

The Prevention Threat of Behavior-based Signature using Pitcher Flow Architecture

Deris Stiawan^{1&2}, Abdul Hanan Abdullah², Mohd. Yazid Idris²

deris.stiawan@gmail.com, {hanan, yazid}@utm.my

¹⁾ *Computer Engineering Department, Faculty of Computer Science, Sriwijaya University, Indonesia*

²⁾ *Faculty of Computer Science & Information System, Univesiti Teknologi Malaysia*

Summary

In recent years, Intrusion Prevention System (IPS) has been widely implemented to prevent suspicious threats. Unlike the traditional Intrusion Detection System, IPS has additional features to secure the computer network system. IPS is an access control device with a prevention function, which enforces a network security policy, is a helpful device that allows for more granular blocking action.

In this paper, we propose a new prediction and prevention method with behavior-based detection, this method is called pitcher flow. We describes the habitual activity of the performance an overall network with a new algorithm for identifying and recognizing the normal behavior of user activities in the internal network. First, we define behavior activity by duration of activity conducted and active connection. Second, we categorize packets into class/type, identifying parameters by classifying the packets. Finally, we use the pitcher flow mechanism to identify and recognize suspicious threats. This paper also describes an algorithm for the complexity of the suspicious response.

Key-words:

Behavior-based detection, Hybrid intrusion prevention, Identify habitual activity.

1. Introduction

In the last few years, the Internet has experienced explosive growth. Along with the widespread evolution of newly emerging services, the quantity and impact of attacks have been continuously increasing as well. Intrusion Prevention System (IPS) has become an essential component of computer security to predict and prevent attacks. They monitor, identify and recognize all real-time packets inbound and outbound. IPS, which proactively combines the firewall technique with that of the Intrusion Detection System, prevents attacks from entering the network by examining various data records and the detection demeanor of the pattern recognition sensor, when an attack is identified, intrusion prevention blocks and logs the offending data.

According to CSI/FBI survey [9], the company business has dollar amount of loss by type of attack.

Meanwhile, to secure the systems, the enterprise uses several technology security systems, and almost 69% of which use intrusion prevention to defend from threat and attack.

The signature is the primary means to identify activity in network traffic, and the host performs the detection of inbound and outbound packets and to block that activity before damage and network resources are accessed. However, IPS can effectively detect suspicious threats that are already known from a list of signatures. Common Vulnerability Exposure (cve.mitre.org) is a list of intruding products, and there are several IPS devices with proprietary standards. For this reason, many IPS vendors dedicate a large number of engineers to continuous observation of suspicious threats and update their product database with new signatures as threats arise.

From our observation, many devices are defined through the process of identifying suspicious threats and rogue activity from inbound network traffic. Unfortunately, computer misuse or malicious activity from inside the network not the main issue in past research, and it is important to understand how to identify a compromised system by inspecting outbound traffic. It is a broader term that encompasses indentifying a variety of suspicious, rogue and malicious threats in outbound user activity. They are, (i) spam e-mail, (ii) theft of intellectual property, (iii) computer zombie from inside network to trigger attack, and (iv) internal system that launches scanning and exploits until it launches a DoS attack against the host on the Internet. Therefore, this action triggers revenge action from outside.

The habitual activity between activity of higher transaction size and concurrent connection, definitely, affects the performance of the overall network. In this paper, we proposed a new system architecture for IPS, named Pitcher Flow, which is a Behavior-based detection mechanism to detect, identify, recognize, and react to a suspicious threat. The proposed method also addresses how to identify common knowledge as an activity profile between new algorithms for identifying the normal behavior of user activities. Our method is expected to help security officers (IT Manager and Administrator) to be aware of status user profile's activities.

2, we present the behavior-based flow with algorithm detection mechanism.

In this section, according to [6] and from the result survey, we combine it between our dataset from capturing data activity. Wherein audit records contain information such as frame protocol (source and destination IP address between port address). Both data stream real-traffic from machine contain over a 100, 000 audit data records.

