HETEREGENEOUS PARAMETERS FOR ACCURACY THREAT

DERIS STIAWAN, MOHD. YAZID IDRIS, ZOHAIR IHSAN, KHALID HUSSAIN, ABDUL HANAN ABDULLAH

Faculty of Computer Science & Information System,

Universiti Teknologi Malaysia, 81310, Johor Bahru, Malaysia

deris@unsri.ac.id, yazid@utm.my, izohair2@live.utm.my, hkhalid2@live.utm.my, hanan@utm.my

ABSTRACT

In this paper, we propose a new approach for increasingly accuracy and precision detection/prevention threat and exploratory phase correlation event in parameters to comprehensive scanning with multiple threat collection. First, we collect data from parameters which we observed previously. Second, find the best method through literature review to integrate the different parameters of collecting information from different structure, label, and variable of data. Finally, we present parameters threat correlation using data mining approach to respond and protect against threat more quickly and updated. Expectation from this research is to obtain a recognition mechanism suspicious threat to be identified before entering and damaging the network.

Keywords: *Parameters Security Violation, Behavior-Based Detection, Intrusion Prevention System, Misuse-Based Detection.*

1. INTRODUCTION

Currently, IDS technologies are not very effective against prediction a new mechanism of attack. There are several limitations, such as performance, flexibility, and scalability. Intrusion Prevention System (IPS) is a new approach system to defense networking systems, which combine the technique firewall with that of the Intrusion Detection properly, which is proactive technique, prevent the attacks from entering the network by examining various data record and detection demeanor of pattern recognition sensor, when an attack is identified, intrusion prevention block and log the offending data. Proposal work in IPS filed by [1], they describe IPS uses to secure the system, the enterprise uses several technology security systems, and almost 54% of them use intrusion prevention to mitigation and defense from threat and attack, and work by [2], they had developed host intrusion prevention to block malicious traffic by applying behavioral techniques that focus on object behavior of threat.

The main contributions this paper is the enhancement of a learning phase and is part of the research have being done [3] and [4], which aim to increasingly accuracy alarm in detection and prevention system. The remaining of the paper is structured as follows: In Section 2 we present and briefly discuss background and related work. Section 3 propose our correlating parameters approach. Section 4, discusses learning process. Section 5 summarized our conclusions and present additional issues on which research can be continued.

2. RELATED WORK

According to annual report CSI/FBI 2009, reported increasing the number of type and volume of attack. These results are similar to the survey conducted by CERT 2009, which conduce seriously concern. Intrusion Detection were developed to identify and report attack in the late 1990s, as hacker attacks and network worm began to affect the internet, detect hostile traffic and send alert but do nothing to stop the attacks [5].

Computer system security has become a major concern over the past few years. Attack, threat or intrusions, against computer system and network have become commonplace events, many system device and other tools are available to help counter the threat of these attack. Analyzed from proposal [6] and [7] highlighted currently countermeasure against from security violation, such as (i) firewall, strengthen in implementing executing rules and policy, but firewall can do nothing about attack from inside network and can not clarify behavior or anomaly attack, (ii) anti virus software. Unfortunately, anti virus very limited ability to pattern recognition of new viruses before the anti-program created by corporate, and (iii) Intrusion Detection, only send the alert to trigger after attacked have entered the network, and do nothing to stop attacks.

The fundamental accuracy for identify threat is low False Positive and False Negative between high True Positive and True Negative rates, which is making a system that could be supervised from performance monitoring, measurement, and reporting side in order to be able to monitor the system before fault, so it could proactive fault monitoring early. The mainly problem in sensor are accuracy and timeliness performance identifies threat, as well as sensitivity, to how effective a particular filter was in blocking knowing and unknown threat response. It was measured in term of false positive and false negative. The alert generated by the sensor, which is the situation trigger alarm (valid and invalid but feasible) from the sensor.

The performance of IPS is measured by how well the system can accurately predict and prevent intrusion and low false positive rate in stream network. As resume in proposal [8], [9] and [10], they present two main intrusion prevention techniques: (i) anomaly detection, and (ii) misuse detection. However, intrusion threat usually curious and unpredictable or evolve continuously. Obviously, to distinguished boundary between normal activity and suspicious activity is difficult. In some instances, peek traffic in network can be effect from packet flooding, broadcast storm or worm action. Nevertheless, some application is voracious with bandwidth (i.e. streaming video, games, and peer to peer). Voracious traffic looks like suspicious activity on screen monitoring. Thus, ambiguity is involved during the process of classifying intrusion from normal activities.

Data warehouse to collecting scattered information in routine update regularly from provider or security community. This data can be useful information to be associated with other. The information, increasingly large of volume dataset and multidimensional data has grown rapidly in recent years. The data set includes signature identification, rules, policy, pattern, method attack, URL blacklist, update patch, log system, list variant of virus and regular expression, all this will be collected and labeled to identify attack patterns and can predict it that would occur. These data set bulk in information and growing from community or security services. Therefore, there is a critical need of data analysis system that can automatically analyze the data to classification it and predict pattern attack future trends.

The problem is how to collecting and integrating information from different structure, label, and variable of data. These data set bulk in information and growing from community or security services, in **Fig, 3**. We illustrate the approach that will be used to integrate this parameter. Therefore, there is a critical need of data analysis system that can automatically analyze the data to classification it and predict pattern attack future trends.

3. EXPLORATORY & OUR APPROACH

Review on proposal Mark in 2005 [25], the study is identifying over 2 million time stamped event per day that are: tapping detection from triggered signatures, host sensors, physical sensor, and application triggers, that are aimed to detecting malicious activity. Thus, they use heterogeneous nature approach to collect consisting records which span physical sensor, network level sensor, host sensor, and applications to the data stored in a central database. According to work [20], [26], and [27] they identify and recognized mitigation threat from habitual activity, this is a combining of our approach with signature database and present algorithm to attack prevention uses module connection to update module state information between sensor signature and traffic analysis module. Furthermore, from our preliminary observed there are correlated with each parameter with others depict in **Fig, 1.**

1. **Public DNS Registry**. DNS has become one of the key on the internet, proposal work by Bin in 2010 [11], also [28], they describe performing a comprehensive analysis on DNS traffic and introduce a new choice to take the measurements of the Internet growth. Security violation in DNS has been particular concern previously researchers, refers from work [12], [13], and [14], the correlation are; (i) From registrant, get info detailed to public URL blacklist provider or community, this content information domain, contact, expired date, name server (NS), IP Addressing pointing, DNS root, DNS Label types, and DNS header Flags, (ii) To IANA, store content information database top level domain which detailed IP Address for subsequently followed up in mapping addressing, (iii) Store info to update information server farm, which is rules block,

allow and alert in Firewall mechanism. And (iv) Traffic flow network provider can take advantage of the registrant info to analyze the growth of the world use a domain name.

