Methods of Intrusion Detection in Information Security Incident Detection: A Comparative Study

Tan Fui Bee
SysArmy Sdn.Bhd
Kuala Lumpur
fuibee@sysarmy.net

Yau Ti Dun
SysArmy Sdn.Bhd
Kuala Lumpur
alan@sysarmy.net

M N M Kahar
Faculty of Computer Systems & Software Engineering,
University Malaysia Pahang (UMP), Kuantan, Pahang, Malaysia.
mnizam@ump.edu.my

Abstract - The advance development in technology have made internet and online application and network usage become one of the important element in human life. With the high demand from the corporate and enterprise, more and more security appliances were developed and deployed, such as IDS (Intrusion Detection Systems), IPS (Intrusion Prevention Systems), Firewall, and SIEM (Security Information and Event Management). All these security tools have serve same purpose which is to safe guard the whole enterprise network. However, every single tool has different ways of detection and accuracy of the detection. This was due to each solution deploy was highly depended with the algorithms reside in the program loaded in the security tools.

These algorithms and methods provide fast and high rate of detection. However, it also produces high false alarm rate (low accuracy) and unable to handle high volume of data. This have attracted researchers to find algorithms and methods that can detect intrusions in a short period of time within a huge volume of data with high accuracy. The objective of this paper is to study and make a comparison among the available intrusion detections algorithms and methods in the intrusion detections. Focus will be given to research that have produced new intrusion detection algorithms, framework and model as well as their gaps in the research. Their research results and gaps can lead to any possible future research to identify new intrusion detection methods.

Keywords – IDS, IPS, SIEM, Intrusion Detection

1. INTRODUCTION

The usage and important of internet has increase significantly in today world. This has led to the increase of demand for the security solution from the enterprise and corporate to safe guard their data and work environment from threats. All the connected computer networks devices regardless inside or outside the enterprise or countries require to speed up all the files, images, video, audio transfer between all individual. This made the "Data" become one of the most elements important in the human life. Therefore, it is extremely important to find an effective way to protect this valuable network infrastructure from any unauthorized or malicious actions. In order to find way to provide the security for the information stored in the computer devices today, there are many tools available in the market for detecting attacks. Such as Intrusion Prevention System (IPS), Intrusion Detections System (IDS) and the more comprehensive solution of Security Incident Event Management (SIEM) tool are useful tool for detecting attacks.

One of the network security/threat prevention technology that examines network traffic flows to detect and prevent vulnerability exploits is known as Intrusion Prevention System (IPS). Attackers are usually exploits vulnerabilities that exist in system and gain control of the application or machine to perform authorizes activities such as transferring of enterprise confidential information or perform damage to the company systems by upload malicious software. Following a successful exploit, the attacker can disable the target application or can potentially access to all the rights and permissions available to the compromised application.

Beside this, to monitor network environment or systems for malicious activity or policy violations, device such as intrusion detection system (IDS) will be use to perform this task. Suspicious activities detected or violation is typically reported either to an administrator by using preconfigure alert or collected centrally using a security information and event management (SIEM) system. A SIEM system is a system that delivery result by combining outputs from multiple sources, and uses alarm filtering techniques to distinguish malicious activity from false alarms.

An intrusion detection system (IDS) are usually used to inspect all inbound and outbound network activity and identifies suspicious patterns that may indicate a network or system attack from someone attempting to break into or compromise a system [20]. There are several ways to categorize an IDS:

- Misuse detection vs. anomaly detection: In misused detection, IDS used technique to analyzes the information it gathers and compares it to large databases of attack signatures that is available. To be accurate, mean the IDS looks for a specific attacked that has already been known and documented. It is more like a virus detection system, misuse detection software is only as good as when the database of attack signatures that it uses to compare packets against are up to date and complete. Whereas, in anomaly detection, system administrator will have to defines the baseline of the normal state of the networks traffic load, breakdown, protocol, and typical packet size in the system. Then, the anomaly detector monitors network segments to compare their state to the normal baseline and look for anomalies.
- Network-based vs. host-based systems: For a network-based system, or NIDS, individual packets that are flowing through a network are analyzed. So, NIDS can detect malicious packets that are designed to be overlooked by a firewalls that only using filtering rules. For a host-based system, it is mainly used to examine all the activities on each individual computer or host.
- Passive system vs. reactive system: For a passive system, the IDS will signal an alert when it detects a potential security breach and logs the information. In a reactive system, the IDS will respond by logging off the suspected malicious source or by sending new rules to the firewall to block network traffic if suspicious activities were detected.

Firewall and IDS are both related to network security, standard firewall only limits the access between networks to prevent intrusion and does not signal any attack from inside or outside the network. Whereas, IDS is differing from a firewall in that a firewall looks out for intrusions in order to stop them from happening but IDS evaluates a suspected intrusion once it has taken place and signals an alarm. An IDS also can be deployed to watches for attacks that originate from within a network. A shortcoming of current approaches in IDS or IPS is that both only attempt to detect and prevent individual attacks and not coordinated distributed attacks [32]. So, that is a lot of research been carried out to centers around improving the ability of systems to detect attacks and the speed of the traffic that can be handled for all these security tools. Both signature-based and anomaly detection sensors are useful tools, but both only provide a snapshot of the event in time after it has transpired [32]. This is to mean that, to know the clues as to what is happening in real-time on a computer network is available but spread out all over the network. For instance, current modern operating systems can have logged system and security events at all time. Servers keep extensive records of their operations, applications write errors, warnings, and failures to their own logs. Firewalls also can program to track how many packets are dropped by different rules typically at the entry and exit points of the network. So, if all this information can be brought together to be analyses at almost real time, it would enable more robust attack detection mechanisms. These is later when the SIEM come into the picture for more comprehensive intrusion detections.

Every SIEM system was designed with the underlying principles to aggregate relevant data from multiple sources, identify deviations from the normal pattern and take appropriate action by creating an event to be attend by the security analyst. The existence of SIEM systems help to comprehend large amounts of the security data and visualization of all the data is the essential part of the SIEM systems. One of the important of SIEM is the Log Management component that used to normalize, classify, prioritize and store collected logs in a storage area. Another component named as Correlation Engine is to analyze events by using all the pre configure rules in the system. And at the end Response system will have to react properly against security threats, threat management and action alarm generation. All the analysis was done based on a knowledge base that includes attack signatures; attack behavior rules; network security policies and network configuration information. So, SIEM is a promising paradigm to reconcile traditional intrusion detection processes along with most recent advances on artificial intelligence techniques in providing automatic and self-adaptive systems [29].

