Revealing the Influence of Feature Selection for Fast Attack Detection

Mohd Faizal Abdollah[†]**, Asrul Hadi Yaacob, Shahrin Sahib, Ismail Mohamad, Mohd Fairuz Iskandar**,

University Technical Malaysia, Ayer Keroh, Melaka, Malaysia, Multimedia University, Melaka, Malaysia and University Technology Malaysia, Skudai Johor, Malaysia

Summary

The success of an intrusion detection system depends on the selection of the appropriate features in detecting the intrusion activity. Selecting unnecessary features may cause computational issues and decrease the accuracy of detection. Furthermore, current research concentrates more on the technique of detection rather than revealing the reason behind the selection. They just used the features without mentioning the influence of the feature inside the system itself. Therefore this research will reveal the influence of the features using statistical approach and comparison approach. The result indicates that the feature selected in the research has a good influence and may be useful in detecting the intrusion activity. After revealing the relation and influence of the features, we propose a set of minimum features that can be used to detect a fast attack.

Key words:

Basic Feature, Derived Feature, Fast Attack, Intrusion Detection System

1. Introduction

Nowadays, as more people make use of the internet, their computers and the valuable data in their computer system contain become more exposed to attackers. Normally, attackers constantly scan the Internet searching for victims' machine that can be broken into, be inserted with malicious code and be controlled in order to suit their malicious purpose using attacks such as worm attack, distributed denial-of-service attack and probing attack. These kind of attacks will cause serious damage to corporate or government agencies such as what happened on the 11th of January 2002 at 2:00 a.m. against Gibson Research Corporation (GRC), USA where the internet access link of the company was completely flooded with SYN/ACK segment due to the reflector-based DDoS attack [1]. Besides that, the losses due to the security breach will slow down the speed of the market development especially in E-Commerce environment. As reported by CSI (Computer Security Institute) and FBI (Federal Bureau of Investigation) in 2002, the total amount of money losses is about 400 million US dollars [2]. Furthermore, by using the investors' reaction in capital markets as a proxy to estimate security breach costs, Cavusoglu et al.(2004) found that on average the lost is approximately 2.1% which equal to 1.65 million average loss within 2 days [3]. Therefore necessary action should be considered to protect the organization from incurring losses due to security breaches.

This can be done by introducing intrusion detection system (IDS) inside the network which has the capabilities to analyze the internet traffic and recognize the intrusion. There are two approach used by the IDS to detect the intrusion activity which are anomaly based system and signature based system. The signature-based IDS, is also often known as a misuse detection system. Signature-based IDS identifies the intrusion by performing simple pattern matching and reporting the situation that matches a pattern corresponding to a known attack type [5]. Meanwhile, the anomaly-based IDS identifies the intrusion by notifying operators of traffic or application content presumed to be different form normal activity on the network or host [6]. The classification is based on heuristic or statistical, rather than patterns or signature, and will detect any type of misuse that falls out with normal system operation. False alarms generated from both systems are the main drawback which delays the implementation of the reactive intrusion detection system.

Based on the anatomy of attack [7], reconnaissance and scanning are an initial stage for the attacker getting information from potential vulnerable machines. For example Nmap and ipsweep are among the most popular tools used to launch both of these attacks. The initial stage can be divided into two different categories which are fast attack and slow attack. Fast attack can be defined as an attack that uses a large amount of packet or connection within a few seconds [8,9]. Meanwhile, the slow attack can be defined as an attack that takes a few minutes or a few hours to complete [10]. Based on the definition, detecting the fast attack is very useful to prevent any early attacks on the network and may help to reduce the possibilities of gaining access, maintaining access and covering tracks. Therefore, this paper will focus on detecting fast attack to reduce possibilities of

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

reconnaissance and scanning activity against the vulnerable machine.

Before introducing intrusion detection system as a defense tool, selecting necessary features is important. It is because the success of the intrusion detection system depends on the decision upon the set of features that the system is going to use for detecting the attacker especially on detecting the fast attack. Because fast attack mechanism takes only a few seconds to launch and the technique that the attacker uses is also different [11], thus the features selected to identify the attacker are also becoming more difficult to construct and differ from one to another [12,13]. Moreover, most of the attackers nowadays are knowledgeable and capable of altering the details of many attacks to avoid the detection of such a system [14].