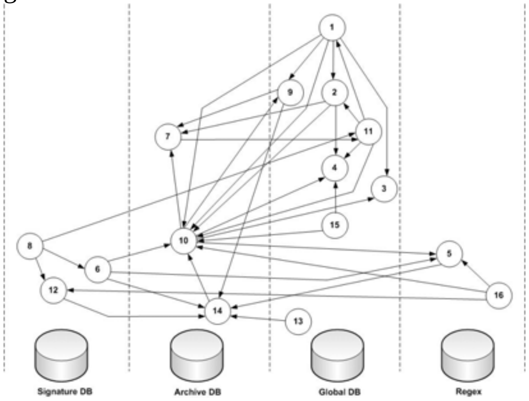

Figure 1: Correlation Parameters

2. **IANA Authority**, According to IANA (http://www.iana.org/protocols), responsibility for maintaining and management numbering of Internet protocols and Internet autonomous system, they provide this service to coordination with the Internet Engineering Task Force (IETF) and Top Level Domain.

3. **Public URL blacklist**. A report from Netcraft (news.netcraft.com), they shows that in August 2009, the total number of websites is over 228 million and there are URLs internet crimes. According to Zhou in 2010 [15], they using URL filtering systems to provide a simple and effective way to protect web security. This concept outplays other filtering methods, such as content scanning and artificial intelligence for its simplicity. The result from experiment [16], they using PhishNet to predicts new malicious URLs from existing blacklist entries. Refers from that, we conclude there are correlation it with rules in server farm configuration and public DNS registry. Furthermore, Security operator can do update rules security management based on this.

4. **Public IP Block lists**. Malicious/ virus like a disease in Internet and IP Address is a unique addressing to delivery packet to destination in over the world. Authorities of IANA to distribute, administration, and management IP Address. Performed work by Wilcox in 2010 [17], they combination of the address study and blacklist allows us to quantify the differences between the IP characteristics of spamming and non-spamming hosts and address prefixes. To summarize the IP visibility data, they using non spamming prefixes exhibit more availability, less volatility, and more uptime than spamming prefixes. Proposal by [36], they called *Nymble* machine to blacklisting a user and notifying the user of blacklist status. Additionally, we combine reports from security communities, suchas:(i) realsecurity.web.officelive.com, (ii) cri.univ-tlse1.fr/blacklists/index_en.php, (iii) www.spamcop.net/bl.shtml, and (iv) spamlinks.net/filter-bl.htm, there are collecting many list of IP bad event. That is very helpful for security operator to update the security rules / policy.

5. **Snort Rules**. Snort, one of the most popular tools in environmental research. Combining the benefits of signature, protocol and anomaly-based inspection. According to [9], Snort is a network intrusion detection system that runs over IP network analyzing real-time traffic for detection of misuse, and subsequently followed by several other researchers. We can identify Concurrent Versions System (CVS) Snort Rules from cvs log (cvs.snort.org/viewcvs.cgi/snort/rules) of snort, there are regularly update rules, event and rules describes in the content update. The correlation with other; (i) Regex, specified using a keyword the keyword PCRE, which stand for Perl Compatible Regular Expression. PCRE is more powerful and complicated, than regex. The regex attribute always contains just a single pattern, not lists of patterns, we can create pcre in regulation expression. Which is snort rule developing in pcre pseudo code, (ii) Log Files, to create cvs a rule in snort is obtained from the log records server and hex decimal from tapping mechanism, and (iii) Alert IDS, IDS produce alert from sensor to recognize and identification pattern attack.

6. **Vulnerability from Common Vulnerability and Exposures**, according to CVE (www.cve.mitre.org), which began in 1999 to adoption of a common naming practice for describing software vulnerabilities and including security tools and service as well as on the fix sites of commercial and open source software package providers. Each month, CVE MITRE receives between 150 until 300 new submissions from ISS, Security Focus, Neohapsis, and the National Infrastructure Protection Center for announcement alert and

advisories, currently CVE identify/compatible enable data exchange between security products and provide a baseline for evaluating coverage of tools and services. Performed work Martin in 2001 [18], describes benefits of CVE compatibility, integrating vulnerability services, tools to provide more complete security provide and alert advisory services. From our preliminary observe there are correlations CVE with SNORT rules, alert IDS, and Log Element for periodic updating of info vulnerability, alert, pattern and patching.

7. **Data pattern attack from Honeypot**. According to project Honey Pot (www.projecthoneypot.org), they project to track online fraud and abuse from more than 170 countries around the world, this project has been online since 2004 and many receives million of email and spammer messages per days from catalogued and shared with law enforcement and security partners. Additionally, from proposal [19], they used Honeypot to capturing and analysis attacker to database analyzer. Performance work by [20], they have integrated the cluster structure visualization technique with, global and local, outlier detection techniques in their honey pot data analyzer tool. Honey pot produce pattern spammer harvesters and attacker behavior.

8. **Signature, dynamic update patch** from antivirus company such as virus signature, alert signature, and attack pattern, the correlation to update CVE, spammer rules and virus definition. Proposal [21], they describes using patch to update the existing version so as to be make the antivirus able to identify more viruses and safeguard the system in an efficient way. The update of the antivirus software is done with the help of data files which contain signatures of the various newly appearing viruses. According to Symantec Signature, extensive database of attack pattern to identify an attacker's attempt to exploit system vulnerability. This confirms the importance of updates provided by vendors. Wherefore, patch was created to cover the bugs that have been discovered by the security community or by the attacker.

