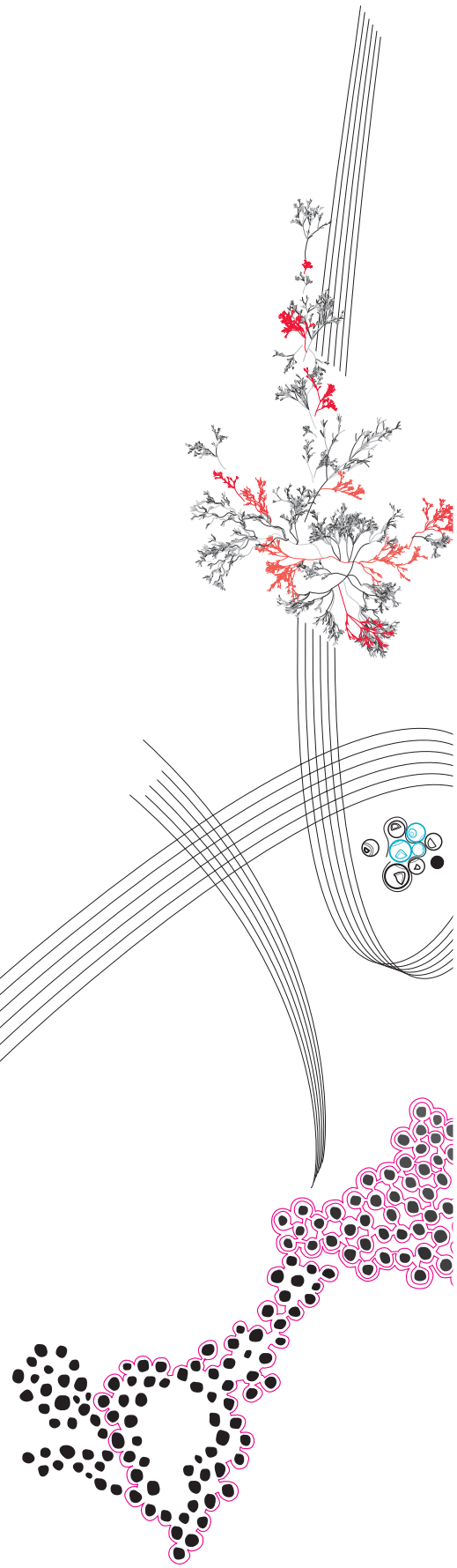

MSc Computer Science
Master Thesis

Enhancing Explainability in Alert Triaging for Improved Security Event Analysis:

Integrating Domain-Specific Knowledge from the MITRE ATT&CK Framework

Remon Hoogendijk

EXAMINATION COMMITTEE:
dr. ir. Thijs van Ede,
dr. ir. Andrea Continella,
dr. ir. José Jair Cardoso de Santanna,

EXTERNAL SUPERVISION:
ir. Mathijs Barendse (KPMG),
ir. Davey Kaak (KPMG)

August, 2024

Department of Computer Science
Faculty of Electrical Engineering,
Mathematics and Computer Science,
University of Twente

UNIVERSITY OF TWENTE.

ABSTRACT

This study attempts to enhance cyber attack analysis by integrating the MITRE ATT&CK Framework into an alert triaging tool, aiming to improve visualization and explanation of entire cyber kill chains. For this study, we attempt to do this for the tool DeepCASE. DeepCASE is an AI triaging tool that clusters events based on contextual similarity. The methodology we used involved mapping security events to the techniques in the MITRE ATT&CK Framework, which are then separated into phases, and plotting these techniques in a graph-like attack structure and after this, using the DeepCASE system to find the most probable attack path in this structure. The evaluation involved testing assumptions and choices made during the development process, such as the mapping to the MITRE ATT&CK Framework, the method used to create the attack graphs, and how paths are found in these graphs. The evaluation results indicate that while the methodology cannot reliably find attack paths in the created graphs, there are many possible adjustments that could improve the system's performance in finding these paths. Improvements such as a better mapping of events to the MITRE ATT&CK Framework, using a dependency graph instead of phases, and integrating other frameworks and tools are all examples of potential improvements to the system.

1. INTRODUCTION

In today's rapidly evolving and growing digital landscape, securing one's data has become more important than ever. Security Operation Centers (SOC) have become the recommended best practice for organizations to manage the security of their digital infrastructure [1].

To battle increasingly numerous and complex cyber attacks, organizations have turned to advanced technologies and artificial intelligence (AI) solutions. SOC operators are supported by AI-driven tools that can automate repetitive tasks, and offer rapid analysis of security events [2]. Among these AI-driven tools is DeepCASE, an innovative research project aimed at reducing the workload of security analysts by clustering security events based on their contextual similarities [3]. Clustering allows analysts to review fewer events while still keeping up high coverage and reviewing fewer events naturally result in a lower workload.

While DeepCASE has shown significant improvement in SOC operations, it highlights the crucial aspect of explainability in the cybersecurity domain. In a domain where trust and transparency are important for the adoption of new tools, understanding and interpreting the decisions made by AI systems such as DeepCASE becomes paramount [4]. While DeepCASE does use techniques to show which events were important for its conclusion, the explanations provided by this technique do not fully align with the needs of the security analyst who require a more domain specific explanation. To fully meet the needs of security analysts, the explanation should connect the flagged events to broader threat patterns. The provided explanation should be easily interpretable and provide a rationale that aligns with the analyst's domain knowledge and workflow.

This research aims to integrate explainability into the DeepCASE using domain-specific knowledge encoded in the MITRE ATT&CK [5] Framework. MITRE ATT&CK is a knowledge base of different adversarial tactics and techniques provided by the MITRE organization. The aim is to offer analysts a more coherent and human-understandable representation of security events resulting in better-informed decision-making.

To ensure our approach aligns with the practical needs of security operators, we conducted preliminary interviews to gain a clear understanding of their current methodologies and to validate our direction towards enhancing explainability.

1.1 Contributions

Concretely, with this paper, we:

- Conducted preliminary interviews with security operators to understand their current approaches and to verify the relevance of our explainability objectives;
- Conducted a comprehensive analysis of integrating the MITRE ATT&CK Framework with the DeepCASE system to enhance the explainability of cybersecurity event analysis;
- Identified challenges in mapping security events to the MITRE ATT&CK Framework and in visualizing these events within a graph-like attack structure;
- Provided a detailed evaluation of the current methodology, highlighting areas where the approach fell short and suggesting specific improvements for future research.

In the spirit of open science, we publish our source code on GitHub.*

2. BACKGROUND

2.1 SOC

The Security Operations Center (SOC) is a centralized environment for monitoring an organization's security events and has become a recommended best practice [1]. A traditional SOC consists of three main components [6]: analysts, tooling, and procedures. Analysts handle various responsibilities from triaging events to conducting threat hunting. Tooling includes systems like IDS, SIEM, and SOAR to monitor and automate responses [2; 3]. Procedures ensure standardized and efficient responses to incidents [7].

2.1.1 Security Events

Monitoring security events is vital for detecting and responding to threats. Tools like Security Monitoring Tools, Firewall logs, and Endpoint Protection Systems track various events, such as contact with new hosts, port scans, or the use of self-signed certificates [3]. The challenge lies in distinguishing between malicious and benign events due to the sophistication of attacks and the large volume of data, which can lead to false positives [8; 9; 10].

3. RELATED WORK

Understanding and visualizing attacker behavior is essential for cybersecurity. Current methods like SAGE and Behavior Nets lead this effort. SAGE uses sequence learning to create attack graphs from intrusion alerts, visualizing potential attack paths and aiding in future attack prediction. Behavior Nets extend Petri Nets to model multi-step attacks by tracking system call dependencies, offering detailed insights into attack sequences. These methods, along with other relevant studies, are discussed in this section.

* https://github.com/RemonHoogendijk/ExplainableDeepCASE

Our work advances these methods by integrating the MITRE ATT&CK Framework into DeepCASE. This enhancement provides domain-specific explanations and graph-like attack structures, improving DeepCASE's interpretability and practical usability for cybersecurity professionals.

3.1 DeepCASE

One work that aims to reduce the workload of security operators is DeepCASE. DeepCASE searches for correlations within sequences of events generated by a specific device. With this context, it clusters events based on the similarity of not only the event itself but also the similarity of their context. A security operator only needs to evaluate a handful of events from a cluster which allows DeepCASE to evaluate the rest of the cluster based on the decision of the security operator [3].

DeepCASE consists of two main parts. A Context Builder and an Interpreter.

3.1.1 Context Builder

The Context Builder identifies which events from the context are relevant for the current event. For this purpose, a Recurrent Neural Network (RNN) that incorporates attention vectors is used [11]. These attention vectors are more commonly used within Natural Language Processing (NLP) to determine the relevant parts of an input sequence with respect to the output [12; 13; 14].

Other state-of-the-art works try to predict the next event using Long Short-Term Memory (LSTM) [15], RNNs [16], or word embeddings [17]. The attention vectors of DeepCASE have the advantage of also being able to show which of the events in the context were relevant for the current event. This is not possible with the other works.

The Context Builder gets trained as if it were to predict the current event using only the context. It looks at which events from the context were relevant for that prediction and which were not. Normally, this setup would be used to predict the next event in the sequence. However, the Context Builder is not designed to predict since the current event is already known, but rather it is designed to identify which parts of the context were relevant for the prediction of the current event. If it's unable to predict the current event, DeepCASE falls back on the security operator.

3.1.2 Interpreter

When the Context Builder has computed the attention for each event in the context, the Interpreter uses this to compare different event sequences and cluster them together. The underlying thought here is that when events generate similar attention vectors they can be treated similarly by the security operators and thus be clustered together [3]. To this end, the interpreter has two main functionalities: Attention Queries, and Clustering.

Attention querying is done to limit the effect of incorrect predictions on the security operator when the Context Builder cannot properly predict the correct event. The attention query asks if the Context Builder cannot predict the correct event with the current context attention vectors, what should the attention vector be to predict the correct event? If after this the correct event is predicted, these values are used and stored. If the Context Builder still cannot properly predict the event, it is passed through for manual inspection.

When each sequence is modeled with the attention vectors the Interpreter clusters together event sequences with similar vectors. Events are clustered together using DBSCAN, a clustering algorithm that does not force each entry into a cluster allowing for individual outliers in case they do not fit a cluster well enough [18].

3.2 Explainable AI

DeepCASE makes use of machine learning in order to see what events in the context are important. The problem with many machine learning or artificial intelligence (AI) models is that they are practically black boxes. Information gets put in, the black box performs its operations, and the output gets generated. However, it is unknown why that output is generated from the input data. In security, knowing why a machine learning system made its decision is advantageous because it allows for transparency and accountability, enabling better error analysis, and detection of vulnerabilities [4].

Explainable Artificial Intelligence (XAI) has been proposed to open the black box and make its internals more understandable to human operators [19]. It does this by either approximating the black box model or by explaining the model in terms of components or input examples. When approximating the black box model, a second model that is simpler and interpretable is trained on the predictions of the original black box model in order to interpret it. When explaining a model in terms of components or input examples, either examples from the data-set are used as an explanation, or the different building blocks of the model are used to explain the model's workings. These building blocks can be anything from model inputs to neurons or layers. Adadi et al. [20] provides a taxonomy of XAI methods and strategies. The methods are classified according to three criteria: (i) Complexity of interpretability, (ii) Scope of interpretability, (iii) Level of dependency from the used ML model.

Complexity of interpretability The complexity of a machine learning model often makes it difficult to interpret. To achieve interpretability, some models are designed to be simple [21; 22; 23; 24], though this may reduce accuracy [25; 26]. Alternatively, complex models can be used for accuracy, with post-hoc methods [27] like natural language explanations [28], visualizations [29], or examples to interpret them [30].

Scope of interpretability This involves understanding a model either globally or locally. Global interpretability seeks to explain the entire logic of a model, which is challenging for complex systems [21; 22; 31; 32; 33; 34]. Local interpretability focuses on explaining specific decisions or predictions, making it more practical, especially for Deep Neural Networks (DNN) [35; 36; 37; 38; 39; 40; 41; 42].

Model related methods Interpretability can be model-specific or model-agnostic. Model-agnostic methods, which are often post-hoc, can be applied to any model. They include visualization, knowledge extraction, influence methods, and example-based explanations. Model-specific interpretability is tailored to particular models and thus limits flexibility in model choice. Due to this, model-agnostic interpretability is often post-hoc. These

interpretability models broadly fall into four categories: (i) Visualisation, which aims at explaining the model patterns through graphical or visual format; (ii) Knowledge extraction, which aims at extracting rules from the model in a comprehensible form; (iii) Influence methods, whose aim is to estimate the importance or the relevance of features; and (iv) Example-based explanation, which selects particular instances in the data-set to explain model behavior.

3.2.1 XAI Stakeholders

When making a machine learning model explainable it is important to realize who this explainability is designed for. For explainable AI there are generally two stakeholders: model users and model designers. In the case of cyber security, three stakeholders can be defined: model users, model designers, and adversaries [43]. These three stakeholders utilize XAI for four different goals. Techniques that are developed and utilized to support model users in making decisions are part of *XAI-enabled user assistance*. These techniques are meant to give control back to the user by helping them understand the model [44]. *XAI-enabled model verification* techniques are developed and utilized to help model designers in debugging and fine-tuning machine learning models. The techniques for *explanation verification & robustness* are meant to help model designers validate the correctness and robustness of the different explainable AI components. They test the functioning of explanations under natural settings [45] and under adversarial settings [46]. Lastly, *offensive use of explanations* is done by adversaries to enhance their capabilities. This can be done by either using explanations to compromise the privacy of the model or by using explanations to compromise the integrity and availability of the model [47; 48].

3.2.2 DeepCASE and XAI

DeepCASE already has some XAI elements in the form of attention vectors. When looking at DeepCASE with the taxonomy of XAI in mind we can make statements about DeepCASE regarding the three criteria: Firstly, the Complexity of interpretability, while DeepCASE's machine learning model is fairly complex its explainability is inherent from its use of attention vectors and not a post-hoc solution. Secondly, the Scope of interpretability, we can safely say that DeepCASE makes use of local interpretability by providing the attention vectors on a per-input basis with the inputs being event sequences. Lastly, in terms of model-related methods, while DeepCASE does not use a post-hoc implementation, attention vectors have been used for other models to provide explanations of feature importance. Looking at these three criteria we can clearly see that DeepCASE already makes use of XAI. However, Jain et al. [49] argue that attention vectors do not provide a sufficient explanation for machine learning models, and when looking at how DeepCASE would be used in practice it is easy to see why. While attention vectors might give clarity on which feature was important for the current prediction, this is not a proper explanation for a security operator, the model user, who might require a more domain-specific explanation in order to trust the model. DeepCASE's current explainability is more focused on model verification while its actual goal should be user assistance since that is the goal of DeepCASE as a whole. In our work, we aim to improve the explainability of DeepCASE by introducing graph-like attack structures using the MITRE ATT&CK Framework.

3.3 Knowledge Frameworks

For DeepCASE, a domain-specific explanation could make use of existing frameworks that aim at understanding and explaining adversary behavior. In cybersecurity, being able to properly defend against adversarial attacks requires an understanding of how adversaries operate, how threats unfold, and how security breaches can escalate into bigger incidents [1]. Different frameworks have arisen that aim at building a knowledge base on tactics, techniques, and procedures used by adversaries. An example of such a framework is the MITRE ATT&CK [5] Framework.

The MITRE ATT&CK Framework is a tool developed by the MITRE Corporation. This framework serves as a detailed knowledge base describing adversarial tactics, techniques, and procedures during various stages of an attack. It provides a structured way to understand threat actor behavior allowing organizations to better defend against, detect, and respond to cyber threats. The framework is organized into matrices that outline specific tactics that adversaries employ in their attacks. This includes tactics such as reconnaissance, initial access, and exfiltration. In total the MITRE ATT&CK Framework contains 14 tactics. For each tactic, multiple techniques are outlined that attackers can employ to fulfill the objective of that tactic [5].

3.4 Attack Graphs

Another way of classifying attacks and getting a clear understanding of attacker behavior using domain knowledge is the use of attack graphs (AG). Attack graphs have been widely used for visual analytics [50; 51] and network hardening [52; 53].

