Automatic incident response solutions: a review of proposed solutions' input and output

Henrik Karlzén
Swedish Defence Research Agency (FOI), Sweden
henrik.karlzen@foi.se

Teodor Sommestad
Swedish Defence Research Agency (FOI), Sweden
teodor.sommestad@foi.se

ABSTRACT

Many organizations are exposed to the risk of cyber attacks that penetrate their computer networks. When such cyber attacks occur, e.g. a ransomware outbreak, it is desirable to quickly respond by containing the threat or limit its consequences. Technologies that support this process have been widely used for decades, including antivirus software and deep-packet inspection firewalls. A large number of researches on cyber security have been initiated to automate the incident handling process further, often motivated by the need to respond to more advanced cyber attacks or the increasing cyber risks at stake. This paper reviews the research on automatic incident response solutions published since the year 2000, in order to identify gaps as well as guide further research. The proposed solutions are categorized in terms of the input they use (e.g. intrusion signals) and the output they perform (e.g. reconfiguring a network) using the D3FEND framework. The solutions presented in 45 papers published in the academic literature are analyzed and compared to four commercially available solutions for automatic response. Many of the 45 papers described input and output in vague terms. The most common inputs were from asset inventories, platform monitoring and network traffic analysis. The most common output was network isolation measures, e.g. to reconfigure firewalls. Commercially available solutions focus more on looking for identifiers in reputation systems and individual analyzing files.

CCS CONCEPTS

• **Security and privacy** → Intrusion/anomaly detection and malware mitigation; Intrusion detection systems.

KEYWORDS

Intrusion response, review, D3FEND

ACM Reference Format:
Henrik Karlzén and Teodor Sommestad. 2023. Automatic incident response solutions: a review of proposed solutions' input and output. In *The 18th International Conference on Availability, Reliability and Security (ARES 2023), August 29–September 01, 2023, Benevento, Italy*. ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3600160.3605066

1 INTRODUCTION

Most organizations are exposed to the risk of being targeted in cyber attacks and many organizations invest resources to be able to detect and respond to these attacks. Tools that offer support and provide system administrators and incident handlers with situational awareness receive a considerable portion of these resources. These tools enable faster and more efficient responses to be executed. They also enable automation of incident response processes. This paper focuses on such automated incident response.

To automate incident response processes is, for many reasons, beneficial and desirable. Automation can reduce the need of expensive and scarce human resources as well as lead to better, quicker and more reliable responses to threats. Accordingly, incident response technologies that automate parts of the incident response process are already present in many organizations. For example, antivirus software, email filters, and advanced firewalls systems automate things previously requiring human resources. Antivirus software use signatures and heuristics to block suspicious files or processes that are present on endpoints in the network; email filters try to recognize phishing emails and drop or block them; advanced firewalls may inspect traffic in the application layer and block or drop traffic if it is associated with threats. However, despite their utility, the incident response technologies used today are unable to respond to all threats organizations face.

First, some threats may be difficult to identify unless logs from multiple machines and network devices are correlated and combined. For instance, an unusual email may need to be correlated with unusual host events and suspicious network traffic to be classified as phishing with enough certainty. Centralized log management systems, often referred to as security information event management (SIEM) systems, enable the correlation of different events in a computer network. A considerable body of research is available on log analysis and log correlation of relevance to SIEM systems [19] [15]. For instance, many solutions correlate alerts using models of the steps attackers may take [15].

Second, some responses may require coordination of different devices and security components in a network. For instance, a phishing email may warrant both blocking of the sender in the local email server and blocking malicious links in the network firewall. Means of executing pre-prepared "playbooks" that activate defenses, addresses this problem. These executions are often realized by security orchestration, automation and response (SOAR) systems. The scholarly literature is not as extensive and deep when it comes to the construction of playbooks that respond to networks attacks. However, best practices for incident handling have been established [18][6], and various ways of specifying responses have been proposed [24]. For example, the language OpenC2 defines actions (e.g. "deny") and targets (e.g. "domain_name") that are relevant [16].

SIEM systems and SOAR systems form a potential foundation of an automated or semi-automated cyber defense system capable

ARES 2023, August 29–September 01, 2023, Benevento, Italy

ACM ISBN 979-8-4007-0772-8/23/08.
https://doi.org/10.1145/3600160.3605066

Figure 1: Scope of the review

of going beyond contemporary automated defenses. More specifically, they may be leveraged to correlate multiple weak intrusion signals, and deploy an appropriate response that involves multiple defenses. Consequently, several large initiatives have been taken along this direction, e.g. the European Defence Fund call "Cyber threat intelligence and improved cyber operational capabilities", the UK initiative for "Autonomous Resilient Cyber Defence" and the DARPA call for "Cyber Agents for Security Testing and Learning".

This paper aims to summarize solutions on automatic and semi-automatic incident response as described in the scholarly literature since the year 2000, and to analyse these solutions in relation to the playbooks of available incident response products. As illustrated in Figure 1, a black box perspective is employed, and the internal logic of the solutions is not addressed as long as it is (semi-) automatic. Instead, the solutions are classified based on what input data they use, what responses they choose from, and how these two relate to actions of available products. More specifically, the paper answers the following research questions:

- Which input data (signals, system information, intelligence, etc.) are used in proposed solutions for automated incident response?
- Which actions (reconfigurations, probes, denylists, etc.) are used in proposed solutions for automated incident response?
- How does the input and output covered in academic papers compare with the playbooks of available products?