9. **Traffic Flow** from Service Provider / Network Provider. As part of the excellent service provided, traffic flow is a conducted. As the framework of the network monitoring, Internet service providers are often involved in security incidents, either as a target of an attack or as one of the defenders, traffic management, anomaly behavior, to have some degree of access and visibility into the second-by-second health and performance internetwork. Some organization using network traffic flow (www.packettrap.com/product) for alert and notifications, application/ anomaly monitoring, log management and remote device control.

10. **Log events** (Server, Web applications, Firewall and network environment), Event logging and log files are playing an increasingly important role in system and network management. A typical syslog message has just the timestamp, hostname, and program name attributes that are followed by a free-form message string, hut only the message string part is mandatory. In most research papers [22], Log file monitoring techniques can be categorized into fault detection and anomaly detection. In the case of fault detection, the domain expert creates a database of fault message patterns. If a line is appended to a log file that matches a pattern, the log file monitor takes a certain action. Refers from Jiang in 2008 [23], introduce a lightweight approach to abstract log lines to execution events, they approach can handle free-form log lines with limited requirements on the format of the log lines. Additionally perform work by [24], analyzing log event for finding of the malware, they using data mining approach to aggregation log event and analysis goal.

11. **Spam Rules** (images spam, spam fingerprint, spam IP Blocklist). Currently spammers are increasing rapidly. They are doing their advertisement free of cost by sending SPAM. From experiment by Sun in 2008 [25], they use client-side honeypot to collect malware and spamm spreading through web browser (such as the Internet Explorer) and Office software (such as word, powerpoint and excel) vulnerabilities, if considered from this work database server logs as the data center of the honeypot machine and analysis machine. In 2010 proposal [17] uses two large datasets, namely a commercial blacklist and an Internet-wide address visibility study to quantify address characteristics of spam and non-spam networks from Autonomous System (AS). They find that spam networks exhibit

significantly less availability and uptime, and higher volatility than non-spam networks. According to [25], their build machine filter to checks incoming source address with its black listed information in mail server, if the address is in black list then it sends all the mail coming from the attacker to the reply generator.

12. **Virus Definition**, Currently, there are hundreds of new malicious codes released to the Internet. The variant of them belong to the traditional computer virus, and others known as worm programs, and some of them are mixed and have been merged into one (traditional virus, worm, and Trojan horses) refers from Bitdefender's and in 2008, Zemao [26] investigate the attacking mechanisms of malicious code. Analyzing virus, worm with DDA models respectively. Anti-virus system model pays mainly attention to five problems as follows: (i) characteristics database, (ii) training of characteristics database, (iii) collection methods of malicious acts, (iv) the judgment algorithms of virus, and (v) virus response [27]. Moreover, virus definition regularly update from anti virus Software Company, this response for anti virus after their research team discovered a new virus variants from the Internet. In this case, algorithm and information signature from virus can be collecting to create a pattern of suspicious. The correlation virus definition with signature update and Regex, in signature-based systems refers to [28], the characteristics of packet including worm can be often represented by regular expressions, and this is very closely related to the payload.

13. **Policy definition**. There are correlation Policy with IDS, in proposal [48], there are four feedback communication correlation with network quarantine channels: (i) policies to the IDS monitor using message flags, (ii) communication to IDS monitor using rules in IDS alert policies, (iii) IDS monitor rules in IDS alert filter, and (iv) IDS using message flags, alert policies and filters in sequence. The policy may be explicit and standard operation, as in a corporate policy, procedure or guideline, or more commonly, implicit as in the configuration of devices governed by a policy. Security operators make a policy definition as a basis to create a rules, the rule would be implemented in the security devices being used (i.e. Firewall / IDS/ Spam Filtering). The security policy must be consistent and well-administered, according to ISO 17799 and ISO 27001, security policy is a crucial step to secure a particular system since it specifies the security properties that must be satisfied and the rules that associate privileges to users, we conclude that standard is closely connected with how to regulate user access from the insides and rules on rights of access other outsiders.

14. **Alert from IDS**. The correlation Alert IDS with policy, Snort rules, log element and traffic flow provider, we observed information alert IDS can be correlation with it, According to [29], IDS having detected such signs, IDSs trigger alerts to report the suspicious knowing with pattern attack, and in 2008 Alsubhi [30], describes IDS alert management techniques. Performed work by [31], proposed alert management module responsible for colleting the alert generated from the self-corrective IDS, this correlated with the alerts, formulating and more general alert based on individual true positive. Additionally, the Intrusion Detection Message Exchange Format drafted by IETF Intrusion Detection Working Group based on RFC 4766, defines a data formats and exchange procedures for sharing information of interest to intrusion detection and response systems, and to the management systems which may need to interact with them (www.ietf.org/old/2009/ids.by.wg/idwg.html). In our observation, we can use defining a relationship metric between alert IDS, the correlation may occur because correlating alerts based on the similarity between alerts attributes, such as time stamp, IP and ports addresses. Alert attributes consist of several fields that provide information about the attack in stream network. This alert information depends on the variant used.

15. **Crawler data**, Web crawlers typically identify themselves to a Web server by using the User-agent field of an HTTP request, one type of robot, or software agent. Currently, rich Internet Applications are already handling much of the data on the web and on portable devices, providing a high degree of interactivity to the user. In 2009, Tong [32] describes the user agent field in crawler developing may include a URL where the Web site administrator may find out more information about the crawler. Represent form

previous work by Duda in 2008 [33], they describes address this problem with implement Search engine with AJAX. Just as a traditional search engine, it contains a crawler, indexer and query processor. From our observed, the crawling web application is one of the key phases of automated web application scanning. The objective of crawling is to collect all possible resources from the server in order to automate vulnerability detection on each of these resources. A resource that is overlooked during this discovery phase can mean a failure to detect some vulnerabilities. The correlations with Public IP block list and public URL black list to know properties and information the IP Address identified problems. Involves with logs element properties, this correlation input information for the defense system, because information like Meta tags, URL, web document, domain and other properties of the websites have visited can be logged for historical. The Analyze of the web crawler's potential threat to the website has describes clearly by [32], they have mechanism rules using policy to manage web crawlers' fetching during the rush hour quickly and effectively.