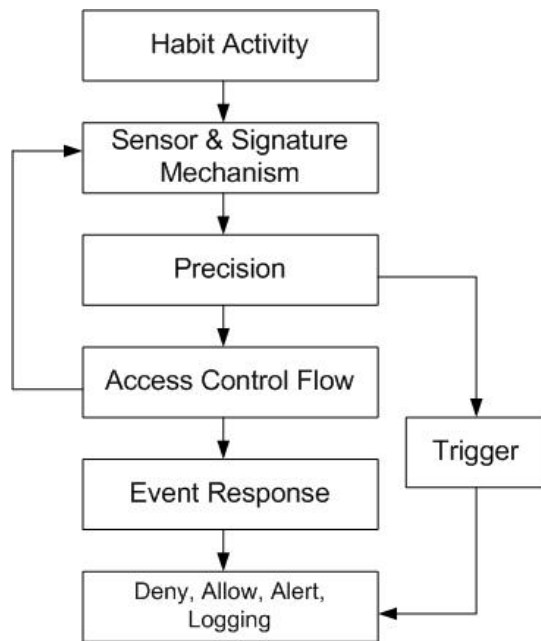


Figure 2 : The flowchart

Meanwhile, the suspicious threat attack from valid inside user is constrained. The emphasis is character behavior activity, however, the composite pattern detection and anomaly-based detection increases the detection demeanor of pattern recognition sensor.

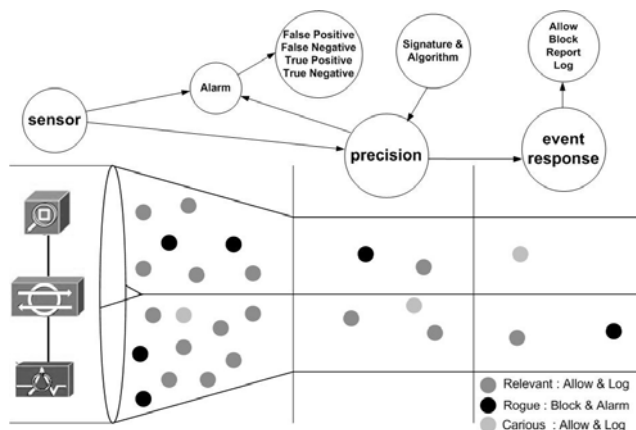


Figure 3. Identify and recognizing suspicious mechanism

Furthermore, from our observations, there are two habitual behavior activities: (i) media rich with activity higher transaction size, and (ii) transactional with activity concurrent connection, as follows :

Table 1. Example Behavior activity level of higher transaction size, between more transactions per connection

Activity	Applications
WWW	Browsers, http
Collaborative Workspaces	Google Apps, Google Readers, blogs.
Download - Upload	P2P, FTP, updates process : System Operation, Anti Virus, Applications
Streaming video	Youtube, Realtime, Quick time, YM Webcam, etc.
Data Replication	backup data, mirroring data in other sites.
Remote Login	SSH access, WinSCP, Putty
Remote VNC	(remote desktop), to remote other PCs in network.
Mail	SMTP, POP, IMAP
Spamming mail	many send to mail address

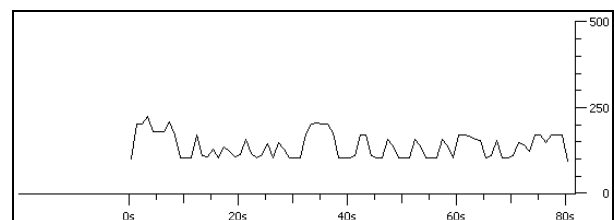
Table 2. Example Behavior activity level of concurrent connection, between higher connection rates

Activity	Applications
E-Commerce	https
Internet Messaging	YM!, mIRC, ICQ, Pidgin, Adium, GTalk, Skype, etc
VoIP	Skype, YM Voice, etc
Game online	Ragnarok, HalfLife, Age of Empires, Ayo Dances, etc
Scanning	Scanning port use script tools

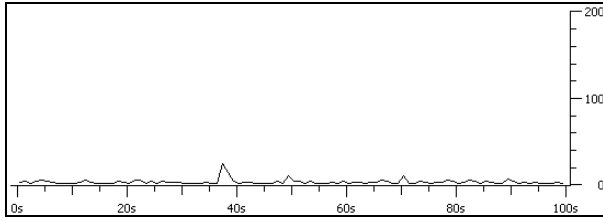
(a) Sensor & Signature Mechanism

From real-traffic, we distinguish between normal activity and malicious activity, such as P2P BitTorrent, and Slammer worm. The similarities are often minimum utilization Ethernet packets.