SIEM systems already become an important security tools and integral component of today's information and communication network environments. SIEM tools become one of the monitoring tools that can perform early detection for any potential attacks that is about to happen [26]. Thus, with the complexity for IT infrastructure and growth of today of network connected world, Monitoring of Security events become one of the important component in most of the organization. SIEM technology is able to provide real-time analysis of security alerts generated by systems, applications and by the network hardware equipment. SIEM solutions come as a form of software, management services and are used to record data security and to generate reports for compliance purposes. The deployment of SIEM tools highly depended on rule-based correlation techniques. Since predefined rules can only detect what has been defined beforehand, there is a demand to enhance SIEM systems using anomaly detection. Dealing with the problem of anomaly detection implies the existence of an underlying concept of "normality". Actually, anomaly-based detection usually relies on two levels of abstraction that includes, (i) model of normal behavior and (ii) deviation degree measure from normality. Anomaly detection models are built upon some algorithms to establish normal behavior standard and identify, as anomaly, any event that deviates from this standard. [1]

The remainder of this paper is organized as follows. Section 2 explains the classification of Intrusion Detection. Section 3 provide literature survey. Section 4 contain the future research areas for anomaly intrusion detections and section 5 concludes the paper.

2. CLASSIFICATION OF INTRUSION DETECTION APPROACHES

There are various Intrusion Detection techniques deployed in different security tools such as IDS, IPS and SIEM. And each of the security tools all has different purpose in the network environment for intrusion detections. So, most enterprise will actually deploy few security tolls to safe guard their environment than depending on single solution. For example, IDS is a device or software application that monitors a network or systems for malicious activity or policy violations but IPS is a network security/threat prevention technology that examines network traffic flows to detect and prevent vulnerability exploits. So, IPS IPS is normally refer to as a system with response capabilities. Whereas, SIEM have the features to support for real-time aggregation, correlation and analysis of events originating from different end points devices such as IDS, IPS, firewalls, servers and anti-virus solutions and generate more relevant alert.

According to [26], a SIEM system combines outputs from multiple sources from the enterprise network, and uses alarm filtering techniques to distinguish malicious activity from false alarms. Any detected activities or violations are typically reported by alert to the system administrator or collected centrally using a SIEM system to be analyzed by security analyst. For example, when SIEM detected a potential issue, it might log additional information, generate an alert and instruct other security controls to stop an activity's progress if the security tools have the feature to re-programmed by scripting. For the most entry level of SIEM tool, the SIEM system either be rules-based or employ a statistical correlation engine to establish relationships between all the event log entries. One of the most widely deployed methods for detecting cyber terrorist attacks and protecting against cyber terrorism employ *signature-based detection techniques*, this method can only detect previously known attacks that have a corresponding signature in the database. So, the signature database has to be manually revised for each new type of attack that is discovered from time to time. These limitations have led to an increasing interest in intrusion detection techniques based on data mining [27]. Data mining-based intrusion detection techniques generally fall into one of two categories; *misuse detection* and *anomaly detection*.

A. Misuse Detection Approaches

For misuse detection in Data Mining-based intrusion detection techniques, each instance or record in a dataset is labelled as 'normal' and a learning algorithm is trained over the labelled data. These techniques are able to automatically retrain intrusion detection models on different input data that include new types of attacks, as long as they have been labelled appropriately. Research in misuse detection has focused mainly on classification of network intrusions using various standard data mining algorithms rare class predictive models, association rules and cost sensitive modelling [11]. Models of misuse are created automatically, and can be more sophisticated and precise than manually created signatures. A key advantage of misuse detection techniques is their high degree of accuracy in detecting known attacks and their variations. Their obvious drawback is the inability to detect attacks whose instances have not yet been observed [11].

B. Anomaly Detection Approaches

Anomaly detection applied to intrusion detection and computer security has been an active area of research since it was originally proposed by [2]. There are many models for the anomaly intrusion detection. These models are based on Social and Technical type of correlation (ST-SIEM) [15], temporal and logical constraints [16], named as 'Hierarchical' correlation approach based on a Complex Event Processing (CEP) as well as based on the network attack graph and correlate the events through the corresponding attack graph distances. There is also Construction of network attack graphs that is based on the application of attacker exploit rules [30]. Anomaly detection approaches build models from all the available normal data and detect deviations from the normal model in observed data. The key advantage of Anomaly detection algorithms is they can detect new types of intrusions as deviations from normal usage. In this problem, given a set of normal data to train from, and given a new piece of test data, the goal of the intrusion detection algorithm is to determine whether the test data belong to "normal" or to an anomalous behavior. However, anomaly detection schemes suffer from a high rate of false alarms. This occurs primarily because previously unseen (yet legitimate) system behaviors are also recognized as anomalies, and hence flagged as potential intrusions [5].

C. Benchmark Dataset KDD'99 and DARPA98

KDD'99 intrusion detection datasets, which are based on DARPA'98 datasets, provides labelled data for researchers working in the field of intrusion detection and is the only labelled dataset publicly available. Numerous researchers employed the datasets in KDD'99 intrusion detection datasets to study the utilization of machine learning for intrusion detection and reported detection rates up to 91% with false positive rates less than 1% [10].

This paper focuses on a comparative study of anomaly schemes for identifying different network systems intrusion detections. Generic Intrusions Detection and Diagnosis Methods that is currently available such as *Decision Tree, Neural Networks, Naive Bayesian classifiers, Fuzzy sets, Artificial Intelligent and etc.* Focus will be given to research that have used the well-known KDD-99 intrusion detection data sets to identify how each component contributes to the overall goals of identifying intrusion detections. Since 1999, KDD'99 has been the most widely used data set for the evaluation of anomaly intrusion detection methods [21, 22]. The KDD'99 Cup data set was prepared by [23] and was built based on the data captured in DARPA'98 IDS evaluation program [24, 25]. This dataset was taken from the Third International Knowledge Discovery and Data Mining Tools Competition (KDD Cup 99). The results of these study can lead to any possible future research to identify new intrusion detection and diagnosis methods.

