Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/cose

Computers & Security

A systematic survey on multi-step attack detection

Julio Navarro [a,b,*], Aline Deruyver [a,b], Pierre Parrend [a,b,c]

[a] ICube Laboratory, Université de Strasbourg, 300 bd Sébastien Brant, 67412 Illkirch, France
[b] Complex System Digital Campus (UNESCO Unitwin), Paris, France
[c] ECAM Strasbourg-Europe, 2 Rue de Madrid, 67300 Schiltigheim, France

ARTICLE INFO

Article history:
Received 13 October 2017
Received in revised form 12 January 2018
Accepted 13 March 2018
Available online 20 March 2018

Keywords:
Advanced persistent threat
Event correlation
Intrusion detection system
Multi-stage attack
Multi-step attack
Network security

ABSTRACT

Since the beginning of the Internet, cyberattacks have threatened users and organisations. They have become more complex concurrently with computer networks. Nowadays, attackers need to perform several intrusion steps to reach their final objective. The set of these steps is known as *multi-step attack*, *multi-stage attack* or *attack scenario*. Their multi-step nature hinders intrusion detection, as the correlation of more than one action is needed to understand the attack strategy and identify the threat. Since the beginning of 2000s, the security research community has tried to propose solutions to detect this kind of threat and to predict further steps. This survey aims to gather all the publications proposing multi-step attack detection methods. We focus on methods that go beyond the detection of a symptom and try to reveal the whole structure of the attack and the links between its steps. We follow a systematic approach to bibliographic research in order to identify the relevant literature. Our effort results in a corpus of 181 publications covering 119 methods, which we describe and classify. The analysis of the publications allows us to extract some conclusions about the state of research in multi-step attack detection. As far as we know, this is the first survey fully dedicated to multi-step attack detection methods as mechanisms to reveal attack scenarios composed of digital traces left by attackers.

1. Introduction

According to ISO/IEC 27000 standard (ISO/IEC, 2016), an attack in the context of computer networks is defined as an “attempt to destroy, expose, alter, disable, steal or gain unauthorized access to or make unauthorized use of an asset”. This definition is broad enough to englobe any malicious intent against a network asset. However, the term *attack* is popularly identified to an individual malicious action. For instance, we speak of *SQL Injection attacks* or *Distributed Denial of Service (DDoS) attacks*, where just one action, possibly repeated, is required to threaten the system. We can call them *single-step attacks*.

Nevertheless, advanced attackers follow a step-by-step method in their attempts to attack a system. There are two reasons to proceed in this way. First, the victims chosen by advanced attackers are usually medium or big organisations, with a complex network topology and different layers of security. Considering that the most important assets in terms of information value are in the less reachable areas of the network, it would be almost impossible to complete an intrusion with success using a single-step attack. Second, if the attack is

[*] *Corresponding author.*
E-mail address: navarrolara@unistra.fr (J. Navarro).

https://doi.org/10.1016/j.cose.2018.03.001
0167-4048/© 2018 Elsevier Ltd. All rights reserved.

decomposed in several steps, it is more stealthy and difficult to be identified by the victim, especially if some of the steps do not pose a risk to the system by themselves.

The described attacks can be called *multi-step attacks* or *multi-stage attacks*. They have evolved and become more stealthy and sophisticated. If they started as a combination of regular attacks, they have later acquired relevance and even their own names, as it is the case of WannaCry, which created havoc in many institutions in May 2017 (Mohurle and Patil, 2017). For being able to detect multi-step attacks, we need to consider the global strategy of the attack and to highlight the causal relationship between traces collected all over the network. The sequence of traces left by the attackers often follows a logical progression (Dain and Cunningham, 2001a), as it is generated from a set of actions performed with a single objective. After each of the steps of the attack, the attacker gains knowledge about the target system (Katipally et al., 2011) and is able to better prepare the subsequent steps.

In this paper we present a systematic survey about multi-step attack detection methods published since the beginning of the field. Even if some seminal work about the importance of linking several events to detect an attack already existed (Vigna and Kemmerer, 1998), Huang et al. (1999) were probably the first ones that pointed out the importance of attack strategies in detection. The new millennium brought an awareness against this kind of attack and the analysis of the global attack strategy became important in security research. Developing mechanisms for multi-step attack detection is critical for protecting the networks, as prevention methods do not suffice. We need to have in mind that there will always be vulnerabilities in a network. If we harden all the network assets against every known vulnerability, we can loose flexibility (Wang and Jajodia). Our interest in detection comes from this limitation. The better are the detection methods, the lower is the risk of having vulnerabilities in the network assets. This challenge is specially hard to solve when attacks are complex and composed of different steps. The aim of this survey is to collect, explain and put in context the detection methods specifically developed against multi-step attacks.

Through a systematic bibliographic research, we have ended up with a selection of 181 publications, which constitutes what we call the *corpus* of this survey. It reflects the published results of 119 different multi-step attack detection methods. Some other publications that are also relevant for multi-step attack detection but that do not propose a specific detection method are mentioned as well. We only consider for the selected corpus the methods aiming to detect real traces of attacks, not the ones hypothesising about possible attack paths on a network. Moreover, we are only interested in the methods considering the whole attack scenario as an ensemble of the individual steps. We have not found previous surveys considering this point of view. There is some work related to correlation of alerts (Salah et al., 2013; Xu and Ning, 2008) or situational awareness (Luh et al., 2016), but they do not focus on the multi-step nature of current attacks, the object of study we are interested in. We have also identified some work about specific types of multi-step attack detection methods, like sequential pattern mining (Husák et al., 2017) or Hidden Markov Models (HMM) (Alghamdi, 2016), but not covering the whole domain.

We start this survey presenting a series of preliminary definitions in Section 2 and a list of challenges faced by multi-step attack detection in Section 3. We then explain in Section 4 the review methodology followed for the identification of the main corpus of the survey. The bibliometrics extracted from the reviewed publications are explained in Section 5. Section 6 constitutes the main part of the survey, where all the reviewed methods are classified by approach and explained. Following the review, we present in Section 7 an analysis of the state of multi-step attack detection in terms of type of data used, type and reproducibility of experiments or origin of knowledge about the attacks. We end this survey with a discussion in Section 8 and the conclusion in Section 9.

2. Preliminary definitions

As in every scientific domain, it is important in the field of computer security to have solid definitions of basic concepts. We present in this section the definitions of some terms extensively used in security detection. We also explain the concept of multi-step attack, the object of the detection methods considered in our study.

2.1. *Clarifying the use of basic terms*

We need to be cautious when using basic terms in the domain of attack detection, because they can be interpreted in many different ways depending on the author. To avoid confusion, we give here the definition of some terms as we use them. The choice is made according to the wide use of the definition in the reviewed literature and our own experience in the security industry.

- **Definition 1. Event.** We consider the definition given by the Standard on Logging and Monitoring published by the European Comission in 2010 (E. Commission, 2010): “An event is an identifiable action that happens on a device”. Events are usually represented as plain text in a format dependent on the used technology. This materialisation of an event, a textual trace stored in the device, is called *log*.
- **Definition 2. Packet.** It is the minimal unit of data exchanged in a network communication protocol. It represents an action that happens between two or more network devices. In cybersecurity, the word *network* is usually preceding the term to specify the environment.
- **Definition 3. Alert.** Also referred to as *alarm*, it is an event generated by a security system in response to the detection of alleged malicious activities or faults (Salah et al., 2013). In other words, it is an indicator that something is not working as it should or that a resource is not properly used. An alert is an event, but not all events are alerts. There exist events reporting normal functioning of systems. An alert supposes an indicator of a suspicious activity by itself and, as its name suggests, it demands the network analyst to be vigilant. Sometimes we refer to the events as “general events”, when we want to emphasize that we refer to any kind of event, including alerts. Alert information is usually represented as a short message with a specific format containing details about the problem.

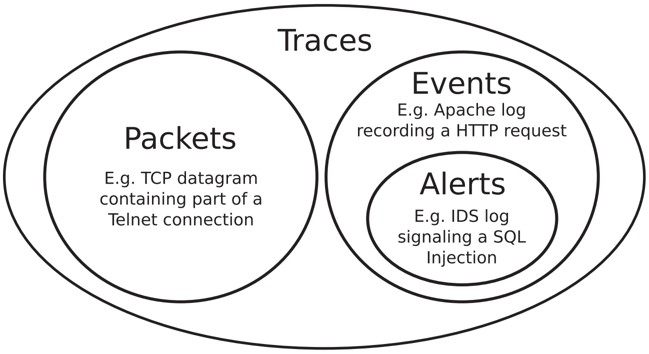
Traces
Packets
E.g. TCP datagram containing part of a Telnet connection
Events
E.g. Apache log recording a HTTP request
Alerts
E.g. IDS log signaling a SQL Injection

Fig. 1 – Diagram representing the scope of traces, packets, events and alerts. Every alert is an event signalling a security incident. Both events and packets are traces.

- **Definition 4. Trace.** This term is used for englobing all the indications of the execution of an action in a network, e.g. events (including alerts) or packets. In Fig. 1 we show a diagram to better understand the definition of trace.
- **Definition 5. IDS.** This acronym stands for Intrusion Detection System and denotes a system for automatic detection of suspicious activities in an IT system. An IDS can be either *network-based* or *host-based* (Vogel and Schmerl, 2011). A network-based IDS monitors network traffic, while a host-based IDS analyses the activity of applications and operating systems running at endpoint level. We are interested in the aspect of an IDS as an observer and reporter of suspicious actions (Cipriano et al., 2011) in the form of alerts, independent of the source of information.

2.2. *Multi-step attacks*

We call *multi-step attack* the ensemble of steps taken by one or several attackers with a single specific objective inside the network, containing at least two distinct actions. If actions are similar between them, it cannot be considered a multi-step attack. For example, in a DDoS attack against a device or service we can find millions of packets but each of them represents a particularisation of the same type of action. We consider thus regular DDoS attacks as multi-agent single-step. We can call them *distributed* or *coordinated* attacks (Zhou et al., 2010). We shall not confound the DDoS attack by itself with the whole process of intrusion into an organisation in order to get control over an endpoint and launch a DDoS attack from it. This last case, represented in the dataset DARPA 2000, is indeed a multi-step attack, and the DDoS attack launched against some victim external to the network is just the final step of it. Some authors (Chen et al., 2006; Yan and Liu, 2004) take the liberty of calling *DDoS attack* to the mentioned multi-step attack, leading to confusion.

The detection methods reviewed here assume there is a relationship between two events belonging to the same attack scenario. Highlighting the links between elementary actions in the form of traces (Brogi and Tong, 2016) should be the first step to multi-step attack detection. Conditions for linking two actions can take many forms. The most used in the literature, as single conditions or in combination with others, are listed below:

- IP addresses of the actors involved in the actions are identical or in the same range. Comparisons are made for source, destination or a combination between these two features (Cipriano et al., 2011; Ourston et al., 2003).
- The actions are usually found together (Mathew and Upadhyaya, 2009).
- There is a certain time difference between the actions or they belong to a specific time window (Sadoddin and Ghorbani, 2009).
- One action prepares for the conditions needed by the other (Ning and Xu, 2010; Wang et al., 2006).
- The type of action is the same. This is usually exploited when linking IDS alerts, where the alert attack type is considered (Fava et al., 2008; Soleimani and Ghorbani, 2012).

2.3. *Different ways to name multi-step attacks*

Depending on the author, multi-step attacks can also be called *multi-stage* (Chen et al., 2006) or *multistage* (Du et al., 2010) attacks. Some others refer to them as *attack strategies* (Huang et al., 1999), *attack plans* (Qin and Lee, 2004), *attack scenarios* (Mathew and Upadhyaya, 2009) or *attack sessions* (Cipriano et al., 2011).

We consider that the term *multi-step attack* is the one better reflecting the difficulties faced in the detection of this type of attack, as the main challenge resides in the fact that the attack is composed of many steps which, to top it all, can be of very different nature. The most relevant characteristic is that they cannot be described by less than two atomic events (Jaeger et al., 2015). The identification of these events as *steps* in the consecution of the attacker's objective captures well the concept.

Since the beginning of the decade (de Vries et al., 2012), the name *Advanced Persistent Threats* (*APT*) has been used to denote multi-step attacks that are specifically crafted against a single victim and where the access of the attacker to the target network is maintained during a long period of time. Not every multi-step attack is an APT. As WannaCry infection (Mohurle and Patil, 2017) has recently demonstrated, even global malware campaigns, that are not targeted, can be based on an attack that can be decomposed in multiple steps affecting several assets in the network.

The term APT has become somehow distorted after having been adopted by the industry and used in marketing campaigns. Moreover, a lack of rigorous definition of what an APT is (Chen et al., 2014) makes this term valid for other contexts, as for attacks using advanced malware. Among the publications proposing a detection method using the term APT for denoting their target, we only include in the corpus of the survey the ones addressing the identification of the global strategy behind the combination of the attack steps (Friedberg et al., 2015).

3. Challenges in multi-step attack detection

Multi-step attacks pose many challenges to detection. Among the challenges we find, of course, all those faced by general intrusion detection. One of them is, for example, the high

Table 1 – Types of trace and number of publications using them.

Type of trace	Number of pub.	Approaches
Only alerts	158	All types
General events	18	Similarity-based (Anming and Chunfu, 2004; Friedberg et al., 2015; Mathew and Upadhyaya, 2009; Skopik et al., 2014), mixed (Abreu et al., 2015), case-based (Eckmann et al., 2002; Giura and Wang, 2012a, 2012b; Jaeger et al., 2015; Kruegel et al., 2001; Vogel and Schmerl, 2011) and causal correlation (Chen et al., 2016)
Traces with triggering alerts	5	Similarity-based (Chen et al., 2006; King et al., 2005; Shaneck et al., 2006; Strayer et al., 2005) and causal correlation (Zhai et al., 2006)

complexity of current network data, which makes hard the identification of information relevant for security. Moreover, the most dangerous attacks happen only rarely, which mean that in each dataset we have only a few examples of attacks. This problem in intrusion detection research, called the Rare Data Problem, and some other statistical problems of security data, such as *skewed distributions* or *imbalanced datasets*, are further explained by Ourston et al. (2003).

The impact of these difficulties seems even worse when multi-step attacks are considered. For a single-step attack, all the information related to it is usually contained in a trace and linked to a vulnerability in a system. The attack can be isolated and compared with similar occurrences. It is more difficult to study and characterise the similarity between attacks as the number of steps involved is incremented. In a multi-step attack we need to infer not only the nature of each of the step, but also the links between them. Even if we can identify the involved steps, we can still miss the attack strategy.

We briefly review here the particular challenges faced by multi-step attack detection:

- Individual steps composing a multi-step attack can seem innocuous.
- We can never assume we have access to a complete library of attack plans. This contrasts with what happens in traditional plan recognition problems where the possible actions are well defined (Qin and Lee, 2004).
- A single attacker can conceive multiple attack plans (Qin and Lee, 2004). Furthermore, the execution of a plan can stop just because the attacker loses interest or is not able to take advantage of the vulnerabilities in the network (Yang et al., 2006).
- The detection of individual steps can be missed due to technical limitations of network devices or due to how they are deployed or configured (Chen et al., 2006).
- The chronological order of actions can become altered when actions are expressed as traces.
- An attacker does not need to follow a precise order for executing a multi-step attack (Zhang et al., 2006), so the set of possible sequences of actions can be very complex.
- Attackers can perform actions just to avoid recognition (Qin and Lee, 2004), even if they are not coherent with the rest of actions in the plan.
- Attack scenarios have to be modelled and represented in a standard language (Michel and Mé, 2002; Templeton and Levitt, 2001).
- An IDS does not usually include explicit information about the root causes of a problem (Chen et al., 2006; Salah et al., 2013), so it is difficult to identify contextual information.
- Much of the features contained in traces are categorical, i.e. there are no predefined order and metric between the possible values. This hinders the application of mathematical methods for attack scenario construction (Julisch, 2001).
- The interval between consecutive stages of an attack can be very high, on the order of hours, days or even months (Chen et al., 2006).
- There are not many standard datasets for the evaluation of multi-step attack detection systems. Furthermore, public research has no access to much of the methods and datasets used by other researchers or private institutions. We shall return to this issue in Section 7.

We show in Table 1 that many of the analysed multi-step attack detection methods use IDS alerts as a source. There are many challenges (Ning and Xu, 2010; Salah et al., 2013; Xu and Ning, 2004) related to the generation of IDS alerts:

- Any medium-size network generates a high volume of them. This poses a big challenge for the development of systems working in real time.
- Alerts indicating a real threat are mixed with false positives and non-relevant ones.
- Different security systems may generate a different alert for the same action.
- If there are systems from different security vendors, the format of the alerts will not probably be the same. A unification method is required for working with the ensemble of the alerts.
- There is an important lack of clear and consistent documentation about the meaning and format of IDS alerts (Zhou et al., 2007).

4. Review methodology

Our bibliographic research is based on a systematic search process inspired by Luh et al. (2016). Systematic search is a way to provide a rigorous and reproducible methodology to literature review. It requires more effort than traditional reviews (Kitchenham, 2004) but it avoids missing relevant but not well-known publications, and that is why we have chosen it as the method for this survey. The expertise in the domain is applied in the selection of the relevant publications and in the recursive identification of references.

We start with the search of a set of keywords related with multi-step attack detection in the most relevant search engines specialised in the scientific literature on Computer Science. We select from the obtained results the publications proposing a multi-step attack detection method. Then, we look for other publications in the references of selected ones. Additionally, we look for some other publications citing the ones found in the previous step.

In this section, we present the methodology used in the search. First, we list the set of criteria determining which work should be included and which one should be excluded. We then explain the three subsequent phases of the search: A) bulk search of keywords in several research engines, B) selection of relevant results and C) recursive search for references.

4.1. *Inclusion and exclusion criteria*

We establish a series of inclusion criteria to determine the work that should be included in the corpus of the survey:

- The authors present a multi-step attack detection method working on real digital traces such as events, alerts or network packets. Methods can be tested by simulation or case studies, but their main purpose should be the analysis of real data.
- The structure of the multi-step attack and the link between its steps should be considered in the detection process.
- The reference has a structure of scientific research publication.
- Methods are exposed in a clear and evident way.
- The publication is written in English and with an understandable style.

Emphasis needs to be given to the source of information for detection, composed of traces generated by the devices in the network. This excludes from the survey, therefore, all methods only based on structural and static network data, which rather belong to the domains of vulnerability analysis or risk assessment. These domains are vast and generate hypothesis about the possible paths of an attack in a network, which can be very useful in preventing further intrusions. Hypotheses are usually coded on an *attack tree*, an abstract representation of the network containing information about vulnerabilities affecting each asset. However, we focus our survey on detection in real time or through forensic investigation, always from the materialisation of actions in the form of traces. This includes work using attack trees but always as an instrument for projection of real traces, as it is presented in Section 6.3.

We also consider a set of exclusion criteria to better identify the reason why some references are excluded. Below we present a list of the categories of publication not considered in the survey:

- *Focused on multi-step attacks but not on detection*. This criterion is met by references addressing multi-step attacks but not proposing a detection method. Some of them can be, for instance, about attack modelling, languages for detection, risk assessment, vulnerability analysis or application of security policies.
- *Only focused on one aspect* of multi-step attack detection such as dropper analysis or C&C identification, among others. They only consider one network asset in the analysis, not having a global perspective for linking the different steps in the attack.
- *Not about multi-step attacks* at all. For example, about coordinated or distributed attacks, such as DDoS attacks, where multiple attackers may have a common objective but they do not necessarily perform multiple steps in the network. We also include here work on other security areas, such as access control, cryptography or wireless security, that can be returned by the automatic search.
- *Not in English*.
- *About single-step attack detection*.
- *Not research*, such as articles published in non-scientific magazines and commercial whitepapers.
- *About flow-based detection*, where the structure of the multi-step attacks is not revealed. In these publications, detection is done analysing the statistical effects on traffic produced by the attack.
- *Slides or posters*.
- *Duplicated*. Documents containing exactly the same content but that were published in different places or whose reference appears twice.
- *Of low quality*. In the selection of the publications, we follow the position defended by Glass (2000), who considers that a systematic survey should include all the references found in the studied domain, both *good* and *bad* ones. Anyway, as there is not a clear benchmark in multi-step attack detection and most of the publications do not offer the possibility of reproducing the experiments (see Section 7.5), it would be difficult to apply a narrow exclusion criterion based on the quality of results. However, we discard some references which are below a minimum level of quality in style, form and presentation; where:
  - there is not enough detail to understand how the method works,
  - text cannot be well understood because of bad use of English or
  - some elements taken from the work made by other authors are not properly credited.

We do not cite in this survey any of the work excluded in terms of quality. In current scientific research, number of citations is such a relevant metric in the evaluation of quality that citing work that we do not consider acceptable would do research a disservice. If the reader wants more information about discarded publications, please contact the corresponding author of this survey.

4.2. *Phase A: Using the research engines*

The objective of this first phase is to search the references conforming the starting point of the systematic bibliographical search. We need to choose a set of keywords frequently appearing in the title of publications about multi-step attack detection. We have chosen the following sets of keywords:

- **Set of keywords 1:** Advanced persistent threat(s) [2 strings]
- **Set of keywords 2:** APT + {analysis, architecture, defense, detection, framework, mechanism, mitigation, prediction, prevention, strategy, system} [11 strings]

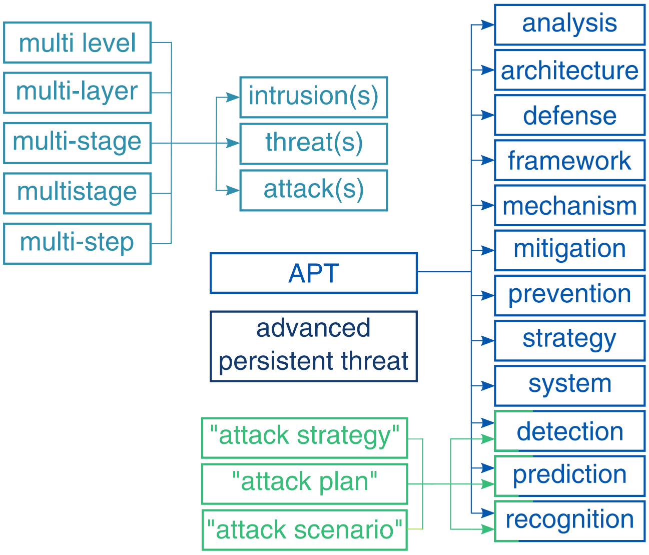

Fig. 2 – Diagram representing the keywords used in the search of Phase A. Different tones correspond to each set of keywords used.