Beside that, huge amounts of network traffic also can slow down the system because the task of identifying and measuring the features inside the network traffic are very tedious [15] especially the complexity of the involved protocol such as UDP and TCP where the number of phenomena can be studied only if indepth knowledge of the protocol detail is exploited. Furthermore, extraneous features inside the network traffic or audit data may be harder for the intrusion detection system to detect the suspicious behavior [16] especially fast attack.

Despite the importance in selecting the most important features, to the best of our knowledge, there are no comprehensive studies and research on classification of the feature that the NIDS might use for detecting the attacker. In addition, most of the researchers used only subsets of features in their research and only briefly mention the chosen feature or reason behind the selection [12]. Most of the researches concentrated only on the technique rather than feature classification which is more important in the NIDS. Furthermore, the influence of features inside the detection model has not been revealed.

Therefore it is critical to overcome the problem on identifying the important feature inside a network traffic to identify the intrusion activity especially fast attack. Furthermore, understanding the relation and influence of the feature may be useful before using the feature for the detection process which has not been concentrated by previous researchers nowadays.

This paper will reveal the influence of the feature using statistical approach and comparison technique from previous researchers to increase the confidence level of the selected feature before using it inside the detection model. After understanding the relation and influence of the feature, we propose a set of minimum feature that can be used in detecting the fast attack.

2. Related Work

There are researchers who concentrate on features selection for the intrusion detection system. The researchers who focus on feature selection concentrated more on KDDCUP99 [17] features. These researchers used multiple techniques to identify the suitable feature for detecting the intrusion activity based on the intrusion class introduced by KDDCUP99. Most of the features used in KDDCUP99 are concentrated on the target host rather that source host. Besides using the KDDCUP99 features, there are other researchers who also used features in detecting the intrusion activity but they only focus on the technique of detection rather that the feature itself. Further explanation of the feature will be divided into 2 subcategories which are KDDCUP99 features and Non KDDCUP Features.

2.1 KDDCUP99 Features

The KDDCUP99 has introduced 41 set of features as depicted in Appendix 1 in detecting 4 classes of attack [16]. The 41 set of features can be grouped into 4 categories which are [13]

a. Basic Feature: Basic feature can be derived from packet header without inspecting the payload.
b. Content Feature: Content Feature can be derived by assessing the payload of the original TCP Packet.
c. Time based Traffic Feature: These feature can be designed to capture properties that mature over a 2 second temporal windows. Time based feature is suitable to detect the fast attack [6].
d. Host based Traffic Feature: Utilized historical windows estimated over the number of connection such a 100 connection without considering time. This feature is suitable to detect the slow attack [6].

Most of the previous research particularly on data mining and neural network who concentrated on feature selection used KDDCUP99 features in identifying the intrusion. Chebrolu et al (2005) used CART and Ensemble techniques to classify the features and manage to introduce 17 features in detecting the probe and DOS activity [16]. Meanwhile Sung and Mukkamala (2003) used PBRM (Performance based Ranking Method) and SVDFRM (Support Vector Machine Function Ranking Method) technique to classify the set of feature [18]. Using PBRM, the researchers were able to classify 8 features for probe and 20 features for DoS which can be categorized as a fast attack. Meanwhile using SVDFRM, the researchers manage to classify 11 features for both of the attacks. Anazida et al, (2006) used rough set (RB) theory to classify the features and manage to classify only 6 features to detect the fast attack [19]. Eventhough they manage to classify 6 features, but they never mentioned the features used to detect the attack for each class of the attack. All the experiment explained earlier was done using the KDDCUP99 data and features introduced by KDDCUP99. By comparing of the previous result, we manage to identify the most useful features to detect the fast attack. Furthermore, understanding the relation and reason behind the selection for each of the feature may help to introduce minimum feature for the detection of the fast attack. Therefore, this research may proposed a set of minimum features which is useful for the detection process. The comparison analysis of the feature will be discussed later in the next section.

2.2 Non KDDCUP99 Features

Besides using the set of features introduced by KDDCUP99, there are researchers who used different sets of features in their research. Lakhina et al, (2005) has combined some field inside IP header and TCP header such as source IP, destination IP and source port [20]. The author concentrates on technique of detection without mentioning the influence of the feature which contributes to the detection process. Payload also has been considered as a parameter that can be used to recognize the network anomalies [21]. By inspecting the payload, the hidden code that is suspicious can be identified and comparison can be made to a normal packet. Although examining the payload is a better approach, it is still not widely used due to some restrictions such as security, privacy and legal issues [22]. Therefore this research does not include payload as one of the feature to detect the fast attack.