16. **Regular Expression pattern**, identify and recognize packet in stream network has become extremely important due to its applications in network security and network monitoring. Currently, regular expressions (Regex) are replacing explicit string patterns as the pattern matching language of choice in packet scanning applications. In our previous work [4], we describes Regex is specified using a keyword "pcre', *Perl Compatible Regular Expression*. PCRE is more powerful and complicated, than regex. The regex attribute always contains a single pattern, not lists of patterns. Some security devices using Regex for recognized pattern of application (Layer 7-filter). This mechanism is effective for identify and recognized application in stream network. According to [34], the packet payload is compared against a set of patterns specified as regular expressions in content scanning. Additionally, Câmpeanu in 2004 [35], proved every pattern expression language is a context-sensitive language. They also establish the relationship between pattern expression languages and context-free languages.

4. LEARNING PROCESS

Data Warehouse (DW) collects information about subject that span the entire information without no longer sorted all information in bulk database. DW is very hard to some clearly information with bulk information store in database. We need intelligent method to extract the information. Obviously, Data Mining (DM) can provide knowledge discovery process with data cleaning, data integration, data transformation and data reduction.

DW collecting scattered information in routine update regularly from provider or security community. This data can be useful information to be associated with other. The information, increasingly large of volume dataset and multidimensional data has grown rapidly in recent years. The data set includes parameters above. These data set bulk in information and growing from community or security services. Therefore, there is a critical need of data analysis system that can automatically analyze the data to classification it and predict pattern attack future trends.

In recent years, learning techniques have been widely used in Intrusion Prevention System (IPS) approach since the self-learning techniques can automatically form an opinion of what the subject's normal behavior is. According to whether, there are some researchers [10], [22] and [36] performed they work based on supervised or unsupervised learning techniques. Supervised learning, regression, linear discriminant analysis (LDA), artificial neural network (ANN) and support vector machine (SVM) theory are four typical supervised neural network. Unsupervised learning, there are no training and label of data. Therefore, the initial step that must be done is to categorize the data into a desired number of classes, which is existing data do not have labels, this label will mark where the data will be sorted.

Some recent researchers have [8], [37], and [38] using DM approach for classification the data. In this initial research, increasingly true alarm rate and classification packet is mainly focuses using knowledge learning method. **Fig, 2**, show our approach, in mark (a) potential useful information and collection of text (refer to **Fig, 1**), mark (b) database collecting with text mining, and mark (c) knowledge based on correlation data and learning behavior-based, and (d) classification (normal, malicious, suspicious) to increase accuracy/ identify: (i) predict pattern attack future trends, (ii) prevent before attack comes to network and (iii) detection security violation.

This method depends on the input information has been collected in a database. The information in the database come from a variety of information collected and stored from time to time. In some cases, the new types of attacks based on previous patterns, especially the attacks from malicious threat, on knowledge process, performed composite and combining the data residing on the database to be sorted, queries and reused as input. The learning process occurs to combine and choose quickly by comparing the fit of the data in the database. We identified these problem in collecting information from different structure, label, and variable of data.

As can be seen from **Fig, 2** in marking (a) is potential useful information and collection of text, from preliminary observed there are many inter-related information about security violation and threat [18], [34], [35] and [38], However, as we discussed in above, collecting data, which is text categorization task involves several sequential steps such as pre-processing of the documents, feature selection, dimensional reduction, document indexing, and inductive classifier learning. Furthermore, in mark (b), phase for find an approach in knowledge-based system, any information and knowledge representation paradigm. Text mining and DM are inherently hard problem in term of computational complexity. An interesting and summary some previously work using text mining help solve problem in security attack [22].

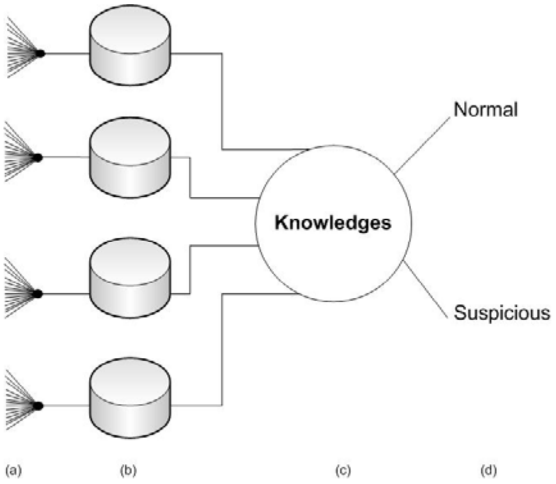

Figure 2: Learning and classification process

5. INTEGRATION CHALLENGING

In this section, we present the main issues and strategies that will be used to overcome the problems in this study. Integration parameters with different structures, labels, and variables of data are problem that must be resolved. Frequent automatic updates will make each event in parameter will always follow the trend of attack. The system will initially check with database for its existence, if it exists, the system will take an appropriate action to trigger alarm. We assumed, each database will contain more than 500,000 event list.

In **Fig. 4**, we depict subclass of parameters with relations and Table 1 present event parameters descriptions. Meanwhile, we illustrated in **Fig. 5** interconnecting datasets from provider or web security community.

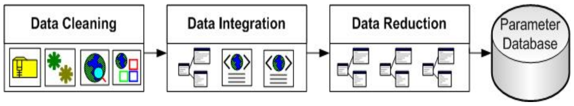

Figure 3: Integration mechanism

We have shown in **Fig. 3**, integration mechanism planning for these steps. Furthermore, four steps approach can be used to overcome this problem, as following;

1. First, data cleaning steps: in this stage all the data filtered first before inclusion in the parameters. Routines attempt to fill in missing values, smooth out noise while identifying outliers, and correct inconsistencies in the data. Classifying / grouping the data, such as : (i) equate data format, (ii) equate data with NTP for time synchronize, and (iii) equate data based on sources.

2. Second, data integration: the merging of data from multiple data stores, the data may also need to be transformed into form appropriate for mining, such as: (i) data integration, and (ii) data transformation. From our observed using the same standard for data format, which is ASCII file type and text comma separated (.csv) is that can be used.

