Intrusion Alert Correlation Technique Analysis for Heterogeneous Log

Robiah Yusof[†], **Siti Rahayu Selamat**[††], **Shahrin Sahib**[†††]

Faculty of Information Technology and Communication,
Universiti Teknikal Malaysia Melaka, Ayer Keroh, Melaka, Malaysia

Summary
Intrusion alert correlation is multi-step processes that receives alerts from heterogeneous log resources as input and produce a high-level description of the malicious activity on the network. The objective of this study is to analyse the current alert correlation technique and identify the significant criteria in each technique that can improve the Intrusion Detection System (IDS) problem such as prone to alert flooding, contextual problem, false alert and scalability. The existing alert correlation techniques had been reviewed and analysed. From the analysis, six capability criteria have been identified to improve the current alert correlation technique. They are capability to do alert reduction, alert clustering, identify multi-step attack, reduce false alert, detect known attack and detect unknown attack.

Key words:
IDS, Alert correlation, Heterogeneous log, capability criteria

1. Introduction

Computer security offers three types of security mechanism to protect a system which are authentication, authorisation and auditing. These three mechanisms are essential for securing the systems against attack. However, if the design and implementation of these mechanisms are flaws, an additional protection is needed. In order to provide an extra layer of defence, IDS have been proposed. Intrusion detection technology has gained increasing acceptance in enterprise networks, with both commercially supported and open source components widely deployed [1]. However, it has few weaknesses such as prone to alert flooding, contextual problem due to attacks are likely to generate multiple related alert, false alert and scalability [4] and correlation is proposed to overcome these weaknesses.

Devices and sensor diversity has resulted in a new difficulty in generating reports due to the unmanageable amount of alert. Inspecting thousand of alerts per day is not feasible, especially if 99% of them are false positives [8]. Referring to Fig. 1, in view of domain perspective, heterogeneous log resources can be gained from host, network, application, sensor and wireless. It can be argued that diverse intrusion detection sensor and heterogeneous log resources provide more complete coverage of the attack space. However, the potential added leverage from devices diversity is not clear that require the security administrator's decision on which reports pertains to the same or to different incidents. This issue has motivated the researcher to do research on correlation of alert produce by intrusion detection sensor and heterogeneous log resources.

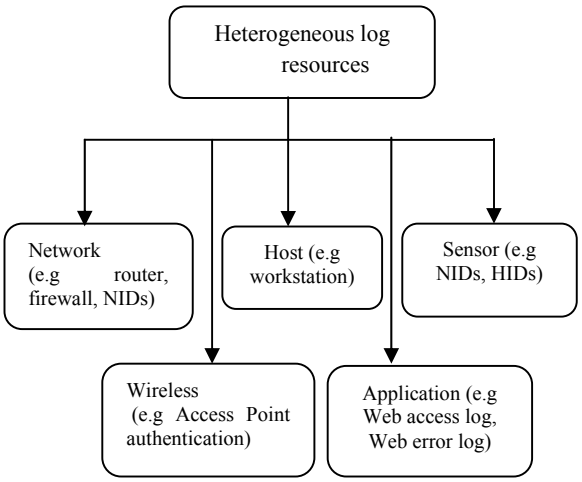

Fig. 1 Domain perspective of heterogeneous log resources

2. What is Alert and Intrusion Correlation?

According to [7], alert is defined as an alarm generated by Intrusion Detection system (IDS), to notify interested parties of interesting event. An event is a low level entity analysed by IDS. Single event can cause multiple alerts and it can be represented in mathematical expression as below:

$$\text{Event} = \{\text{alert}_1, \text{alert}_2, \text{alert}_3 \ldots\ldots \text{alert}_n\}$$

Intrusion correlation refers to the interpretation, combination and analysis of information from all available sources about the target system activity for the purpose of intrusion detection and response. There are two types of intrusion correlation as in Fig. 2; intrusion event correlation and intrusion alert correlation [6].