The answers to these questions can help identify gaps in the current research, e.g. whether research has focused on only some inputs and outputs, and whether other issues are of relevance in practice. Assuming that the published research covers relevant incident handling procedures, the answer to the first question should inform SIEM developers on what data they should include in their solutions, while the second question should inform SOAR developers on what actions their solutions should be capable of executing. In addition, while the answers to these two questions scope the problem of automated responses through the lens of academic research, the answer to the third question ought to indicate the gap between the research and current practice. Ultimately, the answers to the three research questions can assist in developing better automatic and semi-automatic incident response solutions that can help organizations deal with cyber attacks.

A number of literature reviews have covered topics related to automated responses. This includes reviewing research on recommendation systems for cybersecurity incident handling and response [8], machine learning in relation to SOAR systems [13], and the relationship between threat intelligence and intrusion responses [7]. Papers that address the input and output of automatic response solutions are also available. In particular, [1] reviewed and categorized commercial intrusion response solutions and assessed which input and output (i.e. responses) they used. Both input and output were assessed on a high level of abstraction. Input was categorized as network input, host input, or input from a SIEM system. Responses were classified as proactive (three types), reactive (four types), and passive (six types). The review in [1] answers the first two research questions of our paper, but is at a higher level of abstraction, and without papers published since 2010. Another more recent review is presented in [2]. This review focuses mostly on different types of input and detection schemas. It provides a list of intrusion responses that are said to be common, with 18 types of responses (e.g. "enable remote logging" and "lock user account"). It is unclear how it was determined that these are common. A third, even more recent, review is presented in [3]. The review focuses on cyber physical systems and summarizes things such as information sources (e.g. "IDS alerts") and countermeasures (e.g. "blocking IP address") of eight published solutions. Compared to [3], the review in the present paper is more comprehensive and use a more systematic way of summarizing the published proposals.

The remainder of this paper is outlined as follows. Section 2 describes the details of the method of the literature review, e.g. the search strings and inclusion criteria. Section 3 presents the results and answers the research questions. Section 4 discusses these results and Section 5 draws conclusions.

2 REVIEW METHOD

This chapter describes the method used to identify and analyze the research. As recommended in [14][20] for systematic literature reviews, the following is covered: how research was identified (Section 2.1), how inclusion criteria were applied (Section 2.2), how data were extracted and mapped (Section 2.3), and how the extracted data were synthesized (Section 2.4). Additionally, the scholarly data were compared to the functionality of available products (Section 2.5).

2.1 Study identification

Initially, manual searches were conducted to identify papers with clear relevance to the topic. Keywords such as SOAR and "automatic incident handling" were used. Upon review of the resulting papers and papers cited in these papers, the following search string was identified as suitable:

```
( "intrusion response"  OR  "cyber response"
   OR  "incident response"  OR  "incident handling"
```

```
OR "security orchestration, automation and response" )
 AND ( auto* )  AND ( cyber*  OR  network ).
```

The first clause of this search string includes research on the response processes of interest, the second clause limits the scope to papers that cover automation in some sense, and the last clause limits the search to research in the particular domain of interest. The search string was used on title, abstracts, and keywords in the database Scopus on March 9 2023. This yielded 370 records. The relevance of this search string was validated by comparing the resulting records to papers manually identified beforehand. No paper identified as within scope through manual search was left out with this string.

2.2 Inclusion criteria

A number of criteria were used to filter the 370 records. Seven papers were excluded because they were written in other languages than English, 58 records were excluded because they were not scientific publications (e.g. newspaper articles), and 10 records were excluded because they were not associated with any of the subjects "computer science", "engineering", "mathematics", or "decision sciences". The remaining 295 records were reviewed manually. These records were first reviewed based on title, abstracts, and keywords, in order to remove clearly irrelevant or duplicated records. After this 122 records remained. The full text of these 122 was sought, and 107 full text records could be retrieved. The 107 full text records were reviewed and it was again assessed if they covered the right topic and presented an automated solution. After this review, it was concluded that 20 records did not present an automated solution, and one record presented a recommender system. The full text records were also assessed on a scale 0-27 based on how well they described the proposed solution requirements concerning:

- Input in terms of information about the protected system (0-9).
- Input in terms of events or alerts from detection systems (0-9).
- Output in terms of defensive actions or responses (0-9).

Two measures for input were used because most papers focused on a limited part of the problem, and these two types of input cover differ parts of the problem and received different amount of attention in the papers. Few papers covered input and output thoroughly, and only seven papers scored more than three on the scale on each of the three parts. A total score of six (of maximum 27) was set as a threshold for inclusion. This resulted in 45 papers to be included in the review. Remaining papers, not meeting this criterion, were superficial and presented early ideas or focused solely on part of the problem (e.g. describing potential responses without describing input). The 45 papers were assessed as being of reasonable quality, and as a whole constituting a reasonable amount of papers from which to extract data. Hence, no further exclusion was performed based on e.g. the number of citations. However, analysis based on citation count is presented in the discussion (Section 4.1).