3. Third, data reduction, this technique applied to obtain a reduced representation of the data set that is much smaller in volume, yet closely maintains the integrity of the original data: (i) dicreatization & concept hierarchy, and (ii) histogram.

4. Finally, we divide the four database parameters in the priority scale as measured from the level of substance.

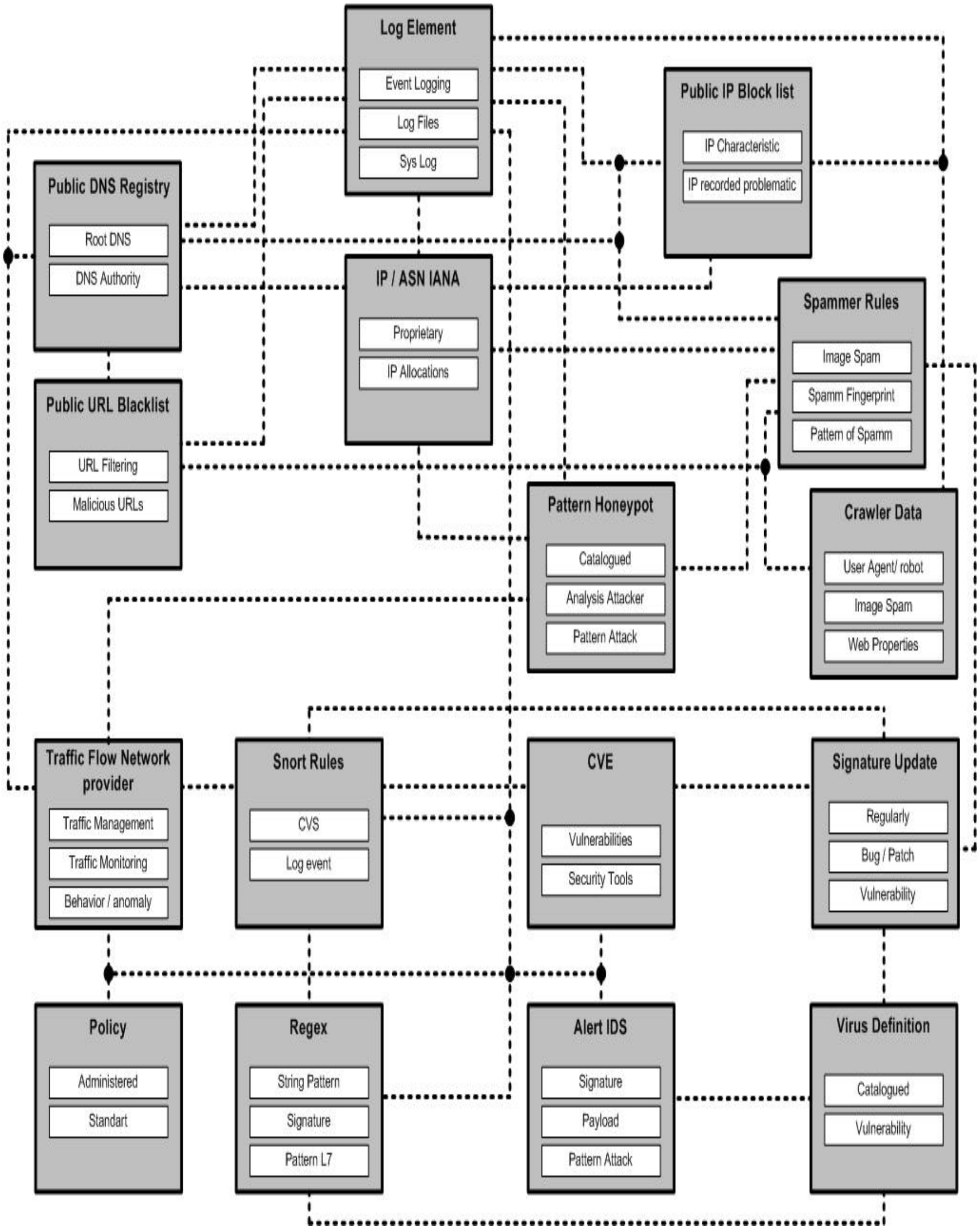

Figure 4: Example of correspond with parameters

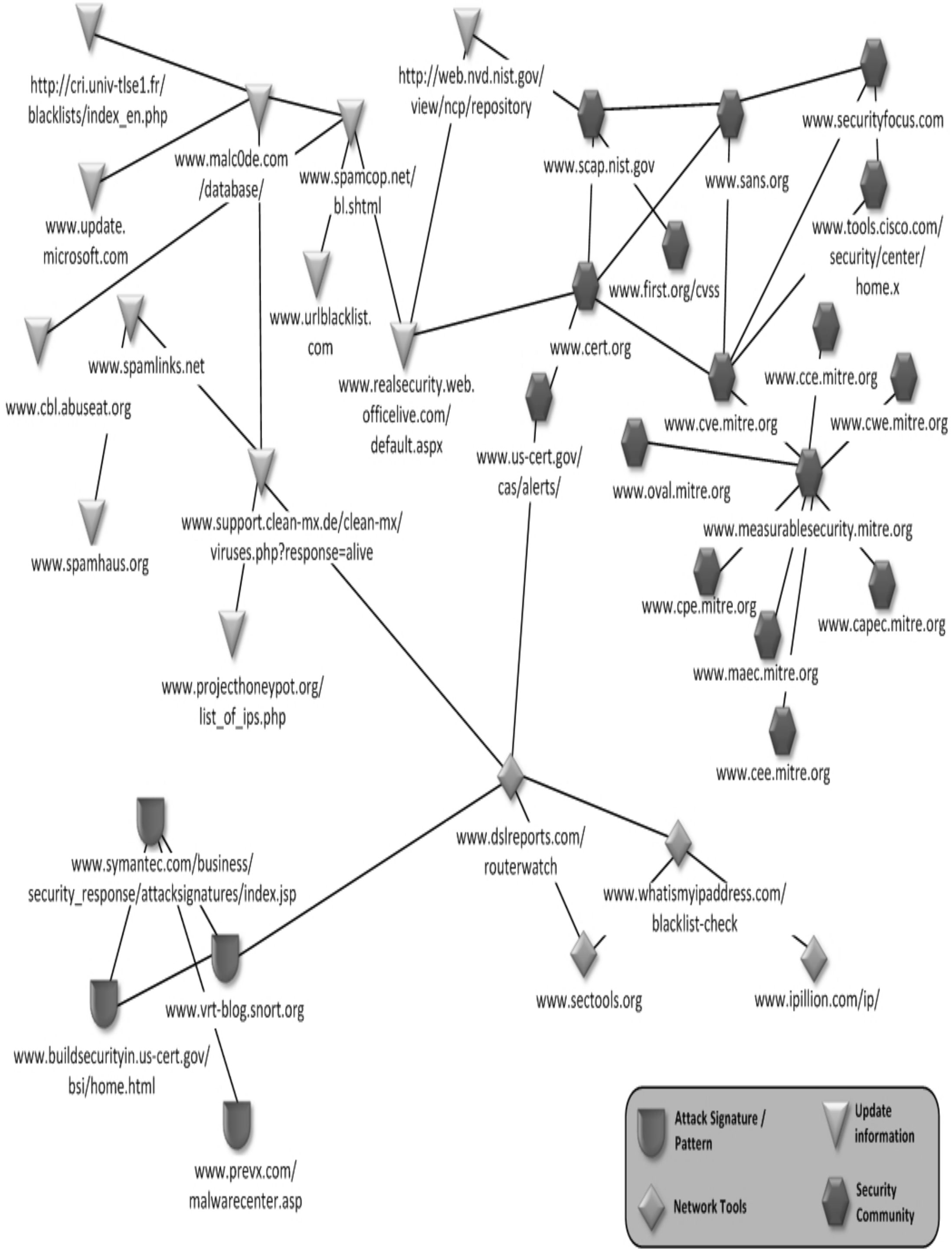

Figure 5: Example of correspond web community

Table 1. Event Parameters

No	Parameters	Attribute	Description
	Public DNS Registry	Domain Name	Name of domain
		Registrant	Usually contain full information about domain owner or company name
		Whois	Name server of registrant
		Referral URL	URL domain of registrant
		Who is	Contains information on using the IP address and the hierarchy of the DNS root registrant
		Admin contact	Contain information of administrative / technical contact domain holder
		Name server	Name server of domain name (sometimes referred to as: ns1, ns2, ns3)
		IP Address	Internet Protocol Addressing of name server
		Status	Report status of domain, active or inactive
		Update date	The last time that domain is extended
		Creation date	Explained when the domain was first made
		Expires date	Report when the domain expiration
2	**IANA Authority**	Organisation name	NetName Information detail
		Contact	Contact NetName
		NetName	Network Name of numbering resources for block IP Address and Autonomous System Registry (AfriNIC, APNIC, ARIN, LACNIC or RIPE NCC)
		InetNum	Ownership the IP Block
		IP Block	Declare the number of IP Addressing given from registry
		Root Zone	The Root Zone Database represents the delegation details of top-level domains, including gTLDs. (http://www.iana.org/domains/root/db/)
3	**Public URL blacklist**	Categories	Divided by groups, such as: porn/ adult, hacking, malware/ warez, hardcore, etc)
		URL	Web address / Domain
		ASN	Autonomous System Name from registry, depict of administrative allocation service providers
		Date	Start – end date of occur event
4	**Public IP block list**	IP Address	Numbering resource from registry
		URL	Web address / Domain
		ASN	Autonomous System Name from registry, depict of administrative allocation service providers
		Date	Start – end date of occur event
5	**Snort Rules**	Sid	Unique numbering based on process
		Name	Name of rules and sid
		Alert	The signature of rules, alert / deny
		Msg	Messages declaration command of Name
		Content	Describes of msg
		Class type	Explain class of sid: bad-unknown
		References	Declaration refers from somewhere
		Proto	Protocol used by Name
6	**Vulnerability from Common Vulnerability and Exposures**	Name	Identify sequences of information with identification event
		Alert	Describes of alert from Source
		Source	References from others security community or security vendor
		Compatibility	Statement from Vendors for CVE Support/compatible
		Description	Detailed information of alert
		Solution	Troubleshooting steps or trick for solve