Manuscript received September 5, 2008.
Manuscript revised September 20, 2008.

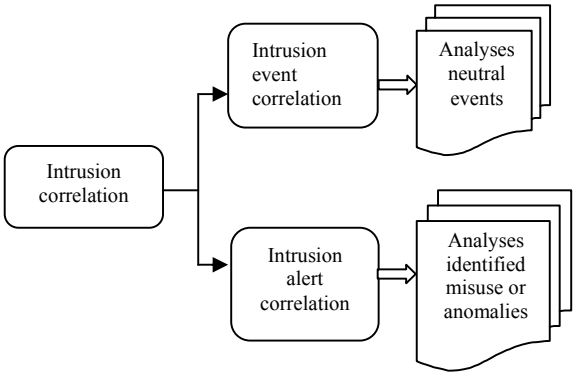

Fig. 2 Types of Intrusion Correlation

The main difference between these two types of intrusion correlations is that intrusion event correlation analyses neutral events, meanwhile intrusion alert correlation analyses identified misuse or anomalies. This relation is also stated in IDMEF specification [3] specifying that each time an analyzer detect an event that match the rule, it sends an alert to its manager(s). It will depend on the analyser as an alert message can correspond to a single detected event, or multiple detected events. Intrusion event correlation is important for forensic investigation whereby intrusion alert correlation is important for security administrator and this study will concentrate on intrusion alert correlation

2.1 Intrusion Alert Correlation

Alert can be produced from various types of sources and it may cause multiple stages of attack. Alert correlation is multi-step processes that receives alerts from one or more IDS as input and produce a high-level description of the malicious activity on the network. According to [7], to achieve good recognition, the data needs to be collected from various sources (for example firewall, web server logs, IDS of multiple manufacturers and so on). Correlation of alerts produced by heterogeneous log resources can provide a number of potential advantages and the most obvious benefit is the reduction in the number of alerts that a security officer must address.

When considering single log resource, a correlation engine can reduce alert volume by organizing numerous alerts that are part of an ongoing attack, namely known as alert threading. In the case of heterogeneous log resources, a correlation engine should recognize when reports from multiple log resources refer to the same incident. Correlation is the degree to which one or more attributes or measurements on the same group of elements show a tendency to vary together. Correlation can enhance detection capability, and give a more complete picture of attacks that an individual sensor or devices may observe only partially without losing the security-relevant information. In addition, correlation can exploit the complementary coverage from several log resources. Reports from several log resources employing diverse analytical techniques may reinforce each other and therefore enhance the confidence of the detection.

3. Classification of Alert Correlation Technique

There are three famous techniques exist in correlating alerts which are Similarity-based, Pre-defined attack scenarios and Pre-requisites and consequences of individual attack [14]. In addition, [12] has introduced one more technique for alert correlation, known as Statistical causal analysis (refer to Fig. 3).

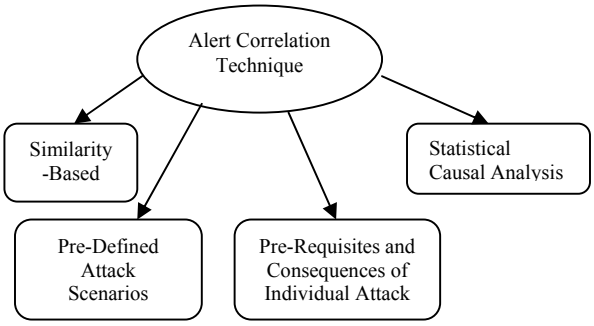

Fig. 3 Classification of Alert Correlation Technique

3.1 Similarity-Based

Similarity-based technique will compare an alert to all alert threads that have similar attributes or features (e.g source IP address, destination IP address, ports) and then correlates alerts with a high degree of feature similarity if match or a new thread is created if none is match [9], [13].