2.3 Extraction

In order to compare the solutions there is a need for a categorization of the inputs and outputs. A number of alternatives were considered for these categorizations. For input, categorizations used in other research (e.g. the ones in [1], [2] and [24]) and well-known data schemas (e.g. Splunk's Common information model and STIX) were considered. For output, schemas proposed for SOAR systems were considered (e.g. OpenC2 [16] and CACAO [4]). However, none of these was found to be ideal for classifying input or output. For instance, Splunk's model is too detailed, while OpenC2 lacks common actions (e.g. the disabling of user accounts). This review found MITRE's D3FEND framework [11], a catalog of defensive cybersecurity techniques, to be suitable for classifying the solutions, and used version 0.12.0-BETA-2 of D3FEND. As shown in Table 1, D3FEND divides countermeasures in six tactics: model, harden, detect, isolate, deceive, and evict. Beneath these six tactics, there are 22 techniques (e.g. "Network Traffic Analysis), with more than 100 sub-techniques (e.g. "DNS Traffic Analysis") in two layers. To illustrate this, Figure 2 outlines the sub-techniques of Network Isolation.

The extraction was performed by reading the included papers and classifying any mentioned input and output using D3FEND. In many cases, this was trivial. For instance, outputs that block external hosts using firewalls were mapped to *Isolate-Network Isolation-Network Traffic Filtering-Inbound Traffic Filtering* (tactic – technique – sub-technique – sub-sub-technique). In other cases, some interpretation was needed. For example, when the solution in [12] responds to brute force guessing on SSH, this was classified as *Detect-Network Traffic Analysis-Administrative Network Activity Analysis* based on the context given in the paper. Another example is when the solution responded to unspecified output from Snort, as in [10], where the detection was classified only on the highest technique-level (*Detect-Network Traffic Analysis*) because of the lack of detail. A note was made when the classification required some interpretation. A note was also made when the input or output of the solution fell outside the scope of D3FEND. For example, D3FEND does not have a response switching over to redundant service, which is discussed in [22]. For a full list of the extracted D3FEND techniques from each paper, see Appendix.

2.4 Synthesis

The extraction form based on the D3FEND techniques allowed for a straightforward analysis of the papers. The proportion of papers using the different techniques for inputs and outputs were calculated to answer the first two research questions. Because of the high abstraction level in the papers, each paper was associated with at most one use of a particular technique, and the highest level of abstraction in D3FEND was used for comparison. For instance, in some cases different variants of Network Isolation were used in a solution. These multiple variants were only counted once (as Network Isolation).

2.5 Comparison to available products

Research question number three concerned the inputs and outputs covered in academic papers in comparison to the actions of available products. To answer this, publicly available playbooks of SOAR

Table 1: Tactics and the topmost technique level of D3FEND.

Model	Harden	Detect	Isolate	Deceive	Evict
Asset Inventory	Application Hardening	File Analysis	Execution Isolation	Decoy Environment	Credential Eviction
Network Mapping	Credential Hardening	Identifier Analysis	Network Isolation	Decoy Object	File Eviction
Operational Activity Mapping	Message Hardening	Message Analysis			Process Eviction
System Mapping	Platform Hardening	Network Traffic Analysis Platform Monitoring Process Analysis User Behavior Analysis			

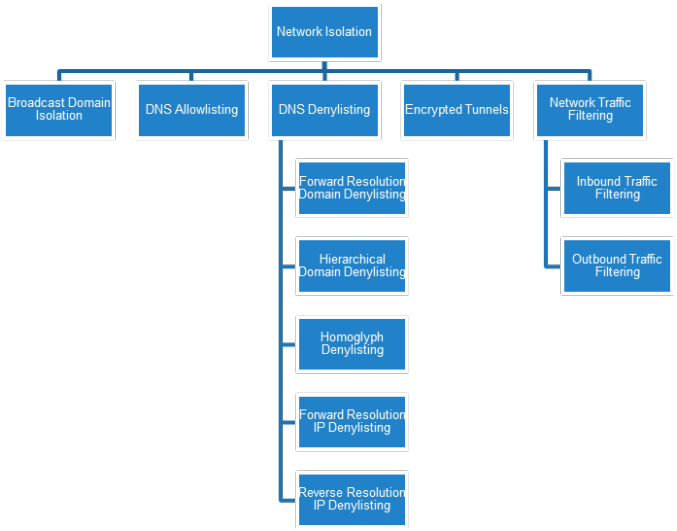

Figure 2: Sub-techniques of the Network Isolation in D3FEND.

systems were retrieved. The default playbooks of the following products were downloaded and processed: InsightConnect from Rapid7 [9], Microsoft Sentinel [17], ThreatConnect's playbooks [26] and Splunk's SOAR Community Playbooks repository [21].

These four were chosen because they are readily available online, and are part of a marketed security product. The descriptions of these playbooks were used to classify them and map them to techniques in the D3FEND framework. As the playbooks, by design, are atomic and related to tangible products or actions, this process was more straightforward than the process of mapping models in the academic papers. However, some playbooks concerned the inner workings of the SOAR system or communications with humans (e.g. reporting an incident). These were not mapped. The relative frequency of different techniques within the repositories was first calculated. The mean of these proportions was used as a proxy for what practitioners rely on. In addition, the number of products or repositories that covered each technique was counted.