- **Set of keywords 3:** {multi level, multi-layer, multi-stage, multi-step, multistage} + {intrusion(s), threat(s), attack(s)} [30 strings]
- **Set of keywords 4:** {“attack plan”, “attack scenario”, “attack strategy”} + {detection, prediction, recognition} [9 strings]

The symbol + denotes the combination of different keywords in the same search. The choices are presented enclosed by curly brackets. An s between parentheses means that a given word is considered in singular and plural. Quotations mean that the set of words enclosed by them is searched as an ensemble: the expression has to appear literally as it is written.

Selected sets of keywords lead to the number of strings to search shown next to each item in the list, between square brackets. This gives us a total of 52 strings to search. A visual representation of the used strings of keywords is shown in Fig. 2.

The next step is to search the strings of keywords in the most important search engines specialised in the scientific literature on Computer Science: IEEE, ACM, Web of Science and Google Scholar. Only the title of each publication is considered in the search. For Google Scholar, we exclude patents and citations. Citations are excluded because they include non-scientific publications, such as news, and because the sources are rarely openly available.

From the obtained results, we just download the references related to cybersecurity. Even if the keywords used are mainly related to cybersecurity, some publications from other domains are found in the search results. For example, the acronym “APT” can refer to concepts in Finances and Medicine.

After merging all the results coming from each of the four research engines, we end up with 432 references.

4.3. *Phase B: Filtering the results*

References found in Phase A are reviewed one by one. We apply the inclusion and exclusion criteria mentioned in Section 4.1 in order to only select the references proposing multi-step attack detection methods.

We follow the method proposed by Meline to generate a “bibliography of candidate studies”. We apply this method in two stages. First, we eliminate the publications clearly meeting one or more exclusion criteria after reviewing titles and abstracts. Then, we read the remaining documents and we include in the survey the publications meeting all inclusion criteria and no exclusion criteria.

After that, we end up with just 50 references to include in the final corpus. That means we have discarded 382 references from the original search.

Excluded work and the reasons we adduce for their exclusion are important to understand the research about multi-step attacks. Because of that, we represent in the pie chart of Fig. 3

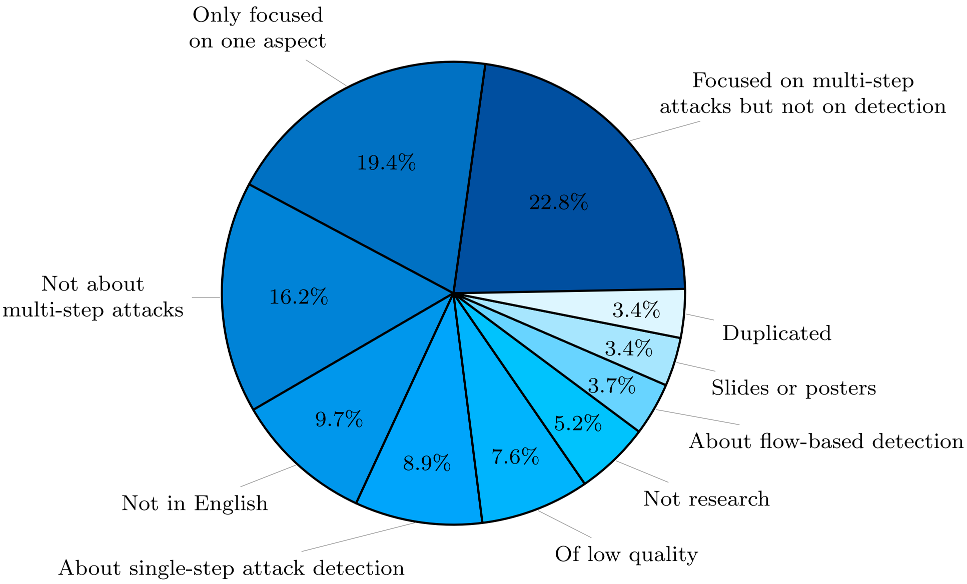

Fig. 3 – Distribution of excluded references according to the main reasons of exclusion.

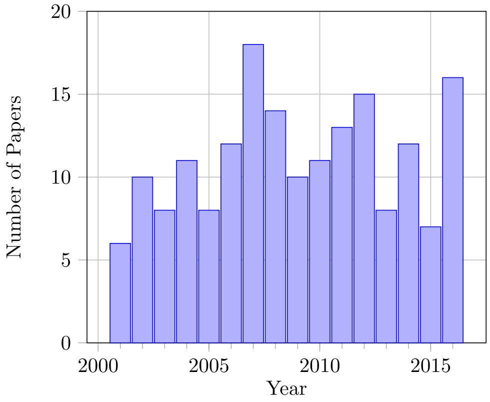

Fig. 4 – Number of publications about multi-step attack detection per year.

the distribution of excluded documents according to the main reason of exclusion. The categories are directly taken from the already cited exclusion criteria.

4.4. *Phase C: Recursive search for references*

After the first search and the filtering process of Phases A and B, we have followed a recursive search of references among the ones cited by the selected publications. This new iteration has returned 65 publications to be included in the corpus. The process has been repeated on the bibliography of these 65 references, after which we have found 8 additional publications. Reviewing the references contained in these 8 publications we have not found further references to include in the survey. This gives a total of 123 publications.

The problem of recursive search of references is that it is oriented "towards the past": found references in one publication are always older than the publication itself. That is why we have looked for new references citing the 123 publications we had already identified. This is easy to do using Google Scholar, where each reference counts with a "cited by" button to access to a list of publications citing the reference. The search has returned 58 additional publications to include in the corpus.

The total number of publications proposing a multi-step attack detection method composing the corpus of this survey is then **181 publications**, covering a total of **119 detection methods**.

5. Bibliometrics

Before presenting each one of the approaches about multi-step attack detection, we introduce a series of statistical analysis of the corpus of publications considered in the survey. We are interested in the number of publications per year, the number of citations and the number of publications written by each author. These metrics could be extracted from a corpus of publications in any scientific domain.

One of the studied aspect is the number of publications per year. The results are represented in the histogram of Fig. 4. They are separated by type of approach in Fig. 5. We have not considered in these figures the documents made public after 2016, because this survey is written before the year 2017 arrives at its end and the results would not be complete.

On the other hand, we want to analyse the most relevant publications and authors in the domain. We have listed the 20 most cited publications in Table 2 and the authors of more than 4 publications about multi-step attack detection in Table 3.

The number of citations included in this survey has been extracted from Google Scholar, which we consider as the most complete search engine in terms of number of references. It is regrettable that this search engine does not provide a better API to extract the metadata and that it blocks automatic

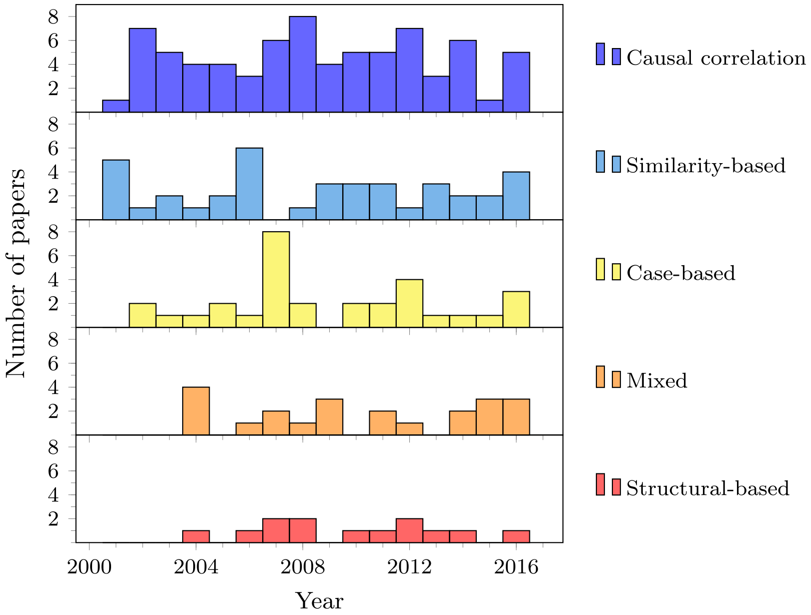

Fig. 5 – Number of publications about multi-step attack detection per year, classified according to the followed approach. Approaches are arranged in decreasing order according to the number of publications in each of them.

Table 2 – Ranking of top 20 entries according to the number of citations in Google Scholar.

Ref.	Title	Year	Cit.
(Cuppens and Miège, 2002a)	Alert correlation in a cooperative intrusion detection framework	2002	941
(Valdes and Skinner, 2001)	Probabilistic alert correlation	2001	932
(Ning et al., 2002a)	Constructing attack scenarios through correlation of intrusion alerts	2002	659
(Valeur et al., 2004)	Comprehensive approach to intrusion detection alert correlation	2004	505
(Julisch, 2003a)	Clustering intrusion detection alarms to support root cause analysis	2003	501
(Eckmann et al., 2002)	STATL: An attack language for state-based intrusion detection	2002	496
(Cuppens, 2001)	Managing alerts in a multi-intrusion detection environment	2001	374
(Ning et al., 2004)	Techniques and tools for analyzing intrusion alerts	2004	359
(Dain and Cunningham, 2001a)	Fusing a heterogeneous alert stream into scenarios	2001	346
(Julisch and Dacier, 2002)	Mining intrusion detection alarms for actionable knowledge	2002	326
(Cheung et al., 2003)	Modeling multistep cyber attacks for scenario recognition	2003	285
(Qin and Lee, 2003)	Statistical causality analysis of INFOSEC alert data	2003	271
(Julisch, 2001)	Mining alarm clusters to improve alarm handling efficiency	2001	239
(Morin and Debar, 2003)	Correlation of intrusion symptoms: an application of chronicles	2003	233
(Ning et al., 2002b)	Analyzing intensive intrusion alerts via correlation	2002	222
(Ning and Xu, 2003)	Learning attack strategies from intrusion alerts	2003	219
(Wang et al., 2006)	Using attack graphs for correlating, hypothesizing, and predicting intrusion alerts	2006	204
(Qin and Lee, 2004)	Attack plan recognition and prediction using causal networks	2004	203
(Ning et al., 2004)	Building attack scenarios through integration of complementary alert correlation method	2004	197
(Geib and Goldman, 2001)	Plan recognition in intrusion detection systems	2001	195

searches when a high number of requests is made. Overcoming these limitations could made Google Scholar an even more important reference for scientific surveys than what it is nowadays.

We see that 2007 is the year with the highest number of publications about multi-step attack detection, closely followed by 2016. This last peak is an indicator that the research about multi-step attacks is still very active nowadays. The recent threat posed by WannaCry makes us think that multi-step attack research will be an important research topic in the near future.

In Table 3 we also include for each author the number of citations of the most cited publication, the number of citations of the least cited publication and the total number of citations. Some important authors, such as Julisch, do not appear in the ranking of Table 3 in spite of their number of citations because they have less than five publications in the

Table 3 – Ranking of top authors according to the number of publications about multi-step attack detection included in the survey.

Author	# of pub.	Citations			References
		Total	Max.	Min.	
Ning, Peng	11	1965	659	2	(Ning and Cui, 2002c; Ning and Xu, 2003, 2004, 2010; Ning et al., 2002a, 2002b, 2004; Xu, 2006; Xu and Ning, 2004; Zhai et al., 2006)
Wang, Li	11	156	39	2	(Li et al., 2007a, 2007b, 2007c, 2007d; Wang et al., 2006, 2007, 2010; Zhang et al., 2007)
Li, Zhitang	11	147	39	2	(Li et al., 2007a, 2007b, 2007c, 2007d; Ma et al., 2008; Wang et al., 2006, 2007; Zhang et al., 2007)
Yang, Shanchieh J.	10	310	77	1	(Byers and Yang, 2008; Du et al., 2009, 2010; Fava et al., 2007, 2008; Holsopple and Yang, 2008; Holsopple et al., 2006; Murphy and Yang, 2010; Yang et al., 2008, 2009)
Ghorbani, Ali A.	8	289	123	7	(Bateni et al., 2013; Ren et al., 2010; Sadoddin and Ghorbani, 2009; Soleimani and Ghorbani, 2008, 2012; Wang et al., 2010; Zhu and Ghorbani, 2006)
Xu, Dingbang	7	973	359	2	(Ning and Xu, 2003, 2004, 2010; Ning et al., 2004; Xu, 2006; Xu and Ning, 2004)
Holsopple, Jared	6	229	77	18	(Du et al., 2010; Fava et al., 2007; Holsopple and Yang, 2008; Holsopple et al., 2006; Yang et al., 2008, 2009)
Lei, Jie	6	99	39	2	(Li et al., 2007b, 2007c, 2007d; Wang et al., 2006, 2007)
Cuppens, Frédéric	5	1499	941	34	(Benferhat et al., 2003; Cuppens, 2001; Cuppens and Miège, 2002; Cuppens et al., 2002a, 2002b)
Cui, Yun	5	1342	659	17	(Cui, 2002; Ning and Cui, 2002c; Ning et al., 2002a, 2002b, 2004)
Sudit, Moises	5	220	77	9	(Holsopple et al., 2006; Mathew et al., 2010; Stotz and Sudit, 2007; Sudit et al., 2005; Yang et al., 2009)
Li, Dong	5	87	39	8	(Li et al., 2007a, 2007c, 2007d; Wang et al., 2007; Zhang et al., 2007)
Meinel, Cristoph	5	83	74	0	(Fayyad and Meinel, 2013; Jaeger et al., 2015; Roschke et al., 2011; Ussath et al., 2016a, 2016b)
Alserhani, Faeiz	5	41	33	0	(Alserhani, 2012, 2013, 2016; Alserhani and Akhlaq, 2011; Alserhani et al., 2010)

Ranking of top authors according to the number of publications about multi-step attack detection included in the survey.

corpus. The full list of authors can be requested to the corresponding author of this survey.

From Tables 2 and 3 we can grasp a better idea of the most relevant and influential authors in the field of multi-step attack detection, who are also the ones leading the most long-lasting projects. The best examples are Peng Ning (Ning and Xu, 2003; Ning et al., 2002a, 2002b, 2004), and Frédéric Cuppens (Benferhat et al., 2003; Cuppens and Miège, 2002a; Cuppens et al., 2002b, 2002c), who in parallel laid the foundations of causal correlation through prerequisites and consequences. Ning, with the help of other collaborators as Yun Cui, has developed this method at the North Carolina State University during 8 years, being the author with the highest number of citations in multi-step attack detection. Although less prolific, Cuppens counts with the most cited publication in the field (Cuppens and Miège, 2002a) and he has explored other types of method, such as clustering (Cuppens, 2001). Li Wang and Zhitang Li (Li et al., 2007a; Wang et al., 2006, 2007, 2010) and the team directed by Ali A. Ghorbani (Bateni et al., 2013; Sadoddin and Ghorbani, 2009; Soleimani and Ghorbani, 2012; Wang et al., 2010; Zhu and Ghorbani, 2006) have also conducted long-lasting projects about multi-step attack detection. Some other authors of extensively developed methods are Qin and Lee (Qin and Lee, 2003, 2004, 2007), the first ones to introduce the idea of attack plan recognition, or Julisch (Julisch, 2001, 2003a, 2003b; Julisch and Dacier, 2002), who developed the concept of root cause analysis.

6. Review of multi-step attack detection methods

In this section we present the final selection of publications proposing multi-step attack detection methods. We refer to this selection as the *corpus* of the survey. The methods are classified according to the main approach they follow. Work in some of the approaches is further classified into categories. The taxonomy of the classification is shown in Fig. 6. We find five kinds of approach:

- **Similarity-based**. The degree of similarity between traces determines the construction of the attack scenarios. We find three categories in this approach: *progressive construction* (by *attribute matching* or *correlation*), *scenario clustering* and *anomaly detection*.
- **Causal correlation**. Detection is focused on the anatomy of multi-step sequences and the causal relationship between their steps. Publications under this approach can be further classify into one out of three categories: *prerequisites and consequences*, *statistical inference* or *model matching*.
- **Structural-based**. Incoming traces are projected to a model of the network, where future attack paths can be predicted.
- **Case-based**. Detection of well-known attack scenarios as an ensemble of traces.
- **Mixed**. More than one of the approaches are followed but none of them stands out among the others.

The distribution of the work in the corpus according to its approach is shown in Fig. 7. Below is a summary of the contributions to each one of the approaches. We finish this section by presenting some work about architectures for detection, which is not included in the corpus of this survey because it does not explain the detection methods involved.

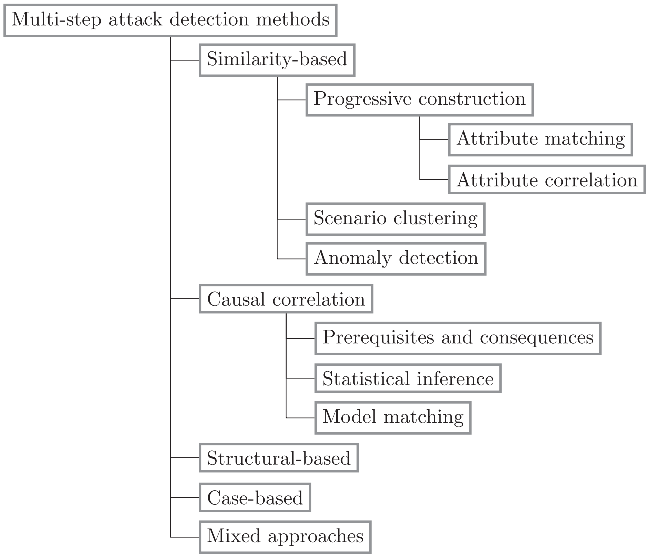

Fig. 6 – Taxonomy of multi-step attack detection classification.

6.1. *Similarity-based methods*

Similarity-based methods propose the composition of scenarios according to the similarity between the individual steps of the attack. They are based on the idea that similar alerts are related to the same root cause (Julisch, 2001; Salah et al., 2013) and they therefore belong to the same attack scenario. Computation of the similarity degree is the central focus of these methods. This distinguishes them from causal correlation methods, which focus on the causal structure of the sequence.

The similarity between traces is computed after one or several attributes or fields among the ones contained in each trace: IP addresses, port numbers, timestamp, type of trace, etc. The metric to do the comparison between fields depends on each detection method. It is generally expressed as a correlation index, which can be binary, i.e. equal or unequal, or based on a more complex correlation function. Some authors consider just the comparison of one feature (Shaneck et al., 2006), while most of them look at a combination of several of them (Chen et al., 2006; Zhu and Ghorbani, 2006). When this is the case, some of the fields can be considered as more important than others through the application of weights.

The main advantage of similarity-based methods is that if the process to determine the link between the traces is correctly chosen, the implementation is easy and can return unknown multi-step attacks. The analysis is just based on comparison between pairs of traces, so systems implementing these methods have usually a good performance. However, choosing how the traces are linked is far from being an easy task. If the linking process is kept simple, only relying on the similarity of few fields, the results will contain too many false positive alerts. On the contrary, a complex linking process based

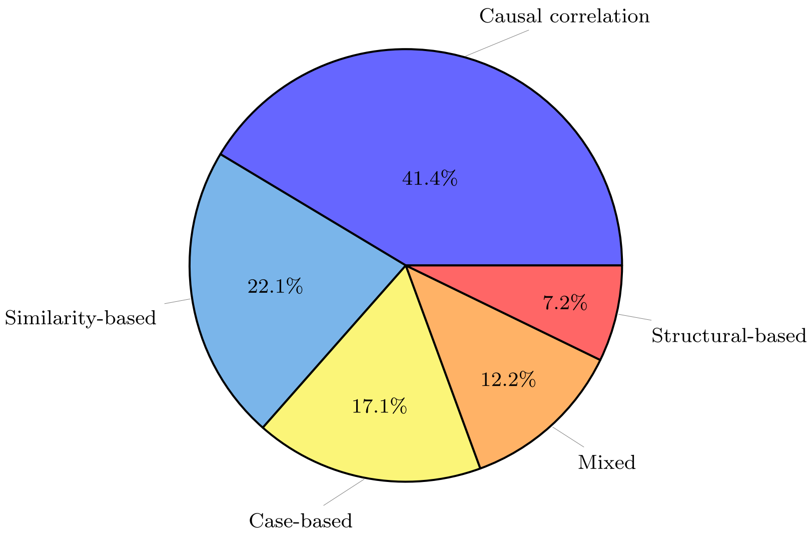

Fig. 7 – Distribution of the publications in the corpus according to the followed approach.

on the application of correlation matrices and using different weights for each field can be too specific to capture the characteristics of the whole range of real multi-step attacks.

We can classify similarity-based methods in three categories depending on how the degree of similarity is used in the creation of the attack scenarios. First of all, we have a set of methods based on *progressive construction*, where the sequence of actions conforming the attack is built step by step by the addition of similar traces. Secondly, methods doing *scenario clustering* just apply a clustering method to all the set of alerts and return the clusters as possible scenarios, without considering the order as it happens in progressive construction. Finally, in *anomaly detection* methods the similarity of incoming sequences of traces is computed against a set of non-malicious traces; and sequences are considered as part of an attack scenario if they differ from normality.

6.1.1. *Progressive construction*

In progressive construction methods, a potential attack sequence is built step by step. Traces are appended to a scenario according to the similarity with the rest of the traces in it. Compared traces are selected from the same time window. The difference with clustering methods is that the order of the actions is an important factor. In progressive construction, the sequences are built step by step and following a logical progression.

The match between the compared fields can be exact or partial. Methods using an exact match between some fields in the events have been considered in the subcategory *attribute matching*. On the other side, methods which consider partial similarity are placed in *attribute correlation*. These last methods calculate a correlation coefficient between events and place them together if the coefficient is above certain threshold. The list of all the progressive construction methods found in the literature is shown in Table 4.

6.1.1.1. Attribute matching. Chen et al. (2006) propose a system for active correlation built on top of a Bro IDS. This allows direct contact with network traffic, as opposed to passive event collection. The system highlights malicious network events using IDS signatures and follows their development as consecutive events, finding matches in terms of IP address or port.