[13] has implemented alert similarity metric using Emerald in three phases as in Fig. 4. In the first phase, the low-level events is aggregated using the attack threads concepts to cluster alert that are part of the same ongoing attack. The alerts are clustered if attribute is overlap which mean that it will only consider attributes that present in both alerts.

The metric for this phase demand that sensor field, attack class, attack name, source and target in both alerts are similar. In second phase, different levels of similarity are expected for different attributes in different situation whereby similar sensor field is dropped and similar alert name is maintained. This phase is to ensure that detection of the same attack by multiple sensors should be fused.

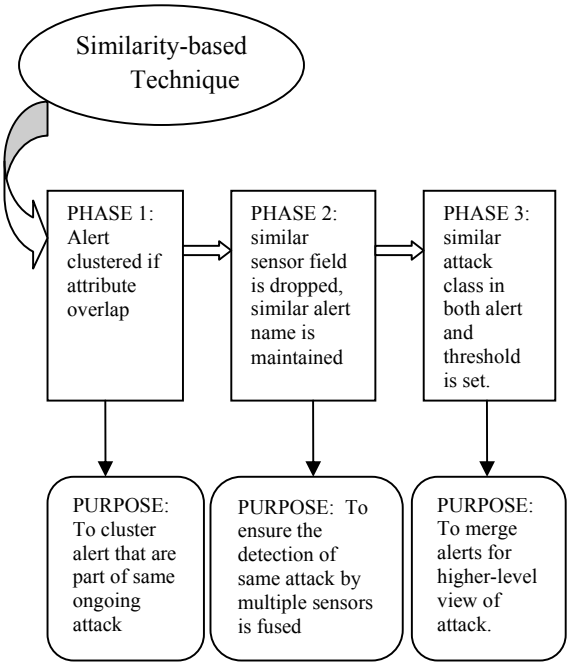

Fig. 4 Similarity-based intrusion alert correlation process by Valdes & Skinner

Then in third phase, it requires similar attack class in both alert. Certain threshold is adjust for example for synthetic threads, sensor id and IP is set high and for multistep attack detection, threshold for attack class is set to low. This phase will merge alerts representing different attack steps to provide a higher-level view of the security state of the system.

3.2 Pre-Defined Attack Scenarios

This technique utilizes the fact that intrusions often require several actions to take place in order to succeed. Every attack scenario has corresponding steps required for the successfulness of the attack. Low- level alerts from IDS are compared against the pre-defined attack scenario before the alerts can be correlated. It is restricted to known attack and misuse detection only and specified by human users or learned through training datasets. The example of implementation is ASL (Attack Specification Language).

[4] has presented a detailed semantic alert model and developed adapter modules to map proprietary alert formats into this model. This alert model was further refined and is now the de-facto format for intrusion detection alerts known as Intrusion Detection Exchange Format (IDMEF). They have proposed a system that performs correlation and aggregation of intrusion detection alerts produced by various sensor as in Fig. 5. In correlation phase, there are two types of correlation which are *duplicate removal* and *consequence correlation*.

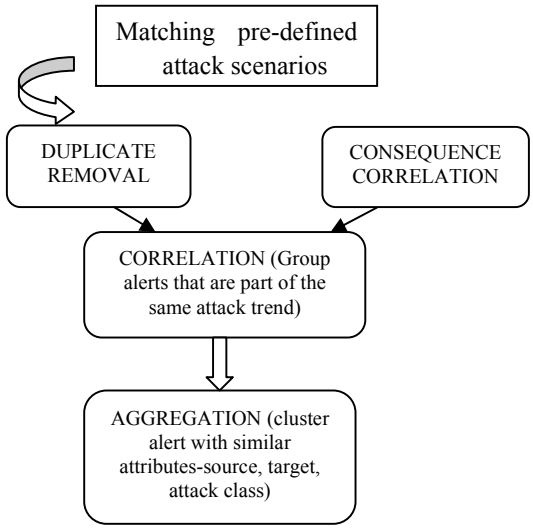

Fig. 5 Pre-defined attack scenarios intrusion alert correlation process by Debar & Wespi