3 INPUT AND OUTPUT OF THE SOLUTIONS

The sections below answer the three research questions. Section 3.1 describes the input of automatic response solutions in the literature, Section 3.2 describes the output of automatic response solutions in the literature, and Section 3.3 compares the solutions in the literature to commercially available products.

3.1 Input to the automatic response solutions

The input used in the models were often vaguely described, e.g. by referring to intrusion detection systems in general and the alerts they produce. The techniques used as input in the papers are listed in Table 2 together with the proportion of papers mentioning them as input.

As the table illustrates, the inputs used vary between papers. The most common inputs are Asset Inventory, followed by Platform Monitoring (e.g. host events) and Network Traffic Analysis (e.g. Snort alerts). Each instance of a technique in the table can correspond to the use of one or more sub-techniques. It should be noted that when several sub-techniques are used in one paper, this counts

Table 2: D3FEND techniques used as input and the proportion of papers using them.

	Technique	Proportion
Model	Asset Inventory	31%
Detect	Network Traffic Analysis	20%
Model	Network Mapping	16%
Model	Operational Activity Mapping	16%
Detect	Platform Monitoring	11%
Model	System Mapping	9%
Detect	File Analysis	2%
Detect	Identifier Analysis	2%
Detect	Message Analysis	2%
Detect	Process Analysis	2%
Detect	User Behavior Analysis	2%

Table 3: D3FEND techniques used as output and the proportion of papers using them.

Tactic	Technique	Proportion
Isolate	Network Isolation	67%
Evict	Process Eviction	33%
Harden	Platform Hardening	18%
Deceive	Decoy Environment	16%
Evict	Credential Eviction	13%
Harden	Credential Hardening	9%
Isolate	Execution Isolation	4%
Deceive	Decoy Object	2%
Harden	Application Hardening	2%
Evict	File Eviction	0%
Harden	Message Hardening	0%

only once towards the topmost technique. Hence, percentages of sub-techniques cannot simply be added together to achieve the technique percentage. The most frequently used sub-techniques of Asset Inventory (31%) are Network Node Inventory (22%) and Software Inventory (9%). The Network Mapping is in many cases Network Traffic Policy Mapping (11%), e.g. firewall configurations. The Platform Monitoring (11%) and Network Traffic Analysis (20%) were often vaguely described, and it was in many cases difficult to identify any particular sub-techniques required. In some papers, it was clarified that Operating System Monitoring (7%) was the main sub-technique of Platform Monitoring. In many cases, the topmost technique level was used. For example, deep packet inspection was related in this review to Network Traffic Analysis (13%), without guessing the particular techniques a deep inspection solution uses. Ambiguity was present also for techniques associated with modeling. For example, two papers (4%) required a System Mapping, but was unclear about what information the mapping should contain.

A few types of input fell outside the scope of D3FEND. Worth mentioning, are [23] and [5], who use the state of an industrial process as input, and a few papers explicitly required an inventory of services running on machines.

3.2 Output from the automatic response solutions

The extraction of D3FEND techniques related to responses to techniques on the highest level of abstraction, except for Message Hardening. The techniques are listed in Table 3, together with the proportion of papers mentioning them as output.

As the table shows, a few techniques dominate. As for inputs, each instance of a technique in the table can correspond to the use of one or more sub-techniques. Network Isolation (67% of papers) was in most cases realized through sub-techniques related to Network Traffic Filtering (62%), in one paper through DNS Denylisting (2%), and in other cases only through unspecified techniques that were associated with the highest abstraction level. Process Eviction (33%) was always done through Process Termination (33%), and sometimes simultaneously through Process Suspension (2%). Platform Hardening (18%) was done through Software Update (18%). Credential Eviction (13%) were often performed through Account Locking (9%). Apart from the tactics Isolate, Evict and Harden, the solutions also elaborate on the Deceive tactic, as follows. Decoy Environment (16%) was more common than Deocy Objects (2%). In many cases, it was difficult to classify the decoy environment, but all of Integrated Honeynet (4%), a Connected Honeynet (2%), and a Decoy Network Resource (2%) were used.

Mapping responses to D3FEND techniques was sometimes difficult, and many solutions proposed techniques not covered by D3FEND. These non-mapped techniques include: shut down a machine (24% of papers), reboot a machine (13%), re-prioritize detection processes (9%), restart processes or services (7%), change network addresses (7%), restore to backup (4%), change priorities for system calls or similar (2%), change an industrial process (2%), reset a network connection (2%), restore a file to backup (2%), roll over to a backup system (2%), take a backup of data (2%), reconfigure or change services (2%), and warn attackers (2%).

3.3 Comparison to available commercial playbooks

A total of 609 playbooks were analyzed and mapped to D3FEND: 211 from Rapid7, 222 from Azure, 40 from ThreatConnect, and 136 from Splunk's Community Playbooks. The ten most frequently used techniques in D3FEND are listed in Table 4. The table details the relative frequency of different techniques within the repositories, the mean of these values, the number of repositories covering the techniques, and the frequency of the technique in published papers.

When it comes to input, Network Mapping, Operational Activity Mapping, System Mapping, and Platform Monitoring are common in papers but seldom or never dealt with in playbooks, while Process Analysis is rare in both. Additionally, Asset Inventory is somewhat common in playbooks but less so than in papers. The same is true for Network Traffic Analysis, while playbooks detect more often using Identifier Analysis (e.g. lookup IP addresses against reputation systems) and File Analysis (e.g. check email attachments) than papers do.