A similar technique, but not directly implemented on the IDS, is followed by Shaneck et al. (2006). They divide the detection process in two stages. The first one combines pattern matching by an IDS engine and profiling to create attack scenarios in a highly restrictive way. In a second step, other alerts or pieces of network traffic with the same IP addresses as the alerts in each scenario are appended to it. The BDB system (King et al., 2005) also starts its operation with a triggering alert, but the authors do not give much detail about the linking process.

IP addresses are also the only features considered by Liu et al. (2008) to link the steps of an attack. In their system, single-step attacks are first assigned to a predefined phase of attack. Alerts are progressively added to a scenario if the source and the destination IP addresses match with the rest of alerts in the scenario. This perspective is also adopted by Ebrahimi et al. (2011), who propose a simple matching algorithm using both source and destination IP addresses.

The technique developed by Brogi and Tong (2016) also aims to find links between elementary attacks but using a method based on tags. These tags are assigned to each of the steps in the potential multi-step attack and propagated through the system by flows of information. Links between the steps are created if they have features with coincident information.

A different perspective is to focus on the detection of the attacker at IP level, such as STARLITE (Strayer et al., 2005). A module is added to an IP traceback system to trace back the source of the attack to the most external router in the network. This system does not consider the attack in terms of a sequence of actions but in terms of a sequence of assets used by the attacker during the infiltration.

6.1.1.2. Attribute correlation. Valdes and Skinner (2001) are one of the first authors in proposing a method based on attribute correlation for multi-step attack detection. Their method is

Table 4 – List of reviewed similarity-based multi-step attack detection methods in the category of progressive construction.

Approach	References	Period	Type of data	Datasets	Knowledge extraction	A_m	A_d	A_k	Rep.	Attack model	Total cit.
Progressive construction by attribute matching	(King et al., 2005)	2005	T	Simulation Private	Manual	No	No	No	No	Yes	104
	(Strayer et al., 2005)	2005	T	Private	Automatic	No	No	–	No	No	25
	(Chen et al., 2006)	2006	T	DARPA 2000 Private	Automatic	Yes	Yes	–	Yes	No	10
	(Shaneck et al., 2006)	2006	T	Private	Automatic	No	No	–	No	No	0
	(Liu et al., 2008)	2008	A	DARPA 2000	Automatic	Yes	Yes	–	Yes	Yes	38
	(Ebrahimi et al., 2011)	2011	A	DARPA 2000	Automatic	Yes	Yes	–	Yes	Yes	8
	(Brogi and Tong, 2016)	2016	A	Case study	Automatic	Yes	No	–	No	Yes	3
Progressive construction by attribute correlation	(Valdes and Skinner, 2001)	2001	A	Private	Manual	No	No	Yes	No	No	932
	(Dain and Cunningham, 2001a, 2001b)	2001	A	DEFCON 8	Supervised	No	Yes	Yes	No	No	477
	(Zhu and Ghorbani, 2006)	2006	A	DARPA 2000	Automatic	Yes	Yes	–	Yes	Yes	123
	(Wang et al., 2006, 2010)	2006–2010	A	DARPA 2000 Private	Manual	Yes	Yes	Yes	Yes	Yes	43
	(Khakpour and Jalili, 2009)	2009	A	DARPA 2000	Supervised	No	Yes	No	No	Yes	1
	(Bateni and Baraani, 2014; Bateni et al., 2013)	2013–2014	A	DARPA 2000 Private	Supervised	Yes	Yes	No	No	Yes	21
	(Pei et al., 2016)	2016	E	Simulation	Automatic	Yes	No	–	No	Yes	1
	(Wang and Chiou, 2016)	2016	A	DARPA 2000	Automatic	Yes	Yes	–	Yes	Yes	1

integrated in the EMERALD system (Porras and Neumann, 1997) and it is based on a similarity matrix that contains the manually crafted correlation indexes between each type of alert. The mechanism they use is not explained in detail, but it is still one of the most cited references in multi-step attack detection. In the system proposed by Dain and Cunningham (2001a, 2001b), attack scenarios are build based on the partial similarity of IDS alerts. Each time an alert arrives at the system, the probability of being assigned to one or another scenario is computed based on a comparison against the last alert in the scenario. The probability depends on several factors: similarity of IP addresses, time interval between the alerts or logic progression between attack types. The similarity score proposed by Khakpour and Jalili (2009) is also based on several features. It also depends on the correlation strength of alert types deduced from a manually crafted database, whose details are not provided in their work.

Similarity is computed by pairs of elementary attacks in the method proposed by Zhu and Ghorbani (2006). The correlation indexes are based on six selected features and extracted from a reference set of alerts. They do so using two methods: Multilayer Perceptron (MLP) and Support Vector Machine (SVM). The resulting indexes are stored by attack type in an Alert Correlation Matrix (ACM), used to deduce the attack scenarios.

The Statistical Filtering algorithm, developed by Wang et al. (2006), is also based on a correlation matrix, but in this case the indexes are manually deduced from experience and they are fixed in time. Even if sequences are automatically extracted according to their frequency, the final verdict about whether they represent an attack scenario relies on the human-crafted correlation indexes. Li Wang continued the development of her idea several years later (Wang et al., 2010) introducing more sophisticated time windows, whose size evolves with time, and a classification of candidate scenarios in three groups according to the relationship between their members. Wang and Chiou (2016) propose in a most recent work a method that also uses a correlation matrix, but they distinguish between forward and backward correlation strength, according to the time order of the alerts.

Pei et al. (2016) propose HERCULE, a system for doing "attack story reconstruction". Their method is inspired by relationships in social networks. They define a long list of possible relationships between events. These relationships are exploited to create graphs of events. Edges in the graphs have a weight value, which is calculated using a quadratic optimisation algorithm. The result returned by the system is a complete graph representing all the events involved in a multi-step attack. A very interesting aspect of their paper is the emulation of an ample group of real APTs from the existing literature.

On the other side, Artificial Immune Systems (AIS) have been much used in single-step intrusion detection (Kim et al., 2007), but the only proposal particularly addressing multi-step attack detection is iCorrelator (Bateni and Baraani, 2014; Bateni et al., 2013). ICorrelator Bateni and Baraani and emulates the human immune system in a three-layer architecture. It is based on *cells*, which are vectors of comparison features, e.g. the similarity between IP addresses. Each cell represents the correlation degree between two alerts. Correlation is based on the set of cells stored in memory, which evolves through supervised learning from an initial set of basic rules.

6.1.2. *Scenario clustering*

The goal of clustering is to discover *natural* groups in a set of elements (Jain, 2010). This is usually done through the application of automatic clustering algorithms. In this category we find the methods applying a clustering algorithm in order to identify groups of similar actions (Table 5). Those groups or clusters are then considered to be potential multi-step attacks. The degree of similarity between traces belonging to the same scenario should be higher than the degree between traces from different scenarios.

As far as we know, the application of clustering to multi-step attack detection was first proposed by Julisch (2001, 2003a, 2003b). His approximative alert clustering method, later integrated in a framework Julisch and Dacier (2002), is based on taxonomies associated to each one of the attributes in the alert. Taxonomies are arranged in the form of trees, where parent nodes include the concepts in children nodes. Alert similarity is computed from the distance between nodes within the tree corresponding to each attribute. The aim of the author is to reduce the huge number of alerts generated by IDS, mostly fruit of persistent configuration errors or particularities of devices. However, the revelation of the root causes of a set of alerts can also lead to the identification of attack scenarios and a better understanding of attacker's intention. Some other work (Wang et al., 2006) has proposed to improve in terms of quality the algorithm developed by Julisch using a genetic algorithm.

In the context of MIRADOR project, Cuppens also developed a clustering method oriented to alert fusion (Cuppens, 2001) which was successfully tested on multi-step attack detection. However, his most important contribution to the domain is on causal correlation, about which we speak in Section 6.2.1.

Clustering of attack tracks using significant services clustering method (Murphy (2009); Murphy and Yang (2010)) uses a similarity matrix based on the services each attack exploits. Clusters of alerts are extracted using Divisive Hierarchical Clustering (DHC) on a social network graph derived from the similarity matrix. Qiao et al. (2012) propose a simple formula for computing the similarity between alerts. They apply a double clustering followed by a loose application of LCS (Longest Common Subsequence).

In JEAN (Judge Evaluation of Attack Intension) system, proposed by Cheng et al. (2011), the process starts building a database of attack session graphs from a training set of IDS alerts using J-Fusion, an algorithm for alert fusion. Once the scenarios are built, the system mines other occurrences of the attack session graphs using a method inspired by the generalised Hough transform, an image processing method to identify geometric forms.

Colajanni et al. (2010) and Manganiello et al. (2011) propose the application of a Self-Organizing Map (SOM), a kind of auto associative neural network, before a second phase of alert clustering using k-means. In a final phase, a correlation index between clusters is calculated and the resulting attack scenarios are represented in the form of oriented graphs. They do not give information about which features are compared.

In a brief paper, Zhang et al. (2015) present a clustering method based on a specific metric between IP addresses. We have not been able to find more details about their work. Their original contribution is that they work with alerts coming from a WAF (Web Application Firewall). A WAF works in a high level

Table 5 – List of reviewed similarity-based multi-step attack detection methods in the categories of scenario clustering and anomaly detection.

Approach	References	Period	Type of data	Datasets	Knowledge extraction	A_m	A_d	A_k	Rep.	Attack model	Total cit.
Scenario clustering	(Cuppens, 2001)	2001	A	Private	Automatic	Yes	No	–	No	No	374
	(Julisch, 2001, 2003a, 2003b; Julisch and Dacier, 2002)	2001–2003	A	Private	Automatic	Yes	No	–	No	Yes	1138
	(Wang et al., 2006)	2006	A	Private	Automatic	Yes	No	–	No	No	14
	(Murphy, 2009; Murphy and Yang, 2010)	2009–2010	A	Simulation	Automatic	Yes	No	–	No	Yes	2
	(Cheng et al., 2011)	2011	A	DARPA 2000 DARPA GCP Simulation	Supervised	Yes	Yes	Yes	Yes	No	16
	(Colajanni et al., 2010; Manganiello et al., 2011)	2010–2011	A	DEFCON 18	Automatic	No	Yes	–	No	Yes	10
	(Qiao et al., 2012)	2012	A	Private	Automatic	Yes	No	–	No	No	7
	(Zhang et al., 2015)	2015	A	Private	Automatic	Yes	No	–	No	No	0
	(Kawakani et al., 2016, 2017)	2016–2017	A	Private	Automatic	Yes	No	–	No	Yes	1
Anomaly detection	(Anming and Chunfu, 2004)	2004	E	DARPA 2000	Supervised	Yes	Yes	Yes	Yes	No	16
	(Mathew and Upadhyaya, 2009)	2009	E	Private	Automatic	Yes	No	–	No	No	0
	(Shin et al., 2013)	2013	A	DARPA 2000	Supervised	Yes	Yes	Yes	Yes	No	30
	(Friedberg et al., 2015; Skopik et al., 2014)	2014–2015	E	Private Simulation	Supervised	Yes	No	No	No	No	59

of abstraction, which can be convenient for the study of the purpose of the attacker.

Finally, Kawakani et al. (2016, 2017) propose a method for hierarchical clustering of graphs representing attack strategies. The graphs are automatically derived from alerts in the same time window by attribute matching of IP addresses. Built clusters can be later used to classify new scenarios.

6.1.3. *Anomaly detection*

Anomaly detection methods (Table 5) learn from a dataset clean of attacks and then consider as a threat the sequences differing from normal behaviour. Similarity comparison is then made against a whole reference dataset, not only between the members composing the scenario. And the results are used differently than in the other similarity-based methods: we do not search the similarities but the differences. It is important to note that abnormal behaviour not necessarily correspond to an attack. The rate of false positives can be high, but anomaly detection methods offer the possibility of finding previously unseen attacks.

Mathew and Upadhyaya (2009) apply Principal Component Analysis (PCA) to build a model from attack-free data. Next, they project new data on the created clusters, so abnormal behaviour is easily identified. They consider a set of different states for the network. Each one of the states is defined by certain features and composed of information from heterogeneous security data.

In the anomaly detection method developed by Friedberg et al. (2015) and Skopik et al. (2014), events from a training set free of attacks are linked in a random way to create hypotheses. They develop a mathematical framework to define hypotheses, rules and anomalies. Events not related to the hypothesis are considered as anomalous and raise alerts. The multi-step attack perspective is only present during the characterisation of the clean dataset, as detection is made event per event.

Hidden Markov Models (HMMs) have also been used in anomaly detection. A set of HMMs representing sequences of normal events are built from a clean dataset. Sequences not corresponding to the trained models are consider as anomalous and thus corresponding to a multi-step attack. Anming and Chunfu (2004) implemented a method based on this in 2004, using the Segmental K-means algorithm to create the HMM from a training dataset of OS audit data. 9 years later, Shin et al. (2013) apply the same method on the same dataset but on IDS alerts.

6.2. *Causal correlation methods*

In causal correlation, the causal progression of the sequences of traces is the key factor in the identification of multi-step attacks. In other words, previous steps determine the ones that follow, and a causal scheme can be derived from this relationship. This completely differs from similarity-based methods, where the found links depend on the inherent similarity between features of each action or set of actions. In progressive construction methods (see Section 6.1.1), causality is considered but just as a result derived from similarity relationships, and not as an element in the analysis.

Causal correlation methods have an important advantage: their process and results can be easily interpreted by a human analyst (Salah et al., 2013). They highlight the character of multi-step attacks as sequences of steps, the most intuitive view we have of these threats so far. Their flexibility leaves a bit of space to find slight variations of known attacks, but not completely unknown ones. There may still be a high number of false positives, but less than in similarity-based methods, as safer hypotheses are made.

Depending on the considered aspect of causality, we find three categories of causal correlation methods: prerequisites and consequences; statistical inference; and model matching. For methods based on *prerequisites and consequences* the causal relationship of individual actions is explicitly coded in a database. *Statistical inference* focuses on the extraction of causality from the frequency of occurrence of actions with respect to other actions. And *model matching* refers to the comparison of data against predefined general models of what should be a multi-step attack.

6.2.1. *Prerequisites and consequences*

In these methods, each alert is supposed to have a series of prerequisites, also called preconditions, and consequences, or postconditions (Table 6). The prerequisites are the conditions to be given for an attack to be successful, while the consequences are the possible effects of the attack (Benferhat et al., 2003; Ning and Xu, 2010). Methods in this category presuppose that each possible alert has a list of known associated prerequisites and consequences. Real alerts are brought together to form hyper-alerts. A hyper-alert is a set of facts, represented by alerts, with the same prerequisites and consequences. Hyper-alerts are then correlated through the automatic identification of prerequisites to consequences, returning a sequence of attacks composing an attack scenario.

The first language created to represent a multi-step attack scenario considering the prerequisites and consequences of each step was LAMBDA (Cuppens and Ortalo, 2000), proposed by Cuppens and Ortalo. It was followed by JIGSAW (Templeton and Levitt, 2001) and ADeLe (Michel and Mé, 2002). The authors of these seminal works did not propose a method for detection, just the language.

Shortly after the development of LAMBDA, Cuppens et al. became the pioneers in the development of prerequisites and consequences methods (Benferhat et al., 2003; Cuppens and Miège, 2002a; Cuppens et al., 2002b, 2002c). They developed their works in the context of MIRADOR project. In their works, the connection of attacks A and B is made if the consequences of A partially match the prerequisites of B. Additional hypotheses, expressed as ontological rules, are needed to the connection of some of the attacks. The extraction of the correlation rules is made in an automatic way from the dataset of individual attacks expressed in LAMBDA language. Once we have the rules, they can be applied to IDS alerts in real time.

The other father of multi-step attack detection based on prerequisites and consequences is Ning, the most cited author in this corpus. His team at North Carolina State University has been the most prolific one in this category of methods, both in terms of number of publications and of time span of the project (Cui, 2002; Ning and Cui, 2002c; Ning and Xu, 2002d,

Table 6 – List of reviewed multi-step attack detection methods based on causal correlation in the categories of prerequisites and consequences and model matching.

Approach	References	Period	Type of data	Datasets	Knowledge extraction	A_m	A_d	A_k	Rep.	Attack model	Total cit.
Prerequisites and consequences	(Benferhat et al., 2003; Cuppens and Miège, 2002; Cuppens et al., 2002a, 2002b)	2002-2003	A	Case study	Manual	Yes	No	Yes	No	No	1125
	(Cui, 2002; Ning and Cui, 2002c; Ning and Xu, 2003, 2010; Ning et al., 2002a, 2002b, 2004; Xu, 2006; Xu and Ning, 2004; Zhai et al., 2006)	2002-2010	A	DEFCON 8 DARPA 2000 DARPA GCP Simulation	Manual	Yes	Yes	Yes	Yes	Yes	1714
	(Cheung et al., 2003)	2003	A	DARPA GCP	Manual	Yes	Yes	No	No	Yes	285
	(Yan, 2005; Yan and Liu, 2004)	2004-2005	A	DARPA 2000	Manual	Yes	Yes	Yes	Yes	Yes	3
	(Wang and Jajodia; Wang et al., 2005, 2006)	2005-2008	A	DARPA 2000 UCSB 2004	Manual	Yes	Yes	No	No	No	257
	(Pandey et al., 2008; Zhou et al., 2007)	2007-2008	A	DARPA 2000 Private	Manual	Yes	Yes	Yes	Yes	Yes	143
	(Alserhani, 2012, 2013, 2016; Alserhani and Akhlaq, 2011; Alserhani et al., 2010)	2010-2016	A	DARPA 2000 Private Simulation	Manual	Yes	Yes	No	No	No	41
Model matching	(Mathew et al., 2010; Stotz and Sudit, 2007; Sudit et al., 2005)	2005-2010	A	Private	Manual	No	No	No	No	No	94
	(Byers and Yang, 2008; Fava et al., 2008; Yang et al., 2008)	2008	A	Simulation	Supervised	Yes	No	No	No	No	90
	(Lee et al., 2008)	2008	A	DARPA 2000	Manual	No	Yes	No	No	No	24
	(Katipally et al., 2011)	2011	A	Simulation	Supervised	No	No	No	No	No	6
	(Chen et al., 2016)	2016	E	Private	Manual	Yes	No	Yes	No	Yes	7
	(Holgado et al., 2017)	2017	A	DARPA 2000 Simulation	Manual	Yes	Yes	Yes	Yes	Yes	0

List of reviewed multi-step attack detection methods based on causal correlation in the categories of prerequisites and consequences and model matching.

2003, 2010; Ning et al., 2002a, 2002b, 2004, 2005; Xu and Ning, 2004, 2006; Zhai et al., 2006). Their method was developed in parallel to the one by the MIRADOR project, and independently according to Ning et al. (2004). The formalism of both methods is different, but the principles behind the correlation process are very similar.

As the proposal by Ning has evolved during a longer time span, many improvements related to graph reduction or analysis have been incorporated. For example, with focused analysis the security expert can focus the analysis on certain values of attributes, while link analysis gives some insight on the connection between attacks (Ning and Xu, 2010; Ning et al., 2002b). Graph reduction can be applied to extract the attack strategy if the method considers the derivation of the extended consequences of alerts (Ning and Xu, 2003). Ning's team has even proposed a different way to express the causal predicates (Xu and Ning, 2004), focusing on triggering events and applying a hierarchical taxonomy to the attributes in the alerts. Finally, they have also proposed the enrichment of IDS alerts with the inclusion of OS-level event logging (Zhai et al., 2006).

A vast research about prerequisites and consequences has been conducted inspired on the work by Cuppens and Ning. Cheung et al. (2003) propose CAML (Correlated Attack Modeling Language), another language for expressing causal sequences of attacks, and use it in the EMERALD project (Porras and Neumann, 1997). In another proposal (Wang and Jajodia; Wang et al. 2005, 2006) the performance of the method proposed by Ning et al. (2002a) is improved by adding a new element called Queue Graph. In this method, alerts are correlated only to the latest copy of each type of alert, not to all the past ones. Their attack graphs contain, apart from the causal information, the vulnerabilities of the attacked environment.

In the works by Pandey et al. (2008) and Zhou et al. (2007), predicates to express preconditions and postconditions are substitute by *capabilities*. A capability can define both the new possibilities of the attacker after performing a single-step attack and the requisites of the attack. This offers a universal way of describing the causal link in multi-step attacks. However, their assignation still requires manual work by a security expert. Another original alternative to substitute predicates (Yan (2005); Yan and Liu (2004)) is to use a case grammar where the relationship between actions is expressed in plain English. In this model, actions are connected as verbs in linguistics.

MARS (Multi-stage Attack Recognition System) (Alserhani (2012, 2013, 2016); Alserhani and Akhlaq (2011); Alserhani et al. (2010)) is the most complete of the multi-step attack detection proposals based on prerequisites and consequences. It is a direct extension of the work by Ning et al. (2002a, 2004). Using the method provided by MARS, Alnas et al. (2013) are able to model the behaviour of Zeus botnet.

6.2.2. *Statistical inference*

Statistical inference is the process of inferring from a dataset the distribution that generated it (Wasserman, 2013). It assumes that the information is already in the dataset of traces, we just need to know where to look. Methods in this category (Tables 7) work with the frequencies of actions and sequences of actions. A statistical model is automatically extracted from a training dataset of traces or from the own data where the detection process is made. The inferred statistics serve to build a probabilistic model that can be used for detection and prediction of further attacks. The most popular method of statistical inference, both in general and in multi-step attack detection, is probably Bayesian inference. Hidden Markov Models (HMM) are also popular to represent the conclusions extracted from statistical inference.

Geib and Goldman (2001) were pioneer in proposing an adaptation of plan recognition, a good established field in AI, to attack scenario detection and prediction. They propose the application of probabilistic reasoning to identify the intentions of the attacker. They do not propose a method to extract the information to be coded in the rules. However, the relevance of their work is that they opened a new way in security research based on plan recognition. The work by Zhuo Ning and Gong is also based on plan recognition (Ning and Gong, 2007). Their method needs a preliminary attack graph with associated belief values, so the detection of unknown attack is not possible.