7	**Data pattern attack from Honey pot**	Malicious IP	IP Address source and destination
		Timing	Time occurrence
		Payload	IP Address, Port Address, and Protocol
		Associate	Illustrate of behavior user or attacker
		CVE	Number unique from CVE (if identify or compatibility with it)
8	**Signature, dynamic update patch**	Type	Virus / Vulnerability/ Patch / Spam
		Sid	Signature identification of name
		Name	Name of type
		Date	Date of issued
		Risk Rating / Severity	Status of threat assessment (low, medium, high)
		CVE	Number unique from CVE (if identify or compatibility with it)
		Solution	Steps or suggestions should be done for fix the risk rating / severity
9	**Traffic Flow**	Graph	Source of traffic flow / looking glass
		Name	Name of graph
		Timing Aggregate	Divide timing of occurrence with daily, weekly and yarly
		IP Address	Addressing of graph
		Status	Describes of traffic load
10	**Log events**	Name	Name of server farm
		Time stamp	Date of issued / time of occurrence
		Payload	Describes IP Address, port address, and MAC address source and destination
11	**Spam Rules**	Name	Name of categories (pharmacies, products advertising, lottery Scams, phishing virus)
		Time stamp	Date of issued / time of occurrence
		Content Type	Divided by types (attachment, bounce, image, multipart, text, HTML, and other)
		Signatures	The information of spam name, describe images and fingerprint.
		Risk Rating / severity	Status of virus assessment (low, medium, high)
		Payload	Describes information from source IP Address / URL
		CVE	Number unique from CVE (if identify or compatibility with it)
12	**Virus Definitions**	Name	Name of virus / malware
		Type	Category of name (Trojan, Worm, Key logger, Phishing, Bot net, Adware, Spyware, others)
		Date	Date of discovered
		Infected	Category entrance to the system (e-mail, java-x, web, software)
		Risk Rating / severity	Status of virus assessment (low, medium, high)
		Payload	Describes IP Address, port address source and destination
		CVE	Number unique from CVE (if identify or compatibility with it)
13	**Policy Definition**	Standard	Standard issued (ISO / Vendors)
		Risk Rating / severity	Status of virus assessment (low, medium, high)
		CVE	Number unique from CVE (if identify or compatibility with it)
14	**Alert form IDS**	Sid	Classifying based on signature identification
		Signatures	Depending from knowledge database
		Time stamp	Time occurrence and event recognized
		Source address	IP Address from source
		Destination address	IP Address to destination
		Proto	Protocol used by the event
		Risk Rating / severity	Status of virus assessment (low, medium, high)
		CVE	Number unique from CVE (if identify or compatibility with it)
15	**Crawler**	URL	Web address / Domain
		Date	Start – end date of occur event
		Payload	Information of IP Address source and destination
		Time stamp	Date of issued / time of occurrence
16	**Regular Expression (Regex)**	Name	Name of applications
		Properties	Information of name applications
		Regex	Layer 7 reguler expression