When it comes to output, Credential Hardening, Platform Hardening, Decoy Environment and Process Eviction are common in papers but seldom or never dealt with in playbooks, while Application Hardening, Message Hardening and Decoy Object are rare

Table 4: The use of techniques in commercial tools.

Tactic	Technique	Proportion of playbooks in repositories using technique					Proportion in commercial repositories	Proportion in published papers
		Rapid7 (N=211)	Azure (N=222)	ThreatC. (N=40)	Splunk (N=136)	Mean (k=4)		
Detect	Identifier Analysis	15%	5%	15%	20%	14%	4/4	2%
Isolate	Network Isolation	24%	12%	3%	14%	13%	4/4	67%
Model	Asset Inventory	19%	3%	0%	10%	8%	3/4	31%
Detect	File Analysis	5%	2%	5%	10%	5%	4/4	2%
Evict	Credential Eviction	8%	3%	0%	1%	3%	3/4	13%
Isolate	Execution Isolation	6%	1%	0%	3%	2%	3/4	4%
Detect	Network Traffic Analysis	0%	2%	0%	7%	2%	2/4	20%
Detect	User Behavior Analysis	1%	2%	0%	4%	2%	3/4	2%
Evict	File Eviction	3%	0%	0%	4%	2%	2/4	0%
Detect	Message Analysis	4%	0%	0%	2%	2%	2/4	2%

or non-existent in both. Additionally, Execution Isolation and Credential Eviction are somewhat common in playbooks but less so than in papers. Conversely, File Eviction is somewhat common in playbooks but not used in papers. Network Isolation is highly common both in papers and playbooks, but less so in playbooks.

A few playbooks were difficult to relate to D3FEND. More specifically, playbooks to restart services, restart machines, and reimage machines were found in the repositories. Additionally, the commercial playbooks frequently use different types of data enrichment, which is not the case in the papers. Approximately a quarter of the playbooks focus on enriching data, e.g. by looking up threat intelligence and combining data from systems in order to present the synthesized result to a human operator.

4 DISCUSSION

The discussion below discusses the homogeneity and maturity of the research (Section 4.1), the relationships between input data and output (Section 4.2), as well as the validity and reliability of this review (Section 4.3).

4.1 Homogeneity and maturity of the research

As described above, the solutions described in the literature cover a wide range of inputs and responses, with clear differences between them. Alternatives to this heterogeneity and variation would have been possible. For instance, the research community could have collectively focused on a particular problem of dealing with a compromised machine, agreed on the type of input and output to work with, and researched techniques for these inputs and outputs. Clearly, this is not how the research community has acted. One possibility is that researchers have tried to defend against different cyber attacks, due to differing domains or due to evolutions of attacks and available defenses (inputs and outputs) over time. However, most papers do not mention a specific domain and instead use a generic network, albeit with some specified cyber-physical systems etc. The differences in inputs and outputs between generic and other domains were very small and only statistically significant for Platform Monitoring (larger in generics). It is also difficult to say much about differences between the inputs and outputs over time, not least since the publication levels were very low before 2016. There are significant differences for Platform Monitoring and Platform Hardening (both nonexistent before 2014). Conversely, some similarities can be seen among overall papers, with input consisting of a basic asset inventory, network alerts, and host alerts; and output as isolating machines or terminating suspicious processes.

Less than half of the included papers were published in a journal. Based on the citation data from Scopus, the average number of citations were 12, with four papers receiving the majority of the citations. No clear difference in the use of different techniques can be observed based on whether the paper is published in journals or has been cited often. However, papers with more citations tend to be more detailed and complete, leaving room for improvement. For instance, the four papers receiving more than half of all citations were associated with more techniques (5.3 per paper) than the other papers (2.5 per paper). As described above, most published papers present conceptual models supported by arguments of usefulness or some hypotheticals used to illustrate how it would work in practice. Some papers present tests or illustrations with small laboratory networks. No paper demonstrates the viability of the solution in a realistic scenario with real data from an operational network. It is not apparent why this is the case, and this presents a clear avenue for future research. Tools capable of emulating cyber threats in realistic settings have been available for more than a decade, e.g. as proof-of-concept demonstrations in cyber defense exercises [25]. In addition, the decision problems are relevant for most organizations today, with technical solutions for input (e.g. asset inventories and network-based detection) and output (e.g. reconfigure firewalls or shut down machines) available or easy to create. Because tests could have been conducted relatively easily, a guess is that no presented solution is mature enough to withstand such tests. Future reviews should delve deeper into the issue of immaturity and determine the reasons why research tends to be at such a foundational level.

4.2 Relationships between input data and output

This paper addresses automatic incident response solutions' inputs and outputs independently from each other. However, in practice,

the output (i.e. action) often depends on certain types of input (e.g. detection methods). For example, a solution that analyses properties of emails in order to detect phishing, will be able to block suspicious domains easily but be poorly equipped to lock accounts that were compromised by the phishing email if no data linking the email to accounts is available. On the conceptual level, the papers discuss these links. However, as the lack of proof-of-concept demonstrations suggests, many papers provide vague descriptions and make little effort to make these links clear. For instance, the inputs needed to determine if a particular action should be taken are largely missing. Additionally, this review treats the automated decision techniques as a black box. In other words, this review disregards the approach a paper's solution chooses a response based on an input, as long as the approach is semi-automated or automated. As such, it is out of scope for this review to consider whether there are particularities depending on the decision model, e.g. if some links between input and output occur depending on the particular type of AI algorithm or other (semi-) automated decision making algorithm. Future reviews are encouraged to look into the black box, for a white box (or gray box) approach. Such research should also consider how real-life complex attacks are performed, as chains of attack techniques, and thereby address a potential drawback of using the technique-based D3FEND framework.