Duplicates are instances of the same attack as detected using rules read from a specified configuration file by two different sensor. Consequences are rules that specify one event should be followed by another type of event. It will link together alerts that are sequential in nature. Once alerts has correlated, aggregation phase will cluster alerts with similar attributes (source, target and attack class). It identifies the source of the attack, the target of the attack and popular attack class. It will group alerts based on certain criteria to aggregate severity level, reveal trends and clarify attacker's intentions. This phase consists of large number of false positive, however there is no specific technique can eliminate this problem. Major weakness of this method are, it requires that human users specify attack scenarios and it is limited to detection of known attacks.

3.3 Pre-Requisites and Consequences of individual attacks

This technique work at a higher level than correlation based on attributes similarities, but a lower level then correlation based on known scenarios. Pre-conditions are defined as the necessary conditions that must exist for the attack to be successful, and the consequences for the attack are defined as conditions that may exist after a specific attack has occurred. It is represented as a logical formula using AND and OR connectives. This technique is not restricted to known attack scenarios and its uncover the causal relationship between alerts. Most of the approaches using this method are focused on the

modeling and detection of multistep attacks to provide a high-level of the attack associated with a security compromise.

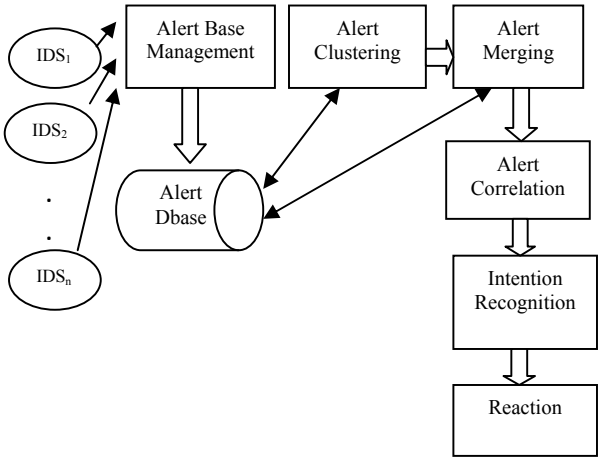

Fig. 6 Pre-requisite and Consequences of individual attack intrusion alert correlation process by Cuppens & Miege

By using these techniques as depicted in Fig. 6, [2] has include five functions including alert base management, alert clustering, alert merging, alert correlation and intention recognition function. In alert base management function, it receives alerts generated by IDS and stores them for further analysis by cooperation module. This alert will be normalized to IDMEF format and store in the relational database. Alert cluster and alert merging function can access the database and will use the similarity function to cluster and merge the alert. Alert correlation function will further analyzes the cluster alerts provided by alert merging function using explicit correlation rules with pre-defined and consequence statement. This approach attempt to generate correlation rules automatically which can introduce correlated alerts that are similar by chance and this could increase the noise in the alert stream.

[10], [11] has implemented causal relationships between alerts using pre-requisite and consequences . Hyper-alerts connected graph are created and graph-manipulation techniques are applied to reduce these connected alerts into manageable alert. This approach can correlate alerts in the middle of attack chain even if the start of the chain is missed. This can be useful in the case of intruder can pass through the access-list of router unnoticable and easily attack the server in the LAN. It can correlate the second and third attack even if the pre-conditions of the second alert were not met. This system can generate graph which is useful in determining the attacker's objective.

3.4 Statistical Causal Analysis

This technique by [12] as shown in Fig. 7, implements anomaly detection and use Granger Causality Test (time series analysis method) to correlate events which emphasis on attack scenario analysis. In order to reduce the volume of raw alerts, it will combine low-level alert based on alert attributes. It uses clustering technique to process low-level alert-data into high-level aggregated alerts. Prioritization alerts is used based on relevance of attacks and impacts on the mission goal. It will then conduct causality analysis to correlate alerts and constructs attack scenario. It is pure statistical causality analysis and does not need pre-defined knowledge about attack scenarios. Hence, new attack scenarios can be identified.