Their influence is evident in the work by Qin and Lee. Their first published multi-step attack detection method (Qin and Lee, 2003) applies a correlation between individual alerts based on the Granger Causality Test, a time series-based causal analysis algorithm. Using this work as the basis, they propose a Bayesian technique (Qin, 2005; Qin and Lee, 2004, 2007) to correlate isolated attack scenarios. A set of Bayesian networks is derived from a manually crafted library of attack trees of high level alerts. The probabilities of the edges in each network are modified according to the probability of correlated scenarios. Jalili et al. are apparently inspired by this idea, but they do not show in their publications (Jemili et al., 2008, 2009) how they obtain the "system indicators of attack consequences and prior knowledge of attack transitions" that they use to link the nodes in their Bayesian models.

A Bayesian model is also used by Ren et al. (2010) but in this case the models are extracted from a training dataset in a first offline phase. All the features in the alerts are tested to find the ones which better represent the relevance of alerts representing attack steps. In a second phase, the selected features and probabilities are used to extract attack scenarios from a real-time stream of alerts. No previous knowledge about the attacks is needed and the Bayesian models are able to evolve during execution. A very similar approach is followed by Kavousi and Akbari (2012, 2014). They present a more abstract view of the resulting attack scenarios, easier to analyse by a human expert. On the other hand, in the algorithm proposed by Marchetti et al. (2011a) the level of correlation between nodes in the Bayesian graph depends on the time difference between the events and on the automatic analysis of past events. In a final step, the graph is pruned to remove the nodes whose correlation probability is lower than a dynamic threshold defined by the current statistics of dataset. Anbarestani et al. (2012) also propose a Bayesian-based method, where the relationship between the traces is only based on the IP addresses, taking the final objective of the attack, i.e. the IP address of vulnerable assets, as a reference to do the comparison.

Li et al. (2007c, 2007d) propose a method to identify the intrusion alerts with the highest probability of being followed by other attacks. The set of probabilities for each alert is computed from a set of candidate sequences extracted from a

Table 7 – List of reviewed multi-step attack detection methods based on causal correlation in the category of statistical inference.

Approach	References	Period	Type of data	Datasets	Knowledge extraction	A_m	A_d	A_k	Rep.	Attack model	Total cit.
Statistical inference	(Geib and Goldman, 2001)	2001	A	No exp.	Manual	No	No	No	No	No	195
	(Qin, 2005; Qin and Lee, 2003, 2004, 2007)	2003-2007	A	DARPA GCP DEFCON 9 Private	Automatic Manual	Yes	Yes	–	Yes	Yes	635
	(Ourston et al., 2003)	2003	A	Private	Supervised	Yes	No	Yes	No	No	136
	(Ning and Gong, 2007)	2007	A	DARPA 2000	Manual	Yes	Yes	Yes	Yes	Yes	2
	(Li et al., 2007c, 2007d)	2007	A	DARPA 2000 DARPA 1999	Supervised	Yes	Yes	Yes	Yes	Yes	54
	(Ma et al., 2008)	2008	A	DARPA 2000	Automatic	Yes	Yes	–	Yes	Yes	14
	(Jemili et al., 2008, 2009)	2009	A	DARPA GCP	Manual	No	Yes	No	No	Yes	2
	(Sadoddin and Ghorbani, 2009)	2009	A	DARPA 2000 Private Simulation	Automatic	Yes	Yes	–	Yes	Yes	47
	(Ahmadinejad and Jalili, 2009)	2009	A	DARPA 2000	Supervised	Yes	Yes	Yes	Yes	Yes	21
	(AmirHaeri and Jalili, 2009)	2009	A	DARPA 2000	Manual	Yes	Yes	Yes	Yes	Yes	0
	(Ren et al., 2010)	2010	A	DARPA 2000 Private	Supervised	Yes	Yes	Yes	Yes	Yes	54
	(Farhady et al., 2010)	2010	A	DARPA 2000	Supervised	Yes	Yes	Yes	Yes	No	2
	(Cipriano et al., 2011)	2011	A	UCSB 2008	Supervised	Yes	Yes	Yes	Yes	No	23
	(Bai et al., 2011)	2011	A	Simulation	Manual	Yes	No	No	No	Yes	5
	(Marchetti et al., 2011a)	2011	A	DEFCON 18	Automatic	Yes	Yes	–	Yes	Yes	15
	(Anbarestani et al., 2012)	2012	A	DARPA 2000	Automatic	Yes	Yes	–	Yes	Yes	8
	(Lagzian et al., 2012)	2012	A	DARPA 2000	Automatic	Yes	Yes	–	Yes	Yes	8
	(Luktarhan et al., 2012)	2012	A	DARPA 2000	Supervised	No	Yes	Yes	No	No	1
	(Man et al., 2012)	2012	A	DARPA 2000	Automatic	Yes	Yes	–	Yes	Yes	0
	(Soleimani and Ghorbani, 2012)	2012	A	DARPA 2000 Private	Supervised	Yes	Yes	Yes	Yes	Yes	13
	(Kavousi and Akbari, 2012), (Kavousi and Akbari, 2014),	2012-2014	A	DARPA 2000	Automatic	Yes	Yes	–	Yes	Yes	12
	(Bahareth and Bamasak, 2013)	2013	A	No exp.	Supervised	Yes	No	No	No	No	4
	(Brahmi and Yahia, 2013)	2013	A	NSA	Automatic	Yes	Yes	–	Yes	No	0
	(Kim and Park, 2014)	2014	A	Private	Automatic	No	No	–	No	No	17
	(Kholidy et al., 2014)	2014	A	DARPA 2000	Supervised	Yes	Yes	Yes	Yes	Yes	7
	(Xuewei et al., 2014)	2014	A	DARPA 2000	Supervised	Yes	Yes	Yes	Yes	Yes	10
	(Lv et al., 2015)	2015	A	Private	Automatic	Yes	No	–	No	Yes	0
	(Li et al., 2016)	2016	A	DARPA 2000	Automatic	Yes	Yes	–	Yes	Yes	0
	(Xian and Zhang, 2016), (Zhang et al., 2016)	2016	A	DARPA 2000 DEFCON 19	Automatic	Yes	Yes	–	Yes	No	0

List of reviewed multi-step attack detection methods based on causal correlation in the category of statistical inference.

training dataset using sliding time windows. The resulting graphs can be applied after verification to a set of new alerts for attack prediction.

There is much work where the candidate sequences are built after a feature match between the alerts. Therefore, in these methods, there is a previous phase of progressive construction by attribute matching (see Section 6.1.1), before deducing the probabilities. For example, the system Nexat (Cipriano et al., 2011) developed by Cipriano et al. groups into sessions the alerts for which there is some connection between source and/or destination IP addresses. Lagzian et al. (2012) also consider only the IP addresses as linking feature, applying Bit-AssocRule, a variation of the Apriori algorithm. In the case of Man et al. (2012), the features matched are the destination IP address, the attack type and the timestamp, while Kim and Park (2014) just use time differences and pairs of IP addresses, not giving much insight about the implementation. Other works (Ma et al., 2008; Xian and Zhang, 2016; Zhang et al., 2016) simply take the ID of the events and apply a sequential mining method to find frequent sequences.

In the work by Sadoddin and Ghorbani (2009), there is also a first phase of attribute matching, by source or destination IP address. However, it selects as potential attacks only the most frequent sequences. Their method, called FSP_Growth, automatically mines frequent patterns of IDS alerts and arranges them in a pattern tree, which is updated even during detection phase. FSP_Growth is based on the FP_Growth algorithm, used by Bai et al. (2011) in the search of frequent sequences. The attack patterns to search are extracted from a Bayesian model of possible attack scenarios.

Finding the most frequent patterns is also the objective of Brahmi and Yah (2013). They apply an improved version of the PrefixSpan algorithm after distributing the alerts and their attributes in multi-dimensional tables. This is not the only variation of PrefixSpan proposed for multi-step attack detection (Lv et al., 2015). Another standard sequential pattern mining algorithm is GSP. It is used in the RMARS framework (Bahareth and Bamasak, 2013) to extract patterns from a training dataset. The patterns can later be used in real time detection. We could not find any published experiments using RMARS.

The correlation matrix proposed by Zhu and Ghorbani (2006) (see Section 6.1.1) is reused by Ghorbani in a later work with Soleimani (Soleimani and Ghorbani, 2012). After a first phase of sequence identification using the correlation matrix, an episode mining algorithm is applied to find combinations of alerts. Multi-step attacks are identified through a supervised Decision Tree (DT) learning method. Other authors have also taken this perspective, focusing on the distribution of time windows (Ahmadinejad and Jalili, 2009). The RTEAS algorithm (AmirHaeri and Jalili, 2009) uses a manually crafted correlation matrix with the indexes relating couples of attack types.

In a very recent approach, Li et al. (2016) propose to group the alerts using a fast fuzz cluster algorithm which relies on similarity between IP addresses, ports and timestamps. Then, a frequent mining method is applied to the dataset, but only considering the alert type as feature.

Hidden Markov Models (HMM) have been also applied to statistical inference for multi-step attack detection. Ourston et al. (2003) were probably the first ones to use HMM in the detection of multi-step attacks. They use IDS alerts, which are previously classified into categories corresponding to each of the usual steps of a complex attack. The definition of an HMM always requires to precise how its steps are linked. In this case, the steps are linked by IP addresses. In the more recent paper by Luktarhan et al. (2012), there is not a clear explanation of how the attack scenario is defined. Farhadi et al., (2010) implement a method similar to the one by Ourston et al. but using classical Markov models, not HMMs. Classical Markov models are also used by Xuewei et al. (2014), who define the relationship between alert types only based on similarities between IP addresses.

Kholidy et al. (2014, 2014b, 2014c) show in a set of three publications that other variations of classical Markov models, such as Variable Order Markov Models (VMM), can be used to predict next steps in a multi-step attack. The approaches they propose highly depend on predefined models or signatures of attacks, but prediction is based on transition probabilities adapted from alert data.

6.2.3. *Model matching*

Methods doing model matching (Table 6) assume that every multi-step attack follows a certain structure. They model this structure and try to find sequences that adapt to them. Model matching methods are different from case-based ones, explained in Section 6.4, because the latter techniques use specific cases of real attacks. Model matching methods use a higher level of abstraction, representing the *skeleton* and general features of a multi-step attack.

For example, a possible model can be derived from the global phases of an APT proposed by the literature (Chen et al., 2014; Luh et al., 2016). The structure of this model would be that of a sequence of ordered stages corresponding to each one of the steps of the attack. Traces, previously associated to a phase, can be linked by their similarity and compared to the model.

This way of working is followed by INFERD (Mathew et al., 2010; Stotz and Sudit, 2007; Sudit et al., 2005). It uses a Guidance Template composed of different stages. Each of them represents a broad category of IDS alerts. INFERD aggregates alerts to the same model if they match the template and they have the same source and/or destination IP addresses.

There exists several systems proposing Markov models for high-level representation of IDS alert sequences. One of the proposal (Byers and Yang (2008); Fava et al. (2008); Yang et al. (2008)) uses variable-length Markov models (VLMM), derived from a set of known multi-step attacks. The projection on the model can depend on the attack description, the category of the attack or the destination IP address. Other methods use HMM for representing abstract models of well-known multi-step attacks (Chen et al., 2016; Fava et al., 2007; Katipally et al., 2011). The one proposed by Lee et al. (2008) considers distributed agents to detect each stage of the attacks. None of the mentioned approaches using HMM takes advantage of the possibilities offered by these models: they just use the formalism of HMM to represent the abstract sequences. But the model is not adapted to changes in the analysed data, or at least the authors do not give details about a training phase. On the other hand, a very recent proposal by Holgado et al. (2018) considers the HMM as a prediction model built from existing data. An HMM is built for each type of attack, instead of having only one general model. The parameters in the model are automatically derived from a training dataset of alerts, which are clustered by the similarity of the attack description with the content of Common Vulnerabilities and Exposures (CVE) documents.

6.3. *Structural-based methods*

Many methods (Table 8) consider the structure of the network as a key element for intrusion detection. They incorporate in their detection engine network information, specially about the vulnerabilities affecting each asset. This information is *structural* in the sense that it only depends on the defended systems and not on the actions of the attackers. The latter are deduced from the former but no real evidence from traces is used. Structural information is usually coded in the form of an attack graph. An attack graph is an abstract representation of the network containing the vulnerabilities of the systems in each node (Sheyner et al., 2002). Structural-based methods project incoming traces into the attack graph representing the defended network. Their objective is to use this projection to predict future steps of ongoing attacks. The process of prediction relies on the hypothesis of possible attack paths, solely deduced from the attack graph.

Noel et al. (2004) were probably the first ones to project received IDS alerts to prebuilt attack graphs in 2004. Vulnerabilities are extracted using a network scan, and they are identified with a known exploit. Alerts are matched to the exploits if their correlativity is above certain threshold. Correlation between alerts depends on the distance between the corresponding exploits.

While some research uses traditional attack trees, in standard (Fayyad and Meinel, 2013; Roschke et al., 2011) or enhanced (Çamtepe and Yener, 2007) form, the network model used in TANDI (Holsopple et al., 2006; Yang et al., 2009) also includes information about the level of access privilege. TANDI is developed at the Rochester Institute of Technology (RIT), the birthplace of some of the most relevant ideas about structural-based multi-step detection. One of these ideas is the *cyber terrain* or *virtual terrain*, proposed by Fava et al. (2007) and Holsopple and Yang (2008). It is a manual model of a network where each asset is associated to its services, its version and the logical connections with other assets. The idea of cyber terrain is also appealing for vulnerability assessment. We consider it as a more important contribution by Fava et al. than their use of HMM for model matching, already cited in Section 6.2.3.

Also born at the RIT, the system proposed by Du et al. (2010) considers four categories: current state of the attacking source, current state of the target, firewall rule configuration and open services at the target. Prediction is made using two methods. The first one uses Transferable Belief Model (TBM) to combine the concepts of Capability and Opportunity assessments developed in the FuSIA system (Holsopple and Yang, 2008). The second one uses Fuzzy inference to merge Variable Lenght Markov Model (VLMM) estimates based on the attributes extracted from the alerts.

Chien and Ho (2012) present a system based on coloured Petri nets. They model each attack plan in an abstract way and detection is made considering a metric of exploit certainty. On the other hand, Zhang et al. (2008) propose to automatically

Table 8 – List of reviewed structural-based multi-step attack detection methods.

Approach	References	Period	Type of data	Datasets	Knowledge extraction	A_m	A_d	A_k	Rep.	Attack model	Total cit.
Structural-based	(Noel et al., 2004)	2004	A	Simulation	Manual	Yes	No	Yes	No	No	158
	(Holsopple et al., 2006)	2006	A	Simulation	Manual	No	No	No	No	No	49
	(Çamtepe and Yener, 2007)	2007	A	Simulation	Manual	Yes	No	No	No	Yes	62
	(Fava et al., 2007; Holsopple and Yang, 2008)	2007–2008	A	Private Case study	Manual	Yes	No	Yes	No	Yes	65
	(Zhang et al., 2008)	2008	A	Simulation	Automatic	Yes	No	–	No	No	34
	(Du et al., 2010)	2010	A	Simulation	Manual	Yes	No	No	No	No	18
	(Roschke et al., 2011)	2011	A	Simulation	Manual	Yes	No	No	No	No	74
	(Chien and Ho, 2012)	2012	A	DARPA 2000	Manual	Yes	Yes	Yes	Yes	Yes	2
	(Lin et al., 2012)	2012	A	Case study	Manual	Yes	No	Yes	No	Yes	9
	(Fayyad and Meinel, 2013)	2013	A	No exp.	Manual	Yes	No	No	No	No	4
	(Luo et al., 2014)	2014	A	Case study	Manual	Yes	No	Yes	No	Yes	5
	(Luo et al., 2016)	2016	A	Case study	Manual	Yes	No	Yes	No	Yes	0

build the trees used for detection. They do that assembling different elements (information about topology, vulnerability scan results, etc.) through the principles of causal correlation.

Some attempts have been made to apply Game Theory to capture the behaviour of an attacker and a defender during the execution of a multi-step attack. Most of these efforts are out of the scope of our survey, as they are only based on structural data and just applied to risk assessment (Haopu, 2016; Liu et al., 2005; Lye and Wing, 2005; Rass et al., 2017; Xupeng et al., 2014). Nonetheless, some of them can be applied in a production environment. In these proposals, the defender reduces the uncertainty of the attack through the signals received from IDS alerts, placing defences in real time in order to avoid damages. For example, Lin et al. (2012) propose a model focused on attackers whose objective is to steal confidential data. And Luo et al. (2014) present an algorithm called RDFP (Responses by Dynamic game tree-based Fictious Play) to be applied in a dynamic game tree.

To end with structural-based methods, we want to mention a method for the very particular context of software-defined home networks (SDHN) (Luo et al., 2016). It is a good example of specialisation in multi-step attack detection. Anyway, the projection method they propose could be adapted to other environments, as the attack graphs can contain information from any type of network.

6.4. *Case-based methods*

A broadly spread approach for intrusion detection is the comparison between the observations and a knowledge-base of previously seen attacks. There are many methods applying this approach to multi-step detection (Table 9). Attacks are in this case represented by scenarios or sequences of actions. The knowledge-base can be manually populated by security experts or attacks can be extracted from a dataset using automatic techniques. Honeypots can aid in the collection of real multi-step attacks for the development of case-based signatures (Vasilomanolakis et al., 2016). Modelling the behaviour of human actors involved in security (Dutt and Kaur, 2013) can also be important to better understand how to develop detection rules and even how to train new security professionals. Newly discovered multi-step attacks, through an inference mechanism or human intervention (Salah et al., 2013), can be added to the database once they have been analysed and a model is built.

It is important to highlight the differences between case-based methods, model matching, structural-based methods and methods based on prerequisites and consequences, as their resemblance can lead to confusion. In case-based methods, models of attacks are built manually and represent a specific type of multi-step attacks. Conversely, in model matching methods models are highly abstract and assumed valid for every multi-step attack, or at least for an ample set of them. In structural-based methods, models are built only upon structural network characteristics. They do not use any information about the attack, but hypothesise it from the vulnerable elements and the possible paths in the network. That means that a multi-step attack is defined as an exploited vulnerability in the network element A followed by another exploited vulnerability in element B, regardless of how they are exploited. Finally, it is true that in methods based on prerequisites and consequences there is a predefined model for each alert, so theoretically all the possible combinations of sequences representing attacks could be built in advance and be applied to detection using a case-based method. However, when building the knowledge database of prerequisites and consequences the steps of the attacks, represented as alerts, are built one by one without defining the sequence of the multi-step attack. The alerts are later assembled in an automatic way, through the linking of prerequisites and consequences. This is what gives methods based on prerequisites and consequences its flexibility and the capability of including more cases than case-based methods.

The clear advantage of case-based methods is that the number of false positives is low: we know what we search and we look for the exact occurrence of it, so it is difficult to miss the mark. Nevertheless, if an attack is not in the database it is not found. Case-based methods are only capable of detecting known multi-step attacks.

Most of the work in case-based detection uses IDS alerts as an input. They are based on a library of handmade attack scenarios, which are generally expressed in the shape of a general graph (Mathew et al., 2005). These libraries can contain additional information that can help in the detection of the attack, such as the criticality of the assets in the network (Soleimani and Ghorbani, 2008).

Attack scenarios can be stored using any standard language for rule representation. For example, Chintabathina et al. (2012) propose A-prolog in order to apply logic programming to case-based detection. XML is also used in some work (Long and Schwartz, 2008). However, there are many languages that have been specifically created to model multi-step attacks. We review here only the languages used in a multi-step attack detection method.

STATL (State Transition Analysis Technique Language), developed by Eckmann et al. (2002), is a transition-based language to develop multi-step attack signatures. It was conceived for centralised case-based detection, but it was later used in other approaches (Valeur et al., 2004). While STATL is specifically thought for being used by a low-level search engine, Morin and Debar (2003) propose to apply a high-level description based on Dousson's chronicle formalism (Dousson, 1994). They propose the integration of chronicle, which was conceived to model dynamic systems, with the M2D2 platform (Morin et al., 2002), a framework for event and structural information management. The chronicle formalism has been later used by other authors (Wang and Ma, 2005; Wang et al., 2004).

Another language for representing multi-step attack signatures is EDL, developed by Meier (2007). It has been later improved by Jaeger et al. (2015). This new version is complemented with a method for automatic derivation of multi-step EDL signatures from taint graphs (Ussath et al., 2016a). These taint graphs are created through semiautomatic correlation of the steps in multi-step attack examples (Ussath et al., 2016b). In EDL, a sequence of steps composing an attack scenario is represented as a sequence of nodes in a coloured Petri net. Although the general idea looks simple, its formal specification includes very innovative mechanisms for rule definition, as the use of moving tokens as search agents.

We can find very simplistic methods (Kannadiga et al., 2007; Katipally et al., 2010; Panichprecha et al., 2007; Xuewei et al.,

Table 9 – List of reviewed case-based multi-step attack detection methods.