The KDD CUP 1999 [20] standard datasets are published for research purpose. It is used in order to assess different feature selection method for Intrusion detection system. The dataset consists of 41 features and a separate feature (42nd features) that labels the connection as 'normal' or type of attack. The data set contains a total of 24 attack types that fall into 4 major categories (DoS, U2R, R2L and Probe) as follows:

- Denial of Service Attack (DoS)/attempt to make a network resource unavailable to its intended users such as suspend services of a host connected to the Internet.
- User to Root (U2R) Attack where an attacker attempts to get unauthorized access of target system
- Remote to User Attack (R2L) where attacker try to control of remote machine by guessing password
- Probing Attack (Probe) where attacker scene/examine the machine to get useful information KddCup'99

3. LITERATURE SURVEY

Security of network are a major concern for most of the enterprise today, especially with the emerging tendency of diversification, sophistication and intelligence on network attack in the connected world today. the. For all the detections techniques available, the principle is as follows: firstly, capturing the data packets that flow through the computer system and cleaning the data packets. Secondly, using an appropriate data analysis algorithm to determine whether the data packets is normal or abnormal data. Lastly, taking the alarm or other measures to warn or prompt the users for non-normal data. Hence, the core of intrusion detection is the algorithm of distinguishing whether the data is normal by analyzing the data packets.

With the emerging and growth of new network attacks, the need for intrusion detection systems to detect novel attacks becomes pressing. One of the hardest task accomplish for novel attacks was no knowledge is available for the attacks. However, a new attack will do something that is different from normal activities. So, if the system has comprehensive knowledge about normal activities and their normal deviations, then all activities that are not normal can be classified as suspicious activities. Hence, the problem can be solved by utilizing the knowledge about unknown attacks from the existing information of normal activities as well as known attacks [5]. Many researches has been done to look into intrusion detection techniques with various methods and algorithms and named it as anomaly detection or outlier detection. How anomaly detection algorithms work was requiring a set of purely normal data to train the anomaly detection model. So, any patterns no observed before from the detection will be implicitly assume "anomalies" traffic that need further investigation. So, most researches pointed the outlier that may be defined as a data point which is very different from the rest of the data, based on some measure. Most studies employ several *outlier detection schemes* to see how efficient these schemes may deal with the problem of anomaly detection.

In [2], research one of the earliest Intrusion Detection Model name as IDES, this model is a real time system that is capable of detecting break-ins, penetrations, and some other forms of computer abuse. IDES is independent of any system, application environment, system vulnerability, or type of intrusion. This IDES model performed on a basis that a security violation can be detected by monitoring a system's audit records for abnormal patterns of system usage. The model includes profiles for representing the behavior of subjects with respect to objects in terms of metrics and statistical models, and rules for acquiring knowledge about this behavior from audit records and for detecting anomalous behavior. The development of a real-time intrusion detection system is motivated by four factors [2]:

a) Most existing systems have security flaws that render them susceptible to intrusions, penetrations, and other forms of abuse; finding and fixing all these deficiencies is not feasible for technical and economic reasons;
b) Existing systems with known flaws are not easily replaced by systems that are more secure -- mainly because the systems have attractive features that are missing in the more-secure systems, or else they cannot be replaced for economic reasons;
c) Developing systems that are absolutely secure is extremely difficult, if not generally impossible; and
d) Even the most secure systems are vulnerable to abuses by insiders who misuse their privileges.

Reference [2] concluded that the IDES model allows intrusions to be detected without knowing about the flaws in the target system. Although the approach can detect most intrusions from the experiment, but it may not capable of detecting gradual modifications of behavior or through subtle forms of intrusion that use low-level features that produce huge data.

Reference [17], in their paper proposed a socio-technical coordinate system to scale and classify cyber security warnings along with security posture levels. Their proposed socio-technical security warning coordinate system was developed to correlate the operational environment metrics with an organization's security posture metrics. All the information was then analyzed by the system for the actual level of security and identify the potential security gaps and unmanaged risks into one picture for the monitor network. The scales used in their scenario are proposed heuristically and further research has been carried to see how these levels should be calibrated in [15].

Figure 1 below was presented in [15] for socio-technical framework for integrating a security risk escalation maturity model into a security information and event management system. The objective of this framework is to develop the foundations for the next generation socio-technical security information and event management systems (ST-SIEMs) enabling socio-technical security operations centers (STSOCs). This STSOCs will be operated with the integration of normal SIEM with the input on security risk information. The primary benefit of the socio-technical framework is to support organizations in overcoming the identified limitations in their security risk escalation maturity and also supporting SOCs in overcoming the limitations of their SIEMs. Metrics will be used to quantify the risk escalation maturity level. These metrics are then used by SIEMs for cross correlating security events before they are disseminated to respective organizations. Generic risk factors were used in today typical SIEMs to calculate security events which may not be necessarily relevant for every organization. The proposed framework by [15] can enable security administrators to effectively and efficiently manage security warnings and to establish necessary countermeasures. *Socio-technical correlation engine* is one of the common and core component of the ST-SIEM system. This engine is designed to perform two types of correlation: social and technical. So, taking the organization security risk escalation capabilities into consideration when correlating events logs and disseminating security warnings is one of the contribution of the ST-SIEM compared to traditional SIEMs. Whereas, most of today SIEM correlation engine is by design focused on the technical attributes of security events which to both FOSS (Free and open source software) and COTS (Commercial off-the-shelf) SIEMs.

