# Early Phase Warning Solution About System Security Based on Log Analysis

Hieu Le Ngoc
*HCMC Open University*
*PhD. Student of Posts and Telecommunications Institute of Technology, HCMC Branch*
Ho Chi Minh City, Vietnam
hieu.ln@ou.edu.vn

Tran Cong Hung
*Post and Telecommunication Institute, HCMC Branch*
Ho Chi Minh City, Vietnam
conghung@ptithcm.edu.vn

Nguyen Duc Huy
*Sales department of VNPT Binh Duong*
Ho Chi Minh City, Vietnam
huynd.bdg@vnpt.vn

Nguyen Thi Thanh Hang
*Department of Information Technology*
*Open University*
Ho Chi Minh City, Vietnam
hangntt2398@gmail.com

***Abstract*— Nowadays when the technology develops day by day, safety information and security system have become one of the biggest issues that many people are concerned about and research in order to find solutions to ensure the security of information systems. At present, the most important thing that network administrators care about is minimizing the damage to the enterprises when cyber security crimes invade the information system or attack the system anytime. Therefore, how they can proactively protect information, ensure the privacy of users, how to improve the confidentiality and security of information in business system. To solve the problem above, the research *"Early warning solution about system security based on log analysis"* is extremely necessary, because it can help to warn and detect early-attacking information when there are abnormal signs in systems via logs. By inheriting the advantages, as well as overcoming the limitations of those previous related topics in the world, in this paper, we would like to develop and to build an early warning application tool capable of interacting, monitoring, adjusting, notifying from the scouting process of the attackers. The new proposed system in this research is based on the characteristics of attack techniques, components and packets through the system, we design a data collection model of Logs input and output. In order to detect and explore abnormal activities in network which are harmful for information systems, we use the log analysis of the information system with the method of analyzing APT target. We study the APT and its characteristic to build the knowledge that we can use an advanced AI algorithm, Deep & Wide Learning algorithm. With logs testing data we conducted in our LAN, we have reached some good results showed the potential of our proposals.**

***Keywords*— Network security; Advanced Persistent Threat – APT; Early warning solutions; Log analysis; Artificial Intelligence (AI); Deep - Wide Learning.***

## I. INTRODUCTION

In line with the strong development trend of the Industrial Revolution 4.0 today, the Internet has increasingly demonstrate an important "vital" position for human life in general and in Technology in particular. Beside the advantages, the Internet also brings many risks for businesses and organizations through the network.

In fact, the current situation of network attacks is growing with the level of danger, which increases day by day. Every day, the world faces thousands of attacks related to cybersecurity and security, costing trillions of dollars due to data theft or hacking into information systems crucial business and organization. According to VNCERT, as of June 25 in 2018, the system has recorded 1,122 phishing attacks, 3,200 interface changes and 857 incidents of malware distribution on the Website. These threats focus on two main groups: stealing sensitive information from organizations, individuals, banks and installing ransomware, money laundering, etc. by targeted attacks APT (Advanced Persistent Threat) - One of the methods of attack done by cyber security criminals. With sophisticated techniques, the main goal of an APT attack is to break into a system to gain continuous access, steal data more than cause damage to the network or organizations, besides APT often targets organizations in high value information industries such as defense, manufacturing and financial industry. Comprehensive prevention of targeted attacks APT still face many difficulties even though organizations and businesses still spend billions of dollars each year on prevention.

From this fact and realizing the importance of ensuring network security against APT attacks is increasing, "Early warning solution about system security based on log analysis" was studied to give early warning to administrators about the type of APT attack based on LOGS system. This solution is implemented based on the method of analyzing APT's targeted attacks and separating Logs and combining AI algorithms, namely Machine Learning - Deep & Wide Learning to analyze data, carry highly effective in early warning of attacks to be faster and more accurate.

In this paper, we are going to introduce four important sections. Section 1, we talk about an overview of network security and targeted APT (Advanced Persistent Threat) attacks. Then, processing Logs and combining AI algorithms, namely Machine Learning - Deep & Wide Learning to analyze data are discussed in Section 2. Section 3 we will propose "Early warning solution about system security based on log analysis" and Experimental research model. Finally, in Section 4 will be the results test solution and the results we achieved after the test.

