International Journal on Advances in Security

2020 vol. 13 nr. 1&2

The *International Journal on Advances in Security* is published by IARIA.
ISSN: 1942-2636
journals site: http://www.iariajournals.org
contact: petre@iaria.org

Responsibility for the contents rests upon the authors and not upon IARIA, nor on IARIA volunteers, staff, or contractors.

Reference should mention:

> *International Journal on Advances in Security, issn 1942-2636*
> *vol. 13, no. 1 & 2, year 2020, http://www.iariajournals.org/security/*

Reference to an article in the journal is as follows:

> *<Author list>, "<Article title>"*
> *International Journal on Advances in Security, issn 1942-2636*
> *vol. 13, no. 1 & 2, year 2020, <start page>:<end page> , http://www.iariajournals.org/security/*

IARIA journals are made available for free, proving the appropriate references are made when their content is used.

Sponsored by IARIA
www.iaria.org

International Journal on Advances in Security

Volume 13, Number 1 & 2, 2020

Zhixiong Chen, Mercy College, USA
Clelia Colombo Vilarrasa, Autonomous University of Barcelona, Spain
Peter Cruickshank, Edinburgh Napier University Edinburgh, UK
Nora Cuppens, Institut Telecom / Telecom Bretagne, France
Glenn S. Dardick, Longwood University, USA
Vincenzo De Florio, University of Antwerp & IBBT, Belgium
Paul De Hert, Vrije Universiteit Brussels (LSTS) - Tilburg University (TILT), Belgium
Pierre de Leusse, AGH-UST, Poland
William Dougherty, Secern Consulting - Charlotte, USA
Raimund K. Ege, Northern Illinois University, USA
Laila El Aimani, Technicolor, Security & Content Protection Labs., Germany
El-Sayed M. El-Alfy, King Fahd University of Petroleum and Minerals, Saudi Arabia
Rainer Falk, Siemens AG - Corporate Technology, Germany
Shao-Ming Fei, Capital Normal University, Beijing, China
Eduardo B. Fernandez, Florida Atlantic University, USA
Anders Fongen, Norwegian Defense Research Establishment, Norway
Somchart Fugkeaw, Thai Digital ID Co., Ltd., Thailand
Steven Furnell, University of Plymouth, UK
Clemente Galdi, Universita' di Napoli "Federico II", Italy
Birgit Gersbeck-Schierholz, Leibniz Universität Hannover, Germany
Manuel Gil Pérez, University of Murcia, Spain
Karl M. Goeschka, Vienna University of Technology, Austria
Stefanos Gritzalis, University of the Aegean, Greece
Michael Grottke, University of Erlangen-Nuremberg, Germany
Ehud Gudes, Ben-Gurion University - Beer-Sheva, Israel
Indira R. Guzman, Trident University International, USA
Huong Ha, University of Newcastle, Singapore
Petr Hanáček, Brno University of Technology, Czech Republic
Gerhard Hancke, Royal Holloway / University of London, UK
Sami Harari, Institut des Sciences de l'Ingénieur de Toulon et du Var / Université du Sud Toulon Var, France
Daniel Harkins , Hewlett Packard Enterprise, USA
Ragib Hasan, University of Alabama at Birmingham, USA
Masahito Hayashi, Nagoya University, Japan
Michael Hobbs, Deakin University, Australia
Hans-Joachim Hof, Munich University of Applied Sciences, Germany
Neminath Hubballi, Infosys Labs Bangalore, India
Mariusz Jakubowski, Microsoft Research, USA
Ravi Jhawar, Università degli Studi di Milano, Italy
Dan Jiang, Philips Research Asia Shanghai, China
Georgios Kambourakis, University of the Aegean, Greece
Florian Kammueller, Middlesex University - London, UK
Sokratis K. Katsikas, University of Piraeus, Greece
Seah Boon Keong, MIMOS Berhad, Malaysia
Sylvia Kierkegaard, IAITL-International Association of IT Lawyers, Denmark
Hyunsung Kim, Kyungil University, Korea
Geir M. Køien, University of Agder, Norway
Ah-Lian Kor, Leeds Metropolitan University, UK
Evangelos Kranakis, Carleton University - Ottawa, Canada
Lam-for Kwok, City University of Hong Kong, Hong Kong
Jean-Francois Lalande, ENSI de Bourges, France
Gyungho Lee, Korea University, South Korea
Clement Leung, Hong Kong Baptist University, Kowloon, Hong Kong
Diego Liberati, Italian National Research Council, Italy

International Journal on Advances in Security
Volume 13, Numbers 1 & 2, 2020

CONTENTS

Investigating the Creation of an Evolvable Firewall Rule Base and Guidance for Network Firewall Architecture, using the Normalized Systems Theory

Geert Haerens

Department of Management Information Systems
Faculty of Business and Economics
University of Antwerp, Belgium
and Engie IT — Dir. Architecture
Email: geert.haerens@engie.be

Herwig Mannaert

Department of Management Information Systems
Faculty of Business and Economics
University of Antwerp, Belgium
Email: herwig.mannaert@uantwerp.be

***Abstract*—A firewall is an essential network security component. It protects network connected company resources from potential malicious traffic. The firewall rule base, the list of filters to be applied to network traffic, can quickly become complex up to the point where companies consider the rule base as unmanageable. The complexity leads to unforeseen and painful side effects when the firewall rule base is changed (add/remove filtering rules). Sufficient literature exists on the root cause of rule base evolvability issues. However, little research is available on how to properly construct a rule base such that the evolvability issues do not occur. Normalized Systems (NS) theory provides proven guidance on how to create evolvable modular systems. In this paper NS is used to study the combinatorics involved when creating a firewall rule base. Based on those combinatorics, an artifact (method) is proposed to create a firewall rule base, that has evolvability in its design. As a network rarely contains only one firewall, the impact of different filtering strategies and changes on multiple firewalls, is studied as well.**

Keywords–Normalized Systems; Firewall; Rule Base; Filtering Strategies.

I. Introduction

This paper is an extended version of “Using Normalized Systems to Explore the Possibility of Creating an Evolvable Firewall Rule Base” [1] Firewalls are an essential component of network security. They have been protecting network-connected resources for over 25 years and will continue to do so for the next decades [2] [3]. Initially, firewalls were used to protect a company against threats coming from the outside (i.e., the “evil Internet”). Such kind of filtering is called North-South traffic filtering [4]. But security breaches are not only caused by access through the Internet. A significant portion of security breaches are caused from within the company network [5] where hacks have become more sophisticated. Getting a foothold on one resource on the internal network and from there on hopping between resources, is a known hacking strategy against which filtering North-South traffic offers no protection. For this reason, protecting the network-connected resources from internal traffic, referred to as East-West traffic [4], is gaining ground.

Networks are becoming more and more complex: they often contain multiple firewalls, which protect numerous network segments. The rule base of those firewalls (i.e., the definitions of which traffic is allowed or not) is becoming equally complex, up to the point where it becomes almost unmanageable. In a survey organized by Firemon [6], 73 % of survey participants stated that their firewall ranges from “somewhat complex” to “out of control”. Further, complexity is the highest-ranked challenge for firewall management [2] [3].

The firewall rule base is a classic example of a system that needs to evolve. It starts with one firewall, and two network segments and filtering rules between them. As the network grows, the number of resources connected to the network grows, the number of services offered on the network grows, and the number of security threats grows. The resulting firewall rule base will enlarge dramatically. This evolution will, at some point, result in a rule base where regular changes (i.e., the addition of a rule or the removal of a rule) result in unforeseen side effects. Those effects are proportional to the size of the rule base: the bigger the system (rule base), the worse it gets [2].

A network rarely contains only one firewall. Large companies have networks containing many firewalls. Valuable IT assets, located in data centers, are protected by multiple layers of firewalls. A single firewall can quickly become a non-evolvable system. Multiple firewalls only make the problem worse. Besides the question on how to create the correct rule and implement it on the rule base, one also has to decide on which firewall(s) this rule should be applied.

Normalized Systems (NS) theory [7]–[11] studies combinatorics in modular systems and provides a set of theorems to design modular systems exhibiting ex-ante proven evolvability. The goal is to avoid so-called combinatorial effects (CE). CE’s are impacts that are proportional to the type of change as well as the size of the system to which the change is applied. When all modules of a system respect the NS theorems, the system will be free of such CE’s. At that point, the system can be considered stable under change for a set of anticipated changes (such as adding and removing components from the system).

Multiple vendors sell tools to analyze a firewall rule base and can even be used to simplify it (e.g., Firemon, Tufin, Algosec). Some academic research on such analyses is available as well. Both industry and academics seem to focus on improving existing rule bases. However, a more ambitious objective would be to avoid this type of problem upfront through the deliberate design of the rule base and incorporate evolvability by design.

This paper will study the combinatorics involved in the firewall rule base. We will propose an artifact (a method), that translates the general NS theorems into a set of firewall rule base principles. When applied, this will result in an ex-ante proven evolvable (free of CE) rule base with respect to the addition and removal of rules to the firewall rule base.

We will start with a literature review and relate work. The remainder of the paper is structured according to the Design Science approach [12] [13]. Therefore, Section III starts by explaining some firewall basics and explains the evolvability issues of a firewall rule base. Section IV describes the artifact goals and design. The artifact is demonstrated (apply changes to a rule base) and evaluated in Section V. Section VI elaborates on different filtering strategies and Section VII will address the problems and possible solutions related to multiple firewalls. In Section VIII, automation and scaling of the propose solution is discussed and a link is made with the concept of Software Defined Network. In Section IX a part of the literature review is revised and weaknesses of the artifact are pointed out. Finally, Section X wraps up the paper and proposes future research.

This article builds on earlier research [11], where the applicability of NS for IT infrastructure systems was being explored. The current paper focuses on a practical case where NS and domain-specific knowledge on firewalls are combined, resulting in a design strategy for an evolvable firewall rule base and network firewall architecture.

II. Literature review and related work

The academic literature about firewalls can be divided into 3 groups. The first group (published roughly before the year 2000) focuses on the performance of the firewall and the hardware used to perform the actual package filtering. The second group (published roughly between 2000 and 2006) focuses on the complexity and issues with the rule base of the firewall. The third group (published roughly after 2006) focuses on the firewall in a Software Define Network (SDN) context, where distributed firewalls and software defined firewalls are used. As this paper focuses on the complexity and issues related to the firewall rule base, the following literature review will only focus on the second group of papers [14]–[25]. To the best of our knowledge, we did not find papers which specifically address and try to solve the evolvability issues of the firewall rule base. Next to academic papers, reports from Forrester and white papers from industry leaders were used as well [2]–[6], [26]–[28]. Those reports include surveys, which give information on the current state-of-affairs. One might think that, because academic publication about rule base issues have diminished after 2006, the problem is solved. However, the surveys provide a different view. Companies are still struggling with their firewall [2]–[6], [26]–[28]. This can be due to the "knowing-doing" gap or because the issue is not fully resolved.

Most papers start by stating that there is a problem with the firewall rule base because of:

- ***Translation issues***: how to convert a high level security policy into a low-level language of firewall rules [14]–[25] [26].
- ***Size of the rule base issues***: a large rule base is considered complex [6] [16] [20] [22] [23].
- ***Error and anomalies issues***: A rule base is error-prone due to complexity and manual interventions [2]–[6], [15], [16], [23], [26]–[28] and can contain firewall rule conflicts or anomalies [6], [14]–[16], [19], [21]–[23], [25], [27].

The ***"Translation-issue"*** is tackled by proposing tools, which could translate high level security concepts into low level firewall rules. FANG [19], FIRMATO [16], LUMETA [18] are artifacts proposed and described, which help translating high level security requirements into a low level firewall rule base. There are however no guarantees that these tools deliver a small and simple firewall rule base free of anamolies [16]. Companies such as TUFIN, ALGOSEC, FIREMON, VMWare also deliver commercial tools, which claim to help managing the complexity of network security. The tools do not prescribe, neither enforce how a rule base should be created in order to be free of anomalies and exhibit evolvability.

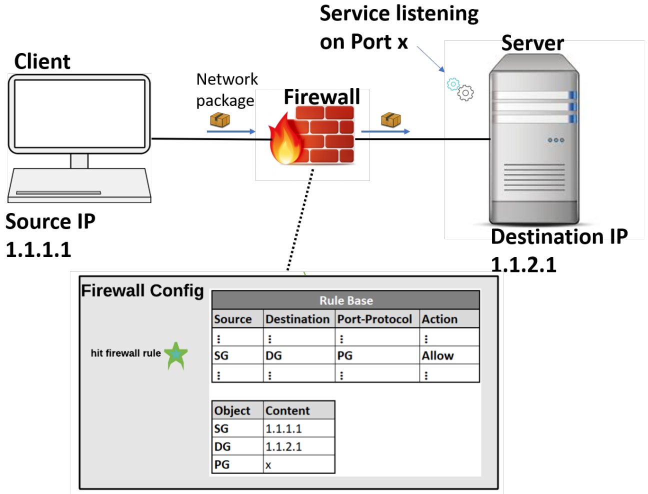
The ***"Size of the rule base issue"*** receives a lot of attention. Effort is put in reducing the rule base to a minimum list of rules, that still answer to the filtering requirements. The motivation for this "reduction of the rule base" is performance, although in [16] it was suggested that the actual size of the rule base is not related to the way the hardware actually applies the rules. This suggests a decorrelation between the size of the rule base and the firewall performance. This point will be revisited in Section IX.

The ***"Error issue"*** due to complexity and manual intervention is recognized and confirmed in recent surveys [2]–[6], [26]–[28]. The academic papers focus more on the technical root causes of the errors, being the anomalies in the rule base. Over time, the definitions of the types of anomalies, their formal definition and proof, have evolved and resulted in a definition of how a firewall rule base should look like in order to remain stable under change: a firewall rule base should only include disjoint rules [15] [21] [22] [23] [24] [25]. Artifacts have been put forward [15] [16], [20]–[22], [25], which allow to scan the rule base for non-disjoint rules and make them disjoint if required. The same artifacts allow to assess the impact of adding a new rule and adjusting the rules in such way that the rule base only contains disjoint rules. However, each time a rule is entered, the whole rule base needs to be scanned to detect potential anomalies between the existing rule base and the new rule. The effort of making a change to the system is thus proportional to the size of the system.

The literature review shows that the problems related to the firewall rule base are well known and the necessary condition to keep the rule base under control (i.e., having disjoint rules) is also known. However, clear architectural guidance on how to create a disjoint rule base as of the moment of conception, is lacking. It is exactly this architectural guidance, making use of NS, which is the main contribution of this paper. By structuring the rules in such a way that they are always disjoint, one can add and remove rules without having to analyze the rule base or worry about unforeseen side effects of the change.

III. General background and problem description

This section explains some fundamental concepts about firewalls, followed by a summary of the issues regarding the evolvability of a firewall rule base. The section continues by explaining the notion of firewall group objects, their value, and related issues. The section continues with a brief explanation of the Zero Trust (ZT) filtering strategy, which is one of the design objectives of the envisioned artifact, and terminates with an introduction to the Normalized Systems Theory.



Rule Base			
Source	Destination	Port-Protocol	Action
⋮	⋮	⋮	⋮
SG	DG	PG	Allow
⋮	⋮	⋮	⋮

Object	Content
SG	1.1.1.1
DG	1.1.2.1
PG	x

Figure 1. Firewall concepts

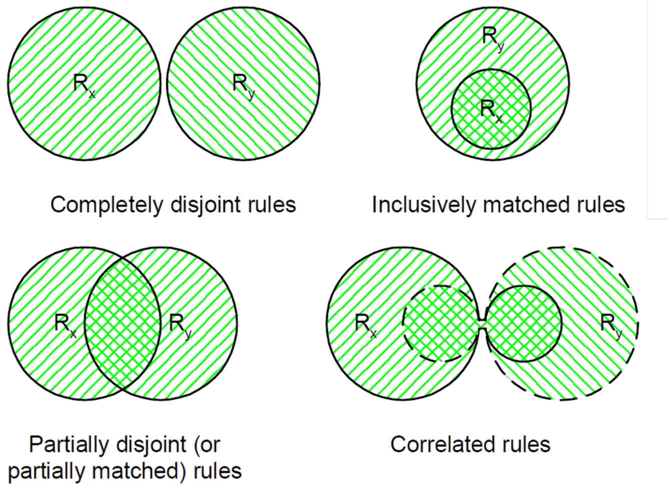


Figure 2. Possible relationships between rules (from [21])

A. *Firewall concepts*

An IP4 TCP/IP based firewall, located in the network path between resources, can filter traffic between the resources, based on the Layer 3 (IP address) and Layer 4 (TCP/UDP ports) properties of those resources [29] [30]. Filtering happens by making use of rules. A rule is a tuple containing the following elements: <Source IP, Destination IP, Destination Port, Protocol, Action>. IP stands for IP address and is a 32-bit number that uniquely identifies a networked resource on a TCP/IP based network. The rule is evaluated by the firewall, meaning that when it sees traffic coming from a resource with IP address =<Source IP>, going to resource =<Destination IP>, addressing a service listening on Port = <Destination port>, using Protocol = <Protocol>, then the firewall will perform an action = <Action>. The action can be "Allow" or "Deny". See Figure 1 for a graphical representation of the explained concepts.

A firewall rule base is a collection of order-sensitive rules. The firewall will evaluate all inbound traffic against the ordered rule base. The firewall starts at the top of the rule base until it encounters the first rule that matches the criteria (Source, Destination, Destination Port, Protocol) of the traffic. The firewall then performs the action as specified in the rule. In a firewall rule, <Source IP>, <Destination IP>, <Destination Port> and <Protocol> can be one value or a range of values. The protocol can be TCP or UDP. In the remainder of this document, the notion of protocol is omitted as it can be included in the Port variable (for example, TCP port 58 or UDP port 58).

B. *Firewall evolvability issues*

As a rule base changes over time, different rules start interfering with each other, resulting in complexity. In [15], the following relations are defined between rules:

- **Disjoint:** Two rules **R1** and **R2** are disjoint (completly or partially), if they have at least one criterion (source, destination, port) that has completely disjoint values (= no overlap or match).
- **Exactly Matching:** Two rules **R1** and **R2** are exactly matched, if each criterion (source, destination, port) of the rules match exactly.
- **Inclusively Matching:** A rule **R1** is a subset, or inclusively matched to another rule **R2**, if there exists at least one criterion (source, destination, port) for which **R1**'s value is a subset of **R2**'s value and for the remaining attributes, **R1**'s value is equal to **R2**'s value
- **Correlated**: Two rules **R1** and **R2** are correlated, if **R1** and **R2** are not disjoint, but neither a subset of the other.

Figure 2 represents the differnet relation in a graphical manner. Exactly matching, inclusively matching and correlated rules can result in the following firewall anomalies [15]:

- *Shadowing Anomaly*: A rule **R1** is shadowed by another rule **R2** if **R2** precedes **R1** in the policy, and **R2** can match all the packets matched by **R1**. The result is that **R1** is never activated.
- *Correlation Anomaly*: Two rules **R1** and **R2** are correlated if they have different filtering actions and **R1** matches some packets that match **R2** and **R2** matches some packets that **R1** matches.
- *Redundancy Anomaly*: A redundant rule **R1** performs the same action on the same packets as another rule **R2** so that if **R1** is removed the security policy will not be affected.

A fully consistent rule base should only contain disjoint rules. Disjoint rule are completely disjoint or partially disjoint. In that case, the order of the rules in the rule base is of no importance and the anomalies described above will not occur [15] [21]–[25]). However, due to several reasons such as unclear requirements, a faulty change management process, lack of organization, manual interventions, and system complexity [13], the rule base will include correlated, exactly matching, and inclusively matching rules. Combined with the order-sensitivity of the rule base, changes to the rule base (the addition or removal of a rule) can result in unforeseen side effects. To be confident that a change will not introduce unforeseen side effects, the whole rule base needs to be analyzed. Therefore, the impact of the change is proportional to the change and the size of the system, being the complete rule base. According to NS, this is a CE. As a result, a firewall rule base containing rules other than disjoint rules, is unstable under change.

C. *Firewall group objects*

A rule base made up of IP's as source/destination and port numbers is difficult to interpret by humans. It is just a bunch of numbers. Modern firewalls allow the usage of firewall objects, called groups, to give a logical name to a source, a destination, or a port, which is more human-friendly. Groups are populated with IP addresses or ports. Groups can be nested.

Using groups should improve the manageability of the firewall. But, using groups can easily result in the introduction of exactly matching, inclusively matching or correlated rules as well.

Example:
"Group_Windows_APP" and "Group_Windows_APPS" could be two groups with each contain the IP addresses of all Windows Application Servers. The latter may have been created without knowledge of the former [6], introducing exactly matching rules. The group memberships may start to deviate from each other, introducing correlated or inclusively matching rules, which could lead to anomalies in the rule base. The group structure must be well designed to avoid this.

D. *Zero Trust*

In [18] [19] [20] Forrester advocates the usage of a Zero Trust (ZT) model:

- Ensure all resources are accessed securely, regardless of location and hosting model,
- Adapt a "least privilege" strategy and strictly enforce access control,
- Inspect and log all traffic for suspicious activity.

The working assumption in the case of protecting network-connected resources is that all traffic towards those resources is considered a threat and must be inspected and secured. A network-connected resource should only expose those services via the network, which are minimally required. Also, each network connected resource should only be allowed access to what it needs.

E. *Introduction to Normalized Systems*

The Normalized Systems Theory [7]–[10] originates from the field of software development. There is a widespread belief in the software engineering community that using software modules decrease complexity and increases evolvability. It is also well known that one should strive towards "low coupling and high cohesion". The problem is that the community does not seem to agree on how exactly "low coupling and high cohesion" needs to be achieved and what the size of a module should be, to achieve low complexity and high evolvability.

The Normalized Systems Theory takes the concept of system theoretic stability from the domain of classic engineering to determine the necessary conditions a modular structure of a system must adhere to in order for the system to exhibit stability under change. Stability is defined as Bounded Input equals Bounded Output (BIBO). Transferring this concept to software design, one can consider bounded input as a certain amount of functional changes to the software and the bounded output as the number of effective software changes. If the amount of effective software changes is not only proportional to the amount of functional changes but also the size of the existing software system, then NS states that the system exhibits a Combinatorial Effect and is considered unstable under change. Normalized Systems Theory proves that, in order to eliminate Combinatorial Effects, the software system must have a certain modular structure, where each module respects four design rules. Those rules are:

- Separation of Concern (SoC): a module should only address one concern or change driver
- Separation of State (SoS): a state should separate the use of a module by another module during its operation
- Action Version Transparency (AVT): a module, performing an action should be changeable without impacting modules calling this action.
- Data Version Transparency (DVT): a module performing a certain action on a data structure, should be able to continue doing this action, even is the data structures has undergone change (add/remove attributes)

Only by respecting those rules, the system can infinity grow and still be able to incorporate new requirements.

Although NS originates in software design, the applicability of the NS principles in other disciplines such as process design, organizational design, accounting, document management, and physical artifacts. The theory can be used to study evolvability in any system that can be seen as a modular system and derive design criteria for the evolvability of such a system. In this paper, NS will be used to study the evolvability of the firewall rule base.

IV. Creating an artifact for an evolvable rule base

This section starts with investigating the modular structure of a firewall rule base, followed by a discussion of the issues that surface when the modular structure is instantiated. The section continues with a set of formal definitions of the firewall rule base components, from which the combinatorics are derived when creating a firewall rule base. The combinatorics are used to distill the design rules for the evolvable rule base. The design rules are translated into the actual artifact.

Based on the analysis of the problem space in the previous section, the objective is:

- To create a rule base compliant with the ZT concept.
- To create a rule base that contains only disjoint rules.
- To create a rule base, making use of firewall group objects to improve readability and manageability.
- To create a rule base that is evolvable for the following anticipated changes: the addition and removal of rules.

NS will be used to structure this evolvable rule base.

A. *Modular structure of the rule base*

A rule base is the aggregation of rules. A rule is an aggregation of Source, Destination, Service, and Action. Source is the aggregation of Clients requiring services. Destination is the aggregation of Hosts offering services. Service is the aggregation of Ports (combination of port number and protocol), which compose a service. Figure 3 represents the implicit modular data structure of a rule base in a firewall. Implicit because firewall vendors do not publish the internal data structure they use. The model corresponds with the type of information one needs to enter to create a rule in a firewall. Therefore, we assume that the model is a sufficient representation of a firewall rule base. In NS terms, the modular structure would be considered as evolvable when "Separation of Concern" is respected (the theorems "Separation of State" and "Data and

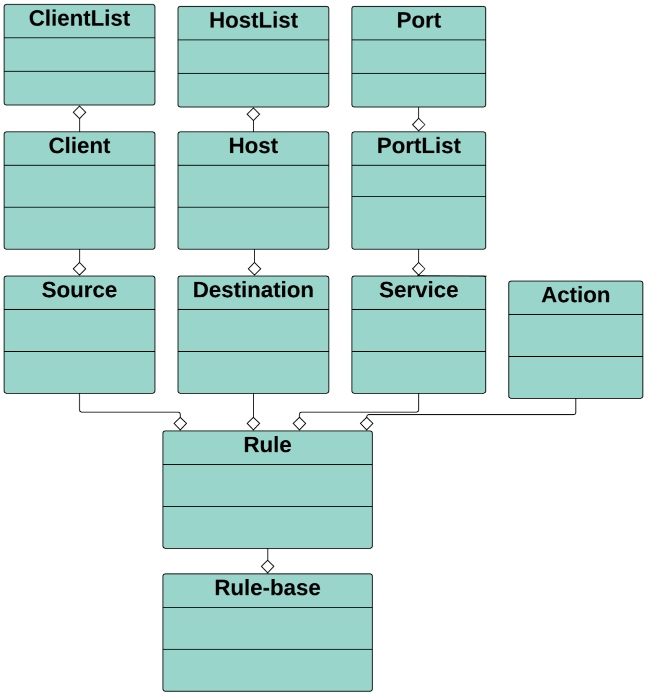


Figure 3. Modular Structure of a rule base

Action Version Transparency" are not relevant for the analysis of the rule base structure). As each of the mentioned modules focusses on one concern, one tends to conclude that the design of a rule base can be considered as stable under change.

B. Module instantiation

If the modular structure of the rule base seems to be stable under change, then where does the problem of non-evolvable rule bases comes from? In this respect, it is important to be aware that a firewall rule base is an order-sensitive system. More specifically, each instantiation of a rule must be given the correct place in the rule base, or the rule will have an impact on existing rules (see Section III). The order sensitivity is the root cause of the evolvability issues when the modular structure is instantiated. Indeed, it seems that —in some specific situations— certain evolvability issues of a modular structure only show up at instantiation time. Therefore, it is interesting to look at the application of the NS theorems at the instantiation level as well. In the context of this research, this would mean that we need to look whether the addition or removal of instantiations (of rules) can result in CE's, and thus evolvability issues, making an operational system unmanageable.

Eliminating the order-sensitivity of the rule base is the key to solving the problem. A firewall rule base should only contain disjoint rules. Disjoint rules have no coupling with other rules and are thus compliant with the "Separation of Concern" theorem of NS.

C. Formal definitions of rule base components

Let **N** represent a Layer 4 TCP/IP based network, in which 2 groups of network connected resources can be defined:

- The hosts, providing network services via TCP/IP ports.
- The clients, requiring access to the services offered by the host.

The network contains a firewall with configuration **F**, which is configured in a way that only certain clients have access to certain services on certain hosts. The ZT principle should be applied, meaning that clients have only access to those services on hosts they have been given explicit access to.

Let **Port** represent a Layer 4 TCP/IP defined port.

- Port.name = the name of the port.
- Port.protocol = the layer 4 TCP/IP protocol, being one of the following two values: TCP or UDP.
- Port.number = the number of the port, represented as an integer ranging from 1 to 2^{16}.

Let **P** represent the list of **Ports**, of length = pj .

$$\begin{cases} \mathbf{P}[1] \ldots \mathbf{P}[pj]. \\ \mathbf{P}[j] \text{ contains a } \mathbf{Port}. \\ 1 \leq j \leq pj. \end{cases}$$

Let **Service** represent a network service accessible via a list of layer 4 TCP/IP ports.

- **Service**.name = name of the service.
- **Service**.ports = list of ports = **P**.

Let **S** represent a list of **Services**, of length = sj.

$$\begin{cases} \mathbf{S}[1] \ldots \mathbf{S}[sj]. \\ \mathbf{S}[i] \text{ contains a } \mathbf{Service}. \\ 1 \leq i \leq sj. \end{cases}$$

Let **Host** represent a network host that provides services.

- **Host**.name = the Fully Qualified Domain Name (FQDN) of the network host.
- **Host**.IP = the IP address of the network host.

Let **H** represent a list of **Hosts**, of length = hj. The length of **H** is a function of the network **N**.

$$\begin{cases} \mathbf{H}[1] \ldots \mathbf{H}[hj]. \\ \mathbf{H}[k] \text{ contains a } \mathbf{Host}. \\ 1 \leq k \leq hj. \\ hj = f_h(\mathbf{N}) \end{cases}$$

Let **Client** represent a network client that requires access to hosted services.

- **Client**.name = the FQDN of the network client.
- **Client**.IP = the IP address of the network client.

Let **C** represent a list of **Clients**, of length = cj. The length of **C** is a function of the network **N**.

$$\begin{cases} \mathbf{C}[1] \ldots \mathbf{C}[cj]. \\ \mathbf{C}[l] \text{ contains a } \mathbf{Client}. \\ 1 \leq 1 \leq cj. \\ cj = f_c(\mathbf{N}) \end{cases}$$

Let **R** represent a firewall rule.

- **R**.Source = a list of Clients **Cs** of length = csj, where
  - ○ $1 \leq csj \leq cj$
  - ○ $\mathbf{Cs} \subset \mathbf{C}$
- **R**.Destination = a list of Hosts **Hd** of length = hdj, where
  - ○ $1 \leq hdj \leq hj$.
  - ○ $\mathbf{Hd} \subset \mathbf{H}$.
- **R**.Ports = a list of Ports = a Service **Sp**
  - ○ where $\mathbf{Sp} \in \mathbf{S}[sj]$.

- **R**.Action = either "Allow" of "Deny".

Let **F**, representing a list of rules **R** of length = fj, be the ordered firewall rule base **F**

- **F**[1] ... **F**[fj]
- **F**[m] contains a firewall rule **R**
- $1 \leq$ m $\leq$ fj
- **F** is order-sensitive. If **R**x is a firewall rule at location y in **F**, then the behavior of the firewall can be different if **R**x is located at position z instead of y, where z:1→ fj and z $\neq$ y. Whether or not the behavior is different depends on the relation **R**x has with the other rules of **F**.

D. Combinatorics

1) Ports: Port numbers are represented by 16-bit binary number and thus go from 1 to 2^{16}. Assuming that only TCP and UDP protocols are considered for OSI Layer 4 filtering, the possible number of values for Ports is equal to $2.2^{16} = 2^{17}$.

2) Services: **S** is the list of all possible services delivered via all ports exposed on the network **N**.
$\mathbf{S_{max}}$ is the largest possible list of services, with length = sj_{max}, in which all possible combinations of possible **Ports** are being used, where

$$sj_{\max} = \sum_{k=1}^{2^{17}} \binom{2^{17}}{k} \tag{1}$$

3) Hosts: The size of the list **H**, hj, is function of the network **N** and expressed as hj = f_h(**N**).
$\mathbf{H_{max}}$ is the list of all possible lists of hosts that are part of **H**. The length of this list is hj_{max}, where

$$hj_{\max} = \sum_{a=1}^{hj} \binom{hj}{a} \tag{2}$$

and where hj = f_h(**N**).

4) Services on Host: The maximum number of Hosts/Services combinations = $\text{hj}_{\text{max}}.\text{sj}_{\text{max}}$ =

$$hj_{\max}.sj_{\max} = \left(\sum_{a=1}^{hj} \binom{hj}{a}\right) \cdot \left(\sum_{k=1}^{2^{17}} \binom{2^{17}}{k}\right) \tag{3}$$

where hj = f_h(**N**).

5) Clients: The size of the list **C**, cj, is a function of the network **N**. and expressed as cj = f_c(**N**).
$\mathbf{C_{max}}$ is the list of all possible lists of clients that are part of **C**. The length of this list is cj_{max} where

$$cj_{\max} = \sum_{a=1}^{cj} \binom{cj}{a} \tag{4}$$

where cj = f_c(**N**).

6) Rules and rule base: In a rule **R**,

- **R**.Source can contain any element of $\mathbf{C_{max}}$.
- **R**.Destination can contain any element of $\mathbf{H_{max}}$.
- **R**.Ports can contain any element of $\mathbf{S_{max}}$.
- **R**.Action is the maximum number of action combinations, being 2 ("Allow" or "Deny")

The firewall rule base $\mathbf{F_{max}}$ contains all possible rules that can be made with $\mathbf{C_{max}}$, $\mathbf{H_{max}}$ and $\mathbf{S_{max}}$

$$fj_{\max} = 2.cj_{\max}.hj_{\max}.sj_{\max} \tag{5}$$

$$fj_{\max} = 2. \left(\sum_{a=1}^{cj} \binom{cj}{a}\right) \cdot \left(\sum_{a=1}^{hj} \binom{hj}{a}\right) \cdot \left(\sum_{k=1}^{2^{17}} \binom{2^{17}}{k}\right) \tag{6}$$

where cj = f_c(**N**) and hj = f_h(**N**)

The possible design space for a rule base is phenomenal. Multiple rules can deliver one particular required functionality. Choosing the right rule is a real challenge. As the network grows and f_c(**N**) and f_h(**N**) grow, choosing the right firewall rule from the design space becomes even more difficult. To gain control over the design space, it needs to be consciously reduced.

E. Designing an evolvable rule base

A rule will be made up of:

- **Cs** representing the Source, where **Cs**$\subset$ $\mathbf{C_{max}}$.
- **Hd** representing the Destination, where **Hd**$\subset$ $\mathbf{H_{max}}$.
- **Sp** representing the Ports, where **Sp** $\in$ $\mathbf{S_{max}}$.
- Action is to be "Allow" as each rule in the rule base explicitly provides access to allowed services on allowed hosts.
- **R** = (**Cs**, **Hd**, **Sp**, "Allow')

Note that the last rule in the rule base **F**, **F**[fj] has to be the default deny rule ($\mathbf{R}_{\text{default\_deny}}$) as, when no rule explicitly provides access to a service on a host, the traffic needs to be explicitly blocked.

$$\begin{cases} \mathbf{R}_{\text{default\_deny}}.\text{Source = ANY,} \\ \mathbf{R}_{\text{default\_deny}}.\text{Destination=ANY,} \\ \mathbf{R}_{\text{default\_deny}}.\text{Port= ANY,} \\ \mathbf{R}_{\text{default\_deny}}.\text{Action = "Deny".} \end{cases}$$

From Section III-B, it is known that:

- A Firewall rule base is order-sensitive.
- Different types of relations/coupling can exist between rules.
- If all rules are disjoint from each other, there is no coupling between the rules.
- If all rules are disjoint, the rule base is no longer order-sensitive.
- If a new rule is added to the rule base and it's disjoint with all existing rules, then the location of the rule in the rule base is not important.

If the whole firewall rule base needs to be checked to see if a rule is disjoint to all existing rules, a CE is being introduced. Introducing a new rule to, or removing a rule from the system should result in work that is proportional to the newly required functionality and not into work, that has no logical link to the required functionality and that requires searching throughout the whole system (being the entire rule base). Or as NS formulates it: the impact of the change should be proportional to the nature of the change itself, and not proportional to the system to which the change is applied.

Disjoint rules have no overlap in source or destination or ports. The following combinations are possible:

- No overlap in sources - do not care about destination and port overlaps.

- No overlap in destinations - do not care about source and port overlaps.
- No overlap in ports - do not care about source and destination overlap.
- No overlap in source-destination combination, do not care about ports.
- No overlap in source-ports combinations, do not care about destinations.
- No overlap in destination-ports combinations, do not care about sources.
- No overlap in source-destination-port combination.

Cs is f_c(**N**) and **Hd** is f_h(**N**). The network is an uncontrollable variable. Trying to find a way to structure **Cs** and **Hd** to allow for disjoint rules starting from this variable, will not yield to anything useful. On the other hand, **Sp** represents the ports and is bound: the nature of TCP/IP limits the number of possible ports and thus all port combinations. It thus makes sense to look for a way to guarantee that there is no overlap at port/service level.

Let us consciously restrict **Sp** to **Su**, so that **Su** only contains unique values.

$$\begin{cases} \exists!\mathbf{Su}[\mathrm{m}] \text{ in } \mathbf{Su} \text{ for m:1}\rightarrow\text{suj.} \\ \mathbf{Su}[\mathrm{u}] \cap \mathbf{Su}[\mathrm{v}] = \emptyset\text{, where u, v:1}\rightarrow\text{suj, and u} \neq \text{v} \end{cases}$$

If each service is represented by 1 port, **Su** will contain 2^{17} elements, which is the max size of **Su** in this restricted case.
The service **Su**[m] can be delivered by many hosts.

Let $\mathbf{Hd}_{\mathbf{Su}[\mathrm{m}]}$ represent the list of hosts that offer service **Su**[m].

$$\begin{cases} \mathbf{Hd}_{\mathbf{Su}[\mathrm{m}]} \subset \mathbf{Hmax} \text{ and } \mathbf{Hd}_{\mathbf{Su}[\mathrm{m}]}[\mathrm{x}] \text{ contains a single host.} \\ \mathbf{Hd}_{\mathbf{Su}[\mathrm{m}]} \text{ contains unique and disjoint elements.} \\ \exists!\mathbf{Hd}_{\mathbf{Su}[\mathrm{m}]}[\mathrm{x}] \text{ in } \mathbf{Hd}_{\mathbf{Su}[\mathrm{m}]}\text{for x:1}\rightarrow\text{hdm} \\ \mathbf{Hd}_{\mathbf{Su}[\mathrm{m}]}[\mathrm{u}] \cap \mathbf{Hd}_{\mathbf{Su}}[\mathrm{v}] = \emptyset\text{, where u, v:1}\rightarrow\text{hdmj, and u} \neq \text{v} \end{cases}$$

Combining hosts and services ($\mathbf{Hd}_{\mathbf{Su}[\mathrm{m}]}$[x],**Su**[m]) where x:1→hdmj, gives a list of tuples that are disjoint. This hold for all m:1→suj. At this point, all services and hosts who deliver the services, form tuples that are disjoint and can thus be used as a basis for creating an order independent firewall rule base. $\mathbf{Cs}_{\mathbf{Hd}_{\mathbf{Su}[\mathrm{m}]}[\mathrm{x}]}$ is the list of clients that have access to service **Su**[m], defined on host $\mathbf{Hd}_{\mathbf{Su}[\mathrm{m}]}$[x].
By using :

- **Su**[m] where m:1→suj, with suj=number of disjoint services offered on the network, for defining **R**.Port
- $\mathbf{Hd}_{\mathbf{Su}[\mathrm{m}]}$[x], x:→hdmj, with hdmj=number of hosts offering **Su**[m], for defining **R**.Destination
- $\mathbf{Cs}_{\mathbf{Hd}_{\mathbf{Su}[\mathrm{m}]}[\mathrm{x}]}$ being the list of clients requiring access to service **Su**[m] on host $\mathbf{Hd}_{\mathbf{Su}[\mathrm{m}]}$[x], of length = cjs, for defining **R**.Source
- "Allow", for **R**.action

disjoint rules are being created, usable for an evolvable firewall rule base.

F. The artifact

What has been discussed in the previous section needs to be transformed into a solution usable in a real firewall. As discussed in Section III-C, firewalls work with groups. Groups can be used to represent the concepts discussed in the previous sections.

1) Starting from an empty firewall rule base **F**. Add as first rule the default deny rule **F**[1]= $\mathbf{R}_{\text{default\_deny}}$ with

$$\begin{cases} \mathbf{R}_{\text{default\_deny}}\text{.Source = ANY,} \\ \mathbf{R}_{\text{default\_deny}}\text{.Destination=ANY,} \\ \mathbf{R}_{\text{default\_deny}}\text{.Port= ANY,} \\ \mathbf{R}_{\text{default\_deny}}\text{.Action = “Deny”.} \end{cases}$$

2) For each service offered on the network, create a group. All service groups need to be completely disjoint from each other: the intersection between groups must be empty.
Naming convention to follow:
   - **S**_*service.name*,
   - with *service.name* as the name of the service.
3) For each host offering the service defined in the previous step, a group must be created containing only one item (being the host offering that specific service).
Naming convention to follow:
   - **H**_*host.name*_**S**_*service.name*,
   - with *host.name* as the name of the host offering the service
4) For each host offering the service from the first step, a client group must be created. That group will contain all clients requiring access to the specific service on the specific host.
Naming convention to follow:
   - **C**_**H**_*host.name*_**S**_*service.name*
5) For each **S**_*service.name*,**H**_*host.name*_**S**_*service.name* combination, create a rule **R** with:

$$\begin{cases} \mathbf{R}\text{.Source =}\mathbf{C}\_\mathbf{H}\_\textit{host.name}\_\mathbf{S}\_\textit{service.name} \\ \mathbf{R}\text{.Destination = }\mathbf{H}\_\textit{host.name}\_\mathbf{S}\_\textit{service.name} \\ \mathbf{R}\text{.Port= }\mathbf{S}\_\textit{service.name} \\ \mathbf{R}\text{.Action = “Allow”} \end{cases}$$

Add those rules to the firewall rule base **F**.
The default rule $\mathbf{R}_{\text{default}}$ should always be at the end of the rule base.

By using the artifact's design principles, group objects are created that form the building blocks for an evolvable rule base. Each building block addresses one concern.
If each service of **Su** is made up of only one Port, then the **Su** will contain maximum 2^{17} elements, resulting in maximum 2^{17} service groups **S**_*service.name* being created. For each host, maximum 2^{17} services can be defined, expressed in **H**_*host.name*_**S**_*service.name* destination groups. According to the artifact, one rule per host and per service, must be created. This reduced the rule base solution space from

$$2.\left(\sum_{a=1}^{cj}\binom{cj}{a}\right).\left(\sum_{a=1}^{hj}\binom{hj}{a}\right).\left(\sum_{k=1}^{2^{17}}\binom{2^{17}}{k}\right) \tag{7}$$

where cj = *fc*(**N**) and hj = *fh*(**N**)
to:

$$fj = hdj.suj + 1 = hdj.2^{17} + 1 \tag{8}$$

with hdj = number of hosts connected to the network.
hdj = *fh*(**N**). The "+1" is the default deny rule $\mathbf{R}_{\text{default\_deny}}$

V. Demonstrate and evaluate artifact

In this section, we will demonstrate the artifact. We will apply different changes on a rule base (add/remove rule) and on the components that make up rule (add/remove a service, add/remove a host, add/remove a client). We also show what happens if rules are aggregated. The section terminate with an evaluation of the proposed artifact.

A. Add and remove a rule

Creating rules according to the artifact's design principles, leads to rules that are disjoint from each other. Disjoint rules can be added and removed from the firewall rule base without introducing CE's.

B. Adding a new service to the network

A new service is a service that is not already defined in **Su**. The new services results in a new definition of a service being added to **Su**. The artifact prescribes that a new group **S**_*service.name* must be created for the new service. The group will contain the ports required for the service. For each new host offering the service, the artifact prescribes to create a new group destination **H**_*host.name*_**S**_*service.name*, and an associated source group **C**_**H**_*host.name*_**S**_*service.name*. The destination groups are populated with only one host (the host offering the service). The source groups are populated with all clients requiring access to the service one specific host. All building blocks to create the disjoint rules are now available. For each host offering the new service, a rule must be created using the created groups. No CE's are being introduced during these operations. Adding the new rules to the rule base does not introduce CE's (see Section V-A).

C. Adding a new host offering existing services, to the network

A new host is a host that is not already defined in **Hd**. The new host results in a new host definition being added to **Hd**. The artifact prescribes that a new group **H**_*host.name*_**S**_*service.name* must be created for each service delivered by the host and a corresponding source group **C**_**H**_*hostname*_**S**_*service.name* must be created as well. The destination groups are populated by their corresponding hosts. The source groups are populated with all clients requiring access to the service on that host. All building blocks to create the disjoint rules are now available. For each service offered by the new host, a rule must be created using the created groups. No CE's are being introduced during these operations. Adding the new rules to the rule base does not introduce CE's (see SectionV-A).

D. Adding a new host offering new services, to the network

Combining Sections V-C and V-B delivers what is required to complete this type of change. The artifact prescribes that new service groups must be created for new services. An equal amount of destination groups needs to be created and each populated by the new host. The same amount of source groups needs to be created and populated by the clients requiring access to one of the new services on the new host. All building blocks to create the disjoint rules are now available. For each combination (new host, new service), a rule must be created using the created groups. No CE's are being introduced during these operations. Adding the new rules to the rule base does not introduce CE's (see Section V-A).

E. Adding a new client to the network

Adding a new client to the network does not require the creation of new rule building blocks or the addition of new rules. The new client only needs to be added to those source groups that give access to the required services/hosts combinations. No CE's are being introduced during these operations.

F. Removing a service from the network

Let **sr** be the service that needs to be removed from the network. The name of the service is **sr**.name=sremove. The service is part of **Su**. The group corresponding with **sr** is **S**_sremove. The hosts offering the service correspond with the groups **H**_*host.name*_**S**_sremove. The clients consuming the service are defined in **C**_**H**_*host.name*_**S**_sremove. All building blocks to identify the rules that require removing from the rule base are now available. For each host offering **sr**, the corresponding rule

$$\begin{cases}\mathbf{R}_{\text{default\_deny}}.\text{Source} = \mathbf{C}\_\mathbf{H}\_\textit{host.name}\_\mathbf{S}\_\text{sremove}\\ \mathbf{R}_{\text{default\_deny}}.\text{Destination}=\mathbf{H}\_\textit{host.name}\_\mathbf{S}\_\text{sremove}\\ \mathbf{R}_{\text{default\_deny}}.\text{Port}=\ \mathbf{S}\_\text{sremove}\\ \mathbf{R}_{\text{default\_deny}}.\text{Action} = \text{“Allow”}\end{cases}$$

must be removed from the rule base. No CE's are being introduced during these operations. Removing rules from the rule base does not introduce CE's (see Section V-A). The service **sr** needs to be removed from **Su** as well as the corresponding group **S**_remove in the firewall.

G. Removing a host from the network

Let **hr** be the host that needs to be removed from the network. The name of the host is **hr**.name=hremove. The host is part of **Hd**. There will be as much destination groups for **hr** as there are services offered by **hr**. They are defined by **H**_hremove_**S**_*service_name*. The same holds form the source groups, defined by **C**_**H**_hremove_**S**_*service.name*. All building blocks to identify the rules that require removal from the rule base are available. For each service offered by **hr**, the corresponding rule

$$\begin{cases}\mathbf{R}_{\text{default\_deny}}.\text{Source} = \mathbf{C}\_\mathbf{H}\_\text{hremove}\_\mathbf{S}\_\textit{service.name}\\ \mathbf{R}_{\text{default\_deny}}.\text{Destination}=\mathbf{H}\_\text{hremove}\_\mathbf{S}\_\textit{service\_name}\\ \mathbf{R}_{\text{default\_deny}}.\text{Port}=\ \mathbf{S}\_\textit{service.name}\\ \mathbf{R}_{\text{default\_deny}}.\text{Action} = \text{“Allow”}\end{cases}$$

must be removed from the rule base. No CE's are being introduced during these operations. Removing rules from the rule base does not introduce CE's (see Section V-A). The host **hr** needs to be removed from **Hd** and the corresponding groups **H**_remove_**S**_*service.name* in the firewall, must be removed as well.

H. Removing a service from a host

Let **sr** be the services with **sr**.name=sremove, which needs removing from host **hr** with **hr**.name = hremove. The service is part of **Su**. The group corresponding with **sr** is **S**_sremove. The destination group for service **sr** on host **hr**, is **H**_hremove_**S**_sremove. The corresponding source group is **C**_**H**_hremove_**S**_sremove. All building blocks to identify the rule

$$\begin{cases}\mathbf{R}_{\text{default\_deny}}.\text{Source} = \mathbf{C}\_\mathbf{H}\_\text{hremove}\_\mathbf{S}\_\text{sremove}\\ \mathbf{R}_{\text{default\_deny}}.\text{Destination}=\mathbf{H}\_\text{hremove}\_\mathbf{S}\_\text{sremove}\\ \mathbf{R}_{\text{default\_deny}}.\text{Port}=\ \mathbf{S}\_\text{sremove}\\ \mathbf{R}_{\text{default\_deny}}.\text{Action} = \text{“Allow”}\end{cases}$$

which require removing from the rule base are available. No CE's are being introduced during these operations. Removing rules from the rule base does not introduce CE's (see Section V-A). The service **sr** does not need to be removed from **Su** and neither does the corresponding group as the service is still offered on other hosts.

I. Removing a client from the network

Let **cr** be a client that needs to be removed from the network. The client is part of **Cs**. Removing a client from the network does not require removing rules from the rule base. The client needs to be removed from the different source groups that provide the client access to specific services on specific hosts. If the services and hosts to which the client has access are known, then the source group from which the client needs to be removed, are known as well. If the services and/or hosts are not known, then an investigation of all the source groups is required to see if the client is part of the group or not. If part of the group, the client needs to be removed. The client also needs to be removed from **Cs**. Determining if a client is part of a source group can be considered as a CE as all source groups require inspection.

J. The impact of aggregations

When following the prescriptions of the artifact, many groups and rules will be created (see Section V-K for more details). The urge to aggregate and consolidate rules into more general rules, will be a natural inclination of firewall administrators as a smaller rule base will be (wrongfully) considered as a less complicated rule base. However, any form of aggregation will result in loss of information. It is because the artifact consciously enforces fine-grained information in the group naming and usages that disjoint rules can be created and the ZT model can be enforced. If due to aggregations it can no longer be guaranteed that rules are disjoint, then a CE-free rule base can no longer be guaranteed either. Aggregation will also lead to violations of the ZT model.

We provide two examples of aggregations.

Aggregation at service level: all hosts offering the same service are aggregated into one destination group. Such an aggregation excludes the possibility of specifying that a client needs access to a specific service on a particular host. A client will have access to the service on all hosts offering the service, desired or not. In such a configuration, ZT can no longer be guaranteed. As long as the services on the network are unique, so will be the port groups. Rules will stay disjoint and the rule base CE-free. The moment that one starts combining ZT and non-ZT rules, non-disjoint rule will pop-up. The rule base can no longer be guaranteed to be CE-free.
Example: if for some reason, it cannot be allowed that a client has access to the service on all hosts and a special service group is being created (no longer disjoint with the existing service group) with a special associated destination group (no longer disjoint with existing destination groups), the rule created with those groups is not disjoint with existing rules in the rule base and the effect of adding this rule to the rule base is no longer guaranteed CE-free.

Aggregation at host level: all services offered on a host are aggregated into one host-bound port/service group. The aggregation method excludes specifying that a client needs access to some of the services on the host. A client will have access to all services defined on the host, desired or not. In such a configuration, ZT can no longer be guaranteed. As long as the destination groups are unique, disjoint rules can still be created. The moment that ZT and non-ZT rules are combined, non-disjoint rule will pop-up. The rule base can no longer be guaranteed CE-free.
Example: if for some reason, it cannot be allowed that a client has access to all services on the host and a special service group is being created (no longer disjoint with existing service groups) with a special associated destination group (no longer disjoint with existing destination group), the rule created with those groups is not disjoint with existing rules in the rule base and the effect of adding this rule to the rule base is no longer guaranteed CE-free.

K. Evaluation

The previous demonstrates that, when applying the artifact, the rules are guaranteed to be disjoint and adding and removing such rules has no unwanted side effects on the existing rule base. Such a rule base will be fine-grained (i.e., having many rules). The size of the rule base might be consider this as a drawback. Large size is often regarded as complex. A large size rule base may also impact firewall performance, as surching for a matching rule in a large rule base, has a direct impact on firewall performance. In Section VIII, the impact of rule base size on performance is further investigated. Some operations on rules may indeed result in CE's at group level, such as adding and removing a client from the network. Aggregations will violate the ZT constraint. Combining aggregation and non-aggregation based rules results in non-disjoint rules and CE's at rule base level.

VI. Filtering strategies

The artifact discussed in the previous section was created to be compliant with the ZT filtering strategy. In this section, we will discuss other filtering strategies: Interconnect strategy and Outbound filtering and see what kind of impact they have on the artifact.

A. Interconnect filtering strategy

The ZT filtering strategy can be considered as an inbound filtering strategy. Only traffic corresponding with exposed services is allowed. The filtering strategy used to interconnect different network segments and control the traffic between those segments is an Interconnect (IC) filtering strategy. The focus is on traffic between network segments, like VLANs or groups of VLANs, and not on the resources connected to those network segments. The rules are different compared to ZT rules. The level of granularity is a network subnet, not the resource. Filtering does not happen at port/service level. This means that there is one less parameter to enforce disjointness between the rules.

The proposed artifact can still be used to create an IC strategy based rule base. The group objects used in an IC strategy rule base would represent the following:

- Destination group: a group containing the IP addresses, expressed in subnets (VLAN's), that make up a logical part of the network.

- Source group: a group of IP addresses expressed in subnets (VLAN's), that make up a logical part of the network.

The VLANs can be organized in different ways. They can be organized according to a physical location or organizational department. In the former case, there is a VLAN per building floor, and the sum of all VLANs represents the building. In the latter case, there are VLANs per organizational unit, grouped in different parts of the building. The sum of all VLANs based organizational units in the building represents the full building.

In ZT based filtering, the most fine-grained component filtering is performed at, is the port. In IC based filtering, the most fine-grained component filtering is performed at, is the VLAN. The design of the rule base will be structured around the VLAN.

Using the artifact previously designed artifact:

- Start from an empty firewall rule base **F**. Add as the first rule; the default deny rule.
- For each VLAN requiring access control, create a destination group. Populate the group with the relevant IP address ranges representing the VLAN. The intersection between all groups must be empty! A VLAN cannot be present in 2 different logical parts of the network and thus in 2 groups. The naming convention of those groups: **D**_VLAN-LogicalName-VLANnr
- For each part of the network, which requires potential shielding from other parts of the network, create a source group. Populate the source group with the VLAN's that require access. The naming convention of those groups: **S_D**_VLAN-LogicalName-VLANnr.
- For each VLAN that requires protection, create a rule:
  - Source: **S_D**_VLAN-LogicalName-VLANnr
  - Destination: **D**_VLAN-LogicalName-VLANnr
  - Protocol: ANY

The **D**_VLAN-LogicalName-VLANnr groups will enforce the disjointness of the rules in the rule base. Add, remove, change operation on a rule base created according to the artifact are compliant with the evolvability conditions. It should be clear that this kind of filtering cannot be combined with ZT based filtering. The disjointness of rule cannot be guaranteed if ZT and IC based rules are used in the same firewall rule base:

- Protocol: violates disjointness
- Destination: ZT rules will be a subnet of IC rules and thus violate disjointness.
- Source: is not used to enforce disjointness

An example of an IC strategy use case is the merger between two companies. Each has their network. As long as the security policies are not aligned between both companies, there is a good reason not to interconnect the two networks directly. The interconnection is best done via a firewall. The firewall will filter between IP ranges, for instance, allowing traffic between the two headquarters, but not yet between remote sites (simplified example, not considering potential IP range overlap, NATing etc.).

As change is the only constant in companies, IC based filtering is complicated. Moves between buildings, reorganization in buildings, add and removal of sites, organizational changes, all make upfront, and stable segmentation of a network difficult. Segmentation rules change, segmentation principles are mixed, and logical network segments no longer become disjoint. The result will be evolvability issues in the rule base(s) and unforeseen side effects due to changes. Till now, the IC problem has been addressed in a network-centric approach. As network segmentation and company organization can result in implementation conflicts, solutions such as identity-based firewalls emerged. In those solutions, IC' based filtering happens based on the identity of the user. When a user tries to connect to certain parts of the network and hits an identity-based firewall enforcing the IC, the firewall will check the identity of the user and will filter based on this identity. This only works if:

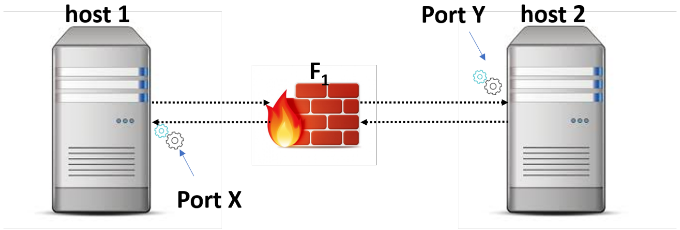


Figure 4. Inbound and outbound on a single firewall

- The firewall can establish the identity of the user associated with the source (who's working on PC with IP = x.y.z.u).
- The firewall has access to a DB containing the identities and has mechanisms to validate the identity.
- The firewall has a set of rules stating which identity has access to which destinations.