Beside payload, timestamp has also been used by researchers as one of the feature in recognizing an intruder inside the network. Normally, most of the researchers used timestamp for the correlation research in detecting the anomalies [23]. The correlation is done by comparing the different log for different host inside the network at the specific time to recognize the anomalies. In this research, timestamp has been selected as one of the important feature used to detect the fast attack. Using timestamp, the time of the intrusion which occurs inside the network can be identified and the derived feature as stated inside KDDCUP99 feature can be computed. Furthermore, timestamp also can be used to detect the intrusion at the initial stage where most of the penetration will use scanning and flooding mechanism to detect the organization network.

Avinash et al (2007) used five tuples as a feature in their research to detect the horizontal scan [24]. The five tuple features include source ip, source port, destination port and destination ip and protocol. Considering only 5 tuple as a feature to detect the intrusion activity is not enough. Considering connection flag especially TCP SYN Flag for TCP connection is important [25] feature to make a good detection. Therefore, we include TCP SYN Flag in our research since TCP SYN flag is important to detect the intrusion activity. Furthermore, by using TCP SYN flag, we can detect an attack at an early stage before large number of half-open connections are maintained by the protected server [26].

From the previous research, none of the researchers mention the usefulness of the features in identifying the attack and how the features can influence the detection of the attackers. Therefore, it is a good opportunity to reveal the influence and usefulness of the features using the statistical approach and comparison technique in identifying the attacker. The statistical approach will be applied to the non KDDCUP99 feature, while comparison technique from previous research will be applied to the KDDCUP99 features. The statistical approach is not applied to the KDDCUP99 features since the previous research has already identified the suitable features used to identify the intrusion activity. For example, dst_host_count is one of the features inside the host based traffic feature. Therefore, the comparison technique is applied to the KDDCUPP99 features to identify the most useful feature in detecting the intrusion activity. After analyze both of the feature, we manage to propose a set of minimum feature which integrate the KDDCUP99 feature with the non KDDCUP99 feature in detecting the fast attack. The propose feature will be explain earlier in next section together with the analysis of the influence of the feature.

3.0 Propose Feature Selection

For intrusion to occur there must be both an overt act by the attacker and a manifestation from the victims. Therefore creating a taxonomy that organizes intrusion from both perspectives; attacker perspective and victim perspective, may help to detect fast attack activities [31]. Therefore, the feature selection for the detection process introduced in this research was motivated by the attacker's perspective and victim's perspective [8]. The features are created by integrating the KDDCUP99 feature and Non KDDCUP99 features. Furthermore, the classification of the feature is based on the KDDCUP99 classification which contains basic feature and time based traffic feature. The host based traffic feature is not used in this research since the features are more concentrated in detecting the slow attack. Instead of using 2 seconds for the temporal windows, we used 1 second for the temporal windows for the time based traffic feature. Table 1 shows the detailed description of the proposed features in detecting the fast attack.

A. Basic Feature

Basic features is also known under the name of Packet Header Features. Basic Feature can be derived from packet header without inspecting the payload. The possible candidates for this feature category includes timestamp, source port, source IP, destination port, destination IP, flag, to name a few.

B. Derived Feature

This feature can be characterized as multiple connection made by the hosts at the same time. Time based traffic feature has been chosen because it has the capabilities to detect the fast attack activity. Time based traffic feature are designed to include the entire derived feature computed with respect to the past t seconds. One second has been chosen in this research to compute the derived feature.

Table 1 : Feature Selection for Fast Attack

Feature	Description	Category
Timestamp	Time the packet was send	Basic Features
Duration	Duration of connection	
IP	Addresses of host	
Protocol Type	Connection protocol (e.g. tcp, udp, icmp).	
Flag	Status flag of the connection	
Service	Source and Destination services	
Src_count	Number of destination host receiving new connection from same source IP presume attacker (AAH)	Derived Features
Srv_count	Number of destination service receiving new connection from presume attacker (AAS)	
Dst_count	Number of connection having the same destination host (AVC)	

Assessing the influence and relation of the features is important before developing the intrusion detection system in detecting the fast attack. Therefore, the next section will discuss the methodology used to assess the influence and relationship of the feature.