On the level of D3FEND techniques, it is worth noting that all solutions that use Platform Monitoring as input produce Network Isolation as output, and 78% of the solutions using Network Traffic Analysis as input produce Network Isolation as output. One interpretation of this is that Network Isolation is the primary kind of defensive measure to employ. Another interpretation is that many Network Isolation measures require knowledge about the assets in the network, events on machines, and analysis of network traffic. For other inputs and outputs, the sample sizes make it difficult to distinguish any clear links. It should be noted that many of the playbooks in commercially available products concern the enrichment of data by connecting alerts to threat intelligence, asset inventories etc. One interpretation of this is that aggregation and normalization of data, in different formats and in different sources, is a substantial part of the problem of producing automated responses. It could be that researchers have avoided this input management part of the problem because they want to produce something generic and tool independent. It could also be that this part of the problem has been avoided simply because it is complicated and not as academically rewarding as describing mathematical reasoning on abstract properties such as signal strength, threat levels, and response costs. Conversely, there may be things that practitioners could learn from the current state of research. For instance, playbook developers could focus more on D3FEND techniques that are common in research but rare in practice, e.g. Network Mapping and Platform Monitoring for input, or Credential Hardening and Decoy Environment for output. Playbook developers could also learn from what inputs and outputs researchers focus on that are not in D3FEND nor in current playbooks. This pertains mostly to outputs, e.g., with researchers often changing priorities for detection processes, and sometimes warning attackers.

4.3 Validity and reliability of this review

This review is vulnerable to many types of bias, and the review makes many interpretations dependent on the authors' knowledge and understanding. Four of these interpretations are discussed below. The four interpretations relate to either the study identification, inclusion criteria, extraction or synthesis and comparison to commercial products.

First, the search string used to identify scholarly literature was intentionally narrowed in order to find a manageable number of records. It is the belief of the authors that the chosen search string and indexing database accurately captured the literature on automatic solutions for responding to cyber incidents. However, the requirement that papers should explicitly mention something related to automated responses excludes many solutions that focus more on the input than on the output. It is likely that a small portion of the many thousands of papers addressing intrusion detection systems also discuss responses without explicitly mentioning that in the abstract. These papers have been missed in this review. However, the authors of this review did not identify any such papers in the preliminary searches while constructing the search string, and were not aware of any such missing papers from previous experience.

Second, abstracts and full text records were used to sort out papers describing solutions of the type addressed in this review. The two authors rated papers independently, and an inclusive approach was taken. This inclusivity is visible in the extracted D3FEND techniques from each of the papers, where seven included papers did not contain anything that could be mapped to D3FEND. Despite the inclusive approach, it is still possible that some relevant papers were excluded in the screening process. In order to minimize differences between the authors' decisions and fine-tune the criteria, there were several discussions on the particularities of the decision making process, as well as some randomly chosen overlapping decisions, e.g. with both authors rating the same paper followed by discussion of rating differences. However, there was no complete pilot-review, where all stages of the review process were trialed in order to align judgements further or fine-tune inclusion and extraction criteria.

Third, the extraction process was associated with a number of judgement calls. As noted above, the papers were not always clear on what they required as input or produced as output, and they often produced outputs not compatible with D3FEND. For instance, a paper could mention that the data in a SIEM system is used as input, and not explicitly describe what data the SIEM system needs to contain. In many cases, the papers describe a generic concept and offer more or less concrete examples of how this concept can be instantiated. For instance, a list of example responses may be provided. In those cases, the examples are used as the scope of the solution. When no concrete examples were presented, no mapping was made to D3FEND techniques. The authors' ability to understand the solutions described in the papers, the use of concrete examples as the definitive scope of the solution, and the decision to ignore vague descriptions, are all threats to validity. However, the decision to focus on the concrete, but ignore vague descriptions, is justified by the assessment that vagueness will anyway be a poor guide for identifying research gaps.

Fourth, the synthesis and comparison to commercial products could have been done in other ways. Alternatives to D3FEND were considered. In retrospect, it is clear that many of the vaguely described solutions were a poor fit with the abstraction level of D3FEND. A model with input as defined by [1] (network input, host input, or input from a SIEM system) would have suited many of the academic papers better. On the other hand, a model on the level of abstraction of Splunk's Common information model is what an operational solution would have to work on and would be more informative to a practitioner, and could have been mapped to the commercial playbooks. D3FEND was chosen because it is a compromise between these two extremes.