An AG is a visual representation of potential attack paths within a computer network or system. Usually, generating attack graphs requires substantial amounts of prior knowledge and published vulnerability reports [54]. Nadeem et al. [54] showed that it is possible to construct accurate attack graphs using prior intrusion alerts on the system using their tool "SAGE". The SAGE tool uses sequence learning to mine patterns from intrusion alerts and model them using an automaton. It is then able to represent this in the form of an attack graph. This way, attack graphs can be used to visualize attacker behavior during previous attacks. The visual representation of attack paths in SAGE served as the basis for the attack graphs in this work, illustrating attack chains in a similar style.

3.5 Petri-Nets

Another method that can be used to visualize behavior in an understandable way is with the help of petri-nets. Just like an AG, petri-nets are graph models and are used to graph out processes due to their ability to compactly represent complex behaviors such as exclusive choices, optional activities, and concurrency [55; 56].

Petri-nets are nets where places can be marked by tokens. Tokens are consumed when transitions are fired and new tokens will be generated at all the outputs. A transition can only fire when at least one token is placed at each input [57].

3.5.1 Behavior Nets

Another form of Petri nets is behavior nets [57]. Behavior nets are a re-implementation of well-known behavior graphs [58; 59]. Behavior nets were used to express malware behavior in terms of API/System calls and their interdependence. A behavior net, much like a Petri net, allows tokens to flow through the net via transitions. The main difference is that in behavior nets the transitions are labeled with observable events, and the tokens can carry contextual information between those transitions. The transitions will become enabled when there are a sufficient number of tokens at the inputs and the contextual information provided by these tokens matches specific patterns or predicates.

Behavior net's ability to track multi-step attacker behavior is achieved by modeling dependencies between system calls. This requires the tracking of specific parameters to those calls. This work investigates a similar idea of tracking multi-step attacker behavior. Our approach is abstracted to the MITRE ATT&CK Framework to cover more types of attacks and remove the requirement of tracking parameters.

4. PRELIMINARY INTERVIEWS

We conducted preliminary interviews with cybersecurity professionals from different security operation centers to address gaps in the literature and gather practical insights that are often not documented in academic sources. While the literature provides a theoretical foundation for integrating explainability into SOC tooling, it often lacks detailed, real-world perspectives on the practical challenges and needs of security operators. These interviews served as an exploratory phase to assess the feasibility and relevance of our proposed approach to combine cyber kill chains with DeepCASE, ensuring that the explainability layer we aim to develop aligns with the actual workflows and decision-making processes of security professionals.

4.1 Participants

For the Interviews, a selection of cybersecurity experts was recruited. In total, five different interviews were held with experts from the Eindhoven Security Hub (ESH), Eviden, and KPMG. From ESH, we interviewed a tier one, tier two, and tier three analyst; From Eviden, we interviewed a tier two analyst; and from KPMG, we interviewed a SOC manager. Each interviewee had a different level or type of experience in the cybersecurity field (Table A2), allowing for different perspectives from different roles and organizations.

4.2 Interview Structure

The interviews were semi-structured but, for the most part, resembled structured interviews. the interview structure included 25 main questions and 26 follow-up questions. A full overview of the questions can be found in Appendix A. The interviews took between 20 and 30 minutes per person. The questions were meant to get the current analysis process clear and see if adding cyber kill chain information during the DeepCASE analysis would improve the explainability of DeepCASE.

Before conducting the interviews, ethics approval was obtained from the Ethics Committee of the University of Twente, ensuring that the study adhered to ethical guidelines.

4.3 Data Collection and Analysis

Interviews were recorded for transcription and then deleted for privacy reasons. The transcriptions of the interviews were encoded in order to analyze which themes and topics were prominently discussed. Thematic analysis was used to identify and report patterns in the qualitative data. The process began with a familiarization phase. During this familiarization phase, significant words, phrases, or segments were identified and highlighted. Each highlighted segment was then assigned a brief label that encapsulated the essence of the segment. These initial codes served as the foundation for identifying broader themes.

The next step involved examining the codes to identify patterns and commonalities. Similar codes were grouped to form overarching themes. Although the thematic analysis provided valuable preliminary insights, the small sample size of five interviews means that the findings are not generalizable. These initial results, however, offer a foundational understanding that can be used for developing our methodology.

Appendix B contains an overview of the key themes identified from the thematic analyst in Table A1. The appendix also shows excerpts from the interview transcripts to illustrate the participants' perspectives for each theme.

4.4 Interview Conclusion

The goal of the interviews was to assess the feasibility and relevance of integrating cyber kill chains with DeepCASE to provide an explainability layer for DeepCASE and to guide the development of this approach. The interviews revealed a clear need for an overview from the analysts, with a direct mention of how linking different tactics from the MITRE ATT&CK Framework would enhance the analysis process. While our approach appears relevant, the interviews also highlight several challenges, such as the lack of information and the difficulty of detecting certain MITRE ATT&CK techniques. Additionally, they portray a one-sided view of the data generated by security events, with certain techniques like brute-force attacks, scans, and web exploits being particularly prominent. These insights will assist in further developing our approach. Analysts can already perform a detailed investigation of alerts but lack an overview of what has occurred on a machine prior to the current event. DeepCASE attempts to provide this overview by using context sequences to perform its clustering, but this relatively small overview cannot span multiple techniques, which is what the analyst lack according to the interviews. This feedback will guide our efforts to enhance DeepCASE, ensuring it offers a broader and more comprehensive view across various techniques, thereby improving its utility and effectiveness for analysts.

The interviews will influence the decisions we make in several key areas:

Event mapping to MITRE ATT&CK Framework: The interviews provided insights into the event landscape that will assist the mapping of events to techniques in the MITRE ATT&CK Framework.

Phases for categorization: Inspired by the discussion on the chain of events and how events become harder to analyze "further" into the MITRE ATT&CK Framework,

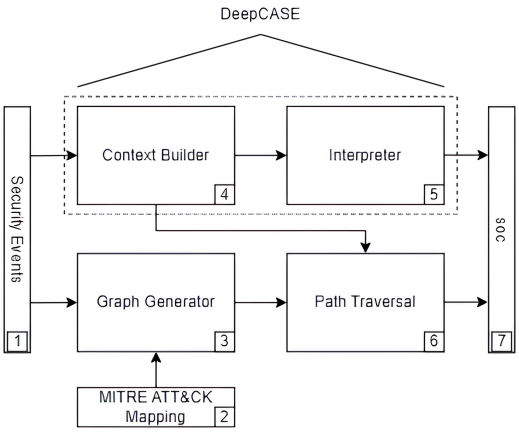

Figure 1. Overview of all system components. 1) Both DeepCASE and the Graph Generator use security events as input. 2) The phases of the events are based on the mapping of security events to MITRE ATT&CK Techniques. 3) The graph generator creates an initial graph based on the temporal location and phase of each event. 4&5) DeepCASE creates its attention values for each context sequence and outputs this for the Path traversal to use. 6) The Path Traversal takes the output of DeepCASE and the generated graph to display the most likely path in the graph. 7) Both the DeepCASE output and the Path Traversal output are sent to a Security Operation Center for analysis

we will categorize techniques into phases based on the MITRE ATT&CK Tactic they belong to.

Graph-like Attack Structure: The interviews highlighted the need for an overview of an entire attack chain. This formed our decision to create this overview in a graph-like structure similar to the attack graphs generated by SAGE.

5. METHODOLOGY

We observed that DeepCASE does not consider enough events to capture an entire attack chain. Cyber attacks often involve multiple steps over hours or days [60], but DeepCASE's default context sequences consist of only 10 events, which is insufficient. Expanding this window to encompass entire attacks is impractical. In this work, we explore how combining multiple DeepCASE sequences can visualize and explain entire cyber kill chains using the MITRE ATT&CK Framework.

To achieve this, we adapted the concept of attack graphs from SAGE by Nadeem et al. [54], diverging in two key areas. First, while SAGE constructs attack graphs using machine learning, our approach uses heuristic algorithms to make the approach inherently explainable. Second, instead of using milestones achieved by the attackers as graph nodes, we use MITRE ATT&CK techniques, ensuring comprehensive coverage and standardization in depicting adversary tactics. Previous works also introduced behavior nets [61] to track multi-step attacker behaviors, which depend on specific system call parameters. Our approach abstracts this idea to MITRE ATT&CK techniques, allowing us to cover more attack types without tracking individual parameters.

The overall plan is to identify the most likely attack path taken by the attacker. We begin by creating attack graphs that display all possible attack paths. Then, we use DeepCASE's sequence analysis and attention values to model the likelihood of each path and identify the most probable attack scenario. This system comprises three main parts that integrate with DeepCASE to display larger attacks in their entirety, as shown in Figure 1. First, events are mapped to the MITRE ATT&CK Framework [5] (indicated by icon 2). This mapping, using various techniques, helps create the attack graphs. Additionally, techniques are split into multiple phases according to their overarching tactics, which will be further explained in Section 5.1. We assume that attackers always aim to advance their attack, so the attack can only move to later phases, not earlier ones. The graph generator, indicated by icon 3, uses the security events and their MITRE ATT&CK mappings to determine the order of events in the graph, based on the defined phases. The resulting graph represents possible attack paths found in the security alerts. Finally, path traversal, indicated by icon 6, uses this graph and the attention values from DeepCASE [3] to identify the most likely attack path within the graph.

5.1 MITRE ATT&CK Mapping

As explained in Section 3.3, each tactic in the MITRE ATT&CK Framework consists of multiple techniques. These techniques are used by an attacker as part of their attacking strategy. The MITRE ATT&CK Mapping links events to their corresponding technique. These techniques are part of an overarching tactic which we sort into phases to show the dependencies between them. Phases are set up in a way that they logically follow each other. E.g., often attackers first perform reconnaissance before gaining initial access and subsequently executing malicious commands. Using this logic, techniques in the reconnaissance tactic precede techniques related to the initial access tactic, which subsequently precedes execution techniques. Here, we make the assumption that attackers always want to advance their attack to a further stage and thus an attack only moves to later phases, not laterally or to earlier ones. This assumption is tested in Sections 6.2 and 6.3.

In some instances, a technique may belong to multiple tactics. In such cases, the technique's role depends on when in the attack chain the technique is used and for what purpose. This dual categorization can influence the later parts of the system, which will be discussed in both Sections 5.2 and 5.3.

Table 1 assigns each tactic to a phase based on their possible sequence and concurrency. 'Reconnaissance' and 'Discovery' are placed at the start, as they gather information before other tactics begin. 'Initial Access' and 'Credential Access' follow, requiring access to the victim machine. Once access is gained, tactics such as executing code, escalating privileges, lateral movement, defence evasion, and persistence are grouped in phase 3, as they can occur in any order after access to a system is acquired. 'Collection' precedes 'Exfiltration,' placing them in phases 4 and 5, respectively. 'Command and Control' also precedes 'Exfiltration' and is included in phase 4. 'Impact,' typically the final action, is grouped with 'Exfiltration' in phase 5. The 'Resource Development' tactic from the MITRE ATT&CK Framework is excluded from these phases as it is not visible on the victim's network.

We assume that the mapping between the MITRE ATT&CK Framework and the security events is something a SOC already has. This study's mapping was made for the events found in the Open-source Collegiate Penetration Testing Competition dataset (CPTC-2018). The mapping is found in Appendix C. We cre-

ated the mapping by comparing the event's description and detection rule with the descriptions of the techniques found in the MITRE ATT&CK Framework. Some events in this dataset do not directly map to any technique found in the MITRE ATT&CK Framework. We labeled these events as 'None' and excluded them from both the graph generation and path traversal steps; they are not included in Appendix C. These unlabeled entries account for 2.9% of the events in the dataset. In the created mapping the same technique occurs often, more specifically, the technique "T1190" Exploit Public-Facing Application. This technique was mapped to 76 out of the 145 unique events. The second most mapped technique is "T1595" Active Scanning with 14 occurrences. This closely matches what was discussed in the preliminary interviews where it was stated that these techniques are very prevalent. Section 6.6 evaluates the mapping of events to MITRE ATT&CK techniques and the subsequent categorizations of tactics to phases based on expert opinion.

The later parts of our approach depend heavily on mapping MITRE ATT&CK Tactics to phases. In Section 6.5, we test the effect of the categorization to phases on the generated paths by creating the paths with an alternative mapping of tactics to phases.

5.2 Graph Generator

The goal is to get a full picture of the attack performed on the victim host. To this end, we model an attack graph to be an overview of all events that took place on a machine and show their connection to one another. We recall that we map events to MITRE ATT&CK Techniques, this has the advantage that rather than treating each event as a single node in a graph, each unique technique becomes a single node in the graph instead. This drastically reduces the number of nodes visualized.

This reduction significantly impacts our dataset. The dataset contains multiple machines with the largest number of events for a single machine being $188,697$. This is the maximum number of nodes in a graph if each event is its own node. By mapping events to techniques, we reduce this number to 227, the number of techniques in the MITRE ATT&CK Framework without 'Resource Development'. The number of unique events in our mapping is 145 (as is shown in Appendix C), which brings the maximum number of nodes down to 145. However, since multiple events are mapped to the same technique, our dataset's maximum number of nodes is 24. Meaning there is a reduction of 99.987% compared to the initial maximum of $188,697$. To further reduce the number of nodes, the graph generation also only uses techniques and no sub-techniques. This removes clutter from the graph while still giving a proper overview of what techniques occurred on the machine. Even though we reduce the number of nodes shown in the graph, no data is lost since each node contains all the events that belong to that technique.

The graph's edges are directional and represent viable steps between techniques. The viability of a step is based on two factors: the technique's phase and the chronological order of its events. Nodes are connected through an edge when the following conditions are met:

(i) The phase of node B is later than that of node A. Since the phases are set up sequentially, the graph preserves this order.

(ii) There is at least 1 event in node B that chronologically occurs after the first event of node A. If the technique does not occur chronologically after another technique, then connecting these nodes is unnecessary.

(iii) There is no node C that satisfies both these criteria and has a phase that occurs between the phases of node A and B. In other words, node C must not be a viable intermediary step between nodes A and B in terms of both phase and chronological ordering. If such a node C exists, it is used instead.

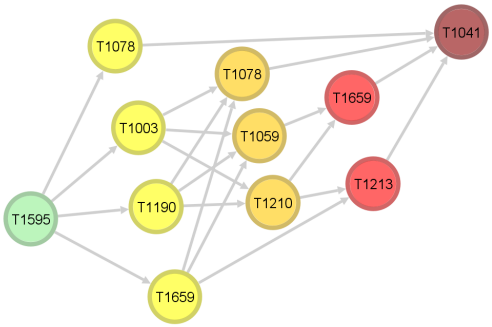

Figure 2. Example of the graphical representation of the potential attack paths found in the security events. Nodes represent the MITRE ATT&CK Techniques performed on the victim machine, and edges represent the potential attack paths between those techniques. Nodes are color-coded based on their phase: Phase 1 - Green, Phase 2 - Yellow, Phase 3 - Orange, Phase 4 - Red, Phase 5 - Dark Red.

The merging of events using the MITRE ATT&CK Framework comes with a problem of connecting nodes. When node A consists of multiple events, there is a chance that some events warrant a connection to node B while others do not. In this scenario, nodes A and B are still connected. This in itself is not a problem during the path traversal since the path traversal uses the graph in combination with DeepCASE. When the path traversal, using DeepCASE, does not detect a path from node A to node B, but this path is present in the attack graph, the path won't be used (see Section 5.3). The problem arises when connections that should be made according to path traversal with DeepCASE are not present in the attack graph. When this occurs, no attack path will be found. One scenario where this could happen is when some events in node A should connect to node B while others should connect to node C, where node C can be placed as an intermediary node between nodes A and B. According to criteria (iii), nodes A and B are not connected with an edge because node C is placed in between them. This affects the path traversal since the nodes are not connected in the graph. DeepCASE might come across events that should be added to the path, but since they are not connected, it will disregard them (see Section 5.3). In Section 6.4, the effect on the generated paths is tested when nodes are connected, even when intermediate nodes are found.

Combining all this, the graph consists of nodes and edges between nodes. Figure 2 shows such a graph. Where the nodes represent techniques used on the target machine and the edges show viable steps between the techniques. The edges are based on the related tactics for a given technique and the chronological ordering of events. Algorithm 1 shows the basic graph generation algorithm.