6. CONCLUSION & FUTURE WORK

Currently, accuracy alarm and prevention from intrusion is mainly focus research, moreover for attack identification and mitigation. In this paper, an approach offered to develop a robust system for identification and mitigation techniques with parameter database. The information, increasingly large of volume dataset and multidimensional data has grown rapidly in recent years. This approach still needs further exploration in future research mainly query correlation each parameters, using data mining approach is one primary our focus.

Preliminary work have been done with collecting from security community, these experiment show the efficiency of the approach. Other experiments are currently being done on larger data sets describing more history on a long time period. In the future research can also include more factors to implement our approach in real environment and benchmarking with other IPS software solution to tested effectiveness on accuracy, attack containing, measurement vulnerabilities, and risk/nearness True Positive and False Positive value.

Acknowledgements
This research is supported by The Ministry of Higher Education and collaboration with Research Management Center (RMC) Universiti Teknologi Malaysia.

REFERENCES:

[1] Ghorbani A.A, *Network Intrusion Detection and Prevention : Concepts and Technique*, Springer, 2009.

[2] M. Shouman, A. Salah, and H.M. Faheem, "Surviving cyber warfare with a hybrid multiagent-based intrusion prevention system," *IEEE Potentials*, 2010, pp. 32-40.

[3] D. Stiawan, A.H. Abdullah, and M.Y. Idris, "Classification of Habitual Activities in Behavior-based Network Detection," *Journal of Computing*, vol. 2, 2010, pp. 1-7.

[4] D. Stiawan, A.H. Abdullah, and M.Y. Idris, "Pitcher Flow : Unified Integration for Intrusion Prevention System," *International Conference on Computer Communication and Management (ICCCM 2011)*, 2011.

[5] Y. Weinsberg, S. Tzur-David, D. Dolev, and T. Anker, "High Performance String Matching Algorithm for a Network Intrusion Prevention System (NIPS)," *High Performance Switching and Routing*, 2006, pp. 147-153.

[6] E. Guillen, D. Padilla, and Y. Colorado, "based Intrusion Detection and Prevention Systems," *Latin-American Conference Communications*, 2009, pp. 1-4.

[7] B. Cao, Z. Zhihong, L. Tie, Y. Zhongde, and L. Jiren, "A Study on Performance Improvement of Gateway Anti-Virus System Based on File Scanning," *Control and Decision Conference 09*, 2009, pp. 2293-2295.

[8] T.S. Chou and T.N. Chou, "Hybrid Classifier Systems for Intrusion Detection," *IEEE Computer Society Seventh Annual Commnucation Networks and Services Research Conference*, 2009, pp. 286-291.

[9] M.A. Aydın, A.H. Zaim, and K.G. Ceylan, "A hybrid intrusion detection system design for computer network security," *Computers and Electrical Engineering*, vol. 35, 2009, pp. 517-526.

[10] S.X. Wu and W. Banzhaf, "The use of computational intelligence in intrusion detection systems : A review," *Applied Soft Computing*, vol. 10, 2010, pp. 1-35.

[11] S. Bin, W. Qiaoyan, and L. Xiaoying, "A DNS based Anti-Phishing Approach," *Science*, 2010, pp. 263-266.

[12] A. Karasaridis, K. Meier-hellstern, and D. Hoeflin, "Detection of DNS Anomalies using Flow Data Analysis," *Communications Society*, 2006, pp. 0-5.

[13] F. Guo and J. Chen, "Spoof Detection for Preventing DoS Attacks against DNS Servers," *IEEE ICDCS'06*, 2006.

[14] R. Perdisci, M. Antonakakis, X. Luo, and W. Lee, "WSEC DNS : Protecting Recursive DNS Resolvers from Poisoning Attacks," *IEEE Proceeding*, 2009, pp. 3-12.

[15] Z. Zhou, T. Song, and Y. Jia, "A High-Performance URL Lookup Engine for URL Filtering Systems," *IEEE ICC 2010*, 2010, pp. 1-5.

[16] P. Prakash, M. Kumar, R.R. Kompella, and M. Gupta, "PhishNet : Predictive Blacklisting to Detect Phishing Attacks," *IEEE INFOCOM*, 2010.

[17] C. Wilcox, C. Papadopoulos, and J. Heidemann, "Correlating Spam Activity with IP Address Characteristics," *IEEE INFOCOM*, 2010, pp. 2-7.

[18] R.A. Martin, "Managing Vulnerabilities in Networked Systems," *Computer*, vol. 34, 2001, pp. 32-38.

[19] U. Thakar, S. Varma, and A.K. Ramani, "HoneyAnalyzer – Analysis and Extraction of Intrusion Detection Patterns & Signatures Using Honeypot," *The Second International Conference on Innovations in Information Technology (IIT'05)*, 2005.

[20] N.V.S. Reddy and U.D. Acharya, with Krishnamoorthi, "A Two-stage Hybrid Model for Intrusion Detection," *IEEE Proceeding, Advanced Computing and Communications, ADCOM*, 2006, pp. 163-165.

[21] A. Badhusha, S. Buhari, S. Junaidu, and M. Saleem, "Automatic Signature files update in Antivirus software using Active Packets," *IEEE Computer Systems and Applications, ACS*, 2001, pp. 457-460.