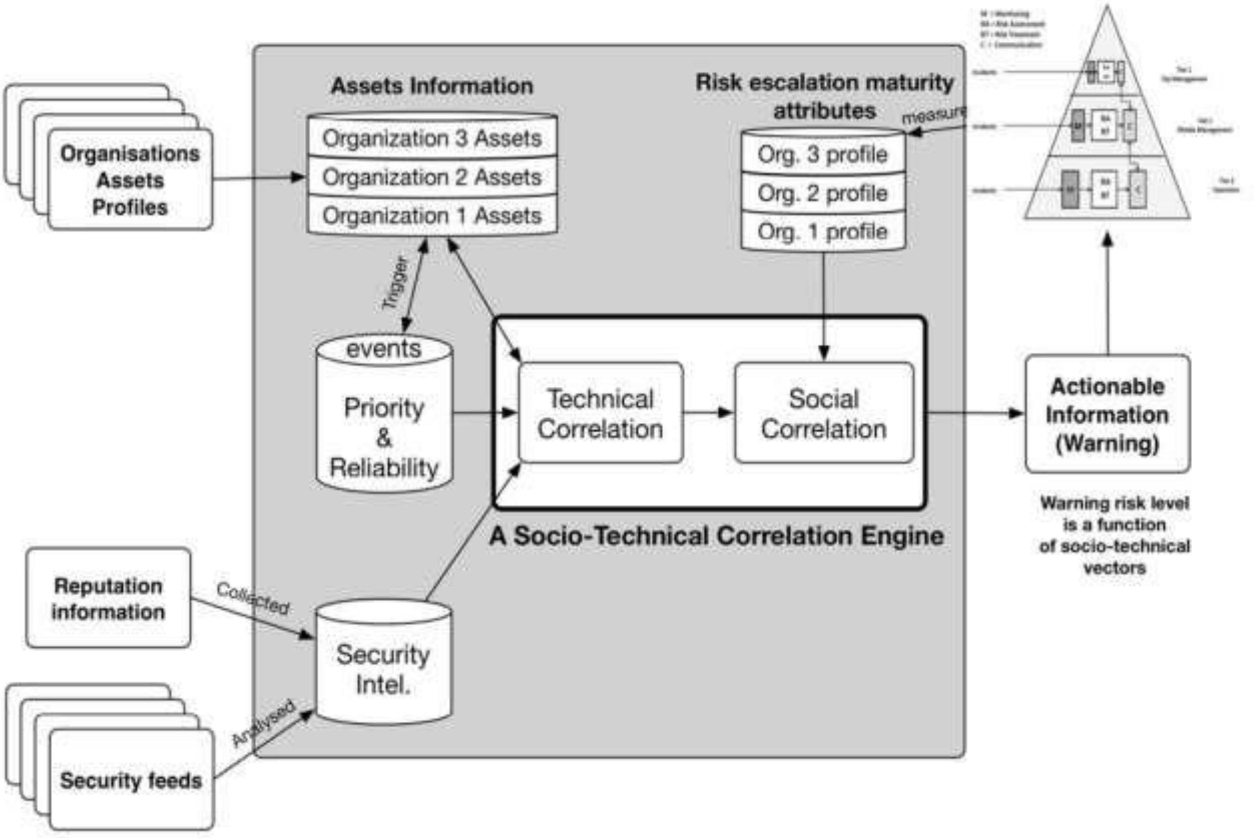

Figure 1: Socio-Technical SIEM Architecture.

[16], presented a generic Intrusion Detection and Diagnosis System (ID^{2}S), which implements a comprehensive alert correlation workflow for detection and diagnosis of complex intrusion scenarios in Large Scale Complex Critical Infrastructures showed in figure 2 below. This ID^{2}S allow detect intrusion scenarios by collecting diverse information at several architectural levels, using distributed security probes, as well as perform complex event correlation based on a Complex Event Processing Engine. The ID^{2}S used hybrid and hierarchical approach for on-line detection and diagnosis process. During the escalation process, a knowledge-based represented by an ontology is used to identified the intrusion symptoms and target and the cause of the intrusion. So, in order to recognize complex intrusion patterns and enrich the semantics of diagnosis, a 'hierarchical' correlation approach based on a Complex Event Processing (CEP) is adopted [16]. The approach captured the causal relationships among the resulting alarms by correlating them on the base of temporal and logical constraints. The ID^{2}S architectural model consists of a collection of software components that can then transform all the raw attack symptoms into high level intrusion scenario alarms/reports to alert the system administrator or/and to perform a recovery action for the system. Every software component in the ID^{2}S architectural model will focuses on different aspects of the overall correlation process The correlation capability is based on a CEP and driven by knowledge-base, which is used to recognize attack scenarios. Their work shows that it is able to detect complex attack scenarios which consist of specific sequence of malicious activities (called 'intermediate attacks') performed by the attacker in order to discover system's vulnerabilities. However, common weakness of this correlation technique is that it requires specific knowledge-based database about the attacks in order to identify their prerequisites and consequences. The hybrid and hierarchical approach is difficult to define all prerequisites and all of the possible consequences. Therefore, the major limitation of this ID^{2}S technique was it cannot correlate unknown attacks since their prerequisites and consequences are not defined.

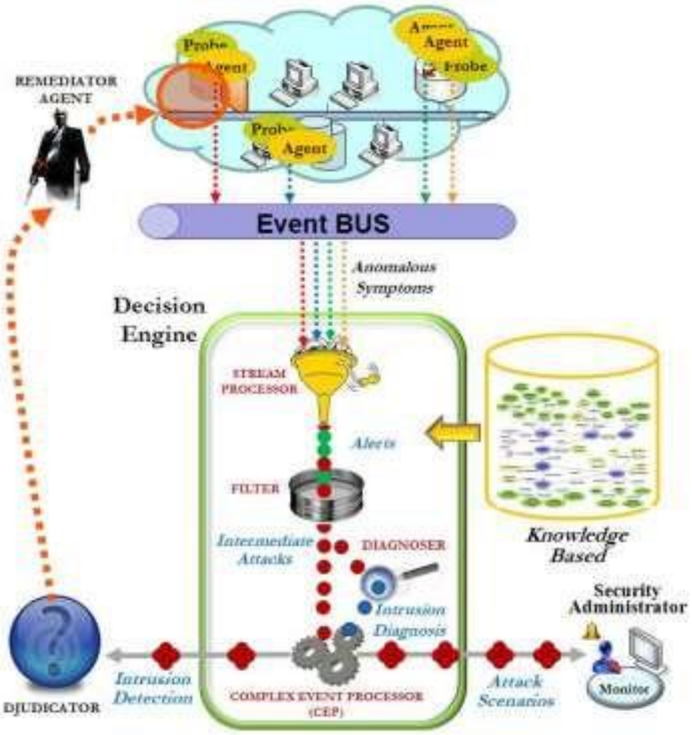

Figure 2: The proposed ID^{2}S framework

Reference [18] proposed an Advanced SIEM that intended to be implemented in the framework of the EU MASSIF project. The EU MASSIF project used the NIST National Vulnerability Database (NVD) [28] that based on Common Vulnerabilities and Exposures (CVE) dictionary to construct the basis of attack graph via known vulnerabilities. This Advanced SIEM have one main analytical components which name as Attack Modelling and Security Evaluation Component (AMSEC) was outlines. The AMSEC based on modeming of malefactors' behavior, building a common attack graph, processing current alerts for real-time adjusting of particular attack graphs, calculating different security metrics and providing security assessment procedures. The AMSEC prototype is intended to complement the SIEM analysis functionality with the capability of attack modelling and security evaluation and it was successfully calculated the security metrics for the evaluate systems.