5 CONCLUSIONS

This paper identified 45 papers describing a technical solution that automatically responds to cyber attacks, and analyzed the papers' input and output using the D3FEND framework. The input to these solutions were a mix of system information and detection alerts. The system information available in an Asset Inventory (31%) and details about Network Mappings (16%) were commonly used. The more common detection inputs were Platform Monitoring (11%) and Network Traffic Analysis (22%). Among responses, Network Isolation (67%), Process Eviction (33%), and Platform Hardening (18%) were the most common. In addition, many solutions proposed measures not covered by D3FEND, e.g. to shut down machines (24%). Commercially available solutions appear to be more focused on Identifier Analysis (e.g. reputation of domains) and File Analysis (e.g. scan email attachments). Commercial repositories also contain many playbooks enriching data by looking up things in different databases.

REFERENCES

[1] Anuar, N.B. *et al.* 2010. An investigation and survey of response options for Intrusion Response Systems (IRSs). *Proceedings of the 2010 Information Security for South Africa Conference, ISSA 2010.* (2010). DOI:https://doi.org/10.1109/ISSA.2010.5588654.

[2] Anwar, S. *et al.* 2017. From intrusion detection to an intrusion response system: Fundamentals, requirements, and future directions. *Algorithms.* 10, 2 (2017). DOI:https://doi.org/10.3390/a10020039.

[3] Bashendy, M. *et al.* 2023. Intrusion response systems for cyber-physical systems: A comprehensive survey. *Computers & Security.* 124, (Jan. 2023), 102984. DOI:https://doi.org/10.1016/j.cose.2022.102984.

[4] CACAO Security Playbooks Version 1.0: https://docs.oasis-open.org/cacao/security-playbooks/v1.0/cs01/security-playbooks-v1.0-cs01.html. Accessed: 2023-05-11.

[5] Erola, A. *et al.* 2017. RicherPicture: Semi-automated cyber defence using context-aware data analytics. *2017 International Conference On Cyber Situational Awareness, Data Analytics And Assessment (Cyber SA)* (Jun. 2017), 1–8.

[6] Grance, T. *et al.* 2008. *Computer Security Incident Handling Guide (SP 800-61).*

[7] Hughes, K. *et al.* 2021. Towards intrusion response intel. *Proceedings of the 2021 IEEE International Conference on Cyber Security and Resilience, CSR 2021.* (2021), 337–342. DOI:https://doi.org/10.1109/CSR51186.2021.9527957.

[8] Husák, M. and Cermak, M. 2022. *SoK: Applications and Challenges of using Recommender Systems in Cybersecurity Incident Handling and Response.* Association for Computing Machinery.

[9] InsightConnect Workflows: 2023. https://github.com/rapid7/insightconnect-workflows. Accessed: 2023-05-29.

[10] Islam, C. *et al.* 2019. An ontology-driven approach to automating the process of integrating security software systems. *Proceedings - 2019 IEEE/ACM International Conference on Software and System Processes, ICSSP 2019* (2019), 54–63.

[11] Kaloroumakis, P.E. and Smith, M.J. 2021. Toward a Knowledge Graph of Cybersecurity Countermeasures. (2021).

[12] Kholidy, H.A. *et al.* 2016. A risk mitigation approach for autonomous cloud intrusion response system. *Computing.* 98, 11 (2016), 1111–1135. DOI:https://doi.org/10.1007/s00607-016-0495-8.

[13] Kinyua, J. and Awuah, L. 2021. Ai/ml in security orchestration, automation and response: Future research directions. *Intelligent Automation and Soft Computing.* 28, 2 (2021), 527–545. DOI:https://doi.org/10.32604/iasc.2021.016240.

[14] Kitchenham, B. 2004. *Procedures for performing systematic reviews.* Citeseer.

[15] Kotenko, I. *et al.* 2022. Systematic Literature Review of Security Event Correlation Methods. *IEEE Access.* 10, (2022), 43387–43420. DOI:https://doi.org/10.1109/ACCESS.2022.3168976.

[16] Mavroeidis, V. and Brule, J. 2020. A nonproprietary language for the command and control of cyber defenses – OpenC2. *Computers and Security.* 97, (2020), 101999. DOI:https://doi.org/10.1016/j.cose.2020.101999.

[17] Microsoft Sentinel and Microsoft 365 Defender: 2023. https://github.com/Azure/Azure-Sentinel/tree/master/Playbooks. Accessed: 2023-05-29.

[18] Mitropoulos, S. *et al.* 2006. On Incident Handling and Response: A state-of-the-art approach. *Computers & Security.* 25, 5 (Jul. 2006), 351–370. DOI:https://doi.org/10.1016/j.cose.2005.09.006.

[19] Navarro, J. *et al.* A Systematic Survey on Multi-step Attack Detection.

[20] Petersen, K. *et al.* 2015. Guidelines for conducting systematic mapping studies in software engineering: An update. *Information and Software Technology.* 64, (2015), 1–18. DOI:https://doi.org/10.1016/j.infsof.2015.03.007.

[21] Phantom Community Playbooks: 2023. https://github.com/phantomcyber/playbooks. Accessed: 2023-05-29.

[22] Piedrahita, A.F.M. *et al.* 2017. Leveraging Software-Defined Networking for Incident Response in Industrial Control Systems. *IEEE Software.* 35, 1 (2017), 44–50. DOI:https://doi.org/10.1109/MS.2017.4541054.

[23] Piedrahita, A.F.M. *et al.* 2018. Virtual incident response functions in control systems. *Computer Networks.* 135, (2018), 147–159. DOI:https://doi.org/10.1016/j.comnet.2018.01.040.