Approach	References	Period	Type of data	Datasets	Knowledge extraction	A_m	A_d	A_k	Rep.	Attack model	Total cit.
Case-based	(Eckmann et al., 2002)	2002	E	Case study	Manual	Yes	No	Yes	No	No	496
	(Kruegel et al., 2001)	2002	E	Private	Manual	Yes	No	No	No	No	131
	(Morin and Debar, 2003)	2003	A	Case study	Manual	Yes	No	No	No	Yes	233
	(Wang and Ma, 2005; Wang et al., 2004)	2004–2005	A	Case study	Manual	Yes	No	No	No	Yes	4
	(Mathew et al., 2005)	2005	A	Private	Manual	Yes	No	No	No	No	36
	(Xiao and Han, 2006)	2006	A	DARPA 2000 Private	Manual	No	Yes	No	No	Yes	7
	(Chien et al., 2007)	2007	A	DARPA 2000	Manual	No	Yes	No	No	No	13
	(Kannadiga et al., 2007)	2007	A	Simulation	Manual	No	No	No	No	Yes	6
	(Li et al., 2007a; Zhang et al., 2007) (Li et al., 2007b; Wang et al., 2007)	2007	A	DARPA 2000 Private	Manual	Yes	Yes	No	No	No	54
	(Panichprecha et al., 2007)	2007	E	Case Study	Manual	Yes	No	No	No	Yes	1
	(Long and Schwartz, 2008)	2008	A	DARPA 2000 DARPA GCP	Manual	Yes	Yes	No	No	No	5
	(Soleimani and Ghorbani, 2008)	2008	A	DARPA 2000	Manual	Yes	Yes	No	No	Yes	8
	(Katipally et al., 2010)	2010	A	No exp.	Manual	No	No	No	No	No	11
	(Xuewei et al., 2010)	2010	A	DARPA 2000	Manual	Yes	Yes	No	No	Yes	9
	(Vogel and Schmerl, 2011; Vogel et al., 2011)	2011	E	Private	Manual	No	No	No	No	No	1
	(Chintabathina et al., 2012)	2012	A	No exp.	Manual	No	No	No	No	No	3
	(Zali et al., 2012a, 2012b)	2012–2013	A	DARPA 2000	Manual	Yes	Yes	Yes	Yes	Yes	22
	(Giura and Wang, 2012a, 2012b)	2012	E	Private	Manual	Yes	No	No	No	No	70
	(Zargar, 2013)	2014	A	DARPA 2000, ISCX	Manual	Yes	Yes	No	No	Yes	14
	(Jaeger et al., 2015; Ussath et al., 2016a, 2016b)	2015–2016	E	Private, Simulation	Manual	Yes	No	Yes	No	Yes	5
	(Navarro et al., 2016)	2016	E	Simulation	Manual	Yes	No	No	No	No	0

2010), just based on pure pattern matching. Some of them embellish the process with a previous phase, of alert aggregation in the case of Xiao and Han (2006). But we can find more complex proposals, where the system incorporates additional mechanisms to improve the detection. For example, MASP (Mining Attack Sequential Pattern) (Li et al., 2007a, 2007b; Wang et al., 2007; Zhang et al., 2007) uses incremental mining of subsequences of known multi-step attacks. Thanks to that, MASP can find variations of the attacks stored in the knowledge base. Zali et al. propose a similar method (Zali et al., 2012a, 2012b). They represent the attack scenarios as Causal Relation Graphs (CRG), with queues of alerts placed in each of the vertex of the graph. This structure eases the implementation for real-time detection and the prediction of missing alerts. The correlation system proposed by Chien et al. (2007) works with primitives representing parts of attack scenarios. Primitives are built from a predefined ontology, but not enough detail is given about how they are created. The authors explain the correlation process in very general terms, but their idea of decomposing the attack scenario in parts is interesting. ONTIDS framework (Zargar, 2013) is also based on an ontology, which works at different levels: context, alert, attack and vulnerability. The connection between all these levels is exploited in the definition of detection rules.

Giura and Wang (2012a, 2012b) propose a model for APT detection that can work with general events, not only with IDS alerts. The stages of the attack are arranged in a layered pyramid, with the goal at the top of the pyramid and the previous steps stratified by layers. In each face of the pyramid we can find different domains: physical, network, application, user, etc. This model is used in a detection framework where correlation rules are based on signatures, profiling or security policies.

Also working with general events, Morwilog (Navarro et al., 2016) is a system inspired by the behaviour of foraging ants. The arrival of events to the system triggers the generation of artificial ants, which look for the best paths in trees that contain possible sequences of malicious events. Morwilog works with the assistance of a human expert, who helps the system to learn through the identification of false positives and real detections. Even if in the paper the event trees are randomly generated, the system is conceived to benefit from manually crafted trees with expert knowledge.

The case-based methods presented so far are centralised: traces are collected in a central point, where detection of attack scenarios takes place. However, there are also some distributed methods where detection of individual steps is done by local agents scattered through the network, who give a coordinated response to multi-step attack detection. This kind of intrusion detection system has been throughly studied as a particular field of security detection (Zhou et al., 2010).

An example of a distributed system applying case-based detection is Quicksand (Kruegel et al., 2001), where handmade multi-step attack signatures are translated into pattern graphs and sent to local agents. Each agent is responsible of detecting the fraction of the signature represented in a node of the graph. Each time an agent detects its assigned pattern, a message is sent up in the tree, until Quicksand identifies the full signature, represented in the root node, and raises an alert. There exists a more recent distributed proposal (Vogel and Schmerl (2011); Vogel et al. (2011)), based on the Petri net principle and using signatures written in EDL (Meier, 2007). Signatures are divided in minimal parts and sent to the local agents. In both mentioned distributed methods, rule division has to be made by hand.

6.5. *Mixed methods*

As we have mentioned earlier, there are some works where several approaches are integrated together in the same system, with no clear prevalence of one over the other (Table 10).

For instance, the framework RTECA (Ramaki et al., 2015) combines frequency analysis, similarity of alert types through a correlation matrix and matching of IP addresses and ports. Frequent sequences are arranged in event trees, which are fed with similar sequences during the execution. The analysis of frequent patterns can be also combined with a phase of clustering (Faraji Daneshgar and Abbaspour, 2016). The creators of RTECA also propose in another paper (Ramaki et al., 2015) a very similar system to their previous one but based on Bayesian networks. Lessons learned from both systems are incorporated in a very recent three-phase framework developed by the same authors (Ramaki and Rasoolzadegan, 2016), where there is no dependence of a predefined correlation matrix.

In ASEA system (Farhadi et al., 2011), statistical relationships between alerts and attribute correlation are merged to apply plan recognition using HMMs. On the other hand, Shittu proposes in the third chapter of her PhD thesis (Shittu, 2016) to combine Bayesian inference with attribute correlation, offering a different perspective to existent Bayesian methods. Furthermore, Du et al. (2009) start by identifying sequences of IDS alerts with the same victim IP address and assigning a severity score to each step. Then, the found sequences are mined using three techniques, each of them based on a totally different mechanism: Longest Common Subsequences (LCS), Fourier transform and social networks.

There is much work combining prerequisites and consequences with other approaches. Ning et al. propose to mix their method with similarity-based methods (Ning and Xu, 2004; Ning et al., 2004). Their purpose is to merge the graphs belonging to the same attack scenario that are mistaken as separate because an IDS alert is missing. Yu and Frincke (2004, 2007) combine causal correlation and statistical inference, applying coloured Petri net with hidden states to multi-step attack detection. Another proposal is that of Saad and Traore (2012), who present a phase of clustering using an intrusion ontology before applying the connection between prerequisites and consequences. A similar method is used by Al-Mamory and Zhang (2007), who additionally propose the application of an Attribute Context-Free Grammar to model the multi-step attacks (Al-Mamory and Zhang, 2008, 2009). This grammar contains information about the level of similarity between the alerts, their prerequisites and consequences and the structure of known attack scenarios.

We find also some publications proposing the combination of methods that we have already mentioned in the context of other approaches. This is mostly done by the same authors who developed the original methods. For instance, causality-based INFERD (Mathew et al., 2010; Stotz and Sudit, 2007; Sudit et al., 2005) is integrated with the structural-based TANDI system

Table 10 – List of reviewed mixed multi-step attack detection methods.

Approach	References	Period	Type of data	Datasets	Knowledge extraction	A_m	A_d	A_k	Rep.	Attack model	Total cit.
Mixed	(Valeur et al., 2004)	2004	A	DARPA 2000 Others	Manual	Yes	Yes	No	No	Yes	505
	(Ning and Xu, 2004; Ning et al., 2004)	2004	A	DARPA 2000	Manual	Yes	Yes	Yes	Yes	Yes	268
	(Yu and Frincke, 2004, 2007)	2004–2007	A	DARPA 2000 DARPA GCP	Supervised	Yes	Yes	Yes	Yes	Yes	113
	(Wang et al., 2006)	2006	A	DARPA 2000	Manual	Yes	Yes	No	No	No	5
	(Al-Mamory and Zhang, 2007, 2008, 2009)	2007–2009	A	DARPA 2000 DEFCON 8	Manual	Yes	Yes	No	No	Yes	36
	(Yang et al., 2009)	2009	A	Private, Simulation	Manual	No	No	Yes	No	No	77
	(Du et al., 2009)	2009	A	Simulation	Automatic	Yes	No	–	No	No	10
	(Farhadi et al., 2011)	2011	A	DARPA 2000	Automatic	Yes	Yes	–	Yes	Yes	27
	(Marchetti et al., 2011b)	2011	A	DEFCON 18	Automatic	No	Yes	–	No	Yes	7
	(Saad and Traore, 2012)	2012	A	DARPA 2000 UCSB 2002	Manual	Yes	Yes	No	No	No	7
	(Ahmed, 2014)	2014	A	DARPA 2000, ISCX	Manual	Yes	Yes	No	No	Yes	1
	(Chen et al., 2014)	2014	A	Private	Automatic	Yes	No	–	No	Yes	1
	(Abreu et al., 2015)	2015	E	No exp.	Manual	No	No	No	No	No	1
	(Ramaki and Rasoolzadegan, 2016; Ramaki et al., 2015)	2015–2016	A	DARPA 2000 DARPA GCP ISCX	Automatic	Yes	Yes	–	Yes	Yes	22
	(Faraji Daneshgar and Abbaspour, 2016)	2016	A	DARPA 2000, ISCX	Automatic	Yes	Yes	–	Yes	Yes	1
	(Shittu, 2016)	2016	A	DARPA 2000 Private	Supervised	Yes	Yes	Yes	Yes	Yes	1

in the work by Yang et al. (2009). Furthermore, the SF algorithm (Wang et al., 2006) is combined with MASP in the SATA platform (Wang et al., 2006). Marchetti, Colajanni and Manganiello propose a framework (Marchetti et al., 2011b) to combine their two approaches, the one using Self-Organizing Maps (Colajanni et al., 2010; Manganiello et al., 2011) and their pseudo-Bayesian algorithm (Marchetti et al., 2011a).

There are some frameworks in the literature proposing a whole end-to-end correlation process focused on multi-step attack detection. In the one proposed by researchers from Palo Alto Research Center and Galois Inc (Abreu et al., 2015), a mixture of methods are applied in different stages, from activity classification to alert ranking. Valeur et al. (2004) introduce a whole correlation system divided in several phases. The phases related to our research are the four aiming to link different alerts in order to compose scenarios: thread reconstruction, session reconstruction, focus recognition and multi-step correlation. WMAPRM (Chen et al., 2014) also merges different detection approaches for the specific case of wireless data, where the level 2 of the OSI model is more relevant than the network level. Another correlation framework is the one proposed by Ahmed (2014), where attack scenario construction is made combining semantic-based clustering and analysis of predefined consequences of alerts. He also proposes previous phases for alert aggregation and verification.

6.6. *Architectures for the detection*

There are some works proposing systems for multi-step attack detection from an architectural point of view but without defining specific methods or algorithms to identify the threat. As we have indicated before, publications in this section are not included in the main corpus of this survey.

We find in this category which is probably the first system conceiving the attack strategy as an important point for intrusion detection, in 1999 (Huang et al., 1999). It is based on a communication protocol between IDS agents, which are distributed through the network and doing local analysis and classical intrusion detection. A master controls their behaviour after the identification of possible strategies. Another most recent proposal of distributed architecture is FCDS (Federated Cyber Defense System) (Bereziński et al., 2012). It is intended to coordinate a Federation of Systems (FoS), a set of connected heterogeneous systems that collaborate in the detection of multi-step attacks.

Very much focused on alert aggregation, Debar and Wespi (2001) conceived an architecture for a console working on top of the Tivoli Enterprise Console (TEC), a commercial product. Their paper does not give much details about the implementation but it is still a much cited architecture of reference for multi-step attack detection.

In the survey about alert correlation written by Salah et al. (2013), an architecture of reference is built from the options proposed in the literature. They propose an end-to-end process composed of four stages: preprocessing, reduction, correlation and prioritisation.

The architecture proposed by Ficco and Romano (2011) takes a perspective of detection based on prerequisites and consequences. Their knowledge base is associated to a specific ontology (Ficco and Romano, 2010) for attack scenario reconstruction.

Bhatt et al. (2014) propose, without showing much detail about the implementation, a framework for APT detection from events using Hadoop.

Directly working with network packets, DFA-AD (Sharma et al., 2016) is a complete architecture to detect APTs. It allows the combination of different detection algorithms, whose results are correlated. A voting process returns the final result.

Finally, it is interesting the idea of *attack narratives* proposed by Mireles et al. (2016). They propose a methodology to associate pieces of traffic to a multi-step APT model.

7. Global state of the field

After the review of all methods in the corpus, in this section we study the domain from a global perspective. We are interested in the information used in multi-step attack detection and which experiments are presented in order to support the validity of the methods. The reason is that they are important points to evaluate the scientific quality of any domain in research. We see in this section that in the past 17 years since the beginning of multi-step attack detection, the field has not been totally consolidated in scientific terms. A sign of this is that the number of citations of each work, shown in the last column of (Tables 4-10), is low or zero for most of the cases.

We start by the analysis of type of trace used by each method in its search of indications left by multi-step attacks. Then, we study the ways used to extract the knowledge about multi-step attacks. We continue by evaluating the experiments made to prove the validity of the methods and the datasets used in them. To conclude this section, we discuss about the reproducibility of the presented methods.

7.1. *Types of trace*

As we have said in Section 2, there are several types of traces reflecting actions that happen in the IT system: packets, events and alerts. The methods studied in this survey analyse the traces in the search of multi-step attacks. It is important to remember the definition of alert as a specific type of event that indicates an alleged malicious activity or fault. Alerts generally come from an IDS.

In Table 1 we show how many publications analyse each type of trace. We distinguish three groups: methods only fed with alerts, those using general events, and those using traces with triggering alerts. This last group can be considered a hybrid between the two first ones. It contains methods using alerts to identify the potential presence of a multi-step attack but later working with other types of traces that are related to the triggering alert.

We can see there is a disparity in the use of the different types of traces: more than 85% of the publications are exclusively focused on the analysis of alerts. According to Brogi and Tong (2016), detecting a multi-step attack is the same as highlighting the links between elementary attacks. It is therefore logical to start the search of multi-step attacks with the individual attacks revealed in the alerts.

Despite this, there are still some methods using general events in the multi-step attack detection process. An alert is also an event, as we saw in Section 2, so alerts can also be considered by these methods. We have found just one method (Mathew and Upadhyaya, 2009) that works with general events but does a distinction between the treatment given to alerts and the rest of events.

There are also other kinds of methods where an alert triggers the detection process and then the attack scenario is built from other types of trace: packets (Chen et al., 2006; Shaneck et al., 2006; Strayer et al., 2005) or events (King et al., 2005; Zhai et al., 2006). These methods take advantage of the alerts as indicators of something suspicious happening and they develop the investigation on other traces from the information contained in this triggering alert, being able to compose the attack scenario with individual elements that are not a threat by themselves.

Further detail of the type of data used by each one of the methods is shown in (Tables 4-10). The following code is used: A: Only alerts. E: General events. T: Traces with triggering alerts.

7.2. *Origin of knowledge about attacks*

Another way of differentiating the publications is by the origin of the information about multi-step attacks in the detection process. According to this, we classify methods into three categories: manual, supervised or automatic extraction. The distribution of the publications in the corpus of the survey according to the mentioned categories is shown in Fig. 8. The category associated to each of the reviewed work is shown in (Tables 4-10), under the column "Knowledge extraction".

In *manual* methods (Cuppens and Miège, 2002a; Eckmann et al., 2002; Ning et al., 2002a; Valeur et al., 2004), the knowledge about the attack is manually coded by an expert. They search for already known attacks or slight variations of them. All reviewed case-based methods (Section 6.4) and all the methods based on prerequisites and consequences (Section 6.2.1) are of this type. It is important to note that we also consider as manual the methods requiring precoded templates of attacks in some stage, even if they include other phases of automatic extraction (Saad and Traore, 2012).

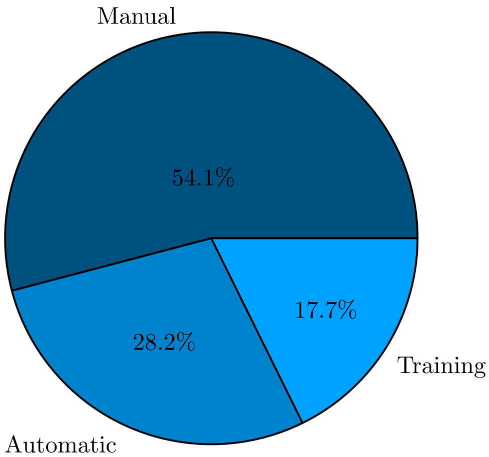

Fig. 8 – Distribution of publications according to the origin of the knowledge about attacks.

Supervised methods (Cheng et al., 2011; Dain and Cunningham, 2001a; Ourston et al., 2003; Yu and Frincke, 2007) count with a preliminary phase of automatic learning from a training dataset. They use supervised machine learning techniques to extract the knowledge about the attacks contained in the training dataset and then apply detection on a test dataset.

Finally, there are other methods approaching detection without any previous knowledge about the attacks to detect. We say that these methods are *automatic* (Cuppens, 2001; Julisch, 2003a; Qin and Lee, 2003; Zhu and Ghorbani, 2006). They learn from the same dataset where detection takes place, in real time. All methods classified under scenario clustering in the taxonomy shown in Fig. 6, for example, are considered as automatic.

The clear advantage of automatic methods over the others is that they could find unknown multi-step attacks. The desired goal of any security system is to be as much autonomous as possible in the detection of attacks. However, it is difficult to reach such a goal considering the current state of research in intrusion detection. Automatic methods are not still fully reliable and return a lot of false positives. It is true that they can eventually inform about an unknown attack, but we should wonder if the work done identifying true positives among the returned alerts is not heavier than the development of rules to detect known cyberattacks.

On the other side, supervised methods greatly depend on the availability of a sound and reliable dataset for the training phase. The dataset has to faithfully represent traces found in real networks in order to be effective in detection. Available datasets are far from reflecting the complex situations we can find in the real world, mainly because of privacy and security issues that impede the organisations to make their network data public.

7.3. *Types of data for experiments*

The validity and effectiveness of the proposed methods are tested by experiments. There are basically three types of data for experiments in the corpus of publications included in this survey: datasets of traces, synthetic data expressly simulated for the tests and attack examples. Among the datasets of traces, we can find public and private ones depending on their open availability to the rest of the research community.

We have classified the 181 publications in the corpus in five categories, according to the type or types of data used in their experiments:

- **Public dataset**. At least an experiment with a public dataset is made. A public dataset can be obtained by any research team, so the experiments can be easily reproduced, provided that the method used is explained with enough detail and, when necessary, expert knowledge is furnished.
- **Private dataset**. Method is tested with experiments based on private datasets of traces, whether also using simulated data (Holsopple et al., 2006; Yang et al., 2009) or examples. Experiments cannot be reproduced exactly as they are presented.

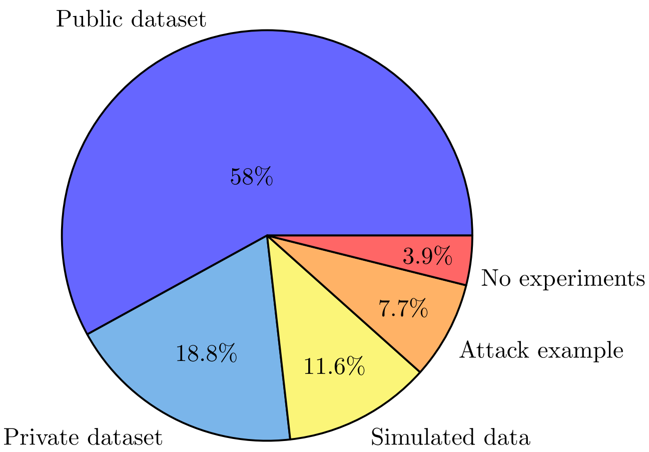

Fig. 9 – Distribution of publications according to the type of data for experiments.

- **Simulated dataset**. Experiments are done using simulated data expressly created to test the method. This data has not been thoroughly tested by third parties, so it is not fully reliable. If the authors make this data openly available, at least other research teams could reproduce the experiment and analyse the validity of the synthetic dataset.
- **Attack example**. Only examples of specific cases are presented to justify the validity of the method.
- **No experiments**. Method is not tested with experiments.

Fig. 9 shows the distribution of publications according to the type of data used in the experiments they carry out. More than half of the publications perform experiments with at least a public dataset. We will talk about existent public datasets below, in Section 7.4. 18.3% of the methods test their experiment using only private datasets containing real traces, which are not shared with the rest of the research community. The remaining quarter of the methods includes methods tested by simulated data, just presenting a case study or not doing experiments at all.

About those methods using private datasets, several of them use data generated by Skaion Corporation (Mathew and Upadhyaya, 2009; Shaneck et al., 2006). Access to this dataset is limited to official U.S. government's research. Some other sources use data collected from their own university (Chen et al., 2014) or company (Julisch, 2001), or from partners (Skopik et al., 2014; Sudit et al., 2005; Zhang et al., 2015).

The type of dataset used in the experiments presented in each publication is shown in (Tables 4-10). We indicate when a private dataset is used ("Private"), when there is just an attack example ("Case study"), when there is a simulation with data expressly created for the experiment ("Simulation") and when there are no experiments ("No exp.").

We should ask ourselves why there is still an ample portion of the publications proposing experiments on private data, which are not reproducible. This is a problem that also extends to the rest of security research. The reason of this secrecy is the sensitive nature of security data. Traces collected from a real network contain a lot of sensitive information, both in terms of security and privacy.

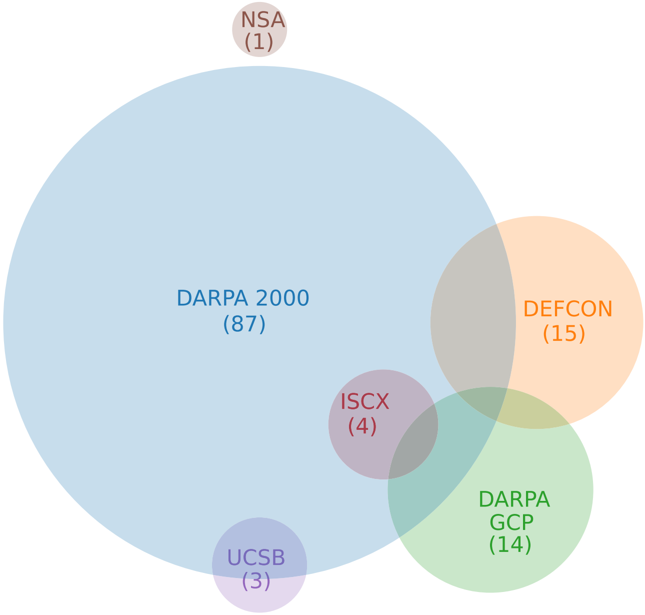

Fig. 10 – Venn diagram representing the usage of public datasets in the corpus of selected publications (Meta-Chart, 2017).