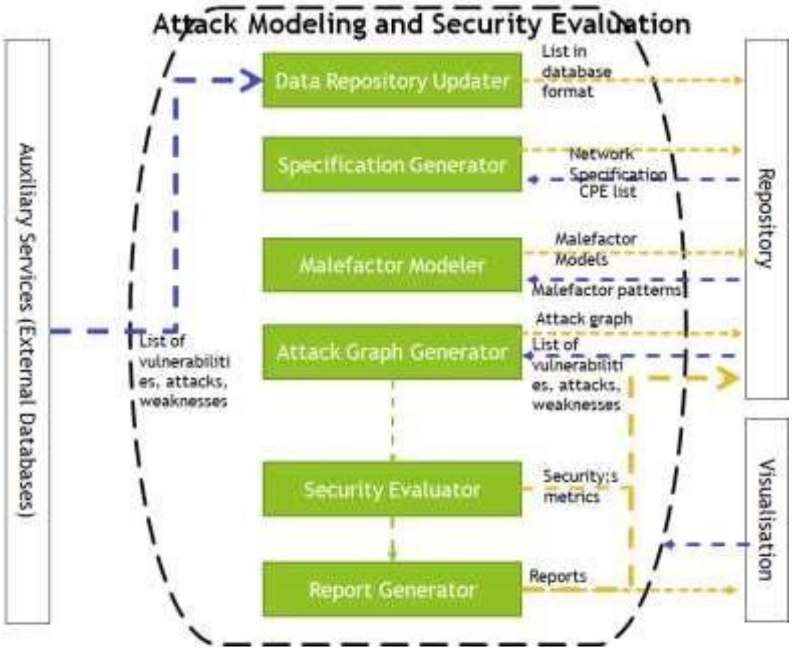

Figure 3: General architecture of Attack Modelling and Security Evaluation Component (AMSEC)

Reference [3], Mahoney. MV and Chan P.K. proposed a learning algorithms based model of normal behavior from attack free network. In the event of any behavior that deviates from the learned normal model from the normal traffic signals possible novel attacks. That is 2 unique aspects in their proposed learning algorithms. First, it is non-stationary, modelling probabilities based on the time since the last event rather than on average rate, which can prevent alarm floods. Second, the algorithms learn protocol vocabularies (at the data link through application layers) in order to detect unknown attacks that attempt to exploit implementation errors in poorly tested features of the target software. The non-stationary features have two non-stationary components developed and tested on the 1999 DARPA IDS evaluation test set, which simulates through a local network under attack. The first component is a packet header anomaly detector (PHAD) which monitors the entire data link, network, and transport layer, without any preconceptions about which fields might be useful. The second component is an application layer anomaly detector (ALAD) which combines a traditional user model based on TCP connections with a model of text-based protocols such as HTTP, FTP, and SMTP. Both systems learn which attributes are useful for anomaly detection, and then use a non-stationary model, in which events receive higher scores if no novel values have been seen for a long time. They evaluate its performance on the DARPA IDS evaluation data set and investigate the contribution of user vs. software anomalies toward detection. Their result classified anomalies into five categories based on the learn model and data set:

a) *Learned signatures* -this learned signature attempts to exploit bugs in the target system. They will first find a security vulnerability is an error, whether in software design, coding, or system configuration. And follow by attackers exploit these errors. That is impossible to test software completely, some errors will always be discovered in software after it has been delivered and put to use. So, errors that are least likely to be discovered and patched are those that occur least often in normal traffic. Any traffic or action that invoke those uncovered error is likely to be unusual. This is what their researcher name it learned signature anomaly. [3]
b) *Induced anomalies* – When that is symptoms of a successful attack, we can observe anomalous behavior from the target after a successful attack even sometimes the anomalous input is missed. This anomalies technique is similar to Forrest's host based anomaly detection technique, except that the symptoms are observed in the output of the target, rather than in the system calls that it makes.
c) *Evasion anomalies* – When they discovered when that is attempts to elude the IDS. Which mean the attacker tries to exploit errors in the IDS to hide the attack, for example, FIN scanning by port sweep to prevent server accesses from being logged. If the attempt backfires, as it does in this case, they call it an evasion anomaly.
d) *Attacker errors* – When that is bugs in the attacking program or software. The attacker might introduce arbitrary variations in the data, which they considered to be errors in the attacking software. Most provide no clues as to the nature of the attack.
e) *User behavior* – Which unexpected client IP addresses were discovered. Finally, behavioral anomaly detection models users rather than software. Most of the U2R attacks are discovered because the exploit software is uploaded on an FTP server normally used for downloads. Many R2L attacks are discovered because the client IP address is not one of the usual users.

Reference [5], proposed a method based on a technique called *pseudo-Bayes estimators* to enhance an anomaly detection system's ability to detect new attacks while reducing the false alarm rate as much as possible. The anomaly detection system called Audit Data Analysis and Mining (ADAM) that was developed at the Center for Secure Information Systems of George Mason University was one of their reference model. ADAM (Audit Data and Mining) is a combination anomaly detector and classifier trained on both attack-free traffic and traffic with labelled attacks. It was also designed to have separate training modes and detection modes. ADAM applies mining association rules techniques to look for the abnormal events in network traffic data, then it uses a classification algorithm to classify the abnormal events into normal instances and abnormal instances. The abnormal instances can be further categorized into attack names if ADAM has gained knowledge about the attacks. With the help of the classifier, the number of false alarms is greatly reduced because the abnormal associations that belong to normal instances will be filtered out. However, the normal instances and attacks that the classifier can recognize are limited to those that appear in the training data. The design composed of three modules: a pre-processing engine, a mining engine and a classification engine. The pre-processing engine sniffs TCP/IP traffic data, and extracts information from the header of each connection according to a predefined schema. The mining engine applies mining association rules to the connection records. It works on two modes: training mode and detecting mode. In training mode, the mining engine builds a profile of the users and systems normal behaviors, and generates labelled association rules, which will be used to train the classification engine. In detecting mode, the mining engine mines unexpected association rules that are different from the profile. The classification engine will classify the unexpected association rules into normal and abnormal events. Some abnormal events can be further classified as attack names. Although mining of association rules has previously been used to detect intrusions in audit trail data. ADAM is unique in two ways:

- It is real-time: it uses an incremental mining (on-line mining) which does not look at a batch of TCP/IP connections, but rather employs a sliding window of time to find the suspicious rules within that window.

- It is an anomaly detection system (instead of a misuse detection system like previous systems that use data mining for intrusion detection. Since it aims to detect anomalous behavior relative to a profile. For this, the technique builds, a profile of "normal" rules, obtained by mining past periods of time in which there were no attacks. Any rule discovered during the on-line mining that also belongs to this profile is ignored, assuming it corresponds to a normal behavior. In this sense, our technique looks for unexpected rules. This helps in reducing the number of false positives flagged by the technique.

[5] shown that the pseudo-Bayes estimator method can enhance ADAM's ability to detect new attacks. The experimental results show that the method helps in detecting new attacks whose properties are different and distinguishable from the normal instances of training data. For new attacks that do not differ much from the normal instances, the method does not work well and will misclassify them as normal instances (misclassification is not unique in pseudo-Bayes, and it exists in every supervised classifier when some classes are not distinguishable). However, the ability to capture new classes is rarely present in classifiers, and that is what they have achieved in their work.