4.0 Methodology

There are two different techniques involve which are statistical approach and comparison approach which were used to reveal the influence of the feature inside the detection model. The statistical approach is used to reveal the influence of the src_count and srv_count feature, while comparison approach is used for assessing the dst_count feature. Next is detailed explanation for both approaches.

4.1 Statistical Approach

For the non KDDCUP99 data, we grounded our exploration of the problem space using a set of real traffic data which was captured from one of the agencies in Malaysia for a period of one day. By using the real traffic and not simulation traffic, we can see the actual behavior of the attacker in launching the fast attack against the network. The real traffic was captured for one day because we made an assumption that normally the attacker will not try to do damage to the system for a period of more that one day. If the attacker stays inside the system for a long period of time, the chances of being detected by the administrator are high.

Figure 1 depicts the methodology used to reveal the influence of the features in this research. We used TCPDUMP application [27] to capture and read the raw network traffic. This module separated the file into their respective protocol such as TCP, UDP and ICMP. In this research we only concentrated on the TCP protocol since TCP protocol is a widely used protocol [25].

Due to huge amounts of network traffic, it is difficult to distinguish the normal and abnormal behavior of network traffic [29]. Therefore we used Bro to distinguish between the normal and abnormal behavior of the network traffic. This technique has been applied by Caulkins et al, (2006) in his research using Snort to distinguish the normal and abnormal behavior of the network traffic [30]. In our research, we used the same approach but different intrusion detection tools which is Bro. Bro default configuration has been applied in this research. The alarm generated by Bro will be captured and the attacker's IP address will be identifed. The IP address of the normal and attacker will be used to find the number of connection made by both of the IP addresses. The output from the time based module [8] will be used for searching and comparing for each of the IP address. Then, the Log Likelihood Ratio Test and Nagelkerke's R^2N test inside the logistic regression model is used to analyze feature influence inside the detection model [33].

For src_count and srv_count, we asses the contribution of the feature using likelihood Ratio Test in equation (1) inside logistic regression. The likelihood ratio test is used by comparing the model with and without a particular predictor. If the value of the likelihood ratio model without the predictor decrease when the predictor is included inside the model, it means that adding the predictor gives a significant contribution to the model in predicting the outcome.

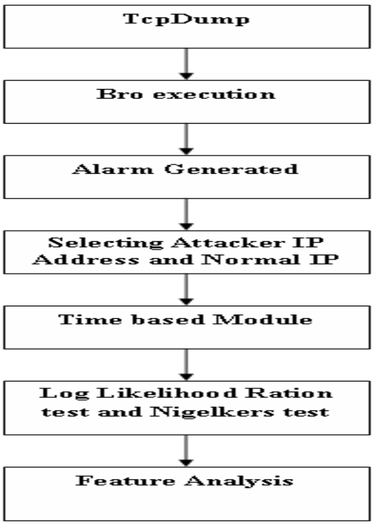
TcpDump
Bro execution
Alarm Generated
Selecting Attacker IP Address and Normal IP
Time based Module
Log Likelihood Ration test and Nigelkers test
Feature Analysis

Fig. 1 : Methodology for Src_count and Srv_count.

$\chi 2$ = 2[LL(New(with predictor)) – LL(Baseline(without predictor))] (1)

Besides the likelihood ratio statistic test, Nagelkerke's R^2N in equation (2) can also be used to indicate whether the feature gives a good prediction on the result or outcome variable. Nagelkerke's is the amendment of the Cox and Snell because the lack of Cox and Snell to reach its theoretical maximum of 1 [33].

$$R^2 = \frac{1 - \left[\frac{-2LL_{null}}{-2LL_k}\right]^{2/n}}{1 - \left(-2LL_{null}\right)^{2/n}} \qquad (2)$$

4.2 Comparison Approach

The comparison approach is used to assess the relation of the dst_count feature in detecting the fast attack. Count is one of the features inside the KDDCUP99 which has the same meaning as dst_count in this research. Therefore, we chose comparison technique to assess the usefulness of the feature since this feature is one of the KDDCUP99 feature.