## II. RELATED WORKS

In this section, we will present some works related to the APT. There have been many scientific research on APT conducted by famous scientists in the world.

Firstly, some APT examples were provided in "Assessing Outbound Traffic to Uncover Advanced Persistent Threat" [1] 2011 by Beth E. Binde et al. such as identified and

described threats, proposed technical methods to reduce mitigate threats,... Besides that, Nart Villeneuve and James Bennett discussed about how advanced detection techniques used to identify command and control communications for malware in the both researches “Detecting APT Activity with Network Traffic Analysis” [2], as well as “Advanced persistent threat detection” [3]. These researches focused on using AI to detect threats and APT activities through analyzing network traffic.

In addition, the paper "Advanced Persistent Threat Detection Based On Network Traffic Noise Pattern and Analysis" [4] by Sin Chun Ng and Majid Bakhtiarib in 2016, has achieved many successes. Especially, some APT attacks were detected by analyzing Zero-Day vulnerabilities. As a result, the connection and flow model of each attack was recorded by traffic monitoring tools. In the future, the authors will deploy a proposal framework with real computer networks, real network peripherals and real computer systems instead of virtual environments. In addition, this study is aimed at analyzing the technique of Knock Onock used by hackers to contact C&C Server on request.

Also in this way, the authors Harikrishnan V N and Gireesh Kumar T with their research “Advanced Persistent Threat Analysis using Splunk” [5], they analyzed APTs by using Splunk-Security Information and Event Management (SIEM), and based on the results of the analysis, they developed a set of APT features. Besides, the authors are also proposing an effective approach to determine the perfection of this APT by using a modern technology called machine learning.

Similarly, in Study A, by analyzing a certain large network volume to detect weak signals related to data and other suspicious APT activities. The results of this paper allow security experts to focus their analysis on a small group of servers among thousands of machines typical for large organizations. In addition, a tutorial entitled "Cyber Security Monitoring and Logging Guide" [8] also presented details on how to monitor and log cyber security events. It provides practical advice on how to effectively manage logs, handle suspicious events, use cybersecurity intelligence and solve challenges. This topic is studied to allow you to prioritize and manage numerous event logs, build an effective network security monitoring process, and learn about where and how you can get help.

In addition, when they talk about AI, there is a famous topic “Deep & Wide Learning for Recommender Systems” by Heng Tze Cheng and other authors. They presented Wide & Deep learning through linear models and deep neural networks to combine the benefits of memorization and generalization for recommender systems. Basing on the result of their results, we use that as a good foundation to develop our research better.

In Vietnam, most of the domestic research based on international research is using Pattern recognition technologies that combine Machine learning to detect cyber system crimes with Syslogs's powerful alert support. Many solutions to build an intrusion warning system have been implemented effectively and exactly. However, these solutions only deploy the system on a small network segment, so that it has not been able to fully evaluate the performance of the system and the system problems that will happen when they are deployed in reality, the identification samples have not been full, diverse attack types and new emerging attack methods are increasing more and more.

## III. Proposal

The purpose of this research is not only to describe the proposed application, but also to build a model to collect input and output logs data for early warning of APT attacks based on the Logs system (Logs in network infrastructure). By researching and applying AI, specifically Machine Learning - Deep & Wide Learning algorithm, we are able to bring high efficiency and more accurate in our solution of early warning about harmful activities.

### A. Model description

- *Core Layer:* This layer includes core network devices such as Router Load-balancing and Firewall.
- *Access Layer*: Users can connect to the network through Switch Access devices here.
- *Convergence Layer:* This layer includes network components such as DMZ area, Layer 3 Router for routing, central Switch device capable of withstanding heavy traffic, and Server Farm area.
- *Server collects Logs:* Located in the Server Farm area, Logs components from the Host, the Device, the network nodes,…will be focused on this Server.
- *Proposed application:* Located in the server area, here is where the gathered Logs data, from which we can easily connect and analyze Logs.