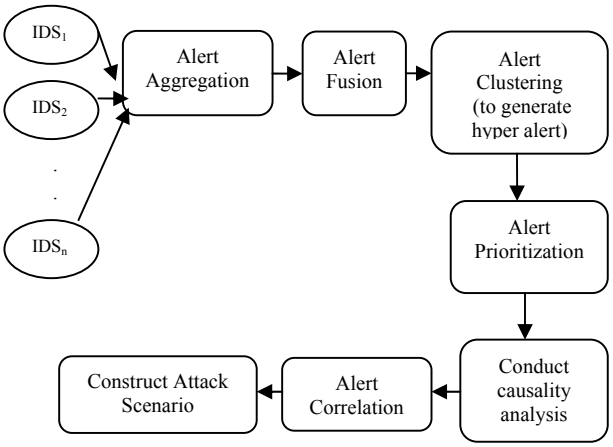

Fig. 7: Statistical Causality Analysis alert correlation process by Qin & Le

This technique declares that every multi-step attack will generate alert that have statistical similarities in their attributes, and this attack steps have causal relationship. [12] run the statistical correlation engine offline with training datasets to compute and store correlations so that it can be used for pattern matching at run-time. This technique is not a feasible solution for the complete correlation process. However it can be utilized as a part of a larger system to pre-process alerts or to provide meta-alert signatures.

4. Discussion and Analysis of Alert Correlation Technique

4.1 Alert Correlation Techniques Comparison

All techniques have their advantages and disadvantageous as summarized in Table 1. Referring to this table, further analysis has been done in order to produce the capability criteria.

Table 1: Alert Correlation Technique Comparison

Technique	Advantage	Disadvantage
Similarity-based	• Can reduce large number of redundant alert generated by multiple sensors.	• False alert can only be detected if multiple sensors can detect the same attack. • Cannot detect multi-step attack • Can detect only misuse detection and not anomaly detection.
Pre-defined attack scenario	• Can reduce large number of redundant alert generated by multiple sensors • Can cluster multiple related alert (contextual alert) • Can detect precise attack as stated in the rules (specification).	• Could generate large number of false positive alarm [5] • it requires that users specify attack scenarios manually • It is limited to detection of known attacks or misuse detection and not anomaly detection. • multi-step attack alert is disregarded
Pre-requisites and Consequences of individual attack	• Multi-step attack can be detected to provide a high-level view of the attack associated with a security compromise • [10] generate useful graph to determine the attacker's objective	• Automatic generation correlation rules can generate large false alarm [2].
Statistical Causality Analysis	• does not need pre-defined knowledge about attack scenarios. • Using anomaly detection technique • new attack scenarios can be identified • can be used as pre-process alerts or meta-alert signatures.	• Not feasible for complete correlation process [7].

4.2 Proposed Criteria for Alert Correlation Technique

Four important intrusion alert correlation techniques has been reviewed and analysed. The objective of this study is to analyse the current alert correlation technique and identify the significant criteria within each technique which can improve the IDS problem. As mentioned by [4], IDS has developed issues on alert flooding, contextual problem, false alert and scalability. The characteristic that shall be analysed in each alert correlation technique is according to the issue listed in Table 2.