Based on the literature review, most of the previous researchers who concentrated on feature selection used KDDCUP99 feature inside their research. The result from the previous researches show that the features selected gives a significant influence in detecting the intrusion activity. Although, different researches give different result, but by comparing all the result, we may find the most popular feature used in detecting the fast attack. The most popular feature indicate that the feature gives a significant contribution and has good relationship in detecting the fast attack activity. Therefore the comparison is solely focus on the KDDCUP99 feature and uses the result from previous researches as input to identify the best feature in detecting the fast attack.

5.0 Feature Analysis

Revealing the reason behind the selection of the feature may increase the confidentiality in selecting the feature inside the detection model. Despite the importance, to the best of our knowledge, most of the researcher more concentrated on detection technique that they use rather than mentioning the reason behind the selection [12] especially for fast attack detection. Therefore, it is a good opportunity to reveal the reason behind the selection of the features since none of the previous researcher focus on this area especially for fast attack detection. This section will reveal the reason behind the selection of the proposed feature. Moreover, the influence and relationship of the feature will also be revealed in this section. The analysis will be divided into 2 parts which is basic feature analysis and derived feature analysis. Below are the detail discussions for both of the analysis.

5.1 Basic Feature Analysis

This section will discuss the reason behind the selection of the basic feature in the research. Timestamp has been used by researchers as one of the feature in recognizing the intruder inside the network especially in log correlation technique. Moreover, using timestamp, we may identify when the intrusion occurs inside the network. Furthermore, timestamp can also be used to detect the intrusion at the initial stage where most of the penetration uses scanning and flooding mechanisms to identify any vulnerabilities inside the organization. This can be done by using timestamp to compute a derived feature such as duration of the certain connection from a host or hosts to targeted destination. Therefore timestamp is chosen as one the basic feature in this research. Duration has been declared as one of the basic feature by KDDCUP99. Using this feature, the length of the connection can be identified and the derived feature can be computed. In this research we chose one second as the duration to compute the derived feature. Therefore, duration is one of the important basic features because using the derived features, the fast attack launch by the attacker can be detected easily.

IP Address also provides significant information to the identification of the attacker [25]. Besides recognizing the attacker, the compromised machine can also be identified using the IP address. This can help the administrator to increase the security of the vulnerable server for future protection. Therefore, IP address has been selected as one of the basic feature for this research. Beside the IP address, Services and Flag also have been selected as a feature to detect the fast attack. This set of feature can be estimated in real-time using conventional low-cost computing system [25]. Therefore these features can be used as a basic feature to detect the fast attack.

As a conclusion, we have encompassed the reason behind the selection of the basic feature which is significant as a guideline for future research in detecting the fast attack.

5.2 Derived Feature Analysis

Src_count, srv_count and dst_count has been categorized as derived feature in detecting the fast attack. Statistical approach has been used to reveal the influence and relation of the src_count and srv_count feature in detecting fast attack. Meanwhile, comparison approach is used to reveal the relation of the dst_count feature. Below are the detail explanation for each of the derived feature.

5.2.1 Dst_count

Dst_count is one of the features inside the KDDCUP99 and has become one of the popular features that have been selected by previous researches. For example, Chebrolu et al (2005) introduced 17 features in detecting fast attack and dst_cout is included as one of the important feature in their research. Other researchers such as Sung and Mukkamala (2003) proposed 8 feature in recognizing probe and 20 feature in recognizing DOS. Both of the attack can be classified as fast attack. Anazida et al, (2006) also proposed 6 feature in detecting the intrusion activity. The features proposed by the previous researchers contain dst_count as one of the important feature in detecting the intrusion activity. Based on the previous researches, we make a comparison to identify the most useful feature in detecting intrusion activity especially fast attack. The result from the comparison, we manage to identify the minimum set of feature that can be used to detect the fast attack. The set of feature selected will be mapped with the proposed feature to identify the relation and reason behind the selection. Figure 2 and 3, illustrate the comparison between result obtained from previous research for DoS and Probing attack respectively. The description of the feature represented inside Figure 2 and 3 is given in Appendix 1.