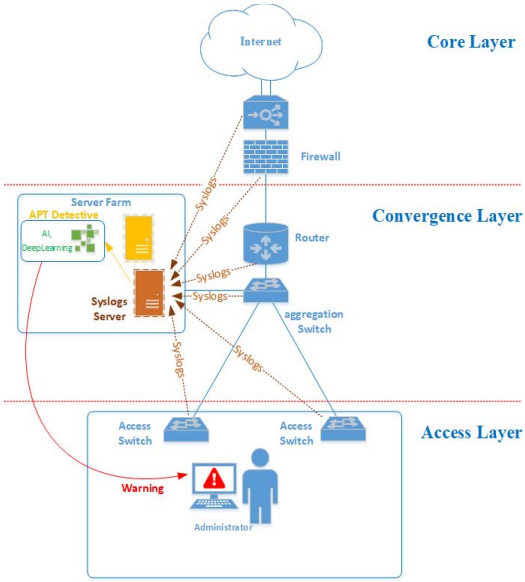


**Figure 1.** Network model of new proposal

### B. Scope of data collection

- *Inputs:* Log data sourced from network devices. Server Syslogs is responsible for collecting raw data including system changes, packets passing through devices, system malfunction and current states of the operating system. Then, this data is filtered according to certain standards which is supported by the Logs software. From the Log data collected above, we need to sort, classify, redefine and homogenize the content for the system to process.

- *Processing technique:* Using the Logs data set, based on pre-attack APT training, from which Deep & Wide Learning application to recognize the Logs with APT pre-attack signs, thereby giving an early attack warning.
- *Output:* After the application has processed, the system will return the results of an early attack signal of APT, from which it can send alerts to users via notification tools such as Email, SMS,…

### C. *Proposal of warning levels*

From the above signs, the system will issue exactly warning levels to the administrator.

- *Warning based on network traffic:* In a short time, a large amount of traffic is recorded by Logs, or exists a received packet with large capacity (Bigsize)
- *Warning based on unusual access:* Telnet, SSH, Remote Desktop
- *System scanning tools:* Ping, Tracer, Nmap Trace

In order to provide the above levels of warning, the system needs to have processing support modules to be able to analyze and give the most accurate results. The main modules used by the system included: Reading Logs and processing data module; Data processing module; Warning module.

- *The reading Logs and processing data module:* This module reads Logs and puts them into the database. These Logs will be modified, and we save the filtered data streams according to the previous configuration. Then, the Logs file is saved as a CSV file and being split into multiple sub-files. Each sub-file contains to 1 from 10 minutes of data streams on the LAN. We can customize and handle Logs reading time, in this paper, we recommend the reading period is 10 minutes and there are around 10.000 data records.
- *Data processing module:* This module will use the Deep & Wide Learning algorithm with training data, which is empirical data based on the above 3 levels of APT warning, the data processing module uses the Deep & Wide Learning algorithm in conjunction with the deep learning Neuron network to classify the layers at risk of attack, virus infection or Trojan. This algorithm **[9]** is a combination of training from linear model and deep learning Neuron network with the strength of memorization and generalization. It is very useful for cases of large data regression and layering over large data with scattered inputs.
- *Warning module:* This module will use the Deep & Wide Learning results in Data processing module, we will have 4 levels of alert level, starting from noticing, reminding, warning and the last is the alarming.

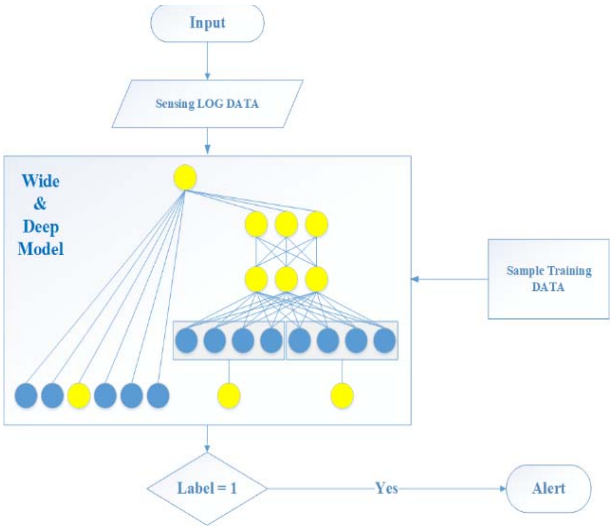


**Figure 2.** Diagram of the proposed algorithm using deep and wide in log analysis.

By proposing an attack warning application based on Logs analysis, this research aims to build a system that can identify dangers before the attack occurrence to warn administrator as earliest as it can. In addition, with the application of the intelligent algorithm proposed as AI, Machine Learning - Deep & Wide Learning in processing raw data Logs for statistical analysis will help administrators have effective tools to support, also as opening up deeper developments in Artificial Intelligence and Machine Learning in information security, towards high accuracy and perfecting the sample data set.

