations yet. In consequence a perspective transformation by calculating a 3x3 warp matrix based on four source and four destination points is executed. The points have to mark equal positions on the image and on a true to scale map to calculate the factor of distortion.

The approach is evaluated using a test environment comprising one corridor delimited by two shelves. The corridor has a width of 2.0m and a length of 4.0m, which compares with a corridor of a typical small store. Raw data (e.g., 20,000 frames) are recorded for 30 different people viewing and buying products from the shelves while traversing the corridor. A maximum of three people is staying in the corridor at the same time. The results reveal issues that are caused by occultation and minor contrast between the people and the background. As a result the approach has a sensitivity of ~61%. The standard deviation between actual and measured position is 0.17m.

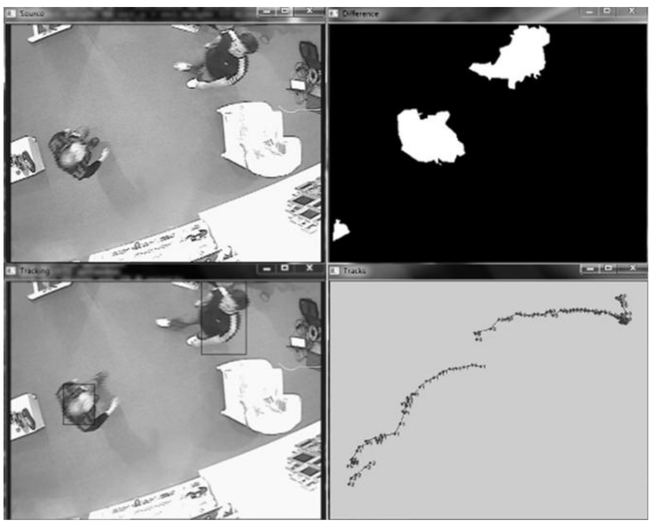


Figure 2. Aerial person tracking

2) Aerial Tracking

By using aerial mounted cameras the size of the observed area depends on the camera's focal distance and altitude. For the detection and tracking of persons within a dedicated area the following set of algorithms is applied. First, background differencing among others described by [23] and Yoshida [24] is used for object detection in single frames being captured by a camera. For this, a reference image is needed, which shows the captured areas without any objects. Comparing the reference image with the actual considered frame excludes all similarities between the two pictures. Differences are highlighted (see Figure 2. , upper right area). After that, image noise is reduced. Eliminating objects that are smaller than the average human shape reveals all objects that could possibly be considered as persons. This step is mostly accomplished by using a template or contour matching algorithm as described by Hu [25] or Zhang and Burkhardt [26]. However, this is not feasible for the presented approach. Contour or template matching algorithms are not able to detect human shapes with high reliability as a result of the varying distance and view of the camera. Besides that, people carrying bags or driving shopping carts as well as disabled people using wheel chairs would not be recognized correctly by the algorithm. Therefore, the detected shapes are filtered by a minimum surface threshold. This leads to significantly better results, i.e., persons can be recognized correctly in most cases.

Subsequently, the continuously adaptive mean shift (camshift) algorithm presented by Bradski [27] is applied for tracking detected persons. The algorithm is based on the mean-shift algorithm originally introduced by Fukunaga and Hostetler [28] and was originally invented for face tracking. Thenceforth, it has been applied for a great variety of tracking purposes.

The mean-shift algorithm is used to track motions of objects by iteratively computing the center of mass of the HUV (hue, saturation, value) vectors within a defined window [29]. For every frame of a video stream the centers of mass are calculated and then defined as new centers of the corresponding windows (see Figure 3). By connecting subsequently occurring centers of windows a trajectory of the movement is obtained. Defining windows as smallest rectangle areas covering shapes of persons extracted by the background differencing approach enables to apply this concept for person tracking purposes.

While the mean-shift algorithm considers windows of static size, the camshift implementation adapts the window size dynamically. This is of great importance for the presented application because persons moving away from or to the center of the observed area occur in different sizes. Using the mean-shift algorithm would lead to an increasing amount of vectors from areas around the considered person. If the amount of these vectors becomes too high, the scope on the person will be lost and errors occur.

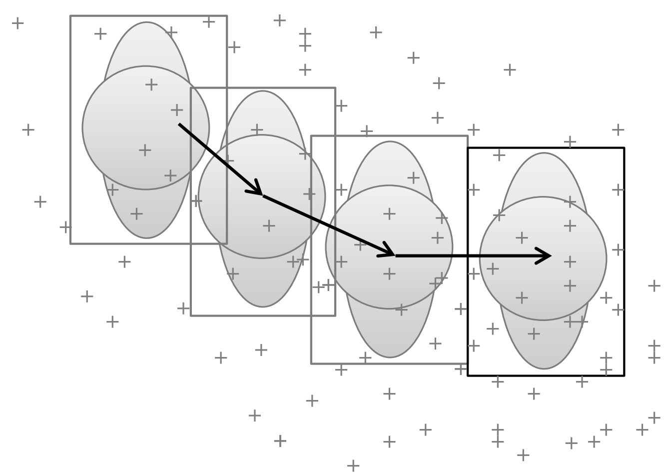


Figure 3. Mean-shift: window shifts

To achieve better results, especially for crowded places the good features to track algorithm by Shi and Tomasi [30] and the optical flow algorithm by Lucas and Kanade [31] are applied as a backup strategy. The good features to track algorithm uses corner detection to find pixels, which differ from those in their surrounding area. Subsequently, the optical flow algorithm tries to find these pixels in the following frame within the surrounding area of their original location on the image.

The combination with a color constancy algorithm (e.g., Barrera et al. [32]) or spatio-temporal rules enables to track persons across several cameras and therefore several corridors. This implies that persons leaving one camera area to another one have to be handed over while crossing an overlapping area (see Figure 4).

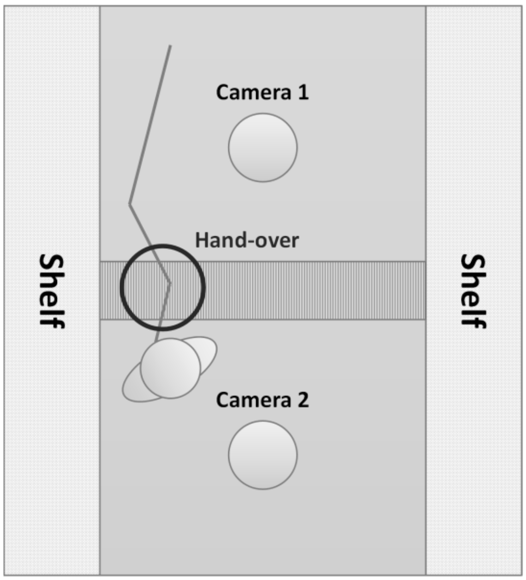


Figure 4. Camera hand-over

An evaluation study was performed for the same test environment that was used for the lateral approach. The camera is placed 3.0m aboveground. The sample footage consists of 16,000 frames showing 30 different people with a maximum of three people walking through the observed area at the same time. Lossless tracking is obtained for ~82% of the observed walks. The average deviation between the real and the automatically determined position is 0.11m.

3) Discussion

The evaluation results reveal that the aerial approach is clearly more accurate and robust. While the lateral approach struggles to overcome contrast and occultation issues the aerial approach doesn't show such problems. Apart from that, a careful evaluation of parameters for the aerial approach is mandatory to gain the described results. This is not necessary for the lateral approach since they are mainly chosen by the algorithms themselves.

Although the delineated tests included a maximum of three people at the same time both approaches are able to handle more people at the same time. Later tests in a grocery showed similar results for transition areas with up to eight persons.

When it comes to expenditures the lateral approach might be preferred by retail managers since the approach is able to use existing surveillance cameras without large-scale alternations. Therefore, expenses for new cameras are reduced to a minimum. Beyond that, the lateral approach regularly requires more cameras that are solely mounted for tracking purposes.

Nevertheless, since the aerial approach provides a better overall reliability its results will be used for the further analyses. Besides, since most other methods like WIFI and RFID lack an inch-perfect accuracy as well as a high reliability especially the aerial approach is considered as a reasonable alternative [13].

B. Activity Data

For the extraction of customer activities the Microsoft Kinect Controller is applied. The controller unifies low costs and high reliability and is therefore widely spread. It allows the tracking of parts of the body and limbs by using an infrared emitter for projecting a distributed grid of 10,000 to 20,000 individual infrared light spots into the physical space. Based on the distortions between the field captured by the infrared camera and a field in empty space the underlying system is able to extract objects. Since the camera features a resolution of 640x480 pixels, the result is an interpolated three-dimensional depth map. By using the Microsoft Software Development Kit positions of limbs of individual persons can be identified.

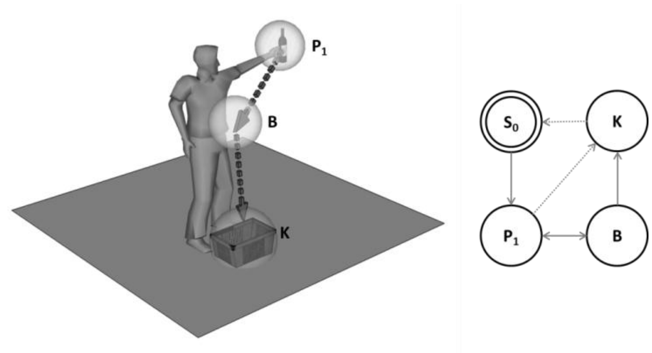


Figure 5. Activity recording

Prior to detecting activities the respective activities have to be determined. To cover the most common interactions of customers with products at the point of sale, this approach is limited to the activities “beholding a product” and “putting a product in the shopping cart”. This reduces the amount of limbs, which have to be considered, to a customer's arms and the position of his body related to the arms. Since detecting gestures based on temporal and spatial movements is disproportionately complex, following Hong et al. [33] the detection task is reduced to limb movements between predefined states, e.g., three-dimensional areas. The states are represented as a finite state machine.

Figure 5 shows an example of a typical motion sequence. The customer moves an arm from a neutral position to a product. The gesture is recorded as "customer reaches for product 1" and the state change from S_0 to P_1. Then, he moves his hand in a pre-defined area in front of his body. The system recognizes a transition from state P_1 to B and, therefore, the gesture "customer moves hand to body". The sequence of both gestures is assigned to the activity "customer beholds product P1".

IV. Customer Stream Analysis

Two groups of stakeholders are differentiated that require information about customers at the point of sale. Services derived from the customer stream analysis are dedicated to the retail management. They provide a more abstract overview of the situation at the point of sale, like information about the quality of a store's structure or the range of products. However, single customer analyses are conducted to provide customer related information to the sales stuff. They require detailed information about individual customers or customer groups, e.g., to coordinate an optimized customer approach.

To analyze customers' stream behavior the DBSCAN algorithm is used to extract regions of interest, i.e., areas that are most interesting to customers. The method was chosen because of its comparably small consumption of resources and its ability to accurately distinguish between high and low density areas, e.g., in contrast to the popular regions algorithm by Giannotti et al. [34]. Movements between areas are modeled as a Markov chain showing transition probabilities and therefore the most likely paths between the regions of interest.

A. DBSCAN

The 'density-based spatial clustering of applications with noise' algorithm originally proposed by Ester et al. [35] was developed to distinguish between clusters and noise in spatial databases.

Figure 6. DBSCAN cluster center search

Clusters are defined as areas with a considerable higher density than outside of the cluster. To distinguish clusters from noise the following steps have to be accomplished. First, an arbitrary point p is selected. Then, all points that can be reached from p are retrieved. If p turns out to be a core point of a cluster a new cluster is formed. Limitations are made regarding the minimum points (minPts) to be reached by p as well as the maximum distance ε between p and the considered neighboring points (see Figure 6). If one of the constraints is not met no new cluster is formed and another randomly chosen point is considered.

The overall datasets of all trajectories extracted by the movement tracking approach are analyzed by the DBSCAN algorithm using a minimum threshold (minPts) of five points and a maximum real world distance (ε) of 0.02 m. The analysis reveals an amount of 551 clusters.

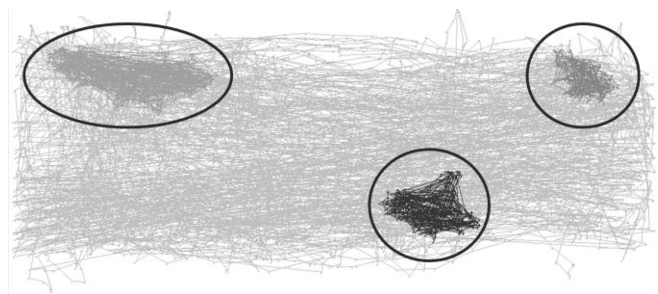


Figure 7. Clusters revealed by DBSCAN algorithm

For further processing all clusters including points from less than 60% of the trajectories are eliminated since they cover positions of too less customers. This step reveals three clusters comprising points of between 60.0% and 90.5% of the trajectories within the test environment (see Figure 7). The areas covered by the clusters are considered as hotspots that are significantly higher visited than other areas of the test retail environment.

B. Markov Chains

A Markov chain comprises states of a system as well as transition probabilities between them [36]. A transition probability is defined as the probability of a system's change from one state to another one. For a first order Markov chain it is only based on the current state. In the presented approach probabilities describe the chances of movements between two clusters. Recursive transitions are neglected because for the presented approach only the succession of movements between different states, i.e., clusters is relevant. That means a transition between two hotspot clusters exists when two temporally succeeding points of a customer trajectory belong to two different clusters. The points do not have to be temporally adjacent points in the database but all of the intermediate points must not be part of another hotspot.

Regarding the movements between clusters, the datasets resulting from the computer vision algorithms described in Section III.A.2 are taken into account. Points that are not part of one of the three considered clusters are ignored.

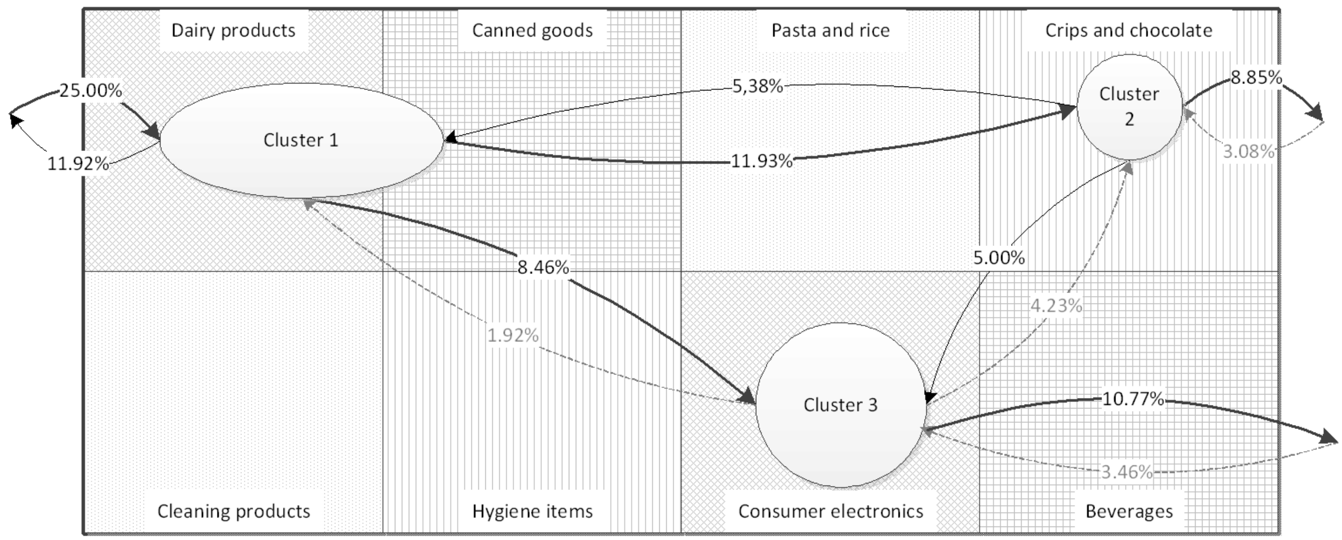


Figure 8. Prototypical retail environment – Scenario 1

C. Results

Considering the clusters and the movements between them allows a closer look on how customers act within a retail store. As an use case two shopping scenarios within the prototypical retail environment are analyzed. The test environment comprises eight product categories (see Figure 8).

The first scenario describes a regular product setup without any advertisements and signs and therefore is used to observe the regular behavior of customers. The second one is based on the findings of the first scenario and includes advertisements for selected products (see Figure 9). Both of the scenarios are compared eventually.

For the first scenario three clusters exceeding the 60% threshold are found. One of them covers the area with shelves containing dairy products. The second, smaller one is located near shelves with crisps ad chocolate. The third one covers the area in front of the shelves with consumer electronics. Figure 8 shows these three clusters as well as the transitions between them. The percentage indicates the proportion of transitions to the total number of transitions being extracted from the customer movements. Transition paths below the limit of a 5% share are greyed out to achieve a clear visualization.

For the given scenario the majority of customers enter the corridor from the left side heading to the first hotspot (dairy products). Afterwards they are more likely moving on to the second one (crisps and chocolate). Then, either they go back to the area of cluster 1 (dairy products) or go on to cluster 3 (consumer electronics).

After that, the customers are most likely leaving the observed area. Besides showing hotspots within the retail environment the graph of Figure 8 also reveals typical paths customers use to move through the store. Looking at visited products it is apparent that products located on the lower left (cleaning products and hygiene items) are less considered. Therefore, advertisements in frequently attended areas are used to call attention for these products.

This idea is seized for the second scenario (see Figure 9). The prototypical retail store is extended by two promotional signs for cleaning and hygiene products. This leads to notable changes of the customers' behavior. While the first scenario leads to three hotspots the second one includes four hotspots. An additional hotspot covers the area between cleaning and hygiene products.

Considering the transitions customers still most likely enter the observed retail environment from the left side attending the area near dairy products first. Afterwards, they are moving on either to crisps and chocolate or to the area in front of the shelves containing consumer electronics.

While most of the consumers move from crisps and chocolate back to the area of dairy products there is also a notable percentage of customers walking to an area in front of cleaning and hygiene products. This could mean that the promotion campaign was successful.

The visualization enables retail managers to get an overview of the movements at the point of sale. Thus, the behavior of customers is monitored and changes are revealed. Besides adding new advertisements existing ones can be evaluated regarding their effectiveness. The same applies to the structure of the environment itself. If the structure is adapted to change the customer flow the movement behavior can be evaluated eventually. If the results don't correspond to the expectation the store might be adjusted iteratively.

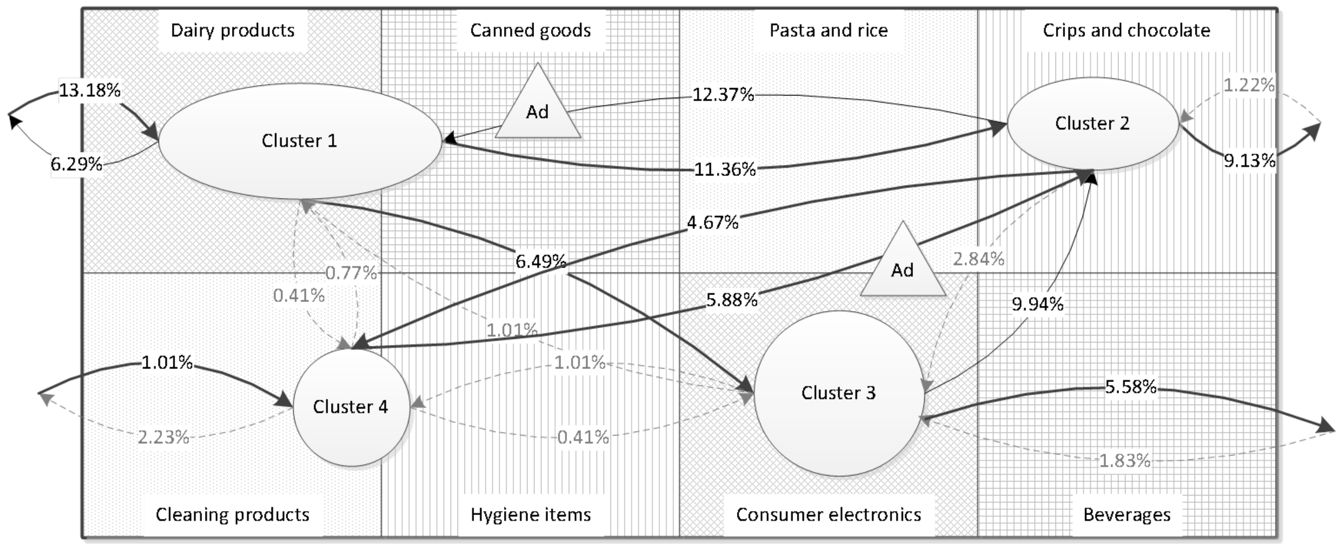


Figure 9. Prototypical retail environment – Scenario 2

V. Single Customer Analysis

Single customer analyses are used to provide information to sales people assisting them to address customers. Therefore, a cockpit is presented that provides relevant customer related information to the sale personnel in real-time. Key figures are used to deliver information about common characteristics of customers. Companionships provide an overview of which customers belong together and estimated behavior and next steps are used to interpret customers' behavior.

A. Key Figures

Key figures are discrete values that provide insight into single dimensions of individual customer behavior. Here, key figures are duration of stay and velocity as well as the number of stops, direction changes and visited sections.

a) Duration of Stay

The duration of stay describes the overall time a customer spent in a section or the entire environment. Therefore, it is an indicator for the interest of customers in certain sections. Customers that are more interested in products of a certain section will spend more time there. In contrast, less interested customers will leave the section faster. Measuring the duration of stay in real-time enables to estimate the time left in the store based on the average time spent.

b) Velocity

A customer's average velocity while walking through the retail environment is derived from the total distance between the entire recorded positions and the total time spent at the store or a section. Since stops distort the average velocity they are considered separately and are excluded from the velocity calculation. The average velocity helps to distinguish between hurrying and slowly traversing customers and is an indicator for the interest of customers in certain areas or the entire store. Besides, it helps to estimate a customer's interest in consultation.

c) Stops

Stops of customers describe a sequence of recorded positions that are close nearby each other or in the same place for a defined amount of time. Considering the average stops per section reveals information about customers' interest for the products exposed in this section. In addition, an above average amount of stops is an indicator for customers that are searching for specific products or comparing them and therefore might be used to address customers different.

d) Direction Changes

Direction changes are another feature of customer movements and indicate how well customers know the retail environment or sections of it. A high number of direction changes in one section indicate that a customer is searching for or comparing a specific product. A high number of changes within the entire environment might be evidence that the customer entered the store for the first time.

Direction changes are calculated as the smaller angle between preceding and succeeding points connected through an apex. If the angle undercuts 60° it is considered as an intentional change in a customer's movement direction.

e) Visited Sections

To calculate the number of visited sections the retail environment has to be separated in clearly delimited areas, e.g., "cereals" or "hygiene items". Then, counting the number of distinct visited sections enables to draw conclusions, which areas of a retail store are visited more or less often by customers. Besides, considering sequences of sections reveal typical sequence patterns. These patterns help to understand in which way customers traverse the retail environment and therefore, which sections are commonly visited in a row.

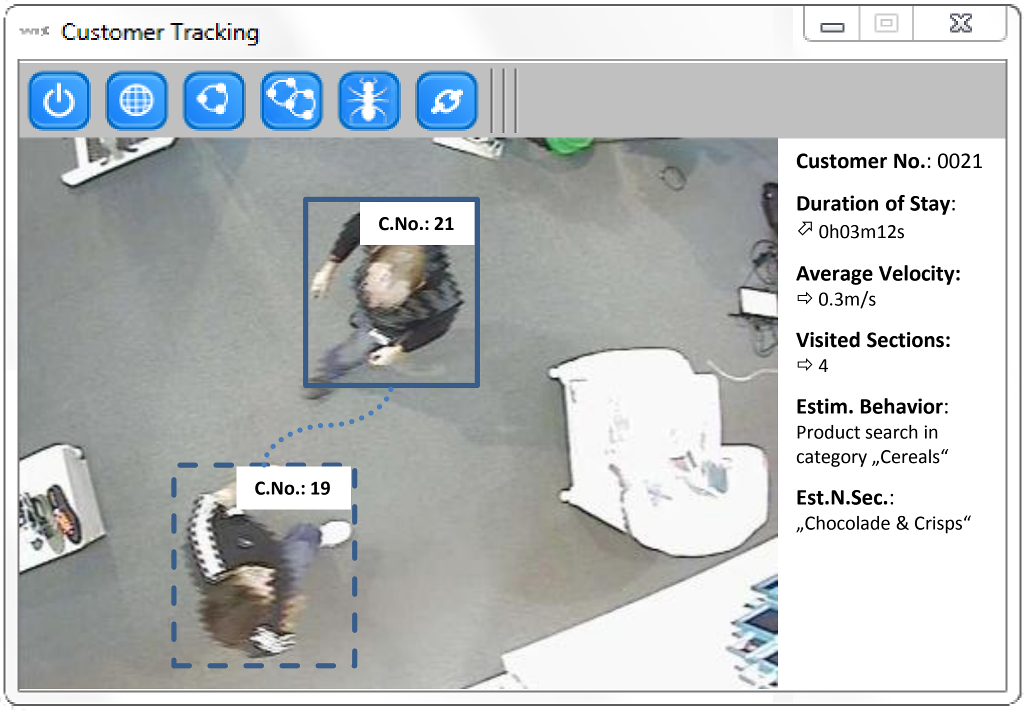


Figure 10. Cockpit from single customer analysis

f) Activities

In contrast to the other key figures activities are recorded through the approach described in Section III.B. Using activity data for product categories reveals information about how customers behave before they buy a specific product. E.g., if a customer views several different products of a product category before he/she puts it into the cart, the activities indicate that the customer is comparing product of a product category to find the most suitable product according to his requirements. In contrast, when a product gains less attention before it is bought it most probably is an item that is purchased habitually.

B. Companions

Companions are people shopping together. To determine companionships all people in the retail environment are continuously observed regarding the distance between each other. If the customers spend most of their time nearby each other the two persons are considered as related.

Figure 10 shows a prototypical cockpit visualizing the companionship of two customers by a dotted line.

C. Estimated Behavior

Estimated behavior describes the behavior that customers show based on their current movements and activities. For that, customer related information such as the number of stops, changes of direction, the average velocity and the activities, e.g., viewed products are determined. A pattern recognition system then uses this information to estimate a predefined class of behavior.

E.g., when a customer shows a significantly higher number of stops, direction changes and recently viewed a lot of products by taking them of the shelves without putting them in his cart his behavior is considered as a product search behavior.

D. Estimated Next Steps

Estimated next steps means the movement behavior that customers will most likely show based on their previous movements. For this, the movement history of a customer is compared to a set of rules being derived from previously analyzed trajectories. The system then estimates the most likely next steps. The approach is based on Markov models to represent transition probabilities between certain areas of a retail environment [36][37].

Figure 11 shows the prediction of the further path, based on a grid, which is superimposed over the shopping environment. The pathway starts at field 5-0 following the grey colored fields. The last performed step is at field 1-F. Based on the underlying model, the system estimates 2-F and 3-F as the most likely next steps.

E. Results

For the described test data the average duration of stay is 3.12 minutes. In that time customers move with an average velocity of 0.48m/s. Compared to the velocity including stopping times the adjusted velocity provides a greater variance (0.024m/s) and therefore enables a better differentiation between different customers. In average, customers perform 7.6 stops and 4.1 significant direction changes. It is striking that the stops and direction changes are mainly located at the beginning of the shopping getting less until the end of it. It looks like customers are losing interest the longer shopping takes. Besides, stops and direction changes are mainly found nearby more expensive products such as consumer electronics.

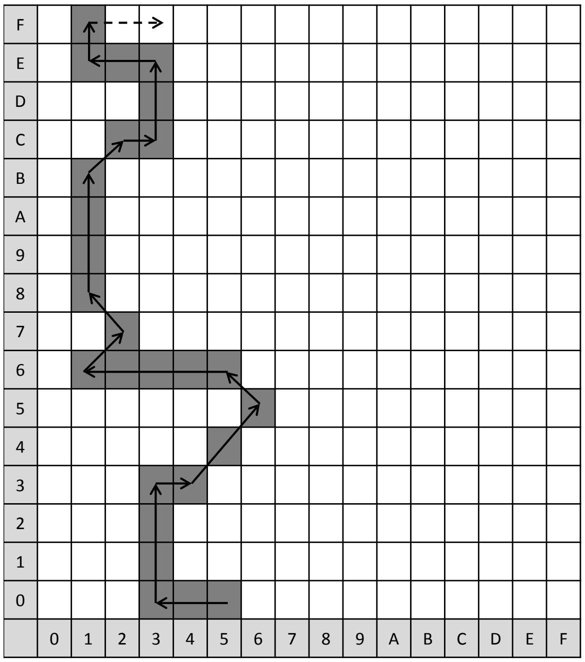


Figure 11. Movement estimation

Since the testing environment comprises only a small area, the average number of visited sections (7.3 of 8 sections) is comparatively high and therefore has less validity. Activities are mainly performed in sections like "crisps and chocolate" as well as "consumer electronics" and therefore corroborate the hypothesis that everyday life products are bought faster.

Using a k-Means clustering algorithm for detecting stereotypic behavior of customers works well for estimating customer's behavior. However, results show solid accuracy for less than 4 behavior classes.

Initial tests of the companionship extraction show that the use of movement data provides reliable results for couples or friends. In contrast, families are harder to recognize since children are moving around more inconsistent. Therefore, distances measured for people staying in the same section are rated higher. As a result the rate of correct assignments is increased to a maximum of 86.1%.

For the estimated next steps Markov chains are tested from first to third order for the pre-defined sections shown in Figure 8 and 9. Besides, an evaluation of Markov chains with uniform grids is performed. The testing results reveal that second order Markov chains using data from grid fields are the best tradeoff between performance and accuracy.

VI. Conclusion and Future Work

The paper introduces an approach for recording and analyzing customer behavior data based on image and infrared sensors.

Surveillance cameras produce large amounts of video data. Intelligent methods for processing these data are crucial in order to gain customer insight especially over a longer period of time. Therefore, methods are introduced for capturing and extracting movements using network cameras and algorithms from the field of computer vision.

The same applies to customer activity data that are rarely considered yet but reveal valuable information. They are extracted by infrared sensors being implemented in the Microsoft Kinect Controller. Although it is difficult to cover an entire retail environment by these controllers the infrared sensor technology turned out to be a reliable way to record activity data.

Both the movement and the activity data are used for customer behavior analysis. Analyses are conducted using sets of pattern recognition algorithms. As a result information for retail managers and sale staff is gained and visualized through cockpits.

Considering longer periods of time might reveal different customer behavior not only for different setups but also for different times of a day, days of the week or seasons. In addition to that, comparative studies of different stores of the same chain are possible. Information gained from these analyses is the basis for planning dynamic product placements or seasonal offers. Besides, the knowledge about customers is considered as an additional source for management information systems. It helps to identify rarely visited areas or products and therefore enables retail store managers to analyze and optimize their shopping environment. New settings can be evaluated by considering the ex-ante and the ex-post change status.

Single customer related information put sales staff in the position of knowing their customers before they address them. Therefore, they are able to react more specific to customer requests and can provide them a better consultation.

Extending the described information by further information such as bought products and socio-demographic factors like gender or age might increase the validity of the information and therefore reveal new possibilities for management and sales force related services. For instance, an extended database allows setting up customer typologies considering different periods of time or different shops.

References

[1] J. Kröckel and F. Bodendorf, "Intelligent Processing of Video Streams for Visual Customer Behavior Analysis," in *ICONS 2012, The Seventh International Conference on Systems*, 2012, pp. 163–168.

[2] H.-F. Li, S.-Y. Lee, and M.-K. Shan, "DSM-TKP: Mining Top-K Path Traversal Patterns over Web Click-Streams," *The 2005 IEEEWICACM International Conference on Web Intelligence WI05*, pp. 326–329, 2005.

[3] I. Nagy and C. Gaspar-Papanek, "User Behaviour Analysis Based on Time Spent on Web Pages," in *Web Mining Applications in Ecommerce and Eservices*, Springer Berlin / Heidelberg, 2009, pp. 117–136.

[4] X. Zhang, W. Gong, and Y. Kawamura, "Customer Behavior Pattern Discovering," in *Advanced Web Technologies and Applications*, J. Yu, X. Lin, H. Lu, and Y. Zhang, Eds. Berlin/Heidelberg: Springer, 2004, pp. 844–853.

[5] P. Underhill, *Why we buy: The Science of Shopping*. Textere, 2000.

[6] B. Rao and L. Minakakis, "Evolution of mobile location-based services," *Communications of the ACM*, vol. 46, no. 12, pp. 61–65, 2003.

[7] J. Hightower, B. Brumitt, and G. Borriello, "The location stack: a layered model for location in ubiquitous computing," *Proceedings Fourth IEEE Workshop on Mobile Computing Systems and Applications*, pp. 22–28, 2002.

[8] G. D. Abowd and E. D. Mynatt, "Charting past, present, and future research in ubiquitous computing," *ACM Transactions on Computer-Human Interaction*, vol. 7, no. 1, pp. 29–58, 2000.