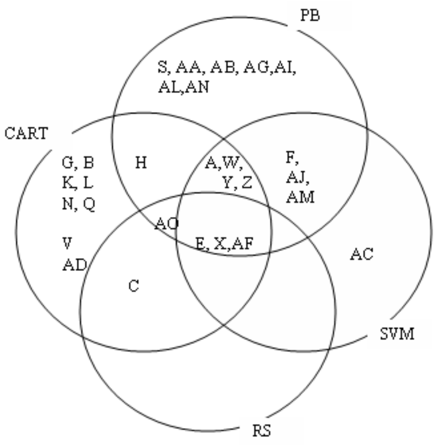

Figure 2 : Feature For DOS Activity

As depicted in Figure 2, feature E, X and AF are important and is constantly selected by all of the four approaches. Feature A, W, Y and Z can also be considered important since it has been selected by 3 different approaches. For this research, we drop features Y and Z due to difficulties in identifying the SYN error using real time application. This constraint is due to the extra packet sent by windows environment which may affect the result of detection [32].

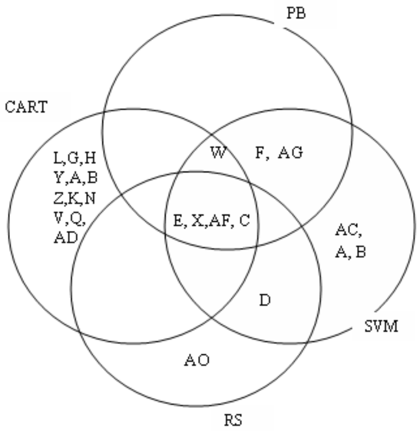

Figure 3 : Feature of Probe Activity

Figure 5 illustrates the features that are useful to detect the probe activities. Features E, X, AF and C are important features as it had constantly been selected by all four approaches. Feature W can also be considered useful since it had been selected by 3 different approaches. As a conclusion, in our research we only select features C, X, A and W for detecting the fast attack. Feature AF has been drop because this feature is a member of host based traffic feature which is useful to detect the slow attack rather than fast attack.

5.2.2 Src_count

This feature is used for detecting fast attack launched by single host to multiple hosts. By selecting this feature, it may help to detect the attacker at the initial stage because host scanning is one of the tools that can be used to launch reconnaissance and scanning activity. Furthermore, there is a question on how the feature can influence the result of the detection of the attacker. This can be validated using statistical value from likelihood ratio test and Nigelkerke’s respectively. Table 2 shows the value of the likelihood ratio statistic after the feature include inside the model. Therefore by referring to equation (1), the value of the baseline model (without the predictor) can be computed. When only the constant was included, -2LL = 295.238. but now src_count has been included this value has been reduce to 14.666. This reduction means that the features has a significant influence at predicting the outcome(attack). Table 1 also show the Nigelkerke’s value for the new model which is 0.976. If the value is close to one, it indicated that the predictor give good influence to the model in predicting the outcome [33]. From the result, it shows that the value is close to one which means that the feature selected in this research gives a good influence to the model in predicting the outcome.

Table 2 : Src_count Analysis

-2 Log likelihood	Nagelkerke R Square
14.666	0.976

5.2.3 Srv_count

Srv_count had also been selected as one of the features in this research. This feature also gives a significant influence to the research and the validation is made using the same test with the previous feature. Table 3 shows the value of the likelihood ratio statistic after the features was included inside the model. Meanwhile, Table 4 show the chi-square (x^2) value of the new model. When only the constant was included, the value of the baseline model is ,-2LL = , 431.176 but now src_count has been included inside the model and this value has been reduce to 35.689. This reduction means that the feature has a significant influence at predicting the outcome (attack). Table 3 also shows the Nigelkerke’s value for the new model which is 0.945. The Nigelkerke’s value is very close to one which means that the feature selected in this research give good influence to the model in predicting the outcome.