The graph generation works as follows: Given a list of events (X) two empty sets are initialized for the edges (E) and the techniques/nodes (N). The algorithm

Table 1. MITRE ATT&CK Tactics placed into phases

Phase 1	Phase 2	Phase 3	Phase 4	Phase 5
Reconnaissance	Initial Access	Execution	Command and Control	Exfiltration
Discovery	Credential Access	Persistence	Collection	Impact
		Privilege Escalation		
		Defense Evasion		
		Lateral Movement		

Algorithm 1 Basic graph generation for a list of events X where each event is denoted by x, returns a list of nodes N and a list of edges E

```
1:  procedure GENERATEGRAPH(X)
2:      E ← ∅
3:      N ← ∅
4:      for all x ∈ X do
5:          if x_t ∉ N then
6:              Create N[x_t]
7:          Append x to N[x_t]
8:      for all a ∈ N do
9:          for all b ∈ N do
10:             if a_start < b_end then
11:                 if a_phase < b_phase then
12:                     Append (a, b) to E
13:     for all (a, b) ∈ E do
14:         for all c ∈ N do
15:             if (a, c) ∈ E and (c, b) ∈ E then
16:                 Remove (a, b) from E
17:     return E, N
```

then iterates over every event in the event list. If the technique x_t for that event already exists in a node, it adds the event to that technique/node. If it is a new technique, it creates a new node and adds the event to it. After this, for each combination of nodes, the following conditions are checked: Is the start time of node A (a_{start}) before the end time of node B (b_{end}), condition (ii); Is the phase if node A (a_{phase}) is earlier than the phase of node B (b_{phase}), condition (i); If both conditions are met, the edge from node A to node B is added to the list of edges. The last for-loop is to assure condition (iii) by removing all connections between node A and node B if a node C is found that connects both nodes. It uses the transitive reduction algorithm for directed graphs proposed by Aho et al. [62]. After this, the list of techniques and the list of edges are returned.

As mentioned before, a technique can be part of multiple tactics. In these cases, these tactics can fall into different phases, resulting in a technique that can be placed at different points in the graph depending on the tactic used. In this scenario, the technique can be added to the graph once for each phase it belongs to. Algorithm 1 shows the base scenario where this is not happening, and each technique is only added once.

5.3 Path Traversal

The path traversal aims to identify the most probable route within the graph according to DeepCASE. Utilizing the graph provides a structured framework for path analysis. As discussed in Section 5.1, a certain order of tactics is expected. Deviations from this expected order, such as an 'Execution' technique preceding 'Initial Access', signify that one of the techniques is not part of the current attack path. Hence, integrating the graph generated by Algorithm 1 equips the path traversal to follow the expected sequence more closely.

DeepCASE's attention mechanism shows the importance of an event in the context in relation to the event directly after the context. The attention is used to indicate whether two nodes are likely connected (high attention value) or not (low attention value). By traversing events based on attention within DeepCASE sequences, we reconstruct the most likely path an attacker exploited according to DeepCASE.

If we were to follow DeepCASE's highest attention one-to-one without using the attack graph generated earlier, the resulting path would jump back and forth between nodes frequently and create side branches to nodes that are not part of the main attack chain. This is because events highlighted by DeepCASE are not always part of the same attack chain. They may be attributed to other attacking attempts by the same attacker, or they can even be caused by a different attacker. DeepCASE can still flag these events as important if its prediction improves by doing so. This does not imply that the two events are linked but instead that these events from the context are, according to DeepCASE, expected to precede the current event. However, since the graph with potential attack paths is already made during the graph generation, DeepCASE can follow that instead and combine DeepCASE's predictions with the knowledge from the MITRE ATT&CK Framework. It's from this logic that we arrive at the path traversal.

The path traversal starts at the event that is selected by the security operator. The node that this event belongs to is selected as the endpoint for the attack path. The path traversal works backward to the start of the path. Using DeepCASE, it looks at the context of the selected event and looks for a parent node of the initially selected node. Since multiple events belonging to one of the parent nodes may be present in the context, we only look at events in the context with an attention value higher than the threshold $\delta = 0.1$. This threshold used on line 8 of Algorithm 2 receives its value based on the length of the context. If a parent node is found with an attention value lower than the threshold, it is considered unimportant and thus disregarded.

$$\delta = \frac{1}{\text{Context length}}$$

All attention values in the context added together always add up to exactly 1. By dividing the total attention by the number of events in the context, the threshold is set to the value each event would have if they all had the same attention value. Any event with a higher attention value than that is considered important for the path traversal. The default context length of 10, set by DeepCASE, results in the threshold value $\delta = 0.1$.

If the path traversal, using DeepCASE, finds the

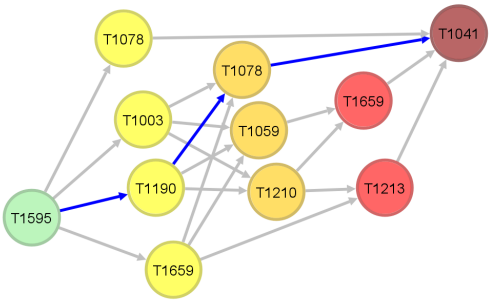

Figure 3. Example of the graphical representation of the potential attack paths found in the security events and an ideal attack path that might be found by the path traversal. Nodes represent the MITRE ATT&CK Techniques performed on the victim machine, grey edges represent the potential attack paths between those techniques, and blue edges represent the attack path found during path traversal. Nodes are color-coded based on their phase: Phase 1 - Green, Phase 2 - Yellow, Phase 3 - Orange, Phase 4 - Red, Phase 5 - Dark Red.

parent node in the context, it traverses to this event and repeats the process. If more than one parent node is found, the node that contains the event with the highest attention is selected. The path traversal will follow the path of the parent node with the highest attention value until in can no longer continue. Because there is no ground truth to guide this process it is unclear what the actual starting point of the attack chain is. Therefor, paths produced using this method may still not show the entire attack chain but a shorter, incomplete path.

If the path traversal cannot find any parent nodes or the parent node's attention value is lower than the threshold, it looks for an event in the context that belongs to the current node and selects the one with the highest attention. After selecting this new event, it tries to look for a parent node again in the new context. It repeats this process until it either reaches the first event of the machine or the next node in the path is found. The path traversal also ends if neither the parent node nor the same node is found. Figure 3 shows an example attack path for the graph in Figure 2. While the grey edges represent the potential attack paths between techniques, the blue edges represent the attack path found during the path traversal with DeepCASE. This path is an ideal scenario where the path covers the entire length of the path. The process that forms an attack path is formalized in Algorithm 2.

The algorithm makes use of a while loop to ensures the algorithm ends when all events on a machine are inspected. This means that the maximum number of iterations for this function is the length of the event list that was used to create the base graph G in Algorithm 1. The algorithm also uses recursion but only calls itself when the next node in the path is found. Since the maximum length of the path is based on the number of phases, the maximum recursion depth is limited to 5. In Section 6.2, the effect of additional edges between phases of the same phase is tested. This affects the maximum recursion depth since it increases the maximum length of the path, which is also discussed in the same section.

Sections 5.1 and 5.2 mention the possibility of techniques falling into two tactics and thus in two phases simultaneously. During the graph generation, we mention the option to add these techniques multiple times to the graph, one for each phase it is part of. For simplicity, we do not show this modification in Algorithms 1 and 2. But if one were to do so, the path traversal must account for the change. Currently, only one node is returned when retrieving the node to which an alert belongs. If the technique is added multiple times, this part of the algorithm would return more than one node. As a result, the path traversal needs to run for each node, which adds additional complexity to the algorithm.

Algorithm 2 Traversal of most likely path given a base graph G, an event e that includes the context e_c and attention e_a given by DeepCASE. Returns the path in the form of a list of edges P

procedure TRAVERSAL($G, e, P \leftarrow \varnothing$)
  $N \leftarrow n | n \in G.nodes, e \in n$
  while e **do**
    $I \leftarrow \{i | i \in e_c\}.sort()$
    $C \leftarrow getChildren(N, G)$
    $S \leftarrow \varnothing$
    for $i \in I$ **do**
      if $e_a[i] < \delta$ **then**
        break
      if $S = \varnothing$ **and** $e_c[i] \in N$ **then**
        $S \leftarrow e_c[i]$
        continue
      $N_2 \leftarrow n | n \in G.nodes, e_c[i] \in n$
      if $N_2 \in C$ **then**
        Append (N_2, N) to P
        $e \leftarrow e_c[i]$
        $P \leftarrow Traversal(G, e, P)$
        return P
    $e \leftarrow S$
  return P

6. EVALUATION

The evaluation aims is to gather information on how the most likely paths generated during the path traversal change depending on these choices and assumptions and, through this, evaluate DeepCASE's capability to analyze entire attack paths using our approach. To achieve this, we established an experiment pipeline:

(i) **Graph Creation:** For each machine in the dataset, we create a graph representing security events. The goal is to capture a comprehensive picture of all security events and their connections on a given machine.

(ii) **Phase Identification:** The latest phase is identified for each graph. This phase represents the most advanced stage of the attack, as defined by the MITRE ATT&CK Framework and the created phases.

(iii) **Path Generation:** For each event in the latest phase, the most likely path through the graph was determined using the path traversal.

(iv) **Dataset Utilization:** To gain a deeper understanding of our proposed approach, we execute several experiments that evaluate different aspects of our approach. During each experiment, we use the same data, the Open-source Collegiate Penetration Testing Competition dataset (CPTC-2018), to ensure consistency across experiments. This setup allows for comparisons using a paired t-test.

(v) **Statistical Export:** Each generated path is fully exported alongside the graph it was generated in. In-

formation such as the path length and completion rate can be gathered from this.

Each experiment is compared with a baseline. This baseline is the pipeline run with all assumptions and choices as described in the methodology. For this comparison, we use the completion rate value. This value is calculated by comparing the length of the found path with the expected path length to see how much of this expected path was completed. Since the dataset does not contain a ground truth for the attack paths, the expected path length is based on the found path. We calculate the distance from the start of the found path to the earliest phase of the graph that can be reached from that point and state that this distance is the missing part of the path. If the found path includes the earliest phase, we state that the path is complete and has a completion rate of 1.0. If the path was not completed, we use the fraction of the expected path.

$$\text{completion rate} = \frac{\text{path length}}{\text{path length} + \text{missing path length}}$$

For each experiment, the completion rate is compared using a paired t-test. The paired t-test calculates the difference for each pair of observations and then assesses whether the average of these differences is significantly different from zero. Each test uses the same hypothesis: the null hypothesis states that the mean difference between the two tests is 0, and the alternative hypothesis states that the mean difference is not equal to 0. This hypothesis is tested with $\alpha = 0.05$ for every experiment.

$$\begin{aligned} H_0 &: u_d = 0 \\ H_1 &: u_d \neq 0 \end{aligned} \tag{1}$$

To compare with the baseline, several experiments were conducted to evaluate the impact of various assumptions and choices:

- **Lateral Movement Between Phases:** This experiment allows lateral movement between phases, testing if allowing such transitions changes the path traversal significantly. It examines the effect of relaxing the assumption that attacks only move forward through phases.
- **Full Movement Between Phases:** Similar to the lateral movement experiment, this experiment allows full movement between phases, including backward transitions. This assesses the impact of removing phase progression constraints entirely.
- **Direct Connections:** This experiment considers keeping the direct connections between events even if intermediate nodes are found. It evaluates what effect removing the connection between these nodes has versus keeping them in place.
- **Alternative Phase Mapping:** This test uses an alternative mapping of tactics to phases to see how sensitive the results are to the initial phase categorization. It helps determine if different phase mappings lead to significantly different path completions.
- **Alternative Technique Mapping:** This experiment involves re-mapping events to different MITRE ATT&CK techniques to see if the choice of technique mapping significantly affects the outcomes. It evaluates the dependency of results on the accuracy and method of technique mapping.

In addition to these experiments, we will also compare the attack graphs generated during the graph generation with the attack graphs generated by SAGE [54].

6.1 Baseline

The results of the tests are shown in Table 2. The baseline has an average completion rate of $\bar{X} = 0.217$ with a standard deviation of $\sigma = 0.319$. This means that our approach, on average, only completes around 22% of the expected path. As a result, the paths generated during the baseline experiment do not reach far into the graph and thus do not give an overview of the entire attack path.

Another observation we can make from the graphs generated during the baseline is that when a machine has more events, it also creates more connections between phases to the point where, with a sufficient number of events, the phases become fully interconnected. This means that every technique in a phase connects to every technique in the next phase, making the creation of connections redundant. However, for machines with fewer events, the graphs show meaningful connections between phases and sometimes skip a phase by connecting to a later phase. This shows that the graph generation is more effective with fewer events.

From the baseline experiment, we can conclude that the current approach is insufficient for analyzing entire attack paths. Especially when a machine has a lot of events.

6.2 Lateral Movement Between Phases

For this experiment, the assumption that phases are only connected sequentially to later phases is changed to allow for connection between nodes of the same phase. In Algorithm 1, line 11 is changed. The line currently states that an edge is only added to the edge list G if the phase of node B is later than the phase of node A. This is changed to allow nodes of the same phase to pass this criteria. As a result of this change, another check had to be added to ensure that node A and node B were not identical. After these changes, line 11 looks like this:

if $a_{phase} \leq b_{phase}$ **and** $a \neq b$ **then**

This change alone can result in the path traversal going back and forth a lot between the same nodes, similar to what it does when it follows the DeepCASE attention values one-to-one as was explained in 5.3. As a result, the maximum recursion depth is no longer 5, but is instead equal to the number of traversed events. To combat this, we only allow edges to be in the graph once. This still increases the maximum recursion depth but limits it to the number of edges in the graph. Limiting edges to only be in the path once is done by changing line 14 of Algorithm 2 to look as follows:

if $N_2 \in C$ **and** $(N_2, N) \notin P$ **then**

While the test did have a statistically significant effect on the completion rate of the generated paths, this effect in practice is very small. This shows that our assumption that attackers always advance their attack is incorrect. Attackers will use multiple techniques in the same phase to gain additional benefits before advancing to the next attack phase.

6.3 Full Movement Between Phases

In this experiment we test the effect on the attack graphs when full movement occurs between phases. Meaning

Table 2. Completion rate results for baseline and individual experiments. The T-test and p-value are relative to the baseline

Experiment	Avg.	Std.	T-test	p-value
Baseline	0.217	0.319	-	-
Lateral Movement Between Phases	0.222	0.323	2.800	0.005
Full Movement Between Phases	0.111	0.314	65.585	0.000
Direct Connections	0.938	0.280	464.083	0.000
Alternative Phase Mapping Random	0.394	0.473	82.492	0.000
Alternative Phase Mapping Expert	0.215	0.321	1.356	0.175
Alternative Technique Mapping	0.563	0.486	157.993	0.000

**NOTE: All values are rounded to three decimal places*

that connections are made to nodes of the same phase, a later phase, and an earlier phase. Line 11 of Algorithm 1 is removed to allow for nodes of any phase to be connected. In turn, a check has to be added to ensure nodes cannot create edges to themselves, similar to what was done in Section 6.2. The line then looks like this:

$$\textbf{if } a \neq b \textbf{ then}$$

This experiment also affects maximum recursion depth, similar to the lateral movement between phases in Section 6.2. For this reason, the same addition to Algorithm 2 is made so the path consists of only unique edges.

Whereas the lateral movement between phases from Section 6.2 had a statistically significant but small, positive effect on the completion rate. We see that removing the restriction between phases altogether has a statistically significant effect that results in a worse completion rate on average. Indicating that using the phases as a criterion to create connections between nodes positively affects the generated paths.

6.4 Direct Connections

One of the criteria for edge formation is that no connection between nodes is made if an indirect path can be found using other nodes that fit between the two initial nodes. For this experiment, we tested the effect of this criteria on the generated attack paths. In Algorithm 1, lines 13 - 16 perform a transitive reduction. Removing these lines from the algorithm will keep the directly connected edges in the graph.

This is expected to allow the attack path to skip multiple nodes and reach the start of the graph faster. Thus, it will have shorter attack paths, but the completion rate should be higher. The results of this experiment are shown in Table 2.

The results of this experiment show a statistically significant increase in the average completion rate. This change is not surprising, as the availability of direct connections between nodes, while reducing the found attack paths, also allows them to traverse further into the graph and thus have a higher completion rate.