[9] D. Ashbrook and T. Starner, "Using GPS to learn significant locations and predict movement across multiple users," *Personal and Ubiquitous Computing*, vol. 7, no. 5, pp. 275–286, 2003.

[10] A. Gutjahr, "Bewegungsprofile und -vorhersage," 2008.

[11] L. Wang, W. Hu, and T. Tan, "Recent developments in human motion analysis," *Pattern Recognition*, vol. 36, no. 3, pp. 585–601, 2003.

[12] A. J. Lipton, H. Fujiyoshi, and R. S. Patil, "Moving target classication and tracking from real-time video," in *Applications of Computer Vision, 1998. WACV'98. Proceedings., Fourth IEEE Workshop on*, 1998, pp. 8–14.

[13] H. Koyuncu and S. H. Yang, "A survey of indoor positioning and object locating systems," *Journal of Computer Science and Network*, vol. 10, no. 5, pp. 121–128, 2010.

[14] D. M. Gavrila, "The Visual Analysis of Human Movement: A Survey," *Computer Vision and Image Understanding*, vol. 73, no. 1, pp. 82–98, 1999.

[15] J. Perl, "A neural network approach to movement pattern analysis.," *Human Movement Science*, vol. 23, no. 5, pp. 605–620, 2004.

[16] G. Andrienko, N. Andrienko, S. Rinzivillo, M. Nanni, and D. Pedreschi, "*A Visual Analytics Toolkit for Cluster-Based Classification of Mobility Data*," pp. 432–435, 2009.

[17] A. Bleiweiss, D. Eshar, G. Kutliroff, A. Lerner, Y. Oshrat, and Y. Yanai, "Enhanced interactive gaming by blending full-body tracking and gesture animation," *ACM SIGGRAPH ASIA 2010 Sketches on - SA '10*, pp. 1–2, 2010.

[18] K. Lai, J. Konrad, and P. Ishwar, "A gesture-driven computer interface using Kinect," *2012 IEEE Southwest Symposium on Image Analysis and Interpretation*, pp. 185–188, 2012.

[19] Z. Ren and J. Meng, "Robust hand gesture recognition with kinect sensor," *MM '11 Proceedings of the 19th ACM international conference on Multimedia*, pp. 759–760, 2011.

[20] H. Fillbrandt, "Videobasiertes Multi-Personentracking in komplexen Innenräumen," Rheinisch-Westfälische Technische Hochschule Aachen, 2008.

[21] N. Dalal and W. Triggs, "Histograms of Oriented Gradients for Human Detection," *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition CVPR05*, vol. 1, no. 3, pp. 886–893, 2004.

[22] C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learning*, vol. 20, no. 3, pp. 273–297, 1995.

[23] M. Piccardi, "Background subtraction techniques: a review," *2004 IEEE International Conference on Systems Man and Cybernetics IEEE Cat No04CH37583*, vol. 4, no. C, pp. 3099–3104, 2004.

[24] T. Yoshida, "Background differencing technique for image segmentation based on the status of reference pixels," *2004 International Conference on Image Processing, 2004. ICIP '04.*, vol. 1, no. 1, pp. 3487–3490, 2004.

[25] M.-K. Hu, "Visual pattern recognition by moment invariants," *IEEE Trans Information Theory*, vol. 8, no. 2, pp. 179–187, 1962.

[26] G. Zhao, N. Zhang, and Z. Liu, "A case investigation on the scaling behaviors in web browsing," *Web Intelligence and Intelligent*, no. 70971089, pp. 160–163, 2010.

[27] G. R. Bradski, "Computer Vision Face Tracking For Use in a Perceptual User Interface," *Interface*, vol. 2, no. 2, pp. 12–21, 1998.

[28] K. Fukunaga and L. Hostetler, "The estimation of the gradient of a density function, with applications in pattern recognition," *IEEE Transactions on Information Theory*, vol. 21, no. 1, pp. 32–40, 1975.

[29] D. Comaniciu and P. Meer, "Mean shift analysis and applications," *Proceedings of the Seventh IEEE International Conference on Computer Vision*, vol. 2, no. 2, pp. 1197–1203 vol.2, 1999.

[30] J. Shi and C. Tomasi, "Good features to track," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 1994, vol. 94, no. June, pp. 593–600.

[31] B. D. Lucas and T. Kanade, "An iterative image registration technique with an application to stereo vision," *International Joint Conference on Artificial Intelligence*, vol. 3, pp. 674–679, 1981.

[32] P. Barrera, J. M. Canas, and V. Matellán, "Visual object tracking in 3D with color based particle filter," *Int Journal of Information Technology*, vol. 2, no. 1, pp. 61–65, 2005.

[33] P. Hong, M. Turk, and T. Huang, "Gesture modeling and recognition using finite state machines," in *Proceedings of the Fourth IEEE International Conference on Automatic Face and Gesture Recognition 2000*, 2000, pp. 410–415.

[34] F. Giannotti, M. Nanni, F. Pinelli, and D. Pedreschi, "Trajectory pattern mining," in *Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining*, 2007, pp. 330–339.

[35] M. Ester, H. P. Kriegel, J. Sander, and X. Xu, "A density-based algorithm for discovering clusters in large spatial databases with noise," in *Proceedings of the 2nd International Conference on Knowledge Discovery and Data mining*, 1996, vol. 1996, pp. 226–231.

[36] A. A. Markov, "Extension of the limit theorems of probability theory to a sum of variables connected in a chain (Reprint in Appendix B)," John Wiley and Sons, 1971.

[37] I. Nižetic, F. Krešimir, and K. Damir, "A prototype for the short-term prediction of moving object's movement using Markov chains," *Proceedings of the ITI 2009 31st International Conference on Information Technology Interfaces*, pp. 559–564, Jun. 2009.

Process Observation as Support for Evolutionary Process Engineering

Stefan Schönig[1], Michael Seitz[2], Claudia Piesche[1], Michael Zeising[1], and Stefan Jablonski[1]
[1] University of Bayreuth
Chair of Applied Computer Science IV
Bayreuth, Germany
{stefan.schoenig, claudia.piesche, michael.zeising, stefan.jablonski}@uni-bayreuth.de

[2] PRODATO Integration Technology GmbH
Erlangen, Germany
michael.seitz@prodato.de

***Abstract* – The Process Observation project is used to generate business process execution logs and guides process participants through process execution. In this contribution, we introduce process evolution as an economic field of application for process observation. There are different needs for process evolution, e.g., to establish more consistent process results, continuously measure and improve process performance or meet accreditation requirements. We will show how process discovery, process guidance and process evidence as the main basic functions of process observation can be applied as support for reasonable process evolution. In this way, process observation can be used to reach a desired evolution stage or rather facilitate the transition between two maturity levels. Furthermore, process observation serves as an implementation for certain evolution stages itself and can additionally be consulted to prove the conformance to quality requirements of maturity levels.**

Keywords – Business Process Observation; Process Monitoring; Process Mining; Process Model Enactment; Process Evolution

I. INTRODUCTION

This article is an extension of a previously published paper [1]. Business process management (BPM) is considered an essential strategy to create and maintain competitive advantage by modeling, controlling and monitoring production and development as well as administrative processes [2] [3]. Many organizations adopt a process based approach to manage various operations. BPM starts with a modeling phase, which is very time and cost intensive. It requires deep knowledge of the underlying application and long discussions with the domain experts involved in the processes in order to cover the different peculiarities of the process [4]. Since process modeling is an expensive and cumbersome task, we identify approaches that promise to reduce the modeling effort.

Furthermore, traditional process management restricts the support of process execution by strictly separating modeling and execution phase. This separation exemplarily flows into the following propositions [35]:

- Schematization of processes: everything, that has not been modeled, is excluded from management and will not be supported by systems.
- Inadequate consideration of exceptions: a priori models are often idealizations. Dynamically occurring exceptions will consequently not be supported by systems.
- By strictly separating modeling and execution phase process models are often incomplete because theory and practice are deviating.
- Inadequate feedback: generally, there is no consideration of feedback from process execution to process modeling. However, knowledge of previous executed process cases should be considered during process modeling.

Latest research achievements in the context of pattern discovery and data mining offer possibilities that promise to overcome the strict separation of modeling phase and execution phase by applying real-time analysis methods of executed processes.

One of these research achievements is process mining. Process mining utilizes information/knowledge about processes whilst execution. The idea is to extract knowledge from event logs recorded by information systems. Thus, process mining aims at the (semi-)automatic reconstruction of process models using information provided by event logs [5]. The computer-aided creation of process models offers huge potential of saving time. By deriving process models from event logs, the appropriateness of process models can be guaranteed to a certain extent, since they are constructed according to the way the processes have actually been executed. During the last decade, many techniques and algorithms for process mining have been developed and evaluated in different domains [6]. The basis for a successful generation of a process model through process mining is an existing and complete process execution log. This is also the big challenge for a successful application of process mining. First of all, not all processes are executed by information systems, i.e., they are executed "external" to computers. In such cases, there is no event log that represents a process available and process mining cannot be applied. In the case that information systems are already used to execute processes there must be guarantees that these event logs record process execution in such a way that processes can be reconstructed. Besides, these event logs must be analyzable in such a way that appropriate process models can be derived. It is obvious: the quality and availability of event logs determine the applicability of process mining techniques.

Our research starts with the assumption that a complete and freely analyzable event log is usually not available. We regard this scenario as the most common one. Thus, one of the major aims of our research is to harvest process execution knowledge. This enables the assembly of a process execution log. This log is built up independently from the existence of information systems that are (at least partly) executing the processes. We developed a special software, a Process Observation (PO) tool, that can be envisioned as a tool that permanently runs on the computers of process participants that asks the process participants “What are you doing right now?”. The participants then have to describe what they are doing. Here, the user does not need any process modeling skills. This is also one very important prerequisite since we assume that just few process participants do show process modeling skills. The recorded data is used by PO to mine for process models. Of course, this process information can be enriched and complemented by event logs from information systems that are involved in process execution. Gathering process execution information comes with the cost that process participants have to record what they are doing. Of course, this means additional work for the process participants. Therefore, PO must offer a stimulus that motivates process participants to work with PO. This stimulus is put into effect by a recommendation service. PO continuously analyzes available process log data to guide the process users. This means, it suggests process steps, documents or tools that the user most probably performed or used. We have experienced that this feature is especially important for users that are still not too familiar with the application; they are thankful that the PO tool recommends possible process entities. This dynamic recommendation service becomes more and more reliable the more process instances have been executed under the guidance of PO. The execution of first instances of a process will therefore not considerably be supported. The effect becomes apparent when a couple of process instances have been executed.

In the following, we want to classify PO. As dimensions for this classification we take the two issues: attaining a process model and executing a process model. We already discussed the two principal approaches to attain a process model. They will be assessed with respect to the amount of effort a process participant has to or is able to invest. The first approach to attain a process model is to create it within a process modeling project. This task is very costly; it usually cannot be performed by process participants but requires process modeling experts. They identify the process through interviews with the domain experts and need to get a good overview over all possible process peculiarities to guarantee the completeness of the process model. Process models can also be attained by the application of process mining techniques. This approach is cheap since only little work from process modelers is required. However, it depends on the existence of event logs representing the execution of processes. These two approaches depict two extreme landmarks: on the one hand processes can be performed within information systems. On the other hand, information systems could not be involved at all. PO bridges the contrary approaches of process execution und thus combines their benefits. It is connectable to process execution systems and can leverage them; also it provides execution support for "external" process execution.

In addition to our previous work of [1], we want to introduce process evolution as an economic field of application for PO in this article. There are different needs for process evolution, e.g., to establish more consistent process results, continuously measure and improve process performance or meet accreditation requirements. We will show how automatic process model generation (process discovery), dynamic recommendations (process guidance) and access to execution logs (process evidence) as the main basic functions of PO can be applied as support for reasonable process evolution. In this way, PO can be used to reach a desired evolution stage or rather facilitate the transition between two maturity levels. Furthermore, PO serves as an implementation for certain evolution stages itself and can additionally be consulted to prove the conformance to quality requirements of maturity levels.

This article shows the following structure: Section II introduces process evolution as an appropriate field of application. Section III introduces business process enactment and Section IV gives an overview over current existing process model enactment approaches. In Section V, we will explain the conceptual details and the general approach of process observation. Furthermore, concrete implementation techniques will be shown in Section VI. Based upon the introduced conceptual details of the previous sections, we will describe different applications for process evolution support and some use cases in Section VII. Section VIII describes the influence of the PO on the current process lifecycle. In Section IX, we will give an overview over related works. In Section X we will finally conclude and give an outlook on further research issues and applications.

II. Evolutionary Process Engineering

Before actually describing our approach in detail we want to introduce process evolution as an economic field of application for process observation.

A. *Introducing Process Evolution*

We perceive process evolution as a certain and natural development each process encounters over time in order to increase its quality and performance. There are different needs for process evolution, e.g., to expand business, establish more consistent and more stable process results, continuously measure and improve process performance or meet accreditation requirements. As an example, the workflow continuum used in [38] can be cited. It shows how a process develops from a unique, exceptional use case (e.g. new type of customer request) to the point of a highly recurrent, well-structured standard workflow. From a quality management perspective, adequate measurements for process evolution are maturity models, such as the Capability Maturity Model Integration (CMMI) [43] or the Business Process Maturity Model (BPMM) [42]. As stated in [37], the evolution of quality is accompanied by (new) requirements to the process model and its enactment for process support. While a low maturity level is satisfied by rudimental modeling of the expected results aiming at a human controlled process execution, a higher maturity level probably requires a formal specification of the organizational and operational structure in order to implement a (partly) system controlled process execution. Anyway, [37] insists on fitting both process model and process support to the quality and performance requirements in order to drive the process in a reasonable way. For example, when it does not make any difference for the success of the process in which way a required document is created it is not worth the effort to design a process model prescribing the user the order of each single step, because this kind of execution support would not be accepted by the users.

In the following section, the maturity levels including an ideal set of climactic evolution stages are described and the requirements for an adequate support of process evolution through the PO are derived.

B. *Requirements for Process Evolution Support*

The maturity model for process evolution in [37] contains five maturity levels (ML1 - ML5) and is based on the perspective oriented process model (POPM). According to POPM, processes can be described by at least five independent perspectives [14]: the functional perspective (work steps), the behavioral perspective (control flow), the data oriented perspective (data), the organizational perspective (people) and the operational perspective (tools). In order to get a general idea about POPM perspectives, we recommend [14] and [15]. Each maturity level in [37] describes a suitable and stable stage of process evolution through dedicated characteristic values of the dimensions process quality, process model and process support. Below, the maturity levels are summarized. Afterwards, requirements for PO to support process evolution are derived.

For ML1 it is sufficient to design a result-oriented process model that covers a list of necessary steps or expected results respectively (functional perspective). This information should be published as support for process participants. It has to be documented properly that the process has actually been executed and the results could be achieved.

According to ML2 the process model should represent a project plan and therefore additionally contain resources and responsibilities (organizational perspective) and a time schedule (behavioral perspective). In order to support the execution at runtime plan deviations should be recognized and communicated. It has to be proved with the help of documentation that the process has really been projected and monitored.

ML3 implies a standardization of the process and accordingly is in need of a reference process model that moreover must include information about the usage and invocation of mandatory tools, applications and services (operational perspective) as well as the default document format and structure (data perspective). During execution the active monitoring of the compliance has to be enabled and supported. The compliance with the standard process is required and must be proved by documentation.

As for ML4, key performance indicators have to be defined, measured and analyzed. Therefore, a formal and (technically) comparable process model with metering points is required that enables the extraction of consistent and stable process logs. The execution has to be supported by an adequate analysis and control tool. It has to be proved that the process targets are actually monitored statistically across several process instances.

ML5 demands the establishment of a systematic continuous improvement process. It necessitates a continuously formal and extensible representation of all occurring process perspective contents across the whole process lifecycle. A system is needed for execution support that allows for respective improvement measures at runtime and is able to reuse them. It has to be proved through documentation that improvements are (probably automatically) incorporated into the reference process model and also implemented successfully in future instances.

It is clearly obvious that with each maturity level a more comprehensive and more detailed process model is needed. Thus, process model engineering can be regarded as an important factor of process evolution and thus constitutes one requirement for its support. After all, the designed process model is used to provide appropriate runtime support. The enactment of process models serves as implementation of evolution stages and is intended to enable the establishment of the desired quality as efficient as possible. Therefore, enactment is considered as further requirement for process evolution support. A maturity level can first be regarded as fulfilled if the demanded quality can be proved. Consequently, documentation of the way the process has actually been executed is yet another requirement for process evolution support.

III. ENACTMENT APPROACHES OF PROCESS MODELS

In the previous section, engineering, enactment, and documentation could be identified as requirements for adequate process evolution support.

Besides engineering and documentation purposes, the enactment of process models is one main target of business process modeling efforts. So it is no wonder that various research activities center on questions on how to best support enactment of process models [14], [15], [24]. Even though most publications deal with particular implementations for process execution support or workflow management systems (WfMS) at it one can recognize the broad relevance of process enactment as concept within the BPM research context. Aside from different presuppositions, i.e. imperative vs. declarative approach, limitation to the execution order vs. perspective-oriented modeling or consideration of business process as serialization of web services, the overall goal is the best possible execution support for processes within organizations. However, the main questions remaining are: How does this support look like? Which type of support is considered as best fitting for a particular situation? Which conceptual approach provides best results regarding efficiency of process execution, quality of resulting products or employee satisfaction?

A first rough answer to these questions can be found for example in [23], where additional to process modeling also process execution issues are discussed and possible solutions are introduced. Similar to database management, the execution of processes is compared with the usage of stored data. Just as the data usage differs immensely between various application scenarios, i.e., "even two different implementation concepts are necessary (normal databases vs. warehouses)", the same observation is valid with process management. Therefore, the conclusion can be drawn, that different use cases for processes need different enactment approaches. To show this is the main contribution of this paper.

A. *Definition of Business Process Enactment*

From a historical perspective, process management and especially process enactment originates from office automation [25] as well as from the concept of BPEL (Business Process Execution Language) where a business process is defined as "a large-grained Web Service which executes a control flow to complete a business goal" [24]. In the same way, the classical workflow management system follows the concept of strictly automated workflow according to an explicit process model. Even [25] regards the enactment of processes only as the generation of corresponding runtime support based on a process model. However, there are example processes which cannot follow this assumption, for example creative processes. Here, the process model rather presents a guideline how to proceed; the sequence of process execution can more or less freely be chosen whereby the process model functions as a "recommended guideline".

Thus, enactment must be seen in a broader sense. It shall not only support a strict default workflow but offer recommendations instead. It happens mainly through use of a process, i.e. enactment is not the same as strict execution, rather more similar to "process usage". It aims at applying or using processes and instantiates a process model in any way, automatically by means of various IT systems (explicit) or through the specific work process of employees within a company (implicit). Therefore, various kinds of support while conducting a process instance are conceivable. Such approaches vary between mere manually controlled and fully automated enactments. Thereby, simple wallpaper models (which is nothing but the printed out process model) with check marks define one end of the spectrum; automatic process execution with WfMS demarcate the other end.

B. *Motivation and Contribution*

As stated before, there is a growing amount of workflow and process enactment approaches, especially considering the broader definition of enactment. However, until now there is a lack of a conceptual evaluation of the quality and adequacy of these approaches. At the same time, enactment is considered an important phase of the business process life cycle [25], [26]. Even though the modeling phase can be seen as the more important part of process management, the implementation of processes gains in importance and deserves closer attention.

The contribution of this section at first consists in widening the view on enactment approaches by introducing a broader definition getting beyond the current limited view of enactment as an automatically supported activity. It is shown that enactment support can be achieved by IT systems as well as by other manual approaches.

IV. Existing Enactment Approaches

This section enfolds the spectrum of enactment approaches from manually controlled process execution to fully IT-supported ones. Thereby, the order of described approaches is somewhat equivalent to the increasing automation of process execution, i.e., external tracking and serialization can manually be carried out but the other three approaches need sufficient support by IT systems. However, the enumeration of following approaches depicts an exemplary selection of enactment approaches and is not exhaustive. The objective of this section is to put in context the concept of Process Observation as Dynamic Guidance approach.

A. *Model-as-is*

Inspired by documentation purpose of process models, external tracking is aiming at supervising the process on an abstract level. Thereby, the model is used as is, i.e., the model is printed out and used as wallpaper or it appears as a map electronically. Responsible agents may mark finished tasks by labels and with it retain control over process execution.

B. *Checklist*

The serialization of process steps happens by defining a checklist that "comprises the main process steps including documents that must be produced and agents that are responsible to perform the corresponding process" [23]. Every process has to be signed after finishing the execution so that, finally, the process is concluded if all steps are signed. A checklist serializes the process steps of a process model, shows what input sources could be used, what results are expected, and who is responsible for the step. The method can be accomplished manually as well as electronically.

Main Process

Process A	Documents: IN: Doc 1 OUT: Doc 2	Comment:	Agent: X Signature:
Process B	Documents: IN: Doc 2 IN: Doc 3	Comment:	Agent: Y Signature:
Process C	Documents: IN: Doc 2 OUT: Doc 4	Comment:	Agent: Z Signature:

Figure 1. Sample Checklist

C. Dynamic Guidance

Aim of this enactment approach is to guide users through a flexible process instance at runtime based on the process model but not restricted to it, i.e. without patronizing the user. Thus, the responsible agent has the possibility to change the order of tasks due to his experience, extraordinary conditions or optimization purposes.

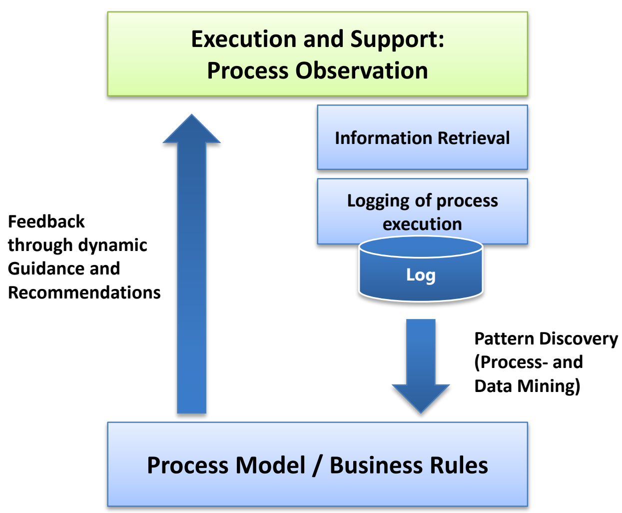


Figure 2. Concept of the Process Observer

The idea for dynamic guidance relies on data- and process mining approaches. Event logs of former process executions are used to find recommendations that are displayed while current process execution within a process instance. The conceptual basis of dynamic guidance is stated in [27] and [40]. It describes an example for the implementation of dynamic guidance using process mining techniques in order to be able to examine various perspectives of process execution logs. The algorithms take the current process scenario into account and offer a best practice recommendation to the user. In contrast to static algorithms, process mining does not need a predefined aim in order to calculate a recommendation. Nevertheless, this approach requires access to log files. However, there are processes that are not electronically executed and controlled. In such cases, the generation of manual logs has to be initiated by process participants by recording their current activities. This information will be enriched and complemented by optional logs from information systems. An example for a dynamic guidance implementation is the Process Observation Project stated in [1]. The conceptual details of this project are presented in Section V and further implementation details are presented in Section VI of the work at hand.

D. Workflow Management System

The Workflow Management Coalition (WfMC) defines a Workflow Management System as: “A system that completely defines, manages and executes “workflows” through the execution of software whose order of execution is driven by a computer representation of the workflow logic.” [28] Insofar, WfMSs are reduced to the strict execution of workflow logic represented by more or less complete process models. They provide support in three functional areas:

- Build-time functions concerned with defining and modeling the workflow process
- Run-time control functions concerned with managing the workflow processes in an operational environment and sequencing the various activities to be handled as part of each process
- Run-time interactions with human users and IT application tools for processing the various activity steps

Likewise, [14] refers to workflow management systems as consisting of two parts (build time and run time). “The build time part allows a modeler to define and maintain all the information necessary to eventually execute workflows. The workflows are finally performed by the run time part of the workflow management system.” Thus, only the run-time functionalities (Workflow Enactment Service) lay within the scope of this paper, are therefore examined in detail and are stated by the workflow reference model as follows [28]:

- interpretation of the process definition
- control of process instances - creation, activation, suspension, termination
- navigation between process activities that may involve sequential or parallel operations, deadline scheduling, interpretation of workflow relevant data
- sign-on and sign-off of specific participants
- identification of work items for user attention and an interface to support user interactions
- maintenance of workflow control data and workflow relevant data, passing work-flow relevant data to/from applications or users
- interaction with external resources (client or invoked application)

One well-known WfMSs today is Declare [29] based on a constraint-based approach and offers with it support for loosely-structured processes.

Within the last decade of research activities on process management a "transition from workflow management systems focusing on automation of structured processes to Business Process Management (BPM) systems aiming at a more comprehensive support of a wider variety of business processes" [25] can be observed. "BPM extends the traditional WFM approach by support for the diagnosis phase […] and allowing for new ways to support operational processes […]" [26]. Because this enhancement does not affect the results of this paper, the term WfMS is used here, but with keeping in mind the concept of business process management systems.

E. Process Navigation System

During the last years, the increasing demand for so-called flexible WfMSs has caused a rethinking on IT-supported process enactment. A flexible business process is understood to restrict participants to a lesser extent [34] and must therefore comprise more possible paths than a rigid one.

In this context, two paradigms are distinguished: the imperative and declarative style of modeling business processes. For an imperative model, every possible path must be foreseen at design (build) time and encoded explicitly. A missing path is considered not allowed. Typical imperative languages include the Business Process Model and Notation (BPMN), Event-Driven Process Chains (EPC) and UML Activity Diagrams. In declarative modeling, on the contrary, only undesired paths and constellations are excluded so that all remaining paths are potentially allowed and do not have to be foreseen individually. The declarative way of modeling is considered best suited for the flexible type of business processes [30].

Declarative modeling is based on constraints that relate events of the process case and exclude or discourage from certain correlations. We argue that both constraints and events must be able to involve all relevant perspectives of business process like, e.g., incorporated data, agents performing the work and utilized tools [31]. On this way it becomes possible to express realistic correlations like, e.g., the actual performing agent of a process step affecting the type of data used in another step [31].

Besides turning to declarative cross-perspective modeling, a Process Navigation System must support different modalities for differentiating between optional and mandatory constraints. Thus, process modelers may distinguish easily between the hard boundaries of a process (mandatory constraints) and best practices (optional constraints) and are therefore able to formulate business process models of a very high expressivity.

Flexible processes usually involve human participants. A Process Navigation System enacting these human-centric processes should therefore be treated as a decision support system instead of a means of automation. As a consequence, an according system should propose actions and support them but it should never enforce them [32]. This requirement goes along with the call for process management systems "to provide directions and guidance rather than enforcing a particular route" whereas they are compared to navigation systems [39]. An important characteristic of decision support is explanation and so-called meta-knowledge [33]. Proposals made by the system need to be justified and explained so that sound choices can be made. For process execution, this means that certain proposed actions must be marked as recommended and that discouraged actions can be traced back to and explained by the according parts of the business process model. This characteristic of a Process Navigation System will be referred to as traceability.

An execution engine that aims at fulfilling the above requirements and therefore allows for declarative, cross-perspective and multi-modal process models and a traceable execution is being developed [41].

V. Process Observation and Guidance through Process Execution

This section is mostly part of our previous work of [1]. Process mining techniques allow for automatically constructing process models. The algorithms are analyzing a process execution log file, in the following referred to as (process) log; this log is usually generated by information systems (IS). However, there are processes that are not executed by information systems. This is an observation that is very important for the classification of our research. Thus, in order to define the application area of our project we have to introduce three different types of process execution support, classified upon the degree of logging and execution support (Fig. 3):

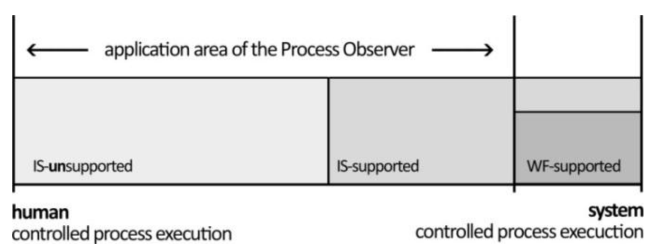


Figure 3. Application area of the Process Observer project

- *IS-unsupported*: Here, processes are executed without the support of any information system. Thus, there is no log for these processes. Furthermore, these processes are also not supported during execution. For example, there is no information system that guides a user through the process.
- *IS-supported*: Here, processes are executed by an information system. Processes of this type are (possibly) logged. However, the information system is not directly

guiding users through the process. The user has to find his way through the information system by himself.

- *WF-supported*: Here, processes are executed by Workflow management systems (WFMS). WFMS build a subset of IS. Typically, they maintain a process log. Additionally, the process participants are guided through process execution with concrete recommendations of how to continue process execution (work list) [16]. The basis for the successful generation of process models through process mining is an existing and complete log. Thus, WF-supported processes are a great source for process mining. Nevertheless, the existence of a process log is the main prerequisite and also the major drawback for a successful application of process mining. Since we assume that in many applications, WF-supported processes will not be encountered, PO turns its attention to IS-supported and IS-unsupported processes (Fig. 3). In order to log IS-unsupported processes, we extend process execution by manual logging. We define the term manual logging as the user action of entering process execution data (e.g., process IDs, documents, and services) as well as of marking process execution events, among other things process start and completion. The action of manual logging is implemented by the PO Logging Client. Finally, our goal is to provide manual logging in such cases when processes are neither IS-supported nor WF-supported. The final aim is then to be able to apply process mining.

A. *Aims of the Process Observation Project*

The challenge of PO is to provide a broader basis for process mining by implying IS-unsupported processes in logs. Therefore, the PO project aims at the adoption and generation of manual logs. The generated manual logs open the opportunity for the automatic generation of process models by process mining techniques even for applications that do not involve information systems. As manual logging is performed by process participants, it means additional work for them. Therefore, PO must offer a stimulus that motivates process participants to support manual logging. Since PO is particularly of interest for IS-unsupported and IS-supported processes, it offers a stimulus with respect to process execution guidance (this is what these two kinds of processes are lacking). PO offers recommendations about how to continue a process execution and offers auto-suggest support. This kind of guidance during process execution is typically exclusively offered by WFMS.

B. *Generation of Manual Logs*

From now on, we generally assume that a complete and freely analyzable log is not available, i.e., we are focusing on IS-(un)supported processes. We regard this scenario as the most common one and it needs to be supported to apply process mining.

1) *Manual Logging:*

The generation of a manual log is initiated by the PO Logging Client. Process participants record what they are currently doing, i.e., they provide information about the process they are currently performing. It is very important that users do not need any process modeling skills to record this information. An important issue is to determine what data the process participants should record. We recommend to record data based upon the different aspects of POPM. We have experienced that most users are very familiar with the approach of describing process in the POPM method. Process participants have to enter data according to the following perspectives:

- *Functional* perspective: name of the current process step, the name of the corresponding superordinate process (if available)
- *Data* perspective: data, i.e., documents or generally information that was used by the current process step as well as the data or documents that were produced
- *Operational* perspective: tools, applications or services that were used during the execution of the currently executed process step
- *Organizational* perspective: information about the process executor (typically, this is that person that is logged into the PO Logging Client), the personal information is enriched by group and role memberships

Besides, process participants have to trace process execution: he has to declare that process execution starts, ends or is aborted.

2) *Merging Logs:*

The application of the PO Logging Client finally results in the generation of a manual log. In the case, that an information system is applied, there might also be an automatic log available. We harness this situation by combining the manual log with the automatic log. Doing this, missing process information of one of the logs can be completed by the other log. In order to be able to combine the two logs, conformance between the recorded data of both logs must be achieved. Therefore, we suggest a component for merging the logs, i.e., locating (matching) and unifying processes that were recorded in the manual log as well as in the automatic log. This results in one consistent log that contains the execution data of IS-unsupported as well as IS-supported processes.

C. *Guidance through process execution*

According to our classification in Fig. 3, many process executions are not assisted by a guidance component, i.e., the participants must decide for themselves which process step they want to perform next. Only WF-supported processes do provide this feature. In this subsection, we will show how PO offers such guidance. It consists of two sub-features: dynamic recommendations and auto-suggest function.

1) *Dynamic Recommendations:*

Dynamic recommendations are generated in the following way. After the completion of a process step, PO immediately starts a process mining algorithm analyzing available log data. It then tries to classify this current process execution into former process executions. If it is successful, PO can recommend the execution of a subsequent process step according to the processes that have been executed formerly. This recommendation service becomes more and more reliable the more process instances have been executed under the guidance of PO. When only a few or even none processes of this type have been executed so far, no recommendations can be made for the particular process. Especially when only a few process instances have been performed so far, the recommendation can be inconsistent. Then, process participants can ignore this recommendation. In order to know about the quality of the recommendation, the number of process instances the recommendation is based upon is displayed in the user interface.