Table 3 : Srv_count Analysis

-2 Log likelihood	Nagelkerke R Square
35.689	0.945

6.0 Conclusion and Future Research

Selecting a good feature is very important because it gives a significant contribution to the intrusion detection system in terms of accuracy of detection. Nowadays, most of the researchers only use the selected features in their research without mentioning the reason of selecting the feature in predicting the intrusion activity. The previous researchers also did not mention the influence of the selected features to the system developed in their research. Understanding the relation and influence of the feature before using them may help to reduce the possibilities of selecting unnecessary feature which may give an impact in detecting the intrusion activity especially fast attack since fast attack is used in the initial stage of an attack where attackers use it to begin their attack inside the network. Therefore identifying the attacker earlier may help the administrator to overcome further damage caused by the attacker. In this research, we manage to reveal the influence of the feature in predicting the detection of the intrusion especially fast attack using statistical approach. In addition, the researcher who concentrates on developing the intrusion detection system can benefit from the results provided by this research as they can consider these features when addressing features selection issues inside their research. Besides validating the influence of the features, we also manage to introduce the minimum feature that can be used to detect the fast attack. Although the rough set theory was able to produce minimum feature, but the features selected did not state the basic features which is important to compute derive features. Furthermore, the research also provides a general detection and did not focus on the fast attack detection.

For our future work, we would like to develop a system using the proposed features and test the system in a real time environment. Furthermore, the future work will also make use of real-traffic from other organizations to produce more accurate results.

References

[1] V. Anil Kumar. (2004). Sophistication in Distributed Denial-of-Service Attack on the Internet. *ACM, Current Science, Vol 87, No. 7,* 10 October 2004.

[2] Fang-Yie Leu & Tzu-Yi Yang. (2003). A Host-based Real Time Intrusion Detection System with Data Mining and Forensic Techniques. *In Proceeding of IEEE Conference, 2003.*

[3] Cavusoglu, H., Mishra B. K. & Raghunathan, S. (2004). The Effect of Internet Security Breach Announcements on Market Value of Breached Firms and Internet Security Developers. *International Journal of Electronic Commerce, Vol 8, pp,4.*

[4] Karl Levitt. (2002). Intrusion Detection: Current Capabilities and Future Directions. *In Proceeding of the 18th Annual Computer Security Applications Conference, IEEE, 2002.*

[5] Wang Y., Huang GX. & Peng DG. (2006). Model of Network Intrusion Detection System Based on BP Algorithm. *Proceeding of IEEE Conference on Industrial Electronics and Applications, IEEE, 2006.*

[6] Wenke Lee. (1999). A Data Mining Framework for Constructing Feature and Model for Intrusion Detection System. PhD thesis University of Columbia.

[7] CEH Training Module, 2005.

[8] Faizal MA., Asrul HY., Shahrin S. (2007). An Earlier Detection Framework for Network Intrusion Detection System. *In Proceeding of the Second International Conference on Advances in Information Technology, Bangkok, 1 – 2 November 2007.*

[9] A. Lazarevic, L. Ertoz, V. Kumar, A. Ozgur and J. Srivastava. (2003). A Comparative Study of Anomaly Detection Schemes in Network Intrusion Detection. *In Proceeding of SIAM International Conference on Data Mining, 2003.*

[10] Shahrin S., Faizal MA., Asrul HY. (2007). Toward Early Detection of Network Intrusion. *In Proceeding of Information Technology and National Security Conference, Riyadh, 1-4 December 2007, Saudi Arabia.*

[11] S. Robertson, E.V. Seigel, M. Miller, M., S. J. Stolfo. (2003). Surveillance Detection in Hugh Bandwidth Environments. *In Proceedings of the DARPA information Survivability Conference and Exposition, IEEE,2003.*

[12] I. V. Onut and A. A. Ghorbani. (2006). Toward a feature classification scheme for network intrusion detection*. In Proceeding of the 4th Annual Communication Networks and Services Research Conferences (CNSR'06), IEEE, 2006.*

[13] H.G. Kayacik, A. N. Z. Heywood, M. I. Heywood. (2005) Selecting Features for Intrusion Detection: A Feature Relevance on KDD 99 Intrusion Detection Datasets. *In Proceeding of the 3rd Annual Conference on Privacy, Security and Trust, St. Andrew, NB, Canada, October 2005.*

[14] D. Caulkins, Lee., J, Wang., M. (2005). Packet-s. Session-Based Modeling for Intrusion Detection Systems. *In Proceedings of the International Conference on Information Technology: Coding and Computing. IEEE, 2005.*

[15] M. Mellia, M. Meo, L. Muscariello. TCP Anomalies: Identifying and Analysis. *Distributed Cooperative Laboratories: Networking, Instrumentation and Measurement, SpringerLink.*

[16] S. Chebrolu, A. Abraham, J. P. Thomas. (2004). Feature Deduction and Ensemble Design of Intrusion Detection System. *Journal of Computer and Security (2004). Elsevier Ltd.*