The direct path gives a simplified version of the attack path that often misses many critical intermediate steps. A detailed inspection of the result shows the creation of unrealistic paths, such as an active scanning technique directly linked to an exfiltration technique without the intermediate steps that show how access was obtained. Given this, it is important to balance the use of direct connections. While they can improve the efficiency and completion rate of the generated paths, care must be taken to ensure that intermediate steps are not omitted. This raises the question of whether the completion rate is a good metric to evaluate our approach. Based on the evaluation, we can say that the completion rate performs well when showing that attack paths do not show an entire attack chain. However, on its own, it cannot show that the found attack paths encompass an entire attack. Other metrics, such as the path length, need to be used in tandem with the completion rate to see if the found path shows the entire attack chain.

6.5 Alternative Phases Mapping

In Section 5.1, we mention the importance of the mapping from events to techniques and the subsequent mapping of tactics to different phases. In this experiment, we evaluate the effect of the mapping from tactics to phases by rerunning the baseline experiment with two different mappings.

The first mapping is formed by having an expert evaluate the original phases and implementing the feedback given through this process. This new mapping is shown in Table 3. The eventual mapping was split up into 6 phases instead of 5 to better show the dependencies between tactics.

The second mapping is randomly assigned. Here, we made the assumption that there were 5 phases before randomly assigning each tactic to one of the phases. The amount of phases were selected because this is the same number of phases as in the baseline, which would make the two mappings more easily comparable. The random mapping can be seen in Table 4.

In both new mappings, the events in the latest phase differ from the baseline experiment which makes a paired t-test is no longer possible. Instead, the completion rates are compared with an independent t-test. The tests use the following hypothesis: the null hypothesis states that the mean of the sets is equal, and the alternative hypothesis states that the means are not equal in either direction. This hypothesis is tested with $\alpha = 0.05$.

$$\begin{aligned} H_0 &: u_1 = u_2 \\ H_1 &: u_1 \neq u_2 \end{aligned} \tag{2}$$

The first mapping, created based on expert feedback, does not show significant changes to the completion rate compared to the baseline. This means that while the paths created change depending on the tactic mapping, changing this mapping does not significantly improve the created paths. One reason for this could be that splitting the MITRE Techniques into phases is not the correct approach. while some tactics depended on each other, such as "Initial Access" occurring before "Execution", other tactics did not have such a dependency. An example here would be how "Persistence" does not depend on "Lateral Movement" even though both occur after "Execution".

That said, we see how using a random mapping of tactics has a significant impact on the created attack

Table 3. Alternative mapping of MITRE ATT&CK Tactics to phases for the evaluation experiment. Based on expert feedback.

Phase 1	Phase 2	Phase 3	Phase 4	Phase 5	Phase 6
Reconnaissance Discovery	Initial Access	Credential Access Execution	Discovery	Persistence Lateral Movement Privilege Escalation Defense Evasion Impact Collection Command and Control	Exfiltration

paths. The completion rate appears to improve, but as stated in Section 6.4, this does not mean the attack paths can now show more of the attack chain. The length of attack paths generated using this method does not exceed 2, and most of the found paths are of length 1. The sample size for the random phases is also much smaller due to the latest phase, phase 5, containing different events. Overall we can say that while the completion rate is impacted positively and the effect on the found paths is drastic, the found paths do not improve.

6.6 Expert Opinion on MITRE ATT&CK Mapping

To enrich our evaluation, we sought expert opinions on the mapping from security events to MITRE ATT&CK techniques. We consulted a cybersecurity professional. The insights gained from this resulted in a new technique mapping, which is found in Appendix D. This new mapping was then used to rerun the baseline experiment in order to view the effect of this mapping compared to the old.

Since the mapping is different, this experiment can also not make use of the paired t-test. For this reason, this experiment uses the same test and hypothesis as in Section 6.5, Hypothesis 2. The results indicate a significant improvement in the average completion rate. This demonstrates that the mapping of MITRE ATT&CK Techniques is crucial in generating paths, to the extent that even minor changes can lead to major improvements.

The improvement could be attributed to the path traversal's ability to find the next node more accurately when the improved technique mapping allows it to find nodes previously discarded as unimportant or the new technique mapping aligns more accurately with the phase mapping and DeepCASE's attention values. This results in paths that show more of an attack chain. A perfect mapping will not necessarily create perfect attack paths if it does not match with the created phase mapping and DeepCASE's attention values. However, we can still see the significant role of technique mapping in path traversal due to the significant impact minor changes can have.

6.7 Comparison of Graph Generation with SAGE

The graph generation proposed in the methodology and the graph generation done by SAGE are different in how they construct the attack graphs, but the overall goal of showing attack paths within the security events is the same. For this reason, we will compare the two graphs. In order to be able to compare the two graphs, the MITRE ATT&CK Technique mapping shown in Appendix C will be applied to the SAGE graphs. After this, the similarity of the two graphs will be calculated using NetworkX's graph edit distance function. This function calculates the steps required to go from one graph to another. The lower this distance is, the more similar the two graphs are. Selecting one of the machines in the dataset lets our graph generator create one network graph with multiple potential paths. This graph is shown in Figure 4. We see that due to the large number of events for this machine, the graph becomes fully connected between layers.

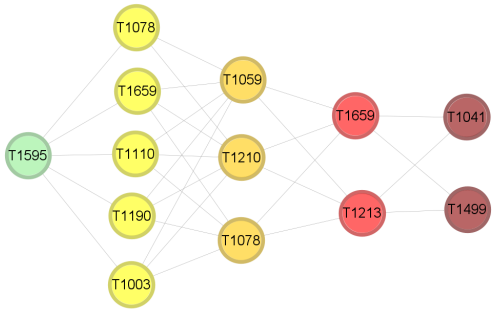

Figure 4. Graph made by the graph generation. Shows the techniques performed for the machine in the dataset with IP-address 10.0.0.22

After the initial application of the technique mapping, the SAGE graph looks like Figure 5. Here, it can be seen that SAGE splits up attack actions into multiple actions even when, according to the technique mapping, the actions are part of the same MITRE ATT&CK Technique. As a result, the graph chains together multiple nodes for the same technique. In doing so, SAGE already shows an attack path for a specific attack, while the graph generated from the graph generator shows an overview of all techniques performed on a machine, after which a path for a specific event can be searched. The graph edit distance between the graph in Figure 4 and Figure 5 is 53 when each operation on a node or edge is counted as 1 step. Since both graphs are relatively small, with 13 and 14 nodes, respectively, this distance of 53 is large. This in itself does not say anything about the quality of either graph but rather that they are not similar.

Since our graph generation does not divide one MITRE ATT&CK Technique into multiple attacker actions, we applied an extra transformation to the graph generated by SAGE. For this transformation, we merged all identical techniques into single nodes, which is similar to how our graph generation works. After this transformation, the new SAGE graph, Figure 6, looks rather different from its original. The graph in Figure 6 already shows an attack path. Noteworthy here is that this attack path can not be found using our graph due to other techniques being placed in between.

This highlights a flaw in our graph generation and shows the advantage of SAGE. By splitting up attacker actions further than the MITRE ATT&CK techniques, it can distinguish paths between specific actions and avoid missing one connection because of another connection.

Table 4. Alternative mapping of MITRE ATT&CK Tactics to phases for the evaluation experiment. Randomly assigned.

Phase 1	Phase 2	Phase 3	Phase 4	Phase 5
Discovery	Privilege Escalation	Reconnaissance	Command and Control	Persistence
Exfiltration	Credential Access	Initial Access		Defense Evasion
	Collection	Execution		Impact
		Lateral Movement		

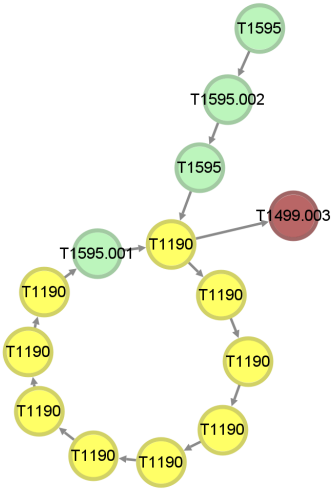

Figure 5. Graph generated by SAGE for the machine in the dataset with IP-address 10.0.0.22 with the MITRE ATT&CK Technique mapping from Appendix C applied to it.

SAGE splits up its graphs into different attacks and creates a unique graph for each attack. While our graph generation does give an overview of all techniques performed on a machine, due to the simplistic heuristics that create the connections between these techniques, connections that should be made can be lost. Because of this, we can say that the graphs generated by SAGE are superior to ours in terms of displaying attack paths due to fewer lost connections.

A potential alternative would be to use SAGE's graph generation instead of our graph generation, though this raises the question of whether DeepCASE is still needed in this process. This is further discussed in Section 7.2.

7. DISCUSSION

One question that arises through this study is the viability of DeepCASE in explaining complete attack paths. Our approach of traversing an attack graph using DeepCASE's attention mechanism indicates that DeepCASE cannot reliably analyze the attack path of an entire attack chain. Several aspects of our approach were effective: Particularly when fewer alerts were present, graph generation could show the connections between different techniques performed on the machine, and the paths displayed in these graphs could show proper attack chains. This shows that, given the right heuristic rule set, a similar working graph generation could work; Secondly, the path traversal was often unable to find a proper path. However, when it did show an attack path, it provided a clear chain of events for analysts to follow. This shows that the concept of displaying the events on the victim machine in the form of an attack chain has a lot of potential; Lastly, the technique mapping was crucial for graph generation and the subsequent path traversal. The evaluation Section 6.6 shows how important an accurate mapping is to these processes and how even slight changes can significantly affect the resulting graphs and attack paths.

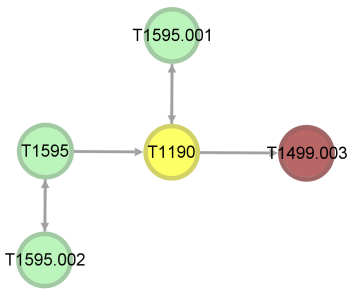

Figure 6. The SAGE graph from Figure 5 where each technique is contained in a single node.

However, there were several ineffective points: The path traversal did not work as effectively as initially hoped. This can have a few causes. Firstly, the method or algorithm could be incorrect, resulting in improper paths. Secondly, while DeepCASE's context sequences and attention system appear suitable for this task, it might be that their design inherently fails to capture larger attack patterns, which is why attack chains can never be properly mapped using these features. Lastly, the heterogeneous and repetitive nature of the data that was used during the study could be a misrepresentation of reality, which is unable to work well with a framework that is supposed to describe actual attacker behavior like the MITRE ATT&CKframework; Another particularly ineffective part was the graph generation when too many alerts were present for a single machine. In the dataset, one particular machine contains over 188 thousand events. The graph generated for this machine cannot show the complex relationships between techniques but instead displays a fully connected graph between layers. In doing so, the graph generation loses some of its effectiveness for the later path traversal. Categorizing MITRE ATT&CK Tactics into different phases also does not represent the dependencies between different tactics as effectively initially predicted. Showing these dependencies with the help of a dependency graph might be able to show these relationships more effectively. Furthermore, the completion rate metric used to show the effect of different assumptions was very effective in indicating when the path generation did not work. However, it was not enough on its own to show when paths did work since the generated paths could get a high completion rate while still being inaccurate. A ground truth or other, more reliable, metric should be considered.

7.1 Limitations

While this study provides valuable insights into the explainability of DeepCASE using knowledge frameworks such as MITRE ATT&CK, it is important to acknowl-

edge its limitations and consider how the process could be improved in the future.

7.1.1 Interview responses

A key limitation of this study is the small sample size of the preliminary interviews, with only five participants, making it difficult to draw strong conclusions.

More participants would likely reduce assumptions, as additional insights would improve the reliability and validity of the findings. For instance, while the interviews highlighted the prevalence of certain event types, a larger sample could have provided broader insights, potentially challenging assumptions like the sequential progression of attacks, which our evaluation later disproved.

The limited sample size also weakens the robustness of verified assumptions, as the small group may not represent the broader population, increasing the risk of bias. Future studies should prioritize recruiting a larger, more diverse participant pool to capture a wider range of perspectives and draw stronger, more nuanced conclusions.

7.1.2 Evaluation method

Another limitation was the evaluation method used. Due to the lack of ground truth and the limited time available, the evaluation was limited to assessing the assumptions made in the study. Ideally, a more comprehensive evaluation method, such as an A/B test with participants, would have been employed alongside the evaluation of assumptions. This A/B test would compare the analysts' performance with and without the kill chains generated during this study.

The practical effect of employing such an A/B test would be significant. It would allow us to directly measure the methodology's actual effectiveness in achieving its initial goal of assisting analysts. For example, by comparing performance metrics such as accuracy, speed, and confidence in threat detection and mitigation, we could have quantitatively demonstrated the benefits (or lack thereof) of using the generated kill chains. This would provide more robust and actionable insights instead of relying solely on theoretical assumptions.

Without this comprehensive evaluation, our findings remain speculative and are based on inferred benefits rather than empirically tested results.

7.1.3 Ground Truth Dataset

Another limitation of this study is the absence of a ground-truth dataset. Without a dataset where the actual kill chains are known, it is impossible to evaluate the accuracy of the generated kill chains. As a result, the effectiveness of the kill chains has to be approximated using the completion rate.

A ground truth dataset would allow us to compare the actual kill chains and the generated ones directly. This comparison would help ensure that the generated kill chains are as close to the true sequences of events as possible. Such a dataset would provide a benchmark against which our methodology could be tested, offering a clear measure of its accuracy and reliability.

However, acquiring a ground truth dataset is difficult. It requires detailed and verified information on the complete sequences of events in various attacks. This dataset type is not readily available and would likely need to be specifically created for this purpose. Developing this dataset would involve extensive collaboration with cybersecurity experts.

7.2 Future Work

Section 6.7, SAGE is compared to the current graph generation. While fundamentally different in their approach, both attempt to do the same: show possible attack paths within the security events. Where SAGE differs greatly is that it shows the attack path for one attack rather than an overview of all attacks. In doing so, SAGE can show attack paths that the current graph generation cannot find. Future work could explore using SAGE as a validation tool for the current approach. By validating and refining the graph generation and path traversal, SAGE can help identify discrepancies, ensuring a more accurate and comprehensive analysis. This integration would strengthen the overall methodology without replacing it.

During the evaluation of mapping MITRE ATT&CK Tactics to phases, it was observed that while some tactics, like "Initial Access" and "Execution," depend on each other, others do not. This makes using phases challenging for graph generation. A potential future approach is to implement a dependency graph, allowing tactics with dependencies to connect while others remain independent. Though this adds complexity, it could enable the generation of attack paths that were previously undetectable, aligning more closely with the paths identified by SAGE in Section 6.7.

8. CONCLUSION

The primary goal of this study was to enhance the explainability of DeepCASE by using attack paths to provide additional context to security analysts. This required determining whether DeepCASE could successfully identify these attack paths. To achieve this, we integrated the MITRE ATT&CK Framework into the DeepCASE system in an attempt to create detailed overviews of performed attacks. By using attack graphs and the DeepCASE context sequences to cover entire attack chains, the research aimed to offer a method to visualize these attack chains.

The approach involved mapping security events to the MITRE ATT&CK Framework and splitting the tactics from the MITRE ATT&CK Framework up into phases. Using the mapping and the phases, generate graphs that display potential attack paths within the performed techniques. Then, with the help of DeepCASE's attention system, attempt to trace back the most likely attack path to show an attack chain to the security analyst.

The study's goal of providing an overview of attack scenarios for security analysts using DeepCASE and the MITRE ATT&CK Framework was partially achieved. The current implementation demonstrated some functionalities that worked well, such as technique mapping and initial graph generation. However, several critical aspects, like path traversal, phase categorization, and the reliability of the completion rate, were less successful. In summary, while integrating the MITRE ATT&CK Framework into DeepCASE shows promise, significant refinements and adjustments are necessary for the system to be fully effective and useful for security professionals. The findings highlight areas for improvement that could make future iterations of the methodology more robust and beneficial.