Such a setup is more user-centric. Access to the network is linked to the identity of the user and not the building or organizational layout. Elegant as the solution may seem, it just shifts the problem from the network space to the identity space. This research will not further investigate this. However, it is worth pointing out that, user identities, identify verification (authentication), identity authorization, identity definition, identity implementation, identity and HR policies, identity synchronization solutions, are among the most complex IS systems of an IT landscape. Researching the associated evolvability issues and proposing solutions is worthy of a separate Ph.D. research.

B. *Inbound and outbound filtering strategy*

An inbound filtering strategy, as ZT, will filter traffic close to the destination. The outbound filtering strategy will filter close to the source. From a security point of view, it makes sense to stop the traffic as early as possible on the network. On a single firewall, the notion of inbound and outbound is relative. A firewall rule base is not aware of inbound or outbound. It only knows source and destination and both can be located on the two sides of the firewall.

The artifact we propose started from a scenario where all sources are located on the left and all destination to the right of the firewall, effectively implementing an inbound filtering strategy. The same artifact can be used in a single firewall setup where sources and destination are located at both sides of the firewall. As long as the artifact is strictly followed, all rules will stay disjoint. There are some dangers involved. Take the case described in Figure 4 where a host1, located on the left side of the firewall, needs to access a host2 on the right side. Host2 also requires access to a service offered by host1. According to the artifact, the 2 following rules would be created.

- **C_H**_host2_**S**_Y, **H**_host2_**S**_Y, **S**_Y, Allow
  - traffic from left to right
  - **H**_host2_**S**_Y contains host2
  - **C_H**_host2_**S**_Y contains host1
- **C_H**_host1_**S**_X, **H**_host1_**S**_X, **S**_X, Allow
  - traffic from right to left
  - **H**_host1_**S**_X contains host1
  - **C_H**_host1_**S**_X contains host2

What the firewall will do internally is look at the content of the groups, not the group names itself, and the rules are internally translated as

- host1, host2, Y, Allow
- hotst2, host1, X, Allow

Both host1 and host2 are member of different group. Interchaning those groups will result in rules which do not follow the logic of the artifact but that do represent the same rules inside the firewall

- **H**_host1_**S**_Y, **C_H**_host1_**S**_Y, **S**_X, Allow
  - host2, host1, X, Allow
- **H**_host2_**S**_Y, **C_H**_host1_**S**_X, **S**_Y, Allow
  - host1, host2, Y, Allow

Group objects are used to increase the manageability of rule bases. The above makes it clear that, if not used correctly, manageability will decrease. Groups created to represent destinations cannot be used to represent souces in rules, and vise versa. This is a manifestation of Separation of Concern. Representing sources and destination are different concerns. They should not be mixed.

Inbound and outbound filtering are also two different concerns. In the above scenario, both are mixed on one firewall yet, no immediate impact seems to surface. The impact will become visible when there are multiple firewalls in the network. This will be discussed in the next section.

VII. Multiple Firewalls

In the previous sections, the assumption was taken that the network only contains one firewall. In this section, we will investigate the impact of multiple firewalls between the source and the destination.

A. *The serial firewall filtering function*

Let **Pa** be a package traveling over the network.

- **Pa**.source = the IP adress of the source sending package **Pa**.
- **Pa**.destination = the IP address of the destination for package **Pa**.
- **Pa**.port = the **Port** targetted on destination **Pa**.destination.

Let $\varphi_{\mathbf{f}}(\mathbf{F_f},\mathbf{Pa})$ be the firewall filtering function that takes rule base $\mathbf{F_f}$ and package **Pa** as input.

$$\begin{cases} \varphi_{\mathbf{f}}(\mathbf{F_f},\mathbf{Pa}) = 0 \text{ if the package is blocked} \\ \text{- there is no rule } \mathbf{R} \text{ in } \mathbf{F_f} \text{ such that the package is allowed} \\ \varphi_{\mathbf{f}}(\mathbf{F_f},\mathbf{Pa}) = 1 \text{ if the package is allowed} \\ \text{- there is a rule } \mathbf{R} \text{ in } \mathbf{F_f} \text{ such that the package is allowed} \end{cases}$$

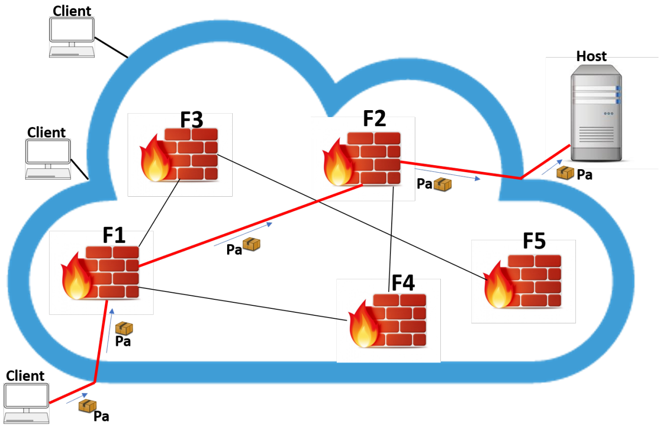


Figure 5. Multiple firewalls in a network

Let f_total be the total amount to firewalls in a given network. Let $\Phi^{\mathbf{s}}{}_{\mathbf{fw}}$ be the serial firewall filtering function for a network path containing fw firewalls in serie. Then

$$\Phi^{\mathbf{s}}{}_{\mathbf{fw}}(\mathbf{Pa}) = \prod_{f=1}^{f=fw} \varphi_{\mathbf{f}}(\mathbf{F_f},\mathbf{Pa}) \tag{9}$$

$$\Phi^{\mathbf{s}}{}_{\mathbf{fw}}(\mathbf{Pa}) = \varphi_{\mathbf{1}}(\mathbf{F_1},\mathbf{Pa}).\varphi_{\mathbf{2}}(\mathbf{F_2},\mathbf{Pa})...\varphi_{\mathbf{fw}}(\mathbf{F_{fw}},\mathbf{Pa}) \tag{10}$$

Where:

$$\begin{cases} \boldsymbol{fw : 1 \rightarrow f\_total} \\ \boldsymbol{\Phi^{s}{}_{fw}}(\mathbf{Pa}) = 0 \text{ if } \mathbf{Pa} \text{ is blocked by at least one of the} \\ \text{ fw firewalls} \\ \\ \boldsymbol{\Phi^{s}{}_{fw}}(\mathbf{Pa}) = 1 \text{ if } \mathbf{Pa} \text{ is allowed by all fw firewalls} \end{cases}$$

See Figure 5 for a graphical representation of these concepts.

B. *Applying the rules on some firewalls*

In a given network, fw and thus $\Phi^{\mathbf{s}}{}_{\mathbf{fw}}$, will differ from the location of the source, destination, and the internal routing of the network. Let us assume that in such a network, all firewalls have an evolvable rule base according to the proposed artifact. The addition of a new resource, host_new offering service S_new, requires the addition of new rules $\mathbf{R_{new}}$, such that host_new is protected according to the ZT filtering strategy. Let us assume that $\mathbf{R_{new}}$ is only implemented on the firewalls in the path between the initially identified sources (members of **C_H**_host_new_**S**_new), and destination host_new. As time moves on, the initially identified sources require modification: a new client needs to access the host, or a client is removed from the network.

According to our artifact, adding or removing a client is just a question of adding and removing the client from the group **C_H**_host_new_**S**_new. In our current scenario, this is no longer the case. If a new client has a different network path towards the host_new compared to the path in which the rule $\mathbf{R_{new}}$ was initially implemented, then the rule $\mathbf{R_{new}}$ must now be implemented on the all firewalls in the path between the new client and host_new as well. In addition, the source group must be updated on all firewalls in all paths between all current clients and host_new. As the possible network paths are a function of the network, and the network can grow infinitely, a CE is being introduced. This is the worst kind of CE, as we will not know upfront where adjustments are required, and the

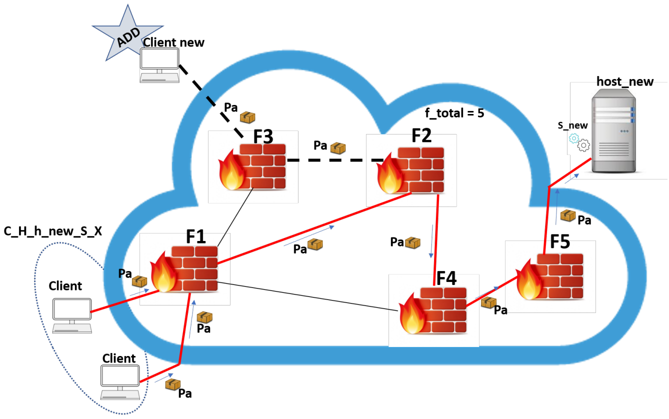


Figure 6. Apply the rules on some firewalls

full investigation of the network is required. An example of the described scenario can be found in Figure 6.

C. Applying the rules on all firewalls

The only way to avoid the problem described in the previous section, is to have all firewalls contain the same rule base. All manipulations of rules must be done on all firewalls simultaneously. As the network grows, so will the number of firewalls, and again, a CE is being introduced. This CE is less aggressive as it is know now the manipulations are required on all firewalls. We have already discussed the impact of the size of the rule base on the firewall. Having to duplicate all rules all over the network will make the rule base even larger and less coherent. Rules are added to firewalls, which will never be activated, and groups contain objects that are not relevant to the context of that specific firewall. The manageability of the firewalls will decrease. All firewalls are addressing the same concern. Normalized Systems learns that this will have a negative impact on evolvability, as can be concluded from the above.

D. Restricting Inbound traffic filtering

The paper “Minimizing the Maximum Firewall Rule Set in a Network with Multiple Firewalls” [31] is closely related to the problem we are trying to solve. According to [31], placing firewalls in a network such that the rule base is minimal, is an NP-complete problem, which requires a heuristics-based approach. Although applying the heuristic-base algorithm described in [31] may minimize the rule base over all firewalls, the evolvability of those rule bases is not discussed.

In Section VI-B, we mentioned that a network with one firewall is combining both inbound and outbound filtering rules. If we have a network with two firewalls that are connected in a back-to-back configuration - meaning the firewalls are directly interconnected and no resources are located in this interconnection - inbound and outbound traffic filtering can be separated. This can be done by adding a new default rule, which states that all outbound traffic is allowed. Figure 7 illustrates the setup, while Algorithm 1 and Algorithm 2 show the construction of the rule bases of F_1 and F_2.

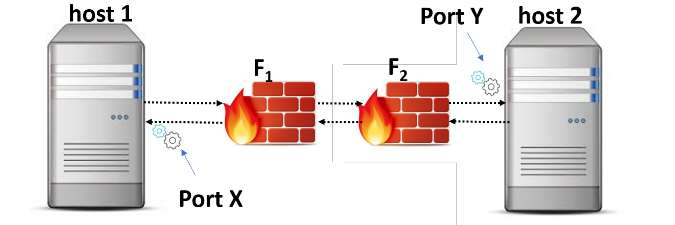


Figure 7. back-to-back firewalls

```
Rules Firewall F1
    R1: C_H_F1Any_S_Any, H_F1Any_S_Any,
      S_Any, Allow
    R2: C_H_host1_S_X, H_host1_S_X, S_X, Allow
    R3: Any, Any, Any, Deny
    with group contents
        in R1: H_F1Any_S_F1Any: any
        in R1: S_F1Any: any
        in R2: C_H_host1_S_X: all hosts needing
          access to host1, host2 in this case
        in R2: H_host1_S_X: host1
        in R2: S_X: port X
```

Algorithm 1: Rule base of F_1

The rules **R**1 on both firewalls are disjoint with respect to the rule base in which they are located as:

- on F_1: **C_H**_F_1Any_**S**_Any - represents all hosts protected by inbound traffic by F_1
- on F_2: **C_H**_F_2Any_**S**_Any - represents all hosts protected by inbound traffic by F_2
- **C_H**_F_1Any_**S**_Any ∩ **C_H**_F_2Any_**S**_Any = ∅

and

- All source groups on F1 are subsets of **C_H**_F_2Any_**S**_Any - represents all hosts protected by inbound traffic by F_2
- All source groups on F2 are subsets of **C_H**_F_1Any_**S**_Any.

Thus, on both F1 and F2, the default outbound rule is disjoint with all other groups.

We see here appearing Separation of Concern. The concern of protecting a resource is only assigned to one firewall. If given to multiple firewalls, evolvability issues will occur. The leads to the following design criteria:

- A firewall should be clearly assigned to protect a set of resources. Those resources are protected by the firewall via the inbound ZT traffic filtering strategy.
- The firewall allows all outbound traffic from the set of resources it protects, to the rest of the network.

```
Rules Firewall F2
    R1: C_H_F2Any_S_Any, H_F2Any_S_Any,
      S_Any, Allow
    R2: C_H_host2_S_Y, H_host2_S_Y, S_Y, Allow
    R3: Any, Any, Any, Deny
    with group contents
        in R1: H_F2Any_S_F2Any: any
        in R1: S_F2Any: any
        in R2: C_H_host2_S_Y: all hosts needing
          access to host2, host1 in this case
        in R2: H_host2_S_X: host2
        in R2: S_Y: port Y
```

Algorithm 2: Rule base of F_2

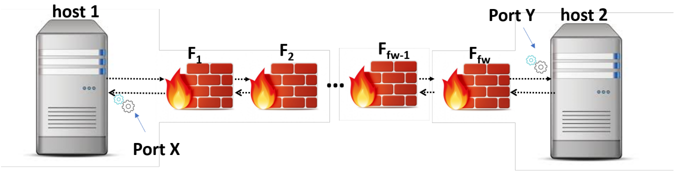


Figure 8. Path with multiple firewalls

- If all firewalls are protecting their resources, there is no need for outbound filtering.

As illustrated, our artifact can be made compliant with such as setup, simply by adding the “default allow” rule and the creation of some extra groups.

The approach described above might be turned around: by default allow all inboud traffic and filter on outbound traffic. Separation of Concerns would be respected. The artifact would need to be revised as disjointness would need to be enforced based on the combination of Service and Source instead of Service and Destination. The same reasoning applies for a the inbound default allow rule. Although technically possible, this filtering strategy would be confusing. Compare with the following scenario: A city needs to close an entry road due to construction works. Traffic will be blocked as close to the yard as possible (inbound filtering). It is impossible to block all roads, which could potentially lead to the city (outbound filtering).

E. *Multiple firewalls revised*

What happens when there are more than 2 firewalls between 2 resources? Figure 8 illustrates the setup. If we apply the design criteria from the previous section, we have to conclude that F_2 to $F_{fw\text{-}1}$ are not allowed to filter inbound traffic. Those concerns are already assigned to F_1 and F_{fw}. Firewall F_2 to $F_{fw\text{-}1}$ must handle other concerns such as:

- **chokepoint**: Use a firewall as a kind of valve: allow all or deny all. This comes in handy in case of network intrusions, and traffic needs to be blocked asap in a simple way, without impacting existing routing.
- **Interconnect filtering strategy**: use those firewalls to control connectivity between network segments (see Section VI-A).

Note that for the Interconnect filtering strategy, Separation of Concern needs to be respected as well. A “IC” firewall should be assigned to handle the interconnect of assigned ranges, and no other “IC” firewall should filter on the same ranges. This can again become quickly complex and evolve into an NP-complete problem. The best advice is to refrain from the usage of “IC” and chokepoint firewalls, limiting the number of firewalls in any network path as much as possible.

VIII. Additional aspects of firewall rules bases

Applying the Normalized Systems Principles results in a fine-grained modular structure. The creation of an evolvable firewall rule base is no exception; it leads to a fine-grained rule base. Creating and managing a large rule base requires automation, and a large rule base may lead to performance issues. In this section, the scalability of an evolvable rule base will be disussed, together with a possible approach to automate the creation and management of an evolvable rule base. The section ends with a reflection on Software Defined Networks (SDN) and Software Defined Firewalls (SDF) and why SDF has interesting evolvabilty features.

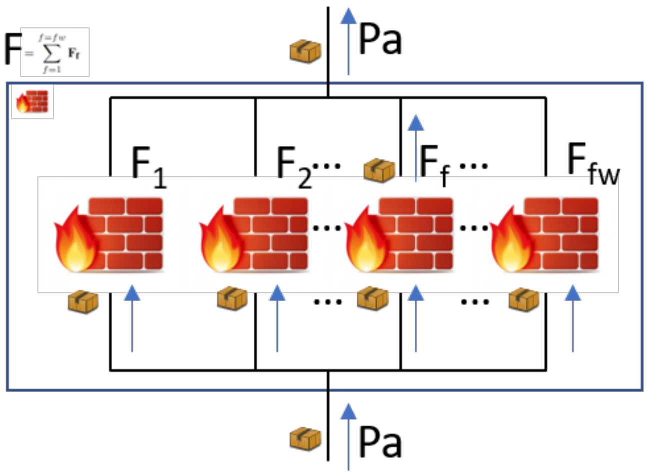


Figure 9. Scaling of Firewalls with normalized rule base

A. *Scaling*

In an evolvable rule base, all the rules are disjoint from each other and every network package can only hit one rule. This rule can be located in the beginning or near the end of the rule base. As there is only one rule that can be hit, the rule base can be split in multiple pieces and distributed parallelly over different firewalls. Let **F** be a firewall rulebase containing only disjoint rules, created according to the artefact described in Section IV-F. As visualized in Figure 9, **F** can be split in fw sub rule bases, which are spread over fw parallel firewalls. Each of the fw rule bases contains the “Default Deny” rule at the end.

A network package will try to pass each of the firewalls, but only one of the firewalls has a rule it can hit.

$$\mathbf{F} = \sum_{f=1}^{f=fw} \mathbf{F_f} \tag{11}$$

Let $\varphi_\mathbf{f}(\mathbf{F_f},\mathbf{Pa})$ be the firewall filtering function that takes rule base $\mathbf{F_f}$ and package **Pa** as input.

- $\varphi_\mathbf{f}(\mathbf{F_f},\mathbf{Pa}) = 0$ if the package is blocked - there is no rule **R** in $\mathbf{F_f}$ such that the package is allowed
- $\varphi_\mathbf{f}(\mathbf{F_f},\mathbf{Pa}) = 1$ if the package is allowed - there is a rule **R** in $\mathbf{F_f}$ such that the package is allowed

Let $\Phi^\mathbf{p}{}_\mathbf{fw}$ be the parallel firewall filtering function for fw firewalls in parallel. Then

$$\Phi^\mathbf{p}{}_\mathbf{fw}(\mathbf{Pa}) = \sum_{f=1}^{f=fw} \varphi_\mathbf{f}(\mathbf{F_f},\mathbf{Pa}) \tag{12}$$

$$\Phi^\mathbf{p}{}_\mathbf{fw}(\mathbf{Pa}) = \varphi_\mathbf{1}(\mathbf{F_1},\mathbf{Pa}) + \varphi_\mathbf{2}(\mathbf{F_2},\mathbf{Pa}) + \ldots + \varphi_\mathbf{fw}(\mathbf{F_{fw}},\mathbf{Pa}) \tag{13}$$

Where:

$$\begin{cases} \boldsymbol{\Phi^p_{fw}}(\mathbf{Pa}) = 0 \textbf{ if Pa} \text{ is blocked by all of the} \\ \quad \text{fw firewalls} \\ \boldsymbol{\Phi^p_{fw}}(\mathbf{Pa}) = 1 \textbf{ if Pa} \text{ is allowed by one rule of one} \\ \quad \text{of the fw firewalls, as:} \\ \exists! \boldsymbol{F_f} \in \boldsymbol{F} \text{ for} f = 1 \rightarrow fw \implies \boldsymbol{R} \in \boldsymbol{F_j} \end{cases}$$

This mechanism shows that the size of the evolvable rule base does not matter, as the solution scales. Firewalls with a non-evolvable rule base cannot scale the same way. Scaling comes with a cost. Modern firewalls allow virtualization, but each virtual instance comes at a cost as well.

In addition to the horiziontal scaling posibilities of an evolvable rule base, the performance of an evolvale rule base can be boosted by moving the most frequently used rules at the top. A firewall vendor such as CheckPoint, suggests to put the rules that are most frequently hit (and applied) at the top of the firewall table. In a rule base that is order-sensitive, this may be a real issue. In a rule base that is not order-sensitive, one could monitor the firewall and see which rules are hit most and move those rules around without having to worry about the potential impact on other rules. Doing this dynamically would even be more powerful as the firewall would be able to reorganize his rules according to the traffic of the day.

B. Automation

The creation of the fine-grained rule base by humans can be an issue. The procedure regarding definitions of the groups needs to be followed strictly, and the creation of a catalog of all possible services is a must. For standard services and tools, lists of assigned ports/protocols and international standardization organizations related to the Internet (like iana.org) exist and can be reused. The management of the groups, their content, and the rules, should be done in a tool outside of the firewall (see Figure 10). This tool could expand the firewall rules in the fine-grained format, according to the naming conventions, performing checks against the group definitions and content via a user-friendly interface. The tool could then push the rules towards the firewall, effectively separating the management of rules and implementation of rules. Such tools exist on the market. Examples are Algosec, Tuffin, Firemon. However, none of those tools consciously restrict the design space and will thus enforce the creation of an evolvable rule base.

Defining a rule for each service may be considered cumbersome. Roles could be created, like ”monitoring and management”, which are a grouping of smaller, disjoint services. The firewall administrator can create a rule specifying this ”monitoring and management” role, to express that the server needs to allow access to all monitoring and management services. The tool will expand this role into the individual rules for each disjoint service. Example:

- ”Monitoring and Management” = SSH + SFTP + FTP + SMTP + TELNET
- Host = x
- Rule : C_Hx_SMaM; Hx_S_MaM; S_MaM; allow
- Will be expanded to :
  - C_Hx_S_SSH; Hx_S_SSH; S_SSH; allow
  - C_Hx_S_SFTP; Hx_S_SFTP; S_SFTP; allow
  - C_Hx_S_FTP; Hx_S_FTP; S_FTP; allow
  - C_Hx_S_SMTP; Hx_S_SMTP; S_SMTP; allow
  - C_Hx_S_TELNET; Hx_S_TELNET; S_TELNET; allow

C. Software Defined Network and Software Defined Firewall

Pushing the inbound filtering strategy discussed in previous section to the limit equals providing each resource with its firewall. This is what is happening in a Software Defined Network (SDN) combined with a Software Defined Firewall (SDF). In an SDN, the network layer is virtualized inside a virtualization layer called the hypervisor. The SDN is decoupled from the actual underlying physical network. In the hypervisor layer, network components such as routers, switches, VLANs, load balancers, firewalls are all defined entirely in software. To each virtual host defined in/on the hypervisor, a virtual firewall can be attached. A package does not enter the network layer of the virtual hosts unless it successfully passes the firewall. SDF is better compared to an Operating System (OS)-based firewall (like IP tables or Windows Group Policies). OS-based firewalls can only perform their filtering function if the package is already ”inside” the host.

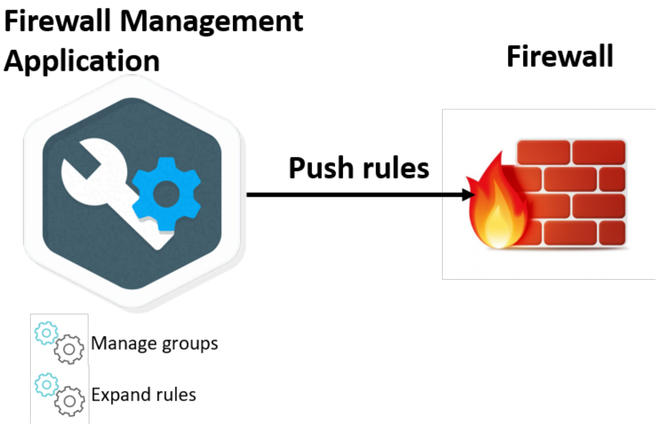


Figure 10. Firewall Management Tool

For an SDF, the rule base is configured by my means of policies. A policy defines the protocol and port that can pass though the firewall. The policies are attached to the firewall. As the firewall is attached to only one host, by default, disjointness for the destination is guaranteed. But, multiple policies can be attached to one host, and in those policies, overlaps and conflicts of protocols/ports and actions can be defined. Again, the conscious restriction of design space is required.

The previously proposed artifact can be adjusted for an SDN context by creating policies for Software Defined Firewalls. The policies are the equivalent of the Service Groups. They must be as fine-grained as possible. For each service exposed on a host, a policy must be created. Policies cannot overlap. Instead of creating a destination group, the polices are being attached to the host. As many policies are attached to the host as there are services offered by the host. Access to the host is provided by giving explicit access of a client to the host. This corresponds to creating a client group as defined in the artifact. Belonging to the group means you can access the host, and the policy attached to the host will check authorized protocols and ports.

A Software Defined Firewall in a Software Defined Context is the best way to guarantee the ZT filtering strategy. SDF also offers the most evolvable setup. Add/remove of hosts to the hypervisor automatically adds/removes the associated host firewalls. Add/remove of rules means add/remove of policies and/or attach/detach of policies. If the policies are created according to the proposed artifact, evolvability is guaranteed.

IX. Discussion

By means of discussion, we will cover two items. First, we will revisit the literature related to the size of the rule base, followed by a reflection on the nature of the CE’s that are still present when applying our proposed artifact.

A. Size of the rule base: revisiting the literature

The ***"Size of the rule base issue"*** is not treated as an issue related to the stability of a system under change. To the best of our knowledge, most contributions do not focus on this point, whereas it is a corner stone of NS. The different artifacts all start with ideas similar to "For each rule in the firewall, do the following ...". One might consider such an approach as a CE in itself. There is attention to reducing the rule base to a minimum list of rules, which still answer to the filtering requirements, motivated to the impact of the size of the rule base on performance. However, in [16] it is suggested that the actual size of the rule base is not related to the way the hardware actually processing the rules. This suggests a decorrelation between the size of the rule base and the firewall performance. If this would be the case, why bother about the reduction of the size of the rule base? In Section VIII we pointed out that a rule base that has built-in evolvability, can scale and thus circument the potential performance issues due to its size. Non-evolvable rule bases cannot scale this way. Scaling does come at a cost. Either in terms of the purchase of more physical firewalls or adding resources to firewalls which allow virtualization. The higher cost will result in a firewall setup which will behave as is exptected. Security always comes at a cost. Further research of the literature and real-life measurements are required to clarify this point.

Looking at the combinatorics of Section IV-D, the design space is enormous. By applying the artifact, there is a conscious reduction of the design space. But the size of the rule base is still large as for each combination (host, service) a rule must be created in the rule base.

$$2^{17}.hdj + 1 \tag{14}$$

The maximum number of services is 2^{17} = 131,072. However, in reality this number will never be reached. A sample in Engie (a multinational and world leader in energy services) on 100 servers revealed that on average 39 services are exposed. The standard deviation in the sample is 14. It can be stated with a statistical probability of 98% that a host exposes less then 67 services. The sample was taken from a population of 1,000 servers. Those 1,000 serves are currently protected by about 890 firewall rules. If the artifact would be applied, it would mean implementing 67,000 rules. However, at Engie, a ZT model at host level is not applied. Instead, ZT at VLAN level is present (still filter at port level, but instad of at host level, filtering happens at VLAN level = a collection of hosts). If the realistic assumption is a made that the 1,000 server are spread over 20 VLAN's, it would mean that 20 x 67 = 1,340 rules are required for an evolvability rule base. This would mean 50% more rules to gain full evolvability.

B. Remaining CE's

The artifact proposed in the paper is not completely free of CE. The evaluation has shown that there is are CE's at the level of groups. However, these CE's are not related to the technical coupling within the rule base but due to the size and topology of the network. The bigger the network, the more objects and rules. Such CE's are considered acceptable given that:

- The actions leading to the CE can be automated (search for, or through, groups)
- The CE is predictable and is the logical effect of the change which needs to be applied (remove a client = look in all groups where the client is present)

CE's which cannot be automated because their impact is not predictable are not acceptable as there is no logical link between the change and the extra work one needs to do to implement the change. For example, the addition of a rule to activate a service on the network that would require the inspection of the whole rule base to find conflicting rules (not related to the newly activated service) would be considered as an unacceptable CE. Note that the proposed artifact facilitates the removal of such unacceptable CE's.

X. Conclusion

Firewall rule bases are typically non-evolvable systems. Tools and literature exist on how to show and potentially reduce the complexity and conflicts in firewall rule bases, but practical guidance on how to make a rule base which has proven evolvability by design, is lacking. Using the NS paradigm and domain specific knowledge, we have proposed an artifact which has the desired evolvability. The most important drawback of the resulting rule base could be the size due to its fine-grained structure, although this should be further analyzed in future research efforts. In addition to the proposed artifact, the evolvability implications of filtering strategies and firewall placement, has been investigated, showing that the Software Defined Firewall, promisses evolvability in a multi firewall network.

What is currently lacking is an acutal tool that could create, push and manage firewall rule bases according to the outlined principles of this paper. Having such a tool is one thing, implementing it and proving that it inhances security and operational efficity related to security, is something completly different. The creation of a tool in combination with the organizational impact, are subject for future research.

Another topic for future reseach is the size of the rule base. More real-life use cases are required to see to what extend existing rule bases can be transformed into evolvalbe rule bases, what the size of those rule bases will be and what the cost if implementing such rule bases would be.

Acknowledgment

The authors would like to thank Stefan Thys, Frederik Leemans and Stefan Biesbroeck for their help in writing and editting the article, and Sam Gozin and Bruno De Becker for providing the Engie operational data.

References

[1] G. Haerens and P. De Bruyn, "Using normalized systems to explore the possibility of creating an evolvable firewall rule-base", The 11th International Conferences on Pervasive Patterns and Applications (PATTERNS), pp. 7-16, May 2019

[2] Firemon whitepaper, "2017 State of the firewall", URL https://www.firemon.com/resources/, [retrieved: April, 2019]

[3] Firemon whitepaper, "2018 State of the firewall", URL https://www.firemon.com/resources/, [retrieved: April, 2019]

[4] M. Bennet, "Zero Trust Security: A CIO's Guide to Defending Their Business From Cyberattacks", Forrester Research June 2017

[5] H. Shel and A. Spiliotes, "The State of Network Security: 2017 to 2018", Forrester Research November 2017

[6] Firemon whitepaper, "Firewall cleanup recommendations", URL https://www.firemon.com/resources/, [retrieved: April, 2019]

[7] H. Mannaert, J. Verelst, and P. De Bruyn, "Normalized Systems Theory: From Foundations for Evolvable Software Toward a General Theory for Evolvable Design", ISBN 978-90-77160-09-1, 2016

[8] 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: Volume 76, Issue 12, pp. 1210-1222, 2011

[9] H. Mannaert, J. Verelst, and K. Ven, “Towards evolvable software architectures based on systems theoretic stability”, Software Practice and Experience: Volume 42, Issue 1, 2012

[10] P. Huysmans, G. Oorts, P. De Bruyn, H. Mannaert, and J. Verelst.- “Positioning the normalized systems theory in a design theory framework”, Lecture notes in business information processing, ISSN 1865-1348-142, pp. 43-63, 2013

[11] G. Haerens, “Investigating the Applicability of the Normalized Systems Theory on IT Infrastructure Systems”, Enterprise and Organizational Modeling and Simulation”, 14th International workshop (EOMAS), pp. 123-137, June 2018

[12] P. Johannesson and E. Perjons, “An Introduction to Design Science”, ISBN 9783319106311, 2014

[13] A.R. Hevner, S.T. March, J. Park, and S. Ram, “Design Science in Information Systems Research”, MIS Quarterly: Volume 38, Issue 1 pp. 75-105, 2004

[14] P. Eronen and J. Zitting, “An expert system for analysing firewall rules”, In Proceedings of the 6th Nordic Workshop on Secure IT Systems (NordSec 2001), pp. 100–107, November 2001.

[15] M. Abedin et al., “Detection and Resolution of Anomalies in Firewall Policy Rules”, In Proceedings of the IFIP Annual Conference Data and Applications Security and Privacy, 2006, LNCS 4127, pp. 15–29

[16] Y. Bartal, A. Mayer, K. Nissim, and A. Wool, “Firmato: A novel firewall management toolkit”, Proceedings of the 1999 IEEE Symposium on Security and Privacy, pp. 17-31, Oakland, California, May 1999

[17] A. Wool, “Architecting the Lumeta firewall analyser”, In Proceedings of the 10the USENIX Security Symposium, Washington DC, August 2001

[18] S. Hinrichs, “Policy-based management: Bridging the gap”, In Proceedings of the 15th Annual Computer Security Applications Conference, Phoenix, Arizona, December 1999, IEEE Computer Society Press.

[19] A. Mayer, A. Wool, and E. Ziskind. “Fang: A firewall analysis engine”, In Proceedings, IEEE Symposium on Security and Privacy, pp. 177-187, IEEE CS Press, May 2000

[20] S. Hazelhurst, “Algorithms for analysing firewall and router access lists”, Technical Report TR-WitsCS-1999-5, Department of Computer Science, University of the Witwatersrand, South Africa, July 1999

[21] E. Al-Shaer and H. Hamed, “Design and Implementation of firewall policy advisor tools”, Technical Report CTI-techrep0801, School of Computer Science Telecommunications and Information Systems, DePaul University, August 2002

[22] E. Al-Shaer and H. Hamed, “Discovery of policy anomalies in distributed firewalls”, In Proceedings of the 23rd Conf. IEEEE Communications Soc. (INFOCOM 2004), Vol 23, No.1, pp. 2605-2616, March 2004

[23] E. Al-Shaer and H. Hamed, “Taxonomy of conflicts in network security policies”, IEEE Communications Magazine, 44(3), March 2006

[24] E. Al-Shaer, H. Hamed, R. Boutaba, and M. Hasan, “Conflict classification and analysis of distributed firewall policies”, IEEE Journal on Selected Areas in Communications (JSAC), 23(10), October 2005

[25] A. Hari, S. Suri, and G.M. Parulkar, “Detecting and resolving packet filter conflicts”, In INFOCOM (3),pp. 1203-1212, March 2000.

[26] D. Monahan EMA, Research Summary: “Network Security Policy Management tools – Tying Policies to Process, Visibility, Connectivity and Migration”, https://web.tufin.com/network-security-policy-management-tools-ema-research, [retrieved: April, 2019]

[27] Algosec whitepaper, “Firewall Management: 5 challenges every company must address”, URL https://www.algosec.com/resources/ [retrieved: April, 2019]

[28] C. Cunningham and J.Pollard, “The Eight Business and Security Benefits of Zero Trust”, Forrester Reseach November 2017

[29] W.R. Stevens, “TCP/IP Illustrated”, Volume 1, the Protocols, Addison-Wesley Publishing Company, ISBN 0-201-63346-9, 1994

[30] H. Zimmermann and J.D. Day, “The OSI reference model - Proceedings of the IEEE”, Volume: 71, Issue: 12, Dec 1983

[31] S.Chen, M. Yoon and Z. Zhang, ”Minimizing the Maximum Firewall Rule Set in a Network with Multiple Firewalls”, IEEE Transactions on Computers, Volume 59, No.2, 2010

Terminology Management in Cybersecurity through Knowledge Organization Systems: an Italian Use Case

Claudia Lanza*, Elena Cardillo†, Maria Taverniti†, Roberto Guarasci*

*University of Calabria, Rende, Italy
Email: c.lanza@dimes.unical.it; roberto.guarasci@unical.it
†Institute of Informatics and Telematics, National Research Council, Rende, Italy
Email: {elena.cardillo;maria.taverniti}@iit.cnr.it

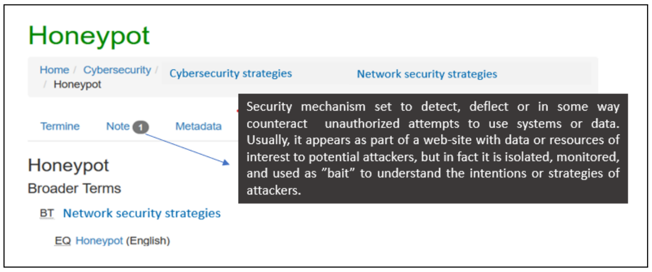
Abstract—**Specialized terminology is usually managed by Knowledge Organization Systems (KOSs), which manipulate and organize concepts and terms through standardized structured techniques. In this paper, an approach to organize, manage, and subsequently update specialized terminologies, specifically related to the domain of Cybersecurity, is proposed. A preliminary analysis and comparison between KOSs showing a higher level of semantic representation, i.e., thesauri and ontologies, is presented in the first section with the objective of clarifying the conceptual framework of these resources. A concrete use case in the domain of Cybersecurity is then described to show the context of application of these two semantic resources, i.e., a project funded by the Institute of Informatics and Telematics of the National Research Council aimed at providing terminology management and representation in the frame of the Italian Cybersecurity Observatory. A transaction between the thesaural and ontological representation of the domain knowledge represents the core of the approach showing the higher qualitative value that ontologies are able to provide to represent the domain of interest, due to the more precise formalization of semantic relationships existing among concepts.**

Keywords- Cybersecurity; KOS; Thesauri; Ontologies; Semantic relations.

I. Introduction

Managing technical terms proper to specialized languages represents one of the main tasks of Knowledge Organization Systems (KOSs). In the context of KOSs, semantic resources, as, for example, thesauri and ontologies, are useful tools to organize domain specific knowledge and to support processes like document indexing, information searching and retrieval and, in some cases, automatic reasoning (e.g., for decision making), above all in those specialized domains where semantic ambiguity between terms represents a step to be avoided. During the last few years some effort has been spent, as shown in Section III, on the definition of ontological models, used in the domain of Cybersecurity, aimed at supporting systems to better identify vulnerabilities and, thus, supporting decision making. Nevertheless, the specificity of the domain and the constant updates of the related information and data, the need for more appropriate semantic resources, based on standards, and highly structured to better represent the domain knowledge, is still evident. This is even more true in the Italian context, where there is a lack of highly semantically structured ways to manage the terminology of this field of study. Taking inspiration by this scenario, the present paper, which is an invited extension of [1], is focused on presenting a preliminary analysis of the main differences existing in the way of organizing and representing the information related to highly specialized domains, targeting the analysis on Cybersecurity. Amongst the KOSs [2] the comparison will focus on two means of semantic knowledge configuration: thesauri and ontologies. The reason why these two types of resources have been selected among others mainly relies on one of the main objectives of the Italian *OCS Project* coordinated by the Cyber Security Observatory of the Institute of Informatics and Telematics, National Research Council (IIT-CNR) [3], presented in detail in Section IV, which provides the understanding of the technical domain of Cybersecurity for a community of users demanding a guided orientation in this field of knowledge. The second purpose of the present work is twofold: (i) to show the results of the above mentioned project, whose main objectives are the development of an Italian and standardized controlled vocabulary, in other words a thesaurus [2] for the Cybersecurity domain, which can be considered a reliable knowledge organization system that structures the information related to specialized domains; (ii) to enhance of its semantic relationships and representation by exploiting a more formal language, i.e., the Web Ontology Language (OWL) [4], the recommended Semantic Web language for authoring ontologies.

The utility of this resource provided in the Italian scenario (and for this reason in Italian language), is specifically addressed to Italian medium-sized companies, citizens, stakeholders and scholars at different levels who need a key access point to better understand and reduce ambiguity dealing with Cybersecurity terminology. The vagueness of certain terms is due to the fact that the majority of them, coming from a domain, which, by essence, is characterized by a predominant usage of English multi word units, are given in their original English version to keep their meaning even when applied to other language use cases and contexts. The present use case implies the involvement of Italian Cybersecurity institutions and training organizations, so the transfer learning process is essential to guarantee the uniformity of key concepts in the Cybersecurity domain either found in sector-oriented magazines and laws or regulations (also in grey literature).



Security mechanism set to detect, deflect or in some way counteract unauthorized attempts to use systems or data. Usually, it appears as part of a web-site with data or resources of interest to potential attackers, but in fact it is isolated, monitored, and used as ”bait” to understand the intentions or strategies of attackers.

Honeypot
Broader Terms
BT Network security strategies
EQ Honeypot (English)

Fig. 1. Thesaurus representation of *Honeypot*.

To give an example, the term *Honeypot* has no corresponding term in Italian language; consequently, to maintain its practical meaning, terminologists in the transfer learning operations should leave the English form as to provide a strong homogeneous informative flow within organizations that are supposed to share common official knowledge (see Figure 1 above to see the use of *honeypot* in a thesaurus structure). To enable users to refer to a uniform resource that spreads specialized information onto several technical databases in a unique modality, the structure of the thesaurus allows the insertion of a Scope Note (*SN*), that is a targeted definition of the terms. This definition is taken from authoritative sources, such as sector-oriented glossaries, standards, official guidelines, etc. This additional feature provides a better unified structure between systems shared under different languages. Moreover, one of the main outcomes of this research activity is strictly linked to the possibility of integrating the Italian thesaurus and the ontology in an automatic threats recognition system, which is intended to monitor terms and concepts and to detect the appearance of new ones without much human effort.

Some of the considered resources to build the source corpus useful to obtain a list of representative terms are hereafter summarized. Representative terms synthesize the concepts belonging to a specific domain and provide the starting model to realize, in a second step, an ontology for Cybersecurity, which is, consequently, based on the structure created for the Italian thesaurus. The ontology has been developed with the goal of representing the classes linked to each other through more precise properties that could, at times, specify the interconnections between them better than a flat visualization that belongs to a thesaural organization of terms. The paper is structured as follows: Section II presents the theoretical background for both thesauri and ontologies in order to highlight which are their main characteristics and the advantages in using them for organizing and representing highly technical domains. Section III gives an overview of the state of the art, presenting related works focused on Cybersecurity information management, both in English and Italian, and on the construction of KOSs. Section IV describes the construction of the Italian thesaurus for Cybersecurity and its enhancement through an ontological representation. Section V provides a discussion about the main advantages derived from exploiting thesauri and ontologies in the described Italian use case. Finally, Section VI sums up the key issues underlined in the paper giving some overall remarks and future perspectives.

II. Background

In this section, a theoretical background is presented to describe and clarify characteristics, purposes, differences and advantages of the two main structured KOSs, i.e., theusauri and ontologies. This will introduce the reader to the approach proposed in Section IV to build such terminological resources for the Cybersecurity domain.

A. Thesauri

Thesauri's main scope is that of structuring information and organizing it in a layered network of semantic connections, and its management and usability is piloted by KOSs functionalities [5][6]. As Soergel affirms in his work, "A thesaurus is a structured collection of concepts and terms for the purpose of improving the retrieval of information. A thesaurus should help the searcher to find good search terms, whether they be descriptors from a controlled vocabulary or the manifold terms needed for a comprehensive free-text search — all the various terms that are used in texts to express the search concept" [7]. The way thesauri are structured follows standardized rules that should be respected, as the ones included in the ISO standards 25964-1:2011 and 25964-2:2013 [8][9], and the objective of uniforming a lexicon meant to be a reference for a community of domain-oriented users is pursued. A thesaurus should provide a reliable and a well structured semantic means to guide the understanding of technical terms representing concepts belonging to a specific field of knowledge. Its indexing function proves to be helpful in the way the users are able to analyze documents according to an informative organization of descriptors. In other words, the abstraction of knowledge occurs indirectly by exploiting terminological units that take on the status of descriptors or indexing units. The latter is the element that language uses to describe, synthesize and extract information from documents [10]. Thesauri's terms undergo both quantitative and quality control. Quantity control refers to thesaurus' terms selection among those that represent in a better way the concepts of the domain of study. These latter become descriptors of the thesaurus (i.e., preferred terms) and usually are followed by the non-preferred terms that act as synonym entries, e.g., *Malicious software* is the preferred term instead of *Malware* in the Italian Cybersecurity thesaurus. In detail, as suggested by the mentioned standard, countable terms have to be expressed in plural form (trees and not tree), and semantically the control is always granted by the respect of the biunivocal relationship existing between terms and concepts (only one concept corresponds to a term and viceversa). That means that the ambiguity of the natural language is controlled and reduced to zero through the use of a limited set of terms (indexing terms) that represent the concepts in a given domain. In this scenario the user who selects a search term and the indexer who chooses indexing terms are both guided to use the same term for the same concept [8]. Thesauri present three main standardized forms of connections that are generated for structuring the information, and five abbreviation codes used to represent such relationships within the controlled vocabulary:

1) Equivalence relation, with the tags Use (USE) and Used For (UF), expresses the synonymy property:
   - Usage:
     Cyber minacce UF *Cyber Threat Actors*; *Cyber Threat Actors* USE *Cyber Minacce*
   - Acronyms:
     Virtual Private Network UF *VPN*; *VPN* USE *Virtual Private Network*
   - Synonymy control:
     Cyber attacks UF *Cibernetic attacks*; *Cibernetic attacks*; USE *Cyber attacks*
2) Hierarchical relation, with the tags Broader Term (BT) and Narrower Term (NT), exists when having two concepts and one of them is part, or is included in the other:
   - Whole/parts:
     Vulnerabilities NT *Software vulneabilities*; *Software vulneabilities* BT *Vulnerabilities*
   - Class/member:
     Logic bombs NT *Elk Cloner*;
     Elk Cloner BT *Logic bombs*
3) Associative relation, with the tag Related Term (RT), covers associations between pairs of concepts that are not hierarchically related [8]:
   Cyber war RT *Cyber weapon*.

The aforementioned standards also guide the way terms should be defined to indicate a unique and unambiguous meaning. The use of a Scope Note is useful when an indexer needs to fix the boundaries of a concept within a domain. Scope Note is marked with the tag SN. An example of SN can be:

Phishing - SN: Tricking individuals into disclosing sensitive personal information through deceptive computer-based means.

The choice to privilege a thesaurus structure instead of other semantic resources, such as glossaries or taxonomies, relies on its peculiarity of managing the representative terms of a specific domain as an entangled network of semantic relations that guide the comprehension of a conceptual model proper of a field of knowledge to be studied [11].

B. Ontologies

The term ontology, which has been borrowed by the Artificial Intelligence (AI) community from phylosophy, gained new definitions and found a broad spectrum of applications in various branches of computer science [12]. In AI, an ontology is considered to be an engineering artefact, which is constituted by a specific vocabulary used to describe a certain reality, plus a set of explicit assumptions regarding the intended meaning of the vocabulary. Gruber defines it as “An explicit specification of a conceptualization” [13], so, in simple words, a formal specification of a domain of knowledge. In order to formally represent a certain domain, ontologies use a set of constructs describing the world in terms of classes, properties, and individuals. To enrich the formalization, other constructs are used for expressing complex descriptions in terms of relations between classes, cardinality, equality, etc. Consequently, it is possible to say that an ontology consists of a set of definitions of classes, properties, and individuals, together with a set of axioms (i.e., formal restrictions) expressing the relations between classes and properties, and a set of facts about particular individuals. Just like thesauri, ontologies define a common vocabulary (for a specific domain) and a shared understanding. We can have different ontologies according to the used level of formalism: (i) light-weight ontologies (i.e., ontologies that represent only the hierarchical level of concepts and relations in a domain, so, more commonly, taxonomies); and (ii) heavy-weight ontologies (i.e., lightweight ontology enriched with axioms used to fix the semantic interpretation of concepts and relations). Ontologies are used to share knowledge between people, agents, and software thanks to their characteristics of enabling the reuse of domain knowledge and making domain assumptions explicit. Another important feature is that through ontologies it is possible to represent both domain knowledge and operational knowledge and reuse them separately, enabling in any case automated reasoning. The importance of an ontology as a means of structuring knowledge is well recognized in different areas, such as, knowledge representation, knowledge management, natural language processing (NLP), multi-agent systems, database integration, web services, and others. The literature is full of significant academic research devoted to the development of a theoretical and practical basis of ontology technology. Among others, the most notable developments have been the world wide web consortium standardization of expressive representational languages for publishing ontologies on the web [14] [15]. From a practical point of view, the methods followed for building ontologies observe basic principles that can be found in guidelines like the one published by Noy and Mcguinness [16] or Bourigault [17].

The OWL language helps in building formal, sound and consistent domain-specific terminologies, and provides a standard web accessible medium for interoperability, access and reuse. OWL uses RDF (Resource Description Framework) for its syntax, the prescribed framework for representing resources in a common format, describing information in the form of *subject-predicate-object* triples, thus enabling to represent them in the form of a graph. Three different OWL sublanguages can be used according to the formalism we want to give to our ontology and to the performances in reasoning and inference we want to obtain: OWL Lite, OWL DL, and OWL Full. The first sublanguage is the least powerful one, in fact it allows to represent taxonomies and uses less constructs (it includes cardinality restrictions). For this reason it has the lowest computational complexity among the OWL sublanguages. The second one, OWL DL (i.e., Description Logic) provides a more formal representation since it imposes restrictions on the usage of OWL/RDF constructors. This sublanguage is used when the maximum decidable expressivity is required and is able to maintain computational completeness (that means that all conclusions are computable). Finally, OWL Full is the most expressive one, since it uses all the OWL language primitives and all of the RDF Schemas (RDFS) and, with respect to the other two sublanguages, it is undecideable, semantically difficult to understand and to work with, and, as a result, standard automatic reasoning techniques cannot be applied. Differently, because of its formalism, OWL DL allows reasoning and inference. Reasoning is the act of making implicit knowledge explicit. To infer knowledge from ontologies, reasoning engines are used, which allow determining also subsumption, classification, equivalence, and identifying ontology inconsistencies [15].

Ontology similarities with thesauri can be easily identified after this theoretical description. In particular, both describe and organize a domain, include concepts and relations between them; they use hierarchies, and describe instances belonging to concepts. Both of them can be applied for information management, for cataloguing and in search engines. However, several differences must be considered. First of all, thesauri had as their original purpose that of being used in librarian contexts as indexing tools and controlled vocabularies. So, it is understandable that they are thought to represent knowledge in a less formal and comprehensive way with respect to ontologies. On the contrary, because of their philosophical origin, ontologies are characterized by a high level of conceptual abstraction, which is accepted, and formal ways of describing domain knowledge. Regarding their structure, as seen above, ontologies are characterized by the explicit representation of the types of relationships and by the use of powerful formalisms, which are not possible to define within thesauri (e.g., axioms, relationships, cardinality). Therefore, to represent hierarchical relations between classes and subclasses, two declared relations are used, i.e., “is-a” and “kind-of”, while, to represent meronymy relations between classes, the “part-of” relation is employed. By contrast, in the thesaurus those relationships are treated as hierarchical relationships. Finally, the associative relations in an ontology are made explicit according to the exact connection (predicate) that exists between two classes. For example, taking up concepts already used in Section II-A , *cyber war* `RT` *cyber weapon*, is specified in an ontology as *cyber war* `uses` *cyber weapon*, where “used” is the ObjectProperty.

The interoperability of semantic resources like thesauri and ontologies, is given by the principle of linked open data [18][19][20], which guarantees a shareable knowledge organization system that can facilitate the coordination among several users for different terminological tasks. To generate a language that can guarantee a higher form of interaction between informative systems, without losing the exact meaning of the shared information, the ontology seems to route towards a constant reuse of the managed information by providing conceptual representations of a domain [21][22].

III. Related works

When terminologists’ activity involves the construction of knowledge organization and representation systems, the phase of taking into account which could be conceived as gold standards represents a key step in order to align the information retrieved by source corpora to texts that represent the reference standards [23]. The research activity presented in this paper starts as a monolingual - Italian - study for Cybersecurity terminology. Therefore, the starting point to develop an Italian controlled vocabulary on Cybersecurity has represented the census of the gold standards. Among the existing examples of Cybersecurity glossaries and vocabularies, of great importance are: for English, the ones contained in the NIST 7298 [24] and ISO 27000:2016 [25] standards for Information and Communication Technologies (ICT) security, and, for Italian, the Italian book ”*Libro Bianco*” (White Book for Cybersecurity) realized by the National Laboratory of Cybersecurity of the Consorzio Interuniversitario Nazionale per l’Informatica (CINI) [26], which thoroughly sheds light on the key issues related to Cybersecurity guidelines and on the latest related episodes that have changed the way to defend informative systems and to conceive some specific concepts proper to Cybersecurity. Another relevant existing resource for Italian is the ”*Glossario Intelligence*” [27], a technical glossary published by the Italian Presidency of the Council of Ministers, which contains several terms belonging to the Cybersecurity domain and which has been used as a basis for the creation of the Italian thesaurus and the ontology for Cybersecurity under investigation.

With respect to ontologies, it is worth mentioning the works targeted at the creation of ontology models for Cybersecurity, i.e., [28][29][30], and the studies focused on the approaches for developing an architecture for Cybersecurity standards [31] and enterprise’s Cybersecurity metrics [32]. In particular, in [33] an ontology has been designed to integrate data from different heterogeneous sources, in the absence of a common terminology, offering a sufficiently complete knowledge on the possible threats, thus allowing organizations to perform reasoning and support decision-making processes related to security. Another study proposed a reference ontology for Cybersecurity operational information, developed, as in our case, in collaboration with Cybersecurity organizations, and which had the aim to review industry specifications. Here, types of Cybersecurity information are defined along with the roles and operation domains (see [34] for details). Finally, a more recent work describes the development of an ontology of metrics for Cybersecurity assessment [35]. This ontology is based on determining the concepts and relations between primary features of initial security data and forming a set of hierarchically interconnected security metrics. Application of the approach is shown on a case study. The main feature of this work is the representation of security metrics as separate instances of the ontology, which allows using the relations between the concepts of ontology for calculating integral metrics reflecting the security state.

Processing the information belonging to specific domains of interest involves the analysis of those documents which semantically tend to represent concepts through a technical language [36]. The creation of terminological databases follows some given criteria linked to gathering the related documents that have to constitute the reference corpus from which terms can be retrieved [36]. To achieve this first informative structure, the corpus firstly aims at including documents that can represent the domain in an official way [37], i.e., the gold standards [38], collecting a terminological standardized repository made up of terms that are meant to be closely specific to the technical field of knowledge under review [39].

To obtain a matching system between the terminology shared by a community of experts from a particular domain and the terms contained in a list derived from the processing of a reference corpus, the documents gathered in the corpus undergo a process of terminology extraction, which shall compare the equivalence between the representative terms of a domain with the ones of the gold standards [40].

This last step is usually implemented by exploiting semi-automatic term extraction tools. Nazarenko *et al.* [41] and Loginova [42] gave in their works detailed lists of several tools for extracting terminologies from texts. With regards to the Cybersecurity domain and the research activity treated in this paper, various existing sources, both in English and in Italian, have been analyzed in order to retrieve an accurate terminological basis from which to build a more sophisticated semantic resource to guide the knowledge representation process. The intent of this project task, as aforementioned, is to provide an Italian resource, firstly conceived as a thesaurus, to configure the terminology of Cybersecurity in a network of semantic relations that can better orientate to a lexical understanding of specialized concepts represented by terms belonging to this field. The goal of this research activity is also based on the reuse of the terms contained in the thesaurus to realize in a consequential way an ontology system that could support the inclusion of customized properties between classes and more comprehensively clarify the associative relationships used in the thesaurus [43][44][45]. This represents the reason why ontologies can usually be considered as resources that can provide a more exhaustive and explicit frame for knowledge representation.

IV. The OCS Project

In this section, the project use case is presented. The first part is focused on the description of the Cybersecurity context and the Italian Cybersecurity Observatory (OCS) scopes and services. The second part presents the thesaurus itself for managing the information about Cybersecurity and its enhancement through its migration into an ontology system.

The main objective of the activity, as mentioned before, is the creation of a thesaurus in Italian language to be used as a semantic tool to organize the terminology related to Cybersecurity, and to be inserted amongst the services of the online platform of the Italian Cybersecurity Observatory [46]. The OCS online platform is a joint work with the experts of the Cybersecurity domain that aims at gathering different services to guide the comprehension of the phenomena occurring in this field of study. For instance, apart from the semantic tools section, to which the Italian thesaurus and the ontology for Cybersecurity belong, this web service includes the analysis and detection of tweets, threats, vulnerabilities, exploits, spam mails, attacks, malware, self-assessment.

The convergence of the semantic tasks with the experts of the domain can be achieved in considering their documentation collections, consisting, among others, of the lists derived from the Common Vulnerabilities Exposure (CVE)[1], or of internal detections of the main cyber attacks, as sources to be used to update the terminology of the domain to be represented.

Indeed, the list of vulnerabilities, the spam detections or the analysis of the latest cyber threats, could represent, in a future perspective, the meeting point between the goal, by the OCS platform, of sharing technical information to defend informative systems and, by terminologists, of providing extra knowledge that can empower the terminological organization of the domain. In this way, both the thesaurus and the ontology can undergo a rethinking phase both on new highly technical term inclusion level and, consequently, the relational one.

A. The Cybersecurity context

The Cybersecurity domain is mainly characterized by a technical terminology. Given that Cybersecurity is a synergy of different sub-fields, the schematization of this specialized domain reflects this high level of heterogeneity. Cybersecurity is permeated by: (i) its multidisciplinary nature that involves Information and Communication Technologies (ICT) and its sub-areas, such as, audiovisual techniques, computer software, electronics; (ii) its specificity with respect to technical and standardized terms; and (iii) its cross-fielding thematic coverage, i.e., computer science field, legislative systems, regulations. Given these premises, the treatment of its internal language, which derives from the textual content extracted from the source corpus documents, is meant to be managed by formal semantic systems in order to obtain shareable standardized lists of the domain's representative terms, organized according to their semantic relations, which, in turn, will orientate the understanding of the conceptual model of the domain [47].