7.4. *Public datasets*

In our research about multi-step attack detection methods, we have found a series of public datasets that we describe below. The number of experiments done with each of the datasets is shown in the Venn diagram of Fig. 10. We should note that in this diagram we consider the total number of experiments using the dataset and that each publication can include several experiments to test the proposed method. There is more information about the datasets used by each method in the Appendix. We must point out that an evaluation of these public datasets in terms of quality is out of the scope of this survey.

7.4.1. *DARPA 2000*

It is undoubtedly the most important public dataset to test multi-step attack detection methods. It was sponsored by DARPA, the Defense Advanced Research Projects Agency in the United States. It was generated using the same test bed network as in DARPA 1998 and 1999 but incorporating multi-step attacks. This allows mixing the traffic and events in these older versions with the ones contained in DARPA 2000 to make the search more difficult (Li et al., 2007c, 2007d). DARPA 2000 was generated by the MIT Lincoln Laboratory in order to be used by the emerging methods for high-level attack analysis and goal recognition (Valeur et al., 2004). It contains network traffic from two networks and BSM audit data coming from the affected Solaris machines. However, researchers generally focus on the alerts returned by an IDS after it has processed the network packets. We have found only one reference (Anming and Chunfu, 2004) using the BSM audit data. DARPA 2000 is actually composed of two datasets, each one containing instances of a slightly different multi-step attack. The first one is LLDOS 1.0, which starts with a phase of probing, followed by the intrusion and culminated on the installation of a DDoS launcher

(Ning et al., 2002a). The second one, LLDOS 2.0, is similar to the first one but more sophisticated. Most of the experiments using public datasets are tested on DARPA 2000 (Ning et al., 2002a; Wang et al., 2006; Zhu and Ghorbani, 2006). Some of them also includes experiments with other public (Cheng et al., 2011; Wang et al., 2005) and private datasets (Ren et al., 2010; Zhou et al., 2007).

7.4.2. *DARPA GCP*

DARPA is also the author of DARPA GCP datasets, derived from their Grand Challenge Problem project. These are examples of attack scenarios composed of alerts coming from a set of heterogeneous sources: network-based IDS, host-based IDS, firewalls and network management systems (Qin and Lee, 2003). The dataset is composed only by alerts (Cheung et al., 2003), which are stored in XML files in IDMEF format. There are several versions of DARPA GCP. We have identified in the corpus of our survey the use of versions 2.0 (Cheung et al., 2003), 3.1 (Qin and Lee, 2004; Ramaki et al., 2015), 3.2 (Cheng et al., 2011; Xu and Ning, 2004) and 4.1 (Long and Schwartz, 2008).

7.4.3. *DEF CON*

The conference DEF CON is yearly organised in Las Vegas, United States since 1993. Each year since 1996 a “capture the flag” contest takes place in the context of the conference (Dain and Cunningham, 2001a). The contest consists of two teams with an opposite objective: one of them runs a set of services and the other one tries to compromise them. The set of DEF CON datasets are built from the traces left by the attackers during the contest. Each dataset is identified by the edition number of the conference. In the corpus of this survey we have observe the use of versions 8 (2000) (Dain and Cunningham, 2001a; Ning et al., 2002b; Xu and Ning, 2006), 9 (2001) (Qin and Lee, 2003), 18 (2010) (Manganiello et al., 2011) and 19 (2011) (Zhang et al., 2016).

7.4.4. *UCSB capture the flag*

This dataset of alerts is the result of another hacking competition yearly sponsored by the University of California, Santa Barbara (UCSB). In this competition, a copy of a multi-host network is presented to each team. The goal is to compromise the network through a series of attacking steps without being detected by the installed security sensors, both signature-based and anomaly-based IDS (Cipriano et al., 2011). Among the multi-step attack detection methods included in this survey, only data from the 2002 (Saad and Traore, 2012), 2004 (Wang and Jajodia) and 2008 (Cipriano et al., 2011) competitions are used. In 2002, it was still a local UCSB competition, and it became international in 2004, acquiring the name iCTF (international Capture The Flag). The framework used in the competitions is now publicly available (U.S. Barbara, 2015).

7.4.5. *NSA*

The National Security Agency of United States also provides a dataset of logs and alerts captured in the period between November 2008 and November 2011, during a security exercise (U.S.M. Academy, 2009). The only method of the corpus tested with this dataset is the one developed by Brahmi and Yahia (2013), who only use the dataset of Snort IDS alerts, and not the logs.

7.4.6. *ISCX UNB intrusion detection evaluation*

It is the most recent dataset used in the experiments of the publications included in the corpus. It was created in 2011 (Ahmed, 2014) by the Information Security Centre of Excellence (ISCX) of the University of New Brunswick (Shiravi et al., 2012), directed by Ali A. Ghorbani. The dataset is composed of labelled network traffic collected over 7 days, and it contains 4 multi-step attacks. Only a few projects have used it for their experiments (Ahmed, 2014; Faraji Daneshgar and Abbaspour, 2016; Ramaki and Rasoolzadegan, 2016), even if it is the most solid proposal of public dataset: not only it is the most recent one but it has been specifically created for multi-step attack detection research by one of the most reputed security laboratory in the world.

7.5. *Reproducibility*

Reproducibility is the ability to replicate the results obtained in a published experiment in similar conditions as were set by the original researcher (Goodman et al., 2016). It is important to distinguish it from *replicability*, which is the ability to reproduce the results but using different data than the one proposed in the original research. Replicability is harder to attain, as the proposed hypothesis also needs to be true for no matter which dataset and not only for the one used in the published experiments. Reproducibility is therefore a minimum standard for defending a claim as scientific (Peng, 2011).

As we could see in Section 7.3, there are publications about multi-step attack detection not proposing experiments at all. Among the ones presenting at least one experiment we have analysed which ones contain reproducible experiments. We define the reproducibility of experiments through a set criteria, similar to the ones proposed by Leek and Peng (2015). We can establish three conditions to be met in order to consider a publication as reproducible in terms of experiments: accessibility of method, accessibility of data and accessibility of knowledge. We briefly explain below these three concepts:

- **Accessibility of method** (A_m): The proposed multi-step attack detection method is exposed in such a clear way that a reader can reproduce its functioning by itself. The explanation of the details of the method can be in plain English or in the form of pseudocode scripts: the important factor is the capability of reproducing how the method works. The maximum degree of accessibility is attained if the publication provides a downloadable implementation of the method (Brogi and Tong, 2016).
- **Accessibility of data** (A_d): At least a public dataset is used in experiments. It can be a well-known public dataset (see Section 7.4) or a new one that is made available by the authors of the publication. We do not consider this condition to be true if only an example of a multi-step attack is shown to prove the validity of the method.
- **Accessibility of models** (A_k): If the detection method relies on models, e.g. a schema of alerts with their prerequisites and consequences, the ones used in the experiment are provided in the publication. There are methods where models are learned from a training dataset. In that case we consider that this condition is accomplished if the training dataset is available. This condition is not applicable to

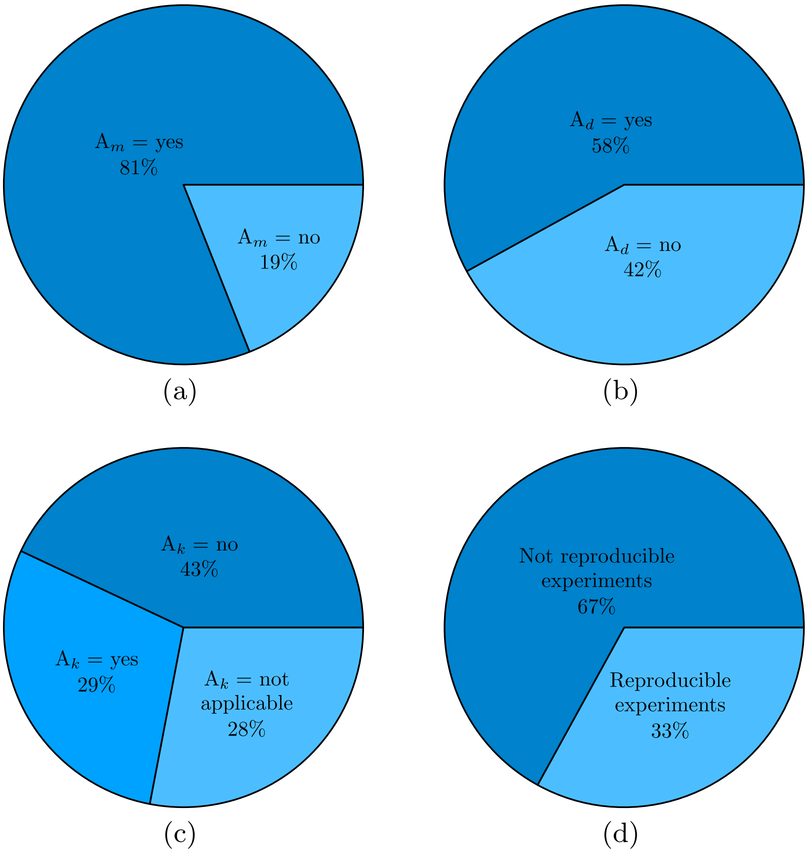

Fig. 11 – Distribution of publications by accessibility of method A_m (a), accessibility of data A_d (b), accessibility of models A_k (c) and reproducibility of experiments (d).

automatic methods, which directly learn the models from the analysed data.

Regarding reproducibility, in (Tables 4-10) we indicate if the method, the data and the models are accessible ("A_m",'A_d','A_k', respectively), if the experiments are reproducible ("Rep.") and if the model of an example attack is provided. Moreover, in Fig. 11 we can see the distribution of publications in the corpus according to their reproducibility and to the accessibility of method, data and models. Most of the publications (67%) do not offer full reproducible experiments. There is even a 19% of publications proposing multi-step attack detection methods that are not accessible by third parties. We examine the implications and possible reasons of this distribution in the next section.

8. Discussion

We have identified the big challenges in research about multi-step attack detection, after a thorough study of the published literature: modelling, automation, dataset structure and reproducibility.

In this section, we discuss all these challenges in detail. We give our vision about how multi-step attack detection should evolve in the near future.

8.1. *How to model multi-step attacks*

Among the 181 publications composing the corpus of this survey, 101 of them show the model of a multi-step attack (see the Appendix for further detail). Out of them, 37 work only with DARPA 2000, so the represented models correspond to the same attacks. Thus, we do not find many different examples of multi-step attack models in the literature about detection. There is a specific literature about multi-step attack modelling that is not the object of this survey, but it is mainly focused on modelling the attacks as possible paths in a network (Daley et al., 2002; Zhang et al., 2006) or on the development of modelling languages (Cuppens and Ortalo, 2000; Michel and Mé, 2002).

Having good models of multi-step attacks is indispensable to guide the development of detection methods. In scientific research, the definition of the studied object is a basic prerequisite. Here the studied objects are the multi-step attacks, which are abstract entities built to correspond to the single goal of an attacker. As we saw in Section 2.2, there does not exist a consensus about which are the features linking two steps. Some authors give more relevance to the IP addresses (Cipriano et al., 2011; Ourston et al., 2003) or to the event type (Fava et al., 2008; Soleimani and Ghorbani, 2012), while some others focus on the causal (Ning and Xu, 2010; Wang et al., 2006) or temporal conditions (Sadoddin and Ghorbani, 2009).

The conclusion is that the research community still does not know how to provide a solid definition of a link between the steps of an attack. Multi-step attack detection lacks from a true debate about what connects the pieces of an attack scenario. The most consistent definition presented so far is that two traces are linked if one is a consequence of the other (Cuppens and Miège, 2002a; Ning et al., 2002a). This has been thoroughly used in the literature (see Section 6.2), but always depending on the manual coding of the pre and post conditions. The link itself become a human construction. On the other hand, in methods where the link is based on common features of the individual steps, the choice of those features seems much conditioned by the immediate problem to solve in each publication.

It is possible that there cannot exist a global multi-step attack detection method considering just one type of link between steps. In that case, a set of methods should be developed, one for each type of link, i.e. each type of multi-step attack.

8.2. *From manual to automatic methods*

Most of the methods included in the corpus of this survey rely on manually coded knowledge about multi-step attacks, as we saw in Fig. 8. The aim in security detection is to develop fully automatic methods, not only to reduce the time wasted by security professionals in the investigation of potential threats but also to avoid human errors in the development of signatures. Moreover, it is the only way of detecting attacks that are still unknown to the intrusion detection community.

Some progress has been made in the development of automatic multi-step attack detection methods, particularly based on clustering (Cuppens, 2001; Julisch, 2003a) and statistical inference (Qin and Lee, 2003; Sadoddin and Ghorbani, 2009). But we are far from a reliable automatic detection. Pattern matching is a highly reliable method to detect well-known attacks and it is still most used in commercial intrusion detection systems. The enrichment with results coming from automatic methods could lead to detection of both known and unknown multi-step attacks.

Only long-term research projects can return solid and reliable security systems. However, we see that 85 out of the 119 methods studied in the corpus publish their results in just one publication, which gives an idea of the lack of continuity in multi-step attack detection research. This seems a problem affecting the whole academic world, due to difficulties linked with current funding of research (Ylijoki, 2003).

8.3. *Limitations on the type of data*

We can see in Table 1 that alerts are by far the most used traces in multi-step attack detection. As we commented in Section 7.1, it is natural to start the search of multi-step attacks as the combination of simpler single-step attacks, represented as IDS alerts. However, we cannot deny that other events not representing a security alert can provide context information that can be fundamental in the identification of an attack scenario. An IDS returns the events that are suspicious by themselves, but multi-step attacks can be composed of harmless steps too.

We consider there are some important reasons behind this preference for the study of IDS alerts over general events. First, it is clear that multi-step attacks are easier to detect if they are composed of elements that suppose a threat by themselves, so alerts are preferred. This follows the classical scientific approach of starting to solve easy problems before facing complex ones.

The second reason is related to the limitations of current networks: IDS alerts have a predefined structure and a limited number of types, which general events have not. Alerts are generated by a specific device, the IDS, which is produced and maintained by a single organisation. It does not matter if the organisation is open source or commercial: each IDS device generates a series of coherent and catalogued set of alerts. On the other side, a system dealing with general events has to manage logs coming from many different devices. Logs are usually stored as plain text and there is no standard format to represent them.

Finally, this preference for IDS alerts is also explained by the composition of public datasets: most of them are composed of IDS alerts or have to be processed by an IDS in order to be used in experiments.

We consider that multi-step attack detection should move beyond the analysis of IDS alerts. Considering other kinds of events can enrich the detection and it can improve the applicability of the methods to real networks. The development of methods able to cope with any type of event is a big challenge.

8.4. *Lack of reproducible research*

In Section 7.5, we have seen how there is a big proportion of work about multi-step attack detection methods where there are no reproducible experiments. We show in Fig. 12a distribution per year of the publications including reproducible experiments. We see that the evolution has not been very favorable: after a huge rise from 2000 to 2004, the presence of reproducible experiments dramatically dropped and it is now stabilised around 40%. This casts doubts on the scientific quality of the work in multi-step attack detection.

The reasons of these results do not necessarily rely on neglects by the authors but on limitations of the domain. A relevant factor to make an experiment reproducible is the pub-

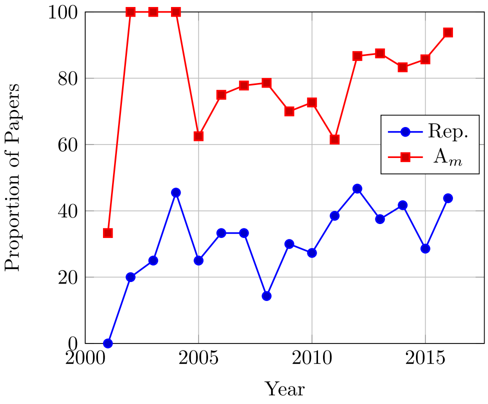

Fig. 12 – Evolution of reproducible experiments (Rep.) and accessibility of method (A_m) in publications about multi-step attack detection.

lication of the data used on it. But this cannot always be done. We have seen in Section 7.4 that public datasets are scarce. And researchers using private datasets cannot publish them for confidentiality reasons. Private institutions financing research in security can provide data to the researchers, but they do not usually want to disclose this data to preserve network confidentiality. Even considering the benefits of producing shareable data (Wicherts and Bakker, 2012), it is difficult to trust data *anonymisation* methods. Although some of them exist (Slagell and Yurcik, 2005), we are far from infallible anonymisation of research data, as it has been proved by several authors (Ji et al., 2014; Narayanan and Shmatikov, 2008). On the other side, test networks, isolated from production ones, are costly to implement.

Creating an artificial dataset for experimentation is hard, because you need to be sure that it follows the same distribution as data generated in real networks (Shiravi et al., 2012). Real data is complex, it is constantly evolving and it highly depends on particularities of the network. Finding its common places to generate a global dataset is a big challenge. We can see some weaknesses in existing public datasets. DARPA 2000, for example, does not contain enough scenarios for the parameter estimation needed in some algorithms (Dain and Cunningham, 2001a).

8.5. *Recommendations*

To end this discussion, we propose below a succinct list of recommendations for research teams about the future of multi-step attack detection. They reflect our personal view about the direction towards which the field should evolve.

- Conduct studies on the links between the steps composing multi-step attacks. A better knowledge of their characteristics can help the development of detection methods.
- Aim to develop methods working with general events. Even if it is harder than working with well-formatted alerts, challenges posed by processing general events can be a source of new questions about the pertinence of used approaches.
- Expose the methods in the clearest possible language, so they can be easily understood and reproduced by the community.
- Publish the datasets used in the experiments when possible.
- Develop new datasets containing multi-step attacks. There are research teams such as the one managed by Ghorbani (Shiravi et al., 2012) that have dedicated some effort to the generation of datasets, as a parallel activity to the creation of detection methods.

9. Conclusion

In this survey we have reviewed published multi-step attack detection methods. We have followed a systematic research method to build a corpus of 181 publications presenting 119 different methods. Methods have been classified according to the approach they follow. We have identified five approaches: similarity-based, causal correlation, structural-based, case-based and mixed. Each method has been briefly explained and placed in context with respect to the other ones. As far as we know, this is the first survey fully dedicated to multi-step attack detection methods as mechanisms to find attack scenarios through the connection of real traces.

The bibliometrics analysis conducted on the corpus of this survey shows that multi-step attack detection is an active field of research. However, further analysis has revealed certain weaknesses in the field. First, there is an important shortage of public datasets. The most used one is DARPA 2000, a dataset containing only two attack scenarios and that is 17 years old, as old as the multi-step attack detection field itself. Secondly, it is difficult to identify a global model of multi-step attacks and to characterise the link between the different steps. Finally, most of the publications do not present reproducible experiments to support the validity of the proposed method.

We have proposed a set of recommendations to improve the development of public research about multi-step attack detection in the future. There are still criminals using multi-step attacks in their network intrusions, probably more than ever. It is important to develop solid detection methods to minimise the threat posed by these criminals. Public availability of detection methods and data, together with credible scientific procedures, are the keys for the progress of multi-step attack detection research.

Acknowledgement

This work was partially supported by the French Banque Publique d'Investissement (BPI) under program FUI-AAP-19 in the frame of the HuMa project, as well as by the ICube SENSAI Project.

REFERENCES

Abreu R, Bobrow D, Eldardiry H, Feldman A, Hanley J, Honda T, et al. Diagnosing advanced persistent threats: a position paper; 2015. p. 193–200. Proceedings of the 26th International Workshop on Principles of Diagnosis (DX-2015).

Ahmadinejad SH, Jalili S. Alert correlation using correlation probability estimation and time windows. Computer Technology and Development; 2009. p. 170–5. 2009. ICCTD'09. International Conference on.

Ahmed SS. Intrusion alert analysis framework using semantic correlation. University of Victoria; 2014. Thesis.

Al-Mamory SO, Zhang HL. Scenario discovery using abstracted correlation graph. IEEE; 2007. p. 702–6. Computational intelligence and security, 2007 International Conference on.

Al-Mamory SO, Zhang HL. Multistep attacks extraction using compiler techniques. IEEE; 2008. p. 183–8. High Performance Switching and Routing, 2008. HSPR 2008. International Conference on.

Al-Mamory SO, Zhang HL. IDS alerts correlation using grammar-based approach. J Comput Virol 2009;5:271–82.

Alghamdi R. Hidden Markov Models (HMMs) and security applications. Int J Adv Comput Sci Appl 2016;7:39–47.

Alnas M, Hanashi AM, Laias EM. Detection of botnet multi-stage attack by using alert correlation model. Int J Eng Sci (IJES) 2013;2(10):24–34.

Alserhani F. A framework for correlation and aggregation of security alerts in communication networks. University of Bradford; 2012. Thesis.

Alserhani F. A framework for multi-stage attack detection. IEEE; 2013. p. 1–6. doi:10.1109/SIECPC.2013.6550973. 2013 Saudi International Electronics, Communications and Photonics Conference (SIECPC).

Alserhani F. Alert correlation and aggregation techniques for reduction of security alerts and detection of multistage attack. Int J Adv Stud Comput Sci Eng 2016;5(2):1.

Alserhani F, Akhlaq M. Event-based correlation systems to detect SQLI activities. Bioplis, Singapore: AINA; 2011. Proceedings of the International Conference on Advanced Information Networking and Applications.

Alserhani F, Akhlaq M, Awan IU, Cullen AJ, Mirchandani P. MARS: Multi-stage Attack Recognition System. IEEE; 2010. p. 753–9. doi:10.1109/AINA.2010.57. 2010 24th IEEE International Conference on Advanced Information Networking and Applications.

AmirHaeri M, Jalili R. RTEAS: a real-time algorithm for extracting attack scenarios from intrusion alert stream. ISC; 2009. 6th International ISC Conference on Information Security and Cryptology (ISCISC'09).

Anbarestani R, Akbari B, Fathi F. An iterative alert correlation method for extracting network intrusion scenarios. Electrical Engineering (ICEE); 2012. p. 684–9. 2012 20th Iranian Conference on.

Anming Z, Chunfu J. Study on the applications of Hidden Markov Models to computer intrusion detection. IEEE; 2004. p. 4352–6. Intelligent Control and Automation, 2004. WCICA 2004. Fifth World Congress on.

Bahareth FA, Bamasak OO. Constructing attack scenario using sequential pattern mining with correlated candidate sequences. The Research Bulletin of JORDAN ACM-ISWSA; 2013.