REFERENCES

1. M. Rosso, M. Campobasso, G. Gankhuyag, and L. Allodi, "Saibersoc: Synthetic attack injection to benchmark and evaluate the performance of security operation centers," in *Annual Computer Security Applications Conference*, ser. ACSAC '20. New York, NY, USA: Association for Computing Machinery, 2020, p. 141–153. [Online]. Available: https://doi.org/10.1145/3427228.3427233
2. M. A. Inam, Y. Chen, A. Goyal, J. Liu, J. Mink, N. Michael, S. Gaur, A. Bates, and W. U. Hassan, "Sok: History is a vast early warning system: Auditing the provenance of system intrusions," in *2023 IEEE Symposium on Security and Privacy (SP)*, 2023, pp. 2620–2638.
3. T. v. Ede, H. Aghakhani, N. Spahn, R. Bortolameotti, M. Cova, A. Continella, M. v. Steen, A. Peter, C. Kruegel, and G. Vigna, "Deepcase: Semi-supervised contextual analysis of security events," in *2022 IEEE Symposium on Security and Privacy (SP)*, 2022, pp. 522–539.
4. A. Nadeem, D. Vos, C. Cao, L. Pajola, S. Dieck, R. Baumgartner, and S. Verwer, "Sok: Explainable machine learning for computer security applications," in *Proceedings of the 2023 IEEE 8th European Symposium on Security and Privacy (EuroS&P)*, L. O'Conner, Ed. United States: IEEE, 2023, pp. 221–240, green Open Access added to TU Delft Institutional Repository 'You share, we take care!' - Taverne project https://www.openaccess.nl/en/you-share-we-take-care Otherwise as indicated in the copyright section: the publisher is the copyright holder of this work and the author uses the Dutch legislation to make this work public.; 2023 IEEE 8th European Symposium on Security and Privacy (EuroS&P) ; Conference date: 03-07-2023 Through 07-07-2023.
5. MITRE, "Mitre att&ck framework," 2015. [Online]. Available: https://attack.mitre.org/
6. F. B. Kokulu, A. Soneji, T. Bao, Y. Shoshitaishvili, Z. Zhao, A. Doupé, and G.-J. Ahn, "Matched and mismatched socs: A qualitative study on security operations center issues," in *Proceedings of the 2019 ACM SIGSAC Conference on Computer and Communications Security*, ser. CCS '19. New York, NY, USA: Association for Computing Machinery, 2019, p. 1955–1970. [Online]. Available: https://doi.org/10.1145/3319535.3354239
7. D. Kelley and R. Moritz, "Best practices for building a security operations center," *Information Systems Security*, vol. 14, no. 6, pp. 27–32, 2006.
8. S. Bhatt, P. K. Manadhata, and L. Zomlot, "The operational role of security information and event management systems," *IEEE security & Privacy*, vol. 12, no. 5, pp. 35–41, 2014.
9. A. Shah, R. Ganesan, S. Jajodia, and H. Cam, "Understanding tradeoffs between throughput, quality, and cost of alert analysis in a csoc," *IEEE Transactions on Information Forensics and Security*, vol. 14, no. 5, pp. 1155–1170, 2018.
10. CISCO, "Anticipating the unknowns - chief information security officer (ciso) benchmark study. technical report," 2019. [Online]. Available: https://ebooks.cisco.com/story/anticipating-unknowns/
11. D. Bahdanau, K. Cho, and Y. Bengio, "Neural machine translation by jointly learning to align and translate," *arXiv preprint arXiv:1409.0473*, 2014.
12. T. Brown, B. Mann, N. Ryder, M. Subbiah, J. D. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell *et al.*, "Language models are few-shot learners," *Advances in neural information processing systems*, vol. 33, pp. 1877–1901, 2020.
13. J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding," *arXiv preprint arXiv:1810.04805*, 2018.
14. A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," *Advances in neural information processing systems*, vol. 30, 2017.
15. M. Du, F. Li, G. Zheng, and V. Srikumar, "Deeplog: Anomaly detection and diagnosis from system logs through deep learning," in *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security*, ser. CCS '17. New York, NY, USA: Association for Computing Machinery, 2017, p. 1285–1298. [Online]. Available: https://doi.org/10.1145/3133956.3134015
16. Y. Shen, E. Mariconti, P. A. Vervier, and G. Stringhini, "Tiresias: Predicting security events through deep learning," in *Proceedings of the 2018 ACM SIGSAC Conference on Computer and Communications Security*, ser. CCS '18. New York, NY, USA: Association for Computing Machinery, 2018, p. 592–605. [Online]. Available: https://doi.org/10.1145/3243734.3243811
17. Y. Shen and G. Stringhini, "{ATTACK2VEC}: Leveraging temporal word embeddings to understand the evolution of cyberattacks," in *28th USENIX Security Symposium (USENIX Security 19)*, 2019, pp. 905–921.
18. M. Ester, H.-P. Kriegel, J. Sander, X. Xu *et al.*, "A density-based algorithm for discovering clusters in large spatial databases with noise," in *kdd*, vol. 96, no. 34, 1996, pp. 226–231.
19. T. Miller, "Explanation in artificial intelligence: Insights from the social sciences," *Artificial intelligence*, vol. 267, pp. 1–38, 2019.
20. A. Adadi and M. Berrada, "Peeking inside the black-box: a survey on explainable artificial intelligence (xai)," *IEEE access*, vol. 6, pp. 52 138–52 160, 2018.
21. B. Letham, C. Rudin, T. H. McCormick, and D. Madigan, "Interpretable classifiers using rules and bayesian analysis: Building a better stroke prediction model," 2015.
22. R. Caruana, Y. Lou, J. Gehrke, P. Koch, M. Sturm, and N. Elhadad, "Intelligible models for healthcare: Predicting pneumonia risk and hospital 30-day readmission," in *Proceedings of the 21th ACM SIGKDD international conference on knowledge discovery and data mining*, 2015, pp. 1721–1730.
23. K. Xu, J. Ba, R. Kiros, K. Cho, A. Courville, R. Salakhudinov, R. Zemel, and Y. Bengio, "Show, attend and tell: Neural image caption generation with visual attention," in *International conference on machine learning*. PMLR, 2015, pp. 2048–2057.
24. B. Ustun and C. Rudin, "Supersparse linear integer models for optimized medical scoring systems," *Machine Learning*, vol. 102, pp. 349–391, 2016.
25. S. Sarkar, T. Weyde, A. d. Garcez, G. G. Slabaugh, S. Dragicevic, and C. Percy, "Accuracy and interpretability trade-offs in machine learning applied to safer gambling," in *CEUR Workshop Proceedings*, vol. 1773. CEUR Workshop Proceedings, 2016.

26. L. Breiman, "Statistical modeling: The two cultures (with comments and a rejoinder by the author)," *Statistical science*, vol. 16, no. 3, pp. 199–231, 2001.
27. Z. C. Lipton, "The mythos of model interpretability: In machine learning, the concept of interpretability is both important and slippery." *Queue*, vol. 16, no. 3, pp. 31–57, 2018.
28. S. Krening, B. Harrison, K. M. Feigh, C. L. Isbell, M. Riedl, and A. Thomaz, "Learning from explanations using sentiment and advice in rl," *IEEE Transactions on Cognitive and Developmental Systems*, vol. 9, no. 1, pp. 44–55, 2016.
29. A. Mahendran and A. Vedaldi, "Understanding deep image representations by inverting them," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 5188–5196.
30. T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado, and J. Dean, "Distributed representations of words and phrases and their compositionality," *Advances in neural information processing systems*, vol. 26, 2013.
31. C. Yang, A. Rangarajan, and S. Ranka, "Global model interpretation via recursive partitioning," in *2018 IEEE 20th International Conference on High Performance Computing and Communications; IEEE 16th International Conference on Smart City; IEEE 4th International Conference on Data Science and Systems (HPCC/SmartCity/DSS)*. IEEE, 2018, pp. 1563–1570.
32. M. A. Valenzuela-Escárcega, A. Nagesh, and M. Surdeanu, "Lightly-supervised representation learning with global interpretability," *arXiv preprint arXiv:1805.11545*, 2018.
33. A. Nguyen, A. Dosovitskiy, J. Yosinski, T. Brox, and J. Clune, "Synthesizing the preferred inputs for neurons in neural networks via deep generator networks," *Advances in neural information processing systems*, vol. 29, 2016.
34. D. Erhan, A. Courville, and Y. Bengio, "Understanding representations learned in deep architectures," 2010.
35. M. T. Ribeiro, S. Singh, and C. Guestrin, "" why should i trust you?" explaining the predictions of any classifier," in *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*, 2016, pp. 1135–1144.
36. ——, "Anchors: High-precision model-agnostic explanations," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 32, no. 1, 2018.
37. J. Lei, M. G'Sell, A. Rinaldo, R. J. Tibshirani, and L. Wasserman, "Distribution-free predictive inference for regression," *Journal of the American Statistical Association*, vol. 113, no. 523, pp. 1094–1111, 2018.
38. D. Baehrens, T. Schroeter, S. Harmeling, M. Kawanabe, K. Hansen, and K.-R. Müller, "How to explain individual classification decisions," *The Journal of Machine Learning Research*, vol. 11, pp. 1803–1831, 2010.
39. K. Simonyan, A. Vedaldi, and A. Zisserman, "Deep inside convolutional networks: Visualising image classification models and saliency maps," *arXiv preprint arXiv:1312.6034*, 2013.
40. M. D. Zeiler and R. Fergus, "Visualizing and understanding convolutional networks," in *Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part I 13*. Springer, 2014, pp. 818–833.
41. B. Zhou, A. Khosla, A. Lapedriza, A. Oliva, and A. Torralba, "Learning deep features for discriminative localization," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 2921–2929.
42. M. Sundararajan, A. Taly, and Q. Yan, "Axiomatic attribution for deep networks," in *International conference on machine learning*. PMLR, 2017, pp. 3319–3328.
43. R. Roscher, B. Bohn, M. F. Duarte, and J. Garcke, "Explainable machine learning for scientific insights and discoveries," *Ieee Access*, vol. 8, pp. 42 200–42 216, 2020.
44. A. Nadeem, C. Hammerschmidt, C. H. Gañán, and S. Verwer, "Beyond labeling: Using clustering to build network behavioral profiles of malware families," *Malware analysis using artificial intelligence and deep learning*, pp. 381–409, 2021.
45. Y.-S. Lin, W.-C. Lee, and Z. B. Celik, "What do you see? evaluation of explainable artificial intelligence (xai) interpretability through neural backdoors," in *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*, 2021, pp. 1027–1035.
46. A.-K. Dombrowski, M. Alber, C. Anders, M. Ackermann, K.-R. Müller, and P. Kessel, "Explanations can be manipulated and geometry is to blame," *Advances in neural information processing systems*, vol. 32, 2019.
47. G. Severi, J. Meyer, S. Coull, and A. Oprea, "{Explanation-Guided} backdoor poisoning attacks against malware classifiers," in *30th USENIX security symposium (USENIX security 21)*, 2021, pp. 1487–1504.
48. X. Zhao, W. Zhang, X. Xiao, and B. Y. Lim, "Exploiting explanations for model inversion attacks (2021)," *arXiv preprint arXiv:2104.12669*, 2021.
49. S. Jain and B. C. Wallace, "Attention is not explanation," *arXiv preprint arXiv:1902.10186*, 2019.
50. M. Angelini, N. Prigent, and G. Santucci, "Percival: proactive and reactive attack and response assessment for cyber incidents using visual analytics," in *2015 IEEE Symposium on Visualization for Cyber Security (VizSec)*. IEEE, 2015, pp. 1–8.
51. M. Chu, K. Ingols, R. Lippmann, S. Webster, and S. Boyer, "Visualizing attack graphs, reachability, and trust relationships with navigator," in *Proceedings of the Seventh International Symposium on Visualization for Cyber Security*, 2010, pp. 22–33.
52. S. Jha, O. Sheyner, and J. Wing, "Two formal analyses of attack graphs," in *Proceedings 15th IEEE Computer Security Foundations Workshop. CSFW-15*. IEEE, 2002, pp. 49–63.
53. K. Kaynar, "A taxonomy for attack graph generation and usage in network security," *Journal of Information Security and Applications*, vol. 29, pp. 27–56, 2016.
54. A. Nadeem, S. Verwer, and S. J. Yang, "Sage: Intrusion alert-driven attack graph extractor," in *2021 IEEE Symposium on Visualization for Cyber Security (VizSec)*, 2021, pp. 36–41.
55. M. Pegoraro, M. S. Uysal, and W. M. P. van der Aalst, "Proved: A tool for graph representation and analysis of uncertain event data," in *Application and Theory of Petri Nets and Concurrency*, D. Buchs and J. Carmona, Eds. Cham: Springer International Publishing, 2021, pp. 476–486.
56. T. Murata, "Petri nets: Properties, analysis and ap-

plications," *Proceedings of the IEEE*, vol. 77, no. 4, pp. 541–580, 1989.
57. J. Starink, M. Huisman, and A. Continella, "Understanding, modeling, and measuring inter-process code injection in malware," 2023.
58. M. Christodorescu, S. Jha, and C. Kruegel, "Mining specifications of malicious behavior," in *Proceedings of the the 6th Joint Meeting of the European Software Engineering Conference and the ACM SIGSOFT Symposium on The Foundations of Software Engineering*, ser. ESEC-FSE '07. New York, NY, USA: Association for Computing Machinery, 2007, p. 5–14. [Online]. Available: https://doi.org/10.1145/1287624.1287628
59. L. Martignoni, E. Stinson, M. Fredrikson, S. Jha, and J. C. Mitchell, "A layered architecture for detecting malicious behaviors," in *Recent Advances in Intrusion Detection: 11th International Symposium, RAID 2008, Cambridge, MA, USA, September 15-17, 2008. Proceedings 11*. Springer, 2008, pp. 78–97.
60. C. Knauer, "How contact centres can leave businesses exposed to cybercrime," *Netw. Secur.*, vol. 2019, pp. 6–9, 2019.
61. J. Starink, "Analysis and automated detection of host-based code injection techniques in malware," Master's thesis, University of Twente, 2021.
62. A. V. Aho, M. R. Garey, and J. D. Ullman, "The transitive reduction of a directed graph," *SIAM Journal on Computing*, vol. 1, no. 2, pp. 131–137, 1972. [Online]. Available: https://doi.org/10.1137/0201008

A. INTERVIEW QUESTIONS

Experience

- What is your position/job within the SOC?
- Can you describe your work?
- What kind of SOC do you work in?
- Can you describe your background and experience in cybersecurity?
- Did you receive any training for assessing security events?
  - Can you name the training?

Decision-making-process

- Can you walk me through the steps you take when assessing/triaging a security event?
- What tools do you use during this process?
- What additional information do you look for? (Both external and internal sources)
  - And where do you look for this information?
  - What benefit does this information provide?

Challenges

- What challenges do you face during your work?
  - Why is this a challenge?
  - Do you think there is a solution?
- Any specific type of events that are more challenging than others?
  - What makes these events more challenging?
- Beyond specific event types. Do you encounter any other challenges while analyzing security alerts?
  - Why is this a challenge?
  - Do you think there is a solution?

Contextual information

- With the term 'contextual information', what comes to mind?
- How important is contextual information when assessing a security event?
  - If it's not very important:
    - Is that a limitation of the type of information, or is it something else?
    - How could this information be made more valuable?
  - If it's very important:
    - What makes it so valuable?
- What type of contextual information do you find most valuable?
  - What makes this information valuable?
- Can you share an example of a situation where the availability of contextual information (or lack thereof) influenced the outcome of an analysis?
  - What contextual information was this?
  - How did it impact the analysis?

Kill chain concepts

- How familiar are you with cyber kill chains and the MITRE ATT&CK framework?
- Do you currently use kill chains in the analysis process? If so, how?
- Do you currently use the MITRE ATT&CK framework in the analysis process? If so, how?
- Do the tools mentioned earlier incorporate either of these frameworks?
- Do you use any other frameworks?
- In your opinion, what are the most important parts or stages of a cyber kill chain for your work?
  - In terms of kill chain phases? (recon, intrusion, exploit, etc.)
  - in terms of the concept in general? (How the framework works more broadly)
- In your opinion, what are the most important parts or stages of the MITRE ATT&CK Framework for your work?
  - In terms of MITRE ATT&CK Tactics or Techniques? (Initial Access, Persistence, or individual techniques)
  - in terms of the concept in general? (How the framework works more broadly)

Feedback

- Do you provide feedback on the tools you work with?
  - No, I do not:
    - Do you not want to, or is there no possibility to do so?
  - Yes, I do:
    - Which tools?
    - Which aspects of these tools?
      - (Functionality, UI, Processes, Other?)
- What type of feedback do you generally give?
  - (or would you want to give?)
  - (or would you give if asked for it now?)
- Are there aspects of tools you currently use that you find particularly effective or lacking?
  - Which tools?
  - What aspects?
  - Why those aspects?