As can be observed by looking at Figure 2, the OCS website, developed for the purposes of spreading the information about Cybersecurity for the Italian community of experts and common users, registered many views on its overall level range. This high number of users coming from several countries denotes the significant interest the organization of the platform has. Nonetheless, the superior percentage of Italian users shows how the target language played an important role in orientating the ways in which the technical structuring of

[1] https://cve.mitre.org/

information about the domain has been set up. The thesaurus and the ontology presented in this paper have been included inside the OCS web page as two tools that provide a semantic outline about the information meant to be structured on the Cybersecurity domain. Even though the numbers reported are not remarkably outstanding, it can be stated that both of them have received attention especially during two Italian events during which they have been presented to an audience.

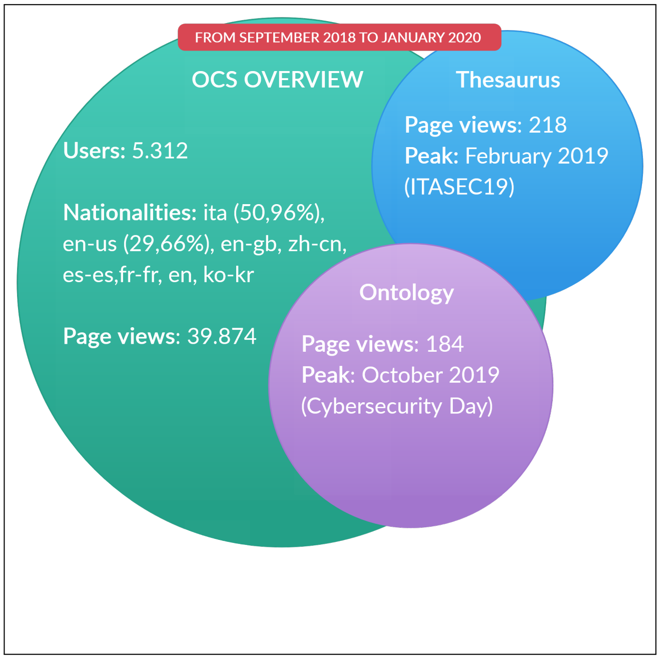


Fig. 2. Statistics OCS website.

B. The Italian thesaurus for Cybersecurity

The main focus of this paper is the creation of the Italian thesaurus on Cybersecurity for the OCS project [46], carried out in collaboration with the Institute of Informatics and Telematics of the National Research Council.

The methodology followed for the realization of the thesaurus covered classical sequences. As primary step, the terminology to be included in the thesaurus has been extracted from reliable sources which made up the corpus characterized by documents distinctively selected for their content oriented to Cybersecurity issues [37]. This collection of texts made the information retrieval highly oriented to the domain to be represented [48], and covered different types of documents, such as, standards and laws [49], Cybersecurity-related magazines or guidelines and certifications. The conceptual content of these documents was meant to be processed to obtain lists of terms (a glossary) sorted according to statistical measures able to provide a first semantic organization [50]. Indeed, the second phase concerned the semi-automatic processing of the information included in the source corpus by exploiting a term extractor software [**?**] (more specifically the Italian native tool, *Text to Knowledge* (T2K)) [51] that provided, as outputs, lists of terms ranked according to their occurrence's value in the texts. Terms selection has been based on frequency, in particular terms with the highest scores in TF-IDF values have been considered as candidate terms to be part of the Italian thesaurus for Cybersecurity. The list of the most representative terms accompanied by their frequency scores has undergone an evaluation process carried out by a group of domain experts.

Indeed, only once having received the validation by domain experts,– the third phase of the methodology –, the terms have been selected as candidate terms to be integrated in the thesaurus and their semantic relations with other terms of the domain, derived from the corpus, have been created. The current Italian Cybersecurity thesaurus contains 246 candidate terms, already validated by domain experts collaborating on the project, and mapped to the taxonomies contained in the main gold standards for Cybersecurity, i.e., NIST 7298 [24] and ISO 27000:2016 [25]. The alignment with the terms contained in the standards for ICT security granted a coordination between the knowledge shared by an international Cybersecurity community of experts and the one represented in the structured thesaurus, which is composed of preferred terms selected amongst those extracted by the T2K tool as the most frequent inside the source documents. In order to carry out a matching configuration with the standards as predictable and stable as possible, the terms included in the standards, and selected with the support of domain experts as key guidance representing the domain, have been translated using the Interactive Terminology for Europe (IATE) term banks [52]. This is considered an important step given the instructive purpose of the application, i.e., the use of the thesaurus in the web portal of the Cybersecurity Observatory. The main entries in the Italian thesaurus for Cybersecurity are the four macro categories finely selected from the extracted glossary, also according to the frequency of terms, and from the mapping with the standards alongside the approval by the domain experts. These macro categories are:

- Cybersecurity;
- Cyberdefence;
- Cyberbullism;
- Cybercriminality.

Almost each of the candidate terms included in the thesaurus network, generated by the semantic relations among the terms, are accompanied by their definitions, i.e., *Scope Note (SN)*, which helps in understanding the terms in their specific contexts giving their definition taken from the source documents [53].

For a better understanding of the actual size of the Italian Thesaurus for Cybersecurity, Table I gives a metrics of the numbers of terms, as well as of the semantic relations (SemRel).

TABLE I. Features of the Italian thesaurus for Cybersecurity.

	Terms	SemRel	Non-preferred Terms	SN
Total	246	280	33	74

TABLE II. Cybersecurity ontology metrics.

Metric	**Total**
Axiom	640
Logical axiom count	316
Declaration axioms count	233
Class count	157
Object property count	37
Data property count	7
Individual count	31
Annotation Property count	5
CLASS AXIOMS	
SubClassOf	58
EquivalentClasses	0
DisjointClasses	24
OBJECT PROPERTY AXIOMS	
SubObjectPropertyOf	7
InverseObjectProperties	1
FunctionalObjectProperty	1
TransitiveObjectProperty	0
SymmetricObjectProperty	1
AsymmetricObjectProperty	0
ObjectPropertyDomain	40
ObjectPropertyRange	39
DATA PROPERTY AXIOMS	
SubDataPropertyOf	1
DataPropertyDomain	8
DataPropertyRange	5
INDIVIDUAL AND ANNOTATION AXIOMS	
ClassAssertion	31
AnnotationAssertion	89

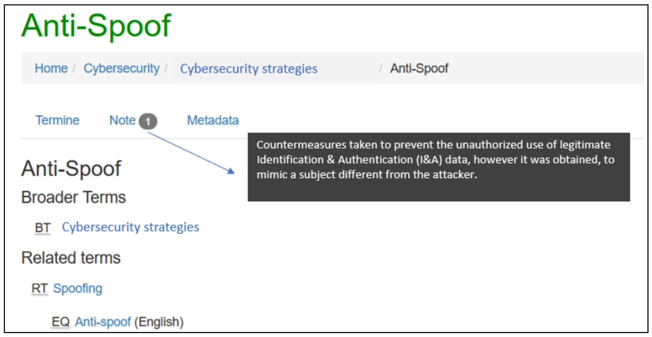
C. *Ontology enhancement*

Another activity of the OCS project has also been focused on the migration of the thesaurus elements into a more formal semantic resource, i.e., an ontology, to better organize and represent the information about Cybersecurity, addressed to users who want to get closer to this field of knowledge [54]. Details on the ontology structure are provided in Table II. Among the main objectives in rengineering a thesaurus into a system working with OWL language there is that referring to the capture of significant real time new terms occurrences in the future, especially following the updates given by the major official sources in the Cybersecurity domain. Indeed, what ontologies allow more than a thesaurus is to exploit the query system operations that enable users to activate reasoning engine operations which are meant to infer semantic connections from several resources given in input as conceptual models. The formalization of a thesaurus into an ontology is a task that has been attracting much interest. In fact, in the literature, different approaches have been proposed for reusing thesaurus semantic content to build ontology meta-models and to populate knowledge bases in different domains, see for example [43][55][56].

The need for migrating the content included in the thesaurus into an ontology lies in the decision to better clarify the associative relationships between the terms of the thesaurus [57]. In particular, the flat modality in which the associative relationship between terms is represented in the thesaurus, i.e., via the RT relation, turned out to be not fully satisfactory in the seek of getting a complete terminological outline for Cybersecurity [58].

As shown in Figure 3 and Figure 4, there is a clear distinction between the two systems used to organize and represent the terminology belonging to Cybersecurity. The example taken into account to represent the differences is referred to the semantic relationship linked to the idea of opposition, i.e., *Spoof* and *Antispoof*: in the thesaurus, even though a definition is present (within the black square), which corresponds to the Scope Note (*SN*), proper to thesauri, giving many details on the context from which terms come from, the ”opposition” is not so well represented because it is only shown through the associative relation (*RT*) [8] between these aforementioned terms without giving other explications on the way the two terms are related as the OWL language does.

On the other hand, in the ontology, these two concepts are connected through the *ObjectProperty* ”HasAsContrary” that helps in considering the *Domain* and the *Range* as linked by a precise relationship.



Anti-Spoof

Home / Cybersecurity / Cybersecurity strategies / Anti-Spoof

Termine Note 1 Metadata

Countermeasures taken to prevent the unauthorized use of legitimate Identification & Authentication (I&A) data, however it was obtained, to mimic a subject different from the attacker.

Anti-Spoof

Broader Terms

BT Cybersecurity strategies

Related terms

RT Spoofing

EQ Anti-spoof (English)

Fig. 3. Thesaurus representation of the semantic relationship that describes opposition.

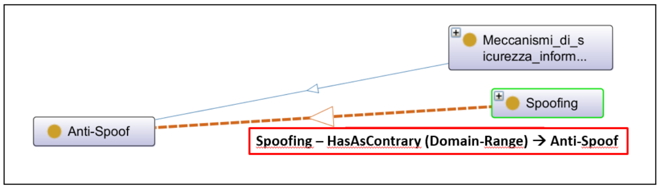
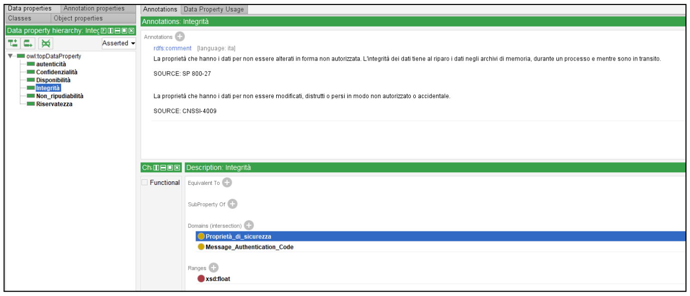

Meccanismi_di_s icurezza_inform...
Spoofing
Anti-Spoof
Spoofing – HasAsContrary (Domain-Range) → Anti-Spoof

Fig. 4. Protégé representation of the semantic relationship that describes opposition.

Another representative case is depicted by Figure 5 and Figure 6, which show how a thesaurus sometimes provides a weak visualization of some attributes associated to a concept. In the following case, the relation that had to be demonstrated was related to several attributes that security properties proper to informative systems own. For this specific purpose, the ontology resource gives more advantages in the visualization of the informative structure allowing a higher accurate organization and representation of the attributes related to the concepts. In detail, the main difference that makes ontologies a good semantic means to represent the conceptual model connected to certain semantic classes is related to the fact that, in this case, the security properties, i.e., integrity, authenticity, confidentiality, availability, reliability, non-repudiation, and privacy, are represented as *Data Properties* and are conceived as attributes. In the thesaurus, as shown in Figure 6, they are related to the hyperonym *BT* ”Data” and are represented as its specific terms, i.e., the *NT* [9].



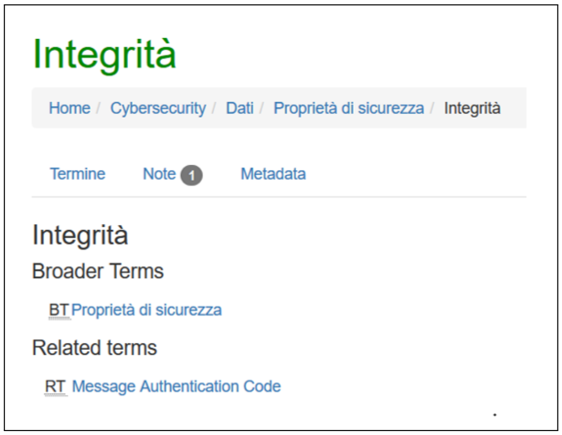
La proprietà che hanno i dati per non essere alterati in forma non autorizzata. L'integrità dei dati tiene al riparo i dati negli archivi di memoria, durante un processo e mentre sono in transito.

SOURCE: SP 800-27

La proprietà che hanno i dati per non essere modificati, distrutti o persi in modo non autorizzato o accidentale.

SOURCE: CNSSI-4009

Fig. 5. Ontology representation of Security properties as *Data Properties*.


Home / Cybersecurity / Dati / Proprietà di sicurezza / Integrità
Termine
Note 1
Metadata

Integrità

Integrità

Broader Terms

BT Proprietà di sicurezza

Related terms

RT Message Authentication Code

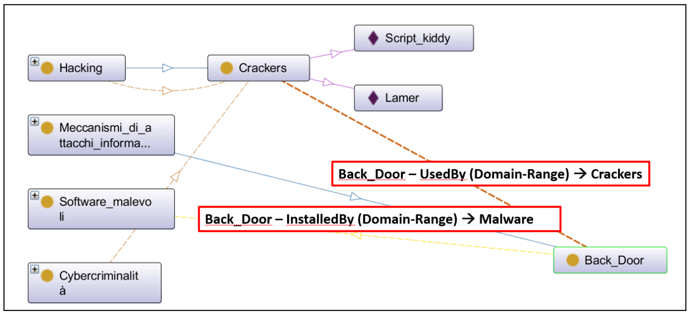
Fig. 6. Thesaurus representation of Security properties as hierarchical relations.

As mentioned before, the ontology has been forged under the basis of the thesaurus structure to organize the Italian terminology about Cybersecurity. For this reason, and considering that the main purpose of the ontology is terminology control and the appropriate semantic representation of the domain concepts, the OWL sublanguage selected for this scenario is OWL Lite. The connections between the terms have been transposed to the ontology object properties, and referred to the information contained in the source corpus documents. To increase the level of accuracy and domain-oriented information representation, the ontology has been enhanced using pattern path-variables configuration [59]. In particular, after having collected a group of passive verb pairs from the Italian Cybersecurity source corpus, a filtering procedure over the most technical ones related to the domain has been launched. From a list of verbs that have been considered domain-dependent, several queries in the source corpus have been run and analysed in order to create the associative connections among the concepts [60]. As Table III shows, the relations that have been retrieved from the reference corpus by using certain pattern paths are very detailed and they are progressively being added to the existing ontology, which has been developed by migrating the content of the Italian Cybersecurity thesaurus. The aim is to guarantee a more precise semantic system that can structure the interconnections among Cybersecurity concepts with the help of interoperable languages. In the case of ontologies, the *Object Properties* are the ones to cover the purpose of providing a more targeted form of associative relationships to represent the information of the domain.

To give a clearer idea of how these associations have been reported into the ontology so far, with the perspective of continuously augmenting the range of relations, Figure 7 gives an highlight of how, for example, the concept *Backdoor* has been connected with *Malware* and *Cracker* by using complementary patterns configurations on Protégé console. Up to now, 160 new associative relations referred to the domain verbal constructions have been selected among the semantic information contained in the texts making up the source corpus, and they are currently being analysed from a linguistic point of view in the co-occurrency level and text scope to increase the semantic relationships meant to be represented. This last passage assumes the activity by terminologists to retrace the semantic entangled network in the text correspondences, and doing so, isolating other constructs as drills of new connections.

TABLE III. Associations retrieval in ontology by patterns configurations.

Associative relationships List			
Aggirare (*by-pass*)	Attaccare (*at-tack*)	Sfruttare (*exploit*)	Attivare (*acti-vate*)
worm→software antivirus	exploit sql injection→web applications	crackers → vulnerabilities	backdoor ← malware
cracker → cybersecurity	script kiddies → DDoS	trojan horses → vulnerabilities	trojan → cyber attacks
virus→ cyber-security	network file systems→ DNS spoofing	spam → bot-net	payload← virus/worm



Back_Door – UsedBy (Domain-Range) → Crackers

Back_Door – InstalledBy (Domain-Range) → Malware

Fig. 7. Additional Object Properties through patterns path-variables.

V. Discussion

Across the phases that have characterized this research activity towards the realization of two semantic sources to monitor and manage the terminology of Cybersecurity in Italian language, the configuring procedure for the thesaurus and ontology development proved to be different in their application to the use case. By describing the steps carried out to build the Italian Cybersecurity thesaurus, the importance of providing a semantic structure that could be as much reliable as possible with respect of the information about the domain has been underlined. This reliability system mainly focuses on the ways a thesaurus can guarantee a reflection of the domain by using a semantic relationship network to connect terms with each other and provide a guided orientation to the organization of the domain knowledge. The connections among the representative terms dependent from the domain of study have been structured following the standard guidelines that give the basis for the arrangement of hierarchical, equivalence and associative interrelations. Nonetheless, the thesaurus outline at times proved to be less accurate in the way it depicts the association among certain terms, in particular for what concerns the usage of the *Related Term* association, which shows some vagueness in how it matches domain-oriented terms mainly because of its lack of deeper semantic descriptions. Therefore, the research activity has been finalized to create another semantic resource able to make the semantic relations between the domain concepts, i.e., the ontology, explicit. We observed that this latter knowledge representation system allowed a more customized organization of the concepts that facilitates the process of combining the semantic links. The ontology has also been implemented, in a latest phase, with the inputs resulted by the execution of some pattern configuration approaches. The use of recurrent variables to be searched in the source corpus proved to be an efficient means through which concepts of the Cybersecurity domain could have been correctly correlated. In fact, we used different domain-oriented verb pairs to show how the structure of the ontology, which has been built following the thesaurus' outline, has been enhanced.

Although thesauri and ontologies belong to the same family of knowledge organization systems and some of their functionalities are the same (e.g., their use for improving information retrieval and knowledge organization), they are built for different purposes. In fact, it has been demonstrated in this contribution that ontologies allow higher formal representation of knowledge for a given domain, by providing explicit relationships between concepts, disjunctions, by applying data properties for each concept or instance and by providing restrictions that avoid ambiguity in the representation of the meaning and the context of use of a concept and their terms in the domain of reference. However, the two semantic resources might be used together or, as widely demonstrated both in this paper and in the literature, one can be reused to build or populate the other, thus they prove to complement each other, improving the end user's search experience.

The natural structural rigidity of thesauri, given by the use of *a priori* defined semantic relationships (hierarchical, associative and equivalence), seems to be a point against these type of controlled vocabularies; by contrast, such weakness seems to be overcome by the flexibility, scalability and reusability of ontologies that, as stressed by the semantic staircase of Blumauer and Pellegrini [61], compared to other KOSs, bring to a highest level of semantic richness thanks to an internal formal description of concepts. This latter combines a system of relations and properties of the concepts themselves.

Despite this, one of the strengths of the thesaurus compared to the ontology, when used in a specialized domain, is its greater capacity to eliminate ambiguity between the terms through the use of synonymy control [2] and the choice of preferred terms, compared to non-preferred terms for representing the concepts. This guarantees a standardization of technical terms in specialized domains, which can help in the process of unifying, and, by consequence, sharing, a specific field of knowledge's terminology.

VI. Conclusion

The objective of this paper concerned the presentation of the main advantages that could be achieved by using two different types of KOSs, i.e., thesauri and ontologies, to organize and represent a technical domain of study. The field of knowledge on which this paper focuses on refers to that of Cybersecurity, and the main task described is specifically linked to its specialized terminology management.

At the beginning of this paper a general overview of Knowledge organization and representation systems has been provided, successively the analysis has been addressed to the thesaurus organization system overview. In detail, the paper underlined the way this semantic monitoring tool has proved to be a reliable system to structure the information derived from heterogenous sources belonging to the Cybersecurity domain, which is widely characterized by technical terms. Concurrently, attention has also been given to the comparison between the modality of representing in the thesaurus some of the relationships existing among terms, which represent the relevant concepts of the domain, with the ones proper to ontologies through OWL language. The perspective has been oriented to provide a demonstrative outline of ontology peculiarities and advantages when using an existing thesaurus, like the one created in the Italian OCS project framework, as a basis for building the meta-model and populating the knowledge base. The perspective of the research activity both for the thesaurus and the knowledge base in OWL is oriented towards a terminological population extension, and this will involve relationships and restrictions where needed, and new evaluations to be executed. Starting from this objective, pattern configurations have been added as means to retrieve additional relations among domain-oriented concepts. Indeed, we have observed that the use of recurrent linguistic structures helped in trace back which could be considered as genre specific relations. Another motivation that lies behind the choice of taking into account pattern constructs is that of improving the preciseness of the associative relationships proper to thesauri that sometimes proved to be rather vague, i.e., *RT* relation. By selecting some verb pairs targeted to the domain of study, is possible, for instance, to create a new range of more accurate

Object Properties in the ontology, and, consequently, enhance the system that has been converted starting from a thesaurus. This last step clearly implies a rearrangement of the source thesaurus, which will continue to be updated in the source texts to provide a representative set of terms that helps in understanding the technical range of information.

Future works will include a translation in other languages (firstly English) to allow, within the OCS project team, the automatic recognition of cyber threats even from non-Italian sources and improve the thesaurus/ontology usability and sharing them also at an international level. Moreover, the remainder of this work targets at taking into account the insertion of several other types of documents to be part of the source corpus. In particular, following the perspective of getting updated on the changes related to the Cybersecurity domain, documents shall be taken from the social media world, adjusting all the analysis related to the processing of information to the treatment of texts written in a specialized form.

REFERENCES

[1] C. Lanza, E. Cardillo, M. Taverniti, and R. Guarasci, "Knowledge representation frameworks for terminology management in cybersecurity: The ocs project use case," in SEMAPRO 2019, The Thirteenth International Conference on Advances in Semantic Processing, U. o. A. S. G. P. L. U. o. H. A. F. Michael Spranger, Hochschule Mittweida, Ed., Porto, Portugal, September 2019.

[2] M. Zeng, "Knowledge organization systems (kos)," Knowledge Organization, vol. 35, pp. 160–182, 01 2008.

[3] Cybesecurity osservatorio. https://www.cybersecurityosservatorio.it\. Accessed: 2019-08-08.

[4] W3C Web Ontology Language (OWL). https://www.w3.org/OWL/. Accessed: 2019-08-08.

[5] R. Davis, H. Shrobe, and P. Szolovits, "What is a knowledge representation?" AI Magazine, vol. 14, p. 17, 03 2002.

[6] A. Miles and S. Bechhofer, SKOS Simple Knowledge Organization System Reference, ser. W3C Recommendation. United States: World Wide Web Consortium, 8 2009.

[7] D. Soergel, "The art and architecture thesaurus (aat): A critical appraisal," Visual Resources, vol. 10, pp. 369–400, 01 1995.

[8] ISO 25964-1:2011 Information and documentation — Thesauri and interoperability with other vocabularies — Part 1: Thesauri for information retrieval, International Organization for Standardization, August 2011.

[9] ISO 25964-2:2013 Information and documentation — Thesauri and interoperability with other vocabularies — Part 2: Interoperability with other vocabularies.

[10] M. Taverniti, "Tra terminologia e documentazione: estrazione automatica di voci indice da corpora documentali della pubblica amministrazione," Ainformazioni, vol. 1-2/2018, pp. 227–238, 2008.

[11] V. Broughton, Essential Thesaurus Construction. Facet, 2006.

[12] N. Guarino, ""formal ontology and information systems"," in *In Proceedings of the International Conference on Formal Ontology in Information Systems (FOIS-1998)*, 1998.

[13] T. R. Gruber, "A translation approach to portable ontology specifications," *Knowledge Acquisition*, vol. 5, no. 2, pp. 199 – 220, 1993. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1042814383710083

[14] A. Gómez-Pérez and O. Corcho, "Ontology specification languages for the semantic web," *IEEE Intelligent Systems*, vol. 17, no. 1, p. 54–60, Jan. 2002. [Online]. Available: https://doi.org/10.1109/5254.988453

[15] S. Staab and R. Studer, *Handbook on Ontologies (International Handbooks on Information Systems)*. SpringerVerlag, 2004.

[16] N. F. Noy and D. L. Mcguinness, "Ontology development 101: A guide to creating your first ontology," Tech. Rep., 2001.

[17] D. Bourigault and N. Aussenac-Gilles, "Construction d'ontologies á partir de textes," pp. 11–14, 01 2003.

[18] A. A. Shiri and C. Revie, "Thesauri on the web: current developments and trends," Online Information Review, vol. 24, no. 4, pp. 273–280, 2000.

[19] D. Soergel, "The art and architecture thesaurus (aat): A critical appraisal," Visual Resources, vol. 10, pp. 369–400, 01 1995.

[20] M. van Assem, V. Malaisé, A. Miles, and G. Schreiber, "A method to convert thesauri to skos," 06 2006, pp. 95–109.

[21] N. Guarino, D. Oberle, and S. Staab, "What is an ontology?" Springer, Berlin, Heidelberg, 05 2009, pp. 1–17.

[22] D. W. Embley, S. W. Liddle, D. W. Lonsdale, and Y. A. Tijerino, "Multilingual ontologies for cross-language information extraction and semantic search," in *ER*, 2011.

[23] A. R. Terryn, V. Hoste, and E. Lefever, "A Gold Standard for Multilingual Automatic Term Extraction from Comparable Corpora: Term Structure and Translation Equivalents," in *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, N. C. C. chair), K. Choukri, C. Cieri, T. Declerck, S. Goggi, K. Hasida, H. Isahara, B. Maegaard, J. Mariani, H. Mazo, A. Moreno, J. Odijk, S. Piperidis, and T. Tokunaga, Eds. Miyazaki, Japan: European Language Resources Association (ELRA), May 7-12, 2018 2018.

[24] R. Kisserl, Glossary of Key Information Security Terms, National Institute of Standards and Technology, May 2013, NISTIR 7298 Revision 2.

[25] ISO/IEC 27000, Information technology — Security techniques — Information security management systems — Overview and vocabulary, International Standard, February 2016.

[26] R. Baldoni, R. De Nicola, and P. Prinetto, Il Futuro della Cybersecurity in Italia: Ambiti Progettuali Strategici Progetti e Azioni per difendere al meglio il Paese dagli attacchi informatici. Laboratorio Nazionale di Cybersecurity (CINI) - Consorzio Interuniversitario Nazionale per l'Informatica, 2018.

[27] Presidenza del Consiglio dei Ministri - Sistema di informazione per la sicurezza della Repubblica, Il linguaggio degli organismi informativi, Glossario intelligence. https://www.sicurezzanazionale.gov.it/sisr.nsf/quaderni-di-intelligence/glossario-intelligence.html\. Accessed: 2019-08-08.

[28] B. Barnett and A. Crapo, "A semantic model for cyber security," 2011.

[29] L. Obrst, P. Chase, and R. Markeloff, "Developing an ontology of the cyber security domain," in STIDS, 2012.

[30] A. Oltramari, L. Cranor, R. Walls, and P. McDaniel, "Building an ontology of cyber security," CEUR Workshop Proceedings, vol. 1304, pp. 54–61, 01 2014.

[31] M. C. Parmelee, "Toward an ontology architecture for cyber-security standards," in *STIDS*, 2010.

[32] A. Singhal and D. Wijesekera, "Ontologies for modeling enterprise level security metrics," ACM International Conference Proceeding Series, 01 2010.

[33] A. Aviad, K. Wecel, and W. Abramowicz, "The semantic approach to cyber security. towards ontology based body of knowledge," vol. 2015, 01 2015, pp. 328–336.

[34] T. Takahashi and Y. Kadobayashi, "Reference ontology for cybersecurity operational information," The Computer Journal, vol. 58, no. 10, p. 2297–2312, 2015.

[35] E. Doynikova, A. Fedorchenko, and I. Kotenko, "Ontology of metrics for cyber security assessment," in *Proceedings of the 14th International Conference on Availability, Reliability and Security*, ser. ARES '19. New York, NY, USA: Association for Computing Machinery, 2019.

[36] A. Condamines, "Sémantique et corpus spécialisés : Constitution de Bases de Connaissances Terminologiques," Habilitation à diriger des recherches, Université Toulouse Le Mirail, Jun. 2003. [Online]. Available: https://halshs.archives-ouvertes.fr/tel-01321042

[37] G. Leech, The state of the art in corpus linguistics, K. Aijmer and B. Altenberg, Eds. London: Longman, 1991.

[38] G. Bernier-Colborne, "Defining a gold standard for the evaluation of term extractors," in Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC'12), 2012, pp. 15–18.

[39] J. Pearson, Terms in Context. John Benjamins, Amsterdam, 1998.

[40] A. Rigouts Terryn, V. Hoste, and E. Lefever, "A gold standard for multilingual automatic term extraction from comparable corpora : term structure and translation equivalents," in Proceedings of the 11th International Conference on Language Resources and Evaluation (LREC 2018). European Language Resources Association (ELRA), 2018, pp. 1803–1808. [Online]. Available: http://www.lrec-conf.org/proceedings/lrec2018/index.html

[41] A. Nazarenko, H. Zargayouna, O. Hamon, and J. Van Puymbrouck, “Evaluation des outils terminologiques : enjeux, difficultés et propositions,” *Traitement Automatique des Langues*, vol. 50, no. 1 varia, pp. 257–281, 2009. [Online]. Available: https://hal.archives-ouvertes.fr/hal-00516698

[42] E. Loginova Clouet, A. Gojun, H. Blancafort, M. Guegan, T. Gornostay, and U. Heid, “Reference Lists for the Evaluation of Term Extraction Tools,” in *Terminology and Knowledge Engineering Conference (TKE)*, Madrid, Spain, Jun. 2012, pp. http://www.oeg–upm.net/tke2012/proceedings. [Online]. Available: https://hal.archives-ouvertes.fr/hal-00816566

[43] E. Cardillo, A. Folino, R. Trunfio, and R. Guarasci, “Towards the reuse of standardized thesauri into ontologies,” in Proceedings of the 5th International Conference on Ontology and Semantic Web Patterns - Volume 1302, ser. WOP’14, 2014, pp. 26–37.

[44] F. Giunchiglia, I. Zaihrayeu, and F. Farazi, “Converting classifications into owl ontologies.”

[45] S.-J. Kang and J.-H. Lee, “Semi-automatic practical ontology construction by using a thesaurus, computational dictionaries, and large corpora,” in *Proceedings of the Workshop on Human Language Technology and Knowledge Management - Volume 2001*, ser. HLTKM ’01. USA: Association for Computational Linguistics, 2001. [Online]. Available: https://doi.org/10.3115/1118220.1118226

[46] Cybesecurity Osservatorio - Thesaurus. https://www.cybersecurityosservatorio.it/it/Services/thesaurus.jspt\. Accessed: 2019-08-08.

[47] J. E. Rowley, J. E. Rowley, and R. J. Hartley, Organizing knowledge: an introduction to managing access to information / Jennifer Rowley and Richard Hartley , 4th ed. Ashgate Aldershot, England ; Burlington, VT, 2008.

[48] C. Barrière, “Semi-automatic corpus construction from informative texts,” in Text-Based Studies in honour of Ingrid Meyer, ser. Lexicography, Terminology and Translation, L. Bowkes, Ed. University of Ottawa Press, January 2006, ch. 5.

[49] G. Zagrebelsky, Il sistema costituzionale delle fonti del diritto, EGES, Ed. Turin: UTET, 1984.

[50] A. Condamines, “L’interprètation en sémantique de corpus : le cas de la construction de terminologies,” Revue française de linguistique appliquée, vol. Vol. XII, no. 2007/1, pp. 39–52, 2007.

[51] F. Dell’Orletta, G. Venturi, A. Cimino, and S. Montemagni, “T2K: a system for automatically extracting and organizing knowledge from texts,” in Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC’14), N. C. C. Chair), K. Choukri, T. Declerck, H. Loftsson, B. Maegaard, J. Mariani, A. Moreno, J. Odijk, and S. Piperidis, Eds. Reykjavik, Iceland: European Language Resources Association (ELRA), may 2014.

[52] IATE European Union Terminology. https://iate.europa.eu/home\. Accessed: 2019-08-08.

[53] C. Lanza, “Italian domain-specific thesaurus as a means of semantic control for cybersecurity terminology,” in The Twelfth International Conference on Advances in Semantic Processing (SEMAPRO 2018), U. o. A. S. G. P. L. U. o. H. A. F. Michael Spranger, Hochschule Mittweida, Ed., Athens, Greece, November 2018.

[54] M. van Assem, M. R. Menken, G. Schreiber, J. Wielemaker, and B. Wielinga, “A method for converting thesauri to rdf/owl,” in *The Semantic Web – ISWC 2004*, S. A. McIlraith, D. Plexousakis, and F. van Harmelen, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2004, pp. 17–31.

[55] M. Nowroozi, M. Mirzabeigi, and H. Sotudeh, “The comparison of thesaurus and ontology: Case of asist web-based thesaurus and designed ontology,” Library Hi Tech, vol. 36, 01 2018.

[56] J. L. D. Kless, L. Jansen and J. Wiebensohn, “A method for re-engineering a thesaurus into an ontology,” in Proceedings of International Conference on Formal Ontology in Information Systems (FOIS 2012), 2012, pp. 133–146.

[57] J. Qin and S. Paling, “Converting a controlled vocabulary into an ontology: the case of gem,” *Inf. Res.*, vol. 6, 2001.

[58] D. Adams, L. Jansen, and S. Milton, “A content-focused method for re-engineering thesauri into semantically adequate ontologies,” *Semantic Web*, 09 2015.

[59] I. Rösiger, J. Bettinger, J. Schäfer, M. Dorna, and U. Heid, “Acquisition of semantic relations between terms: how far can we get with standard nlp tools?” in *Proceedings of Coling 2016: 5th International Workshop on Computational Terminology (CompuTerm)*, Osaka, Japan, 2016.

[60] A. Condamines, “Taking genre into account when analyzing conceptual relation patterns,” *Corpora*, vol. 8, pp. 115–140, 2008. [Online]. Available: https://hal-univ-tlse2.archives-ouvertes.fr/hal-00606250

[61] A. Blumauer and T. Pellegrini, Semantic Web und semantische Technologien: Zentrale Begriffe und Unterscheidungen. Berlin, Heidelberg: Springer Berlin Heidelberg, 2006, pp. 9–25.

Multi-Factor Authentication for Public Displays using the Semantic Ambient Media Framework: Interconnecting Semantic Media and Devices

David Bouck-Standen, Josefine Kipke
Kingsbridge Research Center
Glasgow, United Kingdom
email: {dbs, jfk.student}@kingsbridge.eu

Abstract—**In this contribution, we present an approach to encounter various challenges of the growing amount of media available in our digital society as well as an individual's need to access structure and ordered information presents applying the Semantic Ambient Media Framework. The framework extends digital media, devices and applications, as well as services and with digital meta-data, interconnects them through semantic models, and makes them accessible in a Web-based API. In the event the API is accessed, the framework's services tailor the media, depending on context they are used in, their semantic interconnection with other media, and the specific application, device, and context they are accessed from. A possibility to access the information stored within the Semantic Ambient Media Framework is showing media on public displays. In order for public displays to deliver private, personalized, or sensitive contents or provide access to user-specific functionality, authentication mechanisms are required. On public displays, authentication is subject to a number of risks, especially, if displays offer multi-touch interfaces or grow even larger. In this contribution, we present a multi-factor authentication system for public displays using the Semantic Ambient Media Framework. In our approach, no actual credentials have to be typed in on the public display, as this makes use of the users' personal mobile devices and works with a one-time and location-based code. This contribution illustrates the concept and system architecture of the Semantic Ambient Media Framework in a working scenario together with multi-factor authentication for public displays that protects against threads, such as shoulder surfing, thermal attacks, or smudge attacks, which we also illustrate. We conclude with an outline of future work.**

Keywords-Semantic Media; Pervasive Displays; Multi-factor Authentication; Semantic Repository.

I. Introduction

Feature-rich multimedia systems allow users to produce high amounts of user-generated content. The contexts technology is used in also shift towards mobile and pervasive computing. Today, users use public displays to access public as well as personalized information [1] through interconnected multimedia applications [2]. In order to produce digital media, the users utilize their personal mobile devices equipped with cameras, such as smartphones, tablets, and other devices to interconnect with other systems through the Internet [3], [4].

As each multimedia system uses technologies with different interaction paradigms, they offer different capabilities for presentation, processing, and storing information in their own content repositories [5]. Focusing a vision of a convergence of personal or social information, at least the interconnection of multimedia or media storage systems, or at best a single multi-purpose multimedia repository system would be required [6]. With such a system, media would no longer be isolated for use in a single application or on a single device. Although today Cloud-based solutions already exist, however rather serve the purpose of harvesting data [7], than providing the service described above.

These challenges have been researched in various context-specific domains, as related work (cf. Section II) indicates. With the *Semantic Ambient Media Framework* (SAM.F) [2], this contribution presents a general context-independent approach. SAM.F is a framework that semantically interconnects (a) *digital media*, (b) *devices and applications*, and (c) *services*, which are enriched by digital meta-data in the form of semantic annotations. For both client application development, as well as the extension of framework functionality, SAM.F offers various interfaces for developers.

In SAM.F, digital media consists of, e.g., text, photos, audio, videos, animations, 3D objects, or 3D scenes. These are extended by digital properties, e.g., by classifying the media's content in the internal model of SAM.F. These digital properties include the Meta data extracted from the original file, such as Meta information on MIME type, encoding, or Exif data. For devices, in SAM.F, we model digital properties reflecting, e.g., the devices' capabilities', it's internal hardware, location, capacity, screen size, or screen resolution.

All digital properties are utilized by the services in SAM.F. Client applications running on users' devices access the services of SAM.F through a Web-based API. Each service serves a dedicated purpose, interconnecting devices and applications through the shared use of digital media and services. The service-based architecture of SAM.F is extendable, providing developers with dedicated interfaces and the means to develop new modularized services for SAM.F, as described in detail in this contribution.

In this article, we focus on a working scenario application using SAM.F to provide access to information stored in the framework.

Today, there is a growing number of large multi-touch displays, which are deployed in public spaces, such as public squares, airports, train stations, or in streets or, e.g., museums. These so-called Public displays consist of large multi-touch displays connected to a content providing system, e.g., via the Internet. They offer an entry point for a user to access digital data, such as stored within SAM.F. With the increasing demand for public displays to offer access to personalized or context-specific content or functionalities [8], they offer a supplementary or specialized digital interface device to the users' smartphones or mobile devices that most Internet users possess. Hence, accessing protected data and contexts or user-tailored functionality on public displays presents the need of a secure method for user authentication.

Authentication in general requires a user to enter credentials or use other means for personal identification, only known to or in possession of the user himself. This could be, e.g., classically a username-password combination. More sophisticated methods rely on the uniqueness of a thumbprint, the iris, or other biometrical information unique to the user [9].

The increasing use and functionality of public displays require a solution that protects users and systems against threats. Known threats can be for example *shoulder surfing attacks* (a), where the user is observed while authenticating [10], *thermal attacks* (b), where heat traces resulting from the user's interactions are made visible revealing the sensitive authentication data [11][12], or *smudge attacks* (c) that exploit the residues from finger prints on touch-screens [13]. Research on these techniques indicate that shoulder surfing occurs in daily contexts [10]. All three attack methods have in common that displaying a digital keyboard or using a software keyboard is vulnerable to them. For this reason and to prevent possible attacks exploiting the users' interactions with the systems, systems for biometrical authentication or gaze-touch have been proposed [14], as we explore in related work.

Using additional hardware for public displays, such as bio-scanners or cameras, comes with costs and the need to retrofit most public displays currently deployed. A solution with minimal hardware requirements is more likely to be widely accepted. Thus, one of the challenges of this work is to find a solution that does not require hardware upgrades of public displays.

Modern smartphones are personal devices equipped with different sensors and mostly at least one camera. The smartphone is still on the rise in 2020 due to its multi-functionality and connectivity, as almost 8 out of 10 Internet users in the EU surfed via a mobile or smartphone. The trend toward mobile technology and mobile Internet usage can still be observed globally [15].

In this work, we present a technical solution we developed as a prototype at the Kingsbridge Research Center (KRC), which addresses these challenges with a minimal technical solution. This makes use of a *multi-factor authentication* (MFA) [1]: The first factor is the *ownership* (i) of a personal mobile device, such as a smartphone. The second factor is *knowledge* (ii) of personal credentials, such as the combination of username and password. Using GPS data, we also use the user's and display's *location* (iii) as third factor.

The concept of this work makes use of the interconnection of the devices through the Internet, using the Semantic Ambient Media Framework (SAM.F) [2] as an authoritative facilitator between smartphones and public displays.

In this contribution, we regard related work in Section II, and present a practical scenario in Section III. In Section IV, we outline the system concept and architecture and describe our prototype implementation. In the final section, we summarize our work and illustrate future work.

II. Related Work

Semantic media comprises the integration of data, information and knowledge. This relates to the Semantic Web [16] and aims at allowing computer systems as well as humans to make sense of data found on the Web. This research field is of core interest since it yields naturally structured data about the world in a well-defined, reusable, and contextualized manner.

The field of metadata-driven digital media repositories is related to this work [17] as well. Apart from the goals of delivering improved search results with the help of Meta information or even a semantic schema, SAM.F distinguishes itself from a pure repository by containing and using multiple repositories as internal components, as illustrated below. As Sikos [18] observes, semantic annotations feature unstructured, semi-structured, or structured media correlations. Sikos outlines the lack of structured annotation software, in particular with regard to generating semantic annotations for video clips automatically. SAM.F offers means for both structured and semi-structured semantic annotations. Through an interface, the functionality of SAM.F can be extended to, e.g., automatically annotate media as outlined below, but is not limited to video clips. By these means, SAM.F delivers even more sophisticated features.

In general, SAM.F facilitates collecting, consuming and structuring information through device-independent interaction with semantically annotated media, whereas the linked data research targets sharing and connecting data, information and knowledge on the Web [19]. The concept originally developed by the author [20] was already used in different contexts, e.g., the automatic reconstruction of 3D objects from photo and video footage [2].

Blumenstein et al. [21] outline a technical concept in museum context, that relies on a server-based architecture to provide museum content in a multi-device ecology. SAM.F could be used in similar contexts, as the scenario outlines in Section III, but is not limited to the use in museums.

Ambient systems can provide a platform for displaying of and interaction with media [22]. In this context, the delivery of content on different devices is an important issue in SAM.F, e.g., with respect to the devices' capabilities or their context of use, and SAM.F addresses this challenge by provisioning digital media depending on applications and devices specifications or capabilities. SAM.F also addresses the issue of limited bandwidth of mobile devices.

The Social Web is related to this work, as it makes it easy for people to publish media online. Yadav et al. [23] propose a framework interconnecting Social Web and Semantic Web by semantically annotating and structuring information people share. SAM.F could be used in this way, but focuses on semantically enriched or described instances of media, devices and services.

Semantic frameworks are used in various contexts, such as multimodal representation learning, as proposed by Wang et al. [24]. In their approach, Wang et al. use a deep neural framework to capture the high-level semantic correlations across modalities, which distinguishes this approach from SAM.F.

This work also takes place in the research field of public displays. Related work indicates a general increase in the deployment of public displays [25].

Today, public displays are widely connected in client-server-applications for content serving purposes, and they are connected through the Internet [25]. For example, Memarovic et al. focus on interconnecting displays, e.g., with social media [26]. Our work ties onto related work through its modular client-server-based architecture. As this work depends on Web-based and modular technology, it can be integrated into other existing projects.

The field of multi-factor authentication (MFA) is an important research field for this work. As Ometov et al. [9] recently surveyed state of the art methods for MFA, illustrate technical requirements, and identify commercial, governmental, and forensic applications as three market-related groups of applications for MFA. In context with public displays, this work can potentially be deployed in all three fields, but according to Ometov et al. the use in commercial applications is most likely.

One of the main challenges of MFA is the absence of a correlation between the user identity and the identities of smart sensors and systems or devices, as Ometov et al. observe [9]. Ometov et al. propose a user-friendly process to establish a trust relationship to gain access rights, whereas Mannan et al. [27] propose a concept to use a personal device to strengthen password authentication from untrusted computers. We combine aspects from these theoretical approaches to our technically limited setting, as outlined above, and present a feasible solution for the MFA for public displays using SAM.F.

With the system called Tacita, Shaw et al. [25] demonstrate a system to personalize public display experience by utilizing proximity detection for user's mobile devices, e.g., with iBeacon technology.

Tacita ubiquitously personalizes public displays' content, whereas GTmoPass proposed by Khamis et al. relies on gaze-touch detection through the smartphone and the identification of the display via a Bluetooth Low Energy (BLE) beacon [14].

These approaches are distinguished from the approach presented in this contribution, as the system we develop directly authenticates users and requires a direct user interaction on the public display. It therefore supports direct use, which features the use of public displays in both unauthenticated, as well as authenticated contexts, especially, if personalized or personal information is displayed or the public display is used to access sensitive functionality. In addition, the solution proposed in this article does not require any supplementary hardware, such as BLE beacons or cameras.

The following section illustrates the system's concept and architecture.

III. A PRACTICAL SCENARIO

Together with our project partner, the *Audiovisual Archive of German-language Literature e.V.* in the Hanseatic City of Bremen, Germany, we develop a scenario to add a digital information meta-layer to a physical exhibition planned in a cultural centre in Bremen.

The scenario illustrates both the use of the Semantic Ambient Media Framework (SAM.F) together with the multi-factor authentication for public displays (MFA4PD), which are illustrated in detail in this contribution. This exemplary use-case focuses on the practical implementation. The visualization of information retrieved from SAM.F described in this scenario is not part of our current work. We are planning to contribute this in the future, as it serves for illustration purposes in this article, only.

Emily Walden is 16 years old and is visiting her grandfather Erik Braun in the Hanseatic city of Bremen this weekend. Since she has grown up with technology, Emily is very tech-savvy, while her grandfather owns a smartphone, but only for making phone calls and prefers not to use it at all. He only has little technical affinity.

Erik Braun is interested in literature, while Emily is less enthusiastic about them. Erik Braun became aware of a newspaper article at the cultural centre of the Audiovisual Archive of German-language Literature e.V. in Bremen, which he would like to visit with his granddaughter. That's why he plans their joint excursion through the city so that they will pass by the cultural centre.

After breakfast they start and in the early afternoon arrive at the cultural centre. Through the high glass facade, some exhibits can already be seen, as well as displays on the walls displaying "Literature and I" in large fonts. More as a favour to her grandfather, Emily agrees to a visit and both enter the exhibition, which at this time is especially dedicated to the writers and Nobel-prize winner Günter Grass.

Inside the exhibition, Emily and her grandfather Erik notice a standing desk. She approaches this and learns that this was Grass's standing desk, on which he wrote "The Tin Drum" during his time in Paris. A tablet installed on the standing desk shows a short video, after which the display changes and shows a graph. The graph shows images of the backyard with the entrance to the boiler room, in which Grass wrote the novel. The images are connected by a line to a node with the name "Paris". The node is connected to the node "Günter Grass". Erik and Emily look at the visualization and Emily notices another line connecting the node to a film adaptation of Volker Schlöndorff's Tin Drum. She didn't know there was a film, and she would also watch it with her grandfather instead of reading the book, she joked.

Together they continue through the exhibition and head towards one of the large screens, on which the lettering "Literature and I" is shown. As they approached the display, the wording changes to: "What moves you spontaneously?". Below, they both see various icons displayed below.

Erik, who is a little more reserved than Emily, leaves her to operate the device and watches, as she interacts with the multi-touch display. Spontaneously, she chooses a pictogram with a cooking pot with a trowel.

Immediately the presentation changes and they again see a graph, they recognize from the standing desk. This time, however, Emily notes that the information displayed has to do with cooking. Grandfather Erik, who read Grass but had not yet noticed that the literate had dealt so comprehensively and repeatedly with the topic of hunger and cooking, is also astonished.

On the display, Emily notices a textual hint that tells her, that she can continue exploring the information more deeply from home. She follows the instructions displayed and connects to the Wi-Fi "KulturzentrumBremen" on her Android device. A notification shows up and she opens the mobile's browser, where she completes a short sign-up form in just a few seconds and is signed-in automatically.

Afterwards, she sees a code with five symbols displayed on her smartphone. On the display, Emily selects the symbols shown on her smartphone from a grid of nine symbols shown on the display. The display now indicates that Emily has to check her smartphone again. On the smartphone, a prompt is displayed and Emily confirms, that she wants to log in on a display named "Kulturzentrum Bremen - Bildschirm 4". She quickly cross-checks the name of the display she and her grandfather are standing in front of and confirms the login on her smartphone.

The display immediately changes from the login mask back to the graph and greets Emily with her name. Emily now sees that she can bookmark nodes by selecting them. During their exploration.

Emily, who is interested in politics, touches a node with the description "Victory over the hunger of the world". The depiction changes and shows a video in which the writer Günter Grass made a political statement. Emily immediately notices that this was part of a lengthy interview, knowing the digital video controls from social media, and touches the bookmark icon on the screen.

Together with her grandfather, Emily continues to explore the physical and the digital exhibition. Whenever she finds a medium that is interesting, she stores a digital bookmark.

A few days later, Emily already is back with her parents', the ticket of the cultural center in Bremen falls into her hands while she looks through her receipts in her wallet. She discovers a web link on the ticket and opens it in her web browser. After signing in with her credentials she set up during her visit to the cultural center in Bremen, she again sees the graph view and notices an interconnection between nodes that are familiar. In addition to the bookmarks she recognizes, she sees other nodes connected displaying media associated with the topics at hand.

In the following section, we focus on the system concept and architecture.

IV. System Concept and Architecture

In this section, we illustrate the system concept and architecture of SAM.F. Subsequently, based on the framework, we outline the concept and implementation of multi-factor authentication for public displays.

A. *The Semantic Ambient Media Framework*

SAM.F is a smart media environment, which provides a device-independent access to and interaction with media through devices and applications.

The system's architecture of SAM.F is based on a system concept following these three considerations:

1. Web-based access provides platform-independent use of the services and access to media inside SAM.F and its repositories from the users' devices and applications.
2. a service-based modular architecture features extendibility, which provides developers with a framework to develop their own applications, which can be based on or reference to existing services within SAM.F.
3. the concept of Semantic Media regards media independently of their encoding or modality and automatically transcodes or converts media, where necessary and possible, to meet contexts, applications, and devices specifications or criteria.

In the following sections, we focus on the concept of Semantic Media fundamental to SAM.F. We illustrate the system's architecture and the service concept of SAM.F. Following, the application and device-specific media provisioning is outlined. In addition, technical details on the current implementation of SAM.F are given.

1) Semantic Media

In SAM.F, apart from services delivering media, media themselves are central. Semantic Media consist of plain media, such as text, audio, video, pictures, and 3D media, which are enriched by a dynamic set of semantic annotations. Together, plain media and semantic annotations form *Semantic Media* in SAM.F.

The dynamic set of semantic annotations stored in SAM.F for each media element consist of:

- the original Meta-data of the plain media file. For example, for photos taken with digital cameras, metadata usually contains information on the picture's location, and camera data such as camera make and model, or camera settings, such as camera capture settings. This data might be useful for SAM.F services and adding it to the set of annotations improves accessibility and performance when further processing media.
- data received from automated algorithms. Pictures for example are submitted to a Computer Vision algorithm by SAM.F automatically and in a background process in order to determine semantic annotations describing the media's content.
- data received from client applications. As the main user interaction with media through SAM.F is carried out through client applications, in which the context of use is known, this information is stored in additional semantic annotations. This information is collected automatically in a background process through the use of the SAM.F API Web Services, which implicitly reveal the context of use.
- data received from manual user interactions, such as manual annotations or correcting automatic annotations.

It should be noted that the semantic annotations of Semantic Media may not be complete or available for each media element at all times. This is, e.g., due to the context the media is created in, a foreign source the media is accessed from, or incomplete data entered by the user [28]. This observation presents a challenge we discuss at the end of Section IV in terms of the prototype implementation.

The set of annotations described above is not final and can be extended in context of client applications, devices or services.

In SAM.F, the complete set of semantic annotations are abstracted into the Data Model (see Figure 1) in order to be (i) accessible for all services running inside the framework and (ii) accessible independently of the underlying media repository in the Datastore layer (see Figure 1).

Not all annotations are made available for every client application or device through the API Web Services (see Figure 1), as the API Client Model only contains those properties that are required in the corresponding context. This way, overhead in the access of media through client applications is assumed to be significantly reduced. The effects on performance or bandwidth have however not been measured as part of this work. However, we discuss possible issues arising from this approach at the end of Section IV.

It is one of the hypotheses of this work that the quality of semantic annotations as well as the interconnection of media will be a key issue for realizing appealing scenarios using SAM.F, as, e.g., described in the final section of this article. An approach to achieve this is to gather additional sematic annotations through automated algorithms. As illustrated in Figure 2 and mentioned above, pictures, for example, are submitted to a Computer Vision algorithm. In the current implementation, SAM.F interfaces with Microsoft Cognitive Services. Thus, in the background, SAM.F computes additional semantic annotations, which are then stored in the internal Datastore (see Figures 1 and 2).

2) System Architecture

The architecture of SAM.F consists of a layer-based system concept, as illustrated in Figure 1. Client applications and devices utilized by users connect to the SAM.F *API Web Services* through the *API Security Layer* via the Internet in order to access media stored in SAM.F or interact with services in the *Service Layer* (see Figure 1).

When interacting with SAM.F, client applications as well as devices exchange information with the framework (see Figure 2) using a defined data model. Thus, for any context, the *API Client Model* can be extended to exactly match the needs of the application, device, or context, if necessary. API Web Services offer access to dedicated services provided by SAM.F, as the scenario described above outlines. Internally, SAM.F works with a dedicated *Data Model*, as illustrated in Figure 1. Any data is mapped from the *Datastore*, which includes external (semantic) databases as well as binary data stores, to the internal Data Model, which applies a homogenous model to potentially heterogenic data. Thus, SAM.F features the integration of different repositories and provides a combined access to Semantic Media. For simplification purposes, and in order to reduce the learning curve when implementing client applications accessing SAM.F, the internal Data Model is only used in the *Provider Layer*, which contains, e.g., authentication or data providers to be accessed by the upper Service Layer, and in the *Service Layer*, as shown in Figure 1. Any Semantic Media, together with semantic annotations, provided by a service to a client is mapped to the specific API Client Data Model, as outlined above, and being served through the API Web Services and the API Security Layer to the client application (see Figure 2).

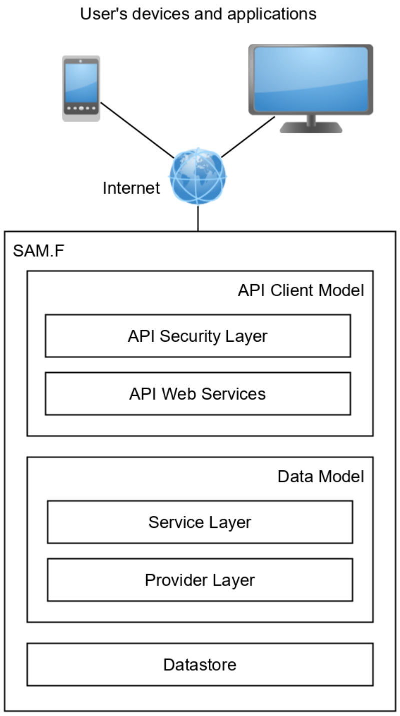


Figure 1: Layered architecture of SAM.F.

With the Data Model only used internally, SAM.F accommodates different models used when storing media in digital repositories. A museum database for example differs significantly from, e.g., the DbPedia's semantic database. To be able to use heterogenous sources simultaneously, different data models are homogenized though the Data Model in SAM.F: by applying the data mapping techniques, the framework uses its own model internally, into which all other models are mapped. Applying data mapping in SAM.F produces constant overhead. However, services and applications, as well as their developers, benefit from only working with data models that are specific to the requirements of the services' or applications' context. This also reduces overhead when loading large sets of Semantic Media.

The range of functions of SAM.F is defined by the functionality provided by services residing in the Service Layer, as illustrated in Figure 1. In the scenario outlined below the developers extend SAM.F by implementing a custom service in order to realize the desired functionality. Thus, in the next section, the SAM.F services are regarded.

3) SAM.F Services

Following the implementation principles of SAM.F, a service features a dedicated set of functions in order to provide a certain functionality, e.g., for a use-case or scenario, as outlined above.