[17] http://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html

[18] A.H Sung, S. Mukkamala. (2003). Identifying Important Feature for Intrusion Detection Using Support Vector Machines and Neural Network. *In Proceeding of International Symposium on Application and the Internet. 2003. pages 209-217.*

[19] Anazida Z., Aizaini MM., Mariyam S. (2006). Feature Selection Using Rough Set in Intrusion Detection. *In Proceeding of the IEEE Region 10 Conference TENCON 2006, IEEE, 2006.*

[20] Lakhina A., Crovella M., Diot MC. (2005). Mining Anomalies Using Traffic Feature Distribution. Proceeding of the ACM SIGCOMM, 2005.

[21] Moore AW and Zuev D. (2005). Traffic Classification Using Statistical Approach". Passive and Active Measurement Workshop 2005.

[22] Karagiannis, T., Papagiannaki, K. & Faloutsos, M. (2005). BLINC: Multilevel Traffic Classification in the Dark. *In Proceeding of ACM SIGCOM'05*. USA.

[23] Chyssler, T., Nadjm-Tehrani, S., Burschka, S. & Burbeck, K. (2004). Alarm Reduction and Correlation in Defense of IP Network. In Proceedings of the 13th IEEE International Workshops on Enabling Technologies: Infrastructure for Collaborative Enterprises (WETICE'04), 2004, pages 229-234.

[24] Avinash, S., Ye, T. & Bhattacharrya, S. (2007). Connectionless Portscan Detection on the Backbone. In Proceeding of Malware Workshop Conjunction with IPCCC, 2007.

[25] Gavrilis, D & Dermatas, E. (2005). Real-Time Detection of Distributed Denial of Service Attack Using RBF Network and Statistical Feature. *International Journal of Computer Network*, Vol 48, pp 235-245.

[26] Xiao, B., Chen, W., He, Y. & Edwin H.-M.S. (2005). *In Proceeding of IEEE* 11^{th} *International Conference of Parallel and Distributed System, Vol 1, pg. 709-715.*

[27] http://www.tcpdump.org

[28] http://www.bro-ids.org

[29] Jalili R., Fatemeh IM., Morteza A., Hamid RS. (2005). Detection of Distributed Denial of Service Attacks Using Statistical Pre-processor and Unsupervised Neural Network. *ISPEC, Springer-Verlag Berlin Heidelberg, 2005.*

[30] Caulkins BD., Joohan L., Morgan CW. (2006). Bootstrapping Methodology for the Session-Based Anomaly Notification Detector (SAND). *ACM, Melbourne 2006.*

[31] McHugh J., Christie A., Allen J. 2000. "Defending Yourself: he Role of Intrusion Detection System". *Proceeding of IEEE, Software, 2000.*

[32] Amol Shukla and Tim Brecht. (2006). TCP Connection Management Mechanisms for Improving Internet Server Performance. In Proceeding of HotWeb 2006.

[33] Andy Field. (2005). *Discovering Statistic Using SPSS, 2nd edn*, Sage Publication London.

APPENDIX 1

Label	Network Data Features	Label	Network Data Features	Label	Network Data Features	Label	Network Data Features
A	Duration	L	logged_in	W	count	AH	Dst_host_same_srv_rate
B	Protocol_type	M	num_compromised	X	Srv_count	AI	Dst_host_diff_srv_rate
C	Service	N	root_shell	Y	Serror_rate	AJ	Dst_host_same_src_port_rate
D	Flag	O	su_attempted	Z	Srv_serror_rate	AK	Dst_host_srv_diff_host_rate
E	Src_byte	P	num_root	AA	Rerror_rate	AL	Dst_host_serror_rate
F	Dst_byte	Q	num_file_creations	AB	Srv_rerror_rate	AM	Dst_host_srv_serror_rate
G	Land	R	num_shells	AC	Same_srv_rate	AN	Dst_host_rerror_rate
H	Wrong_fragment	S	num_access_files	AD	Diff_srv_rate	AO	Dst_host_srv_rerror_rate
I	Urgent	T	num_outbound_cmds	AE	Srv_diff_host_rate		
J	Hot	U	is_host_login	AF	Dst_host_count		
K	Num_failed_login	V	is_guest_login	AG	Dst_host_srv_count		