B. INTERVIEW ANALYSIS

B1 Investigation Process and Methodology

The investigation process and methodology are critical aspects of cybersecurity analysis. This theme encompasses various codes that outline the systematic approach to investigating security incidents, including understanding the responsibilities, identifying root causes, and gathering detailed information.

This theme includes multiple codes. One of the codes is "Causation", which is about understanding the root cause of an event. Participants here talked about the origin of an event with quotes such as:

"You start by looking at the specific use case that triggered the event and why it was triggered."

"Is this normal malware that can be downloaded on accident, or is this malware that is actively being spread by an APT."

Another code in this theme was "Details", which is about the initial and additional details that are gathered by analysts to investigate security events. Here participants talked about what the detailed information from an event is that they look at:

"Type of device or service and what it is being used for."

"You could look at more specific details of the source and the destination hosts that are involved, like the owner, country of origin, and whether it is known to be used for malicious activity."

And the code "Historical", which are quotes about how an is event triggered in the past. Here participants talked about how events got triggered in the past but also if IP addresses that triggered an event were flagged in the past. On top of that alerts that were triggered for specific users were also discussed here.

"When specific users are involved, you look at other alerts that might have been triggered previously for this user."

"Check whether similar activity has been detected and/or reported before."

"... if there are many reports of an IP address, domain or hash and if those reports are trustworthy."

B2 Contextual Understanding

The contextual understanding theme focuses on the broader context surrounding security incidents, including user behavior, attacker identity, and the available context found in surrounding alerts. This helps analysts to make informed decisions based on the full picture.

This theme includes several codes. One of the codes is "Users", which focuses on the behavior. Participants mentioned how user behavior might explain an alert triggering. Or how an alert is most likely more severe when the users behavior does not explain the alert

"So if an infection is caused due to the stupid behavior of a user then that is less exciting than when there is nothing in the context that says that the user is behaving irresponsibly when the alert occurs."

Another code for this theme is "attacker" and it describes how the identity of the attacker and their current and past actions influence an analysis.

"We question if its an internal attacker or one that is external."

"An attacker could be spread out. Like in the case of a DDoS attack."

And the code "Context", which describes how surrounding alerts can help in the analysis. Here it is mentioned how context or the lack of context can say a lot about an alert or the users. Here participants also mention that the type of information they look for within the context differs depending on the type of alert.

"If you see an outgoing request to a suspicious domain but no network traffic in the context that shows someone is behaving normally that is suspicious."

"Depending on the type of alert we see, the context is different."

"So basically what you're trying to find with contextual information. Using the context, see if other steps have taken place."

B3 Framework and Models

The frameworks and models theme revolves around using established methodologies and guidelines to analyze and respond to security incidents. This includes understanding attack chains and common occurrences,.

One of the codes is "Kill Chain", which relates to the stages or steps in an attack chain. Participants mentioned how understanding the kill chain is assist the analysis and how being able to see the connection between tactics would be helpful:

"It's more that I'm trying to determine in my head which stage of an attack it could be. Is it initial access or exploitation."

"What you would actually want is a system that notifies you in the event of a real attack that a reconnaissance has also taken place in advance. At that moment when you know that reconnaissance has taken place and you now see an attack, then it was really targeted."

Another code is "Overview". In this theme the participants discuss the importance of gaining an overview of the situation which can be a clients current situation or a cyber attack. One of the ways this can be done is via frameworks such as the MITRE ATT&CK Framework.

"In order to know how you can best help a customer make their environment more secure and what you need to monitor for, you first need to know what they are vulnerable for and what tactics and techniques can be used to attack their crown jewels."

Another code is "Assistance", which pertains to using frameworks as a guide during investigations. Interviewees highlighted how these frameworks assist analysts and how it can be especially useful for analyst with less experience:

"We often use MITRE because it helps with knowing what to look for during the investigation."

"Usually handy for tier 1 analysts if they don't know what an alert is about."

The last code for this theme is "Occurrence" and it deals common types of alerts. These were often expressed by participants using the techniques found in the MITRE ATT&CK Framework. Participants mention they often see active scanning, brute forces, and exploitation of public facing applications.

"Scans and brute force are by far the most common."

"You see a lot of active scanning and exploitation of public facing applications and brute force and so on."

B4 Severity and Impact

The severity and impact theme deals with assessing the seriousness of security incidents and their effects. This involves evaluating the severity of events and addressing challenges.

The first code is "Severity", which focuses on assessing how serious an event is. Interviewees shared insights into evaluating severity and how the MITRE ATT&CK Framework can be used for this but also how the severity is affected by other factors:

"In my personal experience, the further you get into the miter attack framework in terms of tactics, the more difficult it becomes ... I think the further you are in the kill chain, the more severe it becomes."

"Or with phishing, are credentials harvested? Are these company credentials or not? You can take all that with you and of course the infected host and the infected user. And maybe the number of host or networks. Then the more the worse"

"Is one host infected or is it an entire network or even the entire infrastructure."

Another code in this theme is "Challenges" which discusses the challenges that analysts face. These challenges may be due to the complexity of an alert but could also be due to the lack of information or even the excess of information.

"Detecting and analyzing a scan is a lot more trivial than verifying if command and control traffic is taking place."

"What you often see is that the amount of false positives increases as the amount of information you have goes down."

"Depending on the log source, the amount of information can be limited."

"You collect a lot of internet noise that occurs during the initial access and reconnaissance phase."

Table A1. Overview of identified themes from the preliminary interviews.

Theme	Description
Investigation Process and Methodology	This theme focuses on the systematic approach to investigating security events.
Contextual Understanding	This theme is about the broader context surrounding security events such as user behavior and attacker identity.
Framework and Models	This theme involves the use of established frameworks and models to analyze security events including identifying attack chains.
Severity and Impact	This theme focusses on assessing the severity and impact of security events. Including the importance of additional information in assessing the severity

Table A2. Participants of the preliminary interviews

Participant	Organization	Role	Experience (Years)
01	ESH	Tier 1 Analyst	1
02	ESH	Tier 2 Analist	5
03	ESH	Tier 3 Analist	13
04	Eviden	Tier 2 Analyst	2
05	KPMG	SOC Manager	2

‘

C. ORIGINAL MITRE ATT&CK MAPPING

TABLE A3: Original mapping of Suricata alerts to MITRE ATT&CK Techniques

Alert Message	Technique	Name
ET DROP Spamhaus DROP Listed Traffic Inbound group 39 Classification: Misc Attack	T1133	External Remote Services
ET HUNTING TW Likely Javascript-Obfuscator Usage Observed M1 Classification: Misc activity	T1027	Obfuscated Files or Information
ET INFO DropBox User Content Domain (dl .dropboxusercontent .com in TLS SNI) Classification: Misc activity	T1567	Exfiltration Over Web Service
ET INFO External IP Address Lookup Domain (ipify .org) in TLS SNI Classification: Misc activity	T1614	System Location Discovery
ET INFO External IP Lookup Domain (freegeiop .net in DNS lookup) Classification: Device Retrieving External IP Address Detected	T1614	System Location Discovery
ET INFO External IP Lookup Domain (ipify .org) in DNS Lookup Classification: Misc activity	T1614	System Location Discovery
ET JA3 Hash - Abuse.ch Possible Adware Classification: Unknown Traffic	T1082	System Information Discovery
ET MALWARE Win32/Suspected Reverse Shell Connection Classification: A Network Trojan was detected	T1659	Content Injection
ET MALWARE Windows Microsoft Windows DOS prompt command Error not recognized Classification: A Network Trojan was detected	T1059.003	Command and Scripting Interpreter: Windows Command Shell
ET MALWARE Windows dir Microsoft Windows DOS prompt command exit OUTBOUND Classification: A Network Trojan was detected	T1059.003	Command and Scripting Interpreter: windows Command Shell
ET PHISHING Possible Phishing Redirect Dec 13 2016 Classification: Possible Social Engineering Attempted	T1566	Phishing
ET POLICY Dropbox.com Offsite File Backup in Use Classification: Potential Corporate Privacy Violation	T1567	Exfiltration Over Web Service
ET POLICY PE EXE or DLL Windows file download HTTP Classification: Potential Corporate Privacy Violation	T1105	Ingress Tool Transfer
ET POLICY Possible IP Check api.ipify.org Classification: Potential Corporate Privacy Violation	T1614	System Location Discovery
ET POLICY Vulnerable Java Version 1.8.x Detected Classification: Potentially Bad Traffic	T1203	Exploitation for Client Execution
ET POLICY curl User-Agent Outbound Classification: Attempted Information Leak	T1071.002	Application Layer Protocol: File Transfer Protocols
ET SCAN Behavioral Unusual Port 135 traffic Potential Scan or Infection Classification: Misc activity	T1595.001	Active Scanning: Scanning IP Blocks
ET SCAN Possible Nmap User-Agent Observed Classification: Web Application Attack	T1595.001	Active Scanning: Scanning IP Blocks
GPL EXPLOIT Microsoft cmd.exe banner Classification: Successful Administrator Privilege Gain	T1068	Exploitation for Privilege Escalation
ET POLICY Python-urllib/ Suspicious User Agent Attempted Information Leak	T1041	Exfiltration Over C2 Channel
ET SCAN Possible Nmap User-Agent Observed Web Application Attack	T1595.002	Active Scanning: Vulnerability Scanning
ET SCAN Nmap Scripting Engine User-Agent Detected (Nmap Scripting Engine) Web Application Attack	T1595.002	Active Scanning: Vulnerability Scanning
ET SCAN NMAP SIP Version Detect OPTIONS Scan Attempted Information Leak	T1595.002	Active Scanning: Vulnerability Scanning
ET SCAN Potential SSH Scan Attempted Information Leak	T1595.001	Active Scanning: Scanning IP Blocks
ET SCAN Suspicious inbound to PostgreSQL port 5432 Potentially Bad Traffic	T1595	Active Scanning
ET SCAN Suspicious inbound to MSSQL port 1433 Potentially Bad Traffic	T1595	Active Scanning
ET SCAN Suspicious inbound to mSQL port 4333 Potentially Bad Traffic	T1595	Active Scanning
ET SCAN Suspicious inbound to Oracle SQL port 1521 Potentially Bad Traffic	T1595	Active Scanning
ET SCAN Suspicious inbound to mySQL port 3306 Potentially Bad Traffic	T1595	Active Scanning
ET SCAN Potential VNC Scan 5800-5820 Attempted Information Leak	T1595	Active Scanning
ET SCAN Potential VNC Scan 5900-5920 Attempted Information Leak	T1595	Active Scanning
ET SCAN NMAP SIP Version Detection Script Activity Attempted Information Leak	T1595.002	Active Scanning: Vulnerability Scanning
ET POLICY Http Client Body contains passwd in cleartext Potential Corporate Privacy Violation	T1040	Network Sniffing
ET WEB_SERVER Possible XXE SYSTEM ENTITY in POST BODY. A Network Trojan was detected	T1190	Exploit Public-Facing Application
GPL EXPLOIT .cnf access access to a potentially vulnerable web application	T1190	Exploit Public-Facing Application
GPL WEB_SERVER author.exe access access to a potentially vulnerable web application	T1190	Exploit Public-Facing Application
ET WEB_SPECIFIC_APPS Possible JBoss JMX Console Beanshell Deployer WAR Upload and Deployment Exploit Attempt Web Application Attack	T1059	Command and Scripting Interpreter

Continued on next page

Table A3 – continued from previous page

Alert Message	Technique	Name
ET WEB_SERVER ColdFusion componentutils access Web Application Attack	T1190	Exploit Public-Facing Application
ET WEB_SPECIFIC_APPS Horde type Parameter Local File Inclusion Attempt Web Application Attack	T1190	Exploit Public-Facing Application
ET WEB_SPECIFIC_APPS Request to Wordpress W3TC Plug-in dbcache Directory A Network Trojan was detected	T1190	Exploit Public-Facing Application
ET WEB_SERVER /etc/shadow Detected in URI Attempted Information Leak	T1003.008	OS Credential Dumping: /etc/passwd and /etc/shadow
ET WEB_SPECIFIC_APPS PHP-CGI query string parameter vulnerability Web Application Attack	T1003	OS Credential Dumping
ET WEB_SERVER WEB-PHP phpinfo access Information Leak	T1190	Exploit Public-Facing Application
ET WEB_SERVER Script tag in URI Possible Cross Site Scripting Attempt Web Application Attack	T1659	Content Injection
ET WEB_SERVER PHP SERVER SuperGlobal in URI Potentially Bad Traffic	T1190	Exploit Public-Facing Application
ET WEB_SPECIFIC_APPS Ve-EDIT edit_htmlarea.php highlighter Parameter Remote File Inclusion Web Application Attack	T1190	Exploit Public-Facing Application
ET WEB_SERVER Exploit Suspected PHP Injection Attack (cmd) Web Application Attack	T1190	Exploit Public-Facing Application
ET WEB_SPECIFIC_APPS SAPID get_infochannel.inc.php Remote File inclusion Attempt Web Application Attack	T1190	Exploit Public-Facing Application
ET WEB_SERVER PHP SESSION SuperGlobal in URI Potentially Bad Traffic	T1190	Exploit Public-Facing Application
ET WEB_SPECIFIC_APPS PHP Aardvark Topsites PHP CONFIG PATH Remote File Include Attempt Web Application Attack	T1190	Exploit Public-Facing Application
ET WEB_SPECIFIC_APPS BASE base_stat_common.php remote file include Web Application Attack	T1190	Exploit Public-Facing Application
ET WEB_SPECIFIC_APPS phpSkelSite theme parameter remote file inclusion Web Application Attack	T1190	Exploit Public-Facing Application
ET WEB_SPECIFIC_APPS TECHNOTE shop_this_skin_path Parameter Remote File Inclusion Web Application Attack	T1190	Exploit Public-Facing Application
ET WEB_SPECIFIC_APPS phPortal gunaysoft.php sayfaid Parameter Remote File Inclusion Web Application Attack	T1190	Exploit Public-Facing Application
ET WEB_SPECIFIC_APPS phPortal gunaysoft.php icerikyolu Parameter Remote File Inclusion Web Application Attack	T1190	Exploit Public-Facing Application
ET WEB_SPECIFIC_APPS ProjectButler RFI attempt Web Application Attack	T1190	Exploit Public-Facing Application
ET WEB_SPECIFIC_APPS MODx CMS snippet.reflect.php reflect_base Remote File Inclusion Web Application Attack	T1190	Exploit Public-Facing Application
ET WEB_SPECIFIC_APPS FormMailer formmailer.admin.inc.php BASE_DIR Parameter Remote File Inclusion Attempt Web Application Attack	T1190	Exploit Public-Facing Application
ET WEB_SPECIFIC_APPS MAXcms fm_includes_special Parameter Remote File Inclusion Attempt Web Application Attack	T1190	Exploit Public-Facing Application
ET WEB_SPECIFIC_APPS SERWeb main_prepend.php functionsdir Parameter Remote File Inclusion Web Application Attack	T1190	Exploit Public-Facing Application
ET WEB_SPECIFIC_APPS Sisplet CMS komentar.php site_path Parameter Remote File Inclusion Attempt Web Application Attack	T1190	Exploit Public-Facing Application
ET WEB_SPECIFIC_APPS SERWeb load_lang.php configdir Parameter Remote File Inclusion Web Application Attack	T1190	Exploit Public-Facing Application
ET WEB_SPECIFIC_APPS OpenX phpAdsNew phpAds_geoPlugin Parameter Remote File Inclusion Attempt Web Application Attack	T1190	Exploit Public-Facing Application
ET WEB_SPECIFIC_APPS Possible eFront database.php Remote File Inclusion Attempt Web Application Attack	T1190	Exploit Public-Facing Application
ET WEB_SPECIFIC_APPS p-Table for WordPress wptable-tinymce.php ABSPATH Parameter RFI Attempt Web Application Attack	T1190	Exploit Public-Facing Application
ET WEB_SPECIFIC_APPS AjaxPortal di.php pathtoserverdata Parameter Remote File Inclusion Attempt Web Application Attack	T1190	Exploit Public-Facing Application
ET WEB_SPECIFIC_APPS PointComma pctemplate.php pcConfig Parameter Remote File Inclusion Attempt Web Application Attack	T1190	Exploit Public-Facing Application
ET WEB_SPECIFIC_APPS ProdLer prodler.class.php sPath Parameter Remote File Inclusion Attempt Web Application Attack	T1190	Exploit Public-Facing Application
ET WEB_SPECIFIC_APPS KingCMS menu.php CONFIG Parameter Remote File Inclusion Web Application Attack	T1190	Exploit Public-Facing Application
ET WEB_SPECIFIC_APPS YapBB class_yapbbcooker.php cfgIncludeDirectory Parameter Remote File Inclusion Web Application Attack	T1190	Exploit Public-Facing Application
ET WEB_SPECIFIC_APPS PHP-Paid4Mail RFI attempt Web Application Attack	T1190	Exploit Public-Facing Application
ET WEB_SPECIFIC_APPS OBOphiX fonctions_racine.php chemin_lib parameter Remote File Inclusion Attempt Web Application Attack	T1190	Exploit Public-Facing Application