Utilizing the Data Model, through the Provider Layer, any service might access Semantic Media from the repositories included in SAM.F's Datastore layer. As a result, services may interchange information in a well-defined context.

SAM.F comes with a set of services that are useful to the developer in a Web-based environment and for developing applications in context of mobile use and the use of Semantic Media, explained in more detail below. In this article, we focus on the basic features the SAM.F services consist of:

- an authentication service to identify and authenticate sessions of applications, devices and users.
- a general media service that allows to retrieve or modify Semantic Media elements for a given keyword in a given general context. Media is retrieved both from the internal datastore, as well as external semantic databases housed in the Datastore layer (see Figure 1) and made available through SAM.F.
- the *Application and Device-specific Media Provisioning* (ADMP) service, which transcodes media based on different settings on client retrieval, as outlined below.

In the scenario outlined above, the developers extend the Service Layer of SAM.F (see Figure 1) and add their service to authenticate users on public displays. This service utilizes the modularized architecture of SAM.F and interfaces with the adjacent upper and lower layers. It also makes use of the default user authentication service. Service execution may either be triggered (i) on demand per request, or (ii) internally. This allows services of SAM.F to automatically run in the background without the necessity of user interactions.

4) Application and Device-specific Media Provisioning

Semantic Media in SAM.F can contain various types of plain media. However, their use is determined by the client applications. The devices running these applications are usually limited in their capabilities.

To address these challenges, SAM.F offers an *Application and Device-specific Media Provisioning* (ADMP) for any Semantic Media element retrieved through the API Web Services layer (see Figures 1 and 2).

In general, ADMP transcodes or converts Semantic Media due to specifications given. Trivial examples are the conversion of large photos into thumbnails, including cutting and cropping, if necessary.

ADMP is designed to work in two ways:

- on a per-request basis, in which the application submits the desired parameters (e.g., format, encoding, size, resolution) with every request, or
- on an application or device capability basis. As devices and applications are also represented in the Data Model (see Figure 1) of SAM.F, their capabilities are known to SAM.F. Thus, using per-request parameters can be omitted, if application or device capabilities can be generally set or are valid for multiple requests.

Especially in context of the Web-use of SAM.F and the heterogeneity of devices potentially accessing SAM.F, ADMP's usefulness can be illustrated through these examples, in which the correct parameter settings are

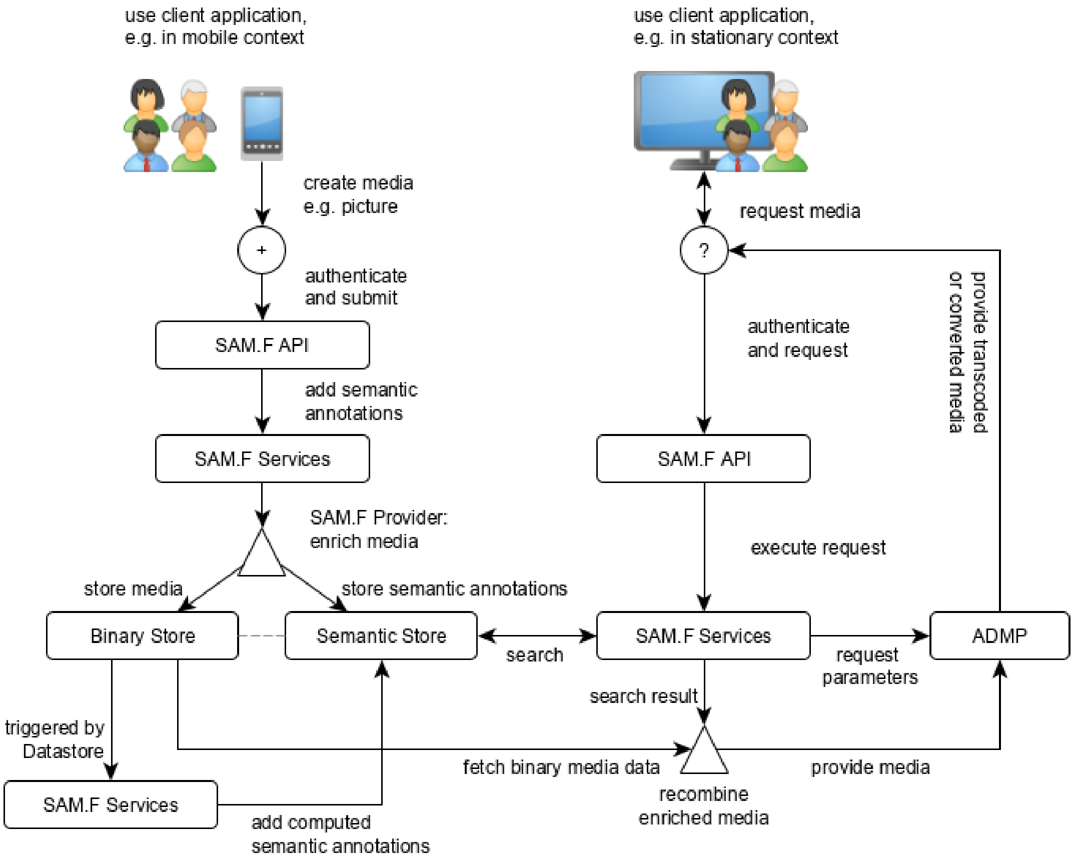


Figure 2: Media creation, enrichment through semantic annotations and retrieval. The Datastore consists of both Binary Store and Semantic Store.

presupposed: A video can be retrieved in different encodings or in matching screen size for the device's resolution. For example, ADMP can provide just the audio track of the video or just the textual transcript. The transcript can also be used to subtitle the video. More challenging 3D objects, which may not be viewed on any device, can be retrieved as a video of the 3D object rotating around the y-axis, or just as a picture in the form of a screenshot of the 3D object.

Reviewing key event-based multimedia applications, Tzelepis et al. [29] observe an enormous potential for exploiting new information sources by, e.g., semantically encoding relationships of different informational modalities, such as visual-audio-text. SAM.F provides these means by transcoding and converting Semantic Media in the background by automated processes.

As a side-effect, using ADMP reduces the use of bandwidth, which is of special interest in mobile contexts.

As these examples indicate, this way of provisioning media though SAM.F provides the means for a vast amount of use-cases. However, the author admits that not all possibilities have been implemented. The ADMP module, which also extends the Service Layer (see Figure 2), can be expanded, as it features an interface with an extendable list of parameters.

5) *Prototype Realization of SAM.F*

A first prototype implementation of SAM.F has been realized at the *Kingsbridge Research Center* (KRC). On the basis of a Windows Server system and its Internet Information Services (IIS) Web server, SAM.F is implemented in C# and runs as IIS Web application. Web services are provided using the Active Server Method File (ASMX) technology.

SAM.F uses an internal database to store all semantic data and semantic correlations. Currently, we follow two main approaches for storage of semantic data.

a) *Working prototype approach*

In our stable prototype approach, semantic annotations used in SAM.F are represented as RDF triples. For performance reasons analyzed under laboratory conditions in experimental settings, SAM.F's internally used RDF data is stored in a NoSQL database, although quantitative performance measurements are future work. SAM.F is compatible to semantic media repositories, e.g., using SPARQL to execute queries. Additionally, other required annotations for external media are stored in the internal datastore of SAM.F. In these terms, external media are media that are made available through SAM.F, but are stored in semantic datastores that are not managed by, but connected to SAM.F.

The approach of combining the automated enhancement of semantic annotations for media and delivering media in a device- or context-specific modality or encoding presents a technical novelty and distinguishes SAM.F from other media frameworks or repositories.

The current prototype has been validated under laboratory conditions. Computations are implemented to be carried out in a complexity of *O(n)*.

Together with our project partner, as outlined below in more detail, we will integrate SAM.F for use in context of research and cultural projects. This will provide the opportunity to evaluate the system under real conditions with regard to functionality and performance.

b) *Experimental approach*

In order to be able to serve client applications connected to the SAM.F API, as depicted in Figure 1, before exposing any information through the API it is mapped into the API Client Model. This well-defined model exposes only those attributes necessary for the given service or context. Thus, a developer can interface an application with SAM.F without prior knowledge of the internal semantic model, or any model applied to a semantic database that might be connected to SAM.F, as outlined in terms of 'external media' in the section above.

Although this approach features a preferable learning curve for developers, it limits the API's capabilities to the information modeled within the API Client Model. It also requires the semantic model to be preset within the implementation of SAM.F.

In our new, highly experimental approach, we are currently implementing an interface that allows the specification of SAM.F's internal model through JSON. Thus, the developer of a client application defines his required model in one or more JSON files, as outlined in Figure 3. From this definition file, SAM.F derives the semantic correlations into an RDF triples, as it sets up the internal model to the developers' specifications.

With the new approach developers are enabled to define their own models, correlations and contexts. However, this is still under research and developers are required to have more understanding of the Semantic Web and semantic queries. We chose JSON notation as shown in Figure 3 because the syntax

```
{
  "properties": {
    "title": {                            <root, properties, title>
      "type": "string",                   <title, type, string>
      "writeable": true                   <title, writeable, true>
    },                                    <root, properties, oncatalogue>
    "oncatalogue": "boolean"              <oncatalogue, type, boolean>
  }
}
```

Figure 3: Basic example for the experimental JSON to RDF conversion.

is more comparable to class modelling, than specifying RDF schema.

We are planning on making this experimental feature available in the future and carry out specific user research with the user group of developers in order to evaluate this highly experimental approach.

6) *Summary*

SAM.F provides Web-based access for devices and applications and features a service-based architecture, which allows for interaction with media, such as, e.g., text, pictures, audio, video, 3D objects, or 3D scenes.

B. *Multi-Factor Authentication for Public Displays*

For this approach and under consideration of the technical limitations outlined above, the following is the starting point for this work:

- users are in possession of a smartphone or equivalent device connected to the Internet. They have already registered an account with credentials known to SAM.F beforehand, as this is a preliminary requirement of this work.
- public displays are connected to the Internet and run on Web-based technology, e.g., showing Web-based contents in a browser-based system.
- the user sojourns in the vicinity of a public display and intentionally starts a private context.

Figure 4 illustrates the system's architecture and the starting point. In a single location, one or more users and one or more displays can be present. A user interacts with a single display and is in possession of a personal mobile device, as depicted in Figure 4. All public displays and user's devices are connected to SAM.F through the internet. However, a direct connection between a smartphone and any public display does not exist.

Inside SAM.F, the *multi-factor authentication for public displays* (MFA4PD) module is hosted. Public displays and user's devices connect to the MFA4PD module through the Internet. In addition, public displays connect to external content providers, which are not illustrated in Figure 4, for simplification purposes.

The system's architecture benefits from the technical limitations outlined above. As there is no direct connection necessary between the personal mobile device of a user and the public display used for authentication, there is no need for the display provider to open up his network for foreign devices. Thus, a multi-device ecology within the network of the display provider is not required, resulting in less administrative effort. Whenever an Internet connection has to be provided, e.g., in the event of poor LTE or cellular reception, provisioning a public hotspot is sufficient.

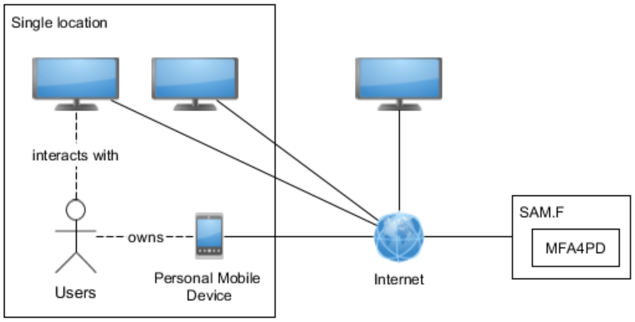


Figure 4. Illustration of the system's architecture and network.

From Figure 4, it can also be observed that no additional hardware, such as, e.g., BLE beacons, is required.

The system concept relies on the interconnection of mobile devices and public displays through SAM.F, which is outlined in the following section.

1) *Interfacing MFA with SAM.F Services*

The architecture of SAM.F, although described here only with reference to the MFA4PD module, consists of a layer-based system concept, as illustrated in Figure 5. A client application, such as the display or mobile application of this work described in more detail below, connect to the SAM.F *API Web Services* through the *API Security Layer*. Data is exchanged between applications and services, which reside in the *Service Logic* layer, in the form provided by the specification of the *API Client Data Model*. Internally, SAM.F works with a dedicated *Data Model*, as illustrated in Figure 5. Any data is mapped from the *Datastore*, which includes external (semantic) databases, as well as binary data stores, to the internal Data Model, which applies a homogenous model to potentially heterogenic data. For simplification purposes, and in order to reduce the learning curve when implementing applications accessing SAM.F, the internal Data Model is only used in the *Provider Layer*, which contains, e.g., authentication or data providers to be accessed by the upper Service Layer, and in the Service Layer, as shown in Figure 5. Any data provided by a service to a client is mapped to the specific API Client Data Model before being served through the API Web Services and the API Security Layer.

Applying data mapping in SAM.F produces constant overhead, but services and applications, as well as their developers, benefit from only working with data models that are specific to the requirements of the services' or applications' context, reducing overhead when loading large sets of data. Data in this respect describes media, devices and services.

In context of this work, SAM.F serves as authentication provider, which validates user credentials via its standard user service. This work extends the Service Layer of SAM.F by adding the MFA4PD module, which implements the authentication process described in the following section.

2) *Authentication Process*

To start a private session on one of the public displays, the user opens up the *mobile application* of MFA4PD on his personal smartphone. The user then enters his credentials previously registered with SAM.F, which the Web application submits to MFA4PD, as illustrated in Figure 5.

At this point, the process of authentication might be enhanced by further means of MFA, such as gathering biometrical data from fingerprint sensors, facial recognition, or voice sensors. These extensions however might require at least a hybrid application deployment for mobile devices, in order to access the appropriate sensor data. For this reason, in this initial approach, we focused on the Web application combining MFA with ownership and knowledge factors.

The users are identified and authenticated by MFA4PD through their credentials. During the entire process, MFA4PD continues to check the actual location of the user. The location is determined from the GPS data, which is accessible through the smartphone's Web browser API. If any location mismatch occurs, the process to establish a secure session or the session itself will be terminated immediately for security reasons. This feature might prove useful, whenever a user leaves the location of a public display. However, we did not evaluate this feature's aspect nor the accuracy required from GPS data in order to work in an everyday scenario, yet.

After the user's location and credentials are validated, SAM.F generates a code consisting of five symbols, which is shown to the user on his mobile device, as illustrated in Figure 5. The code is valid for a short period of time and the specific user only.

In order to authenticate him- or herself on a public display, the user has to enter the one-time code shown on his smartphone (Figure 6, please see next page). The user opens up the login dialogue of the *display application* on the public display, and a grid of symbols is displayed. Within this grid, the user now selects the symbols shown on his or her smartphone. The display application communicates the code back to the MFA4PD module, as shown in Figure 5. This serves two purposes:

a. to identify the display, the user selected from the number of public displays available, and
b. to identify the user, who chose a public display.

However, the session is not yet usable. The last step to enable the session on the public display requires the user to again interact with his smartphone in using the *confirm mechanism*. As illustrated in Figure 5, the MFA4PD module sends an authentication request to the mobile application. The dialogue shown indicates a login event took place, together with the name and location shown on the public display. Without the user confirming the login on the mobile device, the session will not be unlocked. The confirmation screen on the smartphone is illustrated in Figure 7.

The users can now put their smartphones away and start using the public display, until they log out.

An additional timeout mechanism prevents misuse of the session on a public display.

The users can also close the session at any time using their mobile devices, e.g., in case they forgot to select the logout function on the public display. In addition, SAM.F monitors the users' location throughout the entire process and session in order to prohibit misuse of login or automatically logout a session after a user clearly left the screen's location.

Now that the system's architecture and concept have been illustrated, in the following section, this approach is viewed with regard to security.

3) *Prototype Realization*

The prototype consists of three components: (1) the MFA4PD module extending the services of SAM.F, (2) the mobile application and (3) the display application.

SAM.F is developed as Internet Information Services (IIS) application for Windows Servers, as outlined above. The MFA4PD module is implemented as ASMX Web service and a backend-only application, which adds an ASMX Web service to the framework and interfaces with SAM.F.

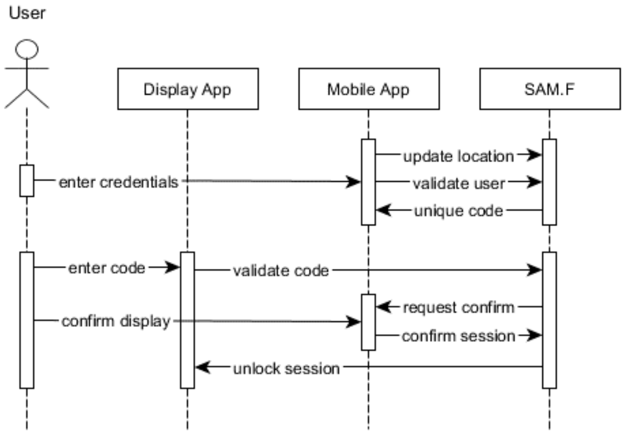


Figure 5. Sequence diagram showing the authentication of one session by a user.

In our initial approach [1] we developed the mobile application as an IIS Web application, which interfaces with the MFA4PD service in order to realize the multi-factor authentication for public displays using SAM.F. It consists from an ASPX form using JavaScript and AJAX to interact with the frameworks service, whereas the graphical user interface can be customized using HTML and CSS. Figure 6 and Figure 7 show the Web application on an Android smartphone.

The display application consists of a graphical component including the necessary HTML, CSS and JavaScript code. It interfaces with the MFA4PD module via JavaScript through AJAX. The display application also comes with a lightweight backend for session management, that also interfaces with the MFA4PD module. This is currently implemented in ASP.NET.

In order to incorporate the display application into an existing application, we provide code snippets that can be integrated into any application. If a target project does not run ASP.NET, the required server-side code can be translated for other frameworks.

In our current approach, we develop a mobile application using Xamarin in addition to the Web application described above, currently focusing on Android devices. Thus, authenticating the mobile application against the SAM.F API is now implemented using token-based authentication. This way, the device ownership factor is strengthened and the session is bound to the user's physical device. In addition, the mobile application can receive push notifications from SAM.F using SignalR, e.g., in order to signal a logout event from a public display directly.

The new mobile application now offers access to the smartphone's sensors. This means although the GPS accuracy and indoor location issues remain analogues to the Web application, the mobile application is now capable of detecting the user's steps. It is our current hypothesis that this can enhance detecting whether a user has left the location of a public display. This is important, e.g., in case a user forgot to logout of his authenticated session on the public display.

The main interface of the Xamarin application with the MFA4PD service is developed as separate Portable Code Library (PCL). Thus, other applications might incorporate public display authentication mechanisms into their own program logic by using the PCL.

We plan to make the prototype available for non-commercial use later this year.

4) *Security*

As outlined above, related work identifies possible means to attack public display authentication, such as shoulder surfing attacks (a), thermal attacks (b), or smudge attacks (c). In the following, we focus on these client-side attacks.

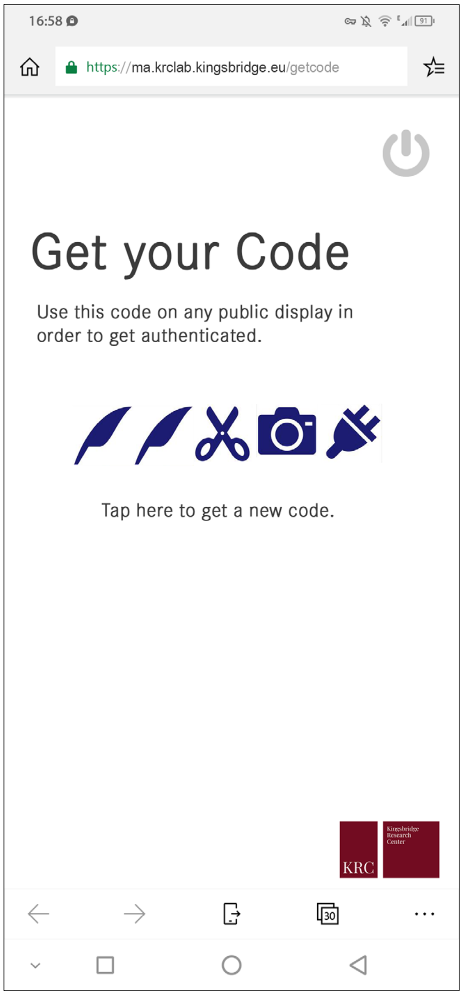
a) Client-side security

Combining ownership and knowledge factors together with the confirm mechanism, only initially entering the user credentials on the mobile application is vulnerable to shoulder surfing attacks. However, once the trust relationship is established between SAM.F and the users' smartphone for the current location, another mobile application login is prohibited for the duration of that trust membership.

In our initial prototype, we use session cookies and device cookies to temporarily store trust relationship data. Device cookies may be subject to manipulation, but session cookies stored on the server would require the attacker to have server access.

We also plan to use the user's location to limit the list of displays he might login-to to the number of displays that actually are in the user's vicinity.

With regard to the one-time code displayed on the mobile application, as well as the user's input of this code on the display application, they are not vulnerable to shoulder surfing attacks. Again, the confirmation mechanism protects the theft of the session. If any irregularity occurs, the user just declines unlocking the session and generates a new code.



16:58

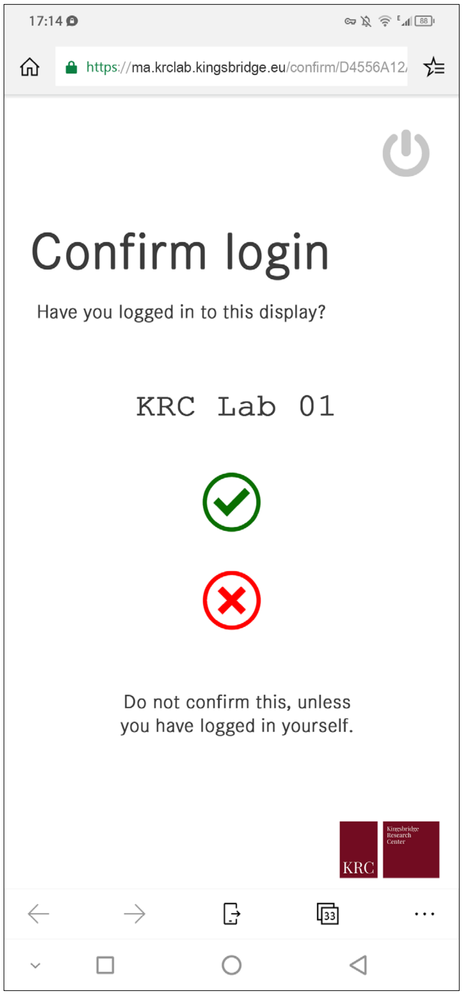
https://ma.krclab.kingsbridge.eu/getcode

Get your Code

Use this code on any public display in order to get authenticated.

Tap here to get a new code.

Figure 6. Screenshot taken from an Android 9 smartphone running MFA4PD Web application inside Edge browser. The symbol code displayed authenticates the user on the public display.



17:14

https://ma.krclab.kingsbridge.eu/confirm/D4556A12/

Confirm login

Have you logged in to this display?

KRC Lab 01

✓

✗

Do not confirm this, unless you have logged in yourself.

Figure 7. Screenshot taken from an Android 9 smartphone running MFA4PD Web application inside Edge browser. The login confirmation screen shows up after entering the code on the public display.

If a user accidentally confirms a session on his mobile application for a code that was used on another display or by somebody else, the simplest way is to just close the session from the user's smartphone immediately. However, this case is unlikely to occur due to the one-time code concept and the narrow time frame, in which a code can be used.

Both thermal attacks and smudge attacks cannot be used on public displays in this approach. Once the one-time code has been used, it is invalidated. The statistical possibility of guessing a one-time code can be decreased by a higher number of symbols used in the one-code, a larger symbol inventory, or a larger grid.

During the entire login process and, in concept, during the entire authenticated session, MFA4PD checks the user's GPS location. If any location mismatch occurs, the process to establish a secure session is aborted. Also, in theory, any ongoing session will be terminated immediately for security reasons. This feature might prove useful in case a user leaves the location of a public display without logging out. However, the evaluation of this feature or the accuracy required from GPS data in order for this concept to work in an everyday scenario will be carried out in future work.

For public displays, we limit the number of entered codes per system, location, and time to additionally prevent brute-force attacks. Again, the time-limited one-time-code limits also the possibility of brute-force attacks.

In summary, with regard to security, the system's concept offers protection on the client side against contemporary threats, such as shoulder surfing attacks, thermal attacks, or smudge attacks. Apart from client-side attacks, backend systems as well as the module's communication might also be vulnerable to attacks, which we outline in the next section.

b) Backend security

The users are identified and authenticated by MFA4PD through their credentials, which they enter inside the mobile app. These currently consist of a combination of a username and password. SAM.F validates the credentials. On credentials mismatch, a login is not possible. Additionally, SAM.F checks the mobile device identification number, which is exposed by both Android and iOS system's API. On mismatch to previous devices used, an additional confirmation is required, which we plan to implement via SMS.

The communication between mobile devices, SAM.F and MFA4PD as well as its public display component is SSL-secured and uses token-based API access. Thus, the communication could suffer from all exploit issues that SSL faces, e.g., a man-in-the-middle attack (MiTM). Means to avoid these vulnerabilities concern the underlying systems running SAM.F or MFA4PD, or accessing them, as well as the networks security configuration.

The database used internally by SAM.F is not accessible directly.

In summary, for the current prototype setup, the security measures are sufficient. In future work, we will also address the question of whether the server can be compromised by other means.

V. Validation and Discussion

In this section, we briefly illustrate validation results for this work as well as future work and discuss our approach.

1) Validation Results and Future Work

We tested the prototype under laboratory conditions with mobile devices running Android with Firefox, Edge and Chrome Web browsers. In all tests, we were able to complete the authentication process. However, an evaluation in an everyday setting with a heterogenic group of users is still pending and scheduled for later this year, as outlined below.

In the evaluation we also focus the question, whether users prefer the Web application or the native mobile application.

Although this work does not focus on implementing more factors at this stage, together with the knowledge factor, the system can be extended with *biometrical* factors, using supplementary sources for MFA, such as fingerprint scanners, facial recognition, or voice biometrics. Mobile devices today offer these types of biometrical sensors. However, Web-based access to these sensors is prohibited by the browser's API. The new Xamarin-based mobile application can however access these sensors and we are planning on extending the mobile application's features.

In addition, the system still has to be evaluated quantitatively with a larger number of users, for example with regard to system's performance, usability, and the user's acceptance. The latter might depend on factors such as the setting the system is applied in.

2) Discussion

Public display authentication using the concept outlined in this contribution presents a feasible way, especially with the limitations of our scenario of minimal and no additional hardware requirements, as well as the limited browser capabilities for mobile devices.

Continuing the development with a mobile application that directly runs on smartphones however, new possibilities emerge.

One possibility is to display a QR code on the public display. After scanning the QR code with the mobile application, the session automatically gets authenticated.

Although implementations exist offering QR code readers for Web-based use, overall device compatibility remains an issue. For this reason, in the purely Web-based approach we did not use a Web-based QR code reader and decided to design the one-time-code mechanism.

In the future, we would have to evaluate, whether users prefer the mobile application over the Web application to authenticate themselves on their mobile devices.

The mobile application requires the users to download the app, but with QR-based authentication provides a lighter and faster authentication mechanism. Added biometrical checks using the mobile app could improve security when identifying users.

The Web application is light-weight, does not require a download nor an installation and could be used as so-called captive portal for local (public) Wi-Fi-hotspots deployed in range of public displays.

Both solutions can also co-exist.

For the purpose outlined, with MFA4PD this article presents a feasible solution that does not require additional hardware. The further development enhances the means, by which the MFA for public displays is achieved, as outlined in the next section.

VI. Conclusion

With the Semantic Ambient Media Framework (SAM.F), this contribution presents a framework that semantically interconnects (a) *media*, (b) *devices and applications*, and (c) *services*. The practical scenario illustrated describes the use of SAM.F together with MFA4PD to provide means of a secure method of multi-factor authentication for public displays and means of content retrieval from the semantic repository.

SAM.F provides Web-based access for devices and applications and features a service-based architecture, which allows for interaction with media, such as, e.g., text, pictures, audio, video, 3D objects, or 3D scenes. The concept of SAM.F regards Semantic Media independently of their encoding and automatically transcodes or converts media, where necessary and possible, to meet contexts', applications', and devices' specifications or criteria.

Using SAM.F also solves the problem of media being isolated for use in a single application or on a single device, as SAM.F interconnects users and their devices through its services and Semantic Media.

SAM.F can be used in contexts where interaction with Semantic Media is intended. Through technological means, SAM.F especially supports mobile contexts, e.g., through the application and device-specific provisioning of Semantic Media. Thus, SAM.F offers an enormous potential for exploiting new information sources, e.g., by the relationships of different informational modalities encoded semantically.

In future work, together with our project partner, the *Society for Audiovisual Archive of German-language Literature* based in the Hanseatic City of Bremen, we will utilize SAM.F as technical foundation to digitally enrich a cultural center for German literature. In this research project, SAM.F will interconnect media from various archives or libraries focusing on German literature and make them available on-site using public displays.

At the cultural center, the physical space will be enriched with digital media served provided by SAM.F. Curators will be enabled to adjust the exhibitions contents on-site by using dedicated functions accessible after authenticating on the public displays. User's will be served with personalized digital contents and personalized view after logging in on public displays.

Public display personalization is achieved through means of identifying the visitor (user) using authentication. Authentication on public displays is vulnerable to various attacks and technically presents a challenge, whenever public displays are connected to protected networks that are inaccessible for other devices and public displays are not equipped with dedicated user authentication hardware.

In this contribution, we present a technical solution, which addresses these challenges with a minimal technical solution. This makes use of a multi-factor authentication (MFA) applying the factors of ownership, knowledge and location. Not requiring any hardware upgrades for public displays, the solution implemented as a prototype makes use of the personal mobile devices of users, connecting them, as well as public displays to SAM.F.

Multi-factor authentication for public displays using SAM.F presents a feasible solution to the security issues public display authentication have. The solution presented securely authenticates users and lets them access private, restricted, or personal contents as well as sensitive functionality from and in SAM.F.

We have technically validated our approach under laboratory conditions. In the future, we plan to evaluate the system with a large number of users under everyday conditions. Research questions in this area also relate to the degree of security measures, that users are willing to accept in their everyday dealings with digital systems, as well as the question of how they perceive security issues with regard to their use of personal and private data and contexts on public displays.

It is our hypothesis that providing meaningful digital content in a body- and space related environment fosters mindful knowledge.

The Kingsbridge Research Center is a non-profit research company based in the United Kingdom. With our research it is one of our goals to strengthen the use of digital technology in public environments in our digital society. We achieve this goal through our scientific and project-oriented work. Currently, our non-profit activities and the development of new future-oriented projects is funded privately. At a time, when many are confronting digitization with skepticism and uncertainty, we are committed to communicating security in the mindful use of these technologies and through fostering awareness.

References

[1] D. Bouck-Standen and J. Kipke, 'Multi-Factor Authentication for Public Displays using the Semantic Ambient Media Framework', in *ADVCOMP'19*, *Best Paper Award*, IARIA, Porto, Portugal, 2019, pp. 30-35.

[2] D. Bouck-Standen, 'Introducing SAM.F: The Semantic Ambient Media Framework', in *AMBIENT'19*, IARIA, Porto, Portugal, 2019, pp. 40-45.

[3] A. Whitmore, A. Agarwal, and L. Da Xu, 'The Internet of Things—A survey of topics and trends', Inf. Syst. Front., vol. 17, no. 2, pp. 261–274, Apr. 2015.

[4] Eurostat, 'Internet use by individuals', 260/2016, Dec. 2016.

[5] S. Vrochidis, B. Huet, E. Y. Chang, and I. Kompatsiaris, Big Data Analytics for Large-Scale Multimedia Search. John Wiley & Sons Ltd, 2019.

[6] C. Vassilakis et al., 'Interconnecting Objects, Visitors, Sites and (Hi)Stories Across Cultural and Historical Concepts: The CrossCult Project', in Digital Heritage. Progress in Cultural Heritage: Documentation, Preservation, and Protection, Cham, 2016, pp. 501–510.

[7] E. Williams and J. Yerby, 'Google and Facebook Data Retention and Location Tracking through Forensic CloudAnalysis'. SAIS 2019 Proceedings. 3, electronic version.

[8] T. Kubitza, S. Clinch, N. Davies, and M. Langheinrich, 'Using Mobile Devices to Personalize Pervasive Displays', SIGMOBILE Mob Comput Commun Rev, vol. 16, no. 4, pp. 26-27, Feb. 2013.

[9] A. Ometov et al., ‘Multi-Factor Authentication: A Survey’, Cryptography, vol. 2, pp. 1-31, 2018.

[10] M. Eiband, M. Khamis, E. von Zezschwitz, H. Hussmann, and F. Alt, ‘Understanding Shoulder Surfing in the Wild: Stories from Users and Observers’, in Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems, New York, NY, USA, 2017, pp. 4254-4265.

[11] Y. Abdelrahman, M. Khamis, S. Schneegass, and F. Alt, ‘Stay Cool! Understanding Thermal Attacks on Mobile-based User Authentication’, in Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems, New York, NY, USA, 2017, pp. 3751-3763.

[12] K. Mowery, S. Meiklejohn, and S. Savage, ‘Heat of the Moment: Characterizing the Efficacy of Thermal Camera-based Attacks’, in Proceedings of the 5th USENIX Conference on Offensive Technologies, Berkeley, CA, USA, pp. 6-6, 2011.

[13] E. von Zezschwitz, A. Koslow, A. De Luca, and H. Hussmann, ‘Making Graphic-based Authentication Secure Against Smudge Attacks’, in Proceedings of the 2013 International Conference on Intelligent User Interfaces, New York, NY, USA, 2013, pp. 277-286.

[14] M. Khamis, R. Hasholzner, A. Bulling, and F. Alt, ‘GTmoPass: Two-factor Authentication on Public Displays Using Gaze-touch Passwords and Personal Mobile Devices’, in Proceedings of the 6th ACM International Symposium on Pervasive Displays, New York, NY, USA, 2017, pp. 8:1-8:9.

[15] DESTATIS, ‘Average internet use by individuals’, Survey on the private use of information and communication technologies (ICT), Online: https://www.destatis.de/EN/Themes/Society-Environment/Income-Consumption-Living-Conditions/Use-Information-Technologies/Tables/use-internet-age-ikt.html, Sep. 2019.

[16] T. Berners-Lee, ‘The Semantic Web’, Sci. Am., pp. 30–37, 2001.

[17] F. Nack, ‘The future in digital media computing is meta’, IEEE Multimed., vol. 11, no. 2, pp. 10–13, 2004.

[18] L. F. Sikos, ‘RDF-powered semantic video annotation tools with concept mapping to Linked Data for next-generation video indexing: a comprehensive review’, Multimed. Tools Appl., vol. 76, no. 12, pp. 14437–14460, Jun. 2017.

[19] C. Bizer, T. Heath, and T. Berners-Lee, ‘Linked data - the story so far’, Int J Semantic Web Inf Syst, vol. 5, no. 3, pp. 1–22, 2009.

[20] D. Bouck-Standen, ‘Construction of an API connecting the Network Environment for Multimedia Objects with Ambient Learning Spaces’, Master Thesis, DOI: 10.13140/RG.2.2.12155.00804, University of Luebeck, Luebeck, Germany, 2016.

[21] K. Blumenstein et al., ‘Bringing Your Own Device into Multi-device Ecologies: A Technical Concept’, in Proceedings of the 2017 ACM International Conference on Interactive Surfaces and Spaces, New York, NY, USA, 2017, pp. 306–311.

[22] P. J. Denning, Ed., The Invisible Future: The Seamless Integration of Technology into Everyday Life. New York, NY, USA: McGraw-Hill, Inc., 2002.

[23] U. Yadav, G. S. Narula, N. Duhan, and B. K. Murthy, ‘An overview of social semantic web framework’, in 2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom), 2016, pp. 769–773.

[24] C. Wang, H. Yang, and C. Meinel, ‘A deep semantic framework for multimodal representation learning’, Multimed. Tools Appl., vol. 75, no. 15, pp. 9255–9276, Aug. 2016.

[25] P. A. Shaw, M. A. Mikusz, P. T. Nurmi, and N. A. J. Davies, ‘Tacita-A Privacy Preserving Public Display Personalisation Service’, UbiComp 2018, pp. 448-451, 2018.

[26] N. Memarovic, I. Elhart, A. Michelotti, E. Rubegni, and M. Langheinrich, ‘Social Networked Displays: Integrating Networked Public Displays with Social Media’, in Proceedings of the 2013 ACM Conference on Pervasive and Ubiquitous Computing Adjunct Publication, New York, NY, USA, 2013, pp. 55-58.

[27] M. Mannan and P. C. van Oorschot, ‘Using a Personal Device to Strengthen Password Authentication from an Untrusted Computer’, in Financial Cryptography and Data Security, Berlin, Heidelberg, 2007, pp. 88-103.

[28] P. Oliveira and P. Gomes, ‘Instance-based Probabilistic Reasoning in the Semantic Web’, in Proceedings of the 18th International Conference on World Wide Web, New York, NY, USA, 2009, pp. 1067–1068.

[29] C. Tzelepis et al., ‘Event-based media processing and analysis: A survey of the literature’, Image Vis. Comput., vol. 53, pp. 3–19, 2016.

Forensic Investigations of Popular Ephemeral Messaging Applications on Android and iOS Platforms

M A Hannan Bin Azhar, Rhys Cox and Aimee Chamberlain
School of Engineering, Technology and Design
Canterbury Christ Church University
Canterbury, United Kingdom
e-mail: hannan.azhar@canterbury.ac.uk; rhys-1998@hotmail.co.uk; aimee.chamberlain@yahoo.co.uk;

Abstract—**Ephemeral messaging applications are growing increasingly popular on the digital mobile market. However, they are not always used with good intentions. Criminals may see a gateway into private communication with each other through this transient application data. This could negatively impact criminal court cases for evidence, or civil matters. To find out if messages from such applications can indeed be recovered or not, a forensic examination of the device would be required by the law enforcement authority. This paper reports mobile forensic investigations of ephemeral data from a wide range of applications using both proprietary and freeware forensic tools. Both Android and iOS platforms were used in the investigation. The results from the investigation uncovered various artefacts from the iOS device including account information, contacts, and evidence of communication between users. The Android device uncovered evidence of communications, and several media files assumed to be deleted within a storage cache in the Android file system. The forensic tools used within the investigations were evaluated using parameters from the National Institute of Standards and Technology's (NIST) mobile tool test assertions and test plan.**

Keywords— Mobile forensics; Digital forensics; NIST measurements; Oxygen Forensics; Ephemeral messaging apps; EMAs.

I. Introduction

The growth of ephemeral messaging applications (EMAs) is also posing a problem to the enforcement of law, with apps being proving a concern for activities like cyberbullying [1] or even high-end criminal activity like terrorism [2]. Criminals may use regular chatting applications, but there is a growing opportunity within the mobile application market for criminals to use ephemeral messaging applications, which allow users to send messages/multimedia etc. to each other with the messages only lasting for a certain period of time [3]. Barker [4] reported that criminals are moving away from dark web interactions and onto EMAs such as Facebook Messenger, Snapchat, and Wire etc. It is thought this is happening because data in these applications is known to delete itself, which is prime for criminal communications. For example, Snapchat allows users to send 'Snaps' to each other containing pictures, which are deleted once the recipient user closes the message [5].

It is not just criminals using EMAs. Mobile phones are an essential part of modern-day life. According to the Global System for Mobile Communications [6], there were five billion mobile users in the world by the second quarter of 2017, with a prediction that another 620 million people will become mobile phone users by 2020, together that would account for almost three quarters of the world population. Due to the increasing popularity in mobile phones, there is naturally an increasing concern over mobile security and how safe communication between individuals or groups is. It is known that EMAs can be used in civil concerns such as evidence of liability in business [7]. Most notably, the United States department of justice imposed heavy restrictions on the use of EMAs by employees in 2018 as part of a scheme to reduce illegal bribery within businesses [8]. The rationale for the use of the applications was reported to be more complex than covering tracks, and that employees themselves had started to turn to using them of their own accord for reasons such as more reliable service. From this, it is clear that the use of EMAs is moving from the use of criminals and privacy advocates, to the general populace as well.

The two most popular EMAs on the mobile market currently are Snapchat and Facebook Messenger. According to Constine [9], Snapchat has a daily user total of 190 million users. According to Noyes [10], Facebook messenger has an average daily user intake of 2.1 billion users. The statistics show a high intake of users within these EMAs. With the ever-growing ephemeral market, it is vital to both civil matters and criminal cases to find out just how truthful the applications are about data being deleted and unrecoverable.

With all the opportunities for new crimes to be committed through growing technology, it is crucial to ensure law enforcement agencies have the appropriate software and methods to deal with these crimes. This paper will report forensic investigations of two mobile phones: one on an Android and the other on iOS platforms. The Android device was rooted; however, the iPhone was not jail broken. This will give an interesting insight to the investigation as different amounts of data may be recovered according to if the device is rooted or not. This paper will be using a variety of forensic tools to extract the mobile devices as well as comparing and examining each tool according to the NIST Measurements Mobile Device Tool Test Assertions and Test Plan [11]. The main contributions include a taxonomy of tools for forensic analysis of mobile platforms, along with hands-on tests of these tools on several Android and iOS messaging apps. The paper's results cover the relative effectiveness of the forensic frameworks, as well as various interesting security findings among the mobile apps.

The remainder of the paper will be organised as follows: Section II will discuss existing research in relation to mobile phone forensics, including forensic tools and ephemeral data. The methodology used during the analysis process will be discussed in Section III. Results and analysis will be reported in Section IV. Finally, Section V will conclude the paper and include possible future work.

II. Literature Review

There is already a vast amount of research on mobile forensics in general, which includes comparing forensic tools, performing different types of mobile acquisitions and focusing on particular pieces of data within the mobile device. There is also work completed on non-EMAs, such as Ovens et al. [12] conducted a forensic analysis on Kik Messenger on iOS. While there have been similar studies in a wide range of apps, the focus of this review is to highlight the findings in extraction of artefacts from the apps, which are specifically ephemeral.

One study undertaken by Sathe et al. [13] provided a broad overview of the available forensic acquisition methods for mobile device forensics, including several freeware options. The study undertook comprehensive analysis of both physical and logical data acquisition options and compared those options via several categories, i.e., Cost, Accuracy, Data Integrity, Training required, OS reliance, Root required etc., all of which would prove useful for identifying the practicality of the tools/techniques in professional scenarios, as well as the forensic soundness of the techniques in question, a pertinent characteristic when dealing with more disruptive techniques, such as those which require root access, as any alteration to the data stored within a device/image may well remove the reliability of a given piece of evidence in the eyes of the court. The results of the study showed that of the chosen forensic tools, AFLogical, Andriller and Dr. Fone toolkit, each provided evidence data in areas, which the other was lacking, leading to the conclusion that the use of multiple forensic tools in a given mobile forensic investigation may well be ideal.

Azhar et al. [14] conducted a forensic experiment of two EMAs: Telegram and Wickr using Autopsy and logically acquiring a database file, as well as performing a RAM dump. Results showed that the application 'Wickr' stored received messages in encrypted "wic" files. The RAM dump recovered username information from Wickr and artefacts from Telegram. This investigation was an Ephemeral application comparison using Android platforms. The investigation more looked into packages and files within the application itself instead of using a mobile forensic tool. This would be interesting for future work as well as perhaps performing the same investigatory analysis on an iOS device.

Al-Hadadi et al. [15] forensically investigated a mobile device, an iPhone 4 running iOS 5.0.1 previously jailbroken by the mobile phone owner, as a part of a real legal case. The case was from the Sultanate of Oman, and the aim of the investigation was to forensically examine the iPhone to determine if the device had been hacked and sent messages over the application 'WhatsApp' out to the owner's contact list. In the investigation, the ISP report of the device was observed and examined, and two forensic tools were used to extract and examine mobile data, one tool being the Universal Forensic Extraction Device's (UFED) physical analyser Cellebrite, and the other being the Oxygen Forensic Suite. The credibility of both tools is highly regarded by computer forensic experts. Results showed that Cellebrite recovered more forensic evidence than Oxygen, including call log artefacts, SMS messages, web history, etc.

Another study, by Umar et al. [16] investigated the specific forensic evidence recoverable from the use of WhatsApp, a popular secure messaging application. The study took a rooted mobile device with Android 5.0.1 and used it in communications with a second device in order to simulate the standard daily use of the application. The mobile device was then forensically analysed by 3 different mobile forensic tools: WhatsApp DB/Key extractor, Belkasoft Evidence and Oxygen Forensic Detective 6.4.0.67. In addition to comparing the results of each tool's analysis, the tools were assessed by various levels of the NIST Mobile device tool test assertions [11], a set of test requirements and guidelines produced to assist in the evaluation of mobile forensic tools. Each tool was assessed by all the baseline assessments and a select number of the optional assessments, before comparing the tools by the number of assessments they passed. The results of the study revealed that Oxygen Forensic Detective provided the most forensically valuable data, managing to identify evidence of the test data in both logical and physical extraction, and passed the most assessment parameters put against it by the NIST guidelines, failing only five out of the twenty-two assertions and functionality tests.

A study undertaken by Naughton et al. [17] provided an investigation into data left by specific apps on mobile and personal devices. Said study utilised two mobile devices, using Android and IOS respectively, alongside a windows 10 based laptop using an Android emulator. The applications selected for the study included various shared and device/OS specific apps, including two ephemeral apps: Snapchat and Instagram. Each device was used to gather forensically valuable data before undergoing forensic analysis, after which the data would be deleted to simulate a criminal covering their tracks and would undergo analysis once more. The study showed detailed information of what each forensic tool could recover from the test devices, with subcategories for each specific application, device and file type. While less focussed on the tools utilised during analysis, this study put heavy focus into the realism of the forensic analysis within the experiment, going as far to consult digital forensic specialists and 15 separate police forces within England and Wales to ensure the experiment would prove as realistic a scenario as possible, a level of justification lacking in every other study found. The results of the study showed that the laptop and the iPhone provided the most forensically valuable data through analysis, and, more relevantly, that both EMAs used in the study, Snapchat and Instagram, provided no recoverable data that the chosen forensic tools, Cellebrite UFED and Autopsy, could identify.

As can be seen from this brief review of the literature, there is not much reported literature in extraction of ephemeral artefacts especially on iOS platforms. This paper will contribute to investigate artefacts recovery from both iOS and Android using wide range of EMAs. Comparisons will be made in evaluation of artefacts recovered using various tools following the guidelines by the NIST Measurements Mobile Device Tool Test Assertions and Test Plan [11]. The paper's results cover the relative effectiveness of the forensic frameworks, as well as various interesting security findings among several Android and iOS messaging apps.

III. METHODOLOGY

Methodology section will detail choice of devices, chosen applications, forensics tools used and investigation process including the testing methodologies using NIST measurements. The investigation was carried out according to the four good practice guidelines of the Association of Chief Police Officers (ACPO) [18]. For example, the third principle of the guidelines state that an audit trail should be recorded throughout the investigation in a manner, such that a third party could recreate the steps taken in the investigation and get the same results.

A. Chosen Devices

The two mobile devices used within the investigation was an iOS device: iPhone 6s [19], and an Android device: Vodafone VF695 [20]. According to Jkielty [21], there is just a 2.3% difference in UK in the market share between Android and iOS devices, with the iOS market having the edge. Vodaphone was running an Android 4.4.2 (KitKat) OS and iPhone had iOS 9. To investigate wide range of exploits, one of the phones was rooted (Android). Having the root level access, it was hoped to gain more access to recover detail artefacts, including deleted files. In case of root level access, while forensic soundness can be questioned, the artefacts could still be valuable giving clues to further direction of investigation, which eventually may lead to gather concrete evidences to be presented in court with sufficient justification.

B. Ephemeral Applications

A wide range of ephemeral messaging apps were selected for the investigation as listed in Table I. Some applications are more popular (Snapchat for iOS and Facebook Messenger for Android) than the others but they all varied in their ephemeral features and target audiences. Details of these applications are given next.

1) iOS Applications: Applications as listed for iOS in Table I were all chosen for different reasons. Snapchat is one of the most popular EMAs. According to Omnicore [28], more than 25% of mobile phone users are on Snapchat, with 71% of the users being aged between 17 to 24. Cyberdust, was chosen due to the difference in its ephemeral features compared to other apps. The encrypted messages within the app delete themselves between users after 24 hours of it being sent [23]. The application also has other uses, such as a "watchdog" feature where users can check their email addresses to see if any data breaches have been completed. Another feature is known as "Stealth Search", where users can search the Internet privately, supposedly without any cookie trackers or trace remnants. This application was selected for the investigation as it creates ephemeral data, and it has many different functions, which allows the user to use the application for multi-purpose functions.

TABLE I. MOBILE DEVICE AND APPLICATIONS

Mobile used	**Ephemeral Messaging Apps**	
	App Name	***Version***
iPhone 6s	Snapchat [22]	10.55.1
	Cyberdust [23]	5.6.1.1049
	Confide [24]	8.3.1
Vodafone VF695	Facebook Messenger [25]	215.1.0.21.101
	Signal [26]	4.39.4
	Wire [27]	3.30
	Confide [24]	5.9.5

The final application, Confide [24], was chosen because of its end to end message encryption between users. Furthermore, the application does not allow screenshots to be taken from users. The messages between users are self-destructing once the recipient has read the message, and the user can only read the message by swiping down on the message on the screen to view it. Furthermore, the user can adjust the settings to change the ephemeral nature of the messages, if a message is not opened within 48 hours, the content of the message will delete itself regardless. All of these features would create an interesting investigation, as the application advertises very strong messaging security, so it would be intriguing to test the security through this forensic investigation.

2) Android Applications: Like iOS, Confide was also used for the Android investigation. Among other applications, Facebook Messenger is one of the most popular EMAs on the market with a similar popularity to Snapchat, as used on the iOS device. According to Google Play[29], as of March 2020 Facebook Messenger has over one billion downloads on the Android market. Facebook Messenger has a recent implementation of a new feature, which is a secret conversations function. It can facilitate encrypted and ephemeral communications between two parties, utilising the signal messaging protocol as previously used in the application 'Signal' description. The ephemeral features exist as a set of optional timers, with 11 delay options between 5 seconds and a day.

Signal is an open source encrypted messaging application with ephemeral capabilities, developed by the company of the same name. As a company, Signal is responsible for producing an encryption-based messaging protocol, also by the same name, which is utilised by multiple other secure messaging applications like WhatsApp and Facebook Messengers secret conversations feature [30]. Signal's ephemeral capabilities come in the form of an optional timer to set for messages, with 11 different settings between five seconds and one week delays for removal. All of this information makes it a perfect EMA for forensic investigation.

The next application was Wire, which is a secure messaging application developed by Wire Swiss [27] and includes ephemeral messaging features. The application is targeted for use in business, with a majority of its promotional descriptions detailing secure communications between teams of employees, and further detailing its free version as ideal for home or family use. The ephemeral capabilities of Wire exist as a set of 6 optional timers between 10 seconds and 4 weeks delay. Its popularity is around twice as much as Confide, with over 1 million downloads on the Google play store.

C. *Forensic Tools*

Oxygen Forensic Detective Enterprise [31] version 10.3.0.100 is a commercial forensic tool that was used to extract and examine both the iOS and the Android phone. Oxygen Forensic Detective is a specialised mobile forensics tool developed by Oxygen Forensics Inc and utilised by professional digital forensic investigators in law enforcement. The specific extraction capabilities for the tool range depending on the device being analysed, but in general it provides several options for extraction depending on the individual device requirements, and provides highly detailed and clear visual representations of the data both in the applications user interface and in the reports it can produce. Various viewers are built into Oxygen Forensic Detective, allowing users to view the contents of files such as SQL databases within the program and make reports specifically from the contents [31].

MOBILedit Forensic Express [32] version 6.1.0.15480 is a commercial forensic tool that was used to extract and examine the iPhone device. MOBILedit can create a logical and physical acquisition of a mobile device and can recover deleted files as well as retrieve mobile data. It is used widely across law enforcement in over 70 countries and is also used in military investigations [32].

Andriller version 3.0.3 [33] is an Android specific proprietary forensic tool developed by the software team of the same name and allows for data extraction from both rooted and unrooted Android devices. This tool was used to extract and examine the Android device. Data extracted from suitable devices is extracted to a directory of the users choosing in the form of several different reports, and folders for shared storage data. Various utility tools come alongside the extraction capabilities, such as a screenshot function, lockscreen decoders and specific database decoders for a specific list of supported applications and sources. For the purposes of this experiment a trial licence was acquired to use the full version for a time period of 30 days.

FTK Imager version 4.1.1.1 is a freeware disk analysis tool produced by AccessData [34] as part of their Forensic Toolkit product range. This forensic tool was used to extract and examine the Android phone. While only a free version of the products AccessData have produced, FTK Imager is still a versatile tool for extracting disk and RAM images, as well as analysing existing forensic images. Lacking elaborate methods of displaying extracted data, FTK displays the filesystem of the chosen image files and provides both plaintext and hexadecimal viewing panes to display file contents. While not intrinsically advertised as a mobile forensic tool, FTK Imager is still capable of analysing an existing image file extracted from a mobile device and could serve as a mobile forensic analysis tool if necessary.

Autopsy 3.0.8 [35] was used to analyse an forensic image file. Autopsy is the graphical frontend for a set of Linux forensics tools called the Sleuthkit. This contains tools that allow for the recovery of deleted data. Autopsy also allows for the processing of unallocated space, which is an important part of the analysis as ephemeral messaging functions rely on the deletion of data. Artefacts such as deleted files sent as attachments to messages can be recovered using Autopsy [14].

Kali Linux is not a forensic tool, instead an operating system that was used for forensic analysis on the Android device. It can produce disk and mobile devices images through the use of the DD command, which serves to create a bit for bit copy of a file or directory. Accessing a mobile device to utilise this method of imaging requires several other tools, Android debug bridge and BusyBox, on top of rooting the device to allow direct access to the mobile devices root directory. As a result, imaging a mobile device with this method is highly questionable in its forensic soundness, however it is still a viable technique in the event a device requires imaging without specialised tools and equipment. While not being assessed as a forensic tool, given its only functionality is copying data bit for bit, both Autopsy and FTK Imager would be using image files produced by Kali Linux for their analysis as an example of full data extraction and analysis with freeware tools [36][37].

D. *NIST Measurements*

NIST, otherwise known as the National Institute of Standards & Technology, is an institution based on technological and scientific advancement. They provide data and professional standards of technology for multiple scientific fields, including the forensic sciences. To ensure the quality and functionality of the tools, equipment and practices utilised [11]. NIST produced a set of standards detailing ten baseline functionality standards and twenty-two optional standards for assessing tools on their suitability for mobile forensic extraction and examination. The main goal of the guidelines is to determine a tool's ability to accurately acquire specific data objects populated onto the feature phone, smart phone, tablet or credit cards. Before proceeding with the examination of the target mobile device for this research, the tools would be assessed with the ten baseline test assertions, MDT-CA-01 through to MDT-CA-10. For example, the first test assertion, MDT-CA-01, indicates if a mobile device forensic tool provides the user with an "Acquire All" data objects acquisition option then the tool should complete the logical or filesystem acquisition of all data objects without error. An accurate acquisition copies means that the bytes of the acquired data object are identical to the bytes of the data object on the device. The NIST guidelines also have some optional assertions focussing on physical extraction ability of a tool, which were omitted for the tests as all versions of the tools used for analysis lacked those features by default.

E. *Testing Methodology*

The iPhone 6s was extracted and examined first using Snapchat on the iPhone device. For this application, three contacts were added and two of those contacts had communication sending picture messages, as well as written messages back and forth. Ten picture messages were exchanged, three written messages were marked as 'saved', while one of other messages was not saved. The username for the mobile owner was 'aimee_test19'. For the Snapchat, the ephemeral artefacts were the picture messages for the investigation.

Cyberdust had a total of eleven messages exchanged. Two of the messages were picture messages. The username for the mobile owner was linked directly to the mobile number of the device instead of an account like Snapchat.

Confide had a total of seven messages exchanged on the iPhone device. Like Cyberdust, here also the username for the mobile owner was linked directly to the mobile number. For the investigation's purpose, only the secure messaging feature was used, where messages were encrypted and deleted after 24 hours.

Oxygen Forensic outputs a GUI home page, which displays the kinds of information that has been extracted, allowing an investigator to navigate around the mobile contents easily. The 'Applications' tile was selected to investigate the three Ephemeral applications mentioned previously, including any data the applications held of the user, conversation data, etc. Once the 'Applications' tile was examined, the 'Passwords' tile was selected and examined. This was to see if any passwords were stored within the three EMAs to test the general security of the applications.