Table 2: Issues analysed in IDS

No	IDS Issue	Description	Propose Solution
1	Alert flooding	IDS are prone to alert flooding as they provide a large number of alerts to the security officer, who then has the difficulties coping with the load.	1. To reduce number of alert generated from IDS.
2	Contextual problem	Attacks are likely to generate multiple related alerts. Current IDS do not make it easy for security officers to logically group related alerts.	1. To group or cluster alert which has a related event or event threaded. 2. To identify multi-step attack
3	False Alert	Existing IDS are likely to generate false positives or false negatives alerts	1. To reduce number of false alerts 2. To identify known attack using misuse detection 3. To identify unknown attack using anomaly detection

Based on the IDS current issues, most of the current alert correlation techniques were developed in order to improve this problem. Therefore the proposed criterion that shall be analysed is according to the capability criteria as listed below:

1. Capability to do alert reduction
2. Capability to do alert clustering
3. Capability to identify multi-step attack.
4. Capability to reduce false alert.
5. Capability to detect known attack
6. Capability to detect unknown attack

Alert reduction is required in order to overcome the problem of alert flooding or large amount of alert data generated by the IDS. This capability criterion is important in order to reduce the security officer's tension in performing troubleshooting when analysing the exact attacker in their environment.

For second criteria, alert clustering is considered as important as it can cluster multiple related alerts and at the same time reduce the number of alert by ignoring the similar alert generated by different sensors or heterogeneous log resources.

The third criteria, most of the alert correlation technique is incapable to detect multi-step attack. Therefore this capability is required as attacker behaviour is becoming more sophisticated and it shall involve one to many, many to one and many to many attacks.

For fourth criteria, most of the IDS have the tendency to produce false alarm. This false alarm reduction criterion is important as it closely related to alert flooding issue.

The fifth and sixth criterion, the capability to detect both known and unknown attack is required to ensure that the alert generated will overcome the issue of alert flooding and false alert. Table 3, is the summary of analysis from each alert correlation techniques match with the proposed capability criteria.

Based on the analysis, all of the techniques have the same capability to reduce and cluster the alert, and detect known attack except for pre-requisite and consequences of individual attack and statistical causality technique. The additional capability of pre-requisite and consequences of individual attack is to identify multi-step attack whereas the unknown attack can be detected using statistical causality technique.

The analysis also shown that, most of the researchers [7],[5],[2] identified that all of these techniques are incapable to reduce false alert. This has given an implication that there is a room for improvement in detecting known and unknown attack, and multi-step attack as these capability criteria shall overcome large number of false alert problem. Based on the study, a proposed an improved solution for alert correlation technique will use combination of all techniques to complement each other weakness. For example, in the alert correlation process for pre-requisites and consequences of individual attack technique in Fig. 6, the alert clustering will use the similarity-based and pre-defined attack scenarios technique.

Table 3: Alert correlation technique versus proposed capability criteria. (capable=√, incapable=×)

Technique Name	Alert Reduction	Alert Clustering	Multi-step Attack	Reduce False Alert	Detect Known Attack	Detect Unknown Attack
Similarity-based	√	√	×	×	√	×
Pre-defined Attack Scenarios	√	√	×	×	√	×
Pre-requisites and consequences of individual attack	√	√	√	×	√	×
Statistical Causality	√	√	×	×	√	√

5.0 Conclusion and Future Works

In this study, researcher have reviewed and analysed the existing alert correlation technique to overcome the IDS's problems discussed. From the analysis researcher propose an improved solution for alert correlation technique based on six capability criteria identified which are capability to do alert reduction, capability to do alert clustering, capability to identify multi-step attack, capability to reduce false alert, capability to detect known attack and capability to detect unknown attack. According to the capability criteria, all alert correlation techniques should be implemented in the alert correlation process. Further improvement should be done on the process of detecting the known and unknown attack, and multi-step attack as these capability criteria shall overcome large number of false alert problem. Therefore further research on intrusion alert correlation technique with these capabilities to detect unknown attack using combination of anomaly and misuse detection approach is required.