Example: A process participant just completed a process step A. This step has already been completed and recorded 10 times before by other agents. On the one hand, step B was executed 7 times after step A; on the other hand, step C was executed 3 times after step A. PO now starts process mining and generates a process model that contains the information that process A shows two subsequent processes B and C. Furthermore, the tool takes into account that step B occurred 7 times and step C occurred 3 times after step A in the log. Thus, a dynamic recommendation is shown to the user suggesting to continue with step B (70%) or step C (30%).

2) Auto Suggest Function:

The second aspect of guidance during process execution is provided by an auto-suggest function. This function helps the process participant to enter required information. PO compares previously recorded process names, data, tool names, etc. with the currently entered term and auto-suggests terms. This function supports two issues: first, the user might nicely be supported through information provision; secondly, by suggesting already used terms, the probability of having to deal with too many aliases in the system is diminished to a certain extent.

Example: Agent 1 is executing a process "Drinking Coffee". Agent 1 starts the process by recording the process name, i.e., Agent 1 enters "Drinking Coffee". The agent starts and completes the process. The process gets a unique identifier and is recorded in the log. Later, Agent 2 also wants to drink coffee and executes this process with support of PO. He starts by typing "Coffee" instead of "Drinking" in the process name row. This would easily result in the recording of a process name like “Coffee Drinking” or just "Coffee". So, aliases are produced without even recognizing. However, in this case an auto suggestion will appear, recommending to choose the process "Drinking Coffee". Agent 2 happily chooses the suggested process and thus ensures homogenous naming of the process step.

3) Visualization and manual mapping of processes:

Example: If the example from the former subsection would occur as described, this would be ideal. However, in many cases same processes will be referenced by different aliases and thus stay unrecognized by PO. In order to handle problems like this, PO offers an administration interface, which allows process administrators to visualize recorded processes. Administrators can start process mining algorithms and thus generate process models visualizing observed processes. Doing this, different aliases of processes can be discovered. However, this must be done manually by the administrator. In order to map different aliases of the same process, the PO administration interface offers a mapping panel. This mapping can be declared valid for multiple processes (Fig. 4). After defining a mapping between processes, a repeated execution of process mining results in the visualization of the amended process model.

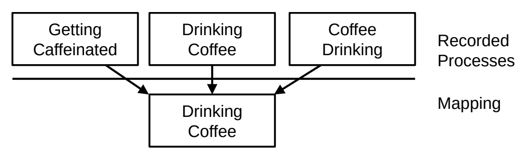


Figure 4. Sample mapping of recorded processes

VI. ARCHITECTURE AND IMPLEMENTATION

The following section mostly relies on our previous work of [1]. In this section, we will describe the architecture and implementation of PO. In the first part, we will show implementation details of the PO Logging Client. After that, process mining implementation and data structures will be explained. Furthermore, we present the administration and mapping components.

A. Process Observation Logging Client

The core of PO is constituted by the PO Logging Client. We decided to choose a web based implementation of the logging interface. This guarantees a great coverage of application scenarios, i.e., PO can be used in almost all applications. If the users are working in a "normal" office, PO can run on a stationary PC or notebook, if users are working "in the field", PO could as well run on a mobile device (e.g., smartphone). For our prototype we chose an implementation based on Microsoft ASP.NET 4.0 and the MS SQL Server 2008 database, but surely any equally equipped database and server technology would be suitable. The core of the web application that implements the PO Logging Client is located on a web server connected to a database. Users have to identify themselves by logging in with their username and password. Users can be assigned to one or more organizational roles. Hence, recommendations and suggestions can be personalized to the users' roles. When users enter process names they want to log, these text strings are immediately sent to PO to test for similar process names. The names of all processes containing a similar string are sent back to the client as a generic list. This list is finally displayed to the user as an auto suggestion list (Fig. 5). The user can select a process from this list. If none of the suggested processes is appropriate, the input process name is added as a new process. Accordingly, all other process data are captured (e.g., superordinate process, current process instance, used and produced data/documents and supporting tools). Finally, the user starts the process.

Figure 5. Example of auto suggestion list

B. Implementation of process mining, data structures and dynamic recommendations

As already described in Section V, PO offers dynamic recommendations of how to continue after finishing a process step. Therefore, a process mining algorithm is executed after each process step. In our prototype we use the alpha algorithm of [4] in order to analyze the available logging information. The algorithm analyzes the log and builds up a dependency graph. Therefore, we used the graph data structure QuickGraph of [17]. For implementation details concerning the alpha algorithm we refer to [4]. The logged execution information results in process models represented as graphs. A node is an instance of a class "Process" containing fields for process name, the executing originator role, used and produced data items as well as supporting tool items. Furthermore, the class contains two fields for the pre- and post-connectors which represent the semantic connection to previous and following processes. This information is also provided by the alpha algorithm. Once a process model has been generated as a graph, PO can use it in order to display recommendations after a user has finished a process step. Therefore, the recently completed process is searched within the process model, i.e., the graph is traversed until the current process ID is identical to the recently completed one (Fig. 6).

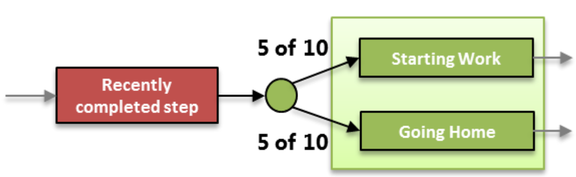


Figure 6. Example of an AJAX modal popup dynamic recommendation

After that, all available edges of this node are examined and their occurrence is counted. Like this, we generate a list, containing the processes that were executed after the recently completed one. Thus, a popup is displayed (Fig. 7), giving the user the possibility to choose the following process step.

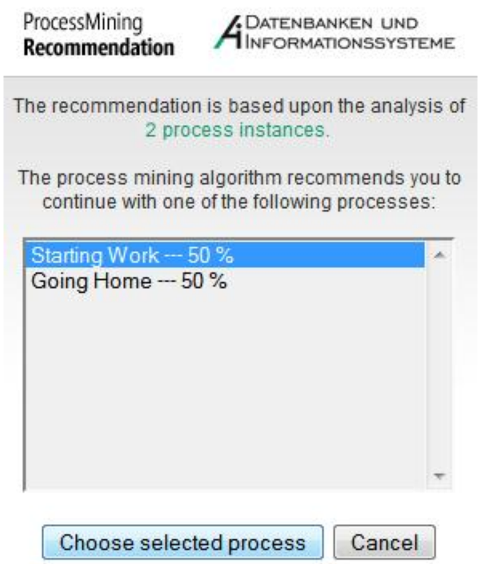


Figure 7. Example of an AJAX modal popup dynamic recommendation

C. Administration interface

Additionally, PO offers an administration interface that allows process administrators to visualize recorded processes as well as defining mappings between logged processes as described in Section V. The application consists of two panels, one for process model selection and visualization and the other one for defining mappings between processes. One could easily imagine additional applications, like agent-role assignments or dataflow applications. Those are planned for future versions.

1) Process visualization:

In order to visualize the generated process model we use basic graph visualization frameworks. In our prototype we used the Graph# framework [18] to display the QuickGraph data structures. The visualization procedure is started by examining the recorded event log for contained composite processes. A process is recognized as composite, if it was chosen as a superordinate process by a process participant during the logging phase of a process with PO. The names of the composite processes are loaded in a tree view. The user selects a composite process that should be displayed from the tree view. The tree view shows the underlying process hierarchy. Processes that are contained within another one can be displayed by extending a process entry. After the selection of an entry, all event log information concerning the selected process is fetched from the database. After that, the alpha algorithm is applied to the resulting event log data. As stated before, the algorithm generates a dependency graph. This graph is finally assigned to the Graph# framework and displayed to the user. Here, the user has various possibilities to scroll within the visualization or to open the model of underlying composite processes by selecting the corresponding process nodes.

2) Mapping definition panel:

Furthermore, the administration interface offers a separate panel to define mappings between logged processes. Therefore, the database provides a separate mapping table with three columns: "superordinate process", i.e., the super process within the mapping is valid, "target process", i.e., the process on which another one is mapped and finally "mapped process", i.e., the process which is mapped. Considering this data model, the mapping panel consists of three columns, too. They appear after the first things first principal. In the first list, the user selects the superordinate process within the mapping should be valid. After this selection, the target process list appears. The list is initialized with all processes occurring within the chosen superordinate process. Like this, the user can choose the target process for the defined mapping. Last but not least, the last list, i.e., a checkbox list, appears. It is again initialized with all processes of the corresponding super process. Here, the user checks all the corresponding boxes of the processes he would like to map on the target process chosen before. Finally, the mapping is applied to the database.

VII. Using Process Observation to Support Evolutionary Process Engineering

In this section, we will show how process discovery (i.e. the automatic generation of process models based upon event logs), process guidance and process evidence as the main basic functions of PO can be applied as support for reasonable process evolution. The presented examples und use cases refer to the maturity levels for process evolution described in Section II.

A. Process Observation Applications for Process Evolution Support

PO can be applied in three respects to support process evolution:

1) Engineering

Firstly, PO can be used to reach a desired evolution stage or rather facilitate the transition between two maturity levels by enhancing the process model (process discovery) conveniently. The engineering function accomplished by the PO is to gradually attain missing parts or details of particular process perspectives according to POPM that are needed to comply with the target maturity level.

Example: ML2 requires a time schedule. In order to engineer the behavior perspective PO inquires the executed process steps over a specific period. In combination with timestamp information the PO derives execution order and dependencies. The attained process model can now be enacted by a project management system and e.g. visualized through a Gantt chart.

2) Enactment

Secondly, PO serves as an implementation for certain evolution stages. Among other information systems, PO provides execution support (dynamic guidance, recommendations) through enactment of extracted process models (see Section V) and thus can be used systematically to meet the requirements for process support as prescribed by the maturity levels.

Example: PO can provide flexible execution support for ML3 by suggesting valid process steps and applicable tools according to the reference process. Since the user himself decides what he actually does, PO cannot prevent deviations from standard but advise him of not being compliant after comparing the event logs with the reference model. Additionally, the user is guided through the process by suggesting best-practice rules as visualized in Fig. 7.

Which maturity levels are further covered by PO and what are other possible enactment approaches is clarified by the two use-cases in the following subsections B and C.

3) Documentation

Thirdly, PO is able to gather information about the way processes have actually been executed and therefore can be consulted to prove the conformance to the quality requirements of the maturity levels (process evidence). Since PO has access to both manual and automatic logs, it is able to provide complete process execution event logs as documentation and audit trail even if the process is partly executed beyond the control of any information system.

Example: IT management is about to analyze the information system landscape and therefore desires a consolidated documentation and quantity structure across all processes about the usage of tools and services (operational perspective) with relation to the respective professional need (functional perspective). Since existing automatic logs are not sufficient, the PO is recording the usage of tools and services as for the "offline" processes that are not provided with any process log or documentation. Through merging these manual logs with the automatic logs a cross-process reporting view is established.

B. Use-Case 1: From ML1 to ML2

Initial situation: The process results are achieved somehow without the help of any IS. The process steps are visualized on wallpaper. It serves as orientation for the participants (ML1).

Target situation: Proper results should be achieved in time (ML2). A project management system (PMS) is intended to support both project manager (PM) and participants. The PM should be provided with up-to-date information about the progress in order to be able to monitor and control the process.

Engineering: The first use case comprises the generation of a manual log (Fig. 8) without an information system being available. The participating agents are executing the corresponding processes under the guidance of PO. The manual log is finally analyzed by process mining algorithms. The resulting process models can be fed into a PMS if wanted and if available. Thus, process models can afterwards be enacted by a PMS.

Both order of the process steps and time flow are engineered by PO through mining of manual logs. Beside the logged in user or role, PO is able to gather start, end and abortion timestamps. When a user declares he is about to start a new step, it could also be possible that PO additionally asks him when it will presumably be completed. By this means, all required runtime information for comparison with the schedule is made available. Already during the engineering process the project manager is able to gather the current project status by consulting the PO logs and align this information with wallpaper.

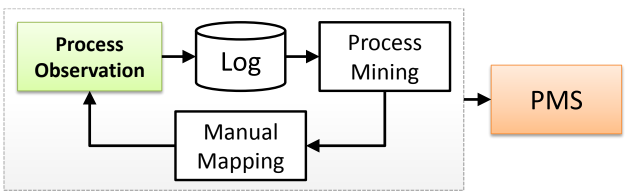


Figure 8. First use case – generation of manual logs

Enactment: We assume that after the transition from ML1 to ML2 the wallpaper is not needed any more and execution support including recognition and communication of project plan deviations (as demanded by ML2) is finally accomplished by the project management system. Therefore, the process structure attained through process discovery serves as project plan template and is enacted by the PMS. Actual due dates can be added manually by PM. At runtime, the progress information is either permanently updated by the PO automatically or maintained by the users and the project manager manually. In both cases the project plan update has to be attained through user interaction.

Documentation: Through combination of manual and automatic logs provided by the PO and PMS it will succeed to prove that the process is managed in such a way that proper results can be achieved in time and thus complies with ML2.

C. *Use-Case 2: From ML4 to ML5*

The second use case comprises the application of PO in parallel to an information system (Fig. 9).

Initial situation:

Process execution is already supported by a fully-fledged WfMS. However, traditional WfMS do not allow for deviating from enacted process models. That's why process participants have to build workarounds in order to be able to execute process steps that had not been considered during process modeling phase, e.g., exceptions or newly occurring process cases. From time to time, process managers have to discuss, probably on the basis of statistically measured KPIs, if the process can be improved somehow (ML4). Thereby, it will be reviewed, if enacted process models had been an adequate basis for the running processes or if additional use-cases or exceptions that had not been considered in process models yet should be introduced in order to improve the performance. Therefore, process managers have to adapt process models manually according to the results of the discussions. Note, that this situation has already been described in the introduction of this article. Traditional process management restricts the support of process execution by strictly separating modeling and execution phase.

Target situation:

If standard process models are not suitable or if a different way promises better results, process participants should have the possibility to deviate from standard during process execution without losing control and overview. There should be an adequate consideration of exceptions: dynamically occurring exceptions should be supported by systems. Furthermore, there should be a systematical support of continuous process improvement (as demanded by ML5).

Engineering:

The research achievements of the previous sections, i.e., Process Observation, offer possibilities to overcome the strict separation of modeling phase and execution phase by applying real-time analysis methods of executed processes. Dynamically occurring exceptions (that are not yet part of enacted process models) are engineered by PO through manual logging and subsequent process mining. On this way, it is possible to overcome the strict separation of modeling and execution phase and process models can be completed automatically. Newly occurring process steps or exceptions are discovered by PO and added to the resulting enhanced process models. Furthermore, there is an adequate feedback: with PO, it is possible to consider feedback from process execution to process modeling. The knowledge of previous executed process cases could be added to process models as recommendations, e.g., which path through the model was the fastest or cheapest one.

Enactment:

We assume that after the transition from ML4 to ML5 the process model enactment of a running WfMS is featured by the application of PO. The intention is to complete the process execution log information mutually. Identical processes are merged to one single process. Process Mining is finally applied to the joint log (Fig. 9). Identified processes can be fed back into information systems. On this way, the resulting process models are enhanced by newly occurring process steps and recommendations. The joint application of WfMS and PO provides a proper solution for process improvement as demanded by ML5.

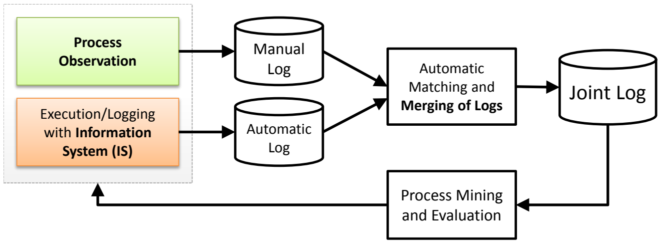


Figure 9. Second use case – merging of logs

Documentation:

The joint application of a WfMS and PO provides a process execution log that contains execution information about process steps that have been executed with support of WfMS as well as under the guidance of PO. On this way, the joint log provides evidence for continuous process improvement. Therefore, we define a value *compliance_level* (1) as the number of matched processes from the manual log and the automatic log divided by the complete number of processes recorded in the manual log. The procedure of calculating this value is the following: the algorithm runs through both logs. It compares each process of the manual log with the processes of the automatic log. If a matching algorithm identified two processes as equal, the numerator *#compliance_level* will be increased by 1. After finishing traversing both log files, the resulting value of *#compliance_level* is divided by the total number of recorded processes within the manual log.

$$compliance\_level = \frac{\#matched\_processes}{\#recorded\ process\ with\ PO} \quad (1)$$

Like this, the calculated value reflects how many processes are already executed with support of the WFMS. For a special organization a *compliance_level* value of 0.9 may be enough. This means, 90% of the executed processes are implemented and supported by the WFMS. Like this, the value reflects the continuous improvement of enacted process models.

VIII. CHANGES WITHIN THE PROCESS LIFECYCLE THROUGH THE APPLICATION OF PROCESS OBSERVATION

In this section, we will describe the impact of the application of PO on different phases in the process lifecycle. As already mentioned, the previous process lifecycle [2] consists of an initial modeling phase that is very time consuming. In this lifecycle, process mining is only used for the evaluation of the process being executed with support of a WFMS. As any WFMS needs at least one predefined process model in order to be operable [4], there is no possibility to support the intense process modeling phase with the automatic process discovery possibilities of process mining. The development of the PO offers the possibility to change this situation. With support of the PO, the lifecycle can be rearranged in the following way (Fig. 10).

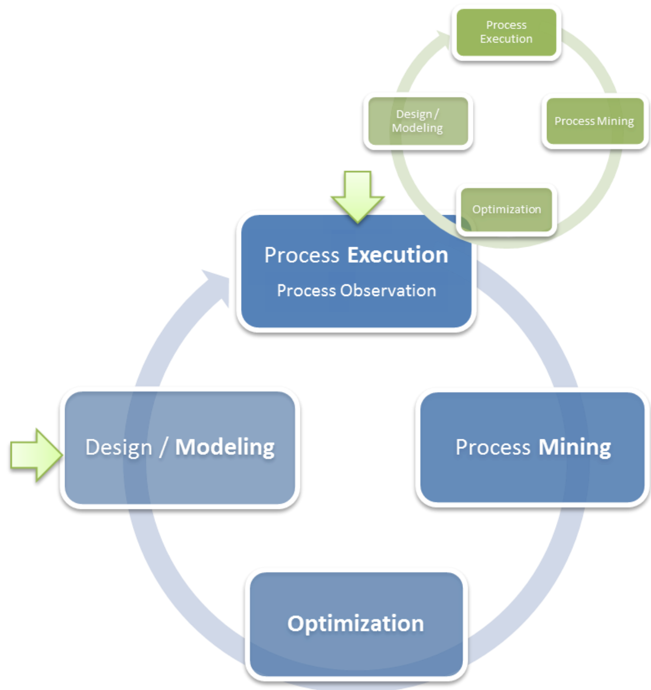


Figure 10. Adapted process lifecycle through the application of PO

The new lifecycle contains two different entry points. In addition to the traditional initial modeling phase, the process lifecycle can be entered by process execution (as usual) accompanied by manual logging, i.e., the generation of a manual log, with PO. Note, that the execution phase consists of a lifecycle in little again: after the execution of every process step, the process observation tool analyses the existing process execution log by process mining methods. The mining results are directly sent back to the process participant who has the possibility to react on the results dynamically during the running process (online process observation). On this way, it is possible to learn from previous mistakes on runtime (Optimization) and to change ones behavior, i.e., the path through the process (Modeling). Having executed a desired number of process instances (cases), the process observation tool can be used to extract process models that can be enacted in other process enactment approaches as stated in Section IV. Here, the results of process mining possibly have to be reworked in a process remodeling phase (Optimization). The benefit of the application of PO consists of the time saving between the previous process modeling phase and the less time consuming remodeling phase.

The previous modeling phase, i.e., the project of process discovery and process definition, had to be operated completely manual. The process management team had to do several interviews with agents, live observations of processes and the tracking of documents, for example. In contrast to that, process discovery with PO is generally more automatable. Merely reworking effort is required in order to annihilate possibly occurring exceptions or execution errors. Based on the results of these first steps, business processes can be evaluated and finally optimized.

IX. Related Work

The last decades of research activities concerning business processes led to a growing amount of process enactment approaches, including a great variety of types of enactment support as well as many different supporting implementations.

Many publications are engaged in the comparison of particular implementations of WfMS and BPMS respectively. [20] offers a good market overview by evaluating the top 25 vendors in the BPMS market. The results are displayed in the so-called Magic Quadrant, spanned by dimensions 'ability to execute' and 'completeness of vision'. Each vendor will therefore correspondingly be classified as niche player, visionary, challenger or leader. However, this evaluation considers only systems fulfilling economically oriented criteria. This approach is a practical one and does not satisfy research interests. Some further efforts like the pattern-based evaluation of scientific WfMS [21] or the evaluation of workflow management systems using meta models [22] examine implementations in a more scientific way and without the use of scoring strategies. For example, [22] introduces a meta model approach for the evaluation of WfMS. A methodology for the selection of workflow products is specified as done by well-known standardized scoring models for software evaluation. This kind of evaluation can only be applied to build-time components used for modeling of organizational entities and the creation of the process model. Furthermore, there must be a high degree of formalization on the user side, because business situations have to be depicted using formal methods.

Even though there are some promising approaches concerning the evaluation of enactment-related implementations, the approaches still remain comparisons of particular implementations. The article at hand stands in the tradition of research in [23] trying to answer questions on: What does "enactment" mean? What alternative enactment approaches can be distinguished?

Besides workflow management system, this article defines various existing types of enactment support, i.e., so-called *model-as-is*, *checklists*, *dynamic guidance* and *process navigation systems*. In this regard, the paper aims at getting over the current limited view of enactment as an automatically supported activity.

The work at hands introduces process evolution as an economic field of application for Process Observation (PO) [1] in order to achieve a higher degree of maturity. In order to support evolutionary process engineering, one of the basic functions of PO is enabling automatic process discovery.

The idea of automating process discovery through event-data analysis was first introduced by Cook and Wolf in the context of software engineering processes [7]. In the following years, Van der Aalst et al. developed further techniques and applied them in the context of workflow management under the term process mining [6]. Generally, the goal of process mining is to extract information about processes from event logs of information systems [8]. There are already several algorithms and even complete tools, like the ProM Framework [9], that aim at generating process models automatically. During the last decade, several algorithms have been developed, focusing different perspectives of process execution data. Van der Aalst et al. give a detailed introduction to the topic process mining and a recapitulation of research achievements in [6] and [8]. For the first prototype of PO, we used the alpha-algorithm of [4] and the heuristics-miner [10].

However, for our future research activity we consider declarative process mining algorithms like [36] appropriate. Declarative process modeling techniques offer the possibility to describe complex applications by comprehendible process models. In contrast to imperative modeling, declarative models concentrate on describing what has to be done and the exact step-by-step execution order is not directly prescribed.

Process mining algorithms rely on complete event logs from information systems. In the case of an incomplete log or even the unavailability of an information system, events can alternatively be recorded by manual activity tracking respectively task management methods. There are several approaches for activity tracking by capturing data on personal task management [11] [12]. However, these approaches are not process based. They are not analyzing execution orders and therefore it is not possible to extract adequate process models that can finally be enacted in process management systems.

In order to discover identical processes between different data storages, we suggest using basic automatic ontology matching algorithms [13]. Process mining is considered as a part of Business Process Management (BPM). BPM relies on a life-cycle where different phases of the process are focused. The traditional approach consists of the following phases: process modeling, implementation, execution and evaluation, started by the modeling step. Despite the successful development and evaluation of the process mining algorithms named above, process mining is ranked among the process evaluation phase [2]. Consider, for example, Enterprise Resource Planning (ERP) systems such as SAP, OpenERP, Oracle, Customer Relationship Management (CRM) software, etc. These systems require a designed process model before they go into service [4]. In these situations, process mining could only be used for process rediscovery and not for real process discovery. Therefore, we aim at assigning process mining to the discovery phase by recording the complete process data covering all aspects of POPM.

X. Conclusion and Outlook

In this article, we introduced process evolution as an economic field of application for process observation. We showed how process discovery, process guidance and process evidence as the main basic functions of process observation can be applied as support for reasonable process evolution. Furthermore, we introduced the spectrum of process model enactment approaches and put in context the concept of Process Observation as Dynamic Guidance approach. Process observation serves as an implementation for certain evolution stages itself and can additionally be consulted to prove the conformance to quality requirements of maturity levels.

Our future research activity in the field of Process Observation will start with the integration of declarative process mining methods and other data mining approaches like association rule mining in order to increase the quality and understandability of extracted process models and the power of the recommendation module. Furthermore, we will include information retrieval methods and implementations like search engines in order to facilitate process information provision. Additionally, we need to include matching methods in order to match and merge identical processes. Furthermore, we will face the problem of recording and logging processes in different granularities. This research faces one of the great challenges of process mining declared during the meeting of the IEEE Task force on process mining at the BPM conference in 2011. Finally, we are looking forward to an extensive application of the PO in an organization, accompanied by a detailed documentation of the practice.

Acknowledgment

The presented system is developed and used in the project “Kompetenzzentrum für praktisches Prozess- und Qualitäts-management”, which is founded by “Europäischer Fonds für regionale Entwicklung (EFRE)”.

References

[1] Schönig, S., Günther, C., and Jablonski, S., "Process Discovery and Guidance Applications of Manually Generated Logs," presented at the The Seventh International Conference on Internet Monitoring and Protection (ICIMP 2012), Stuttgart, Germany, 2012.

[2] Zur Muehlen, M. and Ho D., “Risk management in the BPM lifecycle”. In: Business Process Management Workshops, LNCS, vol. 3812, 2006, pp. 454-466.

[3] Zairi, M., “Business process management: a boundaryless approach to modern competitiveness”. In: Business Process Management Journal, vol. 3, 1997, pp. 64-80.

[4] Van der Aalst, W., Weijters, T., and Maruster, L., “Workflow mining: Discovering process models from event logs”. In: IEEE Transactions on Knowledge and Data Engineering, vol.16 (9), 2004, pp. 1128-1142.

[5] Rozinat, A. and Van der Aalst, W., “Decision mining in business processes”. In: BPM Center Report BPM-06-10, BPMcenter.org, 2006.

[6] Van der Aalst, W. and Weijters, A., “Process mining: a research agenda”. In: Computers in Industry, vol. 53 (3), 2004, pp. 231-244.

[7] Cook, J.E. and Wolf A.L., “Automating process discovery through event-data analysis”. In: 17th International Conference on Software Engineering, Apr. 1995, Seattle, USA.

[8] Van der Aalst, W., Reijers,H., Weijters, A., Van Dongen, B., De Medeiros, A., Songa, M., and Verbeek, H., “Business process mining: An industrial application”. In: Information Systems, vol. 32 (5), 2007, pp. 713-732.

[9] Van Dongen, D., De Medeiros, A., Verbeek, H., Weijters, A., and Van der Aalst, W., “The ProM framework: A new era in process mining tool support”. In: Applications and Theory of Petri Nets, LNCS, vol. 3536, 2005, pp. 444-454.

[10] Weijters, A., Van der Aalst, W., and De Medeiros, A., “Process mining with the heuristics miner-algorithm”. Department of Technology, Eindhoven University of Technology, 2006, Eindhoven, The Netherlands.

[11] Witschel, H., Hu, B., Riss, U., Thönssen, B., Brun, R., Martin, A., and Hinkelmann, K., “A collaborative approach to maturing process-related knowledge”. In: Business Process Management, LNCS, vol. 6336, 2010, pp. 343-358.

[12] Stoitsev, T., Scheidl, S., Flentge, F., Mühlhäuser, M., “From Personal Task Management to End-User Driven Business Process Modeling”. In: Business Process Management, LNCS, vol 5240, 2008, pp. 84-99.

[13] Noy, N. and Musen, M., “Algorithm and tool for automated ontology merging and alignment”. In: Proceedings of the 17th National Conference on Artificial Intelligence, Aug. 2000, Austin, USA.

[14] Jablonski, S. and Goetz, M., “Perspective oriented business process visualization”. In: Business Process Management Workshops, LNCS, vol. 4928, 2008, pp. 144-155.

[15] Jablonski, S. and Bussler, C., “Workflow management: modeling concepts, architecture and implementation”. International Thomson Computer Press, London, 1996, ISBN 1850322228.

[16] Jablonski, S., Igler, M., and Günther, C., “Supporting collaborative work through flexible process execution”. In: 5th International Conference on Collaborative Computing, Nov. 2009, Washington DC, USA.

[17] Halleux, J.d. “QuickGraph, Graph data structures and algorithms for .net”. Last access: Mar. 2012. Available: http://quickgraph.codeplex.com..

[18] Microsoft, “Graph# - Layout algorithms and graph layout control for WPF applications”. Last access: Mar. 2012. Available: http://graphsharp.codeplex.com.

[19] Verbeek, H., Buijs, J., Van Dongen, B., and Van der Aalst, W., “XES, XESame, and ProM 6”. In: Information Systems Evolution, Lecture Notes in Business Information Processing, vol. 72, 2011, pp. 60-75.

[20] Sinur, J. and Hill, J. B., "Magic Quadrant for Business Process Management Suites," Gartner Research ID Number: G00205212, 18.10.2010 2010.

[21] Migliorini, S., Gambini, M., La Rosa, M., and ter Hofstede, A. H. M., "Pattern-Based Evaluation of Scientific Workflow Management Systems," ed, 2011.

[22] zur Muhlen, M., "Evaluation of workflow management systems using meta models," in System Sciences, 1999. HICSS-32. Proceedings of the 32nd Annual Hawaii International Conference on, 1999, p. 11 pp.

[23] Jablonski, S., "Do We Really Know How to Support Processes? Considerations and Reconstruction," in Graph

Transformations and Model-Driven Engineering. vol. 5765, G. Engels, C. Lewerentz, W. Schäfer, A. Schürr, and B. Westfechtel, Eds., ed Berlin, Heidelberg: Springer 2010, pp. 393-410.

[24] Pesic, M., "Constraint-based workflow management systems:shifting control to users," Technische Universiteit Eindhoven, Eindhoven, 2008.

[25] ter Hofstede, A. H. M., v. d. Aalst, W. M. P., Adams, M., and Russell, N., Modern Business Process Automation: YAWL and its Support Environment. Heidelberg: Springer, 2010.

[26] van der Aalst, W. M. P.: "Business process management: a personal view," 2004.

[27] Günther, C., Schönig, S., and Jablonski, S., "Dynamic Guidance Enhancement in Workflow Management Systems," presented at the The 27th Symposium On Applied Computing (SAC 2012), Riva del Garda, Italy, 2012.

[28] Hollingsworth, D., "The Workflow Reference Model," Workflow Management Coalition1995.

[29] Pesic, M., Schonenberg, H., and van der Aalst, W. M. P., "Declare: Full support for loosely-structured processes," 2007, pp. 287-287.

[30] Fahland, D., Lübke, D., Mendling, H., Reijers, H., Weber, B. and Weidlich, M., "Declarative versus Imperative Process Modeling Languages: The Issue of Understandability". In: Enterprise, Business-Process and Information Systems Modeling (10th International Workshop, BPMDS 2009, and 14th International Conference, EMMSAD 2009, held at CAiSE 2009), 2009, pp. 353-366.

[31] Igler, M., Faerber, M., Zeising, M. and Jablonski, S., "Modeling and Planning Collaboration in Process Management Systems using Organizational Constraints". In: The 6th International Conference on Collaborative Computing: Networking, Applications and Worksharing (CollaborateCom 2010), Chicago, IL, USA, 2010, pp. 1-10.

[32] Dourish, P., Holmes, J., MacLean, A., Marqvardsen, P. and Zbyslaw, A., "Freeflow: Mediating Between Representation and Action in Workflow Systems". In: ACM Conference on Computer Supported Cooperative Work (CSCW '96), Boston, MA, US, 1996.

[33] Turban, E., Sharda, R. and Delen, D., "Decision Support and Business Intelligence Systems", 9 ed., Prentice Hall, 2010.

[34] S. Sadiq, W. Sadiq, and M. Orlowska, "Pockets of Flexibility in Workflow Specification," in 20th International Conference on Conceptual Modeling (ER'2001), Yokohama, Japan, 2001.

[35] Vanderhaeghen, D., Fettke, P., "Organisations- und Technologieoptionen des Geschäftsprozessmanagements aus der Perspektive des Web 2.0". In: Wirtschaftsinformatik, no.1, 2010, pp. 17-32.

[36] Schönig, S., Günther, G., Zeising, M., Jablonski, S., Discovering Cross-Perspective Semantic Definitions from Process Execution Logs. In: Proceedings of the Second International Conference on Business Intelligence and Technology (BUSTECH 2012), Nice, France. ISBN: 978-1-61208-223-3.

[37] Seitz, M.; Jablonski, S., "Evolutionäres Prozess-Engineering: Zum angemessenen Grad an Prozessunterstützung" in HMD - Praxis der Wirtschaftsinformatik, vol. 287, D. Ingenhoff, A. Meier, ed: dpunkt Heidelberg, 2012.

[38] Huth, C.; Erdmann, I.; Nastansky, L., "GroupProcess: Using Process Knowledge from the Participative Design and Practical Operation of ad hoc Processes for the Design of Structured Workflows", 34th Annual Hawaii International Conference on System Sciences (HICSS-34)-Volume 9, 2001.