The same extraction process was completed in MOBILedit Forensic Express [8]. Unlike Oxygen, MOBILedit outputs the mobile device extraction into a report. However, there was a contents page produced within the report. There was also a separate section for both 'Applications' and 'Passwords' similarly to Oxygen. Both of those sections were examined. In the next stage of the examination, a general keyword search was made within the Oxygen and MOBILedit in search for artefacts. The keywords searched included 'Snapchat', 'Dust' and 'Confide'. This was completed in case any other information relating to the applications was extracted and missed previously. The application names were used for the searches, as in a real-life scenario the digital forensic investigator may not know the contents of the messages and may be left with no other search options other than the application names.

Next, the Android device was extracted and examined by the nominated forensic tools. Assessment of the supported messaging capabilities within each application was performed and then messaging transcripts for each of the applications were produced, detailing the messages and attachments sent between the Android phone and a personal phone. Each application would be used to produce five distinct text-based messages, exchanges of specifically named image files, and then exchanges of distinct audio messages and document files for the apps that supported audio and file-based attachments.

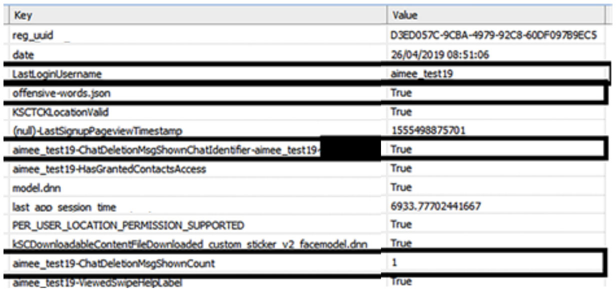
Once the test data was created to the specifications of the transcripts, the device was then forensically analysed, first by the proprietary forensic tools and then the freeware forensic tools. The forensically valuable artefacts were recorded through screenshots and were extracted, if necessary, to identify contents, in the case of the media file attachments. Once thorough analysis of the device was performed with all four chosen tools, the applications were then uninstalled from the device to simulate anti-forensic activity, after which a second stage of analysis was then performed to see if any of the artefacts recovered in the first stage of analysis were still recoverable in a forensically valuable form.

IV. RESULTS

This section covers the key findings from the analysis described in Section III. The results will be broken down into multiple sections: iOS results from forensic tools used to extract the iPhone 6s, Android results from the forensic tools used to extract the Android Vodafone VF695.

A. *Oxygen Forensic for iOS*

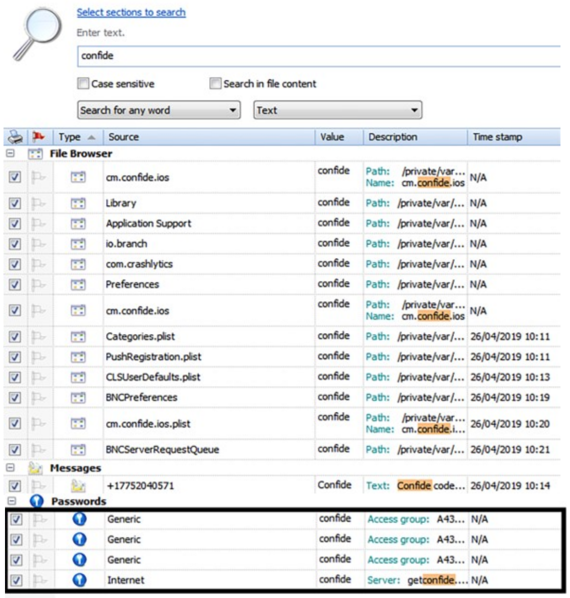
A list of applications on the mobile device was found in the 'Applications' tile using Oxygen. Snapchat was the first application to be investigated. Figure 1 shows Snapchat data. Four areas were highlighted within the figure. This included the login username 'aimee_test19', that was used to log into Snapchat and detection of an 'offensive words' used in messages. The next highlighted section was the evidence that there was messaging communication between a user 'aimee_test19' and another user. The final highlighted section shows a chat deletion message count with a value of one, which indicates that a message was deleted by the user, which was a true case. A general search of the extracted mobile device was conducted using the search feature on Oxygen Forensics. The findings included general application data within the file browser, such as the Snapchat library, stickers, etc.



Key	Value
reg_uuid	D3ED057C-9CBA-4979-92C8-60DF0978 9EC5
date	26/04/2019 08:51:06
LastLoginUsername	aimee_test19
offensive-words.json	True
KSCTCKLocationValid	True
(null)-LastSignupPageviewTimestamp	1555498875701
aimee_test19-ChatDeletionMsgShownChatIdentifier-aimee_test19-[redacted]	True
aimee_test19-HasGrantedContactsAccess	True
model.dnn	True
last app session time	6933.77702441667
PER_USER_LOCATION_PERMISSION_SUPPORTED	True
kSCDownloadableContentFileDownloaded_custom_sticker_v2_facemodel.dnn	True
aimee_test19-ChatDeletionMsgShownCount	1
aimee_test19-ViewedSwipeHelpLabel	True

Figure 1. Snapchat artefacts in Oxygen.

The next application investigated was Cyberdust. Previously, Snapchat appeared in the 'Applications' tile on Oxygen displaying itself as a normal application. However, with Cyberdust only the application folder was recognised, and Cyberdust was not acknowledged as a full application like Snapchat, however the folder proved there was evidence of an application called Cyberdust being present on the mobile device. This could be because the application did not require a username and password to log in, rather the user's mobile number instead, which therefore meant the phone did not identify it as an application in the same way as Snapchat, where it requires a username and password. Figure 2 shows results from a general search of the word ‘dust’.



Select sections to search

Enter text.

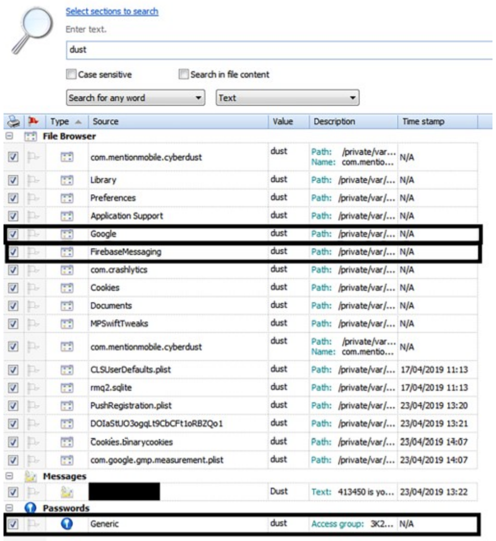
confide

☐ Case sensitive ☐ Search in file content

Search for any word ▾ Text ▾

		Type	Source	Value	Description	Time stamp
⊟		**File Browser**				
☑			cm.confide.ios	confide	Path: /private/var... Name: cm.confide.ios	N/A
☑			Library	confide	Path: /private/var/...	N/A
☑			Application Support	confide	Path: /private/var/...	N/A
☑			io.branch	confide	Path: /private/var/...	N/A
☑			com.crashlytics	confide	Path: /private/var/...	N/A
☑			Preferences	confide	Path: /private/var/...	N/A
☑			cm.confide.ios	confide	Path: /private/var... Name: cm.confide.ios	N/A
☑			Categories.plist	confide	Path: /private/var/...	26/04/2019 10:11
☑			PushRegistration.plist	confide	Path: /private/var/...	26/04/2019 10:11
☑			CLSUserDefaults.plist	confide	Path: /private/var/...	26/04/2019 10:13
☑			BNCPreferences	confide	Path: /private/var/...	26/04/2019 10:19
☑			cm.confide.ios.plist	confide	Path: /private/var... Name: cm.confide.i...	26/04/2019 10:20
☑			BNCServerRequestQueue	confide	Path: /private/var/...	26/04/2019 10:21
⊟		**Messages**				
☑			+17752040571	Confide	Text: Confide code...	26/04/2019 10:14
⊟		**Passwords**				
☑			Generic	confide	Access group: A43...	N/A
☑			Generic	confide	Access group: A43...	N/A
☑			Generic	confide	Access group: A43...	N/A
☑			Internet	confide	Server: getconfide....	N/A

Figure 2. Cyberdust general search Oxygen



Select sections to search

Enter text.

dust

☐ Case sensitive ☐ Search in file content

Search for any word ▾ Text ▾

		Type	Source	Value	Description	Time stamp
⊟		**File Browser**				
☑			com.mentionmobile.cyberdust	dust	Path: /private/var... Name: com.mentio...	N/A
☑			Library	dust	Path: /private/var/...	N/A
☑			Preferences	dust	Path: /private/var/...	N/A
☑			Application Support	dust	Path: /private/var/...	N/A
☑			Google	dust	Path: /private/var/...	N/A
☑			FirebaseMessaging	dust	Path: /private/var/...	N/A
☑			com.crashlytics	dust	Path: /private/var/...	N/A
☑			Cookies	dust	Path: /private/var/...	N/A
☑			Documents	dust	Path: /private/var/...	N/A
☑			MPSwiftTweaks	dust	Path: /private/var/...	N/A
☑			com.mentionmobile.cyberdust	dust	Path: /private/var... Name: com.mentio...	N/A
☑			CLSUserDefaults.plist	dust	Path: /private/var/...	17/04/2019 11:13
☑			rmq2.sqlite	dust	Path: /private/var/...	17/04/2019 11:13
☑			PushRegistration.plist	dust	Path: /private/var/...	23/04/2019 13:20
☑			DOIaStUO3ogqLt9CbCFt1oRBZQo1	dust	Path: /private/var/...	23/04/2019 13:21
☑			Cookies.binarycookies	dust	Path: /private/var/...	23/04/2019 14:07
☑			com.google.gmp.measurement.plist	dust	Path: /private/var/...	23/04/2019 14:07
⊟		**Messages**				
☑			[redacted]	Dust	Text: 413450 is yo...	23/04/2019 13:22
⊟		**Passwords**				
☑			Generic	dust	Access group: 3K2...	N/A

Figure 3. Confide general search Oxygen.

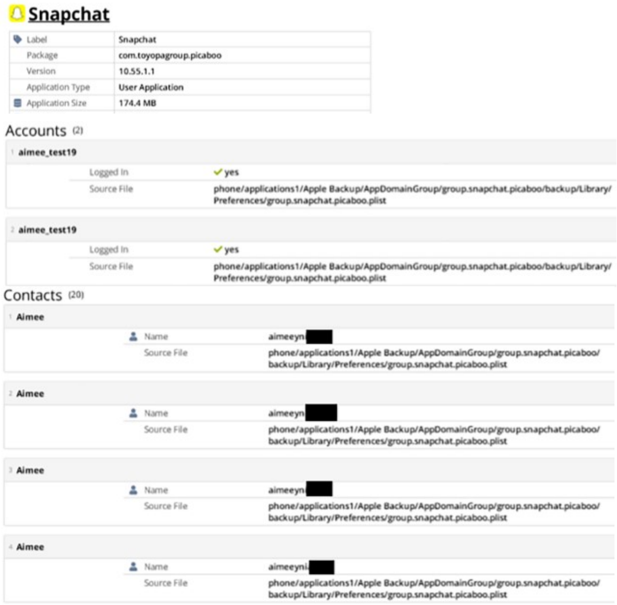
The results from the file browser show private folder pathway names. This acknowledges the existence of the application itself within the mobile device, but it does not have definitive messages between two users. However, as Figure 2 highlights, both ‘Google’ and ‘FireBaseMessaging’ were in the private folders. FireBase, formerly known as Google Cloud Messaging, is a cross-platform cloud solution for messaging [38]. This means that the data from the application could be deleted on the mobile device itself, but data may be uploaded elsewhere in the cloud and therefore access could be granted through that, but this needs to be explored further. For this investigation however, it was proven that the application, Cyberdust, was a messaging application, but there was no evidence of messages between two users. Additionally, Figure 2 highlights a ‘Generic’ password in the search. This shows that the application has stored a password, most likely the user’s password, but has encrypted it with a token.

The last application investigated was Confide. Similarly, to Cyberdust, there was little evidence to prove the application Confide existed under the ‘Applications’ tile. Unlike Snapchat, the only data Confide showed within the Applications tile was a private pathway. Figure 3 shows results from a general search of the word ‘Confide’. The results showed general application files in private folders within the file browser. The number ‘+17752040571’ in Figure 3 is a verification text message from the application itself to verify the user's account. Even though there was evidence that the Confide was installed in the phone, no application specific communication between users or user log in details was recovered. There were however, four passwords that were linked to the application Confide. Three being generic and one being an Internet password. The passwords could have been the user login password, but the passwords were encrypted. Therefore, the passwords were not visible and were secure for the user’s account.

B. MOBILedit Forensic Express for iOS

The next part of the investigation was to examine the mobile device and the applications under examination using MOBILedit Forensic Express. Once the report generated from MOBILedit, the next step in the investigation was to navigate to the applications section of the report focusing on Snapchat, Cyberdust and Confide. The first application investigated was Snapchat. Figure 4 shows the accounts used to log in to Snapchat and the list of contacts and the pathways to “plist”, where the contact’s information was stored.

Figure 4 proves that the mobile device was linked to a Snapchat account with the username ‘aimee_test19’, and both victim and suspect were likely to had communication as the names (username blackened out) appeared on the contact log of the phone. This finding would let further interrogation to the suspect during the investigation. Similarly to Oxygen, MOBILedit also found general application artefacts under private folders, but nothing significant that contributed to the investigation. The next application that was looked at within MOBILedit was Cyberdust. Figure 5 shows Cyberdust application data and the account the mobile device linked to the application. As Figure 5 displays, one account was evidently linked from the mobile device to the application. This proves the mobile user did use the application and also had an account. However, there were no account details recovered from that section of the report and unlike Snapchat, no contacts were found either, when the user did in fact have one contact on the application. However, this may be because the user contact was directly through a mobile number, which was already in the mobile user’s general phone contact list. Therefore, the contact may not have been stored on the application itself.



Snapchat

Label	Snapchat
Package	com.toyopagroup.picaboo
Version	10.55.1.1
Application Type	User Application
Application Size	174.4 MB

Accounts (2)

1 aimee_test19

Logged In	✓ yes
Source File	phone/applications1/Apple Backup/AppDomainGroup/group.snapchat.picaboo/backup/Library/Preferences/group.snapchat.picaboo.plist

2 aimee_test19

Logged In	✓ yes
Source File	phone/applications1/Apple Backup/AppDomainGroup/group.snapchat.picaboo/backup/Library/Preferences/group.snapchat.picaboo.plist

Contacts (20)

1 Aimee

Name	aimeeyn[redacted]
Source File	phone/applications1/Apple Backup/AppDomainGroup/group.snapchat.picaboo/backup/Library/Preferences/group.snapchat.picaboo.plist

2 Aimee

Name	aimeeyn[redacted]
Source File	phone/applications1/Apple Backup/AppDomainGroup/group.snapchat.picaboo/backup/Library/Preferences/group.snapchat.picaboo.plist

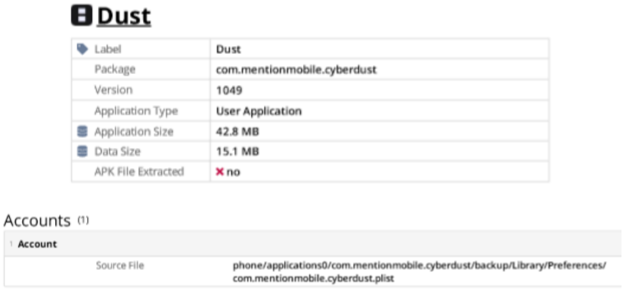
3 Aimee

Name	aimeeyn[redacted]
Source File	phone/applications1/Apple Backup/AppDomainGroup/group.snapchat.picaboo/backup/Library/Preferences/group.snapchat.picaboo.plist

4 Aimee

Name	aimeeyni[redacted]
Source File	phone/applications1/Apple Backup/AppDomainGroup/group.snapchat.picaboo/backup/Library/Preferences/group.snapchat.picaboo.plist

Figure 4. Snapchat data in MOBILedit.



Dust

Label	Dust
Package	com.mentionmobile.cyberdust
Version	1049
Application Type	User Application
Application Size	42.8 MB
Data Size	15.1 MB
APK File Extracted	✗ no

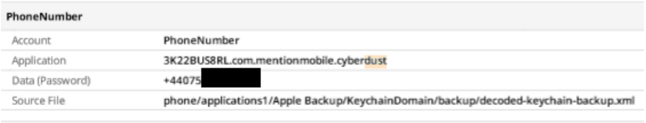
Accounts (1)

1 Account

Source File	phone/applications0/com.mentionmobile.cyberdust/backup/Library/Preferences/com.mentionmobile.cyberdust.plist

Figure 5. Cyberdust application in MOBILedit.

Some data was recovered from the 'Passwords' section within the generated report as shown in Figure 6. The "Password" had the label of "PhoneNumber". The data itself was the mobile user's unencrypted phone number. No other data was found in the passwords section of the report. Since the phone number was stored by the application, it shows evidence of a user account on the mobile device.

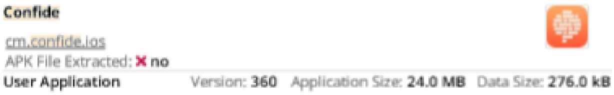


PhoneNumber

Account	PhoneNumber
Application	3K22BUS8RL.com.mentionmobile.cyberdust
Data (Password)	+44075[redacted]
Source File	phone/applications1/Apple Backup/KeychainDomain/backup/decoded-keychain-backup.xml

Figure 6. Phone number recovery in Cyberdust.

The last application MOBILedit investigated on the mobile device was Confide. Figure 7 displays Confide within the application list generated by MOBILedit. Unlike Snapchat and Cyberdust, the generated report displayed no information on contacts or accounts within Confide. Similar to the finding by Oxygen, Figure 6 suggests that there was little evidence that the mobile device had an account with the application.



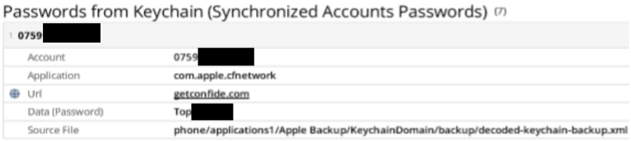
Confide

cm.confide.ios

APK File Extracted: ✗ no

User Application Version: 360 Application Size: 24.0 MB Data Size: 276.0 kB

Figure 7. Confide application data MOBILedit.



Passwords from Keychain (Synchronized Accounts Passwords) (7)

1 0759[redacted]

Account	0759[redacted]
Application	com.apple.cfnetwork
Url	getconfide.com
Data (Password)	Top[redacted]
Source File	phone/applications1/Apple Backup/KeychainDomain/backup/decoded-keychain-backup.xml

Figure 8. Phone number and password artefacts in Confide.

Figure 8 displays the mobile number and the password artefact recovered from the application. The account was the mobile user's unencrypted phone number, and the password was the user password for the created account for the application. The password was also unencrypted. This suggested that the application have stored the user password unsafely.

C. *NIST Measurements for iOS*

MOBILedit met all nine NIST measurement requirements tested in this research, while Oxygen did not, yet Oxygen did meet most of them. Comparisons of all nine test cases have been reported in Table II. MOBILedit provided the user with a "Select All" individual data objects (MDT-CA-02) while completing the logical or filesystem acquisition, it also provided the ability to "Select Individual" data objects (MDT-CA-03) for acquisition; in both of these cases Oxygen failed.

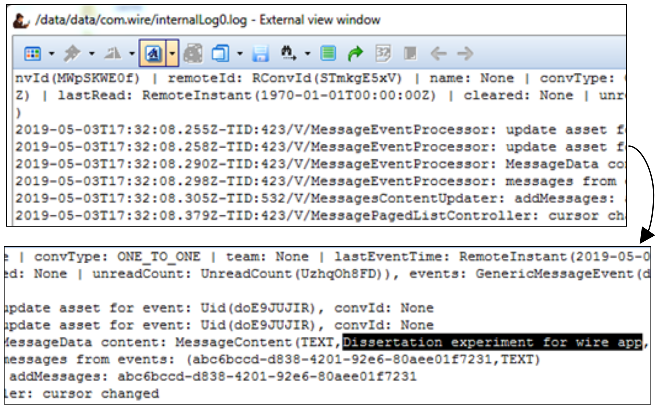
TABLE II. NIST TEST RESULTS (IOS)

Measurements tested	**NIST test assertions applications Were the requirements met? (Y = Yes N = No)**	
	Oxygen Forensic Detective Enterprise	***MOBILedit Forensic Express***
MDT-CA-01	Y	Y
MDT-CA-02	N	Y
MDT-CA-03	N	Y
MDT-CA-04	N	Y
MDT-CA-05	Y	Y
MDT-CA-06	Y	Y
MDT-CA-07	Y	Y
MDT-CA-08	Y	Y
MDT-CA-09	Y	Y

In the fourth test case (MDT-CA-04), where MBOILedit had a success over Oxygen, during data acquisition when connectivity between the mobile and tool was disrupted; a notification was given to alert the user. Both tools could successfully present all supported data elements in useable formats via preview pane or generated report, as required by NIST measurement test id MDT-CA-05. Both tools also reported other test cases, such as reporting equipment related information and hash values for the data objects (MDT-CA-09).

D. *Oxygen Forensic for Android*

The mobile device extracted using Oxygen showed some interesting forensic evidence. A physical acquisition was performed on the device using Oxygen, and it was found that the most relevant pieces of the recovered data were found in Wire, which had records of every single communication stored within a log file by the name of "internalLog0.log" (Figure 9), and a storage cache (Figure 10) for various media files including the image and document files received and the audio message sent, despite those attachments being shown as deleted in application. All three of the identified files could be extracted, and the audio file could be played to hear the original contents of the message.



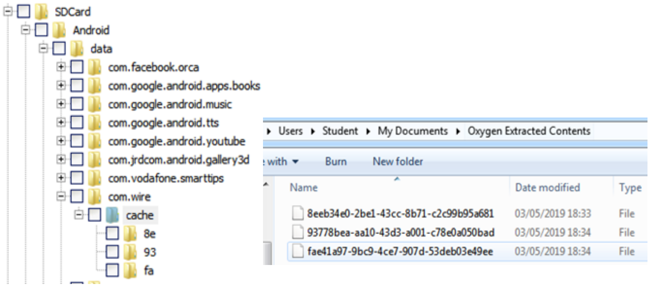
/data/data/com.wire/internalLog0.log - External view window

```
nvId(MWpSKWE0f) | remoteId: RConvId(STmkgE5xV) | name: None | convType: 
Z) | lastRead: RemoteInstant(1970-01-01T00:00:00Z) | cleared: None | unr
)
2019-05-03T17:32:08.255Z-TID:423/V/MessageEventProcessor: update asset f
2019-05-03T17:32:08.258Z-TID:423/V/MessageEventProcessor: update asset f
2019-05-03T17:32:08.290Z-TID:423/V/MessageEventProcessor: MessageData co
2019-05-03T17:32:08.298Z-TID:423/V/MessageEventProcessor: messages from 
2019-05-03T17:32:08.305Z-TID:532/V/MessagesContentUpdater: addMessages: 
2019-05-03T17:32:08.379Z-TID:423/V/MessagePagedListController: cursor ch
```

```
e | convType: ONE_TO_ONE | team: None | lastEventTime: RemoteInstant(2019-05-0
ed: None | unreadCount: UnreadCount(UzhqOh8FD)), events: GenericMessageEvent(d

update asset for event: Uid(doE9JUJIR), convId: None
update asset for event: Uid(doE9JUJIR), convId: None
MessageData content: MessageContent(TEXT,Dissertation experiment for wire app,
messages from events: (abc6bccd-d838-4201-92e6-80aee01f7231,TEXT)
 addMessages: abc6bccd-d838-4201-92e6-80aee01f7231
ler: cursor changed
```

Figure 9. internalLog0.log Wire communications in Oxygen



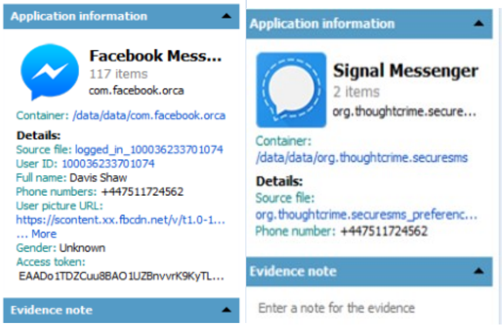
```
⊟ □ SDCard
  ⊟ □ Android
    ⊟ □ data
      ⊞ □ com.facebook.orca
      ⊞ □ com.google.android.apps.books
      ⊞ □ com.google.android.music
      ⊞ □ com.google.android.tts
      ⊞ □ com.google.android.youtube
      ⊞ □ com.jrdcom.android.gallery3d
      ⊞ □ com.vodafone.smarttips
      ⊟ □ com.wire
        ⊟ □ cache
            □ 8e
            □ 93
            □ fa
```

▸ Users ▸ Student ▸ My Documents ▸ Oxygen Extracted Contents

Name	Date modified	Type
8eeb34e0-2be1-43cc-8b71-c2c99b95a681	03/05/2019 18:33	File
93778bea-aa10-43d3-a001-c78e0a050bad	03/05/2019 18:34	File
fae41a97-9bc9-4ce7-907d-53deb03e49ee	03/05/2019 18:34	File

Figure 10. Wire cache and media files in Oxygen

The remaining items of recovered evidence were that of account data, recovered from various log or config-based files within the application data areas of the device storage. This data revealed the username, account ID and mobile number for the registered Facebook Messenger account and the mobile number for the Signal account (Figure 11). Analysis of messenger and Signal program files revealed no data relevant to the conversations undertaken, nor any account information. Keyword search analysis of the image provided few results as shown in Figure 12.



Facebook Mess...
117 items
com.facebook.orca

Container: /data/data/com.facebook.orca

Details:
Source file: logged_in_100036233701074
User ID: 100036233701074
Full name: Davis Shaw
Phone numbers: +447511724562
User picture URL:
https://scontent.xx.fbcdn.net/v/t1.0-1...
... More
Gender: Unknown
Access token:
EAADo1TDZCuu8BAO1UZBnvvrK9KyTL...

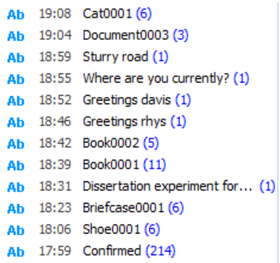
Signal Messenger
2 items
org.thoughtcrime.secure...

Container:
/data/data/org.thoughtcrime.securesms

Details:
Source file:
org.thoughtcrime.securesms_preferenc...
Phone number: +447511724562

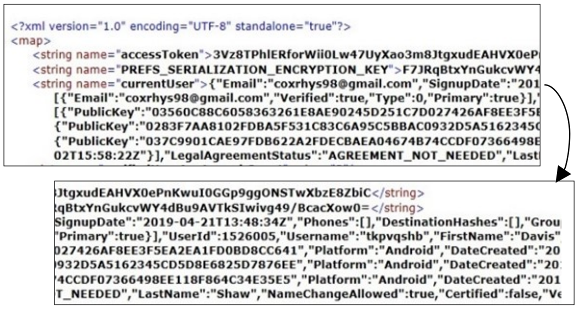
Enter a note for the evidence

Figure 11. Messenger and Signal in Oxygen



```
Ab  19:08  Cat0001 (6)
Ab  19:04  Document0003 (3)
Ab  18:59  Sturry road (1)
Ab  18:55  Where are you currently? (1)
Ab  18:52  Greetings davis (1)
Ab  18:46  Greetings rhys (1)
Ab  18:42  Book0002 (5)
Ab  18:39  Book0001 (11)
Ab  18:31  Dissertation experiment for… (1)
Ab  18:23  Briefcase0001 (6)
Ab  18:06  Shoe0001 (6)
Ab  17:59  Confirmed (214)
```

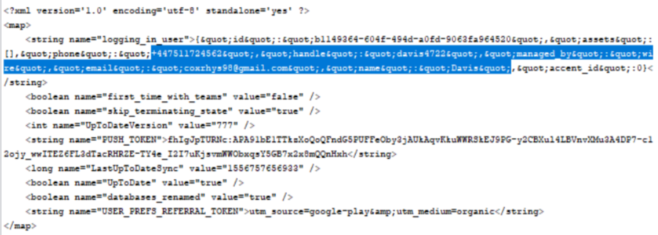
Figure 12. Full Oxygen keyword search results



```
<?xml version="1.0" encoding="UTF-8" standalone="true"?>
<map>
   <string name="accessToken">3Vz8TPhlERforWii0Lw47UyXao3m8JtgxudEAHVX0eP
   <string name="PREFS_SERIALIZATION_ENCRYPTION_KEY">F7JRqBtxYnGukcvWY4
   <string name="currentUser">{"Email":"coxrhys98@gmail.com","SignupDate":"201
      [{"Email":"coxrhys98@gmail.com","Verified":true,"Type":0,"Primary":true}],"
      [{"PublicKey":"03560C88C6058363261E8AE90245D251C7D027426AF8EE3F5E
      {"PublicKey":"0283F7AA8102FDBA5F531C83C6A95C5BBAC0932D5A5162345C
      {"PublicKey":"037C9901CAE97FDB622A2FDECBAEA04674B74CCDF07366498E
      02T15:58:22Z"}],"LegalAgreementStatus":"AGREEMENT_NOT_NEEDED","Last
```

```
JtgxudEAHVX0ePnKwuI0GGp9ggONSTwXbzE8ZbiC</string>
qBtxYnGukcvWY4dBu9AVTkSIwivg49/BcacXow0=</string>
SignupDate":"2019-04-21T13:48:34Z","Phones":[],"DestinationHashes":[],"Grou
"Primary":true}],"UserId":1526005,"Username":"tkpvqshb","FirstName":"Davis"
027426AF8EE3F5EA2EA1FD0BD8CC641","Platform":"Android","DateCreated":"20
932D5A5162345CD5D8E6825D7876EE","Platform":"Android","DateCreated":"20
4CCDF07366498EE118F864C34E35E5","Platform":"Android","DateCreated":"201
T_NEEDED","LastName":"Shaw","NameChangeAllowed":true,"Certified":false,"Ve
```

Figure 13. Confide account data in Oxygen



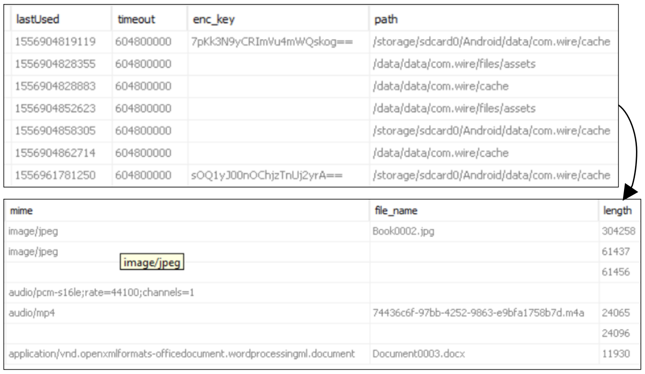
```
<?xml version='1.0' encoding='utf-8' standalone='yes' ?>
<map>
    <string name="logging_in_user">{&quot;id&quot;:&quot;b1149364-604f-494d-a0fd-9063fa964520&quot;,&quot;assets&quot;:
[],&quot;phone&quot;:&quot;+447511724562&quot;,&quot;handle&quot;:&quot;davis4722&quot;,&quot;managed_by&quot;:&quot;wi
re&quot;,&quot;email&quot;:&quot;coxrhys98@gmail.com&quot;,&quot;name&quot;:&quot;Davis&quot;,&quot;accent_id&quot;:0}<
/string>
    <boolean name="first_time_with_teams" value="false" />
    <boolean name="skip_terminating_state" value="true" />
    <int name="UpToDateVersion" value="777" />
    <string name="PUSH_TOKEN">fhIgJpTURNc:APA91bE1TTksXoQoQFndG5PUFFeOby3jAUkAqvKkuWWRSkEJ9PG-y2CBXul4LBVnvXMu3A4DP7-cl
2ojy_wwITEZ6FL3dTacRHRZE-TY4e_I2I7uKjsvmWWObxqsY5GB7x2x8mQQnHxh</string>
    <long name="LastUpToDateSync" value="1556757656933" />
    <boolean name="UpToDate" value="true" />
    <boolean name="databases_renamed" value="true" />
    <string name="USER_PREFS_REFERRAL_TOKEN">utm_source=google-play&amp;utm_medium=organic</string>
</map>
```

Figure 14. Wire account data in Oxygen

Analysis of the Confide program files displayed no data relevant to the conversations but did contain a config file detailing the email registered to the confide account as well as the sign-up date, username and account ID (Figure 13). For

Wire, the username, account ID, mobile number and email address for the registered account were also found (Figure 14).

Analysis of the Wire program files revealed an SQL database named "ZGlobal.db" containing the locations of media files sent/received by the target device within a cache, specifically the jpg image received (Book0002.jpg), the Audio message sent ("Audio test 1") and the document file received (Document0003.doc), as shown in Figure 15.



lastUsed	timeout	enc_key	path
1556904819119	604800000	7pKk3N9yCRImVu4mWQskog==	/storage/sdcard0/Android/data/com.wire/cache
1556904828355	604800000		/data/data/com.wire/files/assets
1556904828883	604800000		/data/data/com.wire/cache
1556904852623	604800000		/data/data/com.wire/files/assets
1556904858305	604800000		/storage/sdcard0/Android/data/com.wire/cache
1556904862714	604800000		/data/data/com.wire/cache
1556961781250	604800000	sOQ1yJ00nOChjzTnUj2yrA==	/storage/sdcard0/Android/data/com.wire/cache

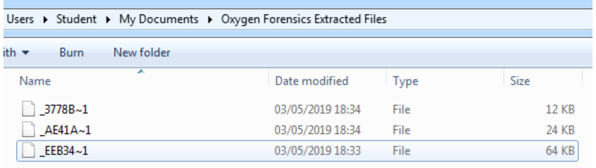
mime	file_name	length
image/jpeg	Book0002.jpg	304258
image/jpeg		61437
		61456
audio/pcm-s16le;rate=44100;channels=1		
audio/mp4	74436c6f-97bb-4252-9863-e9bfa1758b7d.m4a	24065
		24096
application/vnd.openxmlformats-officedocument.wordprocessingml.document	Document0003.docx	11930

Figure 15. Oxygen analysis of ZGlobal.db

Only the images sent by each application were consistently found as they were within device storage. Attempts to extract the media files from the Wire directory using the cache file paths and file names provided by the ZGlobal.db database were successful, and each file could be carved from the image, however both "Book0002.jpg" and "Document0003.doc" were encrypted and could not be opened. The Audio message file on the other hand was unencrypted and once extracted could be played to hear the original message.

E. *Oxygen Anti-forensic for Android*

Upon completing prior testing with the applications installed, all four apps were uninstalled via the Google play store and the device was imaged again for analysis. Both Facebook messenger and Signal were absent from the messengers section of the GUI after uninstallation leaving the account data absent from extraction. The program files for all four applications had also been removed from the file system, however the Wire media cache remained semi intact as recovered data. Searching for the Wire media files by cache name and manually searching for the cache in recovered space did reveal the image and audio message files (Figure 16).

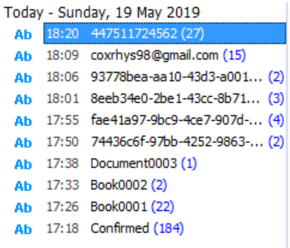


Users ▸ Student ▸ My Documents ▸ Oxygen Forensics Extracted Files
ith ▾
Burn
New folder

Name	Date modified	Type	Size
_3778B~1	03/05/2019 18:34	File	12 KB
_AE41A~1	03/05/2019 18:34	File	24 KB
_EEB34~1	03/05/2019 18:33	File	64 KB

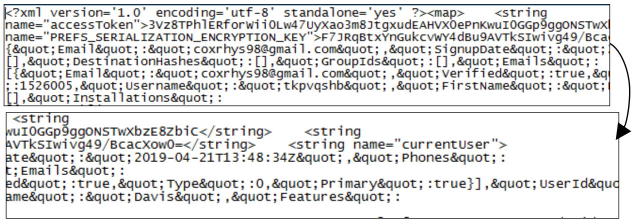
Figure 16. Oxygen extracted deleted Wire media files

Both the identified image and audio files could be extracted, and the audio message could be played to hear its original content. Each application transcript, as well as the email address and mobile number associated with the applications, was then inputted into the search bar, with the results of the search being far less than the prior analysis (Figure 17).



```
Today - Sunday, 19 May 2019
Ab  18:20  447511724562 (27)
Ab  18:09  coxrhys98@gmail.com (15)
Ab  18:06  93778bea-aa10-43d3-a001... (2)
Ab  18:01  8eeb34e0-2be1-43cc-8b71... (3)
Ab  17:55  fae41a97-9bc9-4ce7-907d-... (4)
Ab  17:50  74436c6f-97bb-4252-9863-... (2)
Ab  17:38  Document0003 (1)
Ab  17:33  Book0002 (2)
Ab  17:26  Book0001 (22)
Ab  17:18  Confirmed (184)
```

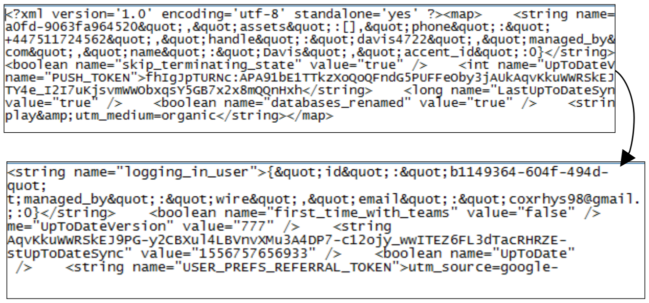
Figure 17. Oxygen Anti-forensics image keyword search result



```
<?xml version='1.0' encoding='utf-8' standalone='yes' ?><map>    <string
name="accessToken">3Vz8TPhlERforWiiOLw47UyXao3m8JtgxudEAHVX0ePnKwuIOGGp9ggONSTwX
name="PREFS_SERIALIZATION_ENCRYPTION_KEY">F7JRqBtxYnGukcvWY4dBu9AVTkSIwivg49/Bca
{&quot;Email&quot;:&quot;coxrhys98@gmail.com&quot;,&quot;SignupDate&quot;:&quot;
[],&quot;DestinationHashes&quot;:[],&quot;GroupIds&quot;:[],&quot;Emails&quot;:
[{&quot;Email&quot;:&quot;coxrhys98@gmail.com&quot;,&quot;Verified&quot;:true,&q
;:1526005,&quot;Username&quot;:&quot;tkpvqshb&quot;,&quot;FirstName&quot;:&quot;
[],&quot;Installations&quot;:
```

```
 <string
wuIOGGp9ggONSTwXbzE8ZbiC</string>    <string
AVTkSIwivg49/BcacXow0=</string>    <string name="currentUser">
ate&quot;:&quot;2019-04-21T13:48:34Z&quot;,&quot;Phones&quot;:
t;Emails&quot;:
ed&quot;:true,&quot;Type&quot;:0,&quot;Primary&quot;:true}],&quot;UserId&quo
ame&quot;:&quot;Davis&quot;,&quot;Features&quot;:
```

Figure 18. Oxygen Anti-forensics image recovered Confide.xml



```
<?xml version='1.0' encoding='utf-8' standalone='yes' ?><map>    <string name=
a0fd-9063fa964520&quot;,&quot;assets&quot;:[],&quot;phone&quot;:&quot;
+447511724562&quot;,&quot;handle&quot;:&quot;davis4722&quot;,&quot;managed_by&
com&quot;,&quot;name&quot;:&quot;Davis&quot;,&quot;accent_id&quot;:0}</string>
<boolean name="skip_terminating_state" value="true" />    <int name="UpToDateV
name="PUSH_TOKEN">fhIgJpTURNc:APA91bE1TTkzXoQoQFndG5PUFFeOby3jAUkAqvKkuWWRSkEJ
TY4e_I2I7uKjsvmWWObxqsY5GB7x2x8mQQnHxh</string>    <long name="LastUpToDateSyn
value="true" />    <boolean name="databases_renamed" value="true" />    <strin
play&amp;utm_medium=organic</string></map>
```

```
<string name="logging_in_user">{&quot;id&quot;:&quot;b1149364-604f-494d-
quot;
t;managed_by&quot;:&quot;wire&quot;,&quot;email&quot;:&quot;coxrhys98@gmail.
;:0}</string>    <boolean name="first_time_with_teams" value="false" />
me="UpToDateVersion" value="777" />    <string
AqvKkuWWRSkEJ9PG-y2CBXul4LBVnvXMu3A4DP7-c12ojy_wWITEZ6FL3dTacRHRZE-
stUpToDateSync" value="1556757656933" />    <boolean name="UpToDate"
 />    <string name="USER_PREFS_REFERRAL_TOKEN">utm_source=google-
```

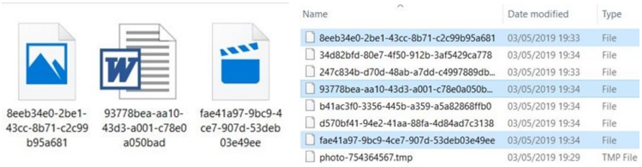
Figure 19. Oxygen Anti-forensics image Wire account data

The Wire messages that had previously appeared within internalLog0.log did not exist, leaving no trace of the text-based communications, however, searches for the mobile number and email address revealed both a recovered copy of the Confide.xml file (Figure 18) and showed a deleted file that appeared to display all the account details for Wire (Figure 19).

F. *Autopsy and FTK for Android*

An Autopsy case file was produced for the Android device and both DD extracted partitions were added as evidence. Analysis of the image files provided similar evidence as Oxygen Forensic Detective: the Confide.xml config file containing the registered email address was discovered, as well as the "ZGlobal.db" database containing cache locations for the Wire media files. Further analysis with a keyword search of the application transcripts also revealed the same data, with the sent media files and entire Wire transcript being identified. Extraction of the media files also proved the same, with both the jpg file and Document file remaining encrypted but the mp4 file remaining audible. The Autopsy analysis differed only in the absence of identified Signal account data and in a lack of Mobile number/Account ID data for Facebook messenger.

FTK Imager was run and both partition images were added as image files for analysis; however, the volume containing the application data stored within mmcblk0.dd was unavailable in analysis. As a result, accessing the "Confide.xml", "ZGlobal.db" and "internalLog0.log" files was impossible. However, partition mmcblk1 was complete and as a result it was possible to access the Wire media cache and extract the media files to the same effect as Oxygen and Autopsy.



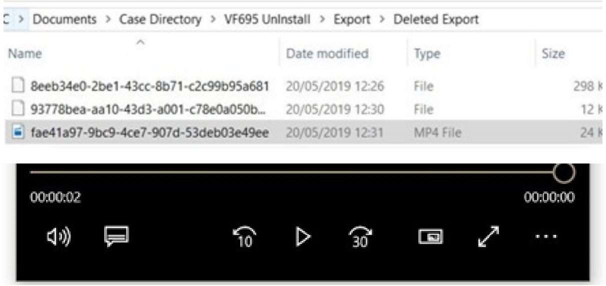
8eeb34e0-2be1-43cc-8b71-c2c99b95a681
93778bea-aa10-43d3-a001-c78e0a050bad
fae41a97-9bc9-4ce7-907d-53deb03e49ee

Name	Date modified	Type
8eeb34e0-2be1-43cc-8b71-c2c99b95a681	03/05/2019 19:33	File
34d82bfd-80e7-4f50-912b-3af5429ca778	03/05/2019 19:34	File
247c834b-d70d-48ab-a7dd-c4997889db...	03/05/2019 19:33	File
93778bea-aa10-43d3-a001-c78e0a050b...	03/05/2019 19:34	File
b41ac3f0-3356-445b-a359-a5a82868ffb0	03/05/2019 19:34	File
d570bf41-94e2-41aa-88fa-4d84ad7c3138	03/05/2019 19:34	File
fae41a97-9bc9-4ce7-907d-53deb03e49ee	03/05/2019 19:34	File
photo-754364567.tmp	03/05/2019 19:29	TMP File

Figure 20. Extracted Wire cache files via Autopsy (Left) and FTK Imager (Right)

G. Autopsy and FTK Anti-forensics Results for Android

Autopsy revealed only slightly fewer results, once again similar to the Oxygen Forensic Detective results. Keyword searches for both the test data transcripts and for the known account details failed to find the "internalLog0.log" file, which had stored the Wire conversations, however it did still manage to find both the deleted "Confide.xml" file and the deleted Wire file containing its account details. Analysis of the Wire cache was also possible, and revealed more deleted records that Oxygen seemed to, enabling the extraction of the media files once again. Both the jpg image file and document files remained unreadable, and the mp4 audio message remained unencrypted and fully audible (Figure 21).



C > Documents > Case Directory > VF695 Uninstall > Export > Deleted Export

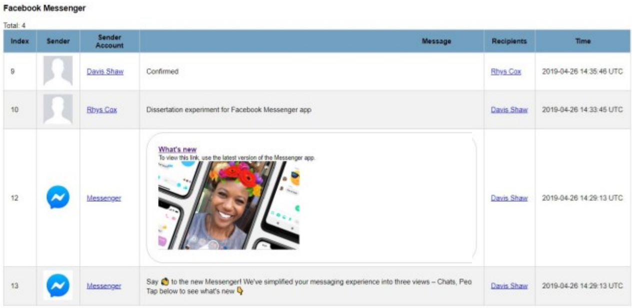
Name	Date modified	Type	Size
8eeb34e0-2be1-43cc-8b71-c2c99b95a681	20/05/2019 12:26	File	298 K
93778bea-aa10-43d3-a001-c78e0a050b...	20/05/2019 12:30	File	12 K
fae41a97-9bc9-4ce7-907d-53deb03e49ee	20/05/2019 12:31	MP4 File	24 K

Figure 21. Autopsy extracted deleted Wire media files

FTK Imager revealed identical results as before, when apps were not deleted. The mmcblk0 partition still appeared partially unreadable, making analysis of partition specific files impossible, but access to the deleted Wire cache was still possible to identify and extract the cache contents. The extracted files behaved as they had before, with all except the mp4 file being encrypted or otherwise unreadable.

H. Andriller Results for Android

The results from Andriller where negligible compared to those in Oxygen, with only account data and Facebook messages being shown in the main report, and no storage data being extracted despite the option being selected before extraction was performed. The account data recovered provided no actual account details, and instead just provided evidence that Facebook messenger and signal were installed, and the Facebook messages extracted were only the unencrypted messages sent between the original Nokia 1 device and the personal device, on top of the account confirmation messages, as shown in Figure 22. Andriller revealed no forensically valuable evidence relevant to the uninstalled applications, account data that was previously extracted was absent and once again was missing shared storage data for manual analysis.



Facebook Messenger

Total: 4

Index	Sender	Sender Account	Message	Recipients	Time
9		Davis Shaw	Confirmed	Rhys Cox	2019-04-26 14:35:46 UTC
10		Rhys Cox	Dissertation experiment for Facebook Messenger app	Davis Shaw	2019-04-26 14:33:45 UTC
12		Messenger	What's new To view this link, use the latest version of the Messenger app.	Davis Shaw	2019-04-26 14:29:13 UTC
13		Messenger	Say 👋 to the new Messenger! We've simplified your messaging experience into three views – Chats, Peo Tap below to see what's new 👇	Davis Shaw	2019-04-26 14:29:13 UTC

Figure 22. Andriller Facebook messages extraction

I. NIST Measurements for tools used for Android

As the baseline test assertions, MDT-CA-1 to 10 are the lowest levels of functionality that NIST determined a mobile forensics analysis tool should have, Oxygen Forensic Detective managed to meet all of the test assertions except MDT-CA-10 (Table III). However, some of the assertions in the other tools such as FTK, Autopsy and Andriller were not relevant and therefore could not be tested.

TABLE III. ANDROID PROPRIETARY TOOLS

NIST Test Guidelines: Oxygen Vs. Andriller			
NIST Base Guidelines		Oxygen Forensic Detective	Andriller 3.0.3
MDT-CA-01		Pass	N/A
MDT-CA-02		Pass	N/A
MDT-CA-03		Pass	Pass
MDT-CA-04		Pass	Fail
MDT-CA-05		Pass	Pass
MDT-CA-06		Pass	Pass
MDT-CA-07		Pass	Pass
MDT-CA-08		Pass	Pass
MDT-CA-09		Pass (Inconsistently)	Fail
MDT-CA-10		N/A	N/A

TABLE IV. ANDROID FREEWARE TOOLS

NIST Test Assertions: FTK Imager Vs. Autopsy			
NIST Base Guidelines		FTK Imager 4.1.1.1	Autopsy 4.8.0
MDT-CA-01		N/A	N/A
MDT-CA-02		N/A	N/A
MDT-CA-03		N/A	N/A
MDT-CA-04		N/A	N/A
MDT-CA-05		Pass	Pass
MDT-CA-06		Fail	Fail
MDT-CA-07		Fail	Pass
MDT-CA-08		Pass	Pass
MDT-CA-09		Fail	Pass
MDT-CA-10		N/A	Pass

As shown in Table IV, being purely analysis tools, both Autopsy and FTK Imager were unable to be assessed by MDT-CA-1 to 4 by default. Andriller failed two of the seven applicable assertions, MDT-CA-4 & MDT-CA-9; Autopsy failed one of the six applicable assertions, MDT-CA-6, and FTK Imager failed three of the five applicable assertions, MDT-CA-6/7/9. Considering both the failed assertions, and the assertions that could not be applied due to a lack of tool functionality, Oxygen Forensic Detective is by far the most reliable by the standards set by NIST, with Andriller second, Autopsy third and FTK Imager fourth.

J. Comparison of tools for iOS

For iOS, both tools used in the mobile investigation output slightly different results. While neither recovered messages from the EMAs tested, both of them recovered artefacts elsewhere. Oxygen and MOBILedit successfully recovered data on all applications: Snapchat, Cyberdust and Confide. While different artefacts and data were detected, the fact that no physical copies of messages were recovered in any application, using either of the forensic tools, proves how efficient EMAs are at protecting user privacy. Oxygen detected offensive words being sent/received, this would be useful within a cyberbullying case, even though the message itself was not recovered. The evidence detected of communication between the mobile user and another contact would also prove useful as the application would be able to tell detectives who the mobile user had been in contact with. This would also be useful in a cyberbullying case, as there would be evidence the 'bully' had contact with the victim.

Furthermore, the detection of Cloud messaging within Cyberdust suggested that although physical messages were not recovered within the application, the messages could have been uploaded elsewhere to a Cloud network and access could be gained through the network. This would provide a chance for messages to perhaps be recovered in a cyberbullying case.

For Confide, Oxygen displays the password in encrypted format, while the MOBILedit shows it in unencrypted format. MOBILedit also recovered an unencrypted version of the registered mobile number, which Oxygen could not. For the Snapchat, MOBILedit detected account data, such as the mobile user's username and the contact list within the application. However, MOBILedit failed to detect other evidences, such as offensive words, evidence of communication between the mobile user and another contact, and the evidence of a message being deleted.

K. Comparison of tools for Android

The application analysis performed revealed that, for the most part, the EMAs are secure enough to keep evidence of user activity and message contents from being identified. Considering the successfully identified/extracted data, the NIST assessments and the overall forensic soundness of the tools and reliant imaging techniques therein, in the case of FTK and Autopsy, Oxygen Forensic Detective appears to be the most capable and reliable tool of the four, able to both non-invasively image suspect devices and analyse the extracted images in detail up to the relevant baseline specifications set by NIST. Furthermore, the evidence analysis shown by Oxygen was rivalled only by Autopsy, which while impressive for a fully freeware tool still required a pre-created image in order to perform analysis. The second freeware analysis tool, FTK Imager, was lacking in its analysis due to an inability to properly analyse the mmcblk0 partition, which contained the majority of the identifiable evidence. As a result, use of FTK Imager as a backup to proprietary tools would be ill advised when Autopsy is far more accessible as an immediate download, instead of FTK Imager's request-based download, and provides more analysis functionality. While not entirely limited to DD images for analysis, without a prior image being obtained through a dedicated imaging tool both FTK Imager and Autopsy would be reliant on the invasive and potentially forensically unsound technique of rooting and DD extracting a device image, which potentially justifiable in court given the right situation still carries great risk of being thrown out as compromised evidence. Despite the potentially evidence unsafe methods required by the freeware tools, both FTK Imager and Autopsy provided more forensically valuable data than Andriller, which did not extract any filesystem data required for the in-depth analysis.

V. CONCLUSION

In this paper, experiments were performed to assess the forensically valuable artefacts that could be recovered from EMAs using various proprietary and freeware tools. The results show that with the rooted Android phone, more artefacts were recovered compared to iOS phone, which was not jailbroken. On iOS platform, no full ephemeral messages were recovered with either of the tools, but other significant artefacts were found, which proved rather interesting to the investigation. One significant finding was that of the Snapchat's 'offensive words' detection, which may help aid evidence in cyberbullying cases to prove inappropriate language may have been used towards a victim. In forensic investigations, the investigators have to look very deep into the data and have a lot of patience, as one small piece of evidence could change the case, such as the offensive word. For iOS, a physical acquisition may have provided a much more thorough investigation to recover deleted data.

The forensic analysis conducted on the Android device also did not recover full ephemeral communications on the applications examined, except for the application 'Wire'. A log file was recovered containing full vocal communication sent and received on the application. Facebook Messenger was acknowledged as an application, and some details of the user were also stored, however no evidence of communication was found. This was interesting, as Snapchat (which is of similar popularity to Facebook Messenger) managed to recover some evidence of communication between two users using a logical acquisition on Oxygen, but it seems Oxygen could not find such communication on Facebook Messenger. This could contradict the fact physical acquisitions are supposed to recover more information. However, it could be the way the application is designed in itself. It does appear that Facebook Messenger has a more secure design, in which messages cannot be recovered even through a physical acquisition.

During experimentation of Android device, automatic tool analysis and analysis of application files revealed that from all four EMAs varying amounts of account data were recovered, of which Confide and Wire provided the most valuable data, then Facebook messenger and then Signal with the least. From this, a moderate range of recoverable forensic evidence has been identified for the four chosen EMAs, displaying where they may be recovered from and what data the evidence specifically relates to. During the anti-forensic investigation, when apps were deleted from the Android phone, some valuable artefacts were recovered. For example, the media files in Wire could still be recovered but the log file was not.

Furthermore, the use of the proprietary and freeware forensic tools, combined with the NIST assessments, provides insight into the capabilities and level of professional functionality that each tool holds, allowing for greater understanding of the available tools in Android and iOS based mobile device analysis and what these tools can do with regard to the extraction and analysis of ephemeral data. In total this study fills the gaps of knowledge that resides in the analysis of both popular EMAs and analysis of those applications via freeware forensic tools to the standard proprietary options. Further research on this topic should focus on better filling the gaps of knowledge regarding the recovery of ephemeral communication data from applications not included in this study, or further research on the applications used within this study to identify if app specific decryption capabilities could assist in identifying ephemeral communications from the applications that did not yield communication evidence.

References

[1] A. Chamberlain and M.A.H.B. Azhar, "Comparisons of Forensic Tools to Recover Ephemeral Data from iOS Apps Used for Cyberbullying", The Fourth International Conference on Cyber-Technologies and Cyber-Systems, CYBER 2019, Porto, Portugal.

[2] R. Graham, "How Terrorists Use Encryption", Combating Terrorism Center at West Point. Available from: https://ctc.usma.edu/how-terrorists-use-encryption/ [Accessed: 01- June- 2020].

[3] C. Cotta, A.J. Fernandez-Lelva, F. Fernandez de Vega and F. Chavez, "Application Areas of Ephemeral Computing: A Survey", in Transactions on Computational Collective Intelligence: David Camacho, University of Malaga, pp. 155-157, 2016.

[4] I. Barker, "Cyber criminals turn to messaging apps following dark web crackdown", Betanews, 2017. [Online]. Available from: https://betanews.com/2017/10/25/criminals-turn-to-messaging/ [Accessed: 01- June- 2020].

[5] T. Alyaha and F. Kausar, "Snapchat Analysis to Discover Digital Forensic Artefacts on Android Smartphone", in 8th International Conference on Ambient Systems, Networks and Technologies, ANT-2017 and the 7th International Conference on Sustainable Energy Information Technology, SEIT 2017, 16-19 May 2017, Madeira, Portugal, pp. 1035-1040, 2017.

[6] GSMA, "Number of Mobile Subscribers Worldwide Hits 5 Billion", [Online]. Available from: https://www.gsma.com/newsroom/press-release/number-mobile-subscribers-worldwide-hits-5-billion/ [Accessed: 01- June- 2020].

[7] D. L. Fisher, M.J. Hamilton and J.K. Southwick, "When Electronic Records Disappear But Legal Issues Linger", Law360, Portfolio Media, Inc., Available from: https://www.pepperlaw.com/publications/when-electronic-records-disappear-but-legal-issues-linger-2018-09-06/ [Accessed: 01- June- 2020].