Bai H, Wang K, Hu C, Zhang G, Jing X. Boosting performance in attack intention recognition by integrating multiple techniques. Front Comput Sci China 2011;5:109–18.

Bateni M, Baraani A. An architecture for alert correlation inspired by a comprehensive model of human immune system. Int J Comput Netw Inf Secur 2014;6:47.

Bateni M, Baraani A, Ghorbani AA. Using artificial immune system and fuzzy logic for alert correlation. IJ Netw Secur 2013;15:190–204.

Bateni M, Baraani A, Ghorbani AA, Rezaei A. An AIS-inspired architecture for alert correlation. Int J Innovative Comput Inf Control 2013;9:231–55.

Benferhat S, Autrel F, Cuppens F. Enhanced correlation in an intrusion detection process. Springer; 2003. p. 157–70. MMM-ACNS.

Bereziński P, Śliwa J, Piotrowski R, Jasiul B. Detection of multistage attack in federation of systems environment. Military Communication Institute; 2012.

Bhatt P, Yano ET, Gustavsson P. Towards a framework to detect Multi-Stage Advanced Persistent Threats Attacks. IEEE; 2014. p. 390–5. doi:10.1109/SOSE.2014.53. 2014 IEEE 8th International Symposium on Service Oriented System Engineering (SOSE).

Bhatt P, Yano ET, Amorim J, Gustavsson P. A cyber security situational awareness framework to track and project multistage cyber attacks. ICCWS; 2014. p. 356–60. Proceedings of the 9th International Conference on Cyber Warfare & Security.

Brahmi H, Yahia SB. Discovering multi-stage attacks using closed multi-dimensional sequential pattern mining. Springer; 2013. p. 450–7. International Conference on Database and Expert Systems Applications.

Brogi G, Tong VVT. TerminAPTor: highlighting advanced persistent threats through information flow tracking. IEEE; 2016. p. 1–5. doi:10.1109/NTMS.2016.7792480. 2016 8th IFIP International Conference on New Technologies, Mobility and Security (NTMS).

Byers SR, Yang SJ. Real-time fusion and projection of network intrusion activity. IEEE; 2008. p. 1–8. Information Fusion, 2008 11th International Conference on.

Çamtepe SA, Yener B. Modeling and detection of complex attacks. IEEE; 2007. p. 234–43. Security and Privacy in Communications Networks and the Workshops, 2007. SecureComm 2007. Third International Conference on.

Chen B, Lee J, Wu AS. Active event correlation in Bro IDS to detect multi-stage attacks. IEEE; 2006. p. 16–50. doi:10.1109/IWIA.2006.2. Fourth IEEE International Workshop on Information Assurance (IWIA'06).

Chen C-M, Guan D-J, Huang Y-Z, Ou Y-H. Anomaly network intrusion detection using Hidden Markov Model. Int J Innovative Comput Inf Control 2016;12:569–80.

Chen G, Zhang Y, Wang C. A wireless multi-step attack pattern recognition method for WLAN. Expert Syst Appl 2014;41(16):7068–76. doi:10.1016/j.eswa.2014.05.029.

Chen P, Desmet L, Huygens C. A study on advanced persistent threats. Springer; 2014. p. 63–72. IFIP International Conference on Communications and Multimedia Security.

Cheng B-C, Liao G-T, Huang C-C, Yu M-T. A novel probabilistic matching algorithm for multi-stage attack forecasts. IEEE J Sel Areas Commun 2011;29(7):1438–48. doi:10.1109/JSAC.2011.110809.

Cheung S, Lindqvist U, Fong MW. Modeling multistep cyber attacks for scenario recognition, vol. 1. IEEE; 2003. p. 284–92. DARPA information survivability conference and exposition, 2003. Proceedings.

Chien S-H, Ho C-S. A novel threat prediction framework for network security. In: Advances in Information Technology and Industry Applications. Springer; 2012. p. 1–9.

Chien S-H, Chang E-H, Yu C-Y, Ho C-S. Attack subplan-based attack scenario correlation. In: Machine Learning and Cybernetics, vol. 4. IEEE; 2007. p. 1881–7. 2007 International Conference on.

Chintabathina S, Villacis J, Walker JJ, Gomez HR. Plan recognition in intrusion detection systems using logic programming. Homeland Security (HST); 2012. p. 609–13. 2012 IEEE Conference on Technologies for.

Cipriano C, Zand A, Houmansadr A, Kruegel C, Vigna G. Nexat: a history-based approach to predict attacker actions. ACM; 2011. p. 383–92. Proceedings of the 27th Annual Computer Security Applications Conference.

Colajanni M, Marchetti M, Manganiello F. Machine learning algorithms for clustering and correlating security alerts. Universita degli Studi di Modena e Reggio Emilia; 2010. Intrusion Detection Systems, Report.

Cui Y. A toolkit for intrusion alerts correlation based on prerequisites and consequences of attacks; 2002.

Cuppens F. Managing alerts in a multi-intrusion detection environment, vol. 1. ACSAC; 2001. p. 22.

Cuppens F, Miège A. Alert correlation in a cooperative intrusion detection framework. IEEE; 2002a. p. 202–15. Security and privacy, 2002. proceedings. 2002 ieee symposium on.

Cuppens F, Ortalo R. LAMBDA: a language to model a database for detection of attacks. Springer; 2000. p. 197–216. International Workshop on Recent Advances in Intrusion Detection.

Cuppens F, Autrel F, Miège A, Benferhat S. Correlation in an intrusion detection process. Springer; 2002b. p. 153–72. Internet Security Communication Workshop.

Cuppens F, Autrel F, Miège A, Benferhat S. Recognizing malicious intention in an intrusion detection process. HIS; 2002c. p. 806–17. In Second International Conference on Hybrid Intelligent Systems (HIS'2002).

de Vries J, Hoogstraaten H, van den Berg J, Daskapan S. Systems for detecting advanced persistent threats: a development roadmap using intelligent data analysis. IEEE; 2012. p. 54–61. doi:10.1109/CyberSecurity.2012.14. 2012 International Conference on Cyber Security (CyberSecurity).

Dain OM, Cunningham RK. Fusing a heterogeneous alert stream into scenarios, vol. 13. Citeseer; 2001a. doi:10.1007/978-1-4615-0953-0_5. Proceedings of the 2001 ACM workshop on Data Mining for Security Applications.

Dain OM, Cunningham RK. Building scenarios from a heterogeneous alert stream, vol. 235. West Point, NY, USA: 2001b. Proceedings of the 2001 IEEE workshop on Information Assurance and Security.

Daley K, Larson R, Dawkins J. A structural framework for modeling multi-stage network attacks. IEEE; 2002. p. 5–10. doi:10.1109/ICPPW.2002.1039705. Proceedings. International Conference on Parallel Processing Workshops, 2002.

Debar H, Wespi A. Aggregation and correlation of intrusion-detection alerts. In: Recent Advances in Intrusion Detection. Davis, CA, USA: 2001. p. 85–103.

Dousson C. Suivi d'évolutions et reconnaissance de chroniques. Université de Toulouse; 1994. Thesis.

Du H, Murphy CT, Bean J, Yang SJ. Toward unsupervised classification of non-uniform cyber attack tracks. IEEE; 2009. p. 1919–25. Information Fusion, 2009. FUSION'09. 12th International Conference on.

Du H, Liu DF, Holsopple J, Yang SJ. Toward ensemble characterization and projection of multistage cyber attacks. IEEE; 2010. p. 1–8. doi:10.1109/ICCCN.2010.5560087. 2010 Proceedings of 19th International Conference on Computer Communications and Networks.

Dutt V, Kaur A. Cyber security: testing the effects of attack strategy, similarity, and experience on cyber attack detection. Int J Trust Manag Comput Commun 2013;1(3–4):261–73.

E. Commission, Standard on logging and monitoring; 2010.

Ebrahimi A, Navin AHZ, Mirnia MK, Bahrbegi H, Ahrabi AAA. Automatic attack scenario discovering based on a new alert correlation method. IEEE; 2011. p. 52–8. Systems Conference (SysCon), 2011 IEEE International.

Eckmann ST, Vigna G, Kemmerer RA. STATL: an attack language for state-based intrusion detection. J Comput Secur 2002;10(1–2):71–103.

Faraji Daneshgar F, Abbaspour M. Extracting fuzzy attack patterns using an online fuzzy adaptive alert correlation framework. Secur Commun Netw 2016;9:2245–60.

Farhadi H, AmirHaeri M, Khansari M. Alert correlation and prediction using data mining and HMM. ISC Int J Inf Secur 2011;3(2).

Farhady H, Jalili R, Khansari M. Attack plan recognition using Markov model. IEEE; 2010. Proceedings of the 7th International ISC Conference on Information Security and Cryptology.

Fava DS, Holsopple J, Yang SJ, Argauer B. Terrain and behavior modeling for projecting multistage cyber attacks. IEEE; 2007. p. 1–7. doi:10.1109/ICIF.2007.4408131. 2007 10th International Conference on Information Fusion.

Fava DS, Byers SR, Yang SJ. Projecting cyberattacks through variable-length Markov models. IEEE Trans Inf Forensics Secur 2008;3(3):359–69.

Fayyad S, Meinel C. New attack scenario prediction methodology. IEEE; 2013. p. 53–9. doi:10.1109/ITNG.2013.16. 2013 Tenth International Conference on Information Technology: New Generations (ITNG).

Ficco M, Romano L. A correlation approach to intrusion detection. MOBILIGHT; 2010. p. 203–15.

Ficco M, Romano L. A generic intrusion detection and diagnoser system based on complex event processing. IEEE; 2011. p. 275–84. Data Compression, Communications and Processing (CCP), 2011 First International Conference on.

Friedberg I, Skopik F, Settanni G, Fiedler R. Combating advanced persistent threats: from network event correlation to incident detection. Comput Secur 2015;48(C):35–57. doi:10.1016/j.cose.2014.09.006.

Geib CW, Goldman RP. Plan recognition in intrusion detection systems, vol. 1. IEEE; 2001. p. 46–55. DARPA Information Survivability Conference & Exposition II, 2001. DISCEX'01. Proceedings.

Giura P, Wang W. A context-based detection framework for Advanced Persistent Threats. IEEE Computer Society; 2012a. p. 69–74. doi:10.1109/CyberSecurity.2012.16. Proceedings of the 2012 International Conference on Cyber Security.

Giura P, Wang W. Using large scale distributed computing to unveil Advanced Persistent Threats. Sci J 2012b;1(3):93–105.

Glass GV. Meta-analysis at 25; 2000. Available from http://www.gvglass.info/papers/meta25.html. Accessed September, 24, 2017.

Goodman SN, Fanelli D, Ioannidis JP. What does research reproducibility mean? Sci Transl Med 2016;8:341ps12.

Haopu Y. Method for behavior-prediction of APT attack based on dynamic Bayesian game. IEEE; 2016. p. 177–82. doi:10.1109/ICCCBDA.2016.7529554. 2016 IEEE International Conference on Cloud Computing and Big Data Analysis (ICCCBDA).

Holgado P, Villagra VA, Vazquez L. Real-time multistep attack prediction based on Hidden Markov Models. IEEE; 2018. IEEE Transactions on Dependable and Secure Computing.

Holsopple J, Yang SJ. FuSIA: future situation and impact awareness. IEEE; 2008. p. 1–8. Information Fusion, 2008 11th International Conference on.

Holsopple J, Yang SJ, Sudit M. TANDI: threat assessment of network data and information. SPIE; 2006. p. 6242. Defense and Security Symposium, International Society for Optics and Photonics.

Huang M-Y, Jasper RJ, Wicks TM. A large scale distributed intrusion detection framework based on attack strategy analysis. Computer Networks 1999;31(23):2465–75. doi:10.1016/s1389-1286(99)00114-0.

Husák M, Kašpar J, Bou-Harb E, Čeleda P. On the sequential pattern and rule mining in the analysis of cyber security alerts. Reggio Calabria: ACM; 2017. p. 22. Proceedings of the 12th International Conference on Availability, Reliability and Security.

ISO/IEC. 27000:2016 standard; 2016.

Jaeger D, Ussath M, Cheng F, Meinel C. Multi-step attack pattern detection on normalized event logs. IEEE; 2015. p. 390–8. doi:10.1109/CSCloud.2015.26. 2015 IEEE 2nd International Conference on Cyber Security and Cloud Computing (CSCloud).

Jain AK. Data clustering: 50 years beyond K-means. Pattern Recognit Lett 2010;31:651–66.

Jemili F, Zaghdoud M, Ahmed MB. Attack correlation and prediction system based on possibilistic networks. Algarve, Portugal: 2008. IADIS'08 International Conference Applied Computing.

Jemili F, Zaghdoud M, Ahmed MB. Attack prediction based on hybrid propagation in Bayesian networks. ICITST; 2009. Proc. of the Internet Technology And Secured Transactions Conference.

Ji S, Li W, Srivatsa M, Beyah R. Structural data de-anonymization: quantification, practice, and implications. ACM; 2014. p. 1040–53. Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security.

Julisch K. Mining alarm clusters to improve alarm handling efficiency. IEEE; 2001. p. 12–21. Computer Security Applications Conference, 2001. ACSAC 2001. Proceedings 17th Annual.

Julisch K. Clustering intrusion detection alarms to support root cause analysis. ACM Trans Inf Sys Secur (TISSEC) 2003a;6(4):443–71.

Julisch K. Using root cause analysis to handle intrusion detection alarms. Universität Dortmund; 2003b. Thesis.

Julisch K, Dacier M. Mining intrusion detection alarms for actionable knowledge. ACM; 2002. p. 366–75. Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining.

Kannadiga P, Zulkernine M, Haque A. E-NIPS: an event-based network intrusion prediction system. 2007. p. 37–52. Information Security.

Katipally R, Gasior W, Cui X, Yang L. Multistage attack detection system for network administrators using data mining. Oak Ridge, Tennessee, USA: ACM; 2010. p. 51. doi:10.1145/1852666.1852722. Proceedings of the Sixth Annual Workshop on Cyber Security and Information Intelligence Research.

Katipally R, Yang L, Liu A. Attacker behavior analysis in multi-stage attack detection system. Oak Ridge, Tennessee, USA: ACM; 2011. p. 63. doi:10.1145/2179298.2179369. Proceedings of the Seventh Annual Workshop on Cyber Security and Information Intelligence Research.

Kavousi F, Akbari B. Automatic learning of attack behavior patterns using Bayesian networks. Telecommunications (IST); 2012. p. 999–1004. 2012 Sixth International Symposium on.

Kavousi F, Akbari B. A Bayesian network-based approach for learning attack strategies from intrusion alerts. Secur Commun Netw 2014;7:833–53.

Kawakani CT, Junior SB, Miani RS, Cukier M, Zarpelão BB. Intrusion alert correlation to support security management. SBSI; 2016. p. 313–20. XII Brazilian Symposium on Information Systems-Information Systems in the Cloud Computing Era.

Kawakani CT, Barbon S, Miani RS, Cukier M, Zarpelão BB. Discovering attackers past behavior to generate online hyper-alerts. iSys-Revista Bras Sistemas Informação 2017;10:122–47.

Khakpour N, Jalili S. Using supervised and transductive learning techniques to extract network attack scenarios. CSI; 2009. p. 71–6. Computer Conference, 2009. CSICC 2009. 14th International.

Kholidy HA, Erradi A, Abdelwahed S. Attack prediction models for cloud intrusion detection systems. Artificial Intelligence, Modelling and Simulation (AIMS); 2014. p. 270–5. 2014 2nd International Conference on.

Kholidy HA, Erradi A, Abdelwahed S, Azab A. A finite state Hidden Markov Model for predicting multistage attacks in cloud systems. IEEE; 2014. p. 14–19. doi:10.1109/DASC.2014.12. 2014 IEEE 12th International Conference on Dependable, Autonomic and Secure Computing (DASC).

Kholidy HA, Yousof AM, Erradi A, Abdelwahed S, Ali HA. A finite context intrusion prediction model for cloud systems with a probabilistic suffix tree. IEEE; 2014. p. 526–31. Modelling Symposium (EMS), 2014 European.

Kim J, Bentley PJ, Aickelin U, Greensmith J, Tedesco G, Twycross J. Immune system approaches to intrusion detection – a review. Natural Comput 2007;6:413–66.

Kim Y-H, Park WH. A study on cyber threat prediction based on intrusion detection event for APT attack detection. Multimedia Tools Appl 2014;71(2):685–98. doi:10.1007/s11042-012-1275-x.

King ST, Mao ZM, Lucchetti DG, Chen PM. Enriching intrusion alerts through multi-host causality. NDSS; 2005.

Kitchenham B. Procedure for undertaking systematic reviews. Tech. rep. Keele University; 2004.

Kruegel C, Toth T, Kerer C. Decentralized event correlation for intrusion detection. Inf Secur Cryptology – ICISC 2001;2002:59–95.

Lagzian S, Amiri F, Enayati A, Gharaee H. Frequent item set mining-based alert correlation for extracting multi-stage attack scenarios. IEEE; 2012. p. 1010–14. doi:10.1109/ISTEL.2012.6483134. 6th International Symposium on Telecommunications (IST).

Lee D, Kim D, Jung J. Multi-stage intrusion detection system using Hidden Markov Model algorithm. IEEE; 2008. p. 72–7. doi:10.1109/ICISS.2008.22. 2008 International Conference on Information Science and Security (ICISS 2008).

Leek JT, Peng RD. Opinion: reproducible research can still be wrong: adopting a prevention approach. Proc Natl Acad Sci 2015;112:1645–6.

Li Y, Xue Y, Yao Y, Zhao X, Liu J, Zhang R. An attack pattern mining algorithm based on fuzzy logic and sequence pattern. Cloud Computing and Intelligence Systems (CCIS); 2016. p. 234–8. 2016 4th International Conference on.

Li Z, Zhang A, Li D, Wang L. Discovering novel multistage attack strategies. Harbin, China: Springer; 2007a. p. 45–56. doi:10.1007/978-3-540-73871-8_6. 3rd International Conference on Advanced Data Mining and Applications.

Li Z, Zhang A, Lei J, Wang L. Real-time correlation of network security alerts. IEEE; 2007b. p. 73–80. e-Business Engineering, 2007. ICEBE 2007. IEEE International Conference on.

Li Z, Lei J, Wang L, Li D. A data mining approach to generating network attack graph for intrusion prediction, vol. 4. IEEE; 2007c. p. 307–11. Fuzzy Systems and Knowledge Discovery, 2007. FSKD 2007. Fourth International Conference on.

Li Z, Lei J, Wang L, Li D. Assessing attack threat by the probability of following attacks. IEEE; 2007d. p. 91–100. Networking, Architecture, and Storage, 2007. NAS 2007. International Conference on.

Lin J, Liu P, Jing J. Using signaling games to model the multi-step attack-defense scenarios on confidentiality. Springer; 2012. p. 118–37. International Conference on Decision and Game Theory for Security.

Liu P, Zang W, Yu M. Incentive-based modeling and inference of attacker intent, objectives, and strategies. ACM Trans Inf Syst Secur (TISSEC) 2005;8(1):78–118.

Liu Z, Wang C, Chen S. Correlating multi-step attack and constructing attack scenarios based on attack pattern modeling. IEEE; 2008. p. 214–19. doi:10.1109/ISA.2008.11. International Conference on Information Security and Assurance, 2008. ISA 2008.

Long J, Schwartz DG. Case-oriented alert correlation. W Trans Comp 2008;7:98–112.

Luh R, Marschalek S, Kaiser M, Janicke H, Schrittwieser S. Semantics-aware detection of targeted attacks: a survey. J Comput Virol Hacking Tech 2016;13:47–85.

Luktarhan N, Jia X, Hu L, Xie N. Multi-stage attack detection algorithm based on Hidden Markov Model. Chengdu, China: Springer; 2012. p. 275–82. doi:10.1007/978-3-642-33469-6_37. International Conference on Web Information Systems and Mining.

Luo S, Wu J, Li J, Guo L. A multi-stage attack mitigation mechanism for software-defined home networks. IEEE Trans Consum Electron 2016;62(2):200–7. doi:10.1109/TCE.2016.7514720.

Luo Y, Szidarovszky F, Al-Nashif Y, Hariri S. A fictitious play-based response strategy for multistage intrusion defense systems. Secur Commun Netw 2014;7(3):473–91. doi:10.1002/sec.730.

Lv Y, Xiang S, Geng J, Li Y, Xia C. An alert correlation algorithm based on the sequence pattern mining. IEEE; 2015. p. 1146–51. Advanced Information Technology, Electronic and Automation Control Conference (IAEAC), 2015.

Lye K-W, Wing JM. Game strategies in network security. Int J Inf Secur 2005;4(1–2):71–86.

Ma J, Li Z, Li W-M. Real-time alert stream clustering and correlation for discovering attack strategies, vol. 4. IEEE; 2008. p. 379–84. Fuzzy Systems and Knowledge Discovery, 2008. FSKD'08. Fifth International Conference on.

Man D, Li X, Yang W, Wang W, Xuan S. A multi-step attack recognition and prediction method via mining attacks conversion frequencies. Int J Wireless Microw Technol 2012;2(2):20.

Manganiello F, Marchetti M, Colajanni M. Multistep attack detection and alert correlation in intrusion detection systems. Springer; 2011. p. 101–10. International Conference on Information Security and Assurance.

Marchetti M, Colajanni M, Manganiello F. Identification of correlated network intrusion alerts. Cyberspace Safety and Security (CSS); 2011a. p. 15–20. 2011 Third International Workshop on.

Marchetti M, Colajanni M, Manganiello F. Framework and models for multistep attack detection. Int J Secur Appl 2011b;5:73–90.

Mathew S, Upadhyaya S. Attack scenario recognition through heterogeneous event stream analysis. IEEE; 2009. p. 1–7. doi:10.1109/MILCOM.2009.5379763. MILCOM 2009-2009 IEEE Military Communications Conference.

Mathew S, Shah C, Upadhyaya S. An alert fusion framework for situation awareness of coordinated multistage attacks. IEEE; 2005. p. 95–104. doi:10.1109/IWIA.2005.3. Third IEEE International Workshop on Information Assurance (IWIA'05).

Mathew S, Upadhyaya S, Sudit M, Stotz A. Situation awareness of multistage cyber attacks by semantic event fusion. IEEE; 2010. p. 1286–91. doi:10.1109/MILCOM.2010.5680121. 2010-milcom 2010 Military communications conference.