[39] van der Aalst, W.M.P., "TomTom for Business Process Management (TomTom4BPM)," in Advanced Information Systems Engineering. vol. 5565, P. van Eck, J. Gordijn, and R. Wieringa, Eds., ed: Springer Berlin / Heidelberg, 2009, pp. 2-5.

[40] Schönig, S., Zeising, M., Jablonski, S., "Adapting Association Rule Mining to Discover Patterns of Collaboration in Process Logs". In: 8th IEEE International Conference on Collaborative Computing: Networking, Applications and Worksharing (CollaborateCom 2012), Pittsburgh, PA, USA, 2012.

[41] Zeising, M., Schönig, S., Jablonski, S., "Improving Collaborative Business Process Execution by Traceability and Expressiveness." 8th IEEE International Conference on Collaborative Computing: Networking, Applications and Worksharing (CollaborateCom 2012), Pittsburgh, PA, USA, 2012.

[42] Object Management Group (OMG), "Business Process Maturity Model (BPMM)", Version 1.0 (June 2008). Last Access: Sep 2012. Available: www.omg.org/spec/BPM/1.0/.

[43] Software Engineering Institute of the Carnegie Mellon University, "Capability Maturity Model Integration (CMMI)". Last Access: Sep 2012. Available: www.sei.cmu.edu/cmmi/.

Towards an Improved IT Service Desk System and Processes: A Case Study

Marko Jäntti
School of Information and Communication Technology
Nathan Campus, Griffith University
170 Kessels Road, Nathan
Brisbane, Queensland 4111, Australia

University of Eastern Finland
School of Computing
Software Engineering Research Group
P.O.B 1627, 70211 Kuopio, Finland
Email: marko.jantti@uef.fi

Anup Shrestha, Aileen Cater-Steel
School of Information Systems
Faculty of Business and Law, University of Southern Queensland
Toowoomba, QLD 4350 Australia
Email: {anup.shrestha, caterst}@usq.edu.au

***Abstract*—An IT service desk provides a Single Point of Contact for the customers and users regarding support requests. The world-wide adoption of IT Infrastructure Library (ITIL) framework has also pushed organizations to improve their service desk operations. However, improving the service desk is a serious challenge for many IT service providers. Many government organizations in Finland have started to use ITIL in their service desks and need help in configuring service desk tools and defining processes. The research problem of this study is: How could an IT service provider's service desk operations be improved by using IT service management best practices? The main contribution of this paper is to present results of a case study on IT service support in Finnish Tax Administration. First, the customer support challenges are described. Second, we present how these challenges were solved by using ITIL-based practices. Third, we show how service desk performance and ITSM training usefulness were measured. Finally, lessons learnt from the case are presented.**

Keywords-Service Desk; service management; IT service; incident management.

I. Introduction

The service desk is a crucial contact point between customers, users, IT service providers and third-party service providers. The service desk is responsible for carrying out incident management and request fulfillment processes. The main objective of the service desk is to restore normal service for users as quickly as possible. Because the service desk is located at the heart of the customer interface, it has a strong impact on customer satisfaction or dissatisfaction.

The service desk is known by many names such as help desk, support center, information center, IT solutions center or technical support. There are also call centers and contact centers that handle contact requests, service requests and complaints but they do not focus on providing specialist support for solving problems. The goal of IT service desk agents is to record, classify, diagnose and resolve service desk cases from customers and users. Service desk cases can be incidents (e.g., software or hardware failures), service requests (e.g., requests for resetting passwords), complaints or feedback. Service desk engineers may also be responsible for identifying requests for a change (a request for change, RFC of an existing service or a feature in a service) or problems (causes of repeating incidents).

The service desk is often the most visible part of the IT organization. It is a function that is under continuous improvement and thus a very fruitful research target. This paper focuses on the improvement of the service desk both from a tool and process perspective. The paper is based on the previously published conference paper on service desk challenges [1]. This paper provides a wider literature review on service support and provides a more detailed case study description.

The IT service desk provides customers and users with support for using IT services. Information system and technology (IST) services can be broadly classified into four categories [2]: *Application services* are services that are delivered via software applications. *Operational services* maintain the IT environment such as installation services for hardware and software, change management, and trouble-shooting services and running the data centre. *Value-enabling services* increase the value of information assets (e.g., consulting, systems design, and help desk). Finally, *infrastructure services* focus on technical capabilities of IT infrastructure, such as capacity and security of IT assets. Customers aim to achieve certain business benefits by using the IT services. Therefore, if services are not available or there are many incidents that cause barriers for service usage, those business benefits may not be fully realized.

One of the most important tasks of the service desk is to effectively communicate with customers. This includes keeping customers and users informed on the progress of their support requests. Frequent communication usually means better customer service experiences. Shaw and Ivens

[3] define customer experience as an interaction between an organization and a customer. They describe the customer experience as “a blend of an organization’s physical performance, the senses stimulated and emotions evoked, each intuitively measured against customer expectations across all moments of contact”. One of the most interesting parts in their work is a customer experience pyramid with four sides: marketing, sales, service and support. Each area consists of building blocks: elements, sub-elements, standards, measures, targets, and improvements.

In the context of IT service management, the service desk is the interface of customer experience. Regardless of the maturity of IT service processes, sophistication of the ITSM tools and the skills and training of the service staff, customer experience in ITSM is provided by the service desk. The service can only be experienced by the customer through the service desk function of the IT service provider. Hence the service desk function must deal with "all" service requests coming through phone calls, the internet, email or in person [4]. The role of ITSM in this environment is not only to provide an effective customer experience but also a consistent one across all channels.

A. *Related work*

The service desk belongs to the services operation phase in the services lifecycle within services computing [5]. Many IT service provider organizations consider the improvement of IT service management processes as a difficult and challenging task. Typically, the process improvement is based on the processes and methods of the IT Infrastructure Library (ITIL). ITIL is the most widely used IT service management framework consisting of a set of best practices for managing IT services. One of the main benefits of ITIL is standardization of the customer support terminology. Lack of standardized terms and terminology differences cause problems for classifying support requests and communication between different service provider organizations.

The service management section of ITIL version 2 consists of two parts [6]: 1) Service Delivery (Service level management, IT financial management, availability management, capacity management, IT service continuity management) and 2) Service Support (service desk function, incident management, problem management, change management, configuration management and release management). ITIL version 3 emphasized the service lifecycle thinking and introduced five core books: Service Strategy [7], Service Design [8], Service Transition [9], Service Operation [10] and Continual Service Improvement [11]. The recent update as ITIL 2011 edition did not introduce radical changes to Service Operation processes. However, the Service Strategy book was completely rewritten.

In order to improve IT service management processes, organizations can use various IT service management frameworks, such as the Control Objectives for IT and related Technology (COBIT) framework [12], Microsoft Operations Framework (MOF) [13], Kapella’s Framework for Incident Management and Problem Management [14], IT Service Capability Maturity Model [15] or IT service management standard ISO/IEC 20000 [16].

Several process improvement guidelines are available for improving IT service processes. Two of the most popular frameworks that support ITSM process improvement are Tudor’s ITSM Process Assessment (TIPA) [17] that uses ITIL and the international standard for process assessment ISO/IEC 15504; and Standard CMMI Appraisal Method for Process Improvement (SCAMPI) [18] that can be used to improve organization’s processes based on the model of CMMI for Services (CMMI-SVC). Moreover, Lahtela et al. [19] have explored how to measure IT service management processes for improvement in practice and discussed the value of real-time measurement information from IT service support processes.

Recently, there has been collaboration between the field of ITSM and process assessment in the International Organization of Standardization (ISO) publication. The ITSM standard ISO/IEC 20000 Technical Report Part 4 has been published showing how ISO/IEC 20000 requirements are translated into a process reference model (PRM) [20]. This model defines the ITSM processes in a life cycle described in terms of process purpose and outcomes along with an architecture describing the relationships between the processes [17]. Such a PRM is a requirement for a conformant assessment using a process assessment model (PAM) based on ISO/IEC 15504 [21]. A PAM provides a detailed model based on one or more PRMs for the purpose of assessing process capability [22]. The recently published Part 8 of standard ISO/IEC 15504 provides an exemplar process assessment model for ITSM [23]. The use of these standards can provide a useful tool to assess process maturity, performance and work products.

These frameworks are crucial to improve the underlying processes of the service desk, i.e., incident management and request fulfillment and the interfaces of these processes with other processes such as problem management and event management. A frequently asked question in IT service management is the differences between ISO/IEC 20000 standard, ITIL and COBIT frameworks and their role in process improvements within the function of service desk.

First, COBIT framework [12] is designed for IT management for governance purposes. COBIT 5 includes 37 governance and management processes that are categorized under five COBIT domains: Evaluate, Direct and Monitor; Align Plan and Organize; Build, Acquire and Implement, Deliver Service and Support; Monitor, Evaluate and Assess. Many of the ITIL processes can be easily found in the COBIT framework. For each process, COBIT defines the goal of the process, control objectives, process inputs and outputs, key activities, roles and responsibilities (usually defined using RACI chart), metrics and a process maturity model. Goals and metrics have been divided into three groups: IT, Process and Activities. Managing Service Desk and Incidents is included in the COBIT framework in the Deliver, Service and Support domain.

While the COBIT format is very compact (4 pages for each process), ITIL provides detailed guidance on how to implement best practices. For example, ITIL v3 Incident Management [10] consists of 18 pages. For each process, ITIL describes process objectives; scope; value to business; policies, principles and basic concepts; process activities, methods and techniques; metrics; roles and responsibilities; critical success factors and risks.

ISO/IEC 20000 is an international standard for IT service management. It defines the auditable requirements for IT service management system. The standard consists of several parts with different purposes. ISO/IEC 20000-1 defines the mandatory (shall) requirements for a service management system. ISO/IEC 20000-2 Part 2: Code of practice for service management provides guidelines for interpreting the Part 1 requirement and addresses the recommended (should) approaches in implementing the service management systems. ISO/IEC TR 20000-3 Information Technology - Service Management - Guidance [24] helps in scope definition and applicability of ISO/IEC 20000-1.

There are several factors that might prevent an effective process improvement in the service desk. First, companies usually use external ITIL consultants to provide training for their employees. These consultants know the ITIL framework and IT service management concepts very well but have limited knowledge on the existing business concepts, methods, tools, services, and the structure of service desk groups. Second, inadequate or too complex IT service management tools may slow down any IT service management process improvement initiative.

Third, lack of process culture and process thinking is a very common phenomenon among IT companies. ITIL is a process oriented framework. Thus, the ITIL implementation team should be well-trained and have excellent process improvement and change management skills. Finally, lack of management support for ITSM project may cause an organization to not allocate sufficient resources for the process/tool improvement. Besides allocating enough resources for improvement work, management needs to motivate and reward people who pass the ITIL Foundation certificate exams and who participate in IT service management in their service support and delivery work activities in a consistent and structured manner as guided by the ITIL framework.

Improvement of the processes however represents only one aspect of service desk challenges. Bardhan et al. [25] state that IT services have aspects such as the high degree of involvement by people in delivery and that they are more or less intangible. Therefore, challenges of the service desk extend beyond the process view and include more subjective "people" aspects such as staff motivation and customer experience to generate business value. Simply put the service desk is the "face" of the IT services in an organization. It is required to be functioning effectively in order to raise the profile of IT in businesses. Hence overcoming the challenges of the service desk should be a key driver in service strategy of any organization.

There is a wide number of IT service management studies available. Previous studies on IT service management have put research efforts on success factors in implementing IT service management [26], [27], measuring ITIL implementation projects [28], implementation of service-oriented IT management [29], creating a taxonomy for ITIL processes [30], ITIL implementation maturity models [31], process maturity self-assessment [32], ITIL process integration architecture [33], and prioritization of service incidents [34].

Many studies have explored service management processes, such as improvement of incident management processes based on ITIL practices [35], creating a mature problem management process [36], service testing [37], service level management [38], change and configuration management [39] and release management challenges [40].

Sharifi et al. [41] present causes why ITIL implementations fail with some references to the service desk. They identify the following factors: spending too much time on complicated process diagrams, not creating work instructions, not assigning process owners, concentrating too much on performance, being too ambitious, allowing departmental demarcation and ignoring reviewing of the ITIL. Mohamed et al. [42] have integrated knowledge management elements to the IT service management providing a framework to attain effectiveness, efficiency and innovation during ITIL implementation in organizations.

There are also studies that have dealt directly with help desk activities such as a knowledge management-centric help desk [43]. Bruton [44] has identified ten key steps for managing the IT Help Desk. The service Desk encapsulates the services provided by the Help Desk. We analyzed the steps of Bruton [44] from the overall viewpoint of IT service management best practices. We list the ten steps for managing IT Help Desk from Bruton [44] and alongside provide our interpretation of evaluating the steps in terms of IT service management (in italics) using the ITIL framework with the focus on the service operation phase of the service lifecycle.

- **Know your resources**. Examine your resources to find out what you are capable of delivering and what you are not. Implement ways of improving your skills, equipment and contacts. *IT service management: In ITSM, knowing your resources is important. However, organizations should also take into account the capabilities required to deliver services. Resources and capabilities combine as Service Assets [6]. Resources are "consumed" as direct inputs for service delivery whereas* *capabilities represent the abilities required to manage resources [4]. Therefore, identifying and understanding performance of service assets is an important step for service management in general. Without the knowledge of resources and capabilities, there is no basis to define service value.*

- **Know your customers**. Identify your customers, both users and non-users of your service and list them in order of priority. *IT service management: Differences between customers and users are clarified in the ITSM discipline. Customers and users are both important stakeholders in service management. Customers represent individuals or groups that "purchase" services which includes definition and agreement of service level targets. On the other hand, users "use" the services at an operational level and may be distinct from customers.*
- **Launch your services**. Launch a set of services that meets the majority of customers' needs. Encourage any customer who finds the services inappropriate to consider the service statement as merely a basis for negotiating a special service level agreement with them. *IT service management: Launching services refers to the service transition phase of the service lifecycle. However, planning for the services to be launched should be initiated during the service design phase in the service catalogue management process. Information on launched services should be available in the service catalogue. There may be two levels in a service catalogue: business level (designed for customers) and technical level (designed for service production teams).*
- **Manage the support workflow**. This requires establishment of effective workflow management in the support department. This should cover call management, query prioritization, job allocation, problem escalation and staff motivation. *IT service management: The key to managing a successful service support workflow is planning during the service design stage. Agreed service levels defined during the service level management process should ensure that the service delivery teams are confident in supporting the services. Service Level Agreements should be frequently reviewed to check whether they are consistently and accurately defined so that service level targets are achievable. After good planning, the most challenging areas in managing the incident workflow are incident classification and prioritization.*
- **Ensure good problem closure techniques**. It is not enough to solve the user's technical problem. Establish query closure methods to encompass ensuring customer satisfaction as well as analyzing recently completed jobs to see what lessons learnt can be extracted from how they were handled. *IT service management: In the ITSM, there are separate closing procedures for incidents and problems. In practice, this means that closing a problem may result in the closure of multiple incidents related to that problem. Problem management staff are responsible for maintaining the known error database before closing the problem so that future incidents and problems can be addressed in a more orderly fashion.*
- **Instant workload status reporting**. Set up your reporting routines to provide instant snapshots of the current workload status as well as routine historical information. This will enable you to make decisions about how to deploy your resources. *IT service management: Modern ITSM tools enable a real time view of the resources that have been spent in the incident management and the number of the open/solved cases as well as performance measurement (average incident resolution and response time). Use of such metrics in the service desk function enables efficient resource consumption and optimizes service operations.*
- **Be proactive - take control**. Look for and implement ways of dealing with the workload proactively instead of just reactively. Don't let the workload control your department, you should control the work. *IT service management: Service management should focus on proactive work instead of reactive. This is one of the key benefits of implementing the ITSM model in the service organisation. Proactive actions are especially visible in the problem management process (major problem reviews, trend analysis, defining preventive actions) and in the Continual Service Improvement process (define what to measure, measure data, process data, analyse data and implement improvements continually). These actions ultimately facilitate the service desk to focus on the critical and urgent issues and eliminate the "fire fighting" attitude.*
- **Regular contacts with customers**. Communicate with your customers, through newsletters, workshops, knowledge provision, and most importantly, personal contact - get out there and mix. *IT service management: Communication plays a very important role in service management. The service desk is responsible for communicating incident resolutions, progress of the resolutions, and updated information on new and existing services. Communication can also occur in the form of service reviews, communicating Service Level Agreements and defining Service Level Requirements from the early stages to meet customer's expectations. The ITIL framework stresses consistent communication in every phase of the service lifecycle.*
- **Conduct surveys**. Establish ways of surveying your customers for their views on how the service is being delivered. Use this information to make improvements to your support service and always report your findings and your subsequent actions to your customers. *IT ser-*

vice management: This is consistent with the previous discussion about the value of communication in service management. Service management may involve several types of surveys: surveys after incident resolution or service request fulfillment or periodic customer satisfaction surveys. Besides customer surveys, employee surveys provide useful information on staff perception about service delivery and service bottlenecks. Reviews should be organized after service delivery, for example, after service deployment or after implementation of new or changed services. Survey results have to be analyzed so that it can be used as input for service improvements.

- **Redo all the above every 4 to 6 months**. Do not rest on your laurels no matter how good you are. Go back to step 1 every four to six months and review your whole set-up. *IT service management: Within service management, there is a continual service improvement phase in the service lifecycle based on Deming's Plan-Do-Check-Act cycle [11]. The Continual Service Improvement philosophy in service management encapsulates the entire service lifecycle from strategy to design to transition to service operation. Thus, the basic ideology is that the improvement of services, products, tools and processes should never stop applies in the service desk function as well.*

These ten steps can be used as a generic checklist while implementing or improving service desk activities. The number of academic studies published in the field of IT service management is rapidly increasing. However, many of them focus on presenting success factors or developing new features for the service management tools. More studies are needed to provide information on service desk improvement projects that have used IT Infrastructure Library as an improvement framework because the transition from a traditional help desk to a service-focused service desk typically involves many challenges.

B. Our Contribution

This paper is related to an IT service management research project Keys to IT Service Management and Effective Transition of Services (KISMET) conducted in Finland. The main contribution of this paper is to

- describe the strengths and challenges regarding IT service desk of Finnish Tax Administration,
- present how identified challenges were solved by using ITSM best practices,
- discuss ways to overcome service desk challenges as part of continual service improvement,
- explain how service desk performance and ITSM training usefulness were measured and
- provide lessons learnt from the case study.

The results of this study might be useful for service managers, service desk managers and IT service management process managers.

The remainder of the paper is organized as follows. In Section II, the research methods of this study are described. In Section III, service desk challenges and activities to improve the service desk are presented. Section IV is the analysis of findings. The discussion and the conclusions are given in Section V.

II. Research Problem & Methodology

This case study is a part of the results of the KISMET project. The research problem of this study is: How could an IT service provider's service desk operations be improved by using IT service management best practices? The research problem was divided into the following research questions:

- What types of tools are used by the IT service desk workers?
- How are self-service methods used in the service desk?
- How has the service desk function been organized?
- What types of challenges exist in the IT service provider's customer support?
- How were these challenges solved by using service management best practices from the IT Infrastructure Library?
- What can be learnt based on IT service management training?

According to Yin [45], a case study is "an empirical inquiry that investigates a contemporary phenomenon within its real-life context". Eisenhardt [46] defines case study research as "a research strategy focusing on understanding the dynamics present within single settings". The settings of this paper is the customer support environment of Tax Administration. A case study research method with a single case was used to address the research problem. Figure 1 describes the research settings of the case study. The study was carried out in Finnish Tax Administration's Kuopio unit.

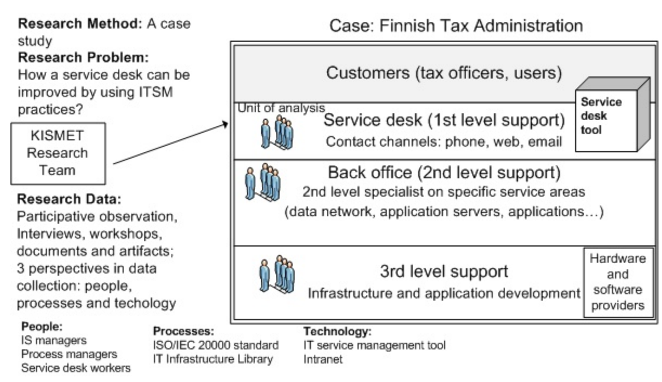


Figure 1. The research settings of the case study

A. The Case Organization and Data Collection Methods

Our case organization is the Information System Management unit of Finnish Tax Administration that provides

IT services (e.g., desktop services, service desk) to the tax administration staff. In Finland, taxation is carried out mainly by four organizations: Ministry of Finance, Tax Administration, Customs Finland and TraFi (Traffic safety agency). Tax Administration provides e-services for tax payers through the Tax.fi website. These services include tax card, revising the tax card, tax return online, notifications of changed bank account details.

In order to demonstrate the complexity of the IT service management, we present a short summary of Tax Administration's 12 organizational units. Each unit has its own special role in Finnish taxation:

- The Individual Taxation Unit (private customers, business owners and self-employed persons)
- the Corporate Taxation Unit (limited companies and corporate customers, customer information and tax control)
- the Tax Collection Unit (payment, collection, recovery and remittance of taxes and the tax account procedure)
- The Tax Auditing Unit (tax auditing activities as a part of tax control, tax audits, EU trade supervision and tax control duties).
- The Joint Services Unit (contact channels, language services ,tax risk management process, coordination of development projects and quality work)
- The IT Services Unit (responsible for application, production and ICT services; directs and oversees the use of ICT)
- The Administrative Unit (HR, financial and general administration tasks),
- The Executive and Legal Unit (Tax Administration's steering and management system),
- The Internal Auditing Unit (internal audits in Tax Administration)
- Communications Unit (communications and marketing)
- The Grey Economy Investigation Unit (producing and distributing information on the grey economy and action against it)
- The Tax Recipients' Legal Services Unit (the rights of tax recipients in taxation matters)

The Tax Administration organization had 5.300 full-time employees in 2011. The number of employees working in IT user support was 70. The organization used a phased approach for implementing service management processes. In the first phase, the focus was on incident management and the ITIL-based service desk service was launched in Spring 2011.

The case study was carried out in August 2011 - March 2012. In order to increase the quality of the case study, researchers used three important principles of data collection: 1) using multiple sources of evidence: three researchers participated in data collection from several sources 2) creating a case study datastore (a case study diary) 3) maintaining a chain of evidence (linking observations to data sources). The following sources of evidence were used:

- Documentation from the case organization (e.g., incident management process description, service support metrics, ITSM tool user guide, service catalogue, service area, event management material, error handling guide).
- Archives (service classification schema, incident and service request records)
- Interviews/discussions (discussions in work meetings between a research team and the case organization, informal coffee table discussions with service support workers, email conversations with service managers, a focused interview on self service methods with an ITSM tool administrator)
- Participative observation (process improvement meetings and workshops (CSI workshop 27 September) and ITSM process training (45 minutes ITSM Introduction, 3 hour Basics of ITSM, ca. 70 participants) organized by the KISMET research team
- Physical artifacts: Service desk tool, intranet

B. *Data Analysis Method*

There are two main techniques to analyze case study data: a case comparison technique and a within-case technique. A within case analysis technique [46] was used in this study to analyze the collected case study data. The within-case analysis focuses on examining each case carefully as a standalone entity before making any generalizations. Additionally, researcher triangulation was used in data analysis. Three case study researchers participated in the data collection and analysis to obtain a richer view on the case organization's behavior. Instead of a formal database, we used Windows folder as a datastore for documents, memos, records and other material that we received from the case organization. The case study findings were validated in weekly meetings with the case organization. At the end of the research period, a case study report was sent to the case organization for validation.

III. Towards an Improved IT Service Desk

We used KISMET (Keys to IT Service Management Excellence Technique) model as a process improvement tool. The model consists of the following seven phases: Create a process improvement infrastructure, Perform a process assessment, Plan process improvement actions, Improve/Implement the process based on ITSM practices, Deploy and introduce the process, Evaluate process improvement and Continuous process improvement. In this paper, we focus on the 'Perform a process assessment' and 'Process Improvement' phases.

A. Perform a Process Assessment

The process assessment focused on observing the current state of the service desk including process goals, inputs, outputs, roles and responsibilities, activities, metrics, and relationships to other processes. The assessement was not carried out as a formal process assessment but was based on ITIL v2 and v3 best practices and ISO 20000-1 requirements for incident management. The assessment was carried out by three ITSM researchers.

Strengths: During the study, we observed the following strengths regarding the service desk and customer support from an IT service management viewpoint:

- Many processes were described by using activity diagrams. Thus, it was easy to get overview of support practices.
- The organization had created an Incident Management process description (including inputs and outputs and process activities)
- The organization had a strong focus on continuous improvement of services (the performance of the service desk was measured with several metrics, frequent surveys were carried out to collect feedback both from customers and staff concerning service desk operations).
- Management support and commitment for improvíng IT service management was highly visible in the organization
- The selected service desk tool supported IT service management principles and a tool administration team had skills and was very motivated to improve the service desk tool based on service management practices.
- Service desk employees were really interested in receiving ITSM training. Some people who were not able to participate in ITSM training days asked whether they could get a short summary or personal training.

Potential improvement areas: The following improvement areas regarding the service desk and customer support were identified. These challenges (P= Process-related challenge, T= Tool-related challenge) should not be considered as weaknesses but as potential ways to make customer support more effective and service-focused.

- **Challenge:** Classification of support requests in the service desk requires clarification. (T) **Improvement suggestion:** Clarify the options in 'Reason for Contact Request' field of the incident record in the service desk tool. Make the difference between service requests and incidents visible. Service area and the type of support requests should be separate fields. Collect concrete examples of both incidents and service request for training purposes.
- **Challenge:** Customers are not able to classify support requests correctly. (T) **Improvement suggestion:** Remove the classification option from customers and simplify the submission of support requests.
- **Challenge:** It is difficult to identify repeating incidents from the service desk system (T). **Improvement suggestion:** Mark the repeating incidents (for example, create an additional 'check box' type data field to an incident record: Repeating incident = x). Use the 'Relate Cases' function to establish relationships between similar cases. Create a problem record based on a repeating incident.
- **Challenge:** The interface between incident management and problem management does not work. People do not understand the difference between incidents and problems. (P) **Improvement suggestion:** Train employees to open a problem record. Establish a simple-to-understand guidelines for problem management including triggers for problem management.
- **Challenge:** Service desk workers record several cases under one incident. (P) **Improvement suggestion:** Train service desk workers to record cases in such a way that one incident record includes only one issue.
- **Challenge:** Improvement ideas are not recorded systematically into the service desk system. (P) **Improvement suggestion:** Improvement ideas should be sent to a Continual Service Improvement team or Change Management team.
- **Challenge:** Lack of a formal Configuration Management Database (CMDB). (P, T) **Improvement suggestion:** Establish a Configuration Management process that is responsible for updating, maintaining and managing a Configuration Management Database (CMDB).

Customer support tools: Based on the interview with a service desk tool administrator (who also held a supervisor role in the Operations Bridge (OB); OB is responsible for observation and monitoring of the IT Infrastructure), we identified the following tools and applications that were used by customer support teams.

- A service desk tool (ITIL-compliant IT service management tool)
- A virtual phone system
- Screenshot application
- Office guidebook
- Information Board (a website of known problems)

A participative observation carried out by the research team revealed that besides the IT service management tool, a large amount of customer support information had been stored on the intranet sites. We found information on configuration items, change requests and problems and errors. This information was valuable for researchers because it showed how customer support practices are performed without ITIL. We expected that it might be difficult to persuade employees to give up the intranet based practices and use the IT service management tool instead.

Self-service methods: Based on interviews, we found out that self service methods in the service desk included a knowledge base; Information Board (a website on known problems); customer portal that enabled 24/7 submission of incidents and service requests, as well as monitoring the progress of customer's own incidents and service requests. The information stored by customer support in the knowledge base includes short tips, advice, workarounds, and operational guidance (in the service desk's internal use).

B. Improving the IT Service Desk: Tools and Processes

This phase started when the research team asked the case organization which challenges they consider most important. In this section, we discuss how these challenges were solved. We focused the research efforts especially on four issues: improving the classification, defining the interface between incident and problem management, raising the awareness of IT service management through training, and improving the measurement of the service desk.

Because the time for research was limited, the process improvement started very soon from improving the classification. Based on discussions with IS management, a tool administrator, service support engineers and observations by researchers, the research team decided that support request classification required changes. There was a 'Reason for contact' field in the support request record. Overlapping categories (failure, error) were removed or renamed in the categories of the 'Reason for contact' field . The new categories included: incident (failure or error), request (feedback), request (improvement idea), request (order), request (other service request), request (advice), request (information/notification).

Additionally, the case organization's representatives asked whether the research team could improve the service area definitions. It was a challenge for a research team to explore the service areas and classify them into three tiers, because the number of service areas was large and researchers did not know the exact role and purpose of some service areas. The improved category tree was submitted to the case organization for validation. As an end result, a new service category tree was established (see Figure 2).

The second improvement area was related to defining the interface between incident and problem management processes. In the meeting (March 8th), there was a discussion on why incident management and problem management have to be separate processes. Two main benefits from having separate processes were identified. First, incident management can resolve cases faster by using workarounds from problem management. Second, problem management enables the detailed investigation of the case although the original incident has been resolved. The research team created a simple procedure for problem management. The procedure addressed how to use the ITSM tool to support the problem management process. The workflow from incidents and problems was explored using the scenario-based

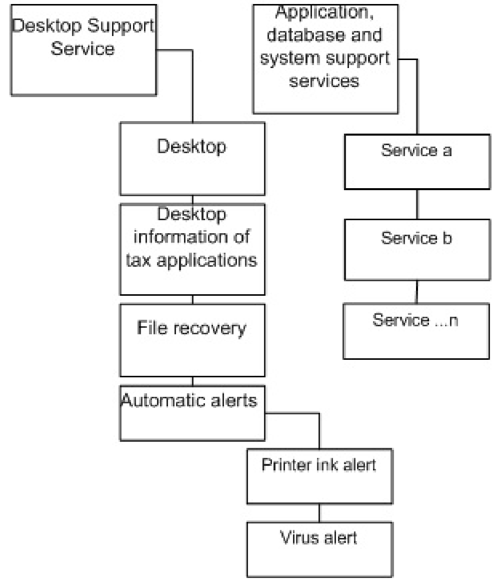


Figure 2. A part of a service category tree

technique. We selected four real customer support cases, analyzed the workflow, and compared the handling procedure to an ITIL-based handling procedure. The improvement work also included the configuration and customization of a problem record (see Figure 3).

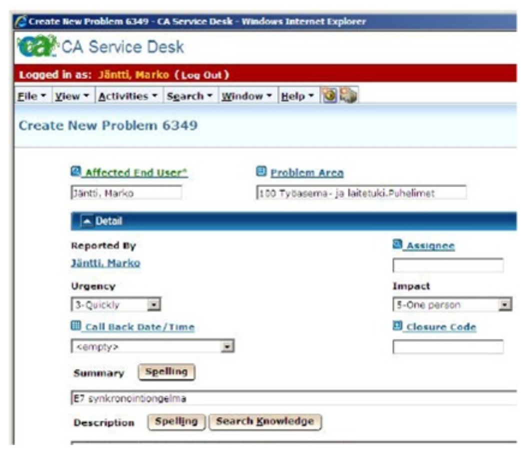


Figure 3. A problem ticket

In the **Problem Management workshop** (February 21-22, 2012), the case organization decided to establish a problem management group that has responsibility to

- search cases that are potential problems
- model the problem management process in the User support services unit
- store essential configuration items to the service desk tool
- record known errors to the knowledge base.

The research team organized ITSM training in the case organization in seven different locations. The following feedback (questions and comments) were collected from IT service management training sessions (Sep. 13th, Sep. 19th, Sep. 20th, Sep. 21th, Oct. 5th, Oct. 7th):

- “What is the current status of the service desk tool and the incident management process?”
- “How support requests should be classified and ITSM concepts integrated to the service desk?”
- “Which interfaces do service desk and incident management have?”
- “How IT service management training for service desk people should be conducted?”
- “What does continual service improvement mean for service desk?”
- “How do other organizations use ITIL?”
- “Is it possible to enable parallel handling of work tasks regarding one incident?”
- “How can incidents be connected to each other?”
- “It was difficult to adopt English ITIL terms, but after training, terms are much more clearer”.
- “Service desk should see the big picture of customer support, for example, whether the issue is related to some existing problem?”
- “We have thought that process frameworks are only for managers. Now, it looks that they are coming to the field, too”.
- “We should think about sending people to ITIL certification courses”.
- “Support requests are complicated and manysided. Many service desk workers log cases incorrectly as problems if they can choose ’problem’ from the contact type list”.
- “ITILv2 looks like a blueprint of a French nuclear plant, it was really good that you (trainer) showed it in the end of the presentation”.
- “Not every support group uses automatic alerts for support cases. It may happen that they do not know that they have received an incident or a problem”.
- “The large number of customer support teams is a problem (routing cases between many teams increases the likelihood that the support case will be lost)”.

After training, we collected feedback from employees how useful they considered IT service management training. Regarding the role of respondents, 14 were IT support engineers, 12 were designers, and 2 were management. The average of responses was 3.46 on a scale of 1 - not valuable at all, to 5 - very useful.