[24] Schlette, D. *et al.* 2021. A comparative Study on Cyber Threat Intelligence: The Security Incident Response Perspective. *IEEE Communications Surveys & Tutorials.* c (2021), 1–1. DOI:https://doi.org/10.1109/comst.2021.3117338.

[25] Sommestad, T. and Hallberg, J. 2012. Cyber security exercises and competitions as a platform for cyber security experiments.

[26] ThreatConnect Playbooks: 2023. https://github.com/ThreatConnect-Inc/threatconnect-playbooks/tree/master/playbooks. Accessed: 2023-05-29.

APPENDIX

Table 5 provides a list of the extracted D3FEND techniques from each paper.

Table 5: The extracted D3FEND techniques from each paper.

D3FEND technique(s)	DOI or reference
	10.3390/a16020112
D3-SUD3-NTFD3-PT, D3-CR	10.1145/3560830.3563732
D3-AVE, D3-SWI, D3-NNI, D3-SU, D3-NTCD, D3-NTF	10.1145/3538969.3538975
D3-NNI, D3-PLLM, D3-NTPM, D3-AVE, D3-NTF	10.1109/GCWkshps56602.2022.10008548
	10.1109/NextComp55567.2022.9932254
D3-NNI, D3-SWI, D3-IPRA, D3-OSM, D3-DE	10.1109/DSC54232.2022.9888808
D3-NTF	10.1109/CSR54599.2022.9850304
	10.32604/cmc.2022.028495
D3-NTF	10.1007/978-3-031-02067-4_6
D3-SU, D3-NTF, D3-PT	10.1016/j.future.2020.09.002
D3-AVE, D3-AM, D3-NTPM, D3-CH, D3-ACH, D3-NTF	10.1109/SSCI50451.2021.9659882
D3-NNI, D3-DI	10.1016/j.cose.2020.101999
D3-SU, D3-NTF	10.1109/GLOBECOM38437.2019.9013291
D3-NTA, D3-PM, D3-NTF, D3-PT	10.1109/ICSSP.2019.00017
D3-NNI, D3-HCI, D3-NTF	10.2197/ipsjjip.27.564
D3-PT	10.5220/0007715201480158
D3-NTF	10.5220/0007359503190327
D3-SPP, D3-OSM, D3-NTA, D3-ISVA, D3-IPCTA, D3-NTF, D3-PT	10.1109/NCS.2018.00007
D3-SU, D3-NTF, D3-PT	10.1109/TDSC.2016.2615622
D3-NNI, D3-NTPM, D3-SWI, D3-AM	10.1145/3230833.3232798
D3-AVE, D3-SVCDM, D3-NTF	10.1109/PDP2018.2018.00081
D3-NTF, D3-DE	10.1016/j.comnet.2018.01.040
D3-CI, D3-NNI, D3-NTPM, D3-SU, D3-NTF, D3-DE	10.1145/3168446
	10.1007/978-3-319-98385-1_14
D3-LLM	10.1007/978-3-319-93767-0_12
D3-OAM, D3-UBA, D3-SFA, D3-NTA, D3-NTF, D3-IOPR	10.1109/CyberSA.2017.8073399
D3-NTF, D3-IHN	10.1109/MS.2017.4541054
	Goodgion J., Mullins B., Active network response using host-based IDS and software defined networking, ICCWS 2017.
D3-NI, D3-PT	10.1109/MILCOM.2016.7795152
D3-ANAA, D3-NTF, D3-PT	10.1007/s00607-016-0495-8
	10.1145/2994539.2994545
D3-UAP, D3-PT, D3-NTF, D3-AL	10.1109/IranianCEE.2016.7585693
D3-SVCDM, D3-SYSDM, D3-SYSVA, D3-AM, D3-SWI, D3-DI, D3-LFP, D3-SU, D3-PT, D3-NTF, D3-KBPI, D3-AL	10.1016/j.cose.2016.06.005
D3-NTA, D3-NNI	10.1109/ICOIN.2016.7427090
D3-SYSM, D3-AM, D3-AVE, D3-LLM, D3-NNI, D3-NTF, D3-ITF, D3-OTFD3-PT, D3-PS, D3-ANCI, D3-AL	10.1016/j.jnca.2015.05.004
D3-MA, D3-CAA, D3-FA, D3-DNSDL, D3-NTF, D3-DNR	10.1109/ARES.2014.46
D3-ORA, D3-HCI, D3-SWI, D3-SU, D3-OSM, D3-SCAD3-NTF, D3-PT	10.1007/s10207-013-0222-9
D3-CHN	10.1007/978-3-642-30507-8_30
D3-NI, D3-DE	10.1080/18756891.2012.733205
D3-NTF, D3-ANCI, D3-PT	10.1109/EC2ND.2010.11
D3-NTA	10.1109/ETCS.2009.255
D3-UAP, D3-NTF, D3-PT, D3-AL, D3-FR,	10.1109/KAMW.2008.4810553
D3-NNI, D3-NTA, D3-NTF, D3-IHN,	10.1109/ICN.2007.92
	10.1016/j.istr.2005.07.002
D3-LLM, D3-SYSM, D3-ODM, D3-NTPM, D3-SYSDM, D3-SVCDM, D3-NTF, D3-PT, D3-AL	10.1109/CSAC.2002.1176302