Meier M. Intrusion Detection effektiv!: modellierung und Analyse von Angriffsmustern. Springer-Verlag; 2007.

Meline T, Selecting studies for systematic review: Inclusion and exclusion criteria, Contemporary issues in communication science and disorders 33.

Meta-Chart, Graphing/Charting and General Data Visualization App. 2017. Available from https://www.meta-chart.com/. Accessed October, 11, 2017.

Michel C, Mé L. ADeLe: an attack description language for knowledge-based intrusion detection. Trusted Information, Springer; 2002. p. 353–68.

Mireles JD, Cho J-H, Xu S. Extracting attack narratives from traffic datasets. Institute of Electrical and Electronics Engineers Inc; 2016. p. 1–6. Cyber Conflict (CyCon US), International Conference on.

Mohurle S, Patil M. A brief study of Wannacry threat: Ransomware attack. Int J 2017;8.

Morin B, Debar H. Correlation of intrusion symptoms: an application of chronicles. Springer; 2003. p. 94–112. International Workshop on Recent Advances in Intrusion Detection.

Morin B, Mé L, Debar H, Ducassé M. M2D2: a formal data model for IDS alert correlation. In: International Workshop on Recent Advances in Intrusion Detection. Zurich, Switzerland: 2002. p. 115–37.

Murphy CT. CACTUSS: clustering of attack tracks using significant services; 2009.

Murphy CT, Yang SJ. Clustering of multistage cyber attacks using significant services. IEEE; 2010. p. 1–7. doi:10.1109/ICIF.2010.5712046. 13th Conference on Information Fusion (FUSION), 2010.

Narayanan A, Shmatikov V. Robust de-anonymization of large sparse datasets. IEEE; 2008. p. 111–25. IEEE Symposium on Security and Privacy, 2008.

Navarro J, Deruyver A, Parrend P. Morwilog: an ACO-based system for outlining multi-step attacks. IEEE; 2016. p. 1–8. 2016 IEEE Symposium Series on Computational Intelligence (SSCI).

Ning P, Cui Y. An intrusion alert correlator based on prerequisites of intrusions. Report. North Carolina State University; 2002c.

Ning P, Xu D. Adapting query optimization techniques for efficient intrusion alert correlation, vol. 11. Springer; 2002d. Proceedings of the 17th IFIP WG.

Ning P, Xu D. Learning attack strategies from intrusion alerts. ACM; 2003. p. 200–9. Proceedings of the 10th ACM conference on Computer and communications security.

Ning P, Xu D. Hypothesizing and reasoning about attacks missed by intrusion detection systems. ACM Trans Inf Syst Secur (TISSEC) 2004;7(4):591–627.

Ning P, Xu D. Toward automated intrusion alert analysis. Springer; 2010. p. 175–205.

Ning P, Cui Y, Reeves DS. Constructing attack scenarios through correlation of intrusion alerts. ACM; 2002a. p. 245–54. Proceedings of the 9th ACM Conference on Computer and Communications Security.

Ning P, Cui Y, Reeves DS. Analyzing intensive intrusion alerts via correlation. Springer; 2002b. p. 74–94. International Workshop on Recent Advances in Intrusion Detection.

Ning P, Cui Y, Reeves DS, Xu D. Techniques and tools for analyzing intrusion alerts. ACM Trans Inf Sys Secur (TISSEC) 2004;7(2):274–318.

Ning P, Xu D, Healey CG, Amant RS. Building attack scenarios through integration of complementary alert correlation method, vol. 4. NDSS; 2004. p. 97–111.

Ning P, Peng P, Hu Y, Xu D. TIAA: a visual toolkit for intrusion alert analysis. North Carolina State University: Center for Advanced Computing and Communication; 2005.

Ning Z, Gong J. An intrusion plan recognition algorithm based on max-1-connected causal networks. Springer-Verlag; 2007. p. 809–16. International Conference on Computational Science.

Noel S, Robertson E, Jajodia S. Correlating intrusion events and building attack scenarios through attack graph distances. IEEE; 2004. p. 350–9. Computer Security Applications Conference, 2004. 20th Annual.

Ourston D, Matzner S, Stump W, Hopkins B. Applications of Hidden Markov Models to detecting multi-stage network attacks. IEEE; 2003. p. 10. doi:10.1109/HICSS.2003.1174909. Proceedings of the 36th Annual Hawaii International Conference on System Sciences, 2003.

Pandey NK, Gupta SK, Leekha S, Zhou J. ACML: capability based attack modeling language. IEEE; 2008. p. 147–54. Information Assurance and Security, 2008. ISIAS'08. Fourth International Conference on.

Panichprecha S, Mohay G, Clark A. Multi-step scenario matching based on unification. Perth, Western Australia: Edith Cowan University; 2007. Australian Digital Forensics Conference, School of Computer and Information Science.

Pei K, Gu Z, Saltaformaggio B, Ma S, Wang F, Zhang Z, et al. HERCULE: attack story reconstruction via community discovery on correlated log graph. IEEE; 2016. p. 583–95. Proceedings of the 32nd Annual Conference on Computer Security Applications.

Peng RD. Reproducible research in computational science. Science 2011;334:1226–7.

Porras PA, Neumann PG. EMERALD: event monitoring enabling response to anomalous live disturbances. SRI; 1997. p. 353–65. Proceedings of the 20th national information systems security conference.

Qiao L-B, Zhang B-F, Lai Z-Q, Su J-S. Mining of attack models in IDS alerts from network backbone by a two-stage clustering method. IEEE; 2012. p. 1263–9. Parallel and Distributed Processing Symposium Workshops & PhD Forum (IPDPSW), 2012 IEEE 26th International.

Qin X. A probabilistic-based framework for INFOSEC alert correlation. Citeseer; 2005. Thesis.

Qin X, Lee W. Statistical causality analysis of INFOSEC alert data. Springer; 2003. p. 73–93. International Workshop on Recent Advances in Intrusion Detection.

Qin X, Lee W. Attack plan recognition and prediction using causal networks. IEEE; 2004. p. 370–9. doi:10.1109/CSAC.2004.7. 20th Annual Computer Security Applications Conference.

Qin X, Lee W. Discovering novel attack strategies from INFOSEC alerts. Springer; 2007. p. 109–57. Data Warehousing and Data Mining Techniques for Cyber Security.

Ramaki AA, Rasoolzadegan A. Causal knowledge analysis for detecting and modeling multi-step attacks. Secur Commun Netw 2016;9:6042–65.

Ramaki AA, Amini M, Atani RE. RTECA: real time episode correlation algorithm for multi-step attack scenarios detection. Comput Secur 2015;49:206–19. doi:10.1016/j.cose.2014.10.006.

Ramaki AA, Khosravi-Farmad M, Bafghi AG. Real time alert correlation and prediction using Bayesian networks. Information Security and Cryptology (ISCISC); 2015. p. 98–103. 12th International Iranian Society of Cryptology Conference on, 2015.

Rass S, König S, Schauer S. Defending against advanced persistent threats using game-theory. PLoS ONE 2017;12(1):e0168675.

Ren H, Stakhanova N, Ghorbani AA. An online adaptive approach to alert correlation. Springer; 2010. p. 153–72. International Conference on Detection of Intrusions and Malware, and Vulnerability Assessment.

Roschke S, Cheng F, Meinel C. A new alert correlation algorithm based on attack graph. Springer; 2011. p. 58–67. Computational intelligence in security for information systems.

Saad S, Traore I. Extracting attack scenarios using intrusion semantics. Springer; 2012. p. 278–92. International Symposium on Foundations and Practice of Security.

Sadoddin R, Ghorbani AA. An incremental frequent structure mining framework for real-time alert correlation. Comput Secur 2009;28(3):153–73.

Salah S, Maciá-Fernández G, Díaz-Verdejo JE. A model-based survey of alert correlation techniques. Computer Networks 2013;57(5):1289–317.

Shaneck M, Chandola V, Liu H, Choi C, Simon G, Eilertson E, et al. A multi-step framework for detecting attack scenarios. Report. University of Minnesota; 2006.

Sharma PK, Moon SY, Moon D, Park JH. DFA-AD: a distributed framework architecture for the detection of advanced persistent threats. Cluster Comput 2016;20(1):1–13.

Sheyner O, Haines J, Jha S, Lippmann R, Wing JM. Automated generation and analysis of attack graphs. IEEE; 2002. p. 273–84. Security and privacy, 2002. Proceedings. 2002 IEEE Symposium on.

Shin S, Lee S, Kim H, Kim S. Advanced probabilistic approach for network intrusion forecasting and detection. Expert Syst Appl 2013;40(1):315–22.

Shiravi A, Shiravi H, Tavallaee M, Ghorbani AA. Toward developing a systematic approach to generate benchmark datasets for intrusion detection. Comput Secur 2012;31:357–74.

Shittu RO. Mining intrusion detection alert logs to minimise false positives & gain attack insight. City University London; 2016. Thesis.

Skopik F, Friedberg I, Fiedler R. Dealing with Advanced Persistent Threats in smart grid ICT networks. IEEE; 2014. p. 1–5. doi:10.1109/ISGT.2014.6816388. 2014 IEEE Innovative Smart Grid Technologies Conference (ISGT), 2014.

Slagell A, Yurcik W. Sharing computer network logs for security and privacy: a motivation for new methodologies of anonymization. In: Security and Privacy for Emerging Areas in Communication Networks. IEEE; 2005. p. 80–9. Workshop of the 1st International Conference on, 2005.

Soleimani M, Ghorbani AA. Critical episode mining in intrusion detection alerts. IEEE; 2008. p. 157–64. Communication Networks and Services Research Conference, 2008. CNSR 2008. 6th Annual.

Soleimani M, Ghorbani AA. Multi-layer episode filtering for the multi-step attack detection. Comput Commun 2012;35(11):1368–79. doi:10.1016/j.comcom.2012.04.001.

Stotz A, Sudit M. INformation Fusion Engine for Real-time Decision-making (INFERD): a perceptual system for cyber attack tracking. IEEE; 2007. p. 1–8. Information Fusion, 2007 10th International Conference on.

Strayer WT, Jones CE, Schwartz BI, Mikkelson J, Livadas C. Architecture for multi-stage network attack traceback. IEEE; 2005. p. 8–785. doi:10.1109/LCN.2005.33. The IEEE Conference on Local Computer Networks 30th Anniversary (LCN'05).

Sudit M, Stotz A, Holender M. Situational awareness of a coordinated cyber attack. SPIE; 2005. p. 114–29. Defense and Security, International Society for Optics and Photonics.

Templeton SJ, Levitt K. A requires/provides model for computer attacks. ACM; 2001. p. 31–8. Proceedings of the 2000 workshop on New security paradigms.

U.S. Barbara, The UC Santa Barbara iCTF Competition. 2015. Available from https://ictf.cs.ucsb.edu/. Accessed September 24, 2017.

U.S.M. Academy, Data capture from National Security Agency (NSA). 2009. Available from http://www.usma.edu/crc/sitepages/datasets.aspx. Accessed September 24, 2017.

Ussath M, Cheng F, Meinel C. Automatic multi-step signature derivation from taint graphs. IEEE; 2016a. p. 1–8. Computational Intelligence (SSCI), 2016 IEEE Symposium Series on.

Ussath M, Cheng F, Meinel C. Event attribute tainting: a new approach for attack tracing and event correlation. IEEE/IFIP; 2016b. p. 509–15. Network Operations and Management Symposium (NOMS), 2016.

Valdes A, Skinner K. Probabilistic alert correlation. In: Recent advances in intrusion detection. Springer; 2001. p. 54–68.

Valeur F, Vigna G, Kruegel C, Kemmerer RA. Comprehensive approach to intrusion detection alert correlation. IEEE Trans Dependable Secure Comput 2004;1(3):146–69.

Vasilomanolakis E, Srinivasa S, García Cordero C, Mühlhäuser M. Multi-stage attack detection and signature generation with ICS honeypots. NOMS; 2016. p. 1227–32. doi:10.1109/NOMS.2016.7502992. 2016 - 2016 IEEE/IFIP Network Operations and Management Symposium.

Vigna G, Kemmerer RA. NetSTAT: a network-based intrusion detection approach. IEEE; 1998. p. 25–34. Computer Security Applications Conference, 1998. Proceedings. 14th Annual.

Vogel M, Schmerl S. Efficient distributed intrusion detection applying multi step signatures, vol. 17. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik; 2011. p. 188–93. OASIcs-OpenAccess Series in Informatics.

Vogel M, Schmerl S, König H. Efficient distributed signature analysis. Springer; 2011. p. 13–25. IFIP International Conference on Autonomous Infrastructure, Management and Security.

Wang C-H, Chiou Y-C. Alert correlation system with automatic extraction of attack strategies by using dynamic feature weights. Int J Comput Commun Eng 2016;5:1.

Wang J, Wang H, Zhao G. A GA-based solution to an NP-hard problem of clustering security events, vol. 3. 2006. p. 2093–7. Communications, Circuits and Systems Proceedings, 2006 International Conference on.

Wang L, Jajodia S, An approach to preventing, correlating, and predicting multi-step network attacks, Intrusion Detection Systems 93.

Wang L, Liu A, Jajodia S. An efficient and unified approach to correlating, hypothesizing, and predicting intrusion alerts. Springer; 2005. p. 247–66. European Symposium on Research in Computer Security.

Wang L, Li Z, Fan J. Learning attack strategies through attack sequence mining method. 2006. p. 1–4. Communication Technology, 2006. ICCT'06. International Conference on, IEEE.

Wang L, Li Z, Wang Q-H. A novel technique of recognizing multi-stage attack behaviour. IEEE; 2006. p. 188–93. doi:10.1109/IWNAS.2006.11. 2006 International Workshop on Networking, Architecture, and Storages (IWNAS'06).

Wang L, Liu A, Jajodia S. Using attack graphs for correlating, hypothesizing, and predicting intrusion alerts. Comput Commun 2006;29(15):2917–33.

Wang L, Li Z, Lei J, Li Y. A novel algorithm SF for mining attack scenarios model. IEEE; 2006. p. 55–61. e-Business Engineering, 2006. ICEBE'06. IEEE International Conference on.

Wang L, Li Z, Lei J. Learning attack strategies through mining and correlation of security alarms. IEEE; 2007. p. 713–16. Integrated Network Management, 2007. IM'07. 10th IFIP/IEEE International Symposium on.

Wang L, Li Z, Li D, Lei J. Attack scenario construction with a new sequential mining technique, vol. 1. IEEE; 2007. p. 872–7. Software Engineering, Artificial Intelligence, Networking, and Parallel/Distributed Computing, 2007. SNPD 2007. Eighth ACIS International Conference on.

Wang L, Ghorbani AA, Li Y. Automatic multi-step attack pattern discovering. Int J Netw Secur 2010;10(2):142–52.

Wang L-M, Ma J-F. Two-stage algorithm for correlating the intrusion alerts. Wuhan Unive J Nat Sci 2005;10:89–92.

Wang L-M, Ma J-F, Zhan Y-Z. Enhancing the content of the intrusion alerts using logic correlation. In: Content Computing. Springer; 2004. p. 137–42.

Wasserman L. All of statistics: a concise course in statistical inference. Springer Science & Business Media; 2013.

Wicherts JM, Bakker M. Publish (your data) or (let the data) perish! Why not publish your data too? Intelligence 2012;40:73–6. doi:10.1016/j.intell.2012.01.004.

Xian M, Zhang Y. A privacy-preserving multi-step attack correlation algorithm. IEEE; 2016. p. 1389–93. Advanced Information Management, Communicates, Electronic and Automation Control Conference (IMCEC), 2016.

Xiao Y, Han C. Correlating intrusion alerts into attack scenarios based on improved evolving self-organizing maps. Int J Comp Sci Netw Secur 2006;6:199–203.

Xu D, Ning P. Alert correlation through triggering events and common resources. IEEE; 2004. p. 360–9. Computer Security Applications Conference, 2004. 20th Annual.

Xu D, Ning P. Correlation analysis of intrusion alerts, Thesis. North Carolina State University; 2006.

Xu D, Ning P. Correlation analysis of intrusion alerts. Intrusion Detection Syst 2008;38:65–92.

Xuewei F, Dongxia W, Jiemei Z, Guoqing M, Jin L. Analyzing and correlating security events using state machine. Computer and Information Technology (CIT); 2010. p. 2849–54. IEEE 10th International Conference on, 2010.

Xuewei F, Dongxia W, Minhuan H, Xiaoxia S. An approach of discovering causal knowledge for alert correlating based on data mining. Dependable, Autonomic and Secure Computing (DASC); 2014. p. 57–62. 2014 IEEE 12th International Conference on.

Xupeng F, Lidong Z, Zhaopeng J, Wenyan B. A game model for predicting the attack path of APT. IEEE Computer Society; 2014. p. 491–5. doi:10.1109/dasc.2014.94. Proceedings of the 2014 IEEE 12th International Conference on Dependable, Autonomic and Secure Computing.

Yan W. Network attack scenarios extraction and categorization by mining IDS alert streams. J. UCS 2005;11:1367–82.

Yan W, Liu F. Semantic scheme to extract attack strategies for Web service network security. In: IP Operations and Management. IEEE; 2004. p. 104–11. Proceedings IEEE Workshop on, 2004.

Yang SJ, Holsopple J, Sudit M. Evaluating threat assessment for multi-stage cyber attacks. Washington, D.C.: IEEE; 2006. p. 1–7. doi:10.1109/MILCOM.2006.302216. MILCOM 2006-2006 IEEE Military Communications conference.

Yang SJ, Byers SR, Holsopple J, Argauer B, Fava DS. Intrusion activity projection for cyber situational awareness. IEEE; 2008. p. 167–72. Intelligence and Security Informatics, 2008. ISI 2008. IEEE International Conference on.

Yang SJ, Stotz A, Holsopple J, Sudit M, Kuhl ME. High level information fusion for tracking and projection of multistage cyber attacks. Inf Fusion 2009;10(1):107–21. doi:10.1016/j.inffus.2007.06.002.

Ylijoki O-H. Entangled in academic capitalism? A case-study on changing ideals and practices of university research. High Educ 2003;45(3):307–35.

Yu D, Frincke D. A novel framework for alert correlation and understanding. Springer; 2004. p. 452–66. International Conference on Applied Cryptography and Network Security.

Yu D, Frincke D. Improving the quality of alerts and predicting intruder's next goal with Hidden Colored Petri-Net. Comput Netw 2007;51(3):632–54.

Zali Z, Hashemi MR, Saidi H. Real-time attack scenario detection via intrusion detection alert correlation. IEEE; 2012a. p. 95–102. doi:10.1109/ISCISC.2012.6408197. 2012 9th International ISC Conference on Information Security and Cryptology (ISCISC).

Zali Z, Hashemi MR, Saidi H. Real-time intrusion detection alert correlation and attack scenario extraction based on the prerequisite-consequence approach. ISC Int J Inf Secur 2012b;4(2).

Zargar ST. ONTIDS: a highly flexible context-aware and ontology-based alert correlation framework, vol. 8352. La Rochelle, France: 2013. p. 161. Foundations and Practice of Security: 6th International Symposium, FPS 2013, Revised Selected Papers.

Zhai Y, Ning P, Xu J. Integrating IDS alert correlation and OS-level dependency tracking. Springer; 2006. p. 272–84. International Conference on Intelligence and Security Informatics.

Zhang A, Li Z, Li D, Wang L. Discovering novel multistage attack patterns in alert streams. IEEE; 2007. p. 115–21. doi:10.1109/NAS.2007.20. 2007 International Conference on Networking, Architecture, and Storage (NAS 2007).

Zhang S, Li J, Chen X, Fan L. Building network attack graph for alert causal correlation. Comput Secur 2008;27(5):188–96.

Zhang Y, Liu T, Shi J, Zhang P, Zhang H, Ya J. An automatic multi-step attack pattern mining approach for massive WAF alert data. Scanning 2015;4514:5.

Zhang Y, Luo X, Luo H. A multi-step attack-correlation method with privacy protection. J Commun Inf Netw 2016;1:133–42.

Zhang Z, Ho P-H, Lin X, Shen H. Janus: a two-sided analytical model for multi-stage coordinated attacks. Busan, Korea: Springer; 2006. p. 136–54. doi:10.1007/11927587_13. 9th International Conference on Information Security and Cryptology.

Zhou CV, Leckie C, Karunasekera S. A survey of coordinated attacks and collaborative intrusion detection. Comput Secur 2010;29:124–40.

Zhou J, Heckman M, Reynolds B, Carlson A, Bishop M. Modeling network intrusion detection alerts for correlation. ACM Tran Inf Syst Secur (TISSEC) 2007;10(1):4.

Zhu B, Ghorbani AA. Alert correlation for extracting attack strategies. IJ Netw Secur 2006;3(3):244–58.

Julio Navarro is PhD student at the CSTB (Complex Systems and Translational Biology) team from the ICube Laboratory at the University of Strasbourg. His research topic is the development of multi-step attack detection methods from network traces. He obtained his MS in Telecommunication Engineering from the University of Granada in 2012 with the highest results. He spent his last year abroad at UCLA. He got the Spanish National Degree Award to the second most outstanding graduate in Engineering. Before starting his PhD, he spent three years in Madrid working as a Security Project Engineer.

Aline Deruyver is tenured senior Associate Professor in computer science at the University of Strasbourg. She is a member of the CSTB (Complex Systems and Translational Bioinformatic) team from the ICube Laboratory of Strasbourg, France. She obtained a PhD in Computer Science from the University of Lille I in 1991 and she obtained an accreditation to supervise research from the University of Auvergne in 1999. She is a member of the Technical Committee TC15 of IAPR: representation of images by graphs and a proofreader in the workshop of the technical committee “Graph Based Representation For Pattern Recognition”. She is an IEEE member.

Dr. Pierre Parrend is the head of the Computer Science and Mathematics department at ECAM Strasbourg-Europe engineer school. He is a member of the CSTB (Complex Systems and Translational Biology) research team at ICube Laboratory of the University of Strasbourg, and the head of the e-laboratory “4PFactory: the factory of the future” of the UNESCO Unitwin Complex System-Digital Campus. His research interests encompass attack detection, software security, artificial immune ecosystems, evolutionary strategies for optimisation and anomaly detection, as well as emergent properties of human organisations.