On March 12, the research team held an **IT service management measurement workshop** with the tool administrator to show the incident/service request listing in MS Excel format. The ratio between incidents and service requests was 114/304. The goal was to identify the differences between service requests and incidents. The tool administrator reported that some cases include both the description of failure but also a request for more detailed investigation of the case. It was agreed that unclear cases shall be recorded as service requests. The research team had prepared Critical Success Factors, Key Performance Indicators and Metrics. The following ITIL metrics were selected as potential process metrics for the incident management process:

Table I
FEEDBACK FROM ITSM TRAINING: HOW USEFUL WAS THE TRAINING? (N=28)

Grade	Description	Responses
1	Not valuable at all	1
2	Less than useful	2
3	Somewhat useful	12
4	Quite useful	9
5	Very useful	4

- Number of opened incidents
- Incident resolution time
- Incidents resolved within SLA targets
- Number of open incidents
- Call response time
- Customer satisfaction on resolved incidents
- Number of opened service requests

The measurement was considered an important part of Continual Service Improvement. The research team also studied whether service level agreements were used in the organization. We observed that there were no SLAs related to incidents in the operation. Another thing we observed was that prioritization of a case was not automatically based on urgency and impact of a case. This is one of the key requirements for incident management. The IT management listed the Service Level Management as a future improvement target.

IV. ANALYSIS

In this section, the analysis of results is presented. Results were categorized according to research questions. In order to maintain the chain of evidence, we describe the source for findings in parentheses (AR= Archives and records, D= Documentation, ID= Interviews and discussions, O= Observation, PA= Physical artefacts, ST= Seminars and training organized by the research group). Lessons learnt are presented after the analysis of each research question.

RQ1: What types of tools are used by the IT service desk workers? (PA, O, ID) Regarding the service desk, we observed during the study that the service desk tool supports the implementation of IT service management well. The tool was called a service desk tool although it might have also supported many other support processes, because it was compatible with the concepts of the ITIL framework. **Lesson 1:** Although the improvement focus would be incident management, take other processes into account while defining the records in the ITSM tool. The self-service interviews and participative observation revealed the usage of tools, such as a virtual phone system, screen-shot application, office guidebook, and information board. The information board was related to problem management because it seemed to store information on known problems. One important observation we made was that there were many datafields in the service desk tool that can be reused between different records (incidents, problems, requests for change) if carefully designed.

RQ2: How are self-service methods used in the service desk? (ID, AR) The organization had invested in automatizing the handling of service requests and electronic forms were well exploited in service request management. Many customer support workers were interested in using the knowledge base module of the IT service desk tool. This led us to define **Lesson 2:** Market self-service tools, such as a knowledge base, not only to customers and users but also to service desk workers. A well-designed and maintained knowledge base available to the service desk can remarkably decrease the incident resolution time. Typically, a self-service portal allowed customers to check the status of their support request and submit new incidents and service requests 24/7. Self-service tools are often considered too strongly as a customers' tool.

RQ3: How has the service desk function been organized? (O, I, D) The service desk and customer support were organized into three levels: Level 1 was responsible for resolving simple requests and incidents and routing more complicated requests to Level 2 that provided more advanced support, such application support for Tax applications etc.). Level 3 consisted of application development and hardware and software vendors as external stakeholders.

Lesson 3: Establish a group that promotes the problem management process and practices in the organization. Regarding incident management roles, we observed that roles and responsibilities and incident management activities were defined in the process description and followed the IT service management terminology. However, we noticed that there was no responsible person or a group for problem management. The organization had assigned a well functioning team (2-3 persons) for configuring the IT service management tool. Detailed process descriptions had been created for ITSM processes. Regarding the results, it has to be mentioned that the organization had focused on service desk and incident management in the first phase of the process improvement cycle (starting from Spring 2011). Therefore, the remaining ITSM processes, such as change management and problem management, were immature.

RQ4: What types of challenges exist in the IT service provider's customer support? (AR, PA, O, ID, D, ST) Many of the challenges seemed to be related to classification of support requests. Service desk workers indicated both in training and in discussions that users and customers have problems in classifying requests. However, nobody had measured the number of incorrectly classified requests.

Lesson 4: Train customer support workers in the early phase to use the 'problem' concept correctly and to identify problems. Based on the knowledge from our previous studies, it was not a surprise that the organization experienced difficulties with problem management. This challenge has been noted also in our previous studies. Some service desk workers had classified a case as a "problem" when the problem was one of the options in Reason for Contact field. Crucial for problem management would be recording information on which incidents are repeating incidents and thus sources of problems. Discussions in training also indicated that some service desk teams were interested in identifying and flagging repeating incidents.

Lesson 5: Establish clear escalation rules and automatic email alerts on new support cases. Comments from ITSM training and participative observation also revealed that incident escalation and routing involves challenges, for example, all teams do not use automatic alerts that send a message to email that they have received a new case. There was also a need to parallel handling of incident tasks in order to accelerate the resolution process.

RQ5: How were customer support challenges solved using service management best practices from the IT Infrastructure Library? (AR, PA, O, ID, D, ST) Creating an effective measurement framework for IT service management is a difficult task. Measurement is visible especially in the ITIL Continual Service Improvement (CSI) book that provides guidance on measuring IT services.

Lesson 6: Establish a CSI process and define the relationship between Measurement activity, Reporting activity and Managing Improvement Ideas activity. Customizing the ITIL CSI is not an easy task because the ITIL CSI does not provide a clear process description or a process diagram. CSI consists of three main areas: measurement, reporting and management of improvement ideas. We organized two workshops related to measurements: a CSI workshop that increased the employee's awareness of CSI practices and a measurement workshop that focused on selecting appropriate process metrics. The problem with some ITIL based metrics was that they were described with complex terminology. There were some metrics the implementation of which would have needed changes to the service desk tool, such as the number of incidents reopened and as a percentage of the total. Regarding the service desk, we observed that there were over 10 metrics. Additionally, we observed the case organization needed a model of how to handle service improvement ideas in a more systematic way. One possibility would be to assign development ideas to the change management team that would open a Request for Change.

Lesson 7: Separate configuration item, the contact request type, and the service area from each other in order to enable reuse of categories. After examining the service desk tool and incident classification rules the research team suggested to the case organization that the service area requires changes and that the using the service desk system should be as simple as possible for customers. Our solution was to separate a configuration item, the contact request type, and the service area from each other. This solution enabled reusing of service area and configuration items in other records, such as in Request for Change and Problem records.

Lesson 8: Start building the CMDB from the most frequently used configuration items, such as applications, operating systems, and desktops. While investigating the incident classification in the service desk tool, we observed that the interface between incident management and configuration management requires clarification. This led us to define a preliminary model of how to store information on configuration items (CIs). We created a table with three columns: CI Family, CI Class and CI Name and started from application CIs. The result was Software Family, 8 CI application classes (for example, Office application, administration application, access management application) and CI names (MS Word). We continued the validation of the CI model by selecting incidents randomly from the service desk tool and exploring which CI the incident is related to and whether the CI is defined in our model.

RQ6: What can be learnt from IT service management training (ST) Service management training serves two main purposes. First, it helps managers to decrease the employees' change resistance to new IT service management terms and processes "...It was difficult to adopt English ITIL terms, but after training, terms are much more clearer...". Additionally, organizing continuous ITSM training is the only way to ensure that ITSM practices are really adopted by the organisation's employees. **Lesson 9:** ITSM training helps to decrease change resistance to ITSM. We observed that after training, some employees seemed to be ready to recommend ITSM principles to their colleagues and were interested in receiving more ITSM training. Training increases the employees' awareness on the objectives and benefits of ITSM implementation and may act as a reminder that employees are also responsible for service management "...We have thought that process frameworks are only for managers. Now, it looks like they are coming to the field, too ...". Employees participate more in the training discussions when they find their own position in the ITSM framework. Usually, service desk employees discuss more in ITIL training than programmers, but why? Is the reason that service desk and incident management are more visible in ITIL than application development? A rival explanation would be that service desk workers are more social than programmers and tend to speak more than programmers.

Second, IT service management training is a valuable source of improvement ideas. This leads to our final lesson **Lesson 10:** ITSM training helps to identify improvement targets. Questions and comments during ITSM training often reveal bottlenecks, challenges and unclear ITIL issues. For example, in our case, ITSM training provided valuable information on improvement areas, such as interfaces between the service desk and other ITSM processes, for example, between incident management and configuration management, change management, and event management. We learnt that face-to-face training resulted in much more discussion compared to remote training, and observed that small training groups (less than 10 people in a group) work very well in Awareness training because people are not afraid of asking questions.

V. Discussion and Conclusion

The service desk plays a very important role in IT service management because it is responsible for daily communication with customers and users. In the IT Infrastructure Library, the service desk is a function that performs the incident management process. This paper aimed to answer the research problem: How could an IT service provider's service desk operations be improved by using IT service management best practices? The main contribution of this study was to explore 1) service desk tools, 2) self-service methods, 3) structure of the service desk function, 4) challenges related to the service desk and customer support and 5) solutions to these challenges based on ITSM practices.

The case organization Finnish Tax Administration was a representative case of a government agency that is implementing service management based on ITIL. The key challenges we identified in service desk operations were related to classification of support requests both from the service desk workers' viewpoint and customers' viewpoint, understanding the differences between incident and problem management processes, and identifying the sources of problems and interfaces between IT service management processes.

The study also presented solutions to identified challenges. Regarding classification of requests and incidents, an improved service category tree was established and categories of the Reason for Contact field in the incident record were clarified. Regarding the interface between incident and problem management, a customized problem record was created, a simple problem management procedure created, and the problem management group created. The improvements on ITSM measurement practices included selection of incident management metrics and two workshops (CSI workshop and Measurement Workshop). Additionally, ITSM training was organized for the case organization's employees. Feedback from training was presented in this paper as well as comments and questions that participants asked during training. Data were analyzed according to the research questions and lessons learnt from researchers' viewpoint were presented as a part of the analysis.

This case study included certain limitations. First, regarding internal validity, data were collected using qualitative research methods. Quantitative methods, such as a customer satisfaction survey on service desk could have provided a richer view on the organization. However, the qualitative case study method is well-suited to research business process-related challenges in organizational context. Additionally, we used a rich set of data sources and three researchers to build a detailed view of the organization and its customer support.

Second, concerning case selection and external validity, the case organization was a partner of the software engineering unit's research project. Thus, convenience sampling was used as a sampling strategy. It is a generally accepted way to obtain case organizations. Third, this study included one service area of the case organization. A larger number of cases and comparison between them based on the preselected categories would have increased the quality of case study. This study can act, for example, as a pilot case study for further studies. During the study we identified several targets for further research. Further research could explore, for example, the role of social media in service desks, impact of service desk improvements on the business, or IT service management process assessment tools.

Acknowledgment

This paper is based on research in KISMET project funded by the National Technology Agency TEKES (no. 70035/10), European Regional Development Fund (ERDF), and industrial partners. We would also like to acknowledge research contribution from University of Southern Queensland, Australia and Griffith University, Australia in this study.

References

[1] M. Jäntti, "Examining challenges in it service desk system and processes: A case study," in *Proceedings of the 7th International Conference on Systems 2012 (ICONS2012)*. CPS Publishing, 2012, pp. 105–108.

[2] J. Peppard, "Managing it as a portfolio of services," *European Management Journal*, vol. 21, no. 4, pp. 467–483, August 2003.

[3] C. Shaw and J. Ivens, *Building Great Customer Experiences*. Palgrave-Macmillan, 2005.

[4] P. Bernard, *Foundations of ITIL*. Zaltbommel, Netherlands: Van-Haren Publishing, 2012.

[5] L.-J. Zhang, J. Zhang, and H. Cai, *Services Computing*. Tsinghua University Press, Beijing and Springer-Verlag GmbH Berlin Heidelberg, 2007.

[6] Office of Government Commerce, *ITIL Service Delivery*. The Stationary Office, UK, 2002.

[7] Office of Government Commerce(a), *ITIL Service Strategy*. The Stationary Office, UK, 2007.

[8] Office of Government Commerce(b), *ITIL Service Design*. The Stationary Office, UK, 2007.

[9] Office of Government Commerce(c), *ITIL Service Transition*. The Stationary Office, UK, 2007.

[10] Office of Government Commerce(d), *ITIL Service Operation*. The Stationary Office, UK, 2007.

[11] Office of Government Commerce(e), *ITIL Continual Service Improvement*. The Stationary Office, UK, 2007.

[12] COBIT 5, *Control Objectives for Information and related Technology: COBIT 5*. ISACA, 2012.

[13] Microsoft, "Microsoft operations framework," http://technet.microsoft.com/en-us/library/cc506049.aspx, December 2012.

[14] V. Kapella, "A framework for incident and problem management," International Network Services whitepaper, 2003.

[15] F. Niessinka, V. Clerca, T. Tijdinka, and H. van Vliet, "The it service capability maturity model version 1.0," CIBIT Consultants&Vrije Universiteit, 2005.

[16] ISO/IEC 20000, *IT Service Management, Part 1: Specification for service management*. ISO/IEC JTC1/SC7 Secretariat, 2005.

[17] B. Barafort, V. Betry, S. Cortina, M. Picard, M. St-Jean, A. Renault, and O. Valdes, *ITSM Process Assessment Supporting ITIL*. Zaltbommel, Van Haren Publishing, 2009.

[18] CMMI, *Standard CMMI Appraisal Method for Process Improvement (SCAMPISM) A, Version 1.3: Method Definition Document*. USA: Software Engineering Institute, Carnegie Mellon University, 2011.

[19] A. Lahtela, M. Jäntti, and J. Kaukola, "Implementing an itil-based it service management measurement system," in *Proceedings of the 4th International Conference on Digital Society*. St. Maarten, Netherlands Antilles: IEEE Computer Society, February 2010, pp. 249–254.

[20] ISO/IEC, *TR 20000-4:2010 - Information technology - Service management - Part 4: Process Reference Model*. Geneva, Switzerland, International Organization for Standardization, 2010.

[21] A. Mesquida, A. Mas, E. Amengual, and J. Calvo-Manzano, "It service management process improvement based on iso/iec15504: A systematic review," *Information and Software Technology*, vol. 54, no. 3, 2012.

[22] ISO/IEC 2004, *ISO/IEC 15504-2:2004 - Information technology - Process Assessment - Part 2: Performing an Assessment*. Geneva, Switzerland, International Organization for Standardization, 2004.

[23] ISO/IEC TS 15504-8:2012(E), *Information technology - Process assessment -Part 8: An exemplar process assessment model for IT service management*. ISO/IEC TC JTC1/SC7 Secretariat, 2012.

[24] ISO/IEC, *ISO/IEC TR 20000-3 Information Technology - Service Management - Guidance on scope definition and applicability of ISO/IEC 20000-1*. ISO/IEC JTC1/SC7 Secretariat, 2010.

[25] I. Bardhan, H. Demirkan, P. Kannan, R. Kauffman, and R. Sougstad, "An interdisciplinary perspective on it services management and service science," *Journal of Management Information Systems*, vol. 26, no. 4, pp. 13–64, 2010.

[26] W.-G. Tan, A. Cater-Steel, and M. Toleman, "Implementing it service management: A case study focussing on critical success factors," *Journal of Computer Information Systems*, vol. 50, no. 2, 2009.

[27] C. Pollard and A. Cater-Steel, "Justifications, strategies, and critical success factors in successful itil implementations in u.s. and australian companies: An exploratory study," *Information Systems Management*, vol. 26, no. 2, pp. 164–175, 2009.

[28] J. Iden and T. R. Eikebrokk, "Understanding the itil implementation project: Conceptualization and measurements," in *Proceedings of 2011 22nd International Workshop on Database and Expert Systems Applications*. Washington, DC, USA: IEEE, 2011.

[29] A. Hochstein, R. Zarnekow, and W. Brenner, "Itil as common practice reference model for it service management: Formal assessment and implications for practice," in *EEE '05: Proceedings of the 2005 IEEE International Conference on e-Technology, e-Commerce and e-Service (EEE'05) on e-Technology, e-Commerce and e-Service*. Washington, DC, USA: IEEE Computer Society, 2005, pp. 704–710.

[30] M. Brenner, "Classifying itil processes; a taxonomy under tool support aspects," in *Business-Driven IT Management, 2006. BDIM '06. The First IEEE/IFIP International Workshop on*, april 2006, pp. 19–28.

[31] R. de Sousa Pereira and M. da Silva, "A maturity model for implementing itil v3," in *Services (SERVICES-1), 2010 6th World Congress on*, july 2010, pp. 399 –406.

[32] R. Meziani and I. Saleh, "e-government: Itil-based service management case study," in *Proceedings of the 12th International Conference on Information Integration and Web-based Applications & Services*, ser. iiWAS '10. New York, NY, USA: ACM, 2010, pp. 509–516.

[33] S. Zhang, Z. Ding, and Y. Zong, "Itil process integration in the context of organization environment," in *Computer Science and Information Engineering, 2009 WRI World Congress on*, vol. 7, 31 2009-april 2 2009, pp. 682–686.

[34] J. Barlow and P. Stewart, *Branded customer service: the new competitive edge*. United States of America: Berret-Koehler Publishers, Inc., 2004.

[35] M. Jäntti, "Lessons learnt from the improvement of customer support processes: A case study on incident management," in *Product-Focused Software Process Improvement*, ser. Lecture Notes in Business Information Processing, W. Aalst, J. Mylopoulos, N. M. Sadeh, M. J. Shaw, C. Szyperski, F. Bomarius, M. Oivo, P. Jaring, and P. Abrahamsson, Eds. Springer Berlin Heidelberg, 2009, vol. 32, pp. 317–331.

[36] M. Kajko-Mattsson, "Problem management maturity within corrective maintenance," *Journal of Software Maintenance*, vol. 14, no. 3, pp. 197–227, 2002.

[37] M. Jantti and T. Kujala, "Exploring a testing during maintenance process from it service provider's perspective," in *Information Science and Service Science (NISS), 2011 5th International Conference on New Trends in*, vol. 2, oct. 2011, pp. 318–323.

[38] M. Kajko-Mattsson, C. Ahnlund, and E. Lundberg, "Cm3: Service level agreement," in *ICSM '04: Proceedings of the 20th IEEE International Conference on Software Maintenance*. Washington, DC, USA: IEEE Computer Society, 2004, pp. 432–436.

[39] C. Ward, V. Aggarwal, M. Buco, E. Olsson, and S. Weinberger, "Integrated change and configuration management," *IBM Syst. J.*, vol. 46, pp. 459–478, December 2006.

[40] H.-M. Sihvonen and M. Jantti, "Improving release and patch management processes: An empirical case study on process challenges," *Proceedings of the International Conference on Software Engineering Advances (ICSEA 2009)*, vol. 0, pp. 232–237, 2010.

[41] M. Sharifi, M. Ayat, A. A. Rahman, and S. Sahibudin, "Lessons learned in itil implementation failure," in *Information Technology, 2008. ITSim 2008. International Symposium*, vol. 1, aug. 2008, pp. 1–4.

[42] M. Mohamed, V. Ribiere, K. O'Sullivan, and M. Mohamed, "The re-structuring of the information technology infrastructure library (itil) implementation using knowledge management framework," *The Journal of Information and Knowledge Management Systems*, vol. 38, no. 3, pp. 315–333, 2008.

[43] L. M. Gonzalez, R. E. Giachetti, and G. Ramirez, "Knowledge management-centric help desk: specification and performance evaluation," *Decis. Support Syst.*, vol. 40, no. 2, pp. 389–405, 2005.

[44] N. Bruton, *How to manage effective IT Help Desk*. Butterworth Heinemann, 2002.

[45] R. Yin, *Case Study Research: Design and Methods*. Beverly Hills, CA: Sage Publishing, 1994.

[46] K. Eisenhardt, "Building theories from case study research," *Academy of Management Review*, vol. 14, pp. 532–550, 1989.

On the Generalization of Normalized Systems Concepts to the Analysis and Design of Modules in Systems and Enterprise Engineering

Peter De Bruyn and Herwig Mannaert
Department of Management Information Systems
Normalized Systems Institute (NSI)
University of Antwerp
Antwerp, Belgium
{*peter.debruyn, herwig.mannaert*}*@ua.ac.be*

***Abstract*—Current organizations need to be able to cope with challenges such as increasing change and increasing complexity in many or all of their aspects. Modularity has frequently been suggested as a powerful means for reducing that complexity and enabling flexibility. However, the proper use of modularity to actually achieve those benefits cannot be considered trivial or straightforward. Normalized Systems (NS) theory has proven to introduce this evolvable and diagnosable modularity in software systems. This paper discusses the generalization of NS concepts to the analysis and design of modules in systems and enterprise engineering as evolvability and diagnosability are deemed to be appealing for most modular structures. In order to do so, this paper highlights the importance of distinguishing blackbox and whitebox views on systems and the fact that a true blackbox requires fully and exhaustively defined interfaces. We further discuss the functional/constructional transformation and elaborate on how NS theory uses the concepts of modularity, stability and entropy to optimize certain properties of that transformation. We argue how some aspects of organizational systems can be analyzed based on the same reasoning, suggesting some viable approaches for Enterprise Engineering. By means of a tentative reflection, we provide a discussion regarding how the concepts of stability and entropy might be interpreted as different manifestations of coupling within modular structures.**

***Keywords*-Normalized Systems; Systems Engineering; Enterprise Engineering; Modularity; Stability; Entropy.**

I. Introduction

Current organizations need to be able to cope with increasing change and increasing complexity in many or all of their aspects. Not only organizations themselves need to deal with this '*changing complexity*' in terms of their organizational structures, business processes, etcetera. Additionally, also the products or services they deliver, and even the software applications supporting these products and business processes, are equally exposed to this changing complexity. In many engineering disciplines, *modularity* has previously been suggested as a powerful means for reducing that complexity by decomposing a system into several subsystems [2], [3]. Moreover, modifications at the level of those subsystems instead of the system as a whole are said to facilitate the overall evolvability of the system. Hence, modularity can be claimed to have properties for tackling both change and complexity in systems.

However, the proper use of modularity to actually achieve those benefits cannot be considered trivial or straightforward. Typical issues involved include the identification and delineation of the modular building blocks (i.e., subsystems), the communication between those modular building blocks (i.e., the interfaces), the assurance of providing compatibility of new versions of a building block with the overall system, etcetera.

In this regard, *Normalized Systems (NS) theory* has recently proven to introduce this diagnosable and evolvable modularity, primarily at the level of software systems [4], [5]. Considering the transformation of basic functional requirements into software primitives (such as data structures, functions, methods, etcetera), which are considered as the basic modular building blocks of software systems, the theory proposes a set of formally proven theorems to design and analyze software architectures. Besides using the concept of modularity, the theory also heavily relies on other traditional engineering concepts such as stability (based on systems theory) [4] and entropy (based on thermodynamics) [6] to optimize those modular structures according to certain criteria. As this approach has proven its value in the past to obtain more maintainable and diagnosable software architectures, it seems appealing to investigate the extent to which we can apply the same approach to other modular systems for which these properties seem beneficial as well.

Indeed, we claim that many other systems could also be regarded as modular structures. Both functional (i.e., requirements) and constructional (i.e., primitives) perspectives can frequently be discerned, modules can be identified and thus the analysis of the functional/constructional transformation seems relevant. For instance, considering organizational systems, Van Nuffel has recently shown the feasibility of applying modularity and NS theory concepts at the business process level [7], [8] while Huysmans did so at the level of enterprise architectures [9]. However, the extension of NS theory to these domains has not been fully formalized yet in several aspects. Consequently, as NS theory proved to be successful in introducing evolvable and diagnosable modularity in software systems, and as it is clearly desirable to extend such properties to other systems, this paper focuses on applying NS theory concepts to the identification and analysis of modular structures in systems engineering (including Enterprise Engineering). This way, the paper can be regarded as a first step towards the formal generalization of NS theory concepts to modularity and system engineering in general.

More specifically, we will focus in the present paper on the illustration and discussion of the following aspects:

1) From an engineering perspective, arguably, many systems can be considered as *modular structures*, including traditional engineering systems, software systems, and organizational systems;
2) In order to profoundly study (and in a second phase, optimize) the structure of such modular systems, it is necessary to formulate *complete and exhaustive interfaces* for each of their constituting modular building blocks as this gives a complete overview of the coupling between them (and, possibly, external systems). This is considered to be an essential part of each systems engineering process;
3) As proposed by Normalized Systems theory, traditional engineering concepts such as *stability* (based on systems theory) and *entropy* (based on thermodynamics) might offer interesting viewpoints for the analysis and optimization of modular systems, each from their own perspective (i.e., evolvability and diagnosability respectively);
4) One way of interpreting both the occurrence of instability and entropy in relation to modular structures, is to consider *coupling* between modular components as their common origin and ground. This further adds to our discussion regarding the importance of fully defined and complete interfaces, as argued under bullet point 2.

This paper mainly further elaborates the reasoning proposed in [1] regarding the need for complete and unambiguous module interfaces (i.e., bullet points 1 and 2) by explicitly relating them to the concepts of coupling, stability and entropy (i.e., bullet points 3 and 4). Also, additional examples (i.e., cases) regarding the consequences of applying our reasoning to organizational systems, will be provided.

The remainder of this paper is structured as follows. Section II discusses the essence of NS theory and how it leverages the concepts of stability and entropy to obtain evolvable and diagnosable modular structures. Next, in Section III, we present some extant literature on modularity (without claiming to be exhaustive), emphasizing the work of Baldwin and Clark. Here, some arguments will also be offered to consider it reasonable to analyze both software and organizational systems from a modularity point of view. Afterwards, we differentiate between blackbox (functional) and whitebox (constructional) views on systems, and offer a more unambiguous definition of modularity by arguing for the need of complete and exhaustively defined interfaces in Section IV. Some useful functional/constructional transformation properties (including stability and entropy) will be discussed in Section V. Emphasizing the usefulness of our approach for Enterprise Engineering, the application of NS stability and entropy reasoning to business processes will be illustrated in Section VI, as well as some additional interface dimensions, which could show up when considering organizational modules (ideally exhibiting a complete interface). In Section VII, some additional examples (i.e., cases) regarding the consequences of applying our reasoning to organizational systems, will be provided. We end this paper by reflecting on the relatedness of the concepts of stability and entropy in terms of modular coupling (Section VIII) and some conclusions (Section IX).

II. Normalized Systems theory

Normalized Systems theory (NS) is a theory about the deterministic creation of evolvable modular structures based on a limited set of proven and unambiguous design theorems, primarily aimed at the design of evolvable software architectures. In order to do so, the theory states that the implementation of functional requirements into software constructs can be regarded as a *transformation* of a set of requirements $\mathcal{R}$ into a set of software primitives $\mathcal{S}$ [10], [4], [5]:

$$\{\mathcal{S}\} = \mathcal{I}\{\mathcal{R}\}$$

Next, the theory argues that the resulting set of primitives can be considered to be a *modular structure* and that ideally (1) the considered design-time transformation should exhibit stability (i.e., evolvability), and (2) the run-time instantiation of implemented primitives should exhibit isentropicity (i.e., diagnosability). For this purpose, a set of theorems is derived based on insights from traditional engineering sciences such as systems theory and thermodynamics. In this section, we will briefly highlight both approaches. First, we will discuss the essence of NS in its initial form, i.e., starting from the stability point of view from systems theory. Next, the recent association and indications towards conformance with entropy concepts from thermodynamics will be highlighted. A preliminary discussion of some real-life NS software implementations can be found in [5].

A. Normalized Systems and Stability

Normalized Systems theory initially originated from the well-known maintenance problems in software applications and the phenomenon that software programs tend to become ever more complex and badly structured as they are changed over time, becoming more and more difficult to adapt and

(a) Container A (b) Container B (c) Container C

Figure 1. Container example for illustrating the concept of entropy. The grey parts represent those spaces filled with gas, the white parts represent those spaces, which are empty.

hence, less evolvable [11]. In order to obtain evolvable modularity, NS theory states that the functional/constructional transformation should exhibit *systems theoretic stability*, meaning that a bounded input function (i.e., bounded set of requirement changes) should result in bounded output values (i.e., a bounded impact or effort) even if an unlimited systems evolution with time $t \rightarrow \infty$ is considered. From this perspective, Mannaert et al. [4] have formally proven that this implies that the modular structure should strictly adhere to the following *principles*:

- *Separation of Concerns (SoC)*, enforcing each concern (here: change driver) to be separated;
- *Data Version Transparency (DvT)*, enforcing communication between data in a version transparant way;
- *Action Version Transparency (AvT)*, requiring that action components can be updated without impacting calling components;
- *Separation of States (SoS)*, enforcing each action of a workflow to be separated from other actions in time by keeping state after every action.

As the systematic application of these principles results in very fine-grained modular structures, NS theory proposes to build information systems based on the aggregation of instantiations of five higher-level software patterns or *elements*, i.e., action elements, data elements, workflow elements, trigger elements and connector elements [10], [4], [5]. Typical cross-cutting concerns (such as remote access, persistence, access control, etcetera) are included in these elements in such a way that it is consistent with the above-mentioned theorems.

A more formal discussion of the stability principles and reasoning as well as some initial case study findings can be found in [4] and [5] respectively.

B. Normalized Systems and Entropy

Recently, efforts were made to explain the above-mentioned findings in terms of *entropy as defined in thermodynamics* [6]. Entropy is a well-known and much debated engineering concept, originating from thermodynamics and referring to its Second Law. As we pointed out in [12], some common interpretations associated with entropy include (1) complexity, perceived chaos or disorder [13], (2) uncertainty or lack of information [14] and (3) the tendency of constituent particles in a system to dissipate or spread out [15]. In [6], it was proposed to primarily employ the statistical thermodynamics perspective on entropy for the extension of NS theory. As such, the definition of Boltzmann [16] was adopted, considering entropy as the number of microstates (i.e., the whole of microscopic properties of a system) consistent with a certain macrostate (i.e., the whole of externally observable and measurable properties of a system).

Consider for example Figure 1, symbolizing a gas container having a boundary in the middle, which is dividing the container into two compartments. The boundary completely isolates both parts of the container as a result of which the gas is solely present in the right side of the container, leaving the left part empty (see Panel (a)). While the macrostate of the container (e.g., its temperature or pressure) is brought about by one particular arrangement or configuration of the gas molecules (i.e., its microstate: the union of the position, velocity, and energy of all gas molecules in the container), many different configurations of molecules (microstates) might result in this same macrostate (hence illustrating the amount of entropy). Now imagine that the shaft between the two components is removed and both components become one single space: the gas (and the energy of its molecules) will expand, dissipate and spread out into the full space, interacting with the second component of the container (see Panel (b)). This interaction (and the removal of the fragmentation, separation and structure between both spaces) moreover increases the degree of entropy as now even a larger set of microstates (configurations of the molecules) can result in a single macrostate. The only way to avoid an increase of entropy throughout time is by introducing structure or effective boundaries between subsystems. Note that such boundaries need to be complete and encompassing, as in the case of partial fragmentation no entropy reduction is obtained (see Panel (c)).

Applying this reasoning to software applications, *microstates* could be defined as binary values representing the correct or erroneous execution of a construct of a programming language. A *macrostate* can then be seen in terms of loggings or database entries representing the correct or erroneous processing of the considered software system. From this perspective, Mannaert et al. [6] have proposed a second set of principles that should be strictly adhered to in order to achieve diagnosability in a modular structure. First, the above described principles of *Separation of Concerns* (now in terms of information units) and *Separation of States* could equally be derived from this entropy reasoning as well. Next, a set of two additional principles were formulated:

- *Data instance Traceability (DiT)*, enforcing the actual version and the values of every instance of a data structure serving as an argument, to be exported (e.g., logged) to an observable macrostate;
- *Action instance Traceability (AiT)*, enforcing the actual version of every instance of a processing function and the thread it is embedded in, to be exported (e.g., logged) to an observable macrostate.

III. Related Work on Modularity

The use of the concept of modularity has been noticed to be employed in several scientific domains such as computer science, management, engineering, manufacturing, etcetera [2], [3]. While no single generally accepted definition is known, the concept is most commonly associated with the process of subdividing a system into several subsystems [19], [20]. This decomposition of complex systems is said to result in a certain degree of complexity reduction [21] and facilitate change by allowing modifications at the level of a single subsystem instead of having to adapt the whole system at once [22], [2], [3].

As such, Baldwin and Clark defined modularity as follows: "*a unit whose structural elements are powerfully connected among themselves and relatively weakly connected to elements in other units*" [3, p. 63]. They conceive each system or artifact as being the result of specifying values for a set of design parameters, such as the height and the vessel diameter in case of a tea mug. The task of the designer is then to choose the design parameter values in such a way, that the 'market value' of the system as a whole becomes maximized. Some of the design parameters might be dependent on one another, as for example the value of the vessel diameter should be attuned to the value of the diameter of a mug. This reasoning is visualized in Figure 2 as well. Here, the use of Design Structure Matrices (DSM) is proposed. Such matrices (as originally elaborated by Steward [17] and Eppinger et al. [18]) typically depict the design parameters in both the rows and columns of the matrix. Each 'x' represents a dependency (hence, coupling) between two design parameters. Additionally, the direction of the dependency is indicated: for instance, in the situation as depicted in Figure 2, the matrix implies that the choice of a particular value for design parameter B (e.g., mug diameter) determines the set of possible choices for the value of design parameter A (e.g., vessel diameter). Obviously, a myriad of types of Design Structures Matrices can appear, according the system under consideration.