References

[1] Anderson, D., Fong, M., & Valdes, A. (2002). Heterogeneous Sensor Correlation: A Case Study of Live Traffic Analysis. *Proceeding 3rd Annual IEEE Information Assurance Workshop.*

[2] Cuppens, F., & Miege, A. (2002). Alert Correlation in a Cooperative Intrusion Detection Framework. *Proceedings of IEEE Symposium Security and Privacy.*
[3] Curry, D., & Debar, H. (2007, March). *Intrusion Detection Message Exchange Format.* Retrieved July 7, 2008, from IETF: http://www.rfc-editor.org/rfc/rfc4765.txt
[4] Debar, H., & Wespi, A. (2001). Aggregation and Correlation of Intrusion Detection Alerts. *Proceedings of the International Symposium on Recent Advances in Intrusion Detection*, (pp. 85-103). Davis, CA.
[5] Valeur,F., (2006). *Real-time Intrusion Detection Alert Correlation.* PhD Thesis, University of California Santa Barbara, USA.
[6] Gorton, D. (2003). *Extending Intrusion Detection with Alert Correlation and Intrusion Tolerance.* MPhil Thesis , Chalmers University of Technology, Department of Computer Engineering, Goteborg, Sweden.
[7] Hattala, A., Sars, C., Addams, R., & Virtanen, T. (2004). Event Data Exchange and Intrusion Alert Correlation in Heterogeneous Networks. *8th Colloqium for Information Systems Security Education.* West Point, New York.
[8] Julisch, K. (2003). Clustering Intrusion Detection Alarms to Support Root Cause Analysis. *ACM Transactions on Information and System Security 6(4), ACM Press* , pp. 443-471.
[9] Julisch, K. (2001). Mining Alarm Clusters to Improve Alarm Handling Efficiency. *Proceedings of the 17th Annual Conference on Computer Security Applications.* New Orleans, LA.
[10] Ning, P., Cui, Y., & Reeves, D. (2002). Analyzing Intensive Intrusion Alerts via Correlation. *Proceedings of the International Symposium on the Recent Advances in Intrusion Detection*, (pp. 74-94). Zurich, Switzerland.
[11] Ning, P., Cui, Y., & Reeves, D. (2002). Constructing Attack Scenarios through Correlation of Intrusion Alerts. *Proceedings of the ACM Conference on Computer and Communications Security*, (pp. 245-254). Washington D.C.
[12] Qin, X., & Le, W. (2003). Statistical Causality of INFOSEC Alert Data. *Proceedings of Recent Advances in Intrusion Detection.*
[13] Valdes, A., & Skinner, K. (2001). Probablistic Alert Correlation. *Proceedings of the Recent Advances in Intrusion Detection (RAID).* Davis, CA.
[14] Zhai, Y., Ning, P., & Xu, J. (2005). *Integrating IDS alert correlation and OS-level dependency tracking.* North Carolina : North Carolina State University.

Robiah Yusof is currently a PhD student at the Universiti Teknikal Malaysia Melaka, Malaysia. She holds Bachelor of Computer Studies (Hons) from Liverpool John Moore's University, UK and a Master degree in Computer Science with honours from the Universiti Kebangsaan Malaysia, Malaysia. Her research interests include intrusion detection, network security, penetration testing and network forensic.

Siti Rahayu Selamat is currently a PhD student at the Universiti Teknikal Malaysia Melaka, Malaysia. She holds Bachelor of Computer Science (Hons) from Universiti Teknologi Malaysia, Malaysia and a Master degree in Computer Science with honours from the Universiti Malaya, Malaysia. Her research interests include network forensic, intrusion detection, network security and penetration testing

Shahrin Sahib received the Bachelor of Science in Engineering, Computer Systems and Master of Science in Engineering, System Software in Purdue University in 1989 and 1991 respectively. He received the Doctor of Philosophy, Parallel Processing from University of Sheffield in 1995. He is a professor and Dean of Faculty of Information Technology and Communication at the Universiti Teknikal Malaysia Melaka. His research interests include network security, computer system security, network administration and network design. He is a member panel of Experts National ICT Security and Emergency Response Center and also Member of Technical Working Group: Policy and Implementation Plan, National Open Source Policy