[8] J. Graham, "WhatsApp, Wickr Seen by Justice Dept. as Tools to Erase Evidence", Available from: https://biglawbusiness.com/whatsapp-wickr-seen-by-justice-dept-as-tools-to-erase-evidence [Accessed: 01- June- 2020].

[9] J. Constine, "Snapchat revives growth in Q1 beat with 190M users", Available from: https://techcrunch.com/2019/04/23/snapchat-q1-2019-earnings/ [Accessed: 01- June- 2020].

[10] D. Noyes, "The Top 20 Valuable Facebook Statistics", Available from: https://zephoria.com/top-15-valuable-facebook-statistics/ [Accessed: 01- June- 2020].

[11] National Institute of Standards and Technology, "Mobile Device Tool Test Assertions and Test Plan", 2016. [Online]. Available from: https://www.nist.gov/system/files/documents/2017/05/09/mobile_device_tool_test_assertions_and_test_plan_v2.0.pdf [Accessed: 01- June- 2020].

[12] K. M. Ovens and G. Morison, "Forensic analysis of kik messenger on ios devices", Digital Investigation, vol. 17, pp. 40-52, 2016.

[13] S. C. Sathe and N. M. Dongre, "Data acquisition techniques in mobile forensics", in 2018 2nd International Conference on Inventive Systems and Control (ICISC), pp. 280–286. doi: 10.1109/ICISC.2018.8399079.

[14] M. A. H. B. Azhar and T. Barton, "Forensic Analysis of Secure Ephemeral Messaging Applications on Android Platforms", Jan. 2017, doi: 10.1007/978-3-319-51064-4.

[15] M. Al-Hadadi and A. AlShidhani, "Smartphone Forensics Analysis: A Case Study", International Journal of Computer and Electrical Engineering, vol. 5, pp. 577-579, 2013.

[16] R. Umar, I. Riadi and G. Zamroni, "Mobile Forensic Tools Evaluation for Digital Crime Investigation", International Journal on Advanced Science, Engineering and Information Technology, vol. 8, pp. 949-955, 2018.

[17] P. Naughton and M. A. H. B. Azhar, "An Investigation on Forensic Opportunities to Recover Evidential Data from Mobile Phones and Personal Computers. The Second International Conference on Cyber-Technologies and Cyber-Systems", CYBER 2017, Barcelona, Spain.

[18] ACPO, "ACPO Good Practice Guide for Digital Evidence", 2012. [Online]. Available from: https://www.digital-detective.net/digital-forensics-

documents/ACPO_Good_Practice_Guide_for_Digital_Evidence_v5.pdf [Accessed: 01- June- 2020].

[19] iPhone 6s, “Wikipedia for iPhone 6s”, [Online]. Available from: https://en.wikipedia.org/wiki/IPhone_6S [Accessed: 01- June- 2020].

[20] Vodafone VF695, “User manual of Vodafone VF695”, [Online]. Available from: https://www.vodafone.com/content/dam/vodcom/devices/smart-first/User%20Manual%20-%20English.pdf [Accessed: 01- June- 2020].

[21] Jkielty, “Android v iOS market share”, 2019, DeviceAtlas, [Online]. Available at: https://deviceatlas.com/blog/android-v-ios-market-share [Accessed: 01- June- 2020].

[22] Snapchat, “Snapchat APP for mobile”, [Online]. Available from: https://www.snapchat.com/l/en-gb/ [Accessed: 01- June- 2020].

[23] Dust, “The APP that protects your assests”, [Online]. Available from: https://usedust.com/ [Accessed: 01- June- 2020].

[24] Confide, “Your Confidential Messenger”, [Online]. Available from: https://getconfide.com/ [Accessed: 01- June- 2020].

[25] Facebook Messenger, “Wikipedia for Facebook Messenger”, [Online]. Available https://en.wikipedia.org/wiki/Facebook_Messenger [Accessed: 01- June- 2020].

[26] Signal Messenger, “Wikipedia for Signal Messenger”, [Online]. Available https://en.wikipedia.org/wiki/Signal_Messenger [Accessed: 01- June- 2020].

[27] Wire App, “Wikipedia for Wire App”, [Online]. Available https://en.wikipedia.org/wiki/Wire_(software) [Accessed: 01- June- 2020].

[28] Omnicore , “Snapchat by the Numbers: Stats, Demographics & Fun Facts”, 2020. [Online]. Available from: https://www.omnicoreagency.com/snapchat-statistics/ [Accessed: 01- June- 2020].

[29] Messenger, “Messenger - Android Apps on Google Play”, [Online], Available at: https://play.google.com/store/apps/details?id=com.facebook.orca [Accessed: 01- June- 2020].

[30] J. Evans, “WhatsApp Partners With Open WhisperSystems To End-To-End Encrypt Billions Of Messages A Day.” [Online]. Available from https://techcrunch.com/2014/11/18/end-to-end-for-everyone/ [Accessed: 01- June- 2020].

[31] Oxygen Forensics, Oxygen Forensic Detective Enterprise, [Online]. Available from: https://www.oxygen-forensic.com/en/products/oxygen-forensic-detective-enterprise [Accessed: 01- June- 2020].

[32] MOBILedit Forensic, MOBILedit Forensic Express, [Online]. Available from: https://www.mobiledit.com/online-store/forensic-express [Accessed: 01- June- 2020].

[33] Andriller, Android Forensic Tools, [Online]. Available from: https://www.andriller.com/ [Accessed: 01- June- 2020].

[34] FTK Imager, AccessData. [Oniline], Available from: https://accessdata.com/product-download [Accessed: 01- June- 2020].

[35] Autopsy. [Online], Available from: https://www.sleuthkit.org/autopsy/ [Accessed: 01- June- 2020].

[36] Andrioid Tools, “Android Forensics: imaging android filesystem using ADB and DD”, [Online], Available from: https://www.andreafortuna.org/2018/12/03/android-forensics-imaging-android-file-system-using-adb-and-dd/ [Accessed: 01- June- 2020].

[37] M. Lohrum, “Live imaging an Android device”, [Online] Available from: http://freeandroidforensics.blogspot.com/2014/08/live-imaging-android-device.html [Accessed: 01- June- 2020].

[38] FireBase Messaging, “Firebase Cloud Messaging”, [Online]. Available from: https://firebase.google.com/docs/cloud-messaging [Accessed: 01- June- 2020].

Enhancing the Resilience of Cyber-Physical Systems by Protecting the Physical-World Interface

Rainer Falk, Steffen Fries
Corporate Technology
Siemens AG
Munich, Germany
e-mail: {rainer.falk|steffen.fries}@siemens.com

***Abstract*—Cyber physical systems operate and supervise physical, technical systems using information and communication technology, also called Operation Technology (OT). Cyber security solutions focus on the OT part, i.e., on the information and communication technology. The focus of cyber security is protection, detection, and respondence to cyber attacks. Cyber resilience aims at delivering an intended outcome despite attacks and adverse cyber events and even failures not directly caused by attacks. Protecting the link between the control systems and the physical world has been addressed only in some very specific cases, e.g., charging of electric vehicles. We propose a physical-world firewall that limits the impact on the physical world of a successful attack of automation systems, thereby enhancing the resilience of cyber-physical system against successful attacks against software-based functionality of its OT systems.**

Keywords—cyber security; cyber resilience; system integrity; cyber physical systems; industrial automation and control system; Internet of Things.

I. Introduction

The common focus of IT security relates to IT-based control equipment and data communication, using e.g., Ethernet, wireless LAN (WLAN), and Internet protocol (IP) communication. In addition to this, in OT systems, also the field level comprising sensors and actuators connected to the Operation Technology (OT) automation and control system has to be considered down to the interface between the control system and the physical world via sensors and actuators.

Traditionally, IT security has been focusing on information security, protecting confidentiality, integrity, and availability of data at rest and data in transit, and sometimes also protecting data in use by confidential computation. In Cyber-Physical Systems (CPS), major protection goals are availability, meaning that automation systems stay productive, and system integrity, ensuring that it is operating as intended. Typical application domains are factory automation, process automation, building automation, railway signaling systems, and power system management. Cyber security is covering different phases during operation as there are protect, detect, and react: Protecting against threats, detecting when an attack has occurred, and recovering from attacks.

We see resilience of cyber-physical systems as an important further protection goal, to limit the effect of potential successful attacks on a cyber-physical system in the physical world. In addition, resilience also addresses system stability to cope with failure scenarios not caused by a successful attack. It can be rather seen as a strategy than a specific technology. Our objective is to increase the robustness with respect to intentional attacks, although resilience in general would consider also accidental failures. This paper, being an extended version of [1], puts the focus on the interface between the OT system, i.e., the automation and control system, and the physical world, proposing an additional layer of defense for cyber physical systems. It can be considered as “physical world firewall”, limiting the access to the physical world by the OT system.

After giving an overview on cyber physical systems and on industrial cyber security in Sections II and III, a new approach on protecting the interface of a CPS between the cyber-world and the physical world is described in Section IV. It is a concept to increase the resilience of a CPS when being under attack. Aspects to evaluate the new approach are discussed in Section V. Section VI concludes the paper.

II. Cyber Physical Systems

A cyber-physical system, e.g., an industrial automation and control system, monitors and controls a technical system. Examples are process automation, machine control, energy automation, and cloud robotics. Automation control equipment with sensors (S) and actuators (A) is connected directly with automation components, or via remote input/output modules. The technical process is controlled by measuring its current state using the sensors, and by determining the corresponding actuator signals.

Figure 1 shows an example of an industrial automation and control system, comprising different control networks connected to a plant network and a cloud backend system. Separation of the network is typically used to realize distinct control networks with strict real-time requirements for the interaction between sensors and actuators of a production cell, or to enforce a specific security policy within a production cell.

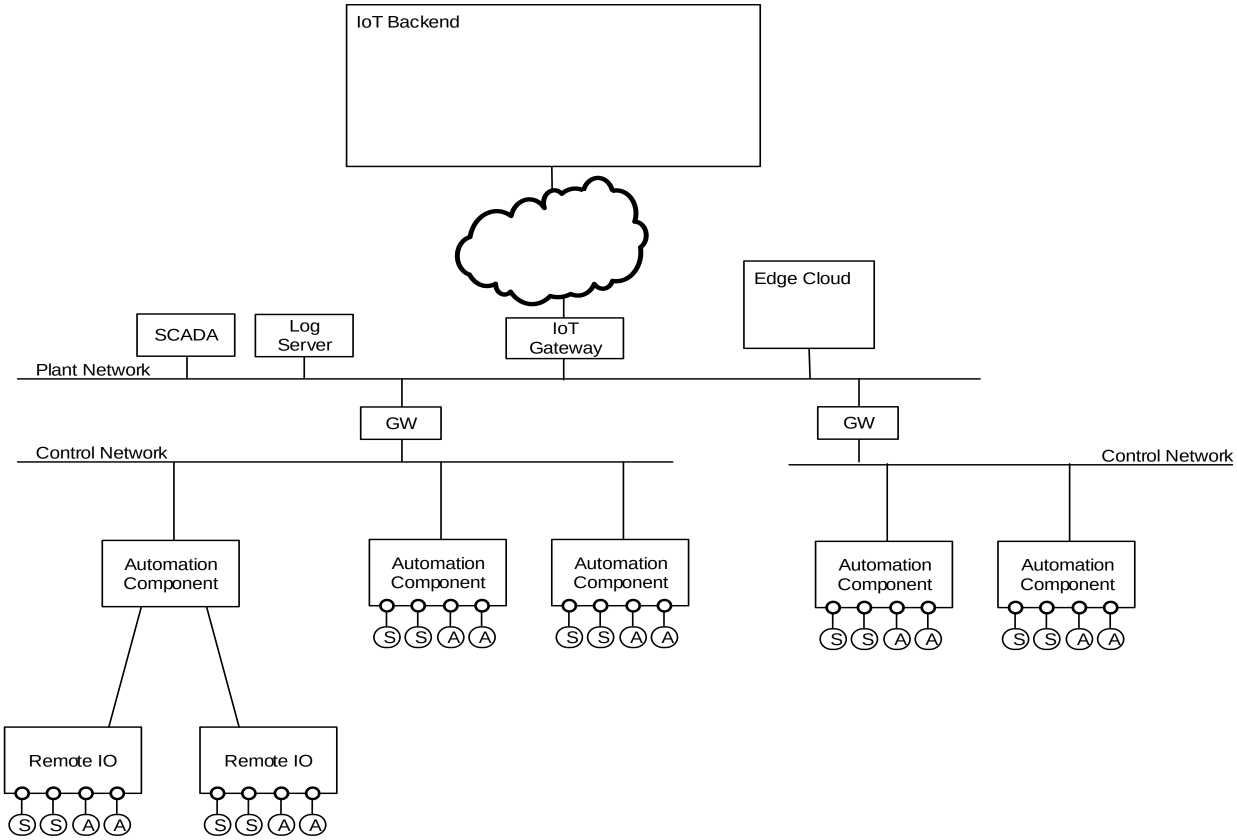


Figure 1. Example CPS System

Such an industrial automation and control system is an example of a cyber-physical system and are utilized in various automation domains, including discrete automation (factory automation), process automation, railway automation, energy automation, and building automation.

Figure 2 shows the typical structure of automation components. The functionality realized by an automation component is largely defined by the firmware/software and the configuration data stored in its flash memory.

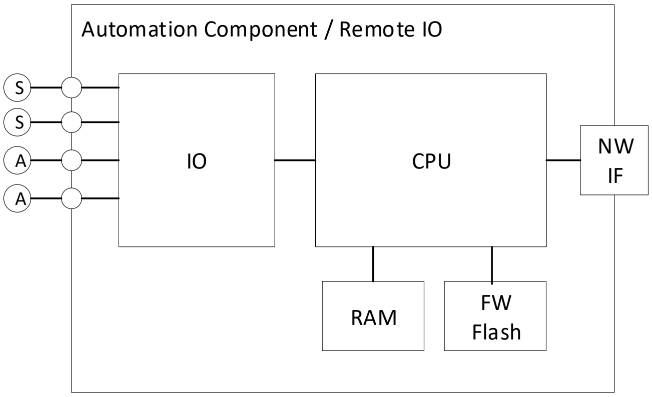


Figure 2. Automation Component

In practice, it has to be assumed that each software component may comprise vulnerabilities, independent of the effort spend to ensure high software quality. This is a reason why automation systems are usually organized in separate security zones. Network traffic can be filtered using network firewalls between different zones, limiting the impact of an impact in one security zone on other connected security zones. In addition, it is often not possible to fix known vulnerabilities immediately by installing a software update, as updates have to be tested thoroughly in a test system before being installed in an operational system, and as an installation is often possible only during a scheduled maintenance window. Also, the priorities of security objectives in different security zones are often different.

In cyber physical systems, the impact of a vulnerability in the OT system may not only affect data and data processing as in classical IT, but it may have an effect also on the physical world. For example, production equipment could be damaged, or the physical process may operate outside the designed physical boundaries, so that the produced goods may not have the expected quality.

III. Industrial Cyber Security

Protecting industrial automation control systems against intentional attacks is increasingly demanded by operators to ensure a reliable operation, and meanwhile also by regulation.

This section gives an overview on industrial security, and on the main relevant industrial security standard IEC 62443 [11] and integrity security requirements.

A. *Industrial CPS Security Requirements*

Industrial security is called also Operation Technology security (OT security), to distinguish it from general Information Technology (IT) security. Industrial systems have not only different security requirements compared to general IT systems, but come also with specific side conditions that prevent that security concepts established in the IT domain can be directly applied in an OT environment. For example, availability and integrity of an automation system often have a higher priority than confidentiality. As an example, high availability requirements, different organization processes (e.g., yearly maintenance windows), and required certifications may prevent the immediate installations of updates.

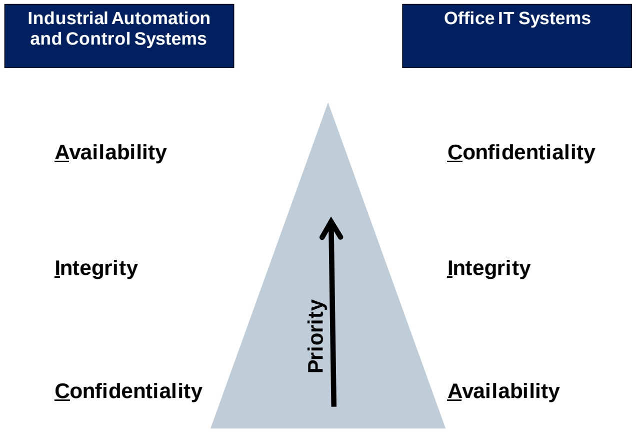


Figure 3. The CIA Pyramid [9]

The three basic security requirements are confidentiality, integrity, and availability. They are also named “CIA” requirements. Figure 3 shows that in common IT systems, the priority is “CIA”. However, in automation systems or industrial IT, the priorities are commonly just the other way around: Availability has typically the highest priority, followed by integrity. Confidentiality is often no strong requirement for control communication, but may be needed to protect critical business know-how. As shown graphically, the CIA pyramid is inverted (turned upside down) in many automation systems.

Specific requirements and side conditions of industrial automation systems like high availability, planned configuration (engineering info), long life cycles, unattended operation, real-time operation, and communication, as well as safety requirements have to be considered when designing a security solution. Note that safety addresses undesired impacts originating from a technical system to the environment, e.g., in the case of a malfunction, while security addresses intentional attacks on the technical system. Often, an important aspect is that the applied security measures do not put availability and integrity of the automation system at risk. Depending on the considered industry (vertical), they may also be part of the critical infrastructure domain, for which security requirements are also imposed for instance by the European Network and Information Systems (NIS) directive [10] or country specific realizations of the directive. Further security requirements are provided by applying standards defining functional requirements, for instance defined in IEC 62443. The defined security requirements can be mapped to different automation domains, including energy automation, railway automation, building automation, process automation.

Security measures to address these requirements range from security processes, personal and physical security, device security, network security, and application security. No single security technology alone is adequate, but a combination of security measures addressing prevention, detection, and reaction to incidents is required (“defense in depth”).

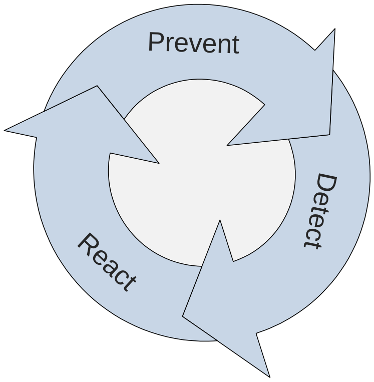
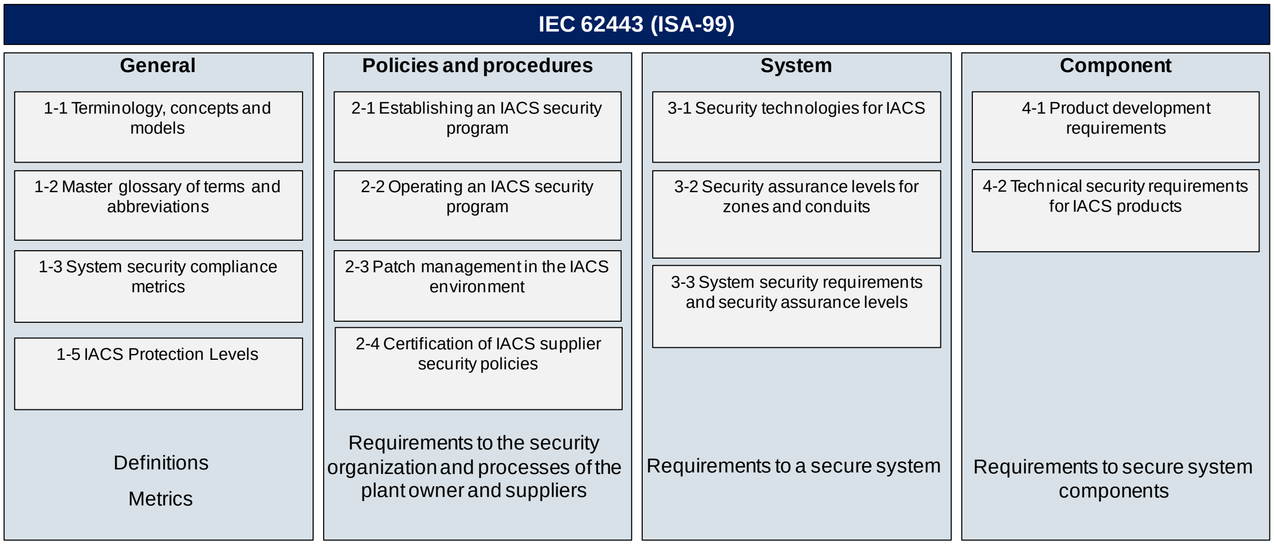


Figure 4. Prevent Detect React Cycle

Also, overall security has to address the areas prevent, detect, and react, see Figure 4. It is not sufficient to only define measures to protect against attacks. The capability has also foreseen to detect attacks, and to define measures to react adequately once an attack has been detected. The physical world firewall described in this paper is targeting the “react” phase, limiting the impact of a successful attack.

B. *Overview IEC 62443 Industrial Security Standard*

The international industrial security standard IEC 62443 [11] is a security requirements framework defined by the International Electrotechnical Commission (IEC). It addresses the need to design cybersecurity robustness and resilience into industrial automation and control systems, covering both organizational and technical aspects of security over the life cycle. Specific parts of this framework are applied successfully in different automation domains, including factory and process automation, railway automation, energy automation, and building automation. The standard specifies security for industrial automation and control systems (IACS) and covers both, organizational and technical aspects of security. Specifically addressed for the industrial domain is



IEC 62443 (ISA-99)

General	Policies and procedures	System	Component
1-1 Terminology, concepts and models	2-1 Establishing an IACS security program	3-1 Security technologies for IACS	4-1 Product development requirements
1-2 Master glossary of terms and abbreviations	2-2 Operating an IACS security program	3-2 Security assurance levels for zones and conduits	4-2 Technical security requirements for IACS products
1-3 System security compliance metrics	2-3 Patch management in the IACS environment	3-3 System security requirements and security assurance levels	
1-5 IACS Protection Levels	2-4 Certification of IACS supplier security policies		
Definitions Metrics	Requirements to the security organization and processes of the plant owner and suppliers	Requirements to a secure system	Requirements to secure system components

Figure 5. IEC 62443 Industrial Security Standard – Overview

the setup of a security organization and the definition of security processes as part of an information security management system (ISMS) based on already existing standards like ISO 27001 [12] or the NIST cyber security framework. Furthermore, technical security requirements are specified distinguishing different security levels for industrial automation and control systems, and also for the used components. The standard has been created to address the specific requirements of industrial automation and control systems.

As shown in Figure 5, different parts of the IEC62443 standard are grouped into four clusters, covering:

- common definitions and metrics;
- requirements on setup of a security organization (ISMS related, comparable to ISO 27001 [12]), as well as solution supplier and service provider processes;
- technical requirements and methodology for security on system-wide level, and
- requirements on the secure development lifecycle of system components, and security requirements to such components at a technical level.

The framework parts address different roles over different phases of the (system) lifecycle: The operator of an automation system operates the automation and control system that has been integrated by the system integrator, using components of product suppliers. In the set of corresponding documents, security requirements are defined, which target the solution operator and the integrator but also the product manufacturer.

According to the methodology described in IEC 62443 part 3-2, a complex automation system is structured into security zones that are connected by and communicate through so-called “conduits” that map for example to the logical network protocol communication between two security zones, see Figure 6.

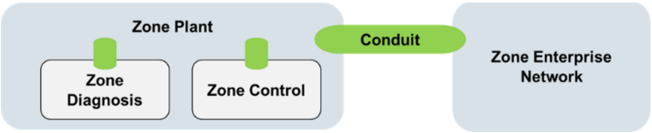


Figure 6. Zones and Conduits

Moreover, this document defines Security Levels (SL) that correlate with the strength of a potential adversary as shown in Figure 7. To achieve a dedicated SL, the defined requirements have to be fulfilled. IEC 62443 part 3-3 defines system security requirements. It supports focusing only on certain facets of security. The security requirements defined by IEC 62443 helps to ensure that all relevant aspects are addressed.

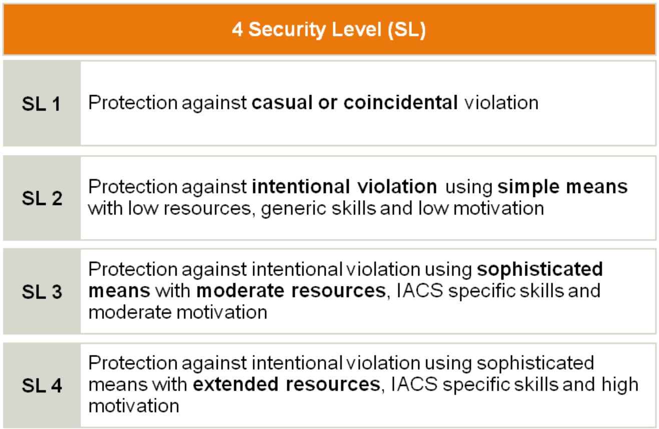
Part 3-3 of IEC 62443 [14], addressing an overall automation system, is in particular relevant for the system integrator. It defines seven foundational requirements that group specific requirements of a certain category:

- FR 1 Identification and authentication control
- FR 2 Use control
- FR 3 System integrity
- FR 4 Data confidentiality
- FR 5 Restricted data flow
- FR 6 Timely response to events
- FR 7 Resource availability

For each of the foundational requirements, there exist several concrete technical security requirements (SR) and requirement enhancements (RE) to address a specific security level. In the context of communication security, these security levels are specifically interesting for the conduits connecting different zones. Related security requirements are defined for the components of an industrial automation and control system in IEC 62443 part 4-2 [15], addressing in particular component manufacturers. The definition of security requirements distinguishes different categories of components, which are “software application”, “embedded device”, “host device”, and “network device”.

Four Security Levels (SL1, SL2, SL3, SL4) are defined that correlate with the strength of a potential attacker as shown in Figure 7. The targeted security level of a zone of the industrial automation and control system is determined based on the identified risk. This allows to tailor the security requirements to the specific needs of an industrial automation and control system.

To reach a dedicated security level, the System Requirements (SR) and potential Requirement Enhancements (RE) defined for that security level have to be fulfilled. The standard foresees that a security requirement can be addressed either directly, or by a compensating countermeasure.



4 Security Level (SL)	
SL 1	Protection against **casual or coincidental** violation
SL 2	Protection against **intentional violation** using **simple means** with low resources, generic skills and low motivation
SL 3	Protection against intentional violation using **sophisticated means** with **moderate resources**, IACS specific skills and moderate motivation
SL 4	Protection against intentional violation using sophisticated means with **extended resources**, IACS specific skills and high motivation

Figure 7. IEC 62443 defined Security Level [9]

The concept of compensating countermeasures allows to reach a certain security level even if some requirements cannot be implemented directly, e.g., as some components do not support the required technical features. This approach is in particular important for existing industrial automation and control systems, so called “brown-field installations”, as existing equipment can be continued to be used.

The security level of a zone or a conduit (a conduit connects zones) is more precisely a security level vector with seven elements. The elements of the vector designate the security level for each foundational requirement. This allows defining the security level specific for each foundational requirement. If, e.g., confidentiality is no security objective within a zone, the security level element corresponding to FR4 “Data confidentiality” can be defined to be SL1 or even none, although SL3 may be required for other foundational requirements (e.g., for FR1, FR2, and FR3). So, the resulting security level vector for a zone could be SL=(3,3,3,1,2,1,3) or SL=(2,2,2,0,1,1,0). The seven elements of the SL-vector correspond to the seven foundational requirements, so that the security level $SL_{FR(i)}$ can be defined separately for each foundational requirement FR(i), i.e., SL = (sl_{FR1}, sl_{FR2}, sl_{FR3}, sl_{FR4}, sl_{FR5}, sl_{FR6}, sl_{FR7}).

Different types of SL vectors are distinguished, depending on the purpose:

– SL-T: A target security level vector is defined by the IACS operator based on his risk assessment, defining which security level shall be achieved by each zone and conduit.

– SL-A: The achieved security level vector designates the current status, i.e., the security level that is actually achieved by each zone and conduit. In particular for brown-field installations, it is common that a targeted security level cannot be set-up immediately. The gap between the targeted and the actually achieved security level can be made transparent.

– SL-C: The security level capability describes the reachable security level a component is capable of, if properly configured, without additional compensating counter measures employed. This also means that depending on the SL-T not all security features of a component may be used in certain installations.

C. *IEC 62443 Integrity Requirements*

One of the seven foundational security requirements defined in Part 3-3 of IEC 62443 [14], targets specifically integrity.

Integrity requirements cover the following areas:

– Overall system integrity

– Communication integrity

– Device integrity

The following examples from IEC 62443-3-3 [14] illustrate some of the integrity-related requirements:

– FR3, SR3.1 Communication integrity: “The control system shall provide the capability to protect the integrity of transmitted information”.

– FR3, SR3.4 Software and information integrity: “The control system shall provide the capability to detect, record, report and protect against unauthorized changes to software and information at rest.”

– FR3, SR3.8 Session integrity: “The control system shall provide the capability to protect the integrity of sessions. The control system shall reject any usage of invalid session IDs.”

– FR5, SR 5.2 Zone boundary protection: “The control system shall provide the capability to monitor and control communications at zone boundaries to enforce the compartmentalization defined in the risk -based zones and conduits model.”

D. *Practical Application of IEC 62443*

The standard IEC 62443[11] is applied successfully by operators, integrators, and manufacturers in various projects. However, it is common that the security documentation and technical designs of real-world deployments are not made public or shared with competitors. Still, some examples for applying the IEC 62443 standards have been made available publicly: An example of a possible application of the IEC 62443 standard to an Ukrainian power plant gives some insight concerning how the standard can be applied in a concrete setting [16]. In particular, it shows that a sound, comprehensive security concept is needed that covers security requirements broadly and at a consistent level addressing both, organizational/procedural and technical security requirements. The German industrial association "Zentralverband Elektrotechnik- und Elektronikindustrie e.V." (ZVEI) published an overview document on IEC 62443 that includes an example, showing the application to a simplified automation system [17]. A further example is provided by a blueprint for the design of secure substations in the power system domain [25]. This blueprint has been certified as IEC 62443-2-4 and IEC 62443-3-3 compliant [26].

E. Resilience

Being resilient means to be able to withstand or recover quickly from difficult conditions [2]. It shifts the focus of "classical" IT/OT security, which puts the focus on preventing, detecting, and reacting on cyber-security attacks, to the aspect to continue to deliver an intended outcome despite an adverse cyber attack is taking place, and to recover quickly to regular operation. More specifically, resilience of a system is the property to be resistant to a range of threats and withstand the effects of a partial loss of capability, and to recover and resume its provision of service with the minimum reasonable loss of performance [3]. It has been addressed in telecommunications, ensuring that subscribers can continue to be served even when one line is out of service. Bodeau and Graubart [6] define resilience guidelines for providers of critical national telecommunications infrastructure in the UK. Kott and Linkov [7] have compiled a book of different contributions addressing various aspects of cyber resilience in networks and systems. Besides an overview on cyber security, metrics to quantify cyber resilience, approaches to assess, analyze and to enhance cyber resilience are described. The notion of resilience is related to risk management, and also to robustness. Risk management, the "classical" approach to cyber security, identifies threats and determines the risk depending on probability and impact of a potential attack. The objective is to put the focus of defined security measures on the most relevant risks. Resilience, however, puts the focus on a reduction of the impact, so that the system stays operational with a degraded performance or functionality even when it has been attacked successfully, and to recover quickly from a successful attack. Robustness is a further related approach that tries to keep the system operational *without* a reduction of the system performance [7], i.e., to withstand attacks.

Figure 8 illustrates the concept of cyber resilience: Even if an attack is carried out, the impact on the system operation, i.e., the performance or functionality of the system, is limited. The effects of an attack are "absorbed", so that the system stays operational, but with limited performance or functionality. A recovery takes place to bring the system up to the regular operation. In adaptation of resilience, the system might be enhanced to better prepare for future attacks. In a cyber-physical environment, a main objective is that the CPS stays operational and that its integrity is ensured. In the context of an industrial automation and control system, that means that (only) intended actions of the system in the physical world continue to take place even when the automation and control system of the CPS should be attacked.

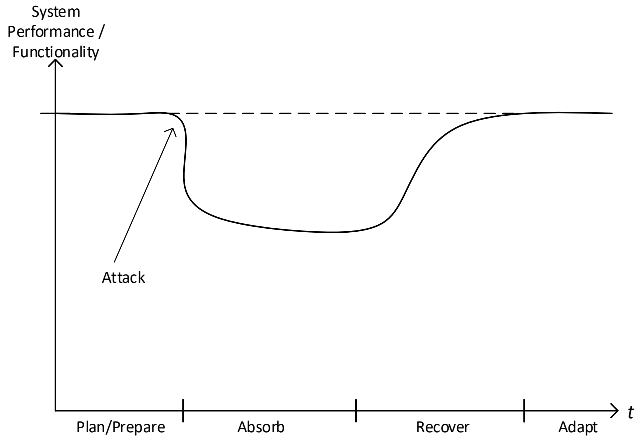


Figure 8. Concept of Cyber Resilience

IV. Protecting the CPS Physical World Interface

Well-known IT security technologies are encryption and access control, protecting data at rest, in transit, and partly even data in use. In cyber-physical systems, this is not enough. Also, the interface between the OT part (automation systems) and the physical world has to be protected, limiting the potential danger that an automation system can have on the physical world when it is attacked. A successful attack on the automation system or control network can have an impact on the physical world [4].

This section describes the concept of a "physical world firewall" that limits the access to the physical world from OT automation systems. The objective is to increase the resilience of cyber-physical systems, by limiting the impact of an attacked automation system on the physical world. It can be seen as a specific approach for increasing cyber resilience, to design for reversibility. This approach means in general that a cyber physical system should be designed in a manner that allows to revert to a safe mode after components have failed or have been compromised [7]. The approach of a physical world firewall, described in the following section, can be both integrated in automation components, or realized as an add-on component to enhance resilience of existing cyber-physical systems (brown-field). It protects the interface between the control system and the physical world, limiting the possible impact of a successful attack on the physical world.

A. Physical-World Firewall

The main idea or the approach is to filter the communication between sensors and actuators on one side, and the control equipment on the other side. This can be called physical-world firewall. It limits in which way a control system, potentially under attack, can impact a physical system in the real world. The filtering takes place directly at the input/output interface, so that it is independent from the software-based functionality of the automation component. Conceptually, it can be considered as a physical world reference monitor to control access to the physical world based on a defined access control policy [8]. However, the physical world firewall described here would be realized independent of the software-based control functionality to ensure that is effective even if the software would be manipulated.

Similar as a communication firewall for data traffic that analyzes and filters data packets (IP packets and IP-based communication, filtering based on network addresses and used protocols), here the actuator and sensor signals are filtered, so that only signals allowed by the signal filter policy are provided.

The allowed signal ranges and dynamic parameters are monitored and limited. If the signal filtering policy is violated, the signal cannot be simply dropped like an IP packet. Instead, a replacement signal is provided. The replacement signal may be a fixed default value, or a clipped maximum/minimum value that is within the allowed value range, or it may be an out-of-range signal or a high-impedance signal that will be detected by an actuator as failure signal, so that the actor can react accordingly).

Figure 9 shows an automation component with an integrated Cyber Physical Controlled Input / Output Interface (CPC IO) that realizes a physical-world firewall functionality. Each input/output channel is monitored separately by the “Value Check” component: It verifies whether the current sensor input value or the current actuator output values are within the given allowed range, and thus are compliant with the defined filtering policy *Pol*.

Besides the value range, also further parameters can be calculated and checked against the defined filtering Policy *Pol*, e.g., statistical parameters as average and variation, and dynamic parameters as a first order or second order derivation, or a transformation as a Fourier transformation. Besides the actual input/output signals, also further data relating to the current operating state of the CPS can be used.

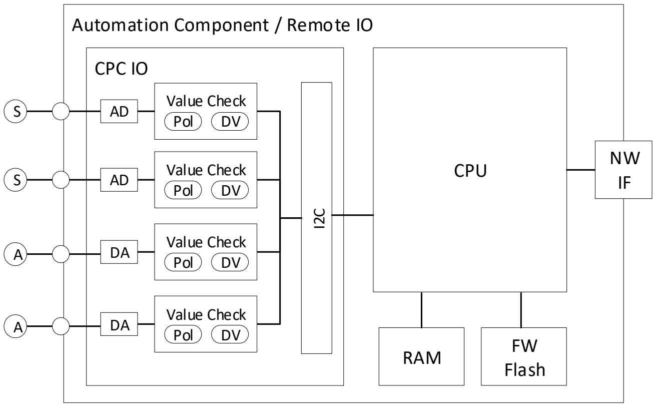


Figure 9. Automation Component with Integrated Physical World Firewall

If the policy is met, the value is allowed, i.e., the unmodified signal is forwarded. Otherwise, the configured default value (DV) is provided as replacement signal, ensuring the CPS system stays operational. It is possible to lock the input/output interface in the case of a detected policy violation. The lock may be permanent, but preferably it can be reset at a reboot of by a manual user interaction.

It is possible that the CPU performs an integrity check as part of a secure boot process or during operation. The CPU subsystem can authenticate towards its CPC IO block after a successful self-integrity check. The CPC IO block can configure a policy depending on the integrity check status of the CPU, limiting the access to the physical world for a manipulated CPU subsystem.

A variant is shown in Figure 10, where the signals of multiple input/output channels are checked in combination. This allows to perform cross-checks between sensor and actuator signals. Moreover, if this approach is applied in a distributed system, it allows to take certain properties of potentially different sensors/actuators into account.

Specifically, if the sensors/actuators used are a mixture of standard (legacy) and specifically hardened, trusted sensors, a potential security assertion can be used in the evaluation of the signals, giving the trusted sensor a higher weight in the evaluation. This is especially advantageous if a larger number of legacy sensors/actuators is already deployed and secure siblings are installed as add-on in a stepwise manner. More information on the basic concept is described in [9].

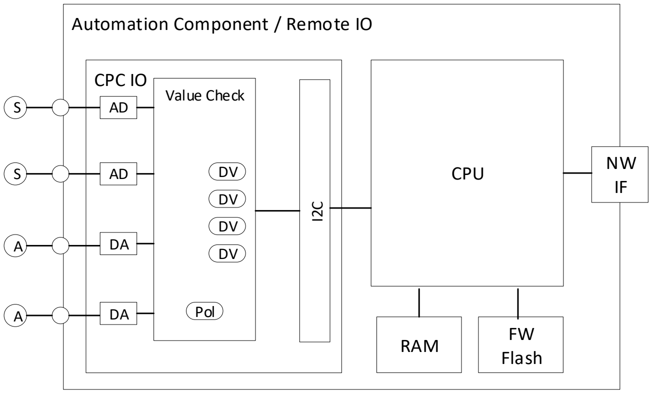


Figure 10. Automation Component with Integrated Physical World Firewall

Both Figure 9 and Figure 10 showed the physical world firewall as an integrated functionality of an automation component. However, it is possible as well to realize the physical world firewall as an add-on component to an existing automation component. This add-on component monitors input/output signals of the automation component between the automation component and the actual sensor/actuator connected to the automation component. The signal is replaced with a replacement signal if the currently observed signal is not complaint with the defined policy *Pol*.

A physical-world firewall realized as add-on component to already existing and deployed automation components can be used in particular within brownfield CPS. A stepwise migration of existing brownfield CPS towards systems with a higher resilience under attack is supported, as the already deployed components of the CPS have not to be replaced.

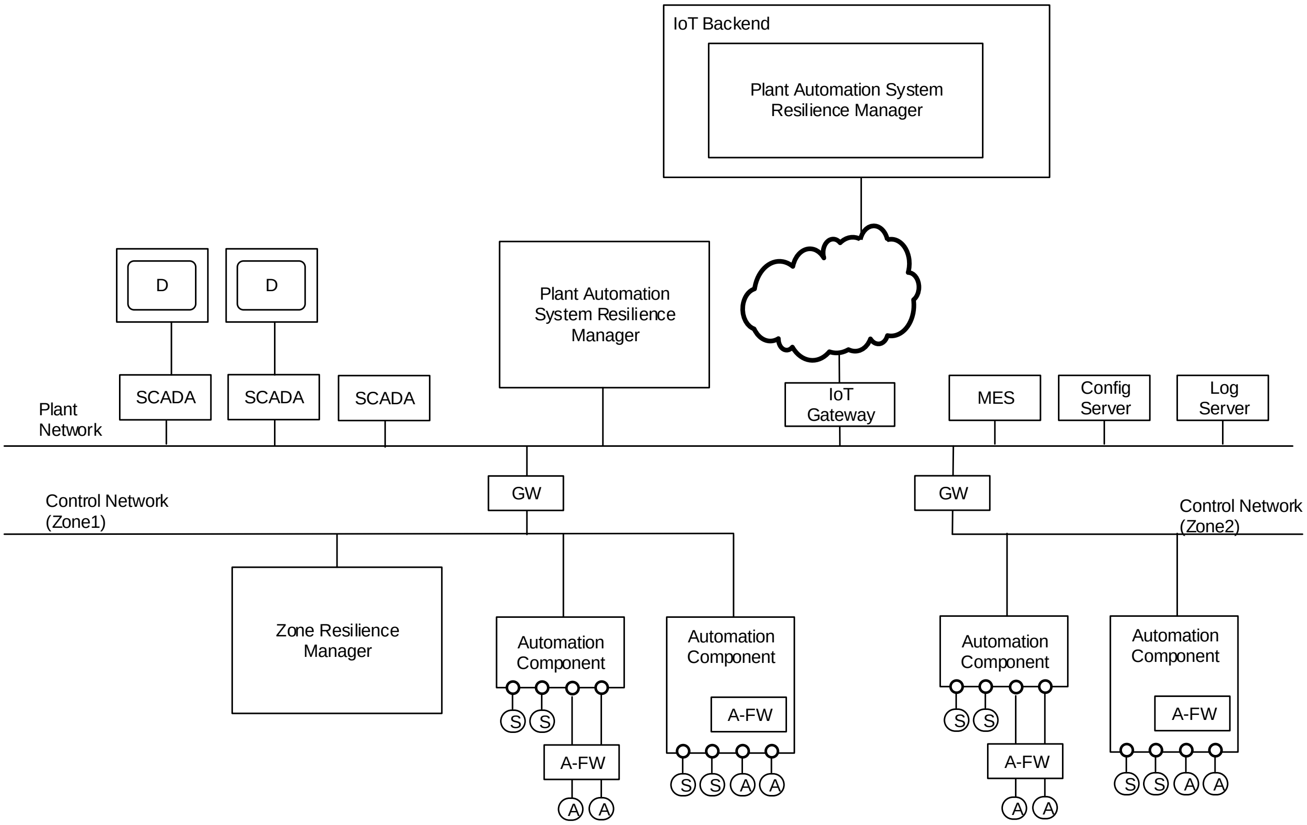


Figure 11. Dynamic Resilience Management

B. Dynamic Resilience Management

The policy of the physical-world firewall can be adapted dynamically, depending on the current operating state of the CPS. This allows to restrict the possibility to influence the physical world even more strictly, as the current state of the production system and the currently performed production step, e.g., cooling or filtering a fluid, can be reflected in the current configuration of the physical-world firewalls.

Resilience managers determine the physical-world firewall policy dynamically, depending on the current state and context of the CPS, see Figure 11.

Resilience managers adapt during operation the current policy configuration of the physical-world firewalls.

The policy adaptation performed by resilience managers can use in particular the following information:

- The current state of the physical world, as obtained by trusted sensor nodes [9].
- The current production batch, the current production step, operating state (e.g., standby, preparation, active, service, alarm). In real-world deployments, the information may be obtained from a Manufacturing Execution System (MES).
- Cyber attacks detected by an integrity monitoring system or an intrusion detection system, supervising the CPS.

The dynamic adaptation allows to enforce tight physical-world firewall policies depending on the current system state and operation.

C. Policy Adaptation for Dynamically Reconfigurable CPS

Cyber-physical systems and industrial automation systems are often rather static. After being put into operation, changes to the configuration happen only rarely, e.g., to replace a defect component, or to install smaller upgrades during a planned maintenance window. To cope with increasing demands for flexible production and increased productivity, also CPS will become more dynamic, allowing for reconfiguration during regular operation. Such scenarios for adaptable, reconfigurable production have been described in the context of Industry 4.0 [19].

An integrity monitoring system for a reconfigurable CPS has to be adapted to the current configuration. Similar as for dynamic resilience management, the policies for physical world firewalls can be adapted depending on the current CPS configuration. The information of the current configuration is usually managed already as part of collecting, storing, and validating production data that describes the production process, so that the information for adapting the policies is available already.

D. Physical World Integrity Monitoring

A further source of information for adapting the filter policy is monitoring the automation site for physical security, using alarm systems, e.g., physical access control and closed-circuit television (CCTV) cameras. If the alarm system detects some unexpected situation, e.g., an intruder, the filter policies of the physical world firewalls can be reconfigured to limit possible damages.

Furthermore, the physical operational properties of automation machinery (e.g., drives, pumps) can be monitored. The acoustic emissions (vibrations) of machines as well as power consumption profile (power fingerprinting) can be monitored. Signal processing algorithms including machine learning (artificial intelligence) can be used to determine whether the machinery is in a normal or exceptional operational state. Also, specific actuations in the physical world can be performed that encode integrity measurements of software and data in control operations, realizing integrity attestation by physical-world actuation signals [18].

If it is detected that the machinery is operating outside the expected operational boundaries, the filter policy of the physical world firewalls can be adapted to limit the impact of the automation system on the physical world accordingly. A restricted physical-world filter policy can be configured dynamically that is foreseen for detected integrity violations.

E. Authenticating Physical Signals

In data communication, the sender of a data packet can be identified by an identifier, e.g., an internet protocol (IP) address or a media access control (MAC) address. The sender may be authenticated cryptographically. A data firewall can filter data packets depending on address information and content. In the physical world, the source of a signal can in general not be identified by an explicit identifier included in the data communication. However, the source is implicitly based on the cabling.

A higher level of confidence can be achieved by performing signal authentication. The sender of a signal can be identified by a sender-specific fingerprint information, e.g., a noise signal. Furthermore, it is possible to actively add a signal marker (signal watermarking), e.g., a coded spread-spectrum signal [20][21][22]. This allows to identify the source of a signal by evaluation properties of the signal. The physical world firewall can identify signals not having the expected fingerprint and block them, i.e., substitute them with a replacement signal. A (physical) signal cannot simply be blocked by not forwarding it. The replacement signal may be a regular signal value, or a specific out-of-range signal value.

The coded spread spectrum signal (signal watermark) can be added to the actual measurement signal close to the analog sensor by adding the watermarking noise signal. However, it is also possible to add actuators in the physical world of the CPS that imprint a watermarking noise signal in the physical world, e.g., by mechanical actuator. Thereby, already deployed sensors (brownfield installation) can capture the watermarking signal, and the sensor measurements can be verified. While having some similarities to the approach described by Ghaeini, Chan, et al. [18], here, the physical world watermarking ensures the reliable identification/authentication of physical signals.

F. Defining Policies for Physical World Firewalls

Even for conventional firewalls filtering network communication, the definition and testing of firewall policies is a huge challenge. The level of security of a network that is actually achieved depends heavily on the ability to manage the available security mechanisms effectively and consistently [23]. This is the case in particular when several firewalls are deployed that have to be configured consistently, and when involving multiple administrative domains. The administration has to be practical, i.e., both efficient and effective, also in such complex environments, with frequent changes and with the complexity of networks consisting of thousands of users and components.

The same applies to policies for physical world firewalls. A further specific side condition is that properties of the physical world have to be understood to come-up with an appropriate policy. This requires a good understanding of the automation system and the physics of the controlled system. A manual configuration of such policies will hardly be practical in real-world deployments. As with other security mechanisms, also physical world firewalls will be introduced stepwise, starting with less critical parts of the CPS and with simple policies to avoid unexpected negative impacts on the regular operation.

For the practical definition of the policies, two approaches seem promising:

- CPS simulation: One important aspect of Industry 4.0 is a digital twin of the physical system that allows to perform simulations in the digital world. Here, the CPS can be simulated under all foreseen operational conditions to derive the filter policies permitting all signals that can be expected in foreseen operational conditions. Also, specific attack scenarios can be simulated.
- Machine learning: The policies can be learnt, similar to network firewalls, where during a learning phase, the policy is automatically determined.

These approaches for determining the filter policies automatically can be enhanced with hand-crafted, manually defined filter policies for interfaces to highly critical physical world components, or for highly critical automation steps. Those tight policies can be adapted specifically to the purpose and foreseen usage of the automation component. So, tight physical world firewall policies can be defined based on a risk assessment, protecting the most relevant components and automation steps.

V. EVALUATION

The security of a cyber system can be evaluated in practice in various approaches and stages of the system’s lifecycle:

- Threat and Risk Analysis (TRA, also abbreviated as TARA) of a cyber physical system (for a system being under design or in operation). In a TRA, possible attacks (threats) on the system are identified. The impact that would be caused by a successful attack (threat) and the probability that the attack happens are evaluated to determine the risk of the identified threats. The risk evaluation allows to prioritize the threats, focusing on the highest, most relevant risks and to define corresponding security measures. Besides technical measures, also organizational and personal security measures can be defined.

- Security checks can be performed during operation or during maintenance windows to determine key performance indicators (e.g., check compliance of device configurations). It can be verified that the defined security measures are in fact in place, and areas requiring increased attention can be identified.
- Security testing (penetration testing, also called pentesting for short) can be performed for a system that has been built, but that is currently not in operation. A pentest can usually not be performed on an operational automation and control system, as the pentest could endanger the reliable operation of the system. Pentesting can be performed during a maintenance window when the physical system is in a safe state, or using a separate test system. The non-operational system is attacked by “white hat” hackers to identify vulnerabilities that need to be addressed.
- Security testing can be performed also on a digital representation of a target system, e.g., a simulation in the easiest case. This digital representation is also called “digital twin”. This allows to perform security checks and pentesting for systems that are not existing yet physically (design phase), or to perform pentesting of operational systems in the digital world without the risk of disturbing the regular operation of the real-world system.

A holistic protection concept has to address measures for all three discussed phases: protect, detect, and react. No single measure or security technology alone can result in an adequate security level. It is always a set of measures that, when used in combination, can reduce the overall risk to an acceptable level.

The security measures presented in this paper, acting on the interface between the cyber world and the physical world, provide an additional security measure that can be used as part of a defense-in-depth security concept. The protection is complementary to well-known security measures that focus on the IT/cyber part, as it operates directly at the interface towards the physical world, not on computer-based control functions as conventional IT security technologies. Even if all security measures in the pure IT/cyber world fail, still the impact on the physical world can be controlled. It can serve as “last line of defense”, allowing to connect cyber systems from the physical world in a tightly controlled way.

A limitation for all evaluations of the effectiveness of an overall security architectural design and of individual security measures is the fact that the threat landscape of attacks seen in practice continuously evolves. Therefore, it is required that a security design allows for being updated to address new attacks. This aspect is in particular important for CPS and automation and control systems having typically a long life-time of typically 10 to 30 years. The defined concept of physical world firewalls supports an update not only to already existing brownfield installations, but also to enhance the security robustness of long-lived systems during operation without directly affecting the control functionality. As CPS are often subject to regulatory approvals, having security measures that can be updated and enhanced along the lifetime without directly affecting regulatory approvals of the control functions is advantageous.

As long as the proposed technology has not been proven in a real-world operational setting, it can be evaluated conceptually by analyzing the impact that the additional security measure would have on the identified residual risks as determined by the TRA of the CPS. A TRA identifies threats against a system, and determines the risk depending on probability and impact. The general effect of the presented security measure is that the impact of a threat, i.e., a successful attack, on the physical world controlled by the CPS is reduced. Whatever attack is ongoing on the IT-based automation and control system, still the possible impact on the real, physical world is limited. So, the measure helps to reduce the risk of threats having an impact on the physical world.

However, TRAs for real-world CPS are not available publicly. Nevertheless, an illustrative example may be given by a chemical production plant performing a specific process like refinery, or a factory producing glue or cement. If the plant is attacked, the attack may target to destroy the production equipment by immediately stopping the process leading to physical hardening of the chemicals / consumables and thus to a permanent unavailability of the production equipment. In this case, trusted sensors could be used to detect a falsified sensor signal, and the physical-world firewall can be used to limit actions in the physical world. Thereby, a physical damage of the production equipment can be avoided. If needed, a controlled shutdown of the production site can be performed.

As the evaluation in a real-world CPS requires significant effort, and as attack scenarios cannot be tested that could really have a (severe) impact on the physical world, a simulation-based approach or using specific test-beds are possible approaches, allowing to simulate or evaluate in a protected test-bed the effect on the physical world of certain attack scenarios with compromised components. The simulation would have to include not only the IT-based control function, but also the physical world impact of an attack. Using physical-world simulation and test beds to evaluate the impact of attacks have been described by Urbina, Giraldo et al. [24].

A major advantage of the physical-world firewall is the property that it can be added to existing brownfield deployments. Legacy equipment, may be 10 or even 20 years in the field, not even been designed with security in mind, and without getting patches. In such cases, the physical-world firewall can be used as an “add-on” security measure for an existing CPS. It can be used as compensating countermeasure to address security requirements defined by industrial security standards like IEC 62443-3.3 [14], where conventional cyber security measures cannot be deployed. However, it can be used also as additional layer of defense in CPS having state-of-the-art security measures integrated, thereby increasing the level of protection even further. The conceptual advantage that the protection acts on a different layer than conventional IT security mechanisms provides an additional, independent layer of defense. As for all security technologies, the confirmation for the actual effectiveness has to come from tests and experience real-world application, starting with smaller pilot tests in real deployments.

VI. Conclusion

With ubiquitous machine-oriented communication, e.g., the Internet of Things and interconnected cyber physical systems (CPS), the integrity of the operation of technical systems is becoming an increasingly important security objective. Protecting such systems against intentional attacks to ensure a reliable operation is demanded by operators, as well as by regulation. There is a need for enhanced protection that can be applied practically both to already deployed installations, where often IT-based functionality cannot be updated practically, as well as to new CPS, which are increasingly open and dynamic.

A CPS comprises the operational cyber-technology and the physical world with which the system interacts. Both parts have to be covered by a security concept and solution. Cyber security puts the focus traditionally on the cyber-part, i.e., on the IT-based automation and control systems. The security of the physical part, like machinery, is protected often by physical and organizational security measures, only. This is challenging for dynamically changing cyber physical systems, that come with the Industrial Internet of Things (IIoT) and Industry 4.0. Cyber systems will become more and more open and dynamic to support flexible production down to lot size 1 (plug-and-work reconfiguration of manufacturing equipment), and to support a flexible adaptation to changing needs like market demand and personalized products.

This paper presented a concept for a new approach that enhances the achieved level of security by protecting the interface between the IT-based cyber-part and the physical world, thereby enhancing the resilience of a CPS being under attack. The CPS may even continue to operate under attack, as the possible negative impact on the physical world is restricted. This allows also to ensure a high availability of the automation system, even under attack, as the automation system has not to be shut down.

The proposed new layer of protection can be applied to new installations (greenfield), e.g., to address the risk of installing malware during update of the software-based functionality. More importantly, it can as well be applied as add-on to already deployed installations (brownfield). It realizes an additional, independent level of protection that can be deployed and updated independently of the actual control systems of the legacy system. Therefore, it can also be applied when a legacy IT-based control system of a CPS cannot be updated with current cyber security technology. This is a demanding problem in many installed CPS, as they are often in use for several decades and are subject to regulations that make updates complicated or even impossible. The proposed solution can be introduced in a complex CPS in a stepwise way, starting with most critical physical world interfaces. Also, the filtering policies can be coarse in an initial usage phase, and it can be updated with increasing sophistication depending on observed attacks, and reflecting the intended operation of the specific CPS and its current operation mode.