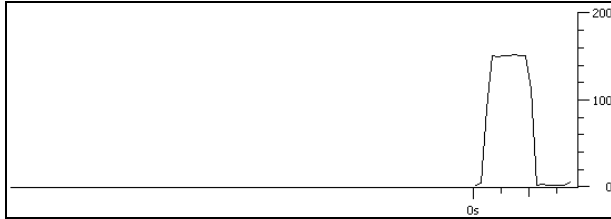
Payload Classification



(a) Bit Torrent p2p. Ethernet : 100%, IP : 100%, TCP : 0.38%, Data : 0.38%, UDP : 99.62%, Data : 99.54%



(b) Slammer worm. Ethernet : 97.43%, IP : 84.16%, TCP : 45.77%, Data : 0.37%, UDP : 36.86%, Data : 0.28%, NetBios session : 32.69%, ARP : 13.27%



(c) Normally activity, containing a few JPEG. Ethernet : 100%, IP : 98.97%, TCP : 0.17 %, UDP : 98.81%, ARP : 0.48%

Figure 4, application-specific bit string of the payload, (a) Bit Torrent P2P, with capture file of two torrent clients communicating without DHT or peer exchange, (b) Slammer worm, sending a DCE RPC packet, (c) Normal activity, a simple capture containing a few JPEG pictures one can reassemble and save to a file.

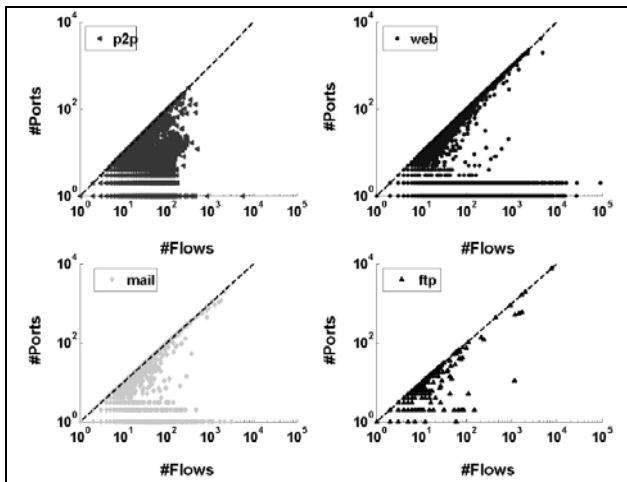


Figure 5. Number of source port versus number of flows per source IP address [10].

To quantify how the number of used source ports may distinguish client from server behavior, the examine the distribution of the source port a host uses in the traces [10]. Figure 5, plots number of flows (x-axis) vs the number of source port (y-axis) that each source IP uses for 15 minutes.

As in show Figure 5. Number of source ports vs number of flows per source IP address in the UN1 trace for a 15 –minute interval for four different applications [10]. In the client-server application (Web, FTP, mail), most points fall on the diagonal or horizontal line for small values in the y-axis (number of used ports). In P2P, point are clustered in between the diagonal and the x-axis.

Table 3 : The notations of parameters used in sensor and signature

Notations	Descriptions
S_1	source IP Address
S_2	source Port Address
Des_1	destination IP Address
Des_2	destination Port number
TCP/UDP	protocol uses
PY	Payload
RG	Regex
MAC_1	MAC source address
MAC_2	MAC destination address
LE	Length packet
URL	URL address
FR	Frame
FL	Flags
WS	Windows size

Previous research has shown that anomalous behavior may be determined by simply inspecting the size of the packet, the identifying the type of attack based on payload size [7]. In this experiment, the payloads were a determined to be a constant size, The applications have a payload classification from their characteristics (string, port, flag, protocol, payload and Regex) that can be examined by a sensor. Table 4 presents sample data of a string payload:

Tabel 4. Sample data string

Application	String	Protocol
BitTorrent	0000000d0600\0x13	TCP/UDP
eDonkey2000	0xe319010000	TCP/UDP
MSN Messenger	"PNG" 0x0d0a	TCP
IRC	"USERHOST"	TCP
YM!	"ymsgr"	TCP
nntp	"ARTICLE"	TCP
SSH	"SSH"	TCP

In Table 5, we define regular expressions (regex) that can be used in selectors to define ranges of values instead of defining each possible value separately. Regex can match with pattern recognizing in layer 7 Application. Therefore, this approach able to combine with global signatures database, which in real-time anomaly analyzer system, proposed by [8].

Table 5. Sample data regex

Suspicious	Regex String
HTTP Overflow	pcr:/"(\w+)\s*=\s*(' ")^2]*\[[^2]*\][^2]*\s*\.;.*new\s+RegExp\s*(\s*\1 new\s+RegExp\('' ")^3]*\[[^3]*\]/smi";
BitTorrent	.*[Ii][Nn][Ff][Oo][Hh][Aa][Ss][Hh]=.*
eDonkey2000	^[C5\D4\E3-\E5].\?.\?.\?.\?([ABCDEF[ABCE\A0\A1\A2\A3\A4]\?[-~]\$)
MSN Messenger	^(ver [-~]*msnftp\x0d\x0aver msnftp\x0d\x0ausr method msnmsgr:)
IRC	^(nick[\x09-\x0d ~]*user[\x09-\x0d ~]*: user[\x09-\x0d ~]*:[\x02-\x0d ~]*nick[\x09-\x0d ~]*\x0d\x0a)
YM!	^(ymsg ypns yhoo).??.??.??.[lwt].*\xc0\x80?
nntp	^20[01][\x09-\x0d ~]*\x0d\x0a[\x09-\x0d ~]*AUTHINFO USER 20[01][\x09-\x0d ~]*news
SSH	([A-Za-z0-9._-]+) sshd\([(0-9)+\):\[^\]]+\)

Regex are 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 observe that identifying suspicious activity, on receiving packet, in Figure 1, system it receives an identification field from each packet header, including S_1 , S_2 , Des_1 , Des_2 , MAC_1 , MAC_2, \dots , and classifies them into classes/types, Identifying parameters by comparing the packets.

Furthermore, we wish to distinguish normal or curious activity by classes/types, such as (i) IP Address between MAC Address, (ii) number of connection, (i.e. 1 to 1 connections, 1 to N connection or N to 1 connection), and (iii) payload. After classifying a packet, it sends the packet to the precision section. Algorithm 1 lists the details.

Algorithm 1 : Identification & declaration Packet data

type packet_data record, Parameters on table 1

precision \Leftarrow identification result

$r \Leftarrow$ risk rating

packet_data = 1 // input data between real-time detection

while packet_data \diamond 0 do

 identification (precision)

 risk_rating (precision, rr)

 trigger (precision)

 event_response (precision, r_r)

end while

(b) Precision

In this section, we describe the process of recognizing a detailed suspicious threat. We identify the habitual activity in the previous section, based on Table 1 and Table 2, which are the behavior activity levels, and next a process identifies the payload, shown in Table 4. Algorithm 2, lists the logic of how identification allows, blocks, logs and reports, respectively.

Algorithm 2 : Identification and recognize

Procedure Identification (precision)

 read (packet_data) // refers from figure 3, table 4 & Algorithm 1

if packet_data = rule **then**

 precision is block

else if packet_data \neq rule **then**

 precision is allow

else if packet_data \neq rule and suspect **then**

 precision is log

else

 precision is report

end if

end procedure

The set of signatures affects the number of attack identifying detected. The accuracy affects the correctness of deciding whether an attack exists in real-traffic, notifying the logging system of an attack based on the list in the database. The sensor trigger alert produces an alarm, which is a False Positive (FP), False Negative (FN), True Negative (TN) or True Positive (TP) in every IPS, the main focus is to analyze the alarm and tune out FPs. The session that generates the through Algorithm 3 is described as follows.