Reference [4] developed at SRI in the EMERALD system that used historical records as its normal training data. It compares distributions of new data to the distributions obtained from those historical records and differences between the distributions indicate an intrusion. Next generation Intrusion Detection Expert System (NIDES) [4], like ADAM, monitors ports and addresses. Instead of using explicit training data, it builds a model of long term behavior over a period of hours or days, which is assumed to contain few or no attacks. If short term behavior (seconds, or a single packets) differs significantly, then an alarm is raised. NIDES does not model known attacks; instead it is used as a component of EMERALD [11], which includes host and network based signature detection for known attacks. The statistical approach used in NIDES is to compare a subject's short-term behavior with the subject's historical or long-term behavior. In the traditional NIDES framework, a subject is a user of a computer system. In the Safeguard context, a subject is an application. In comparing short-term behavior with long-term behavior, the statistical component is concerned both with long term behaviors that do not appear in short-term behavior, and with short-term behaviors that are not typical of long-term behavior. Whenever short-term behavior is sufficiently unlike long-term behavior, a warning flag is raised. They performed five experiments; three experiments - concept, verification, and refinement - were run to determine the most suitable configurations for the Safeguard application of NIDES. Each of these first three experiments include the following steps:

i They set up a predetermined set of configurations for the experiment. Each experiment they ran had a different configuration.
ii They ran TIS-supplied audit data through the statistical analysis component of NIDES to develop baseline profiles of applications.
iii. They performed false-positive tests on the profiles developed to determine if the profile was stable enough to be used in subsequent tests.
iv. They ran tests using data that included masquerading applications to perform true positive testing.
v. They reviewed the results of the experiment to determine the appropriate configuration for subsequent experiments. Their goal was to improve NIDES performance as their experiments progressed.

In the fourth experiment, they ran a cross-profiling experiment to perform additional true-positive testing, using the configuration setting of their refinement experiment. Finally, they performed experiments with application groups using information gathered during the cross-profiling experiment. Overall their results from the first (concept) experiment to their third (refinement) experiment give an improvement. They had higher detection rates and generally lower false-positive rates. The cross-profiling experiment gave them additional insight into similarities of the applications in their test set and how easily one application in the test set could masquerade as another application in the test set. They also gained some understanding of the results one obtains when profiling application groups instead of individual applications - in general, the ability to detect anomalous activity (detection sensitivity) seemed to decrease when they worked with application groupings.

Reference [9] performed anomaly detection on unlabeled data by looking at user profiles and comparing the activity during an intrusion to the non-activity during normal use. They have developed an approach to these problems that examines sequences of user actions (UNIX commands) to classify behavior as normal or anomalous. In the paper they explore the matching function needed to compare a current behavioral sequence to a historical profile. They discussed the difficulties of performing matching in human-generated data and show that exact string matching is insufficient to a single domain. They demonstrated several partial matching functions and examine their behavior on user command data. In particular, they explored two methods for weighting scores by adjacency of matches as well as two growth functions (polynomial and exponential) for scoring similarities. They found that, the best measure is user-dependent and there are indications that it may also depend on choice of instance selection technique and even the identity of the anomalous/intrusive user. There do, however, seem to be indications that polynomials bounded measures and measures biased in support of adjacent matches are, overall, preferable. Both the adjacency and the polynomial bound are evidence in support of the hypothesis that user behavioral sequences are characterized by strong correlations among temporally close command tokens.

Reference [12] classified the KDD99 dataset into four categories and all dataset split into training and testing set of data. According to [12], tree based data mining classification techniques such as Hoeffding tree, j48, Random Forest, Random Tree, RepTree were use in his study on intrusion detection dataset KDD Cup1999 using EKA 3.9 tool. In general result show using 10 fold cross validation Random forest best for Train set and J48 best for test dataset considering their comparative classification accuracy. The big challenge in intrusion detection is to achieve high detection rate and low false alarm. Any single classifier is not sufficient to achieve high accuracy and low false positive or negative. Therefore, more than one classifier can be combined to improve overall performance of attack detection

Reference [13] developed a new approach for network anomaly detection by combining neural network and clustering algorithms. They propose modified Self Organizing Map (SOM) algorithms which initially starts with null network and grows with the original data space as initial weight vector, updating neighborhood rules and learning rate dynamically in order to overcome the fixed architecture and random weight vector assignment of simple SOM. New nodes are created using distance threshold parameter and their neighborhood is identified using connection strength and its learning rule and the weight vector updating is carried out for neighborhood nodes. The k-means clustering algorithm is employed for grouping similar nodes of Modified SOM into k clusters using similarity measures. Performance of the new approach is evaluated with standard KDD cup99 bench mark dataset. The new approach is evaluated using performance metrics such as detection rate and false alarm rate. The modified SOM is used to create the network with the help of distances threshold, connection strength and neighborhood functions and k-means clustering algorithms groups the nodes in the network with the help of similarity measures. The modified self-organizing map has improved 2% higher detection rate compared to the existing SOM but when k-means is deployed it is further increased by 1.5%. It starts with null network and gradually evolves with original data space. The updating of neighborhood function has been improved with the help of connection strength. The learning rate is found to plays the vital role by spreading the map as observed when the learning rate increases the number of output nodes decreases. In particular, the proposed work is found to effective for detecting DOS attacks with 98.5% detection rate.

Reference [14] proposed a flow-based anomaly detection system. Artificial Neural Network (ANN) is an important approach for anomaly detection in their study. They used a Multi-Layer Perceptron (MLP) neural network with one hidden layer and investigate the use of a Gravitational Search Algorithm (GSA) in optimizing interconnection weights of a MLP network. Their proposed GSA-based flow anomaly detection system (GFADS) is trained with a flow-based data set. The trained system can classify benign and malicious flows with 99.43% accuracy and they compare the performance of GSA with traditional gradient descent training algorithms and a particle swarm optimization (PSO) algorithm. The results show that GFADS is effective in flow-based anomaly detection. This study has been done based on centralized processing.

In [32], they developed an alert clustering and classification system which is able to classify IDS alerts and reduces false positive alerts using clustering of genetic algorithms. Alert filtering algorithm were used in their system to improve the accuracy and reduce false positive alerts. Their experiment was performed on dataset of DARPA KDD Cup 98 with two new clustering algorithms. Result were than compared other genetic algorithms and their GA algorithm cluster and classify the alerts with high accuracy and able to reduce the number of false positive alerts considerably.