References

[1] R. Falk and S. Fries, “Enhancing Resilience by Protecting the Physical-World Interface of Cyber-Physical Systems”, The Fourth International Conference on Cyber-Technologies and Cyber-Systems CYBER 2019, September 22, 2019 to September 26, 2019 - Porto, Portugal, [Online]. Available from: https://www.thinkmind.org/index.php?view=article&articleid=cyber_2019_1_20_80033 2020.05.13

[2] P. England, R. Aigner, A. Marochko, D. Mattoon, R. Spiger, and S. Thom, “Cyber resilient platforms”, Microsoft Technical Report MSR-TR-2017-40, Sep. 2017, [Online]. Available from: https://www.microsoft.com/en-us/research/publication/cyber-resilient-platforms-overview/ 2020.05.13

[3] Electronic Communications Resilience&Response Group, “EC-RRG resilience guidelines for providers of critical national telecommunications infrastructure”, version 0.7, March 2008, available from: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/62281/telecoms-ecrrg-resilience-guidelines.pdf 2020.05.13

[4] D. Urbina, J. Giraldo, N. O. Tippenhauer, and A. Cardenas, “Attacking fieldbus communications in ICS: applications to the SWaT testbed”, Singapore Cyber-Security Conference (SG-CRC), IOS press, pp. 75–89, 2016, [Online]. Available from: http://ebooks.iospress.nl/volumearticle/42054 2020.05.13

[5] C. C. Davidson, T. R. Andel, M. Yampolskiy, J. T. McDonald, W. B. Glisson, and T. Thomas, “On SCADA PLC and fieldbus cyber security”, 13th International Conference on Cyber Warfare and Security, National Defense University, Washington, DC, pp. 140–148, 2018

[6] D. Bodeau and R. Graubart, “Cyber resiliency design principles”, MITRE Technical Report, January 2017, [Online]. Available from: https://www.mitre.org/sites/default/files/publications/PR%2017-0103%20Cyber%20Resiliency%20Design%20Principles%20MTR17001.pdf 2020.05.13

[7] A. Kott and I. Linkov (Eds.), “Cyber Resilience of Systems and Networks”, Springer, 2019

[8] E. B. Fernandez, M. VanHilst, D. laRed Martinez, and S. Mujica, An Extended Reference Monitor for Security and Safety, 5th Ibero-American Congress on Information Security, Montevideo, Uruguay, November 2009, [Online]. Available from: http://www.criptored.upm.es/cibsi/cibsi2009/docs/Papers/CIBSI-Dia2-Sesion4(3).pdf 2020.05.13

[9] R. Falk and S. Fries, “Enhancing integrity protection for industrial cyber physical systems”, The Second International Conference on Cyber-Technologies and Cyber-Systems, CYBER 2017, pp. 35–40, November 12 - 16, 2017, Barcelona, Spain, [Online]. Available from: http://www.thinkmind.org/index.php?view=article&articleid=cyber_2017_3_30_80031 2020.05.13

[10] European Commission, “The directive on security of network and information systems (NIS Directive)”, [Online]. Available from: https://ec.europa.eu/digital-single-market/en/network-and-information-security-nis-directive 2020.05.13

[11] IEC 62443, “Industrial automation and control system security” (formerly ISA99), [Online]. Available from: http://isa99.isa.org/Documents/Forms/AllItems.aspx 2020.05.13

[12] ISO/IEC 27001, “Information technology – security techniques – Information security management systems – requirements”, October 2013, [Online]. Available from: https://www.iso.org/standard/54534.html 2020.05.13

[13] NIST, ”Framework for Improving Critical Infrastructure Cybersecurity”, Version 1.1, April 16, 2018, [Online]. Available from: https://nvlpubs.nist.gov/nistpubs/ CSWP/NIST.CSWP.04162018.pdf 2020.05.13

[14] IEC 62443-3-3:2013, “Industrial communication networks – network and system security – Part 3-3: System security requirements and security levels”, Edition 1.0, August 2013

[15] IEC 62443-4.2, “Industrial communication networks - security for industrial automation and control systems - Part 4-2: technical security requirements for IACS components”, CDV:2017-05, May 2017

[16] P. Bock, J.-P. Hauet, R. Françoise, and R. Foley, “Ukrainian power grids cyberattack - A forensic analysis based on ISA/IEC 62443”, ISA InTech magazine, 2017, [Online]. Available from: https://www.isa.org/templates/news-detail.aspx?id=152995 2020.05.13

[17] ZVEI, ”Orientation guideline for manufacturers on IEC 62443”, “Orientierungsleitfaden für Hersteller zur IEC 62443“ [German], ZVEI Whitepaper, 2017, [Online]. Available from: https://www.zvei.org/presse-medien/publikationen/orientierungsleitfaden-fuer-hersteller-zur-iec-62443/ 2020.05.13

[18] H.R. Ghaeini, M. Chan, R. Bahmani, F. Brasser, L. Garcia, J. Zhou, A.-R. Sadeghi, N. O. Tippenhauer, and S. Zonouz, “PAtt: Physics-based Attestation of Control Systems”, 22nd International Symposium on Research in Attacks, Intrusions and Defenses, USENIX, September 23-25, 2019, [Online]. Available from: https://www.usenix.org/system/files/raid2019-ghaeini.pdf 2020.05.13

[19] Plattform Industrie 4.0, “Industrie 4.0 Plug-and-produce for adaptable factories: example use case definition, models, and implementation”, Plattform Industrie 4.0 working paper, June 2017, [Online]. Available from: https://www.plattform-i40.de/PI40/Redaktion/DE/Downloads/Publikation/Industrie-40-Plug-and-Produce.pdf , 2020.05.13

[20] T. Hupperich, H. Hosseini, and T. Holz, “Leveraging sensor fingerprinting for mobile device authentication”, International Conference on Detection of Intrusions and Malware, and Vulnerability Assessment, LNCS 9721, Springer, pp. 377–396, 2016, [Online]. Available from: https://www.syssec.ruhr-uni-bochum.de/media/emma/veroeffentlichungen/2016/09/28/paper.pdf 2020.05.13

[21] H. Bojinov, D. Boneh, Y. Michalevsky, and G. Nakibly, “Mobile device identification via sensor fingerprinting”, arXiv:1408.1416, 2016, [Online]. Available from: https://arxiv.org/abs/1408.1416 2020.05.13

[22] P. Hao, “Wireless device authentication techniques using physical-layer device fingerprint”, PhD thesis, University of Western Ontario, Electronic Thesis and Dissertation Repository, 3440, 2015, [Online]. Available from: https://ir.lib.uwo.ca/etd/3440 2020.05.13

[23] R. Falk and M. Trommer, “Integrated Management of Network and Host Based Security Mechanisms,” 3rd Australasian Conference on Information Security and Privacy, ACISP98, pp. 36-47, July 13-15, 1998, LNCS 1438, Springer, 1998

[24] D. Urbina, J. Giraldo, A. Cardenas, N. O. Tippenhauer, J. Valente, M. Faisal, J. Ruths, R. Candell, and H. Sandberg, “Limiting The Impact of Stealthy Attacks on Industrial Control Systems,” ACM Conference on Computer and Communications Security (CCS), Vienna, Austria, 2016

[25] Siemens, “Secure Substation Manual”, [Online]. Available from: https://www.siemens.com/download?DLA20_114 2020.05.13

[26] Siemens, “Digital Substation Cyber Security”, [Online]. Available from: https://www.siemens.com/download?DLA13_3680 2020.05.13

Searching for Stars
Analyzing and Defining UAV Cyber Risk Assessments

Dillon Pettit
Graduate Cyber Operations
Air Force Institute of Technology
Dayton, Ohio, USA
Email: Dillon.Pettit@afit.edu

Scott Graham
Dept of Computer Engineering
Air Force Institute of Technology
Dayton, Ohio, USA
Email: Scott.Graham@afit.edu

Patrick Sweeney
Dept of Computer Engineering
Air Force Institute of Technology
Dayton, Ohio, USA
Email: Patrick.Sweeney@afit.edu

***Abstract*—The small Unmanned Aerial Vehicle market, commonly called UAVs, has grown immensely in popularity in hobbyist and military inventories. The same core mission set from the hobbyists directly relates to modern global military strategy, with priority on short range, low cost, real time aerial imaging and limited modular payloads. These small devices have the added benefits of small cross sections, low heat signatures, and a variety of transmitters to send real-time data over short distances. As with many new technologies, security seems secondary to the goal of reaching the market as soon as viable. Research indicates a growth in exploits and vulnerabilities, from individual UAV guidance and autopilot controls to the mobile ground station devices that may be as simple as a cellphone application. Even if developers heed calls to improve the security of small UAVs to protect them, consumers are left without meaningful insight into the protections installed when buying new or used UAVs. To date, there is no marketed or accredited risk index for small UAVs, but similar realms of traditional Aircraft operation, Information Technologies, Cyber-Physical Systems, and Cyber Insurance give insight to significant factors required for future small UAV risk assessment. In this research, four fields of risk frameworks are analyzed to determine their applicability to UAV security risk and key components that must be analyzed by a formal UAV framework. This analysis demonstrates that no adjoining field's framework can be directly applied without significant loss of fidelity and that further research is required to score the cyber risks of UAVs, along with potential objectives and avenues for then creation of a new framework.**

Keywords*—*Cyber-physical; Cybersecurity; COTS; Quantitative assessment; Risk; UAV.

I. Introduction

Cybersecurity is the Herculean task to prevent all adversarial attacks over Information Technology (IT) devices with the potential to release information or control deemed valuable to an organization or individual. As computing devices have increased in variety and distribution around the globe, the protection task has grown immensely, with absolute security now accepted as a myth. However, due diligence has been seen to reduce and slow incidents. IT devices have diverged into a multitude of subcategories, including Cyber-Physical Systems (CPSs) and a further subsection of Small Unmanned Aerial Vehicles (sUAVs). While many techniques used to map and defend IT may be extended to sUAVs, CPSs in general have significant differences in internal architecture, external networking, and overall mission sets that influence the effectiveness of common cybersecurity techniques. An important aspect of cybersecurity is risk categorization of individual devices and the conglomeration on a network, which relies on common rating measures for comparison. IT devices still struggle with communication of security characteristics, though certain brands have made strides to separate themselves from the competition. This paper is an extension of the "Zero Stars" paper [1] to define the requirements for a simple rating system for consumers to effectively manage small Unmanned Aerial Vehicle (UAV) risk. The addition of traditional aircraft risk management provides new insights to the current risks facing UAVs that are not being managed by manufactures or consumers.

UAVs have been historically built for military applications and continued by hobbyist enthusiasm. By definition, UAV includes any device that can sustain flight autonomously, which separates it from similar sub-cultures of Remotely Piloted Vehicles (RPVs) and drones [2]. UAVs are usually able to either maintain a hover or move autonomously via computer navigation, whereas RPVs require continuous control instructions throughout flight and drones have even more limited mission and sophistication [2]. Arguably, the first UAV could be considered cameras attached to kites in 1887 by Douglas Archibald as a form of reconnaissance and which William Eddy used the same configuration during the Spanish-American War for reconnaissance [2]. As UAV operations and innovations continued through the Vietnam War, Desert Storm, and especially the Global War on Terror, the size, mission, and shape of UAVs have evolved to support military needs. Criminal uses have also grown with UAV prevalence with ingenious modifications matching latest military exploits [3].

UAVs take a multitude of forms and designs based on mission and user base, from hand-held copters to jet-powered light aircraft. Small UAVs follow the general component break out shown in Figure 1, with six common components on the device and a ground station of some sort. The Basic System is a generalized term for the Operating System (OS), which is usually proprietary to the manufacturer and tailored per vehicle, frequently providing near real time control. Communication Links are most commonly wireless Radio Frequency (RF) bands of 2.4 and 5 GHz. Sensors refer to components that are attached to either aid navigation of the system, such as LIDAR to monitor nearby structures, or for specific mission purposes. Avionics consume sensor input, such as Global Positioning System (GPS) and inertial modules, and provide flight control. For the payload, a weapon component has been seen within military operations, though the vast majority of sUAVs are used for military or hobbyist reconnaissance with only an additional sensor component such as a camera. As defined for UAV, some form of autonomous control is built into the vehicle's navigation, so the autopilot component is logically separated from the Basic System but usually physically combined.

The ground station is split into the Application component and Communication links, though these are typically contained within the same device such as a tablet, phone, or laptop. The complexity and portability of ground stations vary widely from simple RF remote controls to multi-server backends. Examples of these differences can be seen in the common DJI Sciences and Technologies Limited (DJI) brand, which utilizes both manufacture specific hardware and a smartphone application. The software is extremely portable through mainstream app stores and can be updated over secure connections. The hardware connects to the user's smartphone to provide controls to the sUAV with separate antennas and power supply for better coverage. The application can also be used without the hardware through a laptop to program mission states via physical cable. Some DJI models even allow simple remote controls or beacons without application software, though their mission sets are more rudimentary. Each of these configurations introduce risk characteristics by connecting the device to the Internet differently.

The exact definitions of size tiers have not been standardized between countries though they generally consist in some format of very small, small, medium, and large. Very small UAVs exist at a miniaturization of aerodynamics that result in very low Reynolds numbers, meaning the wing interacts with the air more similarly to a fin through water due to viscosity, and are usually less than 20 inches in any dimension. Small UAVs tend to be a range of popular model aircraft used by hobbyists and have at least one dimension greater than 20 inches. While range is limited, their size allows for access or angle of attack at altitudes not normally available to individuals. Medium and Large UAVs are too large for an individual to carry and may even use full runways like light aircraft, which allows for heavier payloads and greater mission duration. Instead of a pilot and sensors, sUAVs are controlled by an autopilot, with varying degrees of autonomy. Autopilots vary greatly by manufacturer, with the most common DJI autopilots closed-source and their specific rules sets proprietary [3].

The rest of the paper is structured as follows. Section II explores current common rating systems for Traditional Aircraft, IT, Critical Infrastructure, and Insurance markets with a focus on the aspects of each that do translate to the sUAV inventory. Section III builds out from the conglomeration of related rating assessments the important objectives that are required for a sUAV specific cybersecurity rating. Section IV analyzes each of the fields for their applicability to a small UAV risk assessment for potential adaptation. We conclude our work in Section V with future work.

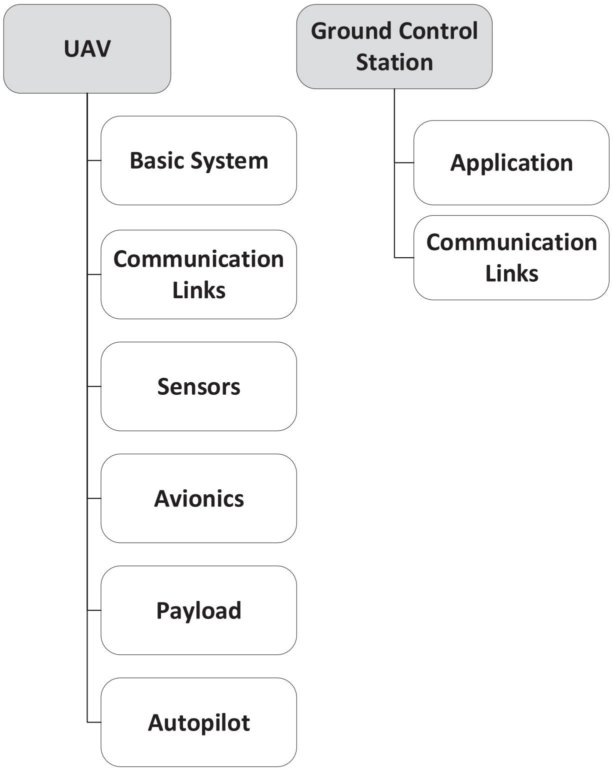


Fig. 1. Components of Typical UAV.

II. Related Work

No current physical or cyber security accreditation exists for UAVs. Since no current process exists to calculate risk, quantitative or qualitative, for sUAVs, there are no star ratings present on the market to be assigned to any sUAV, much less to compare models. Confounding the issue, aerial vehicles were engineered for operational effectiveness first, then marketed with minimal consideration for adversarial interference. Publicized cyber incidents with and against UAVs have been limited with the most well-known consisting of the Iranian incident in 2015 [4]. Whether the United States RQ-170 was captured by Electronic Warfare (EW) or cyber means [4], the incident highlights the vulnerability of UAVs in a combat zone and the need for security in future models to maintain integrity for mission success. With 15,000 UAVs being sold in the United States every month as of 2015 [5], the availability and exploitation of these devices is expected to also rise as the reward to effort ratio grows. The market share of small UAVs manufacturers is as follows: 70% DJI, 7% Parrot, 7% Yuneec [6], with the remaining 14% comprised of all others. DJI and Yuneec are Chinese controlled manufacturers. This domination by China presents yet another avenue of supply chain risk that many organizations and countries with competing interests may want to be wary. Research into the vulnerability of sUAVs has also increased with a multitude of research showing specific risk in areas of Denial of Service (DoS) [7], GPS spoofing [8], and control hijacking [9].

A. Traditional Aircraft Assessment

The invention and market for UAVs stemmed from the traditional aircraft field. Regular aircraft have always been larger to accommodate the weight and thrust requirements needed for carrying pilots. In contrast, unmanned technologies have allowed for the creation of smaller vehicles. With nearly all on-board components being seen on both vehicles, a cyber risk assessment for traditional aircraft could be assumed to be the best translation to sUAVs, especially taking into account cyber-physical aspects that are not seen in other IT fields. Regrettably, the aircraft industry does not currently have any cyber assessments for risk [10]. While industry standards for the design of aircraft information systems exist that incorporate defence in depth (RTCA SC-216 and EUROCAE WG-72), there is no measure of how well these standards were implemented or any comparison between vehicles, and no expected updates to either standard through 2021 [10]. The Aerospace Industries Association (AIA) Civil Aerospace Cybersecurity subcommittee identified that each manufacturer and operator defines their own risk framework and assessment of cyber risk on their aircraft; therefore, there is no commercial aviation cyber safety Cyber Action Team (CAT) to set standards and respond to incidents [10]. As one of the key priorities of the report, the AIA subcommittee published that the industry needs “a risk managed approach...to architect future secure systems” and “better global visibility...to address aviation ecosystem threats and risks” [10].

Since the manufacturers have strict operational regulations but do not have any cyber assessments for aircraft, the Federal Avionics Administration (FAA) has had to incorporate a real-time operational risk assessment to the Aircraft Traffic Management (ATM) system which all traditional aircraft and all larger UAVs connect to for deconfliction of real-time flight plans [11]. Recognizing the need for including smaller UAVs, the FAA has granted funds to the National Aeronautics and Space Administration (NASA) to build and test a new ATM to manage the National Airspace System (NAS) as of 2014 [11]. Building from the ATM risk framework, NASA published the Unmanned Aerial Vehicle Traffic Management Risk Assessment Framework (URAF) which calculates a numerical risk value to correspond to the expected real-time risk associated with collisions per vehicle [12], which is calculated in Figure 2 at the “Conflict?” step. Using Bayesian networks fully defined for every potential component failure based upon the Unmanned Aerial System Traffic Management (UTM)’s input of vehicle and ground sensors, the URAF determines the probability of a collision with another aircraft, a structure, or a human being based on density maps of the United States (US) [12]. The flowchart shown is part of the patent for the system and is nearly a step-by-step reproduction from the current NAS framework, which fails to capture the differences between vehicles and only the operational risk. The small UAV sensor inputs are defined in the patent for the UTM as the required 14 communication protocols, none of which are currently required on small UAVs [13]. Initial tests of the UTM including the URAF was conducted in 2016 at seven FAA testing sites with 17 unique vehicles, though its success was marred with 32.5% non-conforming operations [14]. Non-conforming operations were defined by any position during a vehicle’s mission that broke the operational risk threshold for collision, whether or not collisions actually occurred. The Bayesian network utilized in this testing captured the risk associated through one component failure and calculated from only five sensors [15]. NASA set the goal of initial operation by 2019, which was reached in the form of beta expansion

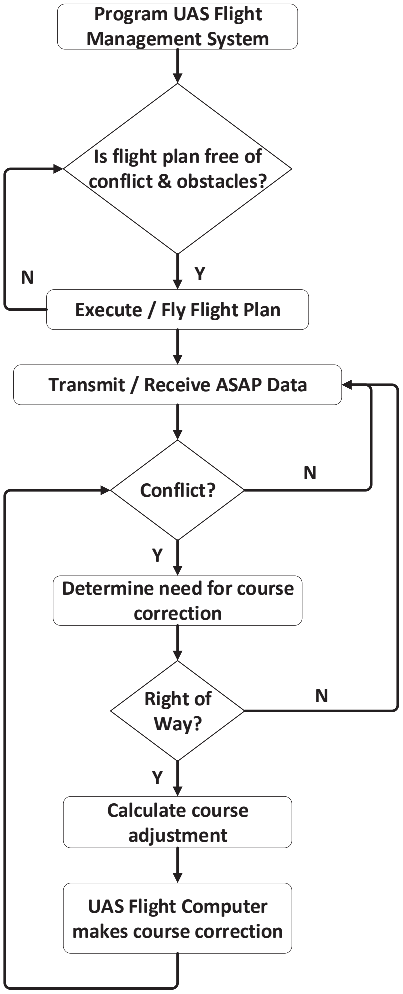

Program UAS Flight Management System
Is flight plan free of conflict & obstacles?
N
Y
Execute / Fly Flight Plan
Transmit / Receive ASAP Data
Conflict?
N
Y
Determine need for course correction
Right of Way?
N
Y
Calculate course adjustment
UAS Flight Computer makes course correction

Fig. 2. FAA’s UTM Control Flow for UAS [11].

to the Low Altitude Authorization and Notification Capability (LAANC) in May of 2019 [16] at over 600 airports, and full operation for massive density operations by 2035 [11]. This beta is not the full UTM design, but a parallel authorization and tracking system of small UAVs. The URAF and the ATM risk framework are both device agnostic, except in terms of size and value [13]. There is no input of vehicle design, securities, or abilities to maneuver, all of which are a factor of cyber risks to aircraft. Due to the increasing automation and computation of aircraft, future risk assessments must individually consider each vehicle.

B. Traditional IT Assessment

UAVs can also be viewed as simply flying computer systems. Traditional IT risk assessments have been around since the early 2000s [17] and have almost solely focused on business devices and networks. While Network Security Risk Model (NSRM) [18] and Information Security Risk Analysis Method (ISRAM) [17] are some of the oldest quantitative risk assessment models, Common Vulnerability Scoring System (CVSS) is the most utilized today [19].

CVSS version 3.1 is an "open framework for communicating the characteristics and severity of software vulnerabilities" [20]. The score is based on three different metrics of a Base ranging from 0.0 to 10.0, tempered by Temporal and Environmental metrics. CVSS is owned and managed by Forum of Incident Response and Security Teams (FiRST) and is a significant information provider to the National Vulnerability Database (NVD). CVSS first gained large-scale usage under their Version 2 score which determined only a base score through metrics for Access Vector, Access Complexity, Authentication, Confidentiality Impact, Integrity Impact, and Availability Impact. Each metric was given a rating from up to three varying responses of severity. CVSSv2 was criticized heavily for vulnerability scoring diversity compared to experimental, lack of interdependence scoring of networks, and lack of correlation between proposed mitigations and actual score improvements [19]. CVSS version 3.0 added mandatory components for Privileges Required, User interaction, and Scope, plus the temporal and environmental metrics to influence the overall score. The current version has grown in use for vulnerability scoring, but still struggles with high false positive rates, poor predictability of future incidents, high sensitivity in regards to Availability Impact compared to all other impacts, and is heavily influenced by software type [21]. Built from CVSS, NVD has been found to lack in predicting mean time to next vulnerability due to the Common Vulnerability and Exploitations (CVEs) recording poor and inconsistent data by vendor and an increasing discovery, across vendors, of zero-day vulnerabilities [22]. The most recent version 3.1 of CVSS, summarized in Figure 3, also updated CVSS's mission from a simple risk severity to more limited vulnerability severity. The Base Metrics are split into three sub-categories due to commonalities in rating or how they are utilized in the underlying algorithms. The same grouping is applied to the Environmental Metrics, where the Modified Base sub-metrics are a repeat of the Base Metrics but updated for a network's individually unique security configuration. Each sub-metric is not equal, but are weighted numerically to best represent the severity that sub-metric conveys to the overall severity. In general practice, the Base Metrics, representing the severity of the attack, are the most heavily weighted as they can singularly push the severity to the extremes of an overall score of 0 or 10. Due to the change of scoring severity over risk and their consistent updating of algorithms, CVSS is a good starting point for known vulnerabilities present within a UAV, but the unique embedded nature of components, the normally informal and ad hoc networks used by UAVs, and unique mission and environment sets mean CVSS is not very likely to give a good perspective of actual vulnerabilities present and therefore directly assess risk.

C. Industrial CPS and Supervisory Control and Data Collection (SCADA)

At the other end of the spectrum for security indexing, sUAVs could be related to larger CPSs which have recently seen a surge in research and regulations to secure their unique networks. Industrial CPS and SCADA have been utilized to gradually reduce required human interaction in safety-compromised work areas and in wide distributed networks. Physical sensors formerly required eyes to read, determine system state, and adjust actuators to keep processes within safety limits and manufacturing effectiveness. These sensors are now directly digitized by network adapters, delivered to Programmable Logic Controllers (PLCs) that determine state, compute new controls, and send signals to actuators to finish the feedback loop. Human-Machine Interface (HMI) screens give a real-time display of the system state with minimal human interaction while smoothly running our critical infrastructure. SCADA systems are owned by corporations that produce or deliver their products to consumers; therefore the networks are not the product themselves, in contrast to home computers or even work stations which are most commonly modelled by IT networks. As CPS stations are utilitarian and usually connected to physical sensors for input, protection schemes need to adjust for their physical process monitoring, closed control loops, attack sophistication, and legacy technology [23]. The first two categories define differences in attack vectors for cyber-to-cyber or cyber-to-physical exploitation. Regular IT exploitation follows a typical path that ends at an IT node with information which is valuable in itself; whereas industrial CPS exploitation usually requires further exploitation to influence physical processes to either ruin or shut down systems [24]. This leads to attack sophistication differences between IT and SCADA risk, since physical process manipulation via PLCs require detailed understanding of systems that are only present in the operational world. While the attack vectors require unique background, the computer systems monitoring and running the physical processes are commonly characterized by legacy equipment

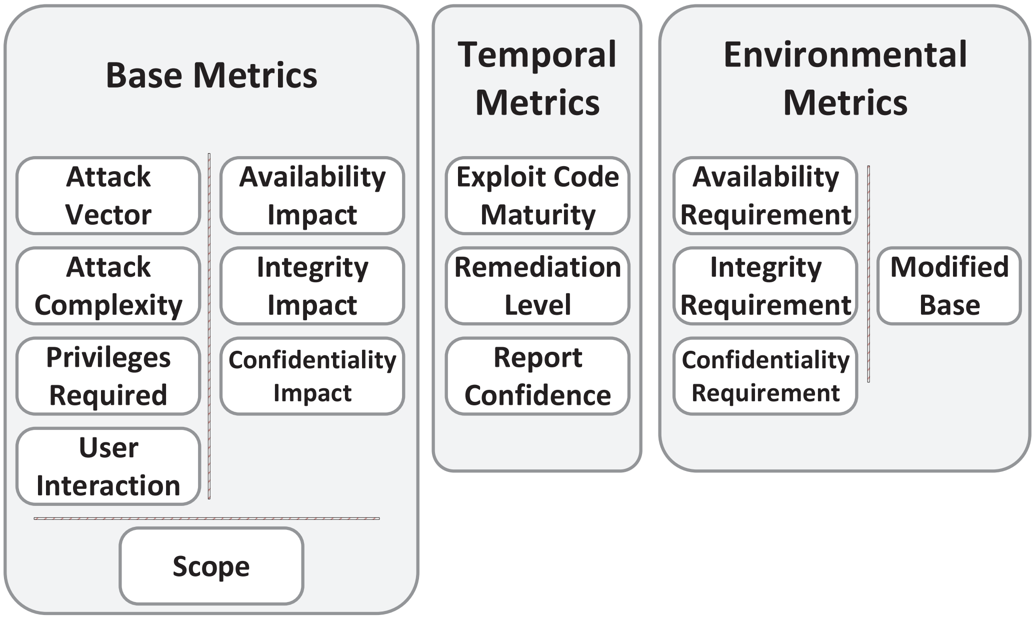


Fig. 3. Metric and Sub-metric breakout of CVSS [20].

with many known vulnerabilities. IT cybersecurity practices push for upgrade cycles on a regular basis to keep pace with manufactures' patching, however industrial systems are unable to upgrade nearly as often and require much larger investment capital to replace legacy systems that are considered permanent fixtures. Research into adding cybersecurity to CPS systems skyrocketed after the discovery of the sophisticated Stuxnet virus in a nuclear plant. The nuclear plant in question has been studied, with its cybersecurity posture matching industry standards and much of the IT standards [25].

Risk assessments building from this impetus, and focusing on more than just nuclear, have attempted to predict the new methods to exploit processes. Most standardized methods merely cover the cyber-to-cyber and physical-to-physical exploitation, which arguably cover the easiest and most common historical attacks [26]. Stuxnet introduced publicly the possibilities of cyber-to-physical exploitation while little is known of possible physical-to-cyber vectors. At the direction of Department of Homeland Security (DHS), National Institute of Standards and Technology (NIST) published the Cybersecurity Framework (CSF) to directly define a risk framework for critical infrastructure in the US [27]. The core of the framework is the process of Identify, Protect, Detect, Respond, and Recover [27]. The framework's first push is to fully define the network currently in operation down to individual sensors with definitions of all system states. The next step is simple cybersecurity fundamentals such as segregation and locking down unnecessary protocols. Once at this steady operational state, the framework directs the effort to setup methods of detection, response, and recovery from attacks. While the framework does reduce the footprint and likelihood of attack, there is no assessment of the risk state of the system nor a method of comparison between systems [26]. Even within unique critical infrastructure systems, it is useful to supervising organizations and protection agencies to compare system risks to more effectively protect nation-wide assets.

TABLE I. Cybersecurity Framework Core and Sub-Categories.

Core Phases	Sub-Categories
Identify	• Asset Management • Business Environment • Governance • Risk Assessment • Risk Management Strategy • Supply Chain Risk Management
Protect	• Identity Management and Access Control • Awareness and Training • Data Security • Information Protection Processes and Procedures • Maintenance • Protective Technology
Detect	• Anomalies and Events • Security Continuous Monitoring • Detection Processes
Respond	• Response Planning • Communications • Analysis • Mitigation • Improvements
Recovery	• Recovery Planning • Improvements • Communications

Attempting to cover the lack of assessment of critical infrastructure systems' cyber risk, Cyber Security Risk Index (CSRI) is a proposed and beta risk assessment specifically using Bayesian Networks since systems should be defined through CSF. To cover the cyber-to-physical risk, the most common technique is to use Markov chains in conjunction with the

Bayesian Networks which allows for distinct states along with probabilities of events. [28]. A major drive to Bayesian networks is the complex states that physical processes may enter, which differ on Mean Time to Shut Down (MTTSD). While the probabilities to reach across the IT network to the PLCs follow well-documented methods and means through NVD or CVSS, detection and vectors at the PLCs require expert weighting and most likely proprietary input [26]. CSRI shows particular promise to the critical infrastructure field since penetration testing is near impossible and simulations are difficult without the hardware in the loop [29]. Detection before shut down is limited within industrial CPS to IT Intrusion Detection Systems (IDSs) that are built to overcome the unique aspects within industrial networks [23]. Even with research progressing to better characterize the risk statically and dynamically present in industrial CPS, there are no open-source rating systems in circulation, though cybersecurity companies specializing in control systems are starting to use them to better define current risk and prioritize defensive actions. While a SCADA risk index has potential for use within the UAV community, the lack of an operational open-source index, the smaller scale of systems, and the shorter lifespan of systems reduce direct applicability to sUAVs.

D. *Cybersecurity Insurance*

As a growing variation of quantitative cyber risk, insurance policies have been diverting some of the risk of exploitation since 1997 when the Internet use globally was only 1.7% of the population [30]. Insurance companies function on a strategy of taking premiums upfront to cover the risk of failure in the future and spread out the cost for the user, whether for disaster, health care, or cyber attack. The Internet has since exploded in size with the total cyber insurance market estimated at $3 to 3.5 billion in 2017 [31], with cyber crimes costing the global economy an estimated $450 billion in 2016 [32]. The companies that issued cyber insurance premiums totaling $1.35 billion in 2016 [33] did so based more on an abstract perception of risk due to a lack of historical data to determine probability and actual monetary damage for previous attacks, especially when the damage is information theft or leakage [34]. The most common and simple equation for insurance is based on the historical average of cost per incident times the probability of incident in the near future [35], which requires the very information that is lacking or obscured for cyber incidents. To reconcile this discrepancy in information, several research models have been developed to validate insurance investment, though fewer have published methods of quantitative risk indexes. Research suggests that cyber insurance is feasible and a positive for security, as long as the premiums charged are tied directly to self-protection strategies employed by the organization [36]. For quantifying this risk versus protections, the largest issue is not previous historical data which will continue to grow over time, but mapping all possible attack vectors in the insured system which requires knowledge of all locations of valuable information and employee accesses and habits [37].

The most promising method to grasp the state of a computer network from the cyber insurance industry is presented by the Cyber Risk Scoring and Mitigation tool (CRISM) which operates continuously as a specially designed IDS [35]. CRISM is designed for IT networks where CVSS and NVD provide comprehensive insight to network vulnerabilities and usage. Inspired by automotive driver insurance programs, users voluntarily install a small device to provide additional operational information to the insurance company for the promise of lower premiums. As shown in Figure 4, CRISM has five phases.

1) Mapping: The first step of CRISM is static analysis of the targeted system to determine all components and links with all currently reported vulnerabilities. This mapping phase consists of determining the data and control links (if different) at a physical and protocol layer, operating system of both ground station and UAV, avionic and embedded systems controlling the UAV, and environment that the UAV lives in for connections and external (not necessarily adversary) radio waves.

2) Vulnerabilities: With all of the mapping laid out statically, the vulnerabilities that are known across all components are then expounded. At the communication links, vulnerabilities can consist of protocol flaws, susceptibility to jamming, and leakage of information. At the OS component, vulnerabilities are better laid out via CVSS and NVD such that the software and hardware vulnerabilities are better reported. The navigation vulnerabilities are based on the probability of false signals being accepted and the combination of sensors relied on reduces risk. Sensors such as Inertial Navigational System (INS) that are much more difficult to spoof than GPS reduce the cyber risk of system, but only if properly checked by the autopilot and the programmed failure state.

3) Attack Vectors: With the mapping and tabulation of known vulnerabilities, attack vectors can be determined by common methods through the entire system and the probability of attacks can be estimated. Attack vectors can be initialized only at input ports, whether on ground station or UAV. Vectors are trimmed by forward progress and ability to cause an effect on the mission.

4) Bayesian Network (BN) Graphs: Bayesian networks are then utilized to build out each vector across nodes to determine probability of forward progress and exploitation probability either through probabilities chosen by the organization or experts in the field.

5) Scoring: Lastly, scoring is completed by tabulating the probabilities of exploitation and its effect to the mission. CVSS does present a usable index for consumers and manufactures, however, it is a vulnerability severity assessment and not a direct correlation to risk indexing.

The ability to add an IDS to a Commercial Off The Shelf (COTS) UAV network is non-trivial due to size-weight constraints, mobile ad hoc network transients, and warranty issues arising from user “tampering”. Due to the light and mobile nature of UAV networks, this device or application

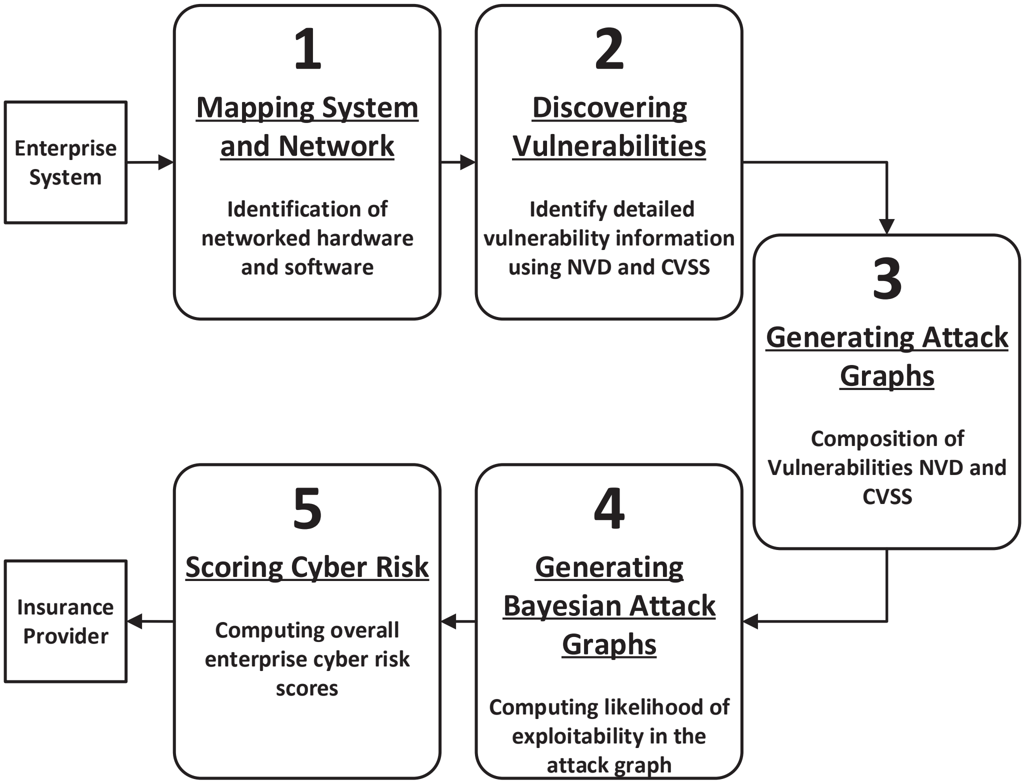

Enterprise System
1
Mapping System and Network
Identification of networked hardware and software
2
Discovering Vulnerabilities
Identify detailed vulnerability information using NVD and CVSS
3
Generating Attack Graphs
Composition of Vulnerabilities NVD and CVSS
4
Generating Bayesian Attack Graphs
Computing likelihood of exploitability in the attack graph
5
Scoring Cyber Risk
Computing overall enterprise cyber risk scores
Insurance Provider

Fig. 4. Five phases of the Cyber Risk Scoring and Mitigation (CRISM) tool [35].

can not be stationary or the network will leave the protected area. Due to the proprietary nature of the majority of UAV manufacturers, tampering or augmenting the device will have secondary consequences that users may be unwilling to accept. Though most UAVs are unable to have an IDS attached or installed, an IDS application installed via hardware or software to the ground controller is a possibility. While CRISM can not be directly applied to UAV risk scoring without the IDS component (which currently does not exist and may not be feasible), their analytic model is very promising in its flexibility to include varying components.

III. Methodology

A difficult problem of risk assessment analysis is determining tool accuracy without historical use data to support it, of interest in this effort because not one of the presented risk assessments cover UAVs. Though accuracy may be the prime measurement, there are other measurements of value to consider which may aid in the process of determining an initial tool deployment until historical data can be realized. Three areas of comparison between these fields of risk assessment that are generally recognized as core to determining viability are as follows: usability, cost, and ease-of-understanding [38]. Unlike accuracy, it is important to note that these measures are qualitative and more prone to variability between observers, but not without value since differences can still be observed and compared.

The first measure, usability, is the measure of how well tailored a tool is to the value it measures. For these risk assessments, this means more specifically how well does the tool measure the risk of small UAVs. To break this down further, usability will be represented by traits of required expertise, flexibility to modifications, and network/device risk coverage. All risk assessments require the user to have some knowledge of the system being measured; however, if the tool uses information on the system that is more abstract or easier to access, then more users in the life-cycle would be able to use the tool. This measure is key if a risk assessment is to be used before operational employment, since not all operational details may be determined. Also the lower the required expertise needed to run the risk assessment, the more likely and more often the tool can be utilized as high expertise users are likely to be rare within any size organization. The second of the three sub-metrics of usability is flexibility. As with almost all computers, components are commonly rearranged and upgraded over time, which changes the cybersecurity risk of the system. Flexibility of the risk assessment to changes in the system and the ability to incorporate non-standard configurations is crucial to properly measuring the risk. The last of the usability sub-metrics is coverage, meaning the ability to measure the entire system for risk. A common saying is that “a chain is only as strong as the weakest link”, and this holds true for computer networks where attackers are smart and rewarded for utilizing the path of least resistance to their target. While it may seem obvious that a risk assessment covers the entire system, history with risk tools has shown that smaller devices or components are commonly ignored even though they present a valuable node in the network [39].

Cost, the second metric, is the measure of how expensive (time and money) the tool is to run. Time is particularly important to smaller and more mobile devices like UAVs since the value of the measurement only lasts until something changes in the system. The monetary cost is also important since it determines the likelihood of an organization actually completing the test and the rate of re-assessing. Monetary cost can be incurred in a variety of methods, the most common being through additional devices to complete the measure and the level of expertise required to assess the measurements.

The last metric to be considered in this paper is that of readability or ease-of-understanding. Outside all of the prior metrics, the assessment needs to be easily communicated afterward to invested parties, such as supervising and regulating entities. Complication of readability can take the form of being too complex where it is impossible to compare systems or too simple where every system appears to have the same value. Users are better able to ingest a rating if it follows a form that they have seen before, such as a star rating or a percent value. Accreditation similar to the European and American automobile safety assessments, which use a number of stars to describe and compare the intrinsic safety quality for the vehicle, would be desirable. All of these criteria should provide a more detailed view into the described domains before determining applicability.

Each of the previously described operational domains use their designated risk assessments simply because they work, to some measure, for their devices. These tools meet an understood baseline that they are effective for their networks, but fall short when sUAVs are the subject. Any assessment that could be applied to sUAVs, but does not have the potential to properly rate the risk for these devices, is rated “Yellow” per category. It is possible for a tool to fall below this “Yellow” baseline and miss key components for a sUAV risk assessment tool, which would then be rated “Red”. This “Red” rating means that significant changes are required to even initialize this tool to rate the risk of sUAVs. In the opposite manner, assessments that properly account for sUAV characteristics and calculate its system’s risk on par with that domain’s specific devices are to be labelled “Green”. A “Green” rating is not to insinuate that all sUAV risk is completely accounted for, but that the tool reaches its own performance baseline with UAVs also. From Section II, it is expected that no assessment will reach “Green” across all or even most metrics since each showed significant failures in applicability to sUAVs.

IV. Analysis

As seen from the build out of other markets’ rating systems, the validity of the rating is based on how holistic the system is examined. The layout of components and a cybersecurity risk index for sUAVs requires additional consideration for adjacent devices and networks plus the environment that the device is operating in since sUAVs are mobile. The environment for UAVs is defined as the system mission and the operational terrain, unlike traditional IT where environment is only the aspects that affect the digital access to a system such as the boundary design. The data link itself may be secure, but consideration for the country, locale, or altitude may change collision rate or noise on the channel and thus effect security. With swarm research as a far end of inter-connectivity of a sUAV, these flying computers use wireless communications that broadcast over the open air to connect to their ground station and to other UAVs. A rating needs to include some factor of the security of these other devices and the connection protocol that allows communications, especially if another ground station or UAV can gain operational control.

Table II shows analyzed applicability of each cybersecurity field to sUAV characteristics, if directly applied as described.

TABLE II. Assessment Applicability to Small UAVs.

	Expertise	Flexibility	Coverage	Cost	Readability
URAF	Green	Red	Red	Yellow	Yellow
CVSS	Yellow	Red	Yellow	Yellow	Green
CSRI	Yellow	Red	Green	Yellow	Red
CRISM	Yellow	Yellow	Yellow	Red	Green

NASA’s URAF shows promise to applicability to sUAVs in terms of expertise required to complete the assessment, given that the Bayesian networks and density maps are pre-populated. Given the working UTM integrating with sUAVs, a real-time assessment of the operational risk should be calculable without much human involvement. However, it is an operational risk assessment that is “device agnostic” so its flexibility and coverage are particularly lacking in assessing cyber risk. Its implementation is expensive since it requires a multitude of ground sensors and manufactures to upgrade models, but this is a requirement of all aircraft so it is not worse applying it to sUAVs. In terms of readability, URAF uses a probability of accident as its score, which may be somewhat easy to use, but communicating the cyber risk is more difficult to tease out.

FiRST’s CVSS provides a scoring system that has been tested and refined for a decade, but fails to assess the key aspects of risk and sUAVs. The tool’s assessment of IT vulnerability severity has been a boon at the enterprise level to prioritize defenses and plan for future improvements. To be applied to sUAV networks, the tool would need to be updated to reflect first the characteristics and market of sUAVs, such as modifications and time in use. In addition, CVSS has explicitly defined themselves away from risk assessment for their own reasons, so the tool would also need to be updated from just severity, or the cost variable of risk, to risk in general. Incorporating likelihood is not easy. However, without it, risk frameworks are unable to compare risk and direct appropriate action.

The proposed risk assessment to CSF, the CSRI, generally misses the goal of a sUAV risk assessment more than the other fields due to its focus on critical infrastructure. While the intent to include CSRI was to observe its ability to incorporate cyber-physical systems and wide area networks of smaller devices, the field of critical infrastructure is inflexible and very slow to change. Nuclear power plants, as the design impetus to CSRI, measure their lifespans in decades and require extreme bureaucratic processes to update networks for fear of network failure or compromise. The ability to fully map out all components and sensors to all system states, cyber and physical, is possible and most likely beneficial, but time and expertise consuming beyond the average user. Corporations may have the expertise and the desire to define their risk minutely, but development and acquisition move too fast for these businesses to stay competitive. To be molded as a sUAV risk assessment, CSRI would require direction to the most important components and provide accurate statistics for the Bayesian networks. A method of rectifying this may be to keep a living document accessible to the public, containing the Bayesian networks for common modifications to configuration and payload.

Lastly and also from the research community, CRISM presented an approach to correct for CSRI's last mentioned failure, adaptation to modifications. By inserting an IDS into a network, a real-time calculation similar to NASA's UTM can be attempted for risk, and unlike UTM, specifically to cyber risk. CRISM suffers from the same restrictions with its Bayesian networks as CSRI, in that likelihood statistics are currently lacking and would need to be provided to the consumer, whether at the acquisition or operational stage. While covering for the flexibility to modifications by tracking live traffic, CRISM lacks the coverage that UTM is building by attaching to the NAS. Without national coverage by regulation, individuals would need to insert the IDS into the sUAV network, which can be difficult due to the mobile and ad hoc nature of sUAVs. While corporations or governments may be willing to cover the additional cost to reduce risk via insurance, it is unlikely individuals will as insurance has not been made viable yet.

V. Conclusion and Future Work

No assessment properly calculated the cyber risk present within a sUAV and all related domains presented in this paper require significant working to be used. Of all of the related domains, CVSS by FiRST appears to have the closest ties to sUAV cyber risk through its presentation of an operational scoring system used by cyber professionals. Though CVSS is no longer defined as a risk assessment, the system was built as one and continues to provide the significant input of severity to the risk imposed by the system on the network. It is conceivable that by updating the definitions of the sub-metrics of CVSS version 3.1 to define sUAV networks over IT components, a new standalone cyber risk assessment may be possible and presented to consumers to more intentionally purchase sUAVs in accordance with their risk frameworks.

This is not to presume that the other assessments in this paper are incapable of adaptation, as described in their analysis. NASA's UTM currently treats sUAVs as indistinguishable from a cyber perspective, which is simply incorrect seeing the wide differences in manufacturers, components, and payloads. Adjusting for cyber-related sensor measurements may be simple enough, however the regulation process will require decades until adoption. Critical infrastructure's CSRI has the most adaptation required as the process to apply to unique sUAVs is significant and the Bayesian statistics required are mostly unknown or unproven. Lastly, the insurance industry's CRISM follows NASA's model with a focus on cyber risk over operational risk, but does not have the backing, funding, or maturity that UTM currently has in the field. National coverage in calculating cyber risk per vehicle would provide unique insights and feedback to corporations to use in their risk framework, but misses the opportunity to provide these inputs at development and acquisition life-cycle phases where they can provide the most effective change.

Future work in the field of sUAV risk assessment requires the building of a quantitative equation for the flying devices or the adaptation from a parallel assessment, as discussed at length in this research. The strongest potential seems to be qualitative characteristics given numerical value and weight, as seen with CVSS and meeting the initial objectives of usability, cost, and ease-of-understanding. Analytical scoring of a sampling of UAVs when paired with missions and environments then would provide validity to the assessment. It is unknown at this time if an analytical-only scoring would provide the best results in light of highly proprietary brands dominating the market and focusing risk assessment at the earliest stages of a system's life-cycle. To focus at the operational stage, a CRISM-like adaptation may be better suited, though the model needs adaptation followed by validation through live testing on hardware in the loop simulation and then networked UAVs. Hardware in the loop is vital to simulations with UAVs due to the physical responses of the system to cyber effects, without which many of the detection methods of cyber-to-cyber and cyber-to-physical attacks are lost. Even with an IDS for UAVs, the Bayesian models would still need to be created for UAVs over the traditional networks and validated to historical data.

Scoring, at this point, is more for internal comparison, but the future expectation is to provide a medium for consumers to easily compare similar sUAVs and influence the manufacturers with their purchases. By providing a single metric per model, mission, and expected environment, the buyer may be better informed based on their individual level of risk acceptance or risk framework, which may be still further offset by insurance premiums. However, until a risk assessment becomes accredited, consumers will be reliant on manufacturer's advertisements and limited personal expertise to compare the risk being introduced to their mission sets. While this trust may be enough for lesser priority missions, the major countries of manufacturing for sUAVs have shown repeated violation of trust and security, and consuming organizations still lack a formal method to utilize their own cyber risk frameworks with sUAV inventories.

Disclaimer: The views expressed in this paper are those of the authors and do not necessarily reflect the official policy or position of the Air Force, the Department of Defense, or the U.S. Government. PA Case Number: 88ABW-2020-0189.

References

[1] D. Pettit, S. Graham, and R. Dill, “Zero Stars: Analysis of Cybersecurity Risk of Small COTS UAVs,” *International Conference on Emerging Security Information, Systems and Technologies, Conference Proceedings*, 2019.

[2] P. G. Fahlstrom and T. J. Gleason, “History and Overview,” in *Introduction to UAV Systems*, 4th ed. West Sussex, United Kingdom: John Wiley Sons, Ltd, 2012, pp. 3–31.

[3] A. Roder, K.-K. R. Choo, and N.-A. Le-Khac, “Unmanned Aerial Vehicle Forensic Investigation Process: DJI Phantom 3 Drone As A Case Study,” *Digital Investigations*, pp. 1–14, 2018. [Online]. Available: arxiv.org/abs/1804.08649

[4] K. Hartmann and K. Giles, “UAV exploitation: A new domain for cyber power,” *International Conference on Cyber Conflict, CYCON*, vol. 2016-Augus, pp. 205–221, 2016.

[5] A. Karp, “Congress to hold UAV safety hearing Oct. 7,” 2015, [Retrieved: September 2019]. [Online]. Available: atwonline.com/government-affairs/congress-hold-uav-safety-hearing-oct-7

[6] Z. Valentak, “Drone Market Share Analysis Predictions for 2018: DJI Dominates, Parrot and Yuneec Slowly Catching Up,” *Drones Globe*, 2017, [Retrieved September 2019]. [Online]. Available: www.dronesglobe.com/news/drone-market-share-analysis-predictions-2018

[7] T. Vuong, A. Filippoupolitis, G. Loukas, and D. Gan, “Physical indicators of cyber attacks against a rescue robot,” *2014 IEEE International Conference on Pervasive Computing and Communication Workshops*, pp. 338–343, 2014.

[8] D. P. Shepard, J. A. Bhatti, T. E. Humphreys, and A. A. Fansler, “Evaluation of Smart Grid and Civilian UAV Vulnerability to GPS Spoofing Attacks,” *Ion Gnss 2012*, pp. 3591–3605, 2012.

[9] T. Reed, J. Geis, and S. Dietrich, “SkyNET: a 3G-enabled mobile attack drone and stealth botmaster,” *Proceedings of the 5th USENIX conference on Offensive technologies (WOOT11)*, p. 4, 2011.

[10] D. Diessner, H. Wynsma, L. Riegle, and P. Morrissey, “Civil Aviation Cybersecurity Industry Assessment & Recommendations, August 2019, Report to the AIA Civil Aviation Council , Civil Aviation Regulatory & Safety Committee AIA Civil Aviation Cybersecurity Subcommittee,” 2019.

[11] P. Kopardekar, “Enabling Civilian Low-altitude Airspace and Unmanned Aerial System Operations by Unmanned Aerial System Traffic Management,” *AUVSI Unmanned Systems 2014*, pp. 1678 – 1683, 2014.

[12] E. Ancel, F. Capristan, J. Foster, and R. Condotta, “Real-time Risk Assessment Framework for Unmanned Aircraft System (UAS) Taffic Management (UTM),” *17th AIAA Aviation Technology, Integration, and Operations Conference*, pp. 1–17, 2017.

[13] P. Kopardekar, “(12) Patent Application Publication (10) Pub. No.: US 2016/0275801 A1,” *United States Patent Applications*, pp. 1–47, 2016.

[14] J. Rios and P. Venkatesan, “NASA UAS Traffic Management National Campaign,” *2016 IEEE/AIAA 35th Digital Avionics Systems Conference*, pp. 1–6, 2016.

[15] L. Barr, R. Newman, E. Ancel, C. Belcastro, J. Foster, J. Evans, and D. Klyde, “Preliminary Risk Assessment Model for Small Unmanned Aerial Systems,” *17th AIAA Aviation Technology, Integration, and Operations Conference*, pp. 1–57, 2017.

[16] F. A. Administration, “Air Traffic Facilities Participating in LAANC,” 2019, accessed December 2019. [Online]. Available: www.faa.gov/uas/programs-partnerships/data-exchange/laanc-facilities

[17] B. Karabacak and I. Sogukpina, “ISRAM: Information Security Risk Analysis Method,” *Computers Security*, vol. 24.2, pp. 147–159, 2005.

[18] M. H. Henry and Y. Y. Haimes., “Comprehensive Network Security Risk Model for Process Control Networks,” *Risk Analysis: An International Journal*, vol. 29.2, pp. 223–248, 2009.

[19] K. Scarfone and P. Mell, “An analysis of CVSS version 2 vulnerability scoring,” in *2009 3rd International Symposium on Empirical Software Engineering and Measurement, ESEM 2009*, 2009, pp. 516–525.

[20] FiRST, “Common Vulnerability Scoring System V3,” 2015, [Retrieved: September 2019]. [Online]. Available: www.first.org/cvss/cvss-v30-specification-v1.8.pdf

[21] A. A. Younis and Y. K. Malaiya, “Comparing and Evaluating CVSS Base Metrics and Microsoft Rating System,” in *Proceedings - 2015 IEEE International Conference on Software Quality, Reliability and Security, QRS 2015*. Institute of Electrical and Electronics Engineers Inc., 2015, pp. 252–261.

[22] S. Zhang, X. Ou, and D. Caragea, “Predicting Cyber Risks through National Vulnerability Database,” *Information Security Journal*, vol. 24, no. 4-6, pp. 194–206, 2015.

[23] R. Mitchell and I.-R. Chen, “A survey of intrusion detection techniques for cyber-physical systems,” *ACM Computing Surveys*, vol. 46, no. 4, pp. 1–29, 2014.

[24] A. J. Chaves, “Increasing Cyber Resiliency of Industrial Control Systems,” *Thesis and Dissertations*, vol. 1563, 2017.

[25] A. Matrosov, E. Rodionov, D. Harley, and J. Malcho, “Stuxnet under the microscope,” *ESET*, 2010.

[26] K. Huang, C. Zhou, Y. C. Tian, S. Yang, and Y. Qin, “Assessing the physical impact of cyberattacks on industrial cyber-physical systems,” *IEEE Transactions on Industrial Electronics*, vol. 65, no. 10, pp. 8153–8162, 2018.

[27] NIST, “Framework for improving critical infrastructure cybersecurity,” *NIST Publications*, vol. 1, no. 1, pp. 1–48, 2018.

[28] S. Haque, M. Keffeler, and T. Atkison, “An Evolutionary Approach of Attack Graphs and Attack Trees: A Survey of Attack Modeling,” in *International Conference on Security and Management*, 2017, pp. 224–229.

[29] J. Shin, H. Son, and G. Heo, “Cyber Security Risk Evaluation of a Nuclear I&C Using BN and ET,” *Nuclear Engineering and Technology*, vol. 49, no. 3, pp. 517–524, 2017.

[30] B. Brown, “The Ever-Evolving Nature of Cyber Coverage,” 2014, [Retrieved: September 2019]. [Online]. Available: www.insurancejournal.com/magazines/mag-features/2014/09/22/340633.htm

[31] C. Stanley, “Cyber Market Estimate,” 2017, interview: 2017-06-26.

[32] L. Graham, “Cybercrime costs the global economy $450 billion: CEO,” 2017, [Retrieved: September 2019]. [Online]. Available: www.cnbc.com/2017/02/07/cybercrime-costs-the-global-economy-450-billion-ceo.html

[33] InsuranceJournal.com, “Cyber Insurance Premium Volume Grew 35% to $1.3 Billion in 2016,” 2017, [Retrieved: September 2019]. [Online]. Available: www.insurancejournal.com/news/national/2017/06/23/455508.htm

[34] J. Yin, “Cyber insurance: Why is the market still largely untapped?” 2015, [Retrieved: September 2019]. [Online]. Available: www.aei.org/publication/cyber-insurance-why-is-the-market-still-largely-untapped

[35] S. Shetty, M. McShane, L. Zhang, J. P. Kesan, C. A. Kamhoua, K. Kwiat, and L. L. Njilla, “Reducing Informational Disadvantages to Improve Cyber Risk Management,” *Geneva Papers on Risk and Insurance: Issues and Practice*, 2018.