Continued on next page

Table A3 – continued from previous page

Alert Message	Technique	Name
ET WEB_SPECIFIC_APPS Enthusiast path parameter Remote File Inclusion Web Application Attack	T1190	Exploit Public-Facing Application
ET WEB_SPECIFIC_APPS Achievo suphp config_atkroot parameter Remote File Inclusion Attempt Web Application Attack	T1190	Exploit Public-Facing Application
ET WEB_SPECIFIC_APPS PHPOF DB_AdoDB.Class.PHP PHPOF_INCLUDE_PATH parameter Remote File Inclusion Web Application Attack	T1190	Exploit Public-Facing Application
ET WEB_SPECIFIC_APPS Mambo Component com_smf smf.php Remote File Inclusion Attempt Web Application Attack	T1190	Exploit Public-Facing Application
ET WEB_SPECIFIC_APPS Possible Mambo/Joomla! com_koesubmit Component 'koesubmit.php' Remote File Inclusion Attempt Web Application Attack	T1190	Exploit Public-Facing Application
ET WEB_SPECIFIC_APPS Joomla AjaxChat Component ajcuser.php GLOBALS Parameter Remote File Inclusion Attempt Web Application Attack	T1190	Exploit Public-Facing Application
ET WEB_SPECIFIC_APPS phptraverse mp3_id.php GLOBALS Parameter Remote File Inclusion Attempt Web Application Attack	T1190	Exploit Public-Facing Application
ET WEB_SPECIFIC_APPS PHP phpMyAgenda rootagenda Remote File Include Attempt Web Application Attack	T1190	Exploit Public-Facing Application
ET WEB_SPECIFIC_APPS Joomla swMenuPro ImageManager.php Remote File Inclusion Attempt Web Application Attack	T1190	Exploit Public-Facing Application
ET WEB_SPECIFIC_APPS Joomla Simple RSS Reader admin.rssreader.php mosConfig_live_site Parameter Remote File Inclusion Web Application Attack	T1190	Exploit Public-Facing Application
ET WEB_SPECIFIC_APPS Joomla Onguma Time Sheet Component onguma.class.php mosConfig_absolute_path Parameter Remote File Inclusion Web Application Attack	T1190	Exploit Public-Facing Application
ET WEB_SPECIFIC_APPS Joomla Dada Mail Manager Component config.dadamail.php GLOBALS Parameter Remote File Inclusion Web Application Attack	T1190	Exploit Public-Facing Application
ET WEB_SPECIFIC_APPS Golem Gaming Portal root_path Parameter Remote File inclusion Attempt Web Application Attack	T1190	Exploit Public-Facing Application
ET WEB_SPECIFIC_APPS KR-Web krgourl.php DOCUMENT_ROOT Parameter Remote File Inclusion Attempt Web Application Attack	T1190	Exploit Public-Facing Application
ET WEB_SPECIFIC_APPS Possible OpenSiteAdmin pageHeader.php Remote File Inclusion Attempt Web Application Attack	T1190	Exploit Public-Facing Application
ET WEB_SPECIFIC_APPS PHP Classifieds class.phpmailer.php lang_path Parameter Remote File Inclusion Attempt Web Application Attack	T1190	Exploit Public-Facing Application
ET WEB_SPECIFIC_APPS DesktopOnNet frontpage.php app_path Parameter Remote File Inclusion Web Application Attack	T1190	Exploit Public-Facing Application
ET WEB_SPECIFIC_APPS DesktopOnNet don3_requiem.php app_path Parameter Remote File Inclusion Web Application Attack	T1190	Exploit Public-Facing Application
GPL EXPLOIT ISAPI .idq access access to a potentially vulnerable web application	T1190	Exploit Public-Facing Application
GPL EXPLOIT iissamples access Web Application Attack	T1190	Exploit Public-Facing Application
GPL EXPLOIT ISAPI .idq attempt Web Application Attack	T1190	Exploit Public-Facing Application
GPL EXPLOIT /msadc/samples/ access Web Application Attack	T1190	Exploit Public-Facing Application
ET WEB_SERVER /system32/ in Uri - Possible Protected Directory Access Attempt Attempted Information Leak	T1190	Exploit Public-Facing Application
ET WEB_SERVER cmd.exe In URI - Possible Command Execution Attempt Attempted Information Leak	T1190	Exploit Public-Facing Application
GPL WEB_SERVER viewcode access Web Application Attack	T1190	Exploit Public-Facing Application
GPL EXPLOIT unicode directory traversal attempt Web Application Attack	T1190	Exploit Public-Facing Application
GPL EXPLOIT fpcount access access to a potentially vulnerable web application	T1190	Exploit Public-Facing Application
ET WEB_SERVER ColdFusion administrator access Web Application Attack	T1190	Exploit Public-Facing Application
GPL WEB_SERVER service.pwd access to a potentially vulnerable web application	T1190	Exploit Public-Facing Application
GPL WEB_SERVER authors.pwd access access to a potentially vulnerable web application	T1190	Exploit Public-Facing Application
GPL EXPLOIT administrators.pwd access access to a potentially vulnerable web application	T1190	Exploit Public-Facing Application
ET WEB_SERVER SELECT USER SQL Injection Attempt in URI Web Application Attack	T1003	OS Credential Dumping
ET WEB_SERVER Possible SQL Injection Attempt SELECT FROM Web Application Attack	T1190	Exploit Public-Facing Application
GPL WEB_SERVER .htaccess access Attempted Information Leak	T1213	Data from Information Repositories

Continued on next page

Table A3 – continued from previous page

Alert Message	Technique	Name
GPL WEB_SERVER .htpasswd access Web Application Attack	T1213	Data from Information Repositories
GPL WEB_SERVER globals.pl access access to a potentially vulnerable web application	T1190	Exploit Public-Facing Application
GPL EXPLOIT .htr access access to a potentially vulnerable web application	T1190	Exploit Public-Facing Application
GPL EXPLOIT iisadmpwd attempt Web Application Attack	T1213	Data from Information Repositories
GPL EXPLOIT /iisadmpwd/aexp2.htr access access to a potentially vulnerable web application	T1190	Exploit Public-Facing Application
ET WEB_SPECIFIC_APPS WEB-PHP RCE PHPBB 2004-1315 Web Application Attack	T1190	Exploit Public-Facing Application
ET WEB_SERVER MYSQL SELECT CONCAT SQL Injection Attempt Web Application Attack	T1213	Data from Information Repositories
ET WEB_SERVER Possible SQL Injection Attempt UNION SELECT Web Application Attack	T1190	Exploit Public-Facing Application
GPL EXPLOIT CodeRed v2 root.exe access Web Application Attack	T1190	Exploit Public-Facing Application
GPL WEB_SERVER / root access Attempted Information Leak	T1190	Exploit Public-Facing Application
ET WEB_SERVER Possible CVE-2014-6271 Attempt Attempted Administrator Privilege Gain	T1210	Exploitation of Remote Services
ET WEB_SERVER Possible CVE-2014-6271 Attempt in Headers Attempted Administrator Privilege Gain	T1210	Exploitation of Remote Services
ET CURRENT_EVENTS QNAP Shellshock CVE-2014-6271 Attempted Administrator Privilege Gain	T1210	Exploitation of Remote Services
ET SCAN Nikto Web App Scan in Progress Web Application Attack	T1595	Active Scanning
ET WEB_SERVER Possible Cherokee Web Server GET AUX Request Denial Of Service Attempt Attempted Denial of Service	T1499.003	Endpoint Devial of Service: Application Exhaustion Flood
ET SCAN Potential SSH Scan OUTBOUND Attempted Information Leak	T1595	Active Scanning
ET SCAN Nmap Scripting Engine User-Agent Detected (Nmap NSE) Web Application Attack	T1595	Active Scanning
ETPRO ATTACK_RESPONSE MongoDB Version Request Successful Administrator Privilege Gain	T1003	OS Credential Dumping
ETPRO ATTACK_RESPONSE MongoDB Database Enumeration Request Successful Administrator Privilege Gain	T1003	OS Credential Dumping
ET WEB_SERVER /bin/bash In URI Possible Shell Command Execution Attempt Within Web Exploit Web Application Attack	T1190	Exploit Public-Facing Application
ET WEB_SERVER /bin/sh In URI Possible Shell Command Execution Attempt Web Application Attack	T1059	Command and Scripting Interpreter
ET SCAN Rapid IMAPS Connections - Possible Brute Force Attack Misc activity	T1110	Brute Force
ET SCAN Rapid POP3S Connections - Possible Brute Force Attack Misc activity	T1110	Brute Force
ET WEB_SERVER Possible MySQL SQLi Attempt Information Schema Access Web Application Attack	T1190	Exploit Public-Facing Application
ET SCAN Sqlmap SQL Injection Scan Attempted Information Leak	T1595	Active Scanning
ETPRO WEB_SERVER SQLMap Scan Tool User Agent Web Application Attack	T1595	Active Scanning
ET SCAN Rapid POP3 Connections - Possible Brute Force Attack Misc activity	T1110	Brute Force
GPL WEB_SERVER DELETE attempt access to a potentially vulnerable web application	T1190	Exploit Public-Facing Application
ET POLICY POSSIBLE Web Crawl using Wget Attempted Information Leak	T1595	Active Scanning
ET SCAN NMAP OS Detection Probe Attempted Information Leak	T1595.001	Active Scanning: Scanning IP Blocks
ET SCAN Rapid IMAP Connections - Possible Brute Force Attack Misc activity	T1110	Brute Force
ET WEB_SERVER /bin/bash In URI, Possible Shell Command Execution Attempt Within Web Exploit Web Application Attack	T1059	Command and Scripting Interpreter
ETPRO SCAN IPMI Get Authentication Request (null seq number - null sessionID) A Network Trojan was detected	T1078	Valid Accounts
ET INFO Executable Download from dotted-quad Host A Network Trojan was detected	T1189	Drive-by Compromise
ETPRO WEB_SERVER Possible Information Leak Vuln CVE-2015-1648 Web Application Attack	T1213	Data from Information Repositories
ET POLICY Executable and linking format (ELF) file download Potential Corporate Privacy Violation	T1204.002	User Execution: Malicious File
ET SCAN Grendel-Scan Web Application Security Scan Detected Attempted Information Leak	T1595	Active Scanning
ET TROJAN Backdoor family PCRat/Gh0st CnC traffic (OUTBOUND) 106 A Network Trojan was detected	T1041	Exfiltration Over C2 Channel

D. ALTERNATIVE MITRE ATT&CK MAPPING

TABLE A4: Readjusted mapping of Suricata alerts to MITRE ATT&CK Techniques. Created using expert feedback.

Alert Message	Technique	Name
ET DROP Spamhaus DROP Listed Traffic Inbound group 39 Classification: Misc Attack	T1071	Application Layer Protocol
ET HUNTING TW Likely Javascript-Obfuscator Usage Observed M1 Classification: Misc activity	T1027	Obfuscated Files or Information
ET INFO DropBox User Content Domain (dl .dropboxusercontent .com in TLS SNI) Classification: Misc activity	T1567	Exfiltration Over Web Service
ET INFO External IP Address Lookup Domain (ipify .org) in TLS SNI Classification: Misc activity	T1614	System Location Discovery
ET INFO External IP Lookup Domain (freegeiop .net in DNS lookup) Classification: Device Retrieving External IP Address Detected	T1614	System Location Discovery
ET INFO External IP Lookup Domain (ipify .org) in DNS Lookup Classification: Misc activity	T1614	System Location Discovery
ET JA3 Hash - Abuse.ch Possible Adware Classification: Unknown Traffic	T1071	Application Layer Protocol: Web Protocols
ET MALWARE Win32/Suspected Reverse Shell Connection Classification: A Network Trojan was detected	T1059	Command and Scripting Interpreter
ET MALWARE Windows Microsoft Windows DOS prompt command Error not recognized Classification: A Network Trojan was detected	T1059.003	Command and Scripting Interpreter: Windows Command Shell
ET MALWARE Windows dir Microsoft Windows DOS prompt command exit OUTBOUND Classification: A Network Trojan was detected	T1059.003	Command and Scripting Interpreter: windows Command Shell
ET PHISHING Possible Phishing Redirect Dec 13 2016 Classification: Possible Social Engineering Attempted	T1566	Phishing
ET POLICY Dropbox.com Offsite File Backup in Use Classification: Potential Corporate Privacy Violation	T1567	Exfiltration Over Web Service
ET POLICY PE EXE or DLL Windows file download HTTP Classification: Potential Corporate Privacy Violation	T1105	Ingress Tool Transfer
ET POLICY Possible IP Check api.ipify.org Classification: Potential Corporate Privacy Violation	T1614	System Location Discovery
ET POLICY Vulnerable Java Version 1.8.x Detected Classification: Potentially Bad Traffic	T1203	Exploitation for Client Execution
ET POLICY curl User-Agent Outbound Classification: Attempted Information Leak	T1071	Application Layer Protocol
ET SCAN Behavioral Unusual Port 135 traffic Potential Scan or Infection Classification: Misc activity	T1595.001	Active Scanning: Scanning IP Blocks
ET SCAN Possible Nmap User-Agent Observed Classification: Web Application Attack	T1595.001	Active Scanning: Scanning IP Blocks
GPL EXPLOIT Microsoft cmd.exe banner Classification: Successful Administrator Privilege Gain	T1068	Exploitation for Privilege Escalation
ET POLICY Python-urllib/ Suspicious User Agent Attempted Information Leak	T1041	Exfiltration Over C2 Channel
ET SCAN Possible Nmap User-Agent Observed Web Application Attack	T1595	Active Scanning
ET SCAN Nmap Scripting Engine User-Agent Detected (Nmap Scripting Engine) Web Application Attack	T1595	Active Scanning
ET SCAN NMAP SIP Version Detect OPTIONS Scan Attempted Information Leak	T1595.002	Active Scanning: Vulnerability Scanning
ET SCAN Potential SSH Scan Attempted Information Leak	T1595	Active Scanning
ET SCAN Suspicious inbound to PostgreSQL port 5432 Potentially Bad Traffic	T1595	Active Scanning
ET SCAN Suspicious inbound to MSSQL port 1433 Potentially Bad Traffic	T1595	Active Scanning
ET SCAN Suspicious inbound to mSQL port 4333 Potentially Bad Traffic	T1595	Active Scanning
ET SCAN Suspicious inbound to Oracle SQL port 1521 Potentially Bad Traffic	T1595	Active Scanning
ET SCAN Suspicious inbound to mySQL port 3306 Potentially Bad Traffic	T1595	Active Scanning
ET SCAN Potential VNC Scan 5800-5820 Attempted Information Leak	T1595	Active Scanning
ET SCAN Potential VNC Scan 5900-5920 Attempted Information Leak	T1595	Active Scanning
ET SCAN NMAP SIP Version Detection Script Activity Attempted Information Leak	T1595.002	Active Scanning: Vulnerability Scanning
ET POLICY Http Client Body contains passwd in cleartext Potential Corporate Privacy Violation	T1040	Network Sniffing
ET WEB_SERVER Possible XXE SYSTEM ENTITY in POST BODY. A Network Trojan was detected	T1190	Exploit Public-Facing Application
GPL EXPLOIT .cnf access access to a potentially vulnerable web application	T1190	Exploit Public-Facing Application
GPL WEB_SERVER author.exe access access to a potentially vulnerable web application	T1190	Exploit Public-Facing Application
ET WEB_SPECIFIC_APPS Possible JBoss JMX Console Beanshell Deployer WAR Upload and Deployment Exploit Attempt Web Application Attack	T1059	Command and Scripting Interpreter