Algorithm 3 : Alarm set

procedure trigger (precision)

read (precision)

if precision = allow **then**

 alarm = TN

else precision = block **then**

 alarm = TP

else if precision = log **then**

 alarm = FP

else

 alarm = FN

end if

end procedure

(c) Access Control Flow

In this section, the list data set from the previous section uses the input in the access control and the risk rating to create the result (information, low, medium, or high). The risk rating then collects and lists it, in this section. The output from this section is the form of rules to generate event response or send it back to sensor for the sensor mechanism to identify. Algorithm 4 lists the details.

Algorithm 4 : Access system

```

Procedure risk_rating (precision, r_r) // to pitcher flow
  filter precision
  if precision = block then
    r_r is high
  else if precision = log then
    r_r is medium or r_r is low
  else precision = report then
    r_r is low or r_r is information
  end if
end procedure

```

(d) Event Response

The event response, in response to the traffic by performing actions, such as : deny, alert, block, and log

Algorithm 6 : Response

```

procedure event_response (precision, r_r)
  read (precision, r_r)
  if precision = allow or r_r = information or r_r = low then
    response is allow
  else if r_r = medium then
    response is allow and log
  else if r_r = high then
    response is block
  end if
end Procedure

```

4. Conclusion

The behavior-based is a complex, multiple problem to identify real-time traffic from internal user, under the variety normal activities. In this paper, a new model has been proposed for identification and recognition behavior-based attack, with analyze real-traffic from habit activity internal user. This approach using the pitcher flow with signature, accuracy, and logging systems to identifying, recognizing and reacting before threat damage and access network resources. The result indicates this approach able to combine with other defense systems between firewall and network monitoring. In the future, we would experiment benchmark algorithm in real-traffic network.

Acknowledgments

This research is supported by the Ministry of Science, Technology and Innovation Malaysia and collaboration with Universiti Teknologi Malaysia.

References

- [1] Thomas Karagiannis, et al, "BLINC: Multilevel Traffic Classification in the Dark" SIGCOMM'05, 2005
- [2] I-Wei Chen, et al, "Extracting Attack Session from Real Traffic with Intrusion Prevention Systems", IEEE Communication Society proceeding", 2009

- [3] Masayoshi Mizutani, et al, "Behavior Rule based Intrusion Detection", CoNEXT Student Workshop'09, 2009.
- [4] Qingbo Yin, et al, "A New Intrusion Detection Method Based on Behavioral Model", Proceedings of the 5th World Congress on Intelligent Control, 2004
- [5] Qingbo Yin, et al, "A New Intrusion Detection Method Based on Behavioral Model", Proceeding of the 5th World Congress on Intelligent Control and Automation, 2004.
- [6] Hyeun-Suk Rhee, et al "Self-efficacy in information security : it influence on end users information security practice behavior", Journal Computer & Security 28, 2009
- [7] Mahoney, M, et al, "Network Traffic Anomaly Detection Based on Packet Bytes", Proceeding ACM-SAC, pp.346-350, 2003
- [8] Taras Dutkevych, et al, "Real-Time Intrusion Prevention and Anomaly Analyze System for Corporate Network", IEEE International Workshop on Intelligent Data Acquisition and Advanced Computing Systems, 2007
- [9] Robert Richardson, "CSI Computer Crime & Security Survey 2008", 2008.
- [10] Thomas Karagiannis, et al, "BLINC: Multilevel Traffic Classification in the Dark", SIGCOMM'05, 2005



Deris Stiawan. Holds an M.Eng from UGM, Indonesia, since 2006, he is Computer Science faculty member at UNSRI, Indonesia. Now since 2009 he joined research in Information Assurance and Security Research Group (IASRG) UTM. His professional profile has derived to the field of computer network and network security, specially focused on intrusion prevention and network infrastructure. He is Ph.D student working in

prevention behavior attack.



Abdul Hanan Abdullah. 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, Universiti Teknologi Malaysia. His research interest is in Information Security. He is also a head of Pervasive Computing Research Group (PCRG) UTM and member of ACM.



Mohd. Yazid Idris. PhD, 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. 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.