[36] J. Bolot and M. Lelarge, “Cyber Insurance as an Incentive for Internet Security,” Tech. Rep.

[37] A. Panou, C. Xenakis, and C. Ntantogian, “RiSKi: A Framework for Modeling Cyber Threats to Estimate Risk for Data Breach Insurance,” *Association for Computing Machinery*, 2017.

[38] I. Stine, M. Rice, S. Dunlap, and J. Pecarina, “A cyber risk scoring system for medical devices,” *International Journal of Critical Infrastructure Protection*, vol. 19, pp. 32–46, 2017. [Online]. Available: doi.org/10.1016/j.ijcip.2017.04.001

[39] A. Pendleton, R. Dill, and D. Pettit, “Surveying the Incorporation of IOT Devices into Cybersecurity Risk Management Frameworks,” *SECUREWARE 2019 Proceedings*, 2019.

End-to-End Secure Application Interactions over Intermediaries on the Example of Power System Communication

Steffen Fries, Rainer Falk
Corporate Technology
Siemens AG
Munich, Germany
e-mail: {steffen.fries|rainer.falk}@siemens.com

***Abstract*—End-to-end security is often a requirement for interacting systems, including energy automation systems. As the term can be interpreted on different layers of the Open System Interconnection (OSI) reference model, it is necessary to clearly define the end points that need to provide or rely on the exchanged data. Connecting client and server applications directly via a transport connection allows the usage of existing security protocols directly, as known from classical Web applications. Typically, Transport Layer Security (TLS) is applied to protect the communication link end-to-end. This approach is utilized in substation automation of energy grids to protect the Transmission Control Protocol (TCP/IP)-based communication between a substation controller and a protection relay applying mutual authentication of the end points. Here, the communicating end points on the application layer terminate in the same entity as the transport layer end points, which essentially provides end-to-end security on a component level. If a direct communication link is not available, communication is realized over an intermediary system. Providing end-to-end security over multiple communication hops, including mutual endpoint authentication (client and a destination application service) as well as integrity and confidentiality of communicated data, deserves specific attention, even if the communication hops with the intermediary are protected hop-by-hop by security protocols like TLS. In power system automation, this kind of communication involving an intermediary is used with publish subscribe protocols, e.g., when integrating Decentralized Energy Resources (DER) or when integrating smart meters in the German Smart Meter Gateway architecture. This paper investigates existing solutions and specifically analyses the end-to-end security approach defined for power system automation within the International Electrotechnical Commission (IEC). A broader application of end-to-end security using session-based communication over intermediaries is desired.**

Keywords—security; device authentication; end-to-end security; multi-hop security; IEC 62351; Publish/Subscribe.

I. INTRODUCTION

Critical Infrastructures (CI) are technical installations that are essential for the daily life of a society and the economy of a country. Examples of CI are provided by technical systems in different application domains like healthcare, telecommunication, transportation, water supply, and power systems. The latter are taken as focus in the context of this paper. In all application domains, there is a clear trend towards increased connectivity and a tighter integration of systems from Information Technology (IT) in common enterprise environments with the Operation Technology (OT) part of the automation systems in the energy and other industrial domains.

This integration enables an enhanced and automated data exchange between industrial systems and IT systems to provide enhanced services. It becomes clear that this integration also requires security measures to avoid negative effects of the formerly isolated OT through control options and to ensure the quality of data provided to the IT for further processing regarding authenticity and integrity but also regarding protection of privacy and potentially know how. Furthermore, this integration also leads to potential physical effects through processing of the provided data. Typically, IT and OT environments have different characteristics in management and operation, which led to distinct domain specific security requirements. This must be considered when designing interconnected cyber-physical systems.

Security in power system communication is getting more momentum [1]. Communication technologies applied in power systems are manifold and comprise, e.g., serial communication in the context of telecontrol. In addition, communication based on the Transport Control Protocol (TCP) is used for monitoring, control, and maintenance of power systems. Multicast Ethernet based communication as further technology is applied in the context of protection relays in substation automation, were real-time capable communication is required. In many scenarios, the security associations established on the transport layer also protect the application layer connection as both terminate at the same entity. But there are scenarios which require multiple consecutive transport connections to exchange application layer data between a sender and a receiver. This paper focuses on the application layer interaction of two entities and the protection of the application data in an ideally transport connection independent manner. The focus is placed on the discussion of secure application layer end-to-end interactions by addressing authenticity, integrity, and confidentiality to ensure reliable control and monitoring of the system.

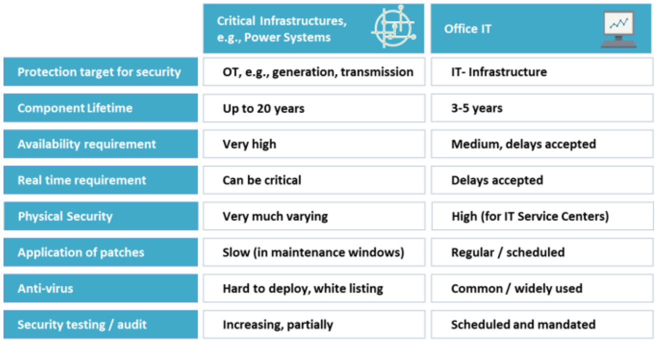
Nevertheless, for the overall system, there are also privacy related considerations that have to be addressed to avoid misuse of the exchanged and collected person-related information. This is obviously necessary for information that can be associated with a single household or a single user, which could be the case for smart meter information, but may also be relevant if provider-based services are used to provide customer-specific information in an online fashion. Although the security discussed here can be leveraged to also address certain privacy properties, privacy specific measures are not in the main focus of this paper.

The remaining part of the paper is structured as follows. Section II provides examples for security requirements for communicating systems, which have been formulated in guidelines and standards or are required by legislation. Section III describes the communication overview of the target scenario and derives high level security requirements to be addressed by specific technical means. These requirements are taken into consideration later in the description of the security approach taken for the integration of Decentralized Energy Resources (DER) into the power system based on IEC 61850. Section IV investigates a selection of existing approaches to provide end-to-end security (message-based and session-based methods). Section V provides more insight into the actual design and application of the protocol defined in IEC 62351-4 to motivate broader application. Section VI provides an evaluation of the investigated application layer security options regarding a derived set of requirements. Section VII concludes the paper with an outlook.

II. Examples of Security Requirements Formulated in Regulation/Guidelines/Standards

Security in communication infrastructures is not a new topic. In office environments or information technology (IT), it is handled as state of the art, and depending on the operational environment certification requirements of specific security processes is mandatory, or at least provides a competitive advantage.

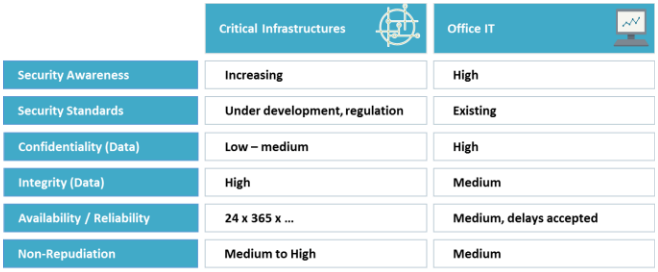
Critical infrastructures or operational technology (OT) on the other hand also rely on communication and utilize increasingly standard communication protocols or standard components whenever possible. This provides some commonalities regarding the utilized technology for communication, but there are distinct differences in the management and operation of these infrastructures as seen in Figure 1.



	Critical Infrastructures, e.g., Power Systems	Office IT
Protection target for security	OT, e.g., generation, transmission	IT- Infrastructure
Component Lifetime	Up to 20 years	3-5 years
Availability requirement	Very high	Medium, delays accepted
Real time requirement	Can be critical	Delays accepted
Physical Security	Very much varying	High (for IT Service Centers)
Application of patches	Slow (in maintenance windows)	Regular / scheduled
Anti-virus	Hard to deploy, white listing	Common / widely used
Security testing / audit	Increasing, partially	Scheduled and mandated

Figure 1. Comparison IT/OT management and operation

These differences in management and operation of the IT systems consequently lead to different high level security requirements as outlined in Figure 2.

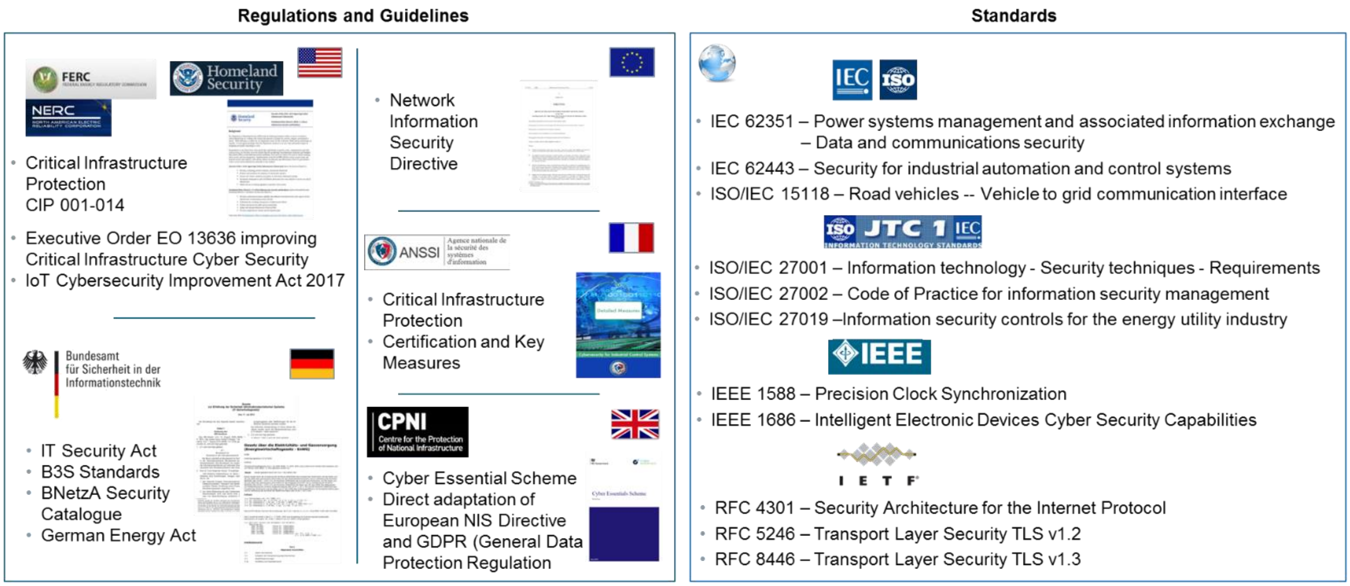


	Critical Infrastructures	Office IT
Security Awareness	Increasing	High
Security Standards	Under development, regulation	Existing
Confidentiality (Data)	Low – medium	High
Integrity (Data)	High	Medium
Availability / Reliability	24 x 365 x …	Medium, delays accepted
Non-Repudiation	Medium to High	Medium

Figure 2. Comparison IT/OT high level security requirements

For critical infrastructures, the European Network and Information System (NIS) Directive [2] requires security measures to be supported by the system operator. This directive has been ratified by the European member states. Germany, for instance, has passed the Information technology (IT) Security Act already in 2015 [3], which requires the definition of domain-specific security standards that have to be implemented by operators of critical infrastructures. For the power system infrastructure, the domain specific security standard is provided by ISO 27019 [4] in conjunction with the IT security catalog of the German BNetzA [5]. Both documents target communication security in terms of authentication of communicating entities in addition to integrity and confidentiality protection of the data exchange, but without specifying specific technical means in terms of security protocols or security mechanisms to be used. A further document to be stated here is the BDEW White Paper [6]. This guideline has been developed by the German Association of Energy and Water Industries ("Bundesverband der Energie- und Wasserwirtschaft" (BDEW), addressing communication security requirements in operations of energy and water utilities. This white paper was one main source for developing ISO 27019.

Security requirements for critical infrastructures are also defined outside Europe, for instance in requirements specified by NIST Cybersecurity framework [7] and specifically for the power system infrastructure by the North American Energy Reliability Council in the NERC Critical Infrastructure Protection (CIP) standards [8]. These documents pose similar requirements, which relate most often to the security processes of an operator and only partly to supporting technology. Common to all requirement documents is that additional standards/specifications are necessary to address the technical implementation of such requirements in components and systems, while ensuring interoperability between different vendor's products. The combination of both, procedural and technical security measures provide the necessary support for reliable operation of critical infrastructure systems.



Regulations and Guidelines

- Critical Infrastructure Protection CIP 001-014
- Executive Order EO 13636 improving Critical Infrastructure Cyber Security
- IoT Cybersecurity Improvement Act 2017

- IT Security Act
- B3S Standards
- BNetzA Security Catalogue
- German Energy Act

- Network Information Security Directive

- Critical Infrastructure Protection
- Certification and Key Measures

- Cyber Essential Scheme
- Direct adaptation of European NIS Directive and GDPR (General Data Protection Regulation

Standards

- IEC 62351 – Power systems management and associated information exchange – Data and communications security
- IEC 62443 – Security for industrial automation and control systems
- ISO/IEC 15118 – Road vehicles -- Vehicle to grid communication interface

- ISO/IEC 27001 – Information technology - Security techniques - Requirements
- ISO/IEC 27002 – Code of Practice for information security management
- ISO/IEC 27019 –Information security controls for the energy utility industry

- IEEE 1588 – Precision Clock Synchronization
- IEEE 1686 – Intelligent Electronic Devices Cyber Security Capabilities

- RFC 4301 – Security Architecture for the Internet Protocol
- RFC 5246 – Transport Layer Security TLS v1.2
- RFC 8446 – Transport Layer Security TLS v1.3

FERC
NERC
Homeland Security
Bundesamt für Sicherheit in der Informationstechnik
ANSSI
CPNI Centre for the Protection of National Infrastructure
IEC
ISO
ISO JTC 1 IEC
IEEE
IETF

Figure 3. Examples for security requirements in regulation and standardization for critical infrastructures as power systems

A standard defining specific technical requirements is provided by the framework IEC 62443 [9]. Beyond other, it describes in two distinct parts technical requirements on system and component level, targeting four different security levels, which relate to the strength of a considered attacker. The framework also addresses also communication security.

Besides these technical requirements, different standards and draft standards exist that address concrete measures for entity authentication, integrity protection, and confidentiality protection on a level ensuring interoperability between different vendors' systems. One example for such a standard protecting specifically TCP/IP based communication is provided by the Transport Layer Security Protocol (TLS 1.2 [10], TLS 1.3 [11]). TLS is a widely used security protocol and most commonly known from the protection of web-based communication, e.g., when accessing a specific web resource. Meanwhile, TLS is applied in further standards to protect domain specific communication protocols.

An example here is the standard ISO 15118 [12], which utilizes TLS to protect the charging related control communication for electric vehicles. A further example is IEC 62351 [16], a framework providing security for data in transit and data at rest in power system automation.

As analyzed in [1] and [13], the necessity to support communication over multiple hops between two entities in power system automation has been emphasized by the support of Decentralized Energy Resources (DER). Integrating DER into the current energy distribution network requires to monitor and control these DER to a similar level as centralized energy generation in power plants to keep the stability of the power network. To cope with the fact that DER are typically operated within a private operator network protected by a firewall, the standard IEC 61850-8-2 [14] defines a communication approach based on the eXtensible Messaging and Presence Protocol – XMPP [15]. Here, both sides, the DER controller, as well as the control center, connect to an intermediate server node, which facilitates the communication between both entities. In this specific case, the standard IEC 62351-4 [16] ensures that the communication between the control center and the DER is secured in an end-to-end fashion. Meanwhile, this standard has been released and will be compared to other existing or currently developed solutions.

The following section elaborates technical means to address these requirements focusing securing communication in an end-to-end fashion.

III. Communication Architecture and Derivation of Technical Security requirements

The discussion of requirements and matching security features and solutions is best done on a concrete use case. Examples for multi-hop communication in power system automation are provided by the integration of DER into distribution networks, the integration of smart meters into a meter data management solution or the connectivity to cloud services providing enhanced data services. Common to all of them is that an intermediary is necessary to support interconnection by providing a rendezvous functionality.

A. *Communication architecture*

For the discussion of end-to-end communication, the integration of DER resources into a power system control network is taken as example, see Figure 4. The lower part of the figure shows the distributed power generators, which may be photovoltaic systems or wind power systems. These are managed by the control function shown in the upper part as control center. The control function may be located at a Distribution Network Operator, a virtual power plant operator, or at a smart energy market operator. All entities are connected via a communication network in which the intermediary XMPP server in the middle provides the connectivity between the control center and the DER controller. All entities essentially work as XMPP clients connecting to the XMPP server, working as dispatcher by facilitating the data exchange between the different XMPP clients. In addition, each XMPP client has a specific functionality from an application perspective. The DER resembles a server, providing power infeed into the distribution network, and provides information by regularly publishing generated power values. The control center in turn works as application client, consuming the generation values to generate a system wide view. Besides the monitoring of generation, the control center may also provide information to the DER devices to control the infeed into the distribution network. For this, the same communication channel is used.

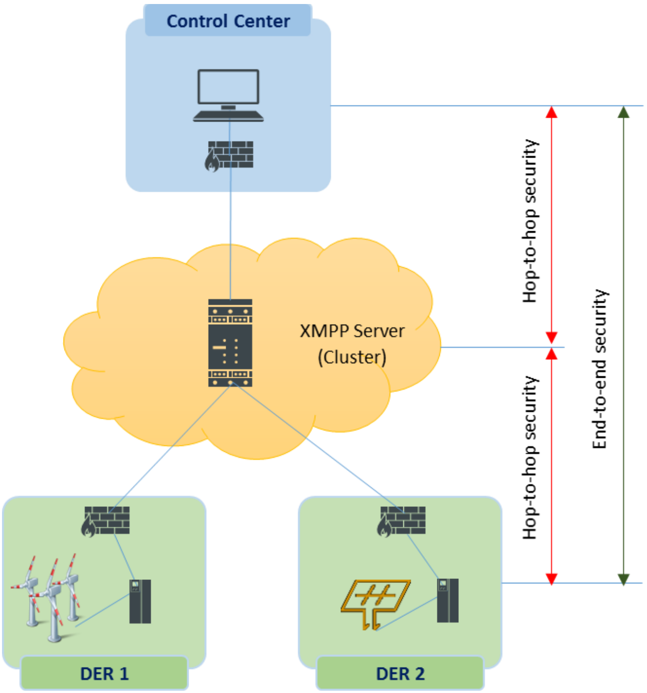


Figure 4. DER Integration based on IEC 61850 over XMPP

The data exchanged between the DER controller and the control center comprises different types of data:

- Customer data, which may be identification information, location data, consumption data or other information belonging to the DER owner.
- Control data, which may be either commands issued by the control center, or event and monitoring information from the DER controller.
- Market data, which may be tariff information provided from a marketplace via the control center or directly (not shown in Figure 1) to the DER controller.

In the context of utilizing IEC 61850 to connect DER to a control center, the communication between the DER controller and the XMPP server is secured using TLS as transport layer security protocol. The same holds for the connection between the control center and the XMPP server. Note that the XMPP server may belong to a different administrative domain and may therefore not be trusted to access the data exchanged between the DER controller and the control center. Hence, the communication relation between the DER controller and the control center is secured at application layer using IEC 62351-4, which will be analyzed in more detail in Section V.

B. *Derivation of Technical Security Requirements*

As stated in the introduction, there are different types of security requirements stemming, on one hand, from the obligation to comply with international and national regulations. On the other hand, security requirements are derived from the system architecture based on a risk-based approach. The international industrial security standard IEC 62443 [9] is a security requirements framework jointly developed by the International Electrotechnical Commission (IEC) and the International Society of Automation (ISA99) to address the need to design cybersecurity robustness and resilience into Industrial Automation and Control Systems (IACS). The standard covers both organizational and technical aspects of security over the life cycle of systems. It can be used in conjunction with ISO/IEC 27019 (the Information Security Management System (ISMS) profile for the energy domain based on ISO 27002) and with IEC 62351, providing specific security solutions. Here, the parts IEC 62443-3-3 (focus on system security requirements) and IEC 62443-4-2 (focus on component security requirements) can be used in the context of a risk-based approach, as they specify technical security requirements for four security levels, corresponding to different strengths of an attacker. For both views, system and component, foundational requirements groups have been defined. For each of the foundational requirements, several concrete technical Security Requirements (SR) and Requirement Enhancements (RE) to address a specific security level exist.

The overall approach applies to the systems and the communication connections are shown in Figure 4. In the context of this paper, the focus is placed on the communication relations, to address the specific target of providing communication security over potentially untrusted nodes. The protection of the communication is addressed by different security requirements focusing on end-to-end security and hop-to-hop security. Note that the hop-to-hop security requirements contribute to the overall system security approach and may be used in conjunction with the end-to-end security. Note that the end-to-end security is intended to be independent of the hop-to-hop security as the endpoints may not have control about the hop-to-hop security setup.

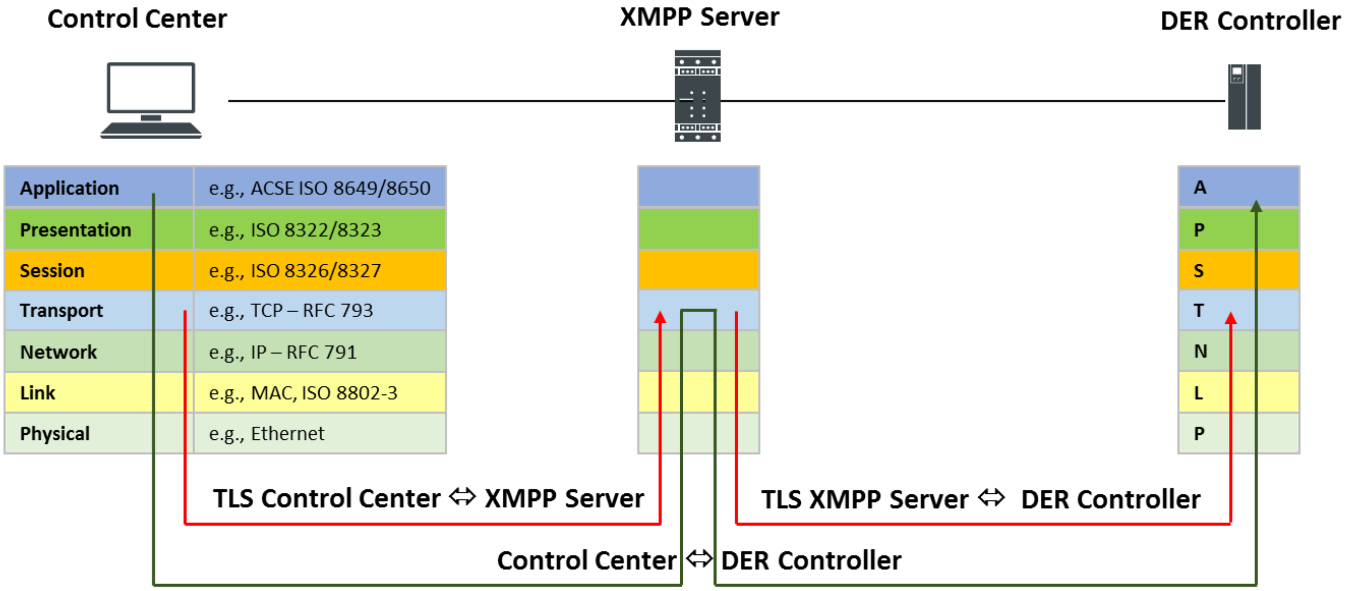


Figure 5. End-to-end-Security and hop-by-hop security according to IEC 62351-4

Figure 5 shows the data exchange between the control center and the DER controller via the XMPP server. The security requirements comprise specifically:

[R 1] End-to-end authentication between the DER controller and the control center to ensure identification and authentication of the communicating endpoints.

[R 2] End-to-end integrity protection to ensure that data in transit has not been tampered with (unauthorized modification) between the DER controller and the control center.

[R 3] End-to-end confidentiality protection to ensure that data in transit has not been accessed (read) in an unauthorized way by the XMPP server. Note that this requirement may not be generally applicable. Use cases exist in which intermediaries need to access the transmitted information. For these use cases the requirements [R 1] and [R 2] may be sufficient.

Hop-to-hop authentication between the XMPP client (DER controller, control center) and the XMPP server is used to identify and authenticate an intermediary system proxying the end-to-end communication between the DER controller and the control center.

IV. Security Measures on Application Layer

This section investigates a selection of existing end-to-end security approaches, which can be used to provide authentication, integrity, and confidentiality. Note that XMPP enhancements to achieve end-to-end security between the clients connected via the XMPP server have already been discussed as part of [13]. The IETF originated drafts discussed in this paper are already outdated and have not been updated in the last years. Therefore, they are not considered further here.

In the following examples of existing standards or standards in development supporting end-to-end security on application layer, are summarized. They are distinguished into message-based approaches and session-based approaches. Message-based approaches are independent of the actual communication session and can be applied to single messages. They typically rely on security credentials, which are setup out of band. These security credentials are applied to the messages directly. Session-based approaches rely on a communication connection, which comprises at least an initialization phase setting up security credentials to be used in the established session only and a data exchange phase. The establishment of the session related security credentials may be bound to long term security credentials of the respective entities. Both approaches have their merits, but also certain drawbacks.

A. *Message-based security*

The following examples target the protection of single messages and do not rely on an established communication connection. They utilize existing security credentials to protect the messages. In general, this type of security is best for occasionally exchanged messages but not necessarily for a consistent data exchange or bulk data exchange. All of the provided examples support the requirements [R 1], [R 2], and [R 3]. Note that confidentiality protection [R 3] is optional.

– IETF RFC 3923 [17] describes end-to-end signing and object encryption utilizing S/MIME to protect the messages exchanged over XMPP connections. This approach is similar to using secure email. It provides end-to-end authentication based on a digital signature and confidentiality protection based on symmetric encryption. As this approach targets message-based communication, without a communication session it will result in a higher per message overhead, as the messages are protected using symmetric encryption, while the key for the symmetric encryption is encrypted with the recipient's public key. This approach has two drawbacks. It is performance intensive due to the use of asymmetric operations and it is bound to RSA as asymmetric algorithm. Newer algorithms like ECDSA based on elliptic curves may not be used.

- W3C defined XML security may also be used to address a secure data exchange on application layer. There are two different standards available, which are already utilized to provide security: XML Signatures [18] and XML Encryption [19]. Both can be used in conjunction, ideally on XML encoded data in so-called XML elements and support the given security requirements. XML encryption allows the encryption of any type of data with symmetric and asymmetric methods. XML signature on the other side applies asymmetric methods to achieve integrity protection and non-repudiation. Note that there exist adequate standards for the binary data representation to safe bandwidth during transfer.
- The IETF working group for JavaScript Object Signing and Encryption (JOSE) defined two further standards, which can be used to protect messages encoded in JavaScript Object Notation (JSON). IETF RFC 7515 [20] specifies JSON Web Signatures, while IETF RFC 7516 [21] defines JSON Web Encryption. The combination of both documents is similar to XML documents developed by W3C for specific JSON encoding.
- A further IETF standard is provided with RFC 8152 [22] defining authentication, integrity protection, and confidentiality protection for Concise Binary Object Representation (CBOR), which enhanced the data model of JSON with a binary representation. This approach allows for enveloping and encryption of arbitrary message blocks.

B. *Session-based security*

The following examples target the protection of communication sessions for application data exchanges. For this, it is assumed that a communication session is established between two entities during which both participants can authenticate and negotiate a set of session keys for protecting further communication. This approach has the advantage for consecutive communication to result in less overhead for the bulk data handling as part of the communication session. This is due to the fact that the combination of symmetric encryption and an additional integrity protection or the direct application of authenticated encryption has a much better performance instead of invoking asymmetric cryptography on a per packet base.

- An IETF standard focusing on object security is RFC 8613. It defines a method for application-layer protection of the Constrained Application Protocol (CoAP), using CBOR Object Signing and Encryption (COSE) called Object Security for Constrained RESTful Environments (OSCOR). This standard defines that client and server establish a shared security context used to process COSE objects. It utilizes pre-shared keys (PSK) for the security context, which are expected to be established out of band or by a different key management protocol. Therefore [R 1] is met with restrictions. For the object protection OSCOR builds on Authenticated Encryption with Associated Data (AEAD). This has to be kept in mind, as it therefore always addresses [R 2] and [R 3].
- IETF draft on Application Layer TLS [24] leverages the existence of a TLS implementation on the communicating entities. The approach utilizes the option of TLS stacks to create and process TLS records based on access to the byte buffer. Based on this, the TLS packets may be transmitted over arbitrary transport connections. The draft targets two different application scenarios, as there is the transport over non-IP networks like Zigbee and the transport over IP based networks. This approach has the advantage that the application layer security immediately benefits from new cipher suites and cryptographic algorithm support by the underlying TLS stack. In addition, several TLS stacks allow key material export using the approach defined in IETF RFC 5705 [25] to leverage the TLS key agreement and to utilize the negotiated key in the context of other protocols. Essentially, ATLS copes with all of the requirements [R 1], [R 2], and [R 3]. Note that when used with TLS 1.3, ATLS will always provide end-to-end confidentiality protected transport.
- Off-the-Record (OTR) [26] is a protocol developed for messenger applications to ensure integrity and confidentiality and most notably plausible deniability. Starting from version 2 of the protocol, peer authentication is also supported. Here, shared keys are utilized to achieve the authentication. The development stopped in 2016. OTR directly addresses the requirements [R 2] and [R 3].
- Signal [27] is another protocol used in messaging systems. It is based on OTR and allows to establish a secure session based on an authenticated triple Diffie Hellman key agreement in which EdDSA signatures are employed for integrity protection during the key establishment phase. The negotiated key material is applied to protect the integrity and confidentiality of the established session based on the Double Ratchet algorithm. It ensures ongoing renewal and maintenance of short-lived session keys. Note that peer authentication is not directly supported by signal. Note also that Signal supports plausible deniability, which

may not be desired in industrial environments to be able to ensure an audit trail. Signal therefore focuses on the requirements [R 2] and [R 3].

- Application Layer Transport Security (ATLS) [28] has been developed by Google in 2017 and is utilized to secure Remote Procedure Calls (RPC). The protocol is defined in a similar way as TLS, consisting of a handshake protocol and a record protocol. It allows for mutual authentication and session integrity and confidentiality. Authentication is bound to an entity rather than an instance (e.g., hostname) as the approach targets mainly cloud environments. Note that there are tradeoffs to TLS described in the specification [28], which relate to privacy concerns for the handshake messages and perfect forward secrecy. Note that these properties are supported out of the box in TLS 1.3, but not in TLS 1.2 and below. ALTS directly addresses the requirements [R 1], [R 2], and [R 3].

V. End-to-End security design in IEC 62351-4

The security requirements derived in Section III.B for providing application layer end-to-end security supporting DER integration are reviewed and enhanced to better address the target scenario to:

[R' 1] Peer authentication between the DER controller and the control center (mutual authentication) based on X.509 certificates.

[R' 2] Integrity protection of exchanged data to ensure that data in transit has not been tampered with.

[R' 3] Optionally, confidentiality protection to ensure that an intermediary cannot access the content of the data exchange. The reason for handling this requirement as optional is based on the necessity in some deployment scenarios that at the security perimeter an inspection of the data may be required. By allowing a mutual authenticated and integrity protected communication connection, the communication may be monitored, e.g., if the control commands cope match a certain system state or to support an audit trail.

[R' 4] Session key management supporting initial key agreement providing perfect forward secrecy (PFS) as well as key update.

Note that it should be possible to use either distinct algorithms for integrity and confidentiality or a combined approach (authenticated encryption). This in general is supported supporting cryptographic agility in the protocol to allow the application of different cryptographic algorithms. Note also that the endpoints typically have no guarantees about what level of transport layer security is enforced along the communication path with multiple hops.

A. Design rational

The design of the final solution specified in IEC 62351-4 already started in 2014. Not all of the security approaches depicted in Section IV were available at this time, but the concept of message-based security and session-based security was defined and applied. The available message-based and session-based approaches were seen to not match the refined requirements in an optimal way. Message based approaches were ruled out as they come with increased processing overhead for a consistent communication connection due to employment of asymmetric key material on a per message base. From the session-based approaches, the messenger-based solutions cannot be applied in industrial communication as they do not provide the necessary means for peer authentication. From the remaining approaches, ATLS would be the closest one from a functionality point of view as it provides an application protocol and transport protocol independent solution. The development of ATLS begun in 2017 and is still an individual draft in the IETF. Moreover, ATLS requires the existence of a TLS implementation on the communication peers.

Based on the review of the existing solutions and the requirements posed for power systems an own solution was seen necessary and the solution was designed based on approaches taken in the design of TLS. This development lead to an update of the standard IEC 62351-4 targeting also multi-hop communication in 2018. The standard meanwhile specifies a transport security profile and an application security profile. The application security targets the provisioning of end-to-end security, as outlined by the requirements above. The following subsections describe the technical preconditions, the session handling, and the packet construction of the protocol.

B. Precondition

The involved endpoints are expected to possess a X.509 certificate and corresponding private key as well as a root certificate trusted by both sides (e.g., bound to the operator) and a common set of Diffie Hellman public parameter. These can be part of the system configuration. Based on the peer certificates and the common root certificate the endpoint authentication can be performed. The Diffie Hellman parameter are then used in a key agreement phase to establish a master key for the application layer context.

As the security targets the application layer a protocol is assumed that supports session handling on application layer in terms of at least initiating a session. In the specific example, this is provided by the Manufacturing Message Specification (MMS [29]) using the *MMS Initiate* and *MMS Initiate Response* messages. MMS in turn is used to carry the IEC 61850 payload to monitor and control the DER resources. As the MMS session is initiated by only one roundtrip, followed by IEC 61850 specific exchanges, the security is expected to proceed in one round trip as well, without adding additional message exchanges.

C. Session Handling

The session handling can be distinguished into the initial key agreement during the session initialization, the key usage phase, and the key update phase. The sequences for the key agreement phase and the key update phase are shown in Figure 6. The key usage phase is neglected, as the application of the negotiated key set is straight forward complying with the agreed cryptographic algorithms for integrity protection and optionally confidentiality protection.

At the beginning of the session, both sides generate a Diffie Hellman key pair to be used in the key agreement resulting in an ephemeral Diffie-Hellman secret. All data necessary for the establishment of the security association between both peers are kept in a data structure called clear token (as the data is transmitted in clear, but integrity protected).

```
ClearToken1::= SEQUENCE {
  sigAlg       AlgorithmIdentifier,
  version      Version DEFAULT {v1},
  assoID       AssoID,
  dHKey        DiffieHellmanSet,
  hmac         ALGORITHM.&id,
  time         TimeStamp,
  encr-mode    CHOICE {
    aea           SET SIZE (1..MAX) OF
      oid         ALGORITHM.&id,
    non-aea       SEQUENCE {
      encr          [0] SET SIZE (1..MAX) OF oid,
      icvAlgID      [1] SET SIZE (1..MAX) Of oid,
    ... },
  confParams   ConfidentialityParms,
  pkCert       PKCert,
  certPath     CertPath OPTIONAL,
  attCert      ACert OPTIONAL,
  ... }
```

Figure 6. Clear token (*ClearToken1*) for key establishment (simplified)

Figure 6 shows the clear token used during connection establishment. Besides the parameter for the session key establishment like the Diffie Hellman values and used certificates also session related information like algorithm identifiers for integrity protection as well as optional confidentiality protection and synchronization information is contained. In addition, the structure allows to also transport an attribute certificate, which may be used to additionally support attribute-based or role-based access control in conjunction with the authentication. To ensure the integrity of the initial exchange, the messages are cryptographically signed.

From each of the handshake messages a fingerprint is taken using a hash function. For this the following procedure is used. The hash h_A is calculated over the concatenation of the current message and the hash of the previous message (the initial message uses “0” as value of the previous message). This fingerprint is used to ensure the right order of messages and to provide additional randomness to the messages. This randomness bases on the generated Diffie Hellman parameter. Note that the calculated hash is never transmitted over the communication connection and only serves as local additional parameter in the final key derivation. Upon reception of the initiation message, the receiver verifies the signature, calculates the hash over the received message and stores the fingerprint h_A. It then generates the response message, from which again the fingerprint is taken by concatenating the response message with the stored fingerprint h_A to calculate the resulting hash h_B. This “running” hash spanning subsequent messages was inspired by the TLS handshake [10].

After providing the signed response to the initiator, both sides can calculate the Diffie-Hellman secret DH_S and utilize it together with the resulting hash h_B as input for the hash based key derivation function HKDF.

This will generate different keys per direction for integrity protection, and for confidentiality protection, resulting in four keys IK_A and IK_B, and EK_A and EK_B. The keys are applied according to the security association. It is necessary that both peers store the hash value h_B to be used in a later key update.

The key update uses a different clear token *(ClearToken2)*, a more simplified structure, as only a restricted set of parameter needs to be transmitted during the data transfer phase. The key update itself can be performed using a single message.

```
ClearToken2::= SEQUENCE {
  version          Version DEFAULT {v1},
  assoID           AssoID,
  time             TimeStamp,
  seq              SequenceNumber,
  iv          [0]  InitializationVector OPTIONAL,
  rekey       [1]  DhPublicKey OPTIONAL,
  reqRekey    [2]  BOOLEAN DEFAULT FALSE,
  changedKey  [3]  BOOLEAN DEFAULT FALSE,
  ... }
```

Figure 7. Clear token (*ClearToken2*) for key update (simplified)

Figure 8 shows the key update triggered by the control center. As in the initial step, the control center generates a fresh Diffie Hellman key pair and utilizes the already received and stored Diffie-Hellman key from the DER controller to immediately to calculate a new Diffie-Hellman secret DH_{S1} and the resulting set of updated session keys for integrity protection. Once this message is received by the DER controller, it can calculate the updated set of keys.

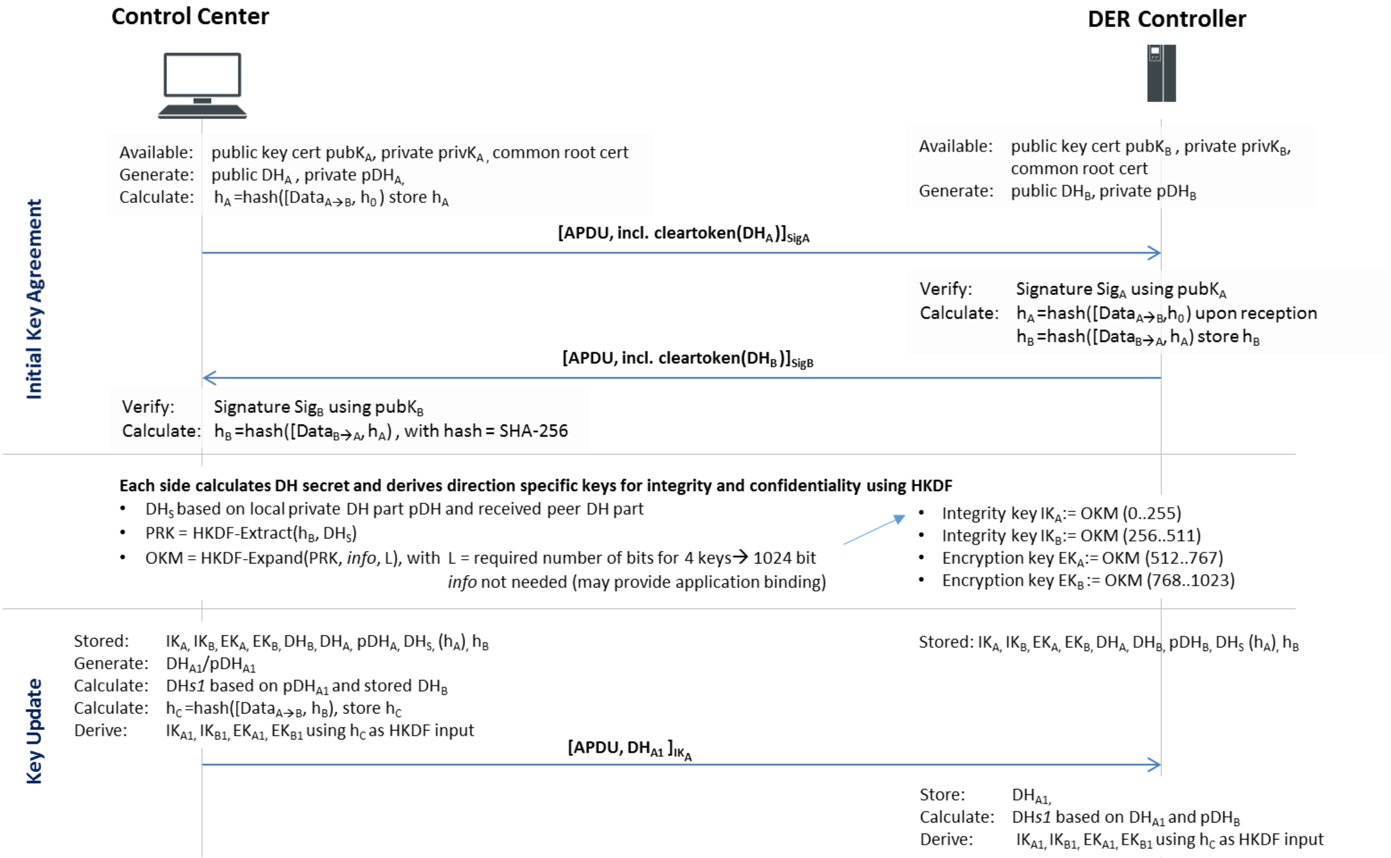

Control Center
DER Controller
Initial Key Agreement
Available: public key cert $pubK_A$, private $privK_A$, common root cert
Generate: public DH_A , private pDH_A,
Calculate: h_A =hash([$Data_{A\rightarrow B}$, h_0) store h_A
Available: public key cert $pubK_B$, private $privK_B$, common root cert
Generate: public DH_B, private pDH_B
[APDU, incl. cleartoken(DH_A)]$_{SigA}$
Verify: Signature Sig_A using $pubK_A$
Calculate: h_A =hash([$Data_{A\rightarrow B}$,h_0) upon reception
h_B =hash([$Data_{B\rightarrow A}$, h_A) store h_B
[APDU, incl. cleartoken(DH_B)]$_{SigB}$
Verify: Signature Sig_B using $pubK_B$
Calculate: h_B =hash([$Data_{B\rightarrow A}$, h_A) , with hash = SHA-256
Each side calculates DH secret and derives direction specific keys for integrity and confidentiality using HKDF
• DH_S based on local private DH part pDH and received peer DH part
• PRK = HKDF-Extract(h_B, DH_S)
• OKM = HKDF-Expand(PRK, info, L), with L = required number of bits for 4 keys→ 1024 bit
info not needed (may provide application binding)
• Integrity key IK_A:= OKM (0..255)
• Integrity key IK_B:= OKM (256..511)
• Encryption key EK_A:= OKM (512..767)
• Encryption key EK_B := OKM (768..1023)
Key Update
Stored: IK_A, IK_B, EK_A, EK_B, DH_B, DH_A, pDH_A, DH_S, (h_A), h_B
Generate: DH_{A1}/pDH_{A1}
Calculate: DHs1 based on pDH_{A1} and stored DH_B
Calculate: h_C =hash([$Data_{A\rightarrow B}$, h_B), store h_C
Derive: IK_{A1}, IK_{B1}, EK_{A1}, EK_{B1} using h_C as HKDF input
Stored: IK_A, IK_B, EK_A, EK_B, DH_A, DH_B, pDH_B, DH_S (h_A), h_B
[APDU, DH_{A1}]$_{IK_A}$
Store: DH_{A1},
Calculate: DHs1 based on DH_{A1} and pDH_B
Derive: IK_{A1}, IK_{B1}, EK_{A1}, EK_{B1} using h_C as HKDF input

Figure 8. End-to-end-Security and hop-by-hop security according to IEC 62351-4

D. Packet construction

Figure 9 shows the packet construction and how the different parts of the messages are protected. Note that the clear token is only integrity protected while the payload of the packet (ADPU in Figure 9). As stated before, the clear token carries all cryptographic parameter necessary to establish the security association.

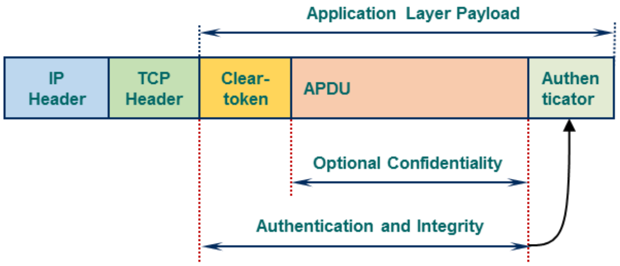


Figure 9. Packet structure of IEC 62351-4 end-to-end security

Although Figure 9 shows the transport over TCP/IP, other transports may be taken. The defined approach has no dependency on the underlying transport protocol. It has to be obeyed that for the session setup one roundtrip is necessary. In the target scenario for embedding DER into the power grid, the *MMS Initiate* and *MMS Initiate Response* message sequence is used to piggyback the secure session establishment. During the session setup, the initial handshake is performed. In the initial setup, the authenticator is provided by invoking the peer certificate and the corresponding private key to calculate a digital signature over the message as indicated in Figure 8 by the *SigA and SigB* indices on the initial handshake messages. For all subsequent messages the authenticator is build using the established session key for integrity protection. Note that the established keys are direction dependent resulting in two keys IK_A and IK_B for the ICV calculation. If confidentiality protection has been negotiated during the initial handshake two additional keys EK_A and EK_B are derived and can be used to encrypt the payload.

VI. EVALUATION

In the following, the different approaches for providing application layer security described in Section IV and Section V are compared regarding their match to the derived requirements [R' 1] to [R' 4].

In addition to the comparison of requirements match, further properties are being investigated. This comprises the effort for the initial handshake and the key update using the notion of Round Trips (RT). Additionally, as the target scenario addresses the integration of DER into the power grid and thus uses longer lasting connections, the potential performance impact based on a qualitative judgement is considered.

TABLE I. EVALUATION OF INVESTIGATED METHODS

Criteria	**Message-based approaches**				**Session-based approaches**					
	IETF RFC 3923 (Sig/Enc)	***XML security (Sig/Enc)***	***IETF JOSE (Sig/Enc)***	***IETF RFC 8152 (Sig/Enc)***	***IETF RFC 8613***	***IETF Draft ATLS***	***OTR***	***Signal***	***Google ALTS***	***IEC 62351-4***
[R' 1]: Peer authentication	X	X	X	X	(based on PSK)	X			X	X
[R' 2]. Integrity protection	X	X	X	X	X	X	X	X	X	X
[R' 3] Optional confidentiality protection	X	X	X	X		(X)	X	X	(X)	X
[R' 4] Mngmt. of session keys					X	X	X	X	X	X
Inital handshakes (using X.509 certificates)	Not applicable	Not applicable	Not applicable	Not applicable	1 RT	TLS 1.2: 2,5 RT TLS 1.3: 2 RT	2 RT	2 RT	2 RT	1 RT
Key Update handshakes	Not applicable	Not applicable	Not applicable	Not applicable	1 RT	TLS 1.2: 2 RT TLS 1.3: 0-1 RT	2 RT	2 RT	1 RT (via sesion resume)	0-0,5 RT
Performance impact	High	High	High	High	Low	Low	Low	Low	Low	Low
Notes	Utilizes RSA only for signatures.	Similar approach availabnle for binary transfer.		Binary transfer	Due to mandatory use of AEAD, no integrity only mode available.	Requires local TLS stack. TLS updates can be applied, but TLS 1.3 is restricted to AEAD.			Supports session resumption. Mandatroy encryption of payload.	No sessison resumption. Session key update with a single message.

For this, it is assumed that asymmetric operations are always applied in message-based approaches, while session-based approaches utilize asymmetric cryptography for a key establishment of a session key, which is used with symmetric crypto algorithms. Note that in the comparison for the key updates, it is stated for ATLS and also for IEC 62351-4, that the update may be performed without additional messages, basically in parallel to the existing data exchange, by stating "0-RT". This leverages the fast that in TLS 1.3 it is possible to send protected communication already in the *ClientHello* message, which can be used in the key update. In IEC 62351-4, the key Update would be signaled in the *ClearToken2* structure, as shown in Figure 7.

Based on the available solutions at the time of starting the specification of IEC 62351-4 in 2015, it was seen that none of existing solutions provides a perfect fit. The message-based approaches were directly ruled out as they have a big influence on the message processing due to the number of necessary asymmetric operations. From the session-based approach, not all of the discussed approaches were available at this time. The ongoing standardization approach of ATLS in the IETF looks promising for applications already utilizing TLS to protect the transport layer communication. Moreover, ATLS directly benefits from updates to the base TLS protocol. Contrary looking at TLS 1.3 integrity only operation will not be supported but may be necessary in power system automation to enable monitoring. IEC-62351-4 on the other hand was tailored to cope with the boundary conditions of the deployment environment resulting in no influence of the target application protocol in terms of additional handshakes. Due to this, the protocol has also less options to be configured or negotiated. This may be beneficial also for other applications as less complexity is often favored. This is visible also in a recently started activity in the IETF by proposing a compact version of TLS 1.3 [30].

VII. CONCLUSIONS

This paper investigates the handling of end-to-end security over intermediate nodes from a system point of view, by investigating existing security requirements and existing solutions. Moreover, the specific use case of incorporating DER into the power grid was taken as main use case for comparing the different approaches. The analysis was divided into message-based approaches and session-based approaches, in which the session-based approaches came out as winner due to the lower performance overhead in long lasting connections. Besides the investigation into existing approaches, the motivation and description of the end-to-end security approach defined in IEC 62351-4 was described.

It establishes an end-to-end security session between two communicating peers with mutual entity authentication resulting in session keys being applied for end-to-end message integrity and confidentiality.

Two points should be obeyed when applying the discussed approach. First, the initial key agreement results in an ephemeral set of session keys, as both sides are expected to generate fresh Diffie Hellman parameters. The key update performed in a single message initiated by either peer results in a semi-static Diffie Hellman key agreement. Depending on the security requirements, the receiver may initiate another key update to ensure the freshness of his Diffie Hellman parameters. The second point relates to potential privacy requirements. The initial key agreement utilizes a clear-text token, which is only integrity protected. Thus, all information contained in the token is potentially readable by an intermediary. As the clear token also contains certificate information, it may allow to identify the communication end points. Applications with similar boundary conditions may leverage this approach in other scenarios or protocol frameworks in industrial communication.

As an outlook to the application of the described approach in IEC 62351-4, it is intended to investigate also the application of other publish-subscribe protocols utilized in automation scenarios like MQTT or AMQP.

In addition to the provided security measures, the application of specific privacy preserving techniques needs to be investigated to specifically address the data exchange with end-user related systems and services to keep their personal relation and data protected.

REFERENCES

[1] S.Fries and R.Falk, “End-to-End Application Security over Intermediaries on the Example of Power System Communication”, Proceedings IARIA Securware 2019, ISBN: 978-1-61208-746-7, pp. 22-27, [Online]. Available from: https://www.thinkmind.org/download.php?articleid=securware_2019_2_10_30063 2020.01.10

[2] European Commission, “The Directive on security of network and information systems (NIS Directive 2016/1148)”, [Online]. Available from: https://eur-lex.europa.eu/eli/dir/2016/1148/oj 2020.01.10

[3] German IT Security Act, official web site (German), [Online]. Available from: https://www.bsi.bund.de/DE/Themen/Industrie_KRITIS/KRITIS/IT-SiG/it_sig_node.html 2020.01.10

[4] ISO 27019: Information technology - Security techniques - Information security controls for the energy utility industry, [Online]. Available from: https://www.iso.org/standard/68091.html 2020.01.10

[5] IT Security Catalog, [Online]. Available from: https://www.bundesnetzagentur.de/ SharedDocs/Downloads/DE/Sachgebiete/Energie/Unternehmen_Institutionen/Versorgungssicherheit/IT_Sicherheit/IT_Sicherheitskatalog_2018.pdf 2020.01.10

[6] BDEW White paper “Requirements for Secure Control and Telecommunication Systems,” BDEW, May 2018, [Online]. Available from: https://www.bdew.de/media/documents/Awh_20180507_OE-BDEW-Whitepaper-Secure-Systems.pdf 2020.01.10

[7] NIST Framework for Improving Critical Infrastructure Cybersecurity, [Online]. Available from: https://nvlpubs.nist.gov/nistpubs/CSWP/NIST.CSWP.04162018.pdf 2020.01.10.

[8] NERC CIP Set of Standards, [Online]. Available from: https://www.nerc.com/pa/Stand/pages/cipstandards.aspx 2020.01.10

[9] IEC 62443, “Industrial Automation and Control System Security” (formerly ISA99), [Online]. Available from: https://www.isa.org/isa99/ 2020.01.10

[10] T. Dierks and E. Rescorla, “Transport Layer Security Protocol version 1.2”, RFC 5246, August 2008, [Online]. Available from: https://tools.ietf.org/html/rfc5246 2020.01.10

[11] E. Rescorla, “Transport Layer Security Protocol version 1.3”, RFC 8446, August 2018, [Online]. Available from: https://tools.ietf.org/html/rfc8446 2020.01.10

[12] ISO/IEC 15118-2:” Road vehicles -Vehicle-to-Grid Communication Interface - Part 2: Network and application protocol requirements”, March 2014, [Online]. Available from: https://www.iec.ch/search/?q=15118, 2020.01.10.

[13] S. Fries, R. Falk, H. Dawidczak, and T. Dufaure, “Decentralized Energy in the Smart Energy Grid and Smart Market – How to master reliable and secure control Secure Integration of DER into Smart Energy Grid and Smart Market,” International Journal of Advances in intelligent Systems, vol. 9 no 1&2, 2016, ISSN: 1942-2679, page 65-75, [Online]. Available from: https://www.thinkmind.org/download.php?articleid=intsys_v9_n12_2016_6 2020.01.10

[14] ISO 61850-x: “Communication networks and systems for power utility automation”, [Online]. Available from: https://www.iec.ch/search/?q=61850 2020.01.10

[15] P. Saint-Andre, ”Extensible Messaging and Presence Protocol (XMPP): Core,” RFC 6120, [Online]. Available from: https://tools.ietf.org/html/rfc6120 2020.01.10

[16] IEC 62351-x Power systems management and associated information exchange – Data and communication security, [Online]. Available from: https://www.iec.ch/search/?q=62351 2020.01.10

[17] P. Saint-Andre, “End-to-End Signing and Object Encryption for the Extensible Messaging and Presence Protocol (XMPP),” RFC 3923, [Online]. Available from: https://tools.ietf.org/html/rfc3923 2020.01.10

[18] W3C: XML Signature Syntax and Processing Version 2.0, June 2015, [Online]. Available from: https://www.w3.org/TR/xmldsig-core2/ 2020.01.10

[19] W3C: XML Encryption Syntax and Processing Version 1.1, April 2013, [Online]. Available from: https://www.w3.org/TR/xmlenc-core1/ 2020.01.10

[20] M. Jones, J. Bradley, and N. Sakimura, “JSON Web Signature (JWS),” RFC 7515, [Online]. Available from: https://tools.ietf.org/html/rfc7515 2020.01.10

[21] M. Jones and J. Hildebrand, “JSON Web Encryption (JWE),” RFC 7516, [Online]. Available from: https://tools.ietf.org/html/rfc7516 2020.01.10

[22] J. Schaad, “CBOR Object Signing and Encryption (COSE),” RFC 8152, [Online]. Available from: https://tools.ietf.org/html/rfc8152 2020.01.10

[23] G. Selander, J. Mattsson, F. Palombini, and L. Seitz, “Object Security for Constrained RESTful Environments (OSCORE)”, RFC 8613, [Online]. Available from: https://tools.ietf.org/html/rfc8613 2020.01.10

[24] O. Friel, R. Barnes, M. Pritikin, H. Tschofenig, and M. Baugher, “Application layer TLS,” IETF Draft, [Online].

Available from: https://tools.ietf.org/html/draft-friel-tls-atls-04 2020.01.10

[25] E. Rescorla, Key Material Exportes fro Transport Layer Security, " RFC 5705, [Online]. Available from: https://tools.ietf.org/html/rfc5705 2020.01.10

[26] Off-the-record Protocol Description version 3, [Online]. Available from: https://otr.cypherpunks.ca/Protocol-v3-4.1.1.html 2020.01.10

[27] Signal protocol, [Online]. Available from: https://signal.org/docs/ 2020.01.10

[28] Application Layer Transport Security, [Online]. Available from: https://cloud.google.com/security/encryption-in-transit/application-layer-transport-security/ 2020.01.10

[29] Manufacturing Message Specification, ISO 9506, [Online]. Available from: https://www.iso.org/standard/37080.html 2020.01.10

[30] E. Rescorla, R. Barns, and H. Tschofenig, "Compact TLS 1.3", IETF Draft, [Online]. Available from: https://datatracker.ietf.org/doc/draft-rescorla-tls-ctls/ 2020.01.10