After drafting a Design Structure Matrix for the system to be engineered (i.e., providing a detailed overview of the dependencies among the different relevant design parameters), modularization is conceived by Baldwin and Clark as the process in which groups of design parameters — highly interrelated internally, but loosely coupled externally — are to be identified as modules and can be designed rather independently from each other, such as for instance the drive system, main board and LCD screen in case of a simplified computer hardware design. In Figure 2, the different modules are indicated by the light and dark grey zones, respectively. As argued by Baldwin and Clark, the modules should thus be ideally fully decoupled: this would mean that, as is the case in Figure 2, there are only 'x's placed within the light and dark grey zones and no 'x's should be found in the white areas representing dependencies between both modules. Nevertheless, as such a situation is mostly only a theoretical ideal, in most realistic settings, dependencies between the distinct modules do occur. In such situations, a set of architectural or design rules (i.e., externally visible information) is typically used to secure the compatibility between the subsystems in order to be assembled into one working system later on, while the other design parameters are only visible for a module itself. Finally, they conclude that this modularity property allows for multiple (parallel) experiments for each module separately, resulting in a higher 'option value' of the system in its totality. Instead of just accepting or declining one system as a whole, a 'portfolio of options' can be considered, as designers can compose a system by purposefully selecting among a set of alternative modules. Systems evolution is then believed to be characterized by the following six modular operators [3]:

- *Splitting* a design (and its tasks) into modules;
- *Substituting* one module design for another;
- *Augmenting*, i.e., adding a new (extra) module to the system;
- *Excluding* a module from the system;
- *Inverting*, i.e., isolating common functionality in a new module, thus creating new design rules;
- *Porting* a module to another system.

Typically, besides traditional physical products, many other types of systems are claimed to be able to be regarded as modular structures as well. First, all different programming and software paradigms can be considered as using modularity as a main concept to build software applications [10]. Whether they are using classes, objects, structures, functions, procedures, etcetera, they all are basically allowing a programmer to compose a software system by aggregating a set of instances from a collection of primitives (available in the concerning programming paradigm) in a modular way. Furthermore, while Baldwin and Clark primarily illustrate their discussion by means of several evolutions in the computer industry, they also explicitly refer to the impact of product modularity on the (modular) organization of workgroups both within one or multiple organizations,

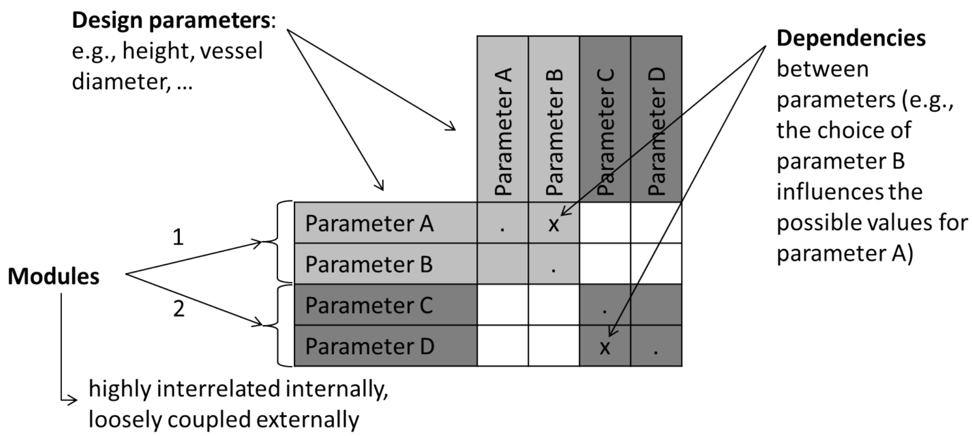


Figure 2. Modularity reasoning as proposed by Baldwin and Clark [3], based on Design Structure Matrices (DSM) from Steward [17] and Eppinger et al. [18].

and even whole industry clusters [3]. Also, Campagnolo and Camuffo [20] investigated the use of modularity concepts within management science and identified 125 studies in which modularity concepts arose as a design principle of complex organizational systems, suggesting that principles based on the concept of modularity offer powerful means to be applied at the organizational level.

Within the field of Enterprise Engineering, trying to give prescriptive guidelines on how to design organizations according to certain (desirable) characteristics, modularity equally proved to be a powerful concept. For instance, Op't Land used modularity related criteria (including coupling and cohesion) to analyze and predict the merging and splitting of organizations [23]. Van Nuffel proposed a framework to deterministically identify and delimit business processes based on a modular and NS theory viewpoint [7], [8], and Huysmans demonstrated the usefulness of modularity with regard to the study of (the evolvability) of enterprise architectures [9].

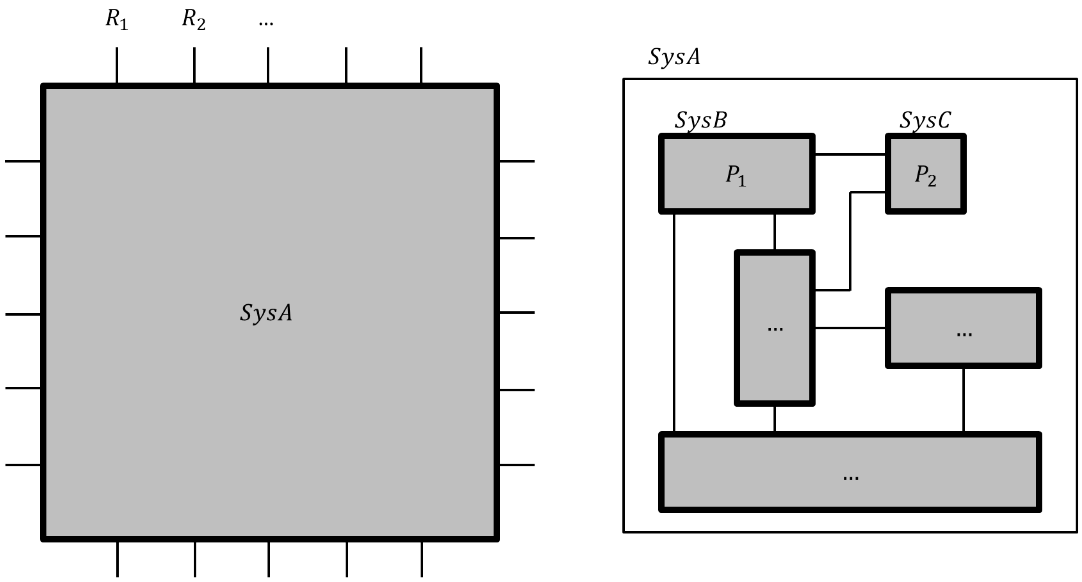
IV. Towards a Complete and Unambiguous Definition of Modules

While we are obviously grateful for the valuable contributions of the above mentioned authors, we will argue in this section that the definition of modularity, as for example coined by Baldwin and Clark [3], already describes an ideal form of modularity (e.g., loosely coupled and independent). As such, before starting our generalization efforts of NS theory to modularity issues in general, we need to clarify a few elements regarding our conceptualization of modularity. First, we will discuss the need to distinguish both functional and constructional perspectives of systems. Next, we will propose to introduce the formulation of an exhaustive modular interface as an intermediate stage, being a necessary and sufficient condition in order to claim 'modularity'. The resulting modules can then be optimized later on, based on particular criteria, which will be our focus of Section V.

A. Blackbox (Functional) versus Whitebox (Constructional) Perspectives on Modularity

When considering systems in general — software systems, organizational systems, etcetera — both a functional and constructional perspective should be taken into account [24]. The *functional perspective* focuses on describing what a particular system or unit does or what its function is [25]. While describing the external behavior of the system, this perspective defines input variables (what does the system need in order to perform its functionality?), transfer functions (what does the system do with its input?) and output variables (what does the system deliver after performing its functionality?). As such, a set of general requirements, applicable for the system as a whole, are listed. The *constructional perspective* on the other hand, concentrates on the composition and structure of the system (i.e., which subsystems are part of the system?) and the relation of each of those subsystems (i.e., how do they work together to perform the general function and adhere to the predefined requirements?) [26].

Equivalently, one could regard the functional system view as a blackbox representation, and the constructional system view as a whitebox representation. By *blackbox view* we mean that only the input and output of a system is revealed by means of an interface, describing the way how the system interacts with its environment. As such, the user of the system does not need to know any details about the content or the inner way of working of the system. The way in which the module performs its tasks is thus easily allowed to change and can evolve independently without affecting the user of the system, as long as the final interface of the system remains unchanged. The complexity of the inner working can also be said to be hided (i.e., information



(a) Blackbox (functional) representation of $SysA$. R_1, R_2, etc. represent the requirements imposed to $SysA$ as a whole.

(b) Whitebox (constructional) representation of $SysA$. Requirements R_1, R_2, etc. are 'realised' through the collaboration of a set of instances of primitives P_1, P_2, etc. in $SysB$, $SysC$, etc. At their turn, $SysB$, $SysC$, etc. are represented here in a blackbox way.

Figure 3. Blackbox (functional) and whitebox (constructional) representations of system $SysA$.

hiding), resulting in some degree of complexity reduction. The *whitebox view* does reveal the inner way of working of a system: it depicts the different parts of which the system consists in terms of primitives, and the way these parts work together in order to achieve the set of requirements as listed in the blackbox view. However, each of these parts or subsystems is a 'system' on its own and can thus again be regarded in both a functional (blackbox) and constructional (whitebox) way.

The above reasoning is also depicted in Figure 3: both panels represent the same system $SysA$, but from a conceptually different viewpoint. Panel (a), depicting the functional (blackbox) view, lists the requirements (boundary conditions) R_1, R_2, $\ldots$ imposed to the system. These are proposed as 'surrounding' the system in the sense that they do not say anything about how the system performs its tasks, but rather discuss what it should perform by means of an interface in terms of inputs and outputs. Panel (b) depicts the constructional (whitebox) view of the same system: the way of working of an aggregation of instantiations of primitives P_1, P_2, $\ldots$ (building blocks), collaborating to achieve the behavior described in Panel (a). Each of the primitives in Panel (b) is again depicted in a blackbox way and could, at their turn, each also be analyzed in a constructional (whitebox) way.

B. Avoiding Hidden Coupling by Strictly Defining Modular Interfaces

Before analyzing and optimizing the transformation between both perspectives, the designer should be fully confident that the available primitives can really be considered as 'fully fledged, blackbox modules'. By this, we mean that the user of a particular module should be able to implement it (i.e., in design-time), exclusively relying on the available interface, thus without having any knowledge about the inner way of working of the concerned module. Stated otherwise, while other authors previously already elaborated on the importance of interfaces for allowing a workable modular design (see e.g., [27]), we primarily stress the importance that the interface of a module should describe any possible dependency regarding the module, needed to perform its functionality. As long as the interface or boundary is not fully articulated, undocumented interaction with other systems can and will occur. In such situation, the study of the functional/constructional transformation or any optimization effort may become worthless. To a certain extent, our reasoning is also somewhat parallel to the motivation of Dijkstra in his argument to abolish 'goto instructions' in programming languages [28]: an incomplete or underspecified interface makes it very difficult if not impossible for an engineer to mentally mimic the actual 'way of working' of the modules (i.e., at run-time) as no clear overview is available on how they affect one another. The importance of a full interface can finally be illustrated by Figure 1(c): only in case the boundary (i.e., interface) is complete, a subsystem (here: gas container compartment) can be properly studied in an isolated way.

Consequently, every interaction of a system with its environment should be properly and exhaustively defined in the interface of a module. While this may seem rather straightforward at first sight, real-life interfaces are rarely described in such a way. Indeed, typical non-functional aspects such as technological frameworks, infrastructure, knowledge, etcetera are consequently also to be taken into account (see Section VI-B). Not formulating these 'tacit assumptions' in an explicit way results in hidden coupling: while the system is claimed to be a module, it actually still needs whitebox inspection in order to be implemented in reality, diminishing the pretended complexity reduction benefits.

Consider for instance a multiplexer for use in a typical processor, selecting and forwarding one out of several input signals. Here, one might conceptually think at a device having for example 8 input signals, 3 select lines and 1 output signal. While this is conceptually certainly correct, a real implementation on a real processor might for example (hypothetically) require 120μ by 90μ CMOS (i.e., material) to make the multiplexer physically operational on the processor, while this is not explicitly mentioned in its conceptual interface. As such, this 'resource dimension' should be made explicit in order to consider a multiplexer as a real blackbox in the sense that the module can be unambiguously and fully described by its interface. A person wanting to use a multiplexer in real-life in a blackbox way, should indeed be aware of this prerequisite prior to his ability of successfully implementing the artifact.

A more advanced example of hidden coupling includes the use of a 'method' in typical object-oriented programming languages, frequently suggested as a typical example of a 'module' in software. Indeed, in previous work, it was argued to consider the multidimensional variability when analyzing the evolvability of programming constructs (such as data structures and processing functions) and that in typical object-oriented programming environments these dimensions of variability increase even further as they make it possible to combine processing actions and data entities into one primitive (i.e., a single class) [4]. Hence, it was argued to start the analysis of object-oriented modular structures already at the level of methods instead of only considering a class as a possible 'module'. However, while it is usually said that such a method in object orientation has an interface, this interface is not necessarily completely, exhaustively and fully defined and thus such a method cannot automatically be considered as a 'real module' according to our conceptualization. Consider for example the constructor of the class in which the method has been defined. Typically, the constructor has to perform certain actions (e.g., making an instantiation (object) of the concerned class) before one can execute the concerned method. Also member variables of the class might introduce hidden coupling: first, they can be manipulated by other methods as well, outside control of the considered method. Second, they have to be created ('exist') before the module can perform its functionality. Finally, employing external libraries in case a method wants to be deemed a genuine module, would imply that ór the library should be incorporated into the module (each time) ór the external library should be explicitly mentioned in the interface. Analogously, a method having a clear interface regarding how to call the method and how the returned values should be captured, can still call (at its turn) another (third) method or rely on a certain external service or technology (e.g., a connection to the financial network of SWIFT, the use of an interface module to an external system, etcetera). As these extra methods, services or technologies are necessary to successfully complete the method, they should actually be included in such a complete and exhaustive interface in addition to the typical parameters and instructions needed to make the method call.

Hence, in our view, one has a genuine module as soon as one is able to define a complete interface, which clearly describes the boundaries and interactions of the subsystem and allows it to be used in a blackbox way. Modularization is then the process of meticulously identifying each dependency of a subsystem, transforming an ambiguously defined '*chunk*' of a system into a clearly defined *module* (of which the borders, dependencies, etcetera are precisely and ex-ante known). Compared to the definition of Baldwin and Clark cited previously, we thus do not require for a module to exhibit already high intramodular cohesion and low intermodular coupling at this stage. Modules having these characteristics are nevertheless obviously highly desirable. However, we are convinced that defining in a first phase such a complete interface, allows to 'encapsulate' the module in an appropriate way and avoid any sort of hidden coupling. Indeed, at least four out of the six mentioned modular operators in Section III require real blackbox (re)usable modules as a conditio sine qua non. More specifically, in order to use the operators Substituting, Augmenting, Excluding and Porting in their intended way, complete and exhaustively defined interfaces are a prerequisite. On the other hand, the modular operators Splitting and Inverting concern the definition of new modules and design rules. Hence, they are precisely focused on the process of defining new modular interfaces themselves, thus usually involving some form of whitebox inspection.

Finally, while defining modules with such a strict interface will not directly solve any interdependency, evolvability, ... issues, it will at least offer the possibility to profoundly study and optimize the 'quality' of the modules (e.g., with regard to coupling and cohesion) in a next stage.

V. Towards Generalizing Normalized Systems to the Function/Construction Transformation

In the same way as the implementation of software is considered as a transformation $\mathcal{I}$ of a set of functional requirements into a set of software primitives (constructs) in [4], the design or engineering of systems in general could be considered as a transformation $\mathcal{D}$ of a set of functional requirements R_j into a set of subsystems or primitives P_i:

$$\{P_i\} = \mathcal{D}\{R_j\}$$

This transformation $\mathcal{D}$ can then be studied and/or optimized in terms of various desirable system properties. In this section, we present a very preliminary discussion on the meaning of several important system properties in this respect.

First, it seems highly desirable to have a linear design transformation that can be *normalized*. This would imply that the transformation matrix becomes diagonal or in the Jordan form, leading to a one-to-one mapping of functional requirements to (a set of) constructional primitives. Such a normalized transformation is explored in [10], [4] for the implementation of elementary functional requirements into software primitives.

Moreover, this approach to systems design or engineering also seems to imply that we should avoid to perform functional decomposition over many hierarchical levels, before starting the composition or aggregation process [29]. Studying and/or optimizing the functional to constructional transformation is a very delicate activity that can only be performed on one or two levels at a time. Therefore, the approach seems to imply a preference for a bottom-up or meet-in-the-middle approach, trying to devise the required system (i.e., the set of functional requirements R_j) in terms instantiations of a set of predefined primitives (i.e., P_i), over a top-down approach.

Next, some other appealing transformation properties might include stability, scalability and isentropicity as we will discuss in the following subsections.

A. Stability

As discussed in [4] and Section II-A for the software implementation transformation, any design transformation can be studied in terms of *stability*. This means that a bounded set of additional functional requirements results only in a bounded set of additional primitives and/or new versions of primitives. As elaborated by Mannaert et al. [4], this would require the absence of so-called combinatorial effects, which result in an impact of additional functional requirements that is proportional to the size of the system. An example of an unstable requirement is for instance a small software application for some recreational sports club that needs to become highly secure and reliable, requiring a completely new and different implementation. In a more traditional engineering context, one could consider the extension of an existing building with additional rooms, which could result in many modifications or even the complete replacement of the plumbing, central heating system, etcetera, of the building.

B. Scalability

The concept of scalability is at least related, and could even be considered to be a special case of stability in this context. Scalability would mean that the increase in value of an existing functional requirement has a clearly defined and limited impact on the constructional view. An example of such a scalable requirement is the amount of concurrent users of a website, which can normally be achieved by adding one or more additional servers. An obvious example of an unscalable requirement in a traditional engineering design is the increase in the number of passengers for an airplane, as this would currently lead to the design of a completely new airplane. In a similar way, an increase of the target velocity or payload capacity of a rocket, generally leads to the design of a completely new and different rocket. One could note here that new rocket manufacturers are indeed trying to scale up existing rocket designs, using more engines or even complete stages in parallel for larger rockets.

C. Isentropicity

With regard to the diagnosability, the concept of isentropicity from statistical thermodynamics can be applied, meaning that an externally observable system state (i.e., a systems macrostate) should completely and unambiguously determine the states of each of the various constituting subsystems (i.e., the systems microstate).

In our view, an isentropic design would therefore imply that the externally observable state of $SysA$ completely and unambiguously determines the states of the various subsystems. An example of such an isentropic design is a finite state machine where the various registers can be read. Indeed, the inputs and register values that are externally observable completely define the internal state of the finite state machine.

VI. Towards the Application of Normalized Systems to Enterprise Engineering

In Sections I and III we argued that not only software applications can be regarded as modular systems, but also many other types of artifacts, such as (for example) organizations. Hence, Sections IV and V focused on a first attempt to extend NS theory concepts to modularity and the systems engineering process in general. In this section, by means of example, we will illustrate some of the implications of our proposed engineering approach when applied to organizational systems. Indeed, several authors have argued for the need of the emergence of an Enterprise Engineering discipline, considering organizations as (modular) systems, which can be 'designed' and 'engineered' towards specific criteria [30], [31], such as (for example) evolvability. More specifically, we will first focus our efforts here at illustrating how the concepts of stability and entropy might offer interesting perspectives to analyze the modular structure of business processes. Next, we will elaborate on the complete and unambiguous definition of organizational modules, as this is in our view a necessary condition to be able to study and optimize the functional/constructional transformation at a later stage.

A. Normalized Systems and Business Process Analysis

When applying the concept of *stability* as considered in NS to business topics, one could consider both business process flows and enterprise architectures. Focusing on business processes, one way to interpret a business process combinatorial effect is the situation in which a single (business process) requirement change leads to N changes in the design of the considered business processes in the business process repository [8].

Consider for instance the handling of a payment incorporating several distinct tasks such as the receiving of an invoice, balance checking, payment execution at the correct date and in the requested format, while taking care of the required accounting and security procedures, etcetera. Suppose that this business process is not separately contained into a single and distinct business process, but instead all kinds of variants are incorporated in all business processes needing to perform payments to (for example) employees, suppliers, moneylenders, and so on. Now further suppose that, for example, a change in legislation would enforce an additional check to be performed for each payment, or a new available payment method would arise. These kind of functional changes would imply an impact of N constructional changes all over the business process repository (i.e., depending on the size of the repository and the number of business processes in which the payment functionality was incorporated). In addition, the precise locations of those modifications within the process repository are unknown upfront and whitebox inspection of each and every process is required to perform the change in a consistent way.

When applying the concept of *entropy* to the level of business process analysis, as we discussed in [12], a first effort should be directed towards interpreting macro- and microstates in such context. Hence, regarding the *macrostate*, typical externally observable properties of a business process, in our view, might include:

- *throughput or cycle time* (how long did the process take to be executed?);
- *quality* and other output related measures (e.g., succesful or non-succesful completion of the process as a whole or the number of defects detected after the execution of the process);
- *costs* involved in the process;
- other *resources* consumed by the process (such as raw materials, electricity, human resources, etcetera).

A typical *microstate*, related to the above sketched macrostate, might then comprise the throughput time of a single task in the process, the correct or erroneous outcome of a single task, the costs related to one activity or the resources consumed by one particular task of the considered business process. Analyzing instantiated business processes in terms of these defined macro- and microstates would then come down to management questions such as:

- which task or tasks in the business process was (were) reponsible for the extremely slow (fast) completion of (this particular instance of) the business process? ;
- which task or tasks in the business process was (were) responsible for the failure of the considered instantiated business process? ;
- which activities contributed substantially or only marginally to the overall cost or resource consumption of the considered business process (cf. cost-accounting and management approaches like Activity Based Costing)?

In case the answer to these questions is unambiguous and clear, the entropy in the system (here: business process repository) is low (or ideally zero) as a particular macrostate (e.g., the extremely long throughput time) can be related to only one or a few microstates (e.g., activity X took three times the normal duration to be carried out, whereas all other activities finished in their regular time span). On the other hand, when no direct answer to these questions can be found, entropy increases: multiple microstates (e.g., prolonged execution of activities X and/or Y and/or Z) could have resulted in the observed and possibly problematic macrostate (e.g., the lengthy execution of the overall process). While we are definitively not the first aiming to relate business processes to entropy (see e.g., [32], [33], [34]), our approach differs in the sense that we consider entropy from the thermodynamics perspective in a run-time environment. An important implication thereof is the fact that entropy in business processes seems to be related to the unstructured aggregation of information and data [12], which could offer interesting research opportunities in (for instance) the accounting domain [35], [36].

B. Towards a Complete and Unambiguous Definition of Organizational Modules

When also considering modules at the organizational level, a considerable effort should equally be aimed at exhaustively listing the interface, incorporating each of its interactions with the environment. This because not including certain dimensions in the interface might evaporate optimization efforts based on stability or entropy reasoning (see Section IV-B and Figure 1(c)).

For instance, when focusing on a payment module, not only the typical 'functional' or 'operational' interface such as the account number of the payer and the payee, the amount and date due, etcetera (typical 'arguments'), but also the more 'configuration' or 'administration' directed interface including the network connection, the personnel needed, etcetera (typical 'parameters') should be included. As such, we might distinguish two kinds of interfaces:

- a *usage interface*: addressing the typical functional and operational (business-oriented) arguments needed to work with the module;
- a *deployment interface*: addressing the typical non-functional, meta-transaction, configuration, administration, ... aspects of an interface.

Although some might argue that this distinction may seem rather artificial and not completely mutually exclusive, we believe that the differences between them illustrate our rationale for a completely defined interface clearly.

While the work of Van Nuffel [8] has resulted in a significant contribution regarding the identification and separation of distinct business processes, the mentioned interfaces still have the tendency to remain underspecified in the sense that they only define the functional 'business-meaning' content of the module but not the other dimensions of the interface, required to fully use a module in blackbox fashion. Such typical other (additional) dimensions — each illustrated by means of an imaginary organizational payment module — might include:

1) Supporting technologies: Modules performing certain functionality might need or use particular external technologies or frameworks. For example, electronical payments in businesses are frequently performed by employing external technologies such as a SWIFT connection or Isabel. In such a case, a payment module should not only be able to interact with these technologies, but the organization should equally have a valid subscription to these services (if necessary) and might even need access to other external technologies to support the services (e.g., the Internet). An organization wanting to implement a module in a blackbox way should thus be aware of any needed technologies for that module, preferably by means of its interface and without whitebox inspection. Suppose that one day, the technology a module is relying on, undergoes some (significant) changes resulting in a different API (application programming interface). Most likely, this would imply that the module itself has to adapt in order to remain working properly. In case the organization has maintained clear and precise interfaces for each of its modules, it is rather easy to track down each of the modules affected by this technological change, as every module mentioning the particular technology in its interface will be impacted. In case the organization has no exhaustively formulated interfaces, the impact of technological changes is simply not known: in order to perform a confident impact analysis, the organization will have to inspect each of the implemented modules with regard to the affected technology in a whitebox way. Hence, technological dependencies should be mentioned explicitly in a module's interface to allow true blackbox (re)use.

2) Knowledge, skills and competences: Focusing on organizations, human actors clearly have to be taken into account, as people can bring important knowledge into an organization and use it to perform certain tasks (i.e., skills and competences). As such, when trying to describe the interface of an organizational module in an exhaustive way, the required knowledge and skills needed for instantiating the module should be made explicit. Imagine a payment module incorporating the decision of what to do when the account of the payer turns out to be insolvent. Besides the specific authority to take the decision, the responsible person should be able (i.e., have the required knowledge and skills) to perform the necessary tasks in order to make a qualitative judgment. Hence, when an organization wishes to implement a certain module in a blackbox way, it should be knowledgeable (by its interface) about the knowledge and skills required for the module to be operational. Alternatively, when a person with certain knowledge or skills leaves the company, the organization would be able to note immediately the impact of this knowledge-gap on the well-functioning of certain modules and could take appropriate actions if needed.

3) Money and financial resources: Certain modules might impose certain financial requirements. For example, in case an organization wants to perform payments by means of a particular payment service (e.g., SWIFT or Isabel), a fixed fee for each payment transaction might be charged by the service company. If the goal is to really map an exhaustive interface of a module, it might be useful to mention any specific costs involved in the execution of a module. That way, if an organization wants to deploy a certain module in a blackbox way, it may be informed about the costs involved with the module ex-ante. Also, when the financial situation of an organization becomes for instance too tight, it might conclude that it is not able any longer to perform the functions of this module as is and some modifications are required.

4) Human resources, personnel and time: Certain processes require the time and dedication of a certain amount of people, possibly concurrently. For example, in case of an organizational payment module, a full time person might be required to enter all payment transactions in the information system and to do regular manual follow-ups and checking of the transactions. As such, an exhaustive interface should incorporate the personnel requirements of a module. That way, before implementing a certain module, the organization is aware of the amount of human resources needed (e.g., in terms of full time equivalents) to employ the module. Equivalently, when the organization experiences a significant decline or turnover in personnel, it might come to the conclusion that it is no longer able to maintain (a) certain module(s) in the current way. Obviously, this dimension is tightly intertwined with the previously discussed knowledge and skills dimension.

5) Infrastructure: Certain modules might require some sort of infrastructure (e.g., offices, materials, machines) in order to function properly. Again, this should be taken into account in an exhaustive interface. While doing so, an organization adopting a particular module knows upfront which infrastructure is needed and when a certain infrastructural facility is changed or removed, the organization might immediately evaluate whether this event impacts the concerning module and vice versa.

6) Other modules or information: Certain modules might use other modules in order to perform their function. For example, when an organization decides to perform the procurement of a certain good, it will probably receive an invoice later on with a request for payment. While the follow-up of a procurement order might be designed into one module, it is reasonable to assume that the payment is designed in a distinct module, as this functionality might also return in other business functions (e.g., the regular payment of a loan). As such, when an organization is planning to implement the procurement module, it should be aware that also a payment module has to be present in the organization to finalize procurements properly. Hence, all linkages and interactions with other modules should be made explicit in the module's interface. When a module (including its interface), used by other modules, is changed at a certain point in time, the adopting organization then immediately knows the location of impact in terms of implemented modules and hence where remedial actions might be required.

In terms of entropy reasoning, the lacking of one or multiple of these above-mentioned dimensions in the organizational module interface can hamper traceability (and hence, diagnosability) regarding the eventually produced outcomes of the organization. For instance, a prolonged throughput time of the payment process can be due to a delay in the communication with one of its external modules or due to the lack of extra knowledge required for the incorporation of an additional legally required check. In case these dimensions are not listed in the module interface, they would typically not be considered as possible causes for the observed result. However, it seems reasonable to assume that an enterprise engineer or dianostician dóes want to be aware of all these possibly problem causing dimensions.

Obviously, it is clear that exhaustively defining the technology, knowledge, financial resources, etc. on which a module depends, will not suffice to solve any of the existing coupling or dependencies among modules. Also, one should always take into consideration that a certain amount of 'coupling' will always be needed in order to realistically perform business functions. However, when the interface of each module is clearly defined, the user or designer is at least aware of the existing dependencies and instances of coupling, knows that ripple-effects will occur if changes affect some of the module's interfaces (i.e., impact analysis) and can perform his or her design decisions in a more informed way, i.e., by taking the interface with its formulated dependencies into account. Consequently, once all forms of hidden coupling are revealed, finetuning and genuine engineering of the concerned modules (e.g., towards low intermodular coupling) seems both more realistic and feasible in a following phase. Indeed, one might deduct that Baldwin and Clark, while defining a module as consisting of powerfully connected structural elements, actually implicitly assumed the existence of an exhaustive set of formulated dependencies before modularization can occur, witness the fact for example that they elaborately discuss Design Structure Matrices [3, chapters 2 & 3]. Our conceptualization is then not to be interpreted as being in contradiction with that of Baldwin and Clark, rather we emphasize more explicitly that the mapping of intermodular dependencies is not to be deemed negligible or self-evident.

VII. On the Feasibility of Applying NS Enterprise Engineering: Some Illustrating Case Studies

In the previous sections, our arguments were mainly illustrated by referring to conceptual examples, such as a tea mug design or imaginary payment module. In this section, we will discuss two short real-life cases further demonstrating the feasibility of how our discussed concepts and reasoning can be applied for analyzing and optimizing realistic organizational problems. Both cases are based on data collected at the administrative department of the authors' research institution and university.

A. Case 1

Our first case analyzes the procedures for the registration of the examination marks of professors and teaching assistants at the end of each semester. In the initial situation, professors and teaching assistants (the 'examiners') were asked to send their grades for each course to the administrative department in one of two possible ways: (1) manually handing in a list of students and their corresponding marks at the secretariat, or (2) sending a mail with a similar list, usually by means of a spreadsheet document. In a next stage, this information would be manually processed by people at the administrative department into the corresponding software applications, generating the student reports afterwards. From the personnel in the administrative department, this processing step required a considerable amount of effort, which kept increasing due to the rising number of students. Hence, an initiative was launched to optimize the way in which the grade administration was performed, aiming to lower the workload on the administrative department. A license for a new software application was bought for this purpose, which was deemed to allow professors and teaching assistants to enter their marks autonomously into the overarching information system by means of a web interface. This way, no direct intervention of the administrative personnel would be required any longer and the students reports would be able to be generated automatically.

Based on our modularity and (NS) enterprise engineering reasoning as discussed above, several remarks can be made. First, as the task 'reporting on the course grades' was embedded in the responsibility of each 'examiner', a necessary 'instability' was noticeable in the form of a *combinatorial effect*. Indeed, the new organizational way of working required all 'examiners' to adapt their individual way of working: no hard-copy forms or direct mails to the secretariat were allowed any longer to register their marks. Instead, as an additional effort compared to the initial situation, each 'examiner' was required to install the correct Internet browser, familiarize himself with the electronic platform, understand the new GUI of the software application, etcetera. One can identify this phenomenon as a combinatorial effect as the considered change (here: the transition to the new application allowing for the web interface) has an impact related to size of the system to which the change has been applied (here: the university having N examining professors and teaching assistants). Such combinatorial effects also have their baleful influence at the organizational performance, such as an increased implementation time of the process optimization directive (i.e., the operation could only succeed after all 'examiners' performed the necessary changes), or an increased risk regarding the incorrect implementation of the directive (i.e., during the implementation efforts of each examiner, errors or inconsistencies might occur).