### D. *Building the simulation enviroment of proposal application by Wide – Deep Learning*

To develop the proposal application, we choose to use Logs recording tools, including Wireshark and KiwiSyslogs. These tools can help us to record either connections in system or suspicious events, and then only save records which are filtered by APT packets. The data recorded by the LOG system are raw data, that's why we need to find the suitable algorithms to analysis effectively the above records. And, exactly for this reason, using Wide & Deep Learning algorithm is extremely reasonable, because the special feature of this algorithm is that it can identify rare cases with low rates, as well as large data sets. In addition, the best advantage of Wide & Deep Learning is a combination between Memorization and Generalization, so it will be easy to detect early attacks from APT with high accurate rates.

Specifically, Wide Model is responsible for remembering all early signs of attack with different data fields such as Duration, Bandwidth,… However, this model cannot generalize those Logs to provide an identification or evaluate results. In contrast to Wide Model, Deep Model do generalize but cannot identify some extraordinary cases of LOGS. Consequently, we have to use both models to learn them all, including unusual cases. This combination is necessary and very useful, because APT is multifarious and there are so many different ways when it attacks the system.

The proposal application is a web-installed application that integrates many modules, runs on both Internet and Intranet environment to process, configure functions for it. There are four main functions in our application, including

Management System, Configure Logfile, Viewing traffic through reading Logs, Adjustment Notifications.

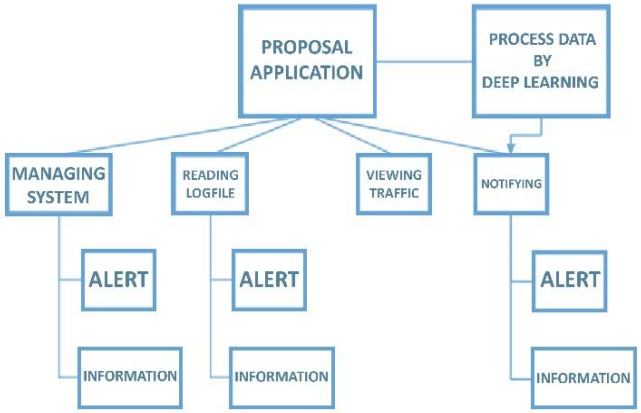


**Figure 3.** Functional diagram of the proposal application

The application of data mining and data science in LAN protection and LAN attack prevention is currently one of the current trends. The combination of Intelligent Algorithms wit existing network monitoring tools is also one of the popular development methods that are highly appreciated to help protect, alert and monitor cyber security in general and LAN especially better.

## IV. RESILTS

To implement the proposed approach above, we have used the network logging tools (Wireshark and KiwiSyslogs) to collect LOGs within the time of 10 minutes. During this test time in the LAN, we have simulated some post-attack and post-hack actions. After selecting logs, we clean the data and use R-Studio integrated Python to analyze.

### *A. Experimental simulation environment proposed application.*

*Method of collecting LOG*: The LOG collection is recorded for about 10 minutes(from 9:00 AM to 9:10 AM). In the process of collecting and recording LOG in the LAN, simulating the stages of pre-attack, reconnaissance, or early attack, includes: Telnet, Nmap-Sweep, SSH, Ping, BigSize, PortStrange, Tracert,…

*Data collected*: is the raw data recorded by the LOG system as raw, including 126,301 log lines stored in the data fields of the INFO fields *(Duration, Source, Destination, Protocol, Length, Info),* separating this field into the required fields based on the APT attack that we mark the logs at risk of pre-attack, and build the training data set in the new INFO fields *(SourcePort, DestinationPort, MessagePhase, Ping, BandWidth, Label)*.