A. Summary of Literature

IPS, IDS and SIEM even were not closely linked, but the three technologies are usually the security tools will be used by majority of the enterprise in order to know and see what is going on in the network. All these technologies passed an approval stage in R&D and become marketable product to fulfil the demand for the emerging need of security tools. However, the changes and improvement in threats also lead to the demand for evolvement of the security solution. So, now we have IPS technology build into firewall technology as the first line of defense to determine and protect against attacks, however that is no obsolete equipment can provide overall protection for an organization. IPS came from the development of another layer to protect firewalls and desktop anti-virus. A common flaw with signature-based and heuristic engines is that they have milliseconds to make a decision, so attacks are only detected on static code. Today complex enterprise need to see an overall evolution of product layers. it was important to have a holistic view, and a good SIEM that ‘can bring things together' and allow you to manage third party software is crucial. SIEM provide the central visibility with log management that can give multiple use cases across security and performance management in an organization.

SIEM, IPS and IDS are still relevant even not closely linked as they can protect against modern attacks. They evolve as the attacks evolve, which adds credibility to the argument that a second layer is needed to protect networks. With regard to IPS solutions, it is firmly believed that it still serves the purpose. Both IDS and IPS are able to intelligently add value into a SIEM solution by analyzing the network traffic and alerting on detected patterns of attack. Technologies such as SIEM, IPS and IDS are all key to securing their own part of the enterprise These technologies is still required to detect or prevent most attacks, but they must be integrated to provide a coherent and contextual view, not lots of large volume of standalone bits and pieces without centralize view of the overall detections in the enterprise.

A SIEM on its own is useless because it has no ability to monitor the raw security events as they happen throughout the enterprise. SIEMs are designed to use log data as recorded by other pieces of software. There is limitation in term of anomalies detection by a correlation engine in IPS or IDS. This is where new technology comes into play. Correlation can detect incidents that poses threat to your enterprise, whereas anomaly detection can detect incidents that belong to zero days' categories. This capability is a breakthrough to the issues previously found in traditional IPS/SIEM/log management solutions. By providing a self-learning technology that uses the taxonomy of the incoming events to build a baseline, modern SIEM solutions can automatically detect deviations from the baseline in order to make alert users for any new, unknown threats. In order to remain competitive, the SIEM providers must deliver these advanced capabilities alongside ease-of-use, which will definitely create a change within the SIEM market. Table 1 below are the summarized general functionality provided by of IDS, IPS which indicate the SIEM does have some advantage over the other solution.

Table 1: General functionality of IDS, IPS and SIEM

IDS	IPS	SIEM
Detection mode only	Active traffic control	Big volume of log collection – Forensics & Retention
Traffic replication required	"Original traffic required"	IT Compliance
Decoupling detection and reaction functionalities	Detection and reaction support	enable centralized analysis and reporting for an organization's security events
IDS as a good assistant for network administration	No administrator assistance needed	Event Correlation -Improve the efficiency of incident handling activities.
Usually used for testing rules	Require strict configuration	User Activity Monitoring

4. FUTURE RESEARCH IN INTRUSION DETECTION FOR NGSOC

Based on the previous summary from the literature survey. It is obvious that SIEM can give a wider view of the network environment within an enterprise. SIEM is a solution that provides a comprehensive view of security posture of an IT infrastructure and it is the core component for all the Security Operation Center (SOC) to correlate events and send alerts to the system owners to take action if found any intrusion. Two key objectives of SIEM is to detect security incidents in near real time and managing logs. SIEM collects information such as logs, events and network flows from various devices on a network, correlates and analyzes the data to detect incidents and abnormal patterns of activity. When successfully deployed and configured, a SIEM helps organizations:

- Discover internal/external threats.
- Monitor (privileged) user activity and access to resources.
- Provide compliance reporting.
- Support incident response.

SIEM system use various conditions to check whether certain events are matching a rule, and depending on the threat, an alarm can be triggered. Typically, there 3 main categories of conditions:

a) Event Based: Y assets log an event targeted at host Y and vulnerability scanner knows that Y is vulnerable. It triggers an alert.
b) Rule Based: If Log X + Log Y + Log Z then create event F, plus if F repeats more than 5 times in 60 seconds. It triggers an alert.
c) Anomaly Based: If the event F exceeds the standard deviation of historic traffic patterns then trigger an alert in cases like malware and DDos.

Once alert is triggered, the team will investigate the alert, escalate the incident to the respective team for remediation. The task to create rules are commonly executing by threat analyst who on continuous basis hunting for indicator of compromise by analyzing threat intelligence feed. One approach to causal event correlation is to apply logical rules that chain together events based on their relevant attributes. But there are several problems with rule-based approaches to event correlation. It can be difficult for complex rule systems to keep pace with online streams of events and maintaining the rule sets needed for constructing attack scenarios from disparate events can be difficult. Also, missing events can prevent rules from assembling a proper attack scenario and attempts at inferring hypothetical missing attacks can lead to irrelevant results [10].

a) Current algorithms and methods in the correlator produces higher false positive events.
b) Current algorithms and methods in the correlator need longer time when doing the event correlation analysis.
c) Current available algorithms and methods in the correlation methodologies are hard to detect abnormal event pattern in isolation.

To further improve anomaly detection mechanism while reducing the false alarm rate in SIEM, further research shall be carried out to develop new incident detection method that can improve the accuracy and reduce false alarm rate when compare to earlier research.

5. CONCLUSION

This paper studied the prior researches that been carried out to improve the incident detection in the IDS, IPS and SIEM. Various work been carried out for numerous types of incidents detection methods in IDS, IPS and SIEM. However, there are little work been done for integrating some well-established incident detection method that can be apply for SIEM. So, this paper will provide fundamental input for type of algorithms and model to be further developed and implement with the SIEM tools available in the market. This new method will be developed as a middleware to integrate with SIEM to complete the NGSOC (Next Generation Security Operation Centre) concept.

REFERENCES

[1] J. M., Estevez-Tapiador, Garcia-Teodoro, and Pand J. E. Diaz-Verdejo, "Anomaly Detection Methods in Wired Networks: A Survey and Taxonomy," Computer Communications, vol. 27, pp. 1569-1584, 2004.

[2] Denning D.E, “An Intrusion Detection Model,” IEEE 1986.

[3] Mahoney. MV and Chan P.K., “Learning Nonstationary Models of Normal Network Traffic for Detecting Novel Attacks.” SIGKDD, 2002.