Second, the new way of working implied a significantly more *complex interface* for all 'examiners' to complete their reporting duties towards the administrative department. In the initial situation, the task 'reporting on the course grades' had an interface, which simply consisted of a plain list with student names and their corresponding grades. In the new situation, this interface became much more complex. Indeed, for successfully handling the task 'reporting on the course grades', the 'examiners' were not only required to provide a list with student results. Instead, 'examiners' were forced to deal with additional concerns on four levels. First, as a 'one-time set-up', the 'examiners' needed to be able to install the correct Internet browser (version) on their computer to be able to access the web interface of the new software application and correctly set-up a VPN client to allow for a secured connection with the university's network. Next, as a 'pre-operation set-up', the 'examiners' needed to perform the correct log-in procedures each time they wanted to report on some course grades, to actually establish the secure network. Third, some 'general competences' were required from the 'examiners' to be able to interact with the browser, the web application (including the new and rather complex GUI) and VPN application, understand the web application's specific coding scheme (e.g., to indicate that a student was legitimately absent), etcetera. Fourth, in case problems arose regarding any of these issues, the 'examiners' were implicitly believed to be able to provide the correct fault and 'exception handling' for all these issues (e.g., dealing with validation errors, anomalies, strange menus, . . .). Consequently, to correctly comply with the new interface, 'examiners' were essentially forced to deal with concerns regarding the (sometimes technical) complexity of the new software application, whereas in the initial situation, people from the administrative personnel were the only ones required to interface with such software application. Hence, in the initial situation, only people from the administrative department needed to have knowledge of these specificities. However, given the more complex interface in the new way of working, many professors and teaching assistants were required to ask for assistance from the administrative personnel due to many problems of often different origins, as they had insufficient knowledge to resolve them independently (i.e., the knowledge implicitly deemed to be present, was not always available).

As a consequence, it was strikingly to note that some people at the administrative department reported that the time and effort required to assist these people, was perceived as exceeding the original effort to manually process all examination results themselves (as was the case in the initial situation). Again, unanticipated baleful consequences for the organizational performance were noticeable. This was clearly not the desired result, as the initial goal was precisely to reduce the overall administrative efforts significantly. Obviously, we do not want to claim in this analysis that automatization efforts are to be deemed counterproductive per se. Rather, we want to illustrate how some ways of working might result in instability issues or problems related to an underspecified or complicated modular interface. For instance, it was noticed that other kinds of 'self-service' automatization efforts (e.g., a new web application allowing students to subscribe themselves and choose their individual course program for each academic year) resulted in very similar problems. In essence, a similar situation arose in which many concerns (deemed to be trivial to deal with for all people) were exported from a centralized department to all people involved in the self-service. In case these concerns however turn out to be non-trivial later on, many problems occur as their complexities are pushed to all involved people via the new interface.

B. Case 2

Our second case focuses on the consequences of a change in the communication policy, such as an adapted company logo and style. Historically, the considered case organization originated from the merge of three existing universities into one larger university. One implication of this organizational merge was the need to 'market' the new university by (amongst others) introducing a new company logo and style (together with the new and correct contact details, VAT-numbers, etcetera), incorporated on (for instance) all official letters (i.e., by means of the printings on 'stationery' or university paper). The traditional routine for sending letters within the organization allowed each of the more than 4000 university members (the 'senders') to autonomously print and send letters by using the 'old' letter style (i.e., on university paper). Analyzing the considered change in the communication policy, we find again a rather large and complex impact of such a (seemingly) small organizational change.

First, as each 'sender' is responsible for taking care of the concern of applying the correct letter lay-out, a change regarding this company logo and style results in a *combinatorial effect*. Indeed, to correctly implement the new company style, each 'sender' should adapt his own way of working, by using and applying the new company logo and style (i.e., confirming to and printing on the new university papers), and hold back from using the old stationery. Hence, the impact of the applied change (here: introducing the new company logo) is dependent on the size of the target system (here: the N 'sending' employees). Also here we could identify additional baleful effects regarding the performance of the organization. One aspect involves again the increased implementation time of the change and the increased risk regarding the correct implementation of the directive (i.e., each and every 'sender' has to adapt his individual way of working and do so in a correct way). Another aspect concerns the tendency of individual university members to 'pile up' an individual 'stock' of stationery on his or her desk due to perceived efficiency reasons. Clearly, this causes a significant loss of stock, proportional to the number of university members, from the moment the change in communication policy has been announced, as the old stock of stationery becomes obsolete.

Additionally, in *entropy* reasoning, some findings (i.e., a macrostate) may suggest that some people seem not to apply the new 'rules' regarding the company logo and lay-out (e.g., given the fact that some clients still employ old VAT-numbers based on recently sent letters printed on and old version of the university paper). However, as no control or logging is kept on who is sending letters on which moment, no clear diagnosis can be made as to who (i.e., which 'sender') is still reluctant to applying the new directive. Indeed, each person sending letters after the new directive was valid, could have sent letters with a wrong letter lay-out and contact details (i.e., multiple microstates) and uncertainty arises.

A possible optimization in the organizational design could for example consist of a situation wherein each person who wants to send a letter, only drafts the letter in terms of its content and afterwards sends this 'e-letter' to the administrative department, which prints the letter on the correct version of the stationery, having the appropriate header and contact details. In such case, the combinatorial effect would be eliminated. Changes regarding the lay-out, logo or general contact information on letters would have an impact limited to the personnel of the administrative department, while the tasks carried out by people who 'send' letters remains unchanged (i.e., only having responsibility regarding the content of letters) and are change independent regarding this concern. Also, in case an error would nevertheless occur, less uncertainty would be present regarding who might have applied the wrong lay-out, as only people from the administrative department are entitled to print-out letters. Finally, also improvement on the organization's performance could be expected: changes in the communication policy would be applied in a more easy, fast and correct way, and no individual stocks of the stationery would be present any longer.

Consequently, the applications of our NS Enterprise Engineering approach to these cases, might show how our reasoning may lead to real-life and relevant organizational problem solving and decision making.

VIII. Interpreting Entropy and Instability as Coupling within Modular Structures

In the previous sections, we elaborated on how NS uses the concepts of stability and entropy to design and optimize the modular structure of software applications and how this reasoning can be generalized to modular systems in general, such as for example organizational systems. In this section, we will take a more broad perspective, reflecting on the essential meaning of stability and entropy in this modularity approach. While in essence, both stability (cf. systems theory) and entropy (cf. thermodynamics) have their origin in distinct theories, we will argue that one way to interpret both concepts is to consider them as pinpointing to other 'symptoms' of the same underlying 'cause' or phenomenon, being the coupling (or interaction) between particles and subsystems of a larger system. In advance, it should be emphasized that the following discussion has a rather explorative, reflective and tentative goal, rather than fully formalizing each aspect. However, given the earlier elaborate discussion and reasoning regarding modularity, coupling, stability and entropy, the authors believe that the given interpretation might clarify some of the claims presented above.

A. Stability and coupling

In fact, the concept of stability from systems theory might be interpreted as referring to *coupling regarding the design parameters or dimensions of modular structures*. Consider for instance a standard example of dynamic (in)stability analysis in traditional engineering: the construction of a bridge. Typically, such buildings or construction 'systems' have a natural frequency of vibration (a so-called 'eigenfrequency'). In case the frequency of the oscillation to which the system is exposed matches this eigenfrequency (e.g., due to gusts of wind), the system absorbs more energy than it does in case of any other frequency: the bridge may be forced into violent swaying motions, leading to mechanical damages and ultimately sometimes even leading to its destruction. In such a case, the system absorbs and aggregates or accumulates the energy throughout its whole structure causing an extreme 'output' instead of the gradual reduction of the wave as it occurs at all other frequencies. Hence, this transformation process can be considered to be instable as a bounded input causes an unbounded and uncontrolled output.

Consequently, engineers of such buildings have to be aware of this coupling in terms of design parameters (e.g., between the buildings eigenfrequency and the frequency of frequently occurring gusts of wind) when devising the system in order to avoid such instability effects. In terms of our modularity approach discussed above, we would require to incorporate this dependency of the 'building system' on the external environment into its 'interface'. Only then could one be able to regard the building as a genuine 'module' as the dependency reveals one type of interaction between the considered system (i.e., the building) and the environment it operates in (i.e., the meteorological conditions). Typically, engineers will try to decouple (i.e., fragment) both aspects by incorporating for example shock mounts to absorb the resonant frequency and compensate (i.e., cancel) the resonance and absorbed energy. Alternatively, the engineers might choose to design the construction in such a way that the building only resonates at certain frequencies not typically occurring. In reality, a combination of both practices will most likely be opted for. Consequently, these remedial measures can be regarded as corresponding to the optimization of the modular arrangements and their interfaces towards a specific criterion (here: the avoidance or reduction of mechanical resonance).

One thing the engineer should definitely avoid at all times is hidden coupling in this respect: the case in which one is not aware of this interaction (i.e., coupling) in terms of design parameters. In the discussed example, not being aware of the eigenfrequency and the danger it implies in terms of instable reactions, might obviously be disastrous. Hence, the coupling of the subsystem with regard to the overall system it operates in, should be mentioned in the interface to safely regard the module as a blackbox. If not all kinds of coupling are documented, whitebox inspection will still be needed to assess the impact of external changes (in our example: a gust of wind).

Another example for illustrating our interpretation of instability in terms of coupling can be found by considering violations towards NS theorems. Consider for instance the *Action version Transparency* theorem. Imagine an action entity A not exhibiting version transparency and being confronted with a mandatory version upgrade: each action entity calling this (new version) of action entity A would then be required to be adapted in terms of its call towards A. As such, employing the assumption of unlimited systems evolution, the impact of this external change can become unbounded or unstable. Also here, it can be clearly seen that the instability is caused by means of coupling within the modular structure in terms of its design parameters. As each action entity calling action entity A has a piece of implemented software code depending on the specificities of action entity A, action entity A is coupled (and interacting) with all of its calling entities through its technical design. Here again, the software engineer would preferably opt for, first, recognizing this dependency (i.e., including it in its interface) and, in a second stage, introducing fragmentation in order to control the coupling. This can for instance be done by employing action entities, which do allow for Action version Transparency, hence isolating each of the considered concerns for this dimension.

More generally, each of the NS theorems from the stability point of view (i.e., *SoC*, *AvT*, *DvT* and *SoS*) can be interpreted as theorems enforcing the decoupling of aspects causing instabilities in the modular design in case they would not be separated. Indeed, *Separation of Concerns* enforces all change drivers to be separated in the design of a software architecture in order to have modules dependent on only one independent concern, not coupled to other change drivers. *Separation of States* decouples various modules in the sense that new (error) states do not get escalated in the design towards other modules. *Action version Transparency* and *Data version Transparency* even precisely aim at the more 'traditional' interface between modules, when demanding that new versions should have an interface not impacting already existing data and action entities.

B. Entropy and coupling

When considering entropy as defined in (statistical) thermodynamics, this concept might be interpreted as referring to *coupling between subparts or particles of a modular system in terms of their run-time execution*. In fact, one way to understand entropy is the natural tendency of particles in a system to interact (see [12] and Section II-B). Indeed, the fact that an increased amount of entropy is associated with an increasing number of possible microstate configurations consistent with one macrostate, can essentially be traced back to the uncontrolled interaction between these particles. Consider for instance again the gas containment illustration as depicted in Figure 1. We know that if the gas is contained in only one out of the two components in the container, entropy will increase as soon as the shaft between the two components is removed: both components become one single space, and the gas will dissipate and spread out into the full space, interacting with the parts in the second component of the container throughout time. Due to the removal of the shaft, both components become '*coupled*' in their run-time dimension and will be subject to their natural tendency to interact with each other throughout time. This interaction (and the removal of the fragmentation or structure within the space) increases the degree of entropy as now a higher number of microstates (configurations of the molecules) can result in a single macrostate.

The only way to avoid such entropy increase, is by introducing structure or fragmentation in this run-time dimension of the modular structure. In our container example, this might be done by maintaining the shaft in the container and hence avoiding additional interaction between the molecules. As repeatedly mentioned before, it is important to note that this additional structure (such as an interface) needs to be complete and exhaustive. This can again be nicely illustrated by Figure 1, Panel (c). While a part of the shaft is still in place and possibly aimed at avoiding interaction, the interaction between both compartments still takes place as the decoupling or fragmentation is incomplete, and entropy increase will occur.

More generally, each of the NS theorems from the entropy point of view (i.e., *SoC*, *AiT*, *DiT* and *SoS*) can be interpreted as theorems enforcing the decoupling of aspects causing entropy generation in their run-time dimension in case they would not be separated. Indeed, *Separation of States* decouples various modules during the run-time execution of a software application in the sense that the statefull calling will generate a persistent state after completion of each action (and hence, information about the microstate is retained and externalized). *Separation of Concerns* enforces concerns (here: information units) to be separated so that each concern of which independent information should be traceable, is contained in a separate module (and hence, state). *Action instance Transparency* and *Data instance Transparency* even specifically aim at the fact that the versions, values and threads of actions and data need to be logged during each instantiation for traceability or diagnosability purposes.

In conclusion, as we have claimed in this section that one way to interpret both the occurrence of instability and entropy in modular structures is their relation to the existence of uncontrolled coupling, the need for completely and exhaustively defined interfaces becomes even more pertinent if one's goal is to obtain a stable modular structure exhibiting isentropicity. Moreover, structures allowing 'hidden coupling' (i.e., still leaving room for certain forms of leakage in the design or run-time dimension) should be considered harmful or 'misleading' in the sense that the designer might be convinced of the stability or isentropicity of his designed structure, while in reality the resulting module boundaries or states do not reflect an isolated part of the system, fully decoupled from the rest of the system.

IX. Conclusion and Future Work

This paper focused on the further exploration and generalization of NS systems engineering concepts to the analysis and design of modules in systems in general, and organizational systems in particular. The current state-of-the-art regarding NS and modularity was reviewed, primarily focusing on the seminal work of Baldwin and Clark. Subsequently, we argued that, first, a distinction should be made between blackbox (functional) and whitebox (constructional) perspectives of systems. As the practical blackbox (re)use of modules requires the absence of any hidden coupling, the need for complete and exhaustively defined interfaces was argued for. Next, a discussion of some properties of the functional/constructional transformation was proposed. As we believe that this systems engineering approach might be useful for optimizing other modular structures (including organizations) as well, we discussed some of the implications of this reasoning when applied to Enterprise Engineering (such as stability, entropy and complete interfaces) and provided two short real-life cases for illustrative purposes. We concluded that our conceptualization of is not in contradiction with that of Baldwin and Clark, but rather emphasizes an additional intermediate design stage when devising (organizational) modules. Also, from a broader modularity viewpoint, we provided an interpretation of the traditional engineering concepts of instability and entropy as being manifestations of coupling in modular structures (regarding their design or run-time dimensions, respectively).

Regarding our applications towards Enterprise Engineering, a limitation of this paper can be seen in the fact that no guarantee is offered that the identified additional interface dimensions will reveal all kinds of hidden coupling in every organization. Therefore, additional research (e.g., extra case studies) with regard to possible missing dimensions seems to be required. In addition, our application of modularity and NS concepts to the organizational level was limited to the definition of completely defined organizational modules and the illustration of the existence of instability and entropy generation at a business process level. A completely defined stable and isentropic functional/constructional transformation on the organizational level (as it exists on the software level) was still out of scope in this paper. Furthermore, future research at our research group will be aimed at identifying and validating organizational blackbox reusable modules, exhibiting exhaustively defined interfaces and enabling the bottom-up functional/constructional transformation.

Regarding the discussion of our interpretation of entropy and stability as different manifestations of coupling, it should be noted that it is to be considered as a mainly exploratory effort, needing further formalization in future research.

Acknowledgment

P.D.B. is supported by a Research Grant of the Agency for Innovation by Science and Technology in Flanders (IWT).

References

[1] P. De Bruyn and H. Mannaert, "Towards applying normalized systems concepts to modularity and the systems engineering process," in *Proceedings of the Seventh International Conference on Systems (ICONS) 2012*, 2012, pp. 59 – 66.

[2] C. Y. Baldwin and K. B. Clark, "Managing in an age of modularity," *Harvard Business Review*, vol. 75, no. 5, pp. 84–93, 1997.

[3] ——, *Design Rules: The Power of Modularity*. Cambridge, MA, USA: MIT Press, 2000.

[4] H. Mannaert, J. Verelst, and K. Ven, "The transformation of requirements into software primitives: Studying evolvability based on systems theoretic stability," *Science of Computer Programming*, vol. 76, no. 12, pp. 1210 – 1222, 2011, special Issue on Software Evolution, Adaptability and Variability.

[5] ——, "Towards evolvable software architectures based on systems theoretic stability," *Software Practice and Experience*, vol. Early View, 2011.

[6] H. Mannaert, P. De Bruyn, and J. Verelst, "Exploring entropy in software systems: towards a precise definition and design rules," in *Proceedings of the Seventh International Conference on Systems (ICONS) 2012*, 2012, pp. 93 – 99.

[7] D. Van Nuffel, H. Mannaert, C. De Backer, and J. Verelst, "Towards a deterministic business process modeling method based on normalized systems theory," *International Journal on Advances in Software*, vol. 3, no. 1-2, pp. 54–69, 2010.

[8] D. Van Nuffel, "Towards designing modular and evolvable business processes," Ph.D. dissertation, University of Antwerp, 2011.

[9] P. Huysmans, "On the feasibility of normalized enterprises: Applying normalized systems theory on the high-level design of enterprises," Ph.D. dissertation, University of Antwerp, 2011.

[10] H. Mannaert and J. Verelst, *Normalized systems: re-creating information technology based on laws for software evolvability*. Koppa, 2009.

[11] M. Lehman, "Programs, life cycles, and laws of software evolution," *Proceedings of the IEEE*, vol. 68, no. 9, pp. 1060 – 1076, sept. 1980.

[12] P. De Bruyn, P. Huysmans, G. Oorts, and H. Mannaert, "On the applicability of the notion of entropy for business process analysis," in *Proceedings of the Second International Symposium on Business Modeling and Software Design (BMSD) 2012*, 2012, pp. 128 – 137.

[13] G. Anderson, *Thermodynamics of Natural Systems*. Cambridge University Press, 2005.

[14] C. Shannon, "A mathemathical theory of communication," *Bell System Technical Journal*, vol. 28, pp. 379–423, 1948.

[15] H. Leff, "Thermodynamic entropy: The spreading and sharing of energy," *American Journal of Physics*, vol. 64, no. 10, pp. 1261–1271, Oct 1996.

[16] L. Boltzmann, *Lectures on gas theory*. Dover Publications, 1995.

[17] D. Steward, "The design structure system: A method for managing the design of complex systems," *IEEE Transactions in Engineering Management*, vol. 3, no. 28, pp. 71 – 84, 1981.

[18] S. Eppinger, "Model-based approaches to managing concurrent engineering," *Journal of Engineering Design*, vol. 2, no. 4, pp. 283 – 290, 1991.

[19] H. Simon, *The Sciences of the Artificial*, 3rd ed. Cambridge, Massachusetts: MIT Press, 1996.

[20] D. Campagnolo and A. Camuffo, "The concept of modularity within the management studies: a literature review," *International Journal of Management Reviews*, vol. 12, no. 3, pp. 259 – 283, 2009.

[21] H. Simon, "The architecture of complexity," in *Proceedings of the American Philosophical Society*, vol. 106, no. 6, December 1962.

[22] R. Sanchez and J. Mahoney, "Modularity, flexibility, and knowledge management in product and organization design," *Strategic Management Journal*, vol. 17, pp. 63–76, 1996.

[23] M. Op't Land, "Applying architecture and ontology to the splitting and allying of enterprises," Ph.D. dissertation, Technical University of Delft (NL), 2008.

[24] G. M. Weinberg, *An Introduction to General Systems Thinking*. Wiley-Interscience, 1975.

[25] L. Bertalanffy, *General Systems Theory: Foundations, Development, Applications*. New York: George Braziller, 1968.

[26] M. Bunge, *Treatise on Basic Philosophy: Vol. 4: Ontology II: A World of Systems*. Boston: Reidel, 1979.

[27] J. Rowson and A. Sangiovanni-Vincentelli, "Interface-based design," in *Proceedings of the 34th Design Automation Conference*, 1997.

[28] E. Dijkstra, "Go to statement considered harmful," *Communications of the ACM*, vol. 11, no. 3, pp. 147–148, 1968.

[29] P. De Bruyn, D. Van Nuffel, P. Huysmans, and H. Mannaert, "Towards functional and constructional perspectives on business process patterns," in *Proceedings of the Sixth International Conference on Software Engineering Advances (ICSEA)*, Barcelona, Spain, 2011, pp. 459–464.

[30] J. L. G. Dietz, *Enterprise Ontology: Theory and Methodology*. Springer, 2006.

[31] J. Hoogervorst, *Enterprise Governance and Enterprise Engineering*. Springer, 2009.

[32] J.-Y. Jung, "Measuring entropy in business process models," in *3rd International Conference on Innovative Computing Information and Control, 2008 (ICICIC '08)*, june 2008.

[33] J.-Y. Jung, C.-H. Chin, and J. Cardoso, "An entropy-based uncertainty measure of process models," *Information Processing Letters*, vol. 111, no. 3, pp. 135 – 141, 2011.

[34] J. Trienekens, R. Kusters, D. Kriek, and P. Siemons, "Entropy based software processes improvement," *Software Quality Journal*, vol. 17, pp. 231–243, 2009, 10.1007/s11219-008-9063-6.

[35] J. Ronen and G. Falk, "Accounting aggregation and the entropy measure: An experimental approach," *The Accounting Review*, vol. 48, no. 4, pp. 696–717, 1973.

[36] A. R. Abdel-Khalik, "The entropy law, accounting data, and relevance to decision-making," *The Accounting Review*, vol. 49, no. 2, pp. 271–283, 1974.

Optimally Controlled Nonintrusive Broadcasting for Path Key Establishment Mechanism in Wireless Sensor Networks

Aishwarya Mishra
School of Information Technology
Illinois State University
Normal IL 61790 USA
amishra@ilstu.edu

Tibor Gyires
School of Information Technology
Illinois State University
Normal IL 61790 USA
tbgyires@ilstu.edu

Yongning Tang
School of Information Technology
Illinois State University
Normal IL 61790 USA
ytang@ilstu.edu

***Abstract*—Random Key Predistribution Scheme (RKPS) guarantees any pair of neighboring nodes in a Wireless Sensor Network (WSN) can build a secure connection either directly if a common key found, or indirectly through a Path Key Establishment Mechanism (PKEM). When a sensor node resorts to PKEM to establish a secure connection with a neighboring node, it needs to broadcast a keyrequest to all securely connected nodes. However, unbounded broadcasting in PKEM can potentially cause unnecessary or duplicated broadcast message forwarding, which can intrusively incur disruptive power consumption on all involved sensor nodes in a highly resource constrained WSN. Such negative impact can be much worse if exploited by a malicious adversary to launch power exhaustion Denial of Service (DoS) attacks to sabotage a secured WSN. Thus, it is essential to convert unbounded broadcasting in PKEM to a nonintrusive broadcasting with optimally minimal message forwarding boundary in a WSN. Previous research empirically identified bounds on PKEM for small networks, which may not be suitable for densely deployed WSNs with much higher sensor node populations. In this paper, we tackle this problem by applying theoretical results to identifying the upper bound of diameter on a WSN when represented as a Erdős-Rényi random graph. We then verify the performance of a broadcast bounded PKEM through simulations. The performance evaluation shows the effectiveness of the optimally bounded PKEM.**

Keywords- sensor networks; random key predistribution; graph diameter; random graph; theoretical bound.

I. INTRODUCTION

A Wireless Sensor Network (WSN) comprises of a large population of inexpensive sensor nodes that form an ad-hoc wireless network to transmit information. The sensor nodes can be deployed in a large geographical area to detect or measure physical quantities such as temperatures, magnetic anomalies, chemicals or motion in their immediate environment. Several important WSN applications require to operate in a hostile environment, where adversaries may attempt to sabotage the WSN via different means, such as unhindered physical access, eavesdropping, message deception triggered exhaustive power consumption.

Securing WSNs is a highly demanded but also highly challenging task, especially when implemented in highly resource constrained sensor nodes. Conventional security mechanism based on public-key cryptography requires extensive computations that are infeasible on current sensor node platforms. Consequently, Symmetric key cryptography has been explored for securing WSNs due to its low computational requirement.

Random Key Predistribution Scheme (RKPS) [2] has been proposed and effectively used to secure WSNs. RKPS relies on predistributing a random subset of keys (keyring) from a large set of keys (keypool) on each sensor node before a WSN is deployed. RKPS uses a small keyring to achieve secure communication between a sensor node and all its neighboring nodes within its transmission range. However, the small keyring size can only be controlled to guarantee that a sensor node is able to authenticate and thus trust a small fraction of its neighboring nodes after its deployment. If the sensor node is unable to find a common key with a neighboring node, it will broadcast an authentication request (keyrequest) to all its trusted nodes who could either authenticate the incoming keyrequest or forward the request to their trusted sensor nodes for authentication. The process will continue until the keyrequest is authenticated.

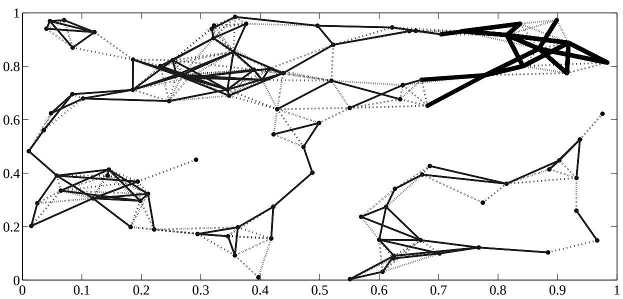


Figure 1: PKEM in progress over a RKPS secured WSN.

The keyrequest forwarding in RKPS is analogical to a flooding broadcast mechanism. Optimally bounding keyrequest broadcasting in Path Key Establishment Mechanism (PKEM) [2] is the focus of this paper. We discuss this mechanism further in Section III while detailing a general model of basic RKPS.

Figure 1 shows a RKPS enabled WSN, where (1) thick links (as shown on the upper right corner) represent indirect secure connections created by PKEM, (2) the solid lines represent direct secure connection between two neighboring sensor nodes with at least one common key, and (3) the dashed lines represent unsecured connections between two sensor nodes without a common key. Sensor nodes connected by dashed lines can resort to PKEM to authenticate each other and construct secure connections indirectly. The diagram plots the Cartesian coordinates of sensor node locations that uniformly distributed on a unit square to represent a sensor field.

RKPS can be used to secure a WSN using a limited keyring size with the cost of network communication overhead, which grows significantly in large-scale WSNs. Thus, it is desirable to control the communication overhead in PKEM, especially if the sensor nodes are allowed to forward a keyrequest unconditionally. For handling a keyrequest, a sensor node needs to consume its limited battery power for receiving, processing (e.g., searching common keys) and forwarding or acknowledging the request. In PKEM, a keyrequest can be initiated by any sensor node, which might be exploited by an adversary to inject bogus keyrequests. In response, a large number of sensor nodes in the attacked WSN will potentially consume a large amount of power to authenticate the bogus keyrequests being injected. Such an exploited weakness in PKEM can easily result to a Denial-of-Service (DoS) attack, which can easily exhaust the battery power on many involved sensor nodes.

The communication overhead in PKEM can be effectively controlled through the use of a Time-To-Live (TTL) parameter on each packet. However, the TTL value needs to satisfy multiple conflicting constraints. Firstly, the TTL needs to be sufficiently large to ensure that (1) PKEM can perform its intended function, and (2) every keyrequest can be forwarded to a sensor node that can authenticate it. On the other hand, this TTL should be as small as possible to limit the communication overhead in PKEMS. Furthermore, a mechanism should exist to enforce that the TTL value does not exceed a value decided before the deployment of a WSN. This is to ensure that an adversary should not be able to inject keyrequest into the WSN with arbitrarily (large) TTLs to launch the DoS attack. Finally, the TTL value should be adaptive for the varying size of a WSN due to newly added nodes or gradual death of existing nodes.

In this paper, we show how to model a secured WSN as a connected Erdős-Rényi random graph such that a keyrequest in RKPS originated from any sensor node can be guaranteed to reach every other sensor node within the WSN. The keyring and keypool sizes are chosen to ensure that a sensor node connects securely to its neighbor nodes with a predictable probability, which can further ensure the resulting graph is connected.

In our previous work [1], we introduced the problem of identifying the maximum TTL (MAXTTL) and applied the related results from Erdős-Rényi random graph theory to identify MAXTTL for a WSN with full-visibility. More specifically, the diameter of the Erdős-Rényi random graphs can be used to calculate the maximum value of the TTL. If one calculates the shortest paths between every pair of nodes within a random graph, the diameter would be the longest of these shortest paths. The full visibility case, discussed further in the next section describes a deployment of the sensor network where every sensor node can connect to any other sensor node within the network. In this paper we expand upon our conference paper to give more details about the problem, our simulation approach and results.

The rest of the paper is organized as the following. Section II discusses the related work. Section III provides the background of PKEM and derives the theoretical bound of flooding radius in PKEM. Section IV presents our simulation design and results. Finally, Section V and Section VI present some discussion and conclude the paper.

II. Related Work

In this section, we will review the work that has been proposed to obtain a suitable TTL

Basic RKPS was first introduced in [2] and its security characteristics have been extensively studied. A variety of schemes have built upon the basic RKPS by combining it with other key predistribution schemes for improving its resilience to node compromise. Reference [4] reviews the basic RKPS and its derivative schemes and also surveys the state-of-art in sensor network security. PKEM overhead applies to the basic RKPS and all its derivative schemes that trade network overhead for reduced keyring size. Since our work is fundamental to RKPS itself and also remains a component of its derivatives, we will work with a model of the Basic RKPS that includes all the elements of the scheme that have remained invariant in its derivative schemes also. Instead of reviewing the specifics of each scheme based on RKPS we dedicate this space to reviewing research that has investigated the characteristics of PKEM or has provided some guidance on its possible values.

The original work in [2] reported empirical observations that the keypath length did not exceed a constant number for the range of node populations between 1000 and 10000, in their simulations. However it did not provide analytical guidance on how PKEM will behave for larger node populations and different node neighborhoods. The first reference to use a TTL limited PKEM appears in [5], where it was recommended to set the keypaths based on the average path lengths in the trust graph. This conclusion was based on empirical observations on an experiment setup similar to the original work in [2]. It was also noted that a majority of the keypath lengths were much smaller than the observed average and the maximum. However, it did not analytically characterize the asymptotic behavior of the PKEM path-lengths and how it evolves with node population, deployment density or average node connectivity. Another contribution of [5] was the explicit statement of the assumptions related to the minimum degree of the underlying connectivity graph. The node connectivity in the original work and many of the derivative schemes has been assumed to be much higher than that supportable by state-of-art MAC layer protocols such as IEEE 802.15.4 Zigbee [11] standard, popular on several sensor node platform implementations.

Apart from empirical observations related to TTL, interesting progress has been made in the investigation of the validity of Erdős-Rényi Random Graphs [7] that discussed the application of graph theory to RKPS in the context of sensor networks and produced validating results for specific ranges of its parameters.

The work in [12] applies random graph theory to RKPS to propose a graph theoretic framework for parametric decision making for RKPS, optimal keyring size, and network transmission energy consumed in PKEM etc. It provides some analytical formulations on the basis of the diameter of de Bruijn's graphs [27] but did not provide any analytical guidance on how to find the diameter of a RKPS trust graph. Our work may be considered supplementary to this research since reliable estimation of the diameter of RKPS trust graph may allow exact derivation of some of the quantities mentioned in this work. Our work instead focuses on proposing supporting theory on the asymptotic bounds of the diameter of the sensor network configuration, which relates it to RKPS parameters such as node population, probability and node neighborhood.

We also note the recent theoretical investigation in Uniform Random Intersection Graphs that model Random Key Predistribution Scheme under the full-visibility assumption. Related work in [23]–[25] investigate several interesting properties of Uniform Random Intersection Graphs and formally prove its connectivity properties and node degree. Finally, we note that [22] and [26] study the diameter of Uniform Random Intersection Graph and solves a problem very similar to ours from a theoretical point of view. We could have chosen to base our analytical model on the basis of [22], however, the focus of this paper is inclined towards the investigation and extension of Erdős-Rényi random graph theory as applied to RKPS implementation on sensor networks.

For the construction of our simulator we used the guidance from [6] that discusses the construction of a high performance simulation for Key Predistribution Schemes for WSN in Java. Our experiment design replicates the experiment set up in [2], and we simulated node populations between 1000 to 10000 sensor

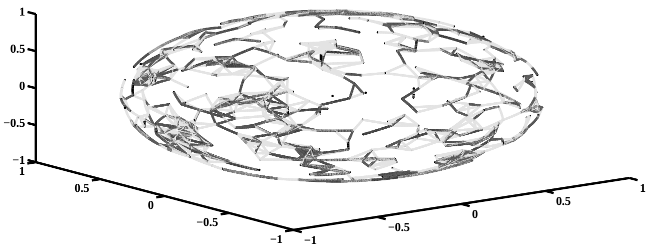


Figure 2: Trust Graph Illustration: Lighter edges represent wireless connectivity and darker edges represent secure connectivity.