[22] R. Vaarandi, "A Data Clustering Algorithm for Mining Patterns From Event Logs," *World Wide Web Internet And Web Information Systems*, 2003, pp. 119-126.

[23] Z.M. Jiang, A.E. Hassan, P. Flora, and G. Hamann, "Abstracting Execution Logs to Execution Events for Enterprise Applications," *IEEE Proceeding The 8th on Quality Software*, 2008, pp. 181-186.

[24] R. Gabriel, T. Hoppe, A. Pastwa, and S. Sowa, "Analyzing Malware Log Data to Support Security Information and Event Management : Some Research Results," *IEEE International Conference on Advances in Databases, Knowledge, and Data Application*, 2009, pp. 108-113.

[25] S. Das and I.I.T. Roorkee, "Reducing the Effect of Distributed Directory Harvest Attack and Load of Mail Server," *IEEE Region 10 Colloquium and the Third ICIIS*, 2008, pp. 1-6.

[26] C. Zemao, Z. Junge, W. Xiaoping, and T. Weimin, "Analyze and Model the Primitive Attacking Mechanisms of Malicious Codes," *IEEE Region 10 Colloquium and the Third ICIIS*, 2008, pp. IEEE Region 10 Colloquium and the Third ICIIS.

[27] X.Z. Chen, Q.-hua Zheng, and X.-hong Guan, "Multiple behavior information fusion based quantitative threat evaluation *," *Computer & Security*, 2005, pp. 218-231.

[28] K. Tatara, "Analyzing Maximum Length of Instruction Sequence in Network Packets for," *IEEE ICMUE*, 2008, pp. 485-489.

[29] T. Pietraszek and A. Tanner, "Data mining and machine learningd Towards reducing false positives in intrusion detection," *Information Security Technical Report (2005)*, vol. 10, 2005, p. 169e183.

[30] K. Alsubhi, E. Al-shaer, and R. Boutaba, "Alert Prioritization in Intrusion Detection Systems," *IEEE proceeding Network Operations and Management Symposium*, 2008, pp. 33-40.

[31] M. Sourour, B. Adel, and A. Tarek, "Collaboration between Security Devices toward improving Network Defense," *Seventh IEEE/ACIS International Conference on Computer and Information Science (icis 2008)*, May. 2008, pp. 13-18.

[32] W. Tong, "A Research On A Defending Policy Against the Webcrawler ' s Attack," *IEEE, ASID 2009*, 2009.

[33] C. Duda and G. Frey, "AJAXSearch : Crawling , Indexing and Searching Web 2 . 0," *ACM Proceeding VLDB*, 2008, pp. 2-5.

[34] R.H. Katz, "Fast and Memory-Efficient Regular Expression Matching for Deep Packet Inspection," *ANCS 06*, 2006, pp. 93-102.

[35] C. Câmpeanu and S. Yu, "Pattern expressions and pattern automata," *Information Processing Letters*, vol. 92, 2004, pp. 267-274.

[36] J. Zhang, M. Zulkernine, and A. Haque, "Random-Forests-Based Network Intrusion," *MAN and Cybernetics*, vol. 38, 2008, pp. 649-659.

[37] A. Foroughifar, M.S. Abadeh, A. Momenzaideh, and M.B. Pouyyan, "Misuse Detection via a Novel Hybrid System," *2009 Third UKSim European Symposium on Computer Modeling and Simulation*, 2009, pp. 11-16.

[38] T. Dutkevych, A. Piskozub, and N. Tymoshyk, "Real-Time Intrusion Prevention and Anomaly Analyze System for Corporate Networks," *IEEE International Workshop on Intelligent Data Acquisition and Advanced Computing Systems Technology and Application*, 2007, pp. 599-602.

AUTHOR PROFILES:

Deris Stiawan. Holds an M.Eng from University of Gadjah Mada, Indonesia, since 2006, he is Computer Science faculty member at University of Sriwijaya, Indonesia. He is member of IEEE and currently pursuing his Ph.D degree at Faculty of Computer Science & Information System, Universiti Teknologi Malaysia (UTM) working in intrusion prevention system. He holds C|EH licensed from EC-Council and joined research group Information Assurance and Security Research Group (IASRG) in the area of Intrusion Prevention and Detection (ITD). His professional profile has derived to the field of computer network and network security, specially focused on intrusion prevention and network infrastructure.

Mohd. Yazid Idris. Ph.D, is a senior lecturer at Faculty of Computer Science and Information System. He obtained his M.Sc and Ph.D in the area of Software Engineering, and Information Technology (IT) Security in 1998 and 2008 respectively. In software engineering, he focuses on the research of designing and development of mobile and telecommunication software. His main research activity in IT security is in the area of Intrusion Prevention and Detection (ITD). He is currently active in various academic activities and involves in university-industry link initiative in both areas, and recently received a prestigious award in the mobile software invention by the government of Malaysia and telecommunication leading industry.

Zohair Ihsan,is PhD researcher at Universiti Teknologi Malaysia. His research area includes Information and Communication Security, Intrusion Prevention and Detection System using Semantic Technologies. He received his MS(CS) degree from COMSATS Institute of Information Technology Wah Cantt. Pakistan. Currently he is associated with the Information Assurance & Security Research Group (IASRG) at UTM. Formerly he was working as Assistant Professor in Department of Computer Science at COMSATS Institute of Information Technology Attack Pakistan.

Khalid Hussain, He is PhD. Scholar in Universiti Teknologi Malaysia. Receive the MS (CS) degree fro COMSATS Institute of Information Technology Islamabad Pakistan in 2007. He is Assistant Professor and In-Charge Research Faculty of Computing Riphah International University Islamabad Pakistan. His research interest is Information Security and Wireless Network Security. He is also member of Pervasive Computing Research Group (PCRG) UTM and Dependable and Secure Communication System.

Abdul Hanan Abdullah. Ph.D, Receive the B.Sc. and M.Sc from San Francisco, California, and Ph.D degree from Aston University, Birmingham, UK, in 1995. He is a Professor at Faculty of Computer Science & Information System, UTM. His research interest is in Information Security. He is also a head of Pervasive Computing Research Group (PCRG) UTM and he is a senior member of ACM and IEEE.