[4] Anderson, Debra, Teresa F. Lunt, Harold Javitz, Ann Tamaru, Alfonso Valdes, "Detecting unusual program behavior using the statistical component of the Next generation Intrusion Detection Expert System (NIDES)", Computer Science Laboratory SRI-CSL 95-06 May 1995. http://www.sdl.sri.com/papers/5/s/5sri/5sri.pdf

[5] Barbará, D., N. Wu, S. Jajodia, "Detecting Novel Network Intrusions using Bayes Estimators", First SIAM International Conference on Data Mining, 2001, http://www.siam.org/meetings/sdm01/pdf/sdm01_29.pdf

[6] Roesch, Martin, "Snort - Lightweight Intrusion Detection for Networks", Proc. USENIX Lisa '99, Seattle: Nov. 7-12, 1999.

[7] Neumann, P., and P. Porras, "Experience with EMERALD to DATE", Proceedings 1st USENIX Workshop on IntrusionDetection and Network Monitoring, Santa Clara, California, April 1999, 73-80, http://www.csl.sri.com/neumann/det99.html

[8] Helman,P., Bhangoo, J. “A Statistically Base System for Prioritizing Information Exploration Under Uncertainty,” IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans, 27(4):449-466, 1997.

[9] Lane, T., Brodley, C. E., “Sequence Matching and Learning in Anomaly Detection for Computer Security,” AAAI Workshop: AI Approaches to Fraud Detection and Risk Management, 43-49, 1997.

[10] Günes Kayack, H., Nur Zincir-Heywood, A., and Heywood,M. I., "Selecting Features for Intrusion Detection: A Feature Relevance Analysis on KDD 99 Intrusion Detection Datasets." Dalhousie University, Faculty of Computer Science, 6050 University Avenue,
Halifax, Nova Scotia. B3H 1W5

[11] Razaq,A., Tianfield, H., and Barrie,P., “A Big Data Analytics Based Approach to Anomaly Detection” in 2016 IEEE/ACM 3rd International Conference on Big Data Computing, Applications and Technologies

[12] Ahmad.B., “Intrusion Detection With Tree-Based Data Mining Classification Techniques BY Using KDD DataSet.” European Journal of Computer Science and Information Technology Vol.5, No.6, pp.11-18, December 2017

[13] Aneetha. AS., and Bose. Dr. S., “The Combined Approach For Anomaly Detection Using Neural Networks And Clustering Techniques.” Computer Science & Engineering: An International Journal (CSEIJ), Vol.2, No.4, August 2012

[14] Jadidi, Z., Muthukkumarasamy, V, and Sheikhan M., “Flow-Based Anomaly Detection Using Neural Network Optimized with GSA Algorithm.” 2013 IEEE 33rd International Conference on Distributed Computing Systems Workshops

[15] Alsabbagh, B., and Kowalski,S., “A Framework and Prototype for A Socio-Technical Security Information and Event Management System (ST-SIEM)” in 2016 European Intelligence and Security Informatics Conference

[16] Ficco, M. and Romano, L.,"A Generic Intrusion Detection and Diagnoser System Based on Complex Event Processing." in 2011 First International Conference on Data Compression, Communications and Processing.

[17] Kowalski,S., Barabanov,R., and Hoffmann, R., “ Cyber Security Alert Warning System:A socio-techinal coordinate system proposal” 2012 Third International Workshop on Security Measurements and Metrics

[18] Kotenko, I., Chechulin A., “Computer Attack Modeling and Security Evaluation based on Attack Graphs.” The 7th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications 12-14 September 2013,
Berlin, Germany

[19] S.A. Hofmeyr, S.Forrest, A.Somayaji” Intrusion Detection using Sequences of System Calls.” Journal of Computer Security, vol. 6, no. 3, pp. 151-180, 1998.

[20] https://www.webopedia.com/TERM/I/intrusion_detection_system.html

[21] M Tavallaee, E Bagheri, W Lu, AA Ghorbani, A Detailed Analysis of the KDD CUP 99 Data Set, in Proceedings of the IEEE Symposium on Computational Intelligence in Security and Defense Applications (Ottawa, 2009)

[22] KDD, KDD Cup, http://www.kdd.org/kdd-cup/view/kdd-cup-1999, Accessed April 2018

[23] SJ Stolfo, W Fan, W Lee, A Prodromidis, PK Chan, Cost-Based Modeling for Fraud and Intrusion Detection: Results From the JAM Project, in Proceedings of the 2000 DARPA Information Survivability Conference and Exposition (DISCEX '00), Hilton Head, SC,
2000

[24] RP Lippmann, DJ Fried, I Graf, JW Haines, KR Kendall, D McClung, D Weber, SE Webster, D Wyschogrod, RK Cunningham, MA Zissman, Evaluating intrusion detection systems: the 1998 DARPA off-line intrusion detection evaluation, Hilton Head, 25–27 January 2000, vol 2 (IEEE, Amsterdam, 2000), pp. 10–12 21.

[25] MIT Lincoln Labs, DARPA Intrusion Detection Evaluation, 1998.
http://www.ll.mit.edu/mission/communications/ist/corpora/ideval/index.html, Accessed April 2018

[26] S.Asanger and A.Hutchison. “Experiences and Challenges in Enhancing Security Information and Event Management Capability using Unsupervised Anomaly Detection.” 2013 International Conference on Availability, Reliability and Security, pp 654-661.

[27] D. Gupta., S. Singhal., S. Malik., and A. Singh., “Network Intrusion Detection System Using various data mining techniques.” International Conference on Research Advances in Integrated Navigation Systems (RAINS - 2016), April 06-07, 2016, R. L. Jalappa Institute of Technology, Doddaballapur, Bangalore, India.

[28] NIST National Vulnerability Database (NVD). http://nvd.nist.gov/ Accessed April 2018.

[29] G.S.-Tangil, E.Palomar, A.Ribagorda, and I.Sanz., “Providing SIEM systems with self-adaptation.” Information Fusin 21 (2015) pp 145-158

[30] [30] F.Alserhani. “ A framework for Multi-stage Atrack Detection.” Conference: Electronics, Communications and Photonics Conference (SIECPC), 2013 Saudi International

[31] H. Bahrbegi., A.H. Navin., A.A.A. Ahrabi., M.K. Mirnia.,and A.Mollanejad., “ A New System to Evaluate GA-based Clustering
Algorithms Algorithms in Intrusion Detection Alert Management System. Second World Congress on Nature and Biologically Inspired Computing Dec. 15-17,2010 in Kitakyushu, Fukuoka, Japan 2010

[32] C.Abad, Jed. T., Cigdem. S., William. Y., “Log Correlation for Intrusion Detection: A Proof of Concept.” 19th Annual Computer Security Applications Conference (ACSAC), Pp 255-264. 2003