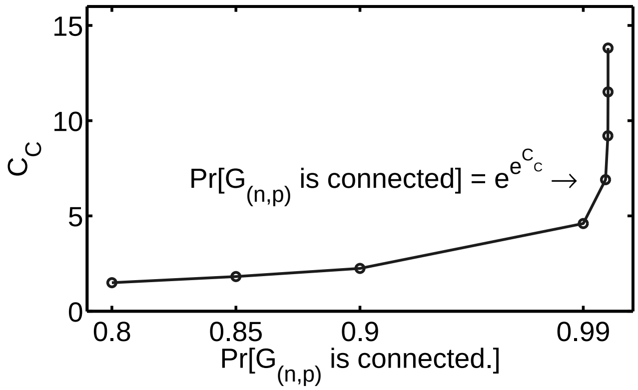


Figure 3: Variation of C_c for desired probability of graph connectivity in Eq. 3.

nodes. Finally, we used a number of open source software libraries for high precision arithmetic, graph theory algorithms and MATLAB for experiment design, statistical analysis and visualization. We discuss these in further detail in Section IV.

III. Theoretical Bound Analysis

For a given WSN with node population n, RKPS models it as a Erdős-Rényi random graph such that appropriate keyring size k and keypool size K can be selected to ensure the formed secure network remain connected.

A. Trust Graph

A sensor network can be modeled as a pair of overlaid graphs, where the underlying graph represents the wireless connectivity among sensor nodes, and the overlay graph shows the secured wireless connectivity among sensor nodes. In the underlying wireless connectivity graph, each sensor node is denoted as a vertex and each wireless connection between two neighboring nodes is represented as a link. In the overlay secured wireless connectivity graph, also referred as Trust Graph, only secured wireless connections remain and are represented as secured links. Figure 2 shows a trust graph overlaid on the top of the connectivity graph of a WSN deployed on the surface of a unit sphere. A trust graph is a sub-graph of the underlying wireless connectivity graph, since its edges only exist if an underlying connectivity graph edge exists. The figure plots the Cartesian coordinates of sensor node locations distributed on the surface of a model sphere, with unit radius. We discuss this modeling approach further in Section IV, Deployment Model.

B. Generalized RKPS Model

In this section, we describe the general model of basic RKPS [2].

Before a WSN is deployed, RKPS allocates a small random subset of keys (keyrings) on each sensor node from a large universal set of random keys (keypool), where each subset may overlap with other subsets with a small probability p. Once deployed, each sensor initiates a shared key discovery protocol with its neighboring nodes by sending a keyrequest containing unique identifiers for each key in its keyring. The neighboring node with common key will respond back by encrypting a random number with a common key (challenge), which will be decrypted by the requesting node and sent back (response) to complete the authentication process. Subsequently, the identified common key can be used to negotiate a shared session encryption key.

The small keyrings only allow a fraction of neighboring nodes to directly authenticate the received keyrequest. A sensor node that unable to authenticate a targeted neighboring node would resort to a path key establishment mechanism (PKEM), where it forwards the keyrequest to its authenticated and thus trusted neighboring nodes. These neighboring nodes would either authenticate the targeted node or forward it to their trusted neighboring nodes until a transitively trusted node authenticates the targeted neighboring node.

The deployment model of a WSN is generally assumed to be uniformly random, and thus the neighboring nodes of any given sensor node cannot be predicted. An Erdős-Rényi random graph is denoted as $G_{(n,p)}$, where n is the number of vertices and p

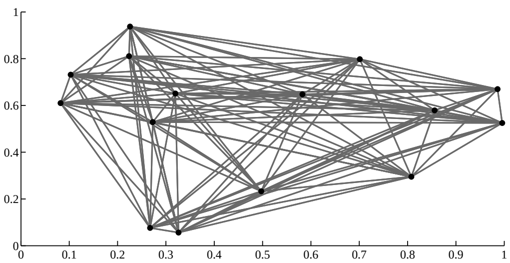


Figure 4: Sensor Network Model with full-visibility assumption.

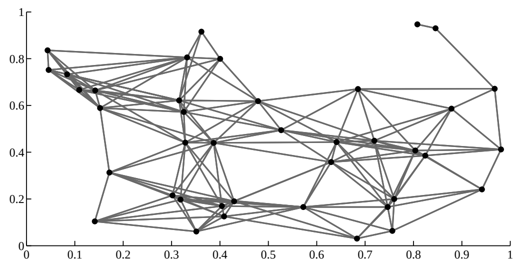


Figure 5: Sensor Network Model with limited-visibility assumption.

represents the probability that a vertex is connected to any other vertex within the graph.

For a random graph $G_{(n,p)}$, we have

$$if \quad p = \frac{\ln(n)}{n} + \frac{C_c}{n} \tag{1}$$

$$then \lim_{n\to\infty} P(G_{(n,p)} \ is\ connected) = e^{e^{-C_c}} \tag{2}$$

where C_c is a constant and should be chosen such that $P(G_{(n,p)} \ is\ connected)$ is close to 1.

Prior research [2] on RKPS has recommended choosing the value of C between 8 and 16, as shown in Figure 3 which can yield the desired value of p, and further derive the keyring size (k) for a given keypool size (K).

It is essential to note that the Erdős-Rényi graph theory assumes that within the graph any node can be connected to another one, i.e., every node can see any others within the network (full-visibility model). However, in sensor networks a sensor node is only connected to a small subset $n_a : n_a \ll n$ of the randomly deployed nodes that are within its transmission range (limited-visibility model). Figure 4 visually illustrates a WSN modeled under the full-visibility assumption in Erdős-Rényi random graph theory. Figure 5 shows a sensor network modeled under the limited-visibility, encountered in practical sensor network deployments. In both, Figure 4 and Figure 5, the nodes represented by the plotted points are only joined by edges if there is connectivity between them. The dimensions represent distance and the plot models the Cartesian coordinates of each sensor's deployment location on a two dimensional square sensor field. Note that in case of full-visibility Figure 4, each node is connected to every other node and its location is immaterial. In contrast, the location and neighborhood of a sensor in the limited-visibility Figure 5 case governs its connectivity with other sensors in the network.

In order to overcome the lack of connectivity of the limited-visibility case, the work in [2] proposed adjusting p to the effective probability (p_a), with which a node can connect to any of its neighboring nodes, such that the average degree d of the nodes in the graph remains constant as shown by Eq. 3.

$$d = (n_a - 1)p_a = np \tag{3}$$

With this calculated value of p_a, the work in [2] derived k according to the following equation:

$$p_a = 1 - \frac{(K-k)!^2}{K!\,(K-2k)!} \tag{4}$$

Research results identifying the upper bound on the random graph diameter with the parameter C_C controlled with the proposed range have been proposed in Theorem 4 in [8], where we have $p \geq c\ln(n)/n$.

C. Diameter of a Sparse Random Graph

Several studies have analytically investigated the upper bound of Erdős-Rényi random graphs for various ranges of n and p. For example, the work presented in [8] reviews the analytical results on various ranges of p, in terms of n. Moreover, it derives the asymptotic bounds on the diameter of Erdős-Rényi random graphs at its critical threshold where both n and p satisfy the relationship mentioned in Eq. 1. Theorem 3 in [8] states that given the relationship in Eq. 5, the diameter of the graph is concentrated on at most three values around value indicated in Eq. 6.

$$\frac{np}{\ln n} = c \geq 2 \tag{5}$$

$$diam(G_{(n,p)}) \leq \lceil \frac{\ln n}{\ln np} \rceil + 1 \tag{6}$$

Furthermore, for ranges of $c \leq 2$, Theorem 4 [8] predicts the upper bound on the diameter indicated by Eq. 7 as follows.

$$\left\lceil \frac{\ln(\frac{cn}{11})}{\ln(np)} \right\rceil \leq diam(G_{(n,p)}) \leq \left\lceil \frac{\ln(\frac{33c^2}{400})n\ln(n)}{\ln(np)} \right\rceil + 2\left\lfloor \frac{1}{c} \right\rfloor + 2 \quad (7)$$

The above formula gives the upper-bound on the diameter of sparse random graph, where $p \geq c\ln(n)/n$. It worths noting that c in Eq. 7 can be greater than 2 since it depends upon the value of C_c as shown in Eq. 1. Figure 7 shows how c varies with the value of C_c. A sufficiently high value of C_c can be chosen to get $c \geq 2$. Our experiment design takes this precondition into account to interpret the observed results.

IV. Simulation Design and Results

To investigate the effective diameter of a trust graph for the RKPS configurations discussed in Section III, we created a sensor network simulator along the directions discussed in [11]. Our simulator derives the keying size based on [2] as discussed in Eq. 4, and allows for reasonable variations in the sensor network deployment densities.

A. Simulation Model

The simulation has been constructed in MATLAB and Java. The implementation of the WSN model in Java allows us to take advantage of efficient thread-safe data structures to model a WSN, sensor keyrings and sensor nodes. We have implemented several different experiments using MATLAB to collect the data from Java simulation and interpret it to construct the result visualizations. With the support of several MATLAB toolboxes, we were able to implement efficient graph algorithms for the calculation of all-pair shortest paths and the size of a WSN.

As shown in Figure 6, our simulation model comprises of a MATLAB driver script that calls multiple MATLAB functions to prepare simulation arguments and pass them to the Java simulation model. At the completion of the simulations, the state of the Java simulation is imported into Matlab in the form of an adjacency matrix representation of the trust graph and the connectivity graph.

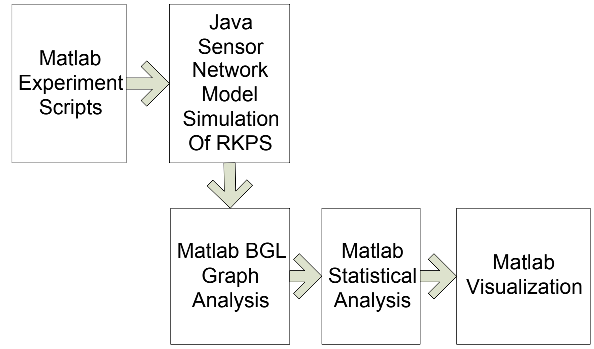


Figure 6: Simulation Model.

Adjacency Matrix is a widely used standard representation of a graph in many graph theory algorithms. Informally, the adjacency matrix is a square symmetric matrix with n rows and n columns, where n represents the number of nodes within the graph. Each row corresponds to a node within the graph and contains a value of 1 for each edge that connects that node to another node in the graph. The adjacency matrix can be used as input for the all-pair shortest path graph algorithms available in MATLAB-BGL [20].

MATLAB-BGL provides a MATLAB wrapper for the standard Boost Graphics Library (BGL) [9]. BGL is a comprehensive C++ library implementing almost all known efficient graph algorithms. BGL is well-established and well-reviewed by the development community, and becomes the standard library used for graph theory calculations.

The diameter of a trust graph is the longest path among all shortest paths between any pair of vertices in the graph. Various so-called all-pair shortest path algorithms allow us to measure the shortest number of hops between every pair of sensor nodes in the network. The choice of a suitable algorithm to identify all-pairs shortest paths depends upon the sparseness of the graph and the space complexity of an algorithm with a given graph size.

Assuming V to be the number of vertices and E to be the number of edges in a graph respectively, the complexity of Johnsons [18] all-pair shortest path algorithm is $O(V^2 \ln V + VE)$. Floyd-Warshall [17] provides another approach with the complexity of $O(V^3)$, which is independent of the number of edges. Floyd-Warshall algorithm is generally well-suited for dense graphs with a large number of edges. However, it requires matrix multiplication that is relatively difficult to implement for large-scale networks utilizing built-in data types of common programming languages like C++ and Java. Johnsons algorithm on the other hand has lower space complexity and is therefore well-suited for calculation of the all-pair shortest paths for large-scale networks, even when the number of calculations required for dense graphs may be higher than for Floyd-Warshall. We utilized BGL implementation of Johnsons all-pair shortest paths to calculate the diameter of networks in the range of 1000 to $10,000$ nodes.

We designed experiments with various ranges of n and C_c to validate whether the obtained trust graph from simulations follows the theoretical results. We generated random topologies for creating WSNs by varying the number of nodes from 1000 to 10000. The required probability to obtain full connectivity were calculated based on Eq. 1 and Eq. 3.

B. Verification of Theoretical Results

We utilized the Boost Graph Library [9] and Matlab-BGL [20] toolkit for MATLAB to verify the theoretical results on several instances of random graphs for various values of n when $1 \leq c \leq 2$. Our simulation results as presented below confirm the theoretical results as shown in Section III based on full visibility RKPS models with a keypool size of 100000.

Please note that the diameter values remain relatively stable for large increments of n, which should allow the future extension of a WSN, even with the current controlled diameter. We also observe that the observed value is well-below the value predicted by the theory, which would make it robust against transmission failures in the shortest path.

As discussed earlier, most empirical studies of RKPS have assumed a value of C_c in the range of 8 to 16. Figure 7 and Figure 8 plot the value of c as in Eq. 5 and Eq. 6 showing that c can be assumed to be higher than 2 for lower ranges of n and higher ranges of C_c as in Eq. 1. These values are coincident with the range assumed in prior research on RKPS schemes.

The value of C_c in Eq. 1 has significant impact upon weather c in Eq. 5 is in a range where the diameter of the random graph remains $O(\ln(n)/\ln(np))$. Figure 7 implies that lower values of C_c in Eq. 1 will not allow the diameter of the graph to remain small.

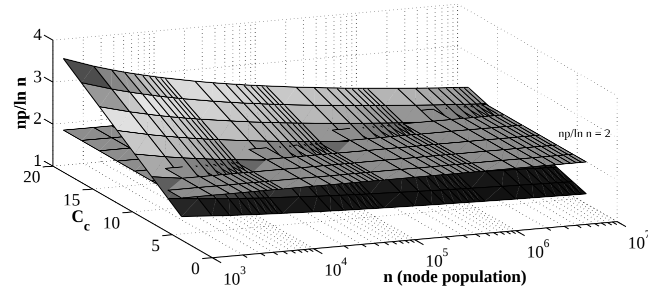


Figure 7: Plot of $np/\ln(n)$ showing the value of c in Eq. 5 for various ranges of n and C_c.

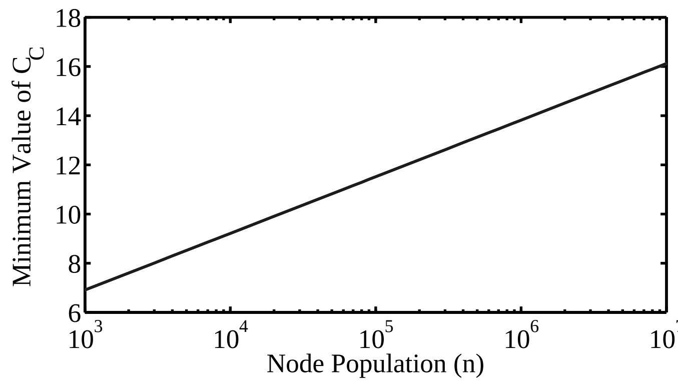


Figure 8: Plot showing $C_n \geq \ln(n)$, values of C_c where $np/\ln(n) = 2$.

C. Keyring Size Calculation

The keyring size was calculated for a keypool of 100000 on the basis of [2] as described by Eq. 5. To work with factorials of large numbers as required by Eq. 5, we utilized the JScience scientific library to support numbers of arbitrary precision. Further, to improve the precision and the performance of our simulations, we created a symbolic fraction that allows canceling of factors in numerator and denominator of a fraction before calculation of its final value. This reduces the loss of precision due to floating point arithmetic operations and enabled us to reproduce the exact calculations for keyring sizes as published in [2]. Since the presence of factorials in Eq. 5 does not permit further simplification for obtaining keyring size through a formula, we utilize a hit and trial approach for calculation of the keyring size for a given keypool size.

To speed up the simulations, we observe that the increase in keyring size would result in higher probability of connectivity. We exploit this monotonicity

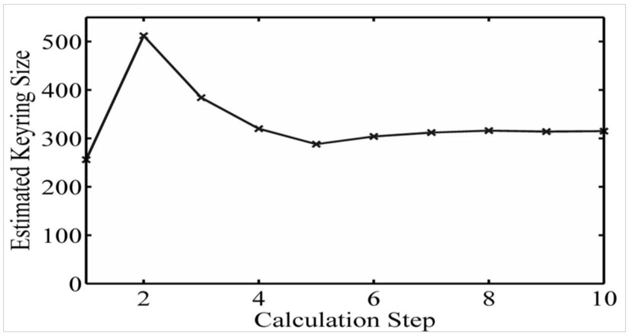


Figure 9: Calculation steps for a keypool size of 10^5 with desired probability of 0.63.

of Eq. 5 to devise a simple binary search to converge on the keyring size with the least number of trials. Figure 9 shows the calculation steps required for typical values of probability and keypool size. Firstly, we rapidly increase keyring sizes in steps of 256 to obtain a probability higher than the required; then we gradually alternate between increasing and decreasing the keyring size to converge upon the size of the keyring, which would give the exact desired probability of connectivity for a given keypool size.

D. Deployment Model

Basic RKPS simulation model uses a unit square as the deployment area. Node density was assumed to be uniform and simulated by varying the transmission range of the sensor node model. More recent work in [11] showed the boundary effect in the context of simulating key predistribution schemes for WSNs. Boundary effect occurs at the borders of the sensor network, where the sensor nodes do not have the average neighborhood connectivity as available to nodes closer to the center. Boundary Effect can significantly influence the degree distribution of the trust graph in simulations but its impact in practical deployments is considerably less as the network grows larger. A recommended elimination [11] of this effect is to modify the WSN deployment model on a spherical surface, which results in a uniformly distributed node population in a WSN. Eliminating the boundary effect also allows us to produce a sensor network model with homogeneous node connectivity, which can be further mitigated if the boundary nodes resort to dynamic range extension as suggested by [5].

To simulate a spherical deployment field, we followed the directions from the work in [7], and modeled our node deployment using Ziggurat method due to Marsaglia [13]. This method allowed us to generate a uniform distribution of three dimentional points on the surface of a sphere. We calculated the node distances using the great circle arc length, with assumption that the node range is a disk shaped area on the surface of the sphere. This is equivalent to the transmission range of a sensor node on a planner surface.

V. Results and Discussion

Figure 10 shows a plot of our simulations on MATLAB, where the diameter of the generated random graph closely follows the the theoretical expectation as described in Section II. The theoretical predictions of the figure is a composite generated on the basis of Eq. 6 and Eq. 7. For the points that satisfy the precondition in Eq. 5 we have used Eq. 6 and Eq. 7 for the points for the others. As shown in Figure 10 the practical diameter of the trust graph is co-incident to the theoretical expectations with an error of ± 1.

Figure 11 and Figure 12 show the long range predictions of the analytical tools that we have discussed in this paper. The predict shows that the diameter of a WSN will increase very slowly with the increase of network size, and will remain constant for large ranges of node populations. This further shows that setting a maximum limit to the TTL employed by PKEM will not interfere with the extensibility of the sensor network. More sensor nodes can be deployed later with the same TTL setting to continue operation of the network. Moreover, the stability of TTL for largely varying network sizes also show that the network will be robust against failures or compromise of a large percentage of sensor nodes, and PKEM operation will not be impacted by a limited TTL. On the other hand, this also indicates that controlling TTL would only provide limited control over the number of nodes visited by a keyrequest and the consequent power consumption of PKEM. The number of nodes that may receive a PKEM request rises rapidly with the increment of TTL in a large-scale WSN.

Figure 12 shows the node degrees may rise as high as 140, which is prohibitively high for current sensor node platforms. We notice that several methods have been proposed to mitigate this problem, including range extension [5]. Further investigation of the diameters of practical sensor network deployments should be undertaken using simulations and analytical models,

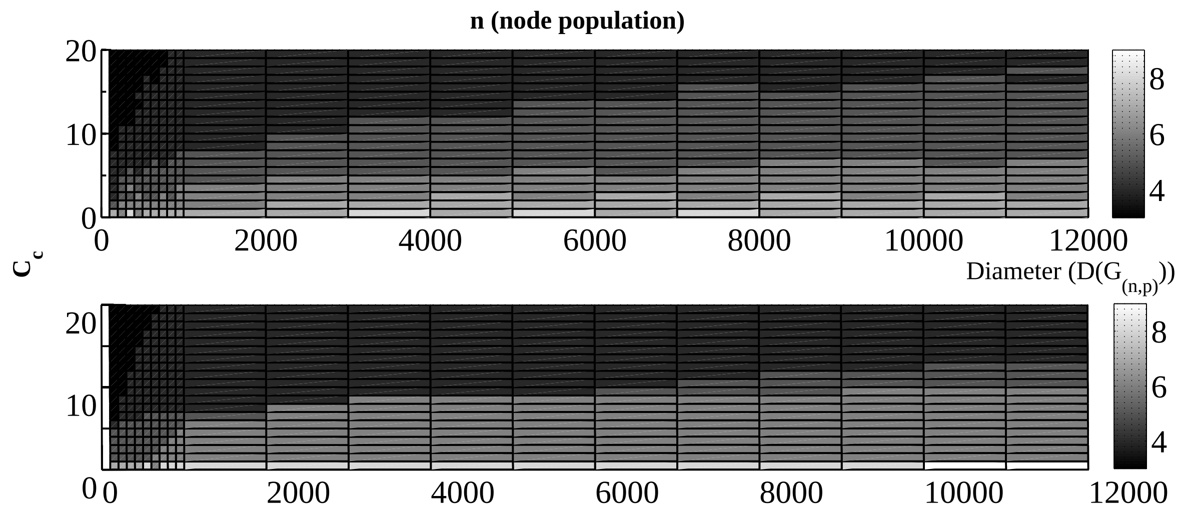


Figure 10: The comparison of practical and theoretical graph diameter.

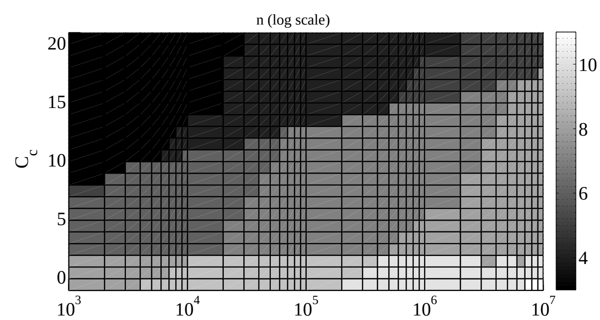


Figure 11: Log scale plot of diameter for large network sizes n and C_c.

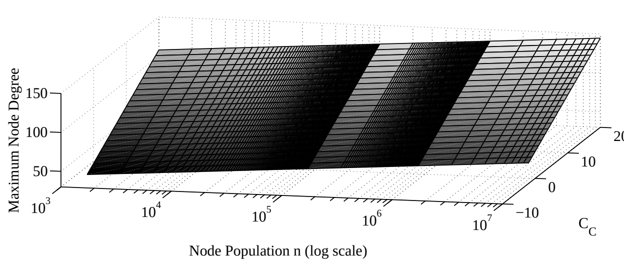


Figure 12: Log scale plot of maximum node degree for large network sizes n as predicted by [10].

which specifically address the limited-visibility sensor network deployment model, such as Random Geometric Graphs(RGG) [16].

VI. Conclusion and Further Work

To conclude, this paper formally studies the communication overhead of PKEM and its possible improvement through Erdős-Rényi random graph theory. PKEM employs a variant of flooding broadcasting and specifically an instance of probabilistic broadcasting [19]. We have shown that the theory on the diameter of the Erdős-Rényi random graph can be used to limit the overhead of the PKEM without impacting its function in RKPS. While we have focused on PKEM specifically, the key revocation protocol for RKPS also relies on broadcasting. Thus, our results can be directly applied to limit the overhead of the key revocation protocol also.

We have presented and tested an analytical model that provides a simplified guidance on the TTL setting in PKEM for sensor network deployments under full visibility setting. We have shown that certain assumptions regarding the modeling of the trust graph are necessary to preserve its properties as applicable in an Erdős-Rényi random graph. Lastly, we have studied the predictions of our analytical model for large scale deployment and identified their impact on the feasibility of large scale sensor networks.

In this paper we have studied the solution of the MAXTTL problem for the full-visibility case where a sensor can potentially communicate (see) with any other sensor within the network. A majority of practical sensor network deployment confirm to the limited visibility case where a sensor can only communicate (see) other sensor nodes within its transmission range. We intend to extend this work for further for practical sensor network deployments under the limited visibility assumption. Recent work in modeling practical sensor networks deployments have utilized Random Geometric Graph theory. We intend to explore the theoretical results on Random Geometric Graphs to find guidance on the diameter of a RKPS trust graph under the limited visibility assumption.

We chose this problem to trigger a discussion of the energy consumption of RKPS when PKEM transmissions are also taken into account. Optimally controlling the transmission overhead in RKPS is critical to its eventual success as a security scheme for WSNs. Competing public-key cryptography schemes generally require much smaller number of transmissions, and may eventually become viable on somewhat more powerful sensor node platforms. Finally, bounded keyrequest broadcasting and methods to securely limit its overhead in RKPS are essential to mitigate adversarial DoS attacks. These DoS attacks are not defendable because RKPS cannot identify weather a keyrequest originates from an authentic sensor node or an adversary. For achieving this function, it would require an authentication scheme that is at least as secure as RKPS, preferably with lesser overhead.

Randomized broadcasting (or gossiping) has been considered as another method to lower the transmission complexity of RKPS, and may be more suitable for implementation on some sensor network configurations and node populations. However, PKEM based on randomized broadcasting trades latency and reliability for lower transmission complexity. Finally, we expect to provide a skeleton of theoretical assumptions, which may facilitate the application of results in Erdős-Rényi random graph theory to the problem of broadcasting at large, and the application adopting its upper bound on diameter in bounding the TTL values for flood broadcasting at large.

Our work also shows that the secure connectivity and diameter of the trust graph is intimately related to the deployment density of a WSN, and the average node connectivity. A poorly connected graph would result in a sparser trust graph, and may result in unreliable operation of PKEM with a limited TTL. Sparser trust graphs may require PKEM to broadcast packets with higher TTL values that exposes it to undesired potential DoS attacks. A predefined upper bound and sensor network configuration with a smaller diameter would effectively prevent an adversary from exploiting this mechanism.

References

[1] A. Mishra, T. Gyires and Y. Tang, ”Towards A Theoretically Bounded Path Key Establishment Mechanism in Wireless Sensor Networks,” in Proceedings of the IARIA ICN 2012, The Eleventh International Conference on Networks, Saint Gilles, Reunion, 2012.

[2] L. Eschenauer and V. D. Gligor, ”A key-management scheme for distributed sensor networks,” in Proceedings of the 9th ACM conference on Computer and communications security, Washington, DC, USA, 2002.

[3] P. Erdős and A. Rényi, ”On the evolution of random graphs,” Magyar Tud. Akad. Mat. Kutat Int. Kzl, vol. 5, pp. 17-61, 1960.

[4] Y. Xiao, V. K. Rayi, B. Sun, X. Du, F. Hu, and M. Galloway, ”A survey of key management schemes in wireless sensor networks,” Comput. Commun., vol. 30, 2007, pp. 2314-2341.

[5] J. Hwang and Y. Kim, ”Revisiting random key predistribution schemes for wireless sensor networks,” in Proceedings of the 2nd ACM workshop on Security of ad hoc and sensor networks, Washington DC, USA, 2004, pp. 43-52.

[6] T. M. Vu, R. Safavi-Naini, and C. Williamson, ”On applicability of random graphs for modeling random key predistribution for wireless sensor networks,” in Proceedings of the 12th International Conference on Stabilization, Safety, and Security of Distributed Systems, NewYork, NY, USA, 2010.

[7] V. Tuan Manh, W. Carey, and S.-N. Reihaneh, ”Simulation modeling of secure wireless sensor networks,” in Proceedings of the Fourth International ICST Conference on Performance Evaluation Methodologies and Tools, Pisa, Italy, 2009, pp. 1-10.

[8] F. Chung and L. Lu, ”The Diameter of Sparse Random Graphs,” Advances in Applied Mathematics, vol. 26, pp. 257-279, 2001.

[9] J. G. Siek, Lie-Quan Lee, and Andrew Lumsdaine. ”The Boost Graph Library: User Guide and Reference Manual”: C++ In-Depth Series. Addison-Wesley Professional, December 2001.

[10] R. Durrett, Random Graph Dynamics. New York, NY: Cambridge University Press 2006.

[11] J. Zheng and M. J. Lee, ”A comprehensive performance study of IEEE 802.15. 4,” Sensor network operations, 2004, pp. 218-237.

[12] A. C. F. Chan, ”A graph theoretic approach for optimizing key pre-distribution in wireless sensor networks,” in Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks, 2009. WiOPT 2009. 7th International Symposium on, 2009, pp. 1-8.

[13] G. Marsaglia and W. Tsang, The Ziggurat method for generating random variables, 2000.

[14] B. Bollobás, Random Graphs: Academic Press, London, 1985.

[15] C. Blundo, A. De Santis, A. Herzberg, S. Kutten, U. Vaccaro, and M. Yung, ”Perfectly-secure key distribution for dynamic conferences,” in Advances in cryptologyCRYPTO92, 1993, pp. 471-486.

[16] M. Penrose, Random geometric graphs: Oxford University Press, 2003.

[17] M. J. Atallah and M. Blanton, Foundations of Algorithms and Theory of Computation: Taylor & Francis, 2009.

[18] Paul E. Black, ”Johnson’s Algorithm”, Dictionary of Algorithms and Data Structures, National Institute of Standards and Technology 2004[12/01/2012].

[19] Y. Sasson, et al., ”Probabilistic broadcast for flooding in wireless mobile ad hoc networks,” 2003, pp. 1124-1130 vol. 2.

[20] D. Gleich, ”MatlabBGL. A Matlab Graph Library,” Institute for Computational and Mathematical Engineering, Stanford University. Available: http://www. stanford. edu/ dgleich/programs/matlab˙bgl, 2008[12/01/2012].

[21] P. Roberto Di, V. M. Luigi, M. Alessandro, P. Alessandro, and R. Jaikumar, ”Redoubtable Sensor Networks,” ACM Trans. Inf. Syst. Secur., vol. 11, pp. 1-22, 2008.

[22] K. Rybarczyk, ”Diameter, connectivity, and phase transition of the uniform random intersection graph,” Discrete Mathematics, vol. 311, pp. 1998-2019, 2011.

[23] M. Karosńki, et al., ”On Random Intersection Graphs: The Subgraph Problem,” Comb. Probab. Comput., vol. 8, pp. 131-159, 1999.

[24] S. R. Blackburn and S. Gerke, ”Connectivity of the uniform random intersection graph,” Discrete Mathematics, vol. 309, pp. 5130-5140, 2009.

[25] M. Deijfen and W. Kets, ”Random intersection graphs with tunable degree distribution and clustering,” Probab. Eng. Inf. Sci., vol. 23, pp. 661-674, 2009.

[26] O. Yağan and A. M. Makowski, ”Random Key Graphs Can They be Small Worlds?,” in Networks and Communications, 2009. NETCOM’09. First International Conference on, 2009, pp. 313-318.

[27] N. G. de Bruijn and P. P. Erdős, ”A combinatorial problem,” Koninklijke Netherlands: Academe Van Wetenschappen, vol. 49, 1946, pp. 758-764.

iaria journals

www.iariajournals.org

International Journal On Advances in Intelligent Systems

ICAS, ACHI, ICCGI, UBICOMM, ADVCOMP, CENTRIC, GEOProcessing, SEMAPRO, BIOSYSCOM, BIOINFO, BIOTECHNO, FUTURE COMPUTING, SERVICE COMPUTATION, COGNITIVE, ADAPTIVE, CONTENT, PATTERNS, CLOUD COMPUTING, COMPUTATION TOOLS, ENERGY, COLLA, IMMM, INTELLI, SMART, DATA ANALYTICS

issn: 1942-2679

International Journal On Advances in Internet Technology

ICDS, ICIW, CTRQ, UBICOMM, ICSNC, AFIN, INTERNET, AP2PS, EMERGING, MOBILITY, WEB

issn: 1942-2652

International Journal On Advances in Life Sciences

eTELEMED, eKNOW, eL&mL, BIODIV, BIOENVIRONMENT, BIOGREEN, BIOSYSCOM, BIOINFO, BIOTECHNO, SOTICS, GLOBAL HEALTH

issn: 1942-2660

International Journal On Advances in Networks and Services

ICN, ICNS, ICIW, ICWMC, SENSORCOMM, MESH, CENTRIC, MMEDIA, SERVICE COMPUTATION, VEHICULAR, INNOV

issn: 1942-2644

International Journal On Advances in Security

ICQNM, SECURWARE, MESH, DEPEND, INTERNET, CYBERLAWS

issn: 1942-2636

International Journal On Advances in Software

ICSEA, ICCGI, ADVCOMP, GEOProcessing, DBKDA, INTENSIVE, VALID, SIMUL, FUTURE COMPUTING, SERVICE COMPUTATION, COGNITIVE, ADAPTIVE, CONTENT, PATTERNS, CLOUD COMPUTING, COMPUTATION TOOLS, IMMM, MOBILITY, VEHICULAR, DATA ANALYTICS

issn: 1942-2628

International Journal On Advances in Systems and Measurements

ICQNM, ICONS, ICIMP, SENSORCOMM, CENICS, VALID, SIMUL, INFOCOMP

issn: 1942-261x

International Journal On Advances in Telecommunications

AICT, ICDT, ICWMC, ICSNC, CTRQ, SPACOMM, MMEDIA, COCORA, PESARO, INNOV

issn: 1942-2601