Continued on next page

Table A4 – continued from previous page

Alert Message	Technique	Name
ET WEB_SERVER ColdFusion componentutils access Web Application Attack	T1190	Exploit Public-Facing Application
ET WEB_SPECIFIC_APPS Horde type Parameter Local File Inclusion Attempt Web Application Attack	T1190	Exploit Public-Facing Application
ET WEB_SPECIFIC_APPS Request to Wordpress W3TC Plug-in dbcache Directory A Network Trojan was detected	T1190	Exploit Public-Facing Application
ET WEB_SERVER /etc/shadow Detected in URI Attempted Information Leak	T1003.008	OS Credential Dumping: /etc/passwd and /etc/shadow
ET WEB_SPECIFIC_APPS PHP-CGI query string parameter vulnerability Web Application Attack	T1190	Exploit Public-Facing Application
ET WEB_SERVER WEB-PHP phpinfo access Information Leak	T1592	Gather Victim Host Information
ET WEB_SERVER Script tag in URI Possible Cross Site Scripting Attempt Web Application Attack	T1059.007	Command and Scripting Interpreter: JavaScript
ET WEB_SERVER PHP SERVER SuperGlobal in URI Potentially Bad Traffic	T1190	Exploit Public-Facing Application
ET WEB_SPECIFIC_APPS Ve-EDIT edit_htmlarea.php highlighter Parameter Remote File Inclusion Web Application Attack	T1190	Exploit Public-Facing Application
ET WEB_SERVER Exploit Suspected PHP Injection Attack (cmd) Web Application Attack	T1059	Command and Scripting Interpreter
ET WEB_SPECIFIC_APPS SAPID get_infochannel.inc.php Remote File inclusion Attempt Web Application Attack	T1059	Command and script Interpreter
ET WEB_SERVER PHP SESSION SuperGlobal in URI Potentially Bad Traffic	T1190	Exploit Public-Facing Application
ET WEB_SPECIFIC_APPS PHP Aardvark Topsites PHP CONFIG PATH Remote File Include Attempt Web Application Attack	T1190	Exploit Public-Facing Application
ET WEB_SPECIFIC_APPS BASE base_stat_common.php remote file include Web Application Attack	T1059	Command and script Interpreter
ET WEB_SPECIFIC_APPS phpSkelSite theme parameter remote file inclusion Web Application Attack	T1059	Command and script Interpreter
ET WEB_SPECIFIC_APPS TECHNOTE shop_this_skin_path Parameter Remote File Inclusion Web Application Attack	T1059	Command and script Interpreter
ET WEB_SPECIFIC_APPS phPortal gunaysoft.php sayfaid Parameter Remote File Inclusion Web Application Attack	T1059	Command and script Interpreter
ET WEB_SPECIFIC_APPS phPortal gunaysoft.php icerikyolu Parameter Remote File Inclusion Web Application Attack	T1059	Command and script Interpreter
ET WEB_SPECIFIC_APPS ProjectButler RFI attempt Web Application Attack	T1059	Command and script Interpreter
ET WEB_SPECIFIC_APPS MODx CMS snippet.reflect.php reflect_base Remote File Inclusion Web Application Attack	T1059	Command and script Interpreter
ET WEB_SPECIFIC_APPS FormMailer formmailer.admin.inc.php BASE_DIR Parameter Remote File Inclusion Attempt Web Application Attack	T1059	Command and script Interpreter
ET WEB_SPECIFIC_APPS MAXcms fm_includes_special Parameter Remote File Inclusion Attempt Web Application Attack	T1190	Exploit Public-Facing Application
ET WEB_SPECIFIC_APPS SERWeb main_prepend.php functionsdir Parameter Remote File Inclusion Web Application Attack	T1059	Command and script Interpreter
ET WEB_SPECIFIC_APPS Sisplet CMS komentar.php site_path Parameter Remote File Inclusion Attempt Web Application Attack	T1059	Command and script Interpreter
ET WEB_SPECIFIC_APPS SERWeb load_lang.php configdir Parameter Remote File Inclusion Web Application Attack	T1059	Command and script Interpreter
ET WEB_SPECIFIC_APPS OpenX phpAdsNew phpAds_geoPlugin Parameter Remote File Inclusion Attempt Web Application Attack	T1059	Command and script Interpreter
ET WEB_SPECIFIC_APPS Possible eFront database.php Remote File Inclusion Attempt Web Application Attack	T1059	Command and script Interpreter
ET WEB_SPECIFIC_APPS p-Table for WordPress wptable-tinymce.php ABSPATH Parameter RFI Attempt Web Application Attack	T1059	Command and script Interpreter
ET WEB_SPECIFIC_APPS AjaxPortal di.php pathtoserverdata Parameter Remote File Inclusion Attempt Web Application Attack	T1059	Command and script Interpreter
ET WEB_SPECIFIC_APPS PointComma pctemplate.php pcConfig Parameter Remote File Inclusion Attempt Web Application Attack	T1059	Command and script Interpreter
ET WEB_SPECIFIC_APPS ProdLer prodler.class.php sPath Parameter Remote File Inclusion Attempt Web Application Attack	T1190	Exploit Public-Facing Application
ET WEB_SPECIFIC_APPS KingCMS menu.php CONFIG Parameter Remote File Inclusion Web Application Attack	T1059	Command and script Interpreter
ET WEB_SPECIFIC_APPS YapBB class_yapbbcooker.php cfgIncludeDirectory Parameter Remote File Inclusion Web Application Attack	T1059	Command and script Interpreter
ET WEB_SPECIFIC_APPS PHP-Paid4Mail RFI attempt Web Application Attack	T1059	Command and script Interpreter
ET WEB_SPECIFIC_APPS OBOphiX fonctions_racine.php chemin_lib parameter Remote File Inclusion Attempt Web Application Attack	T1059	Command and script Interpreter

Continued on next page

Table A4 – continued from previous page

Alert Message	Technique	Name
ET WEB_SPECIFIC_APPS Enthusiast path parameter Remote File Inclusion Web Application Attack	T1059	Command and script Interpreter
ET WEB_SPECIFIC_APPS Achievo suphp config_atkroot parameter Remote File Inclusion Attempt Web Application Attack	T1059	Command and script Interpreter
ET WEB_SPECIFIC_APPS PHPOF DB_AdoDB.Class.PHP PHPOF_INCLUDE_PATH parameter Remote File Inclusion Web Application Attack	T1059	Command and script Interpreter
ET WEB_SPECIFIC_APPS Mambo Component com_smf smf.php Remote File Inclusion Attempt Web Application Attack	T1059	Command and script Interpreter
ET WEB_SPECIFIC_APPS Possible Mambo/Joomla! com_koesubmit Component 'koesubmit.php' Remote File Inclusion Attempt Web Application Attack	T1059	Command and script Interpreter
ET WEB_SPECIFIC_APPS Joomla AjaxChat Component ajcuser.php GLOBALS Parameter Remote File Inclusion Attempt Web Application Attack	T1059	Command and script Interpreter
ET WEB_SPECIFIC_APPS phptraverse mp3_id.php GLOBALS Parameter Remote File Inclusion Attempt Web Application Attack	T1059	Command and script Interpreter
ET WEB_SPECIFIC_APPS PHP phpMyAgenda rootagenda Remote File Include Attempt Web Application Attack	T1059	Command and script Interpreter
ET WEB_SPECIFIC_APPS Joomla swMenuPro ImageManager.php Remote File Inclusion Attempt Web Application Attack	T1059	Command and script Interpreter
ET WEB_SPECIFIC_APPS Joomla Simple RSS Reader admin.rssreader.php mosConfig_live_site Parameter Remote File Inclusion Web Application Attack	T1059	Command and script Interpreter
ET WEB_SPECIFIC_APPS Joomla Onguma Time Sheet Component onguma.class.php mosConfig_absolute_path Parameter Remote File Inclusion Web Application Attack	T1059	Command and script Interpreter
ET WEB_SPECIFIC_APPS Joomla Dada Mail Manager Component config.dadamail.php GLOBALS Parameter Remote File Inclusion Web Application Attack	T1059	Command and script Interpreter
ET WEB_SPECIFIC_APPS Golem Gaming Portal root_path Parameter Remote File inclusion Attempt Web Application Attack	T1059	Command and script Interpreter
ET WEB_SPECIFIC_APPS KR-Web krgourl.php DOCUMENT_ROOT Parameter Remote File Inclusion Attempt Web Application Attack	T1059	Command and script Interpreter
ET WEB_SPECIFIC_APPS Possible OpenSiteAdmin pageHeader.php Remote File Inclusion Attempt Web Application Attack	T1059	Command and script Interpreter
ET WEB_SPECIFIC_APPS PHP Classifieds class.phpmailer.php lang_path Parameter Remote File Inclusion Attempt Web Application Attack	T1059	Command and script Interpreter
ET WEB_SPECIFIC_APPS DesktopOnNet frontpage.php app_path Parameter Remote File Inclusion Web Application Attack	T1059	Command and script Interpreter
ET WEB_SPECIFIC_APPS DesktopOnNet don3_requiem.php app_path Parameter Remote File Inclusion Web Application Attack	T1059	Command and script Interpreter
GPL EXPLOIT ISAPI .idq access access to a potentially vulnerable web application	T1190	Exploit Public-Facing Application
GPL EXPLOIT iissamples access Web Application Attack	T1190	Exploit Public-Facing Application
GPL EXPLOIT ISAPI .idq attempt Web Application Attack	T1190	Exploit Public-Facing Application
GPL EXPLOIT /msadc/samples/ access Web Application Attack	T1190	Exploit Public-Facing Application
ET WEB_SERVER /system32/ in Uri - Possible Protected Directory Access Attempt Attempted Information Leak	T1083	File and Directory Discovery
ET WEB_SERVER cmd.exe In URI - Possible Command Execution Attempt Attempted Information Leak	T1059	Command and Scripting Interpreter
GPL WEB_SERVER viewcode access Web Application Attack	T1190	Exploit Public-Facing Application
GPL EXPLOIT unicode directory traversal attempt Web Application Attack	T1083	File and Directory Discovery
GPL EXPLOIT fpcount access access to a potentially vulnerable web application	T1190	Exploit Public-Facing Application
ET WEB_SERVER ColdFusion administrator access Web Application Attack	T1078	Valid Accounts
GPL WEB_SERVER service.pwd access to a potentially vulnerable web application	T1190	Exploit Public-Facing Application
GPL WEB_SERVER authors.pwd access access to a potentially vulnerable web application	T1190	Exploit Public-Facing Application
GPL EXPLOIT administrators.pwd access access to a potentially vulnerable web application	T1190	Exploit Public-Facing Application
ET WEB_SERVER SELECT USER SQL Injection Attempt in URI Web Application Attack	T1190	Exploit Public-Facing Application
ET WEB_SERVER Possible SQL Injection Attempt SELECT FROM Web Application Attack	T1190	Exploit Public-Facing Application
GPL WEB_SERVER .htaccess access Attempted Information Leak	T1083	File and Directory Discovery
GPL WEB_SERVER .htpasswd access Web Application Attack	T1083	File and Directory Discovery
GPL WEB_SERVER globals.pl access access to a potentially vulnerable web application	T1190	Exploit Public-Facing Application
GPL EXPLOIT .htr access access to a potentially vulnerable web application	T1190	Exploit Public-Facing Application

Continued on next page

Table A4 – continued from previous page

Alert Message	Technique	Name
GPL EXPLOIT iisadmpwd attempt Web Application Attack	T1190	Exploit Public-Facing Application
GPL EXPLOIT /iisadmpwd/aexp2.htr access access to a potentially vulnerable web application	T1190	Exploit Public-Facing Application
ET WEB_SPECIFIC_APPS WEB-PHP RCE PHPBB 2004-1315 Web Application Attack	T1190	Exploit Public-Facing Application
ET WEB_SERVER MYSQL SELECT CONCAT SQL Injection Attempt Web Application Attack	T1190	Exploit Public-Facing Application
ET WEB_SERVER Possible SQL Injection Attempt UNION SELECT Web Application Attack	T1190	Exploit Public-Facing Application
GPL EXPLOIT CodeRed v2 root.exe access Web Application Attack	T1190	Exploit Public-Facing Application
GPL WEB_SERVER / root access Attempted Information Leak	T1005	Data from Local System
ET WEB_SERVER Possible CVE-2014-6271 Attempt Attempted Administrator Privilege Gain	T1548	Abuse Elevation Control Mechanism
ET WEB_SERVER Possible CVE-2014-6271 Attempt in Headers Attempted Administrator Privilege Gain	T1548	Abuse Elevation Control Mechanism
ET CURRENT_EVENTS QNAP Shellshock CVE-2014-6271 Attempted Administrator Privilege Gain	T1548	Abuse Elevation Control Mechanism
ET SCAN Nikto Web App Scan in Progress Web Application Attack	T1595.002	Active Scanning: Vulnerability Scanning
ET WEB_SERVER Possible Cherokee Web Server GET AUX Request Denial Of Service Attempt Attempted Denial of Service	T1499.002	Endpoint Devial of Service: Service Exhaustion Flood
ET SCAN Potential SSH Scan OUTBOUND Attempted Information Leak	T1595	Active Scanning
ET SCAN Nmap Scripting Engine User-Agent Detected (Nmap NSE) Web Application Attack	T1595	Active Scanning
ETPRO ATTACK_RESPONSE MongoDB Version Request Successful Administrator Privilege Gain	T1548	Abuse Elevation Control Mechanism
ETPRO ATTACK_RESPONSE MongoDB Database Enumeration Request Successful Administrator Privilege Gain	T1548	Abuse Elevation Control Mechanism
ET WEB_SERVER /bin/bash In URI Possible Shell Command Execution Attempt Within Web Exploit Web Application Attack	T1059	Command and Scripting Interpreter
ET WEB_SERVER /bin/sh In URI Possible Shell Command Execution Attempt Web Application Attack	T1059	Command and Scripting Interpreter
ET SCAN Rapid IMAPS Connections - Possible Brute Force Attack Misc activity	T1110	Brute Force
ET SCAN Rapid POP3S Connections - Possible Brute Force Attack Misc activity	T1110	Brute Force
ET WEB_SERVER Possible MySQL SQLi Attempt Information Schema Access Web Application Attack	T1190	Exploit Public-Facing Application
ET SCAN Sqlmap SQL Injection Scan Attempted Information Leak	T1595.002	Active Scanning: Vulnerability Scanning
ETPRO WEB_SERVER SQLMap Scan Tool User Agent Web Application Attack	T1595.002	Active Scanning: Vulnerability Scanning
ET SCAN Rapid POP3 Connections - Possible Brute Force Attack Misc activity	T1110	Brute Force
GPL WEB_SERVER DELETE attempt access to a potentially vulnerable web application	T1190	Exploit Public-Facing Application
ET POLICY POSSIBLE Web Crawl using Wget Attempted Information Leak	T1595	Active Scanning
ET SCAN NMAP OS Detection Probe Attempted Information Leak	T1595	Active Scanning
ET SCAN Rapid IMAP Connections - Possible Brute Force Attack Misc activity	T1110	Brute Force
ET WEB_SERVER /bin/bash In URI, Possible Shell Command Execution Attempt Within Web Exploit Web Application Attack	T1059	Command and Scripting Interpreter
ETPRO SCAN IPMI Get Authentication Request (null seq number - null sessionID) A Network Trojan was detected	T1046	Network Service Discovery
ET INFO Executable Download from dotted-quad Host A Network Trojan was detected	T1105	Ingress Tool Transfer
ETPRO WEB_SERVER Possible Information Leak Vuln CVE-2015-1648 Web Application Attack	T1213	Data from Information Repositories
ET WEB_SERVER Suspicious Chmod Usage in URI Attempted Administrator Privilege Gain	T1068	Exploitation for Privilege Escalation
ET POLICY Executable and linking format (ELF) file download Potential Corporate Privacy Violation	T1105	Ingress Tool Transfer
ET SCAN Grendel-Scan Web Application Security Scan Detected Attempted Information Leak	T1595	Active Scanning
ET TROJAN Backdoor family PCRat/Gh0st CnC traffic (OUTBOUND) 106 A Network Trojan was detected	T1071	Application Layer Protocol