*Training log data:* Because reconnaissance attacks, (or early attacks by APT) are always scattered, and when taken on the recorded LOG data, the ratio of connections at risk of being in the early attack group are few. So, to implement this Deep & Wide – Learning algorithm, we firstly define the fields of basic log data (Base Feature Columns): Protocol, Type, MessagePhase, Ping, Destination, Source, Info, Label.

Besides the basic feature columns, we also have the data fields used for Wide model, which is basically a linear model with relationships between attribute groups (Crossed Columns):

- Influence group between [Protocol] and [Type].
- Influence group between [Type] and [Label].
- Influence group between [BandWidth] and [Protocol] and [Type].

We continue to build Deep model with feature columns, which are columns of element type and properties of type numbers:

- Deep feature: Protocol, Type, Ping, MessagePhase and Label.
- Continuous type features: Duration, length, BandWidth, DestinationPort, SourcePort.

### *B. Experiments and simulation results*

We have 13 data fields with 126,301 observations ( Log stream recorded from the network). For example, in the first case, we take 75% of the dataset as test-dataset, 25% of the dataset will be the training dataset.

```
> str(trainData)
'data.frame':    94725 obs. of  13 variables:
 $ Duration       : num  9260 9269 6550 5995 9433 ...
 $ Source         : Factor w/ 259 levels "::","0.0.0.0",..: 101 10
1 99 101 34 85 97 101 62 101 ...
 $ Destination    : Factor w/ 247 levels "00:32:10:d1:29:62",..: 7
8 82 132 66 83 132 120 78 83 78 ...
 $ Protocol       : Factor w/ 88 levels "0x0004","AFS (RX)",..: 82
16 70 75 79 70 37 82 75 53 ...
 $ Length         : int  90 82 488 54 1466 424 64 106 1466 82 ...

 $ Info           : chr  "55443  >  51717 Len=48" "Standard query
0x321d AAAA flora.web.telegram.org" "NOTIFY * HTTP/1.1 " "50391  >
443 [ACK] Seq=19642 Ack=1073751 Win=527104 Len=0" ...
 $ Type           : Factor w/ 4 levels "HTTPS","Nmap",..: 2 2 2 2
2 2 2 2 2 2 ...
 $ SourcePort     : int  469 474 474 220 474 474 474 474 474 47
4 ...
 $ DestinationPort: int  49 460 460 49 460 460 460 460 460 460 ...
 $ MessagePhase   : Factor w/ 2 levels "NULL","yes": 1 1 1 1 1 1 1
1 1 1 ...
 $ Ping           : Factor w/ 2 levels "NULL","yes": 1 1 1 1 1 1 1
1 1 1 ...
 $ BandWidth      : num  0.01 0.01 0.07 0.01 0.16 0.15 0.5 0.01 0.
17 0.01 ...
 $ Label          : Factor w/ 9 levels "BigSize","Nmap",..: 2 2 2
2 2 2 2 2 2 2 ...
```

**Figure 4.** Train dataset from 25% of the collected Log.

```
> str(testData)
'data.frame':    31576 obs. of  13 variables:
 $ Duration       : num  0.161 0.174 0.246 0.247 0.248 ...
 $ Source         : Factor w/ 259 levels "::","0.0.0.0",..: 148 88
200 200 88 88 144 104 148 88 ...
 $ Destination    : Factor w/ 247 levels "00:32:10:d1:29:62",..: 8
0 144 80 80 197 197 133 80 80 144 ...
 $ Protocol       : Factor w/ 88 levels "0x0004","AFS (RX)",..: 79
79 75 75 75 75 16 16 79 79 ...
 $ Length         : int  1494 571 54 1506 54 54 96 140 1494 18
0 ...
 $ Info           : chr  "Server Hello" "Client Hello" "443  >  17
320 [ACK] Seq=1 Ack=518 Win=28160 Len=0" "[TCP Out-Of-Order] 44
3  >  17320 [ACK] Seq=1 Ack=518 Win=28160 Len=1452" ...
 $ Type           : Factor w/ 4 levels "HTTPS","Nmap",..: 1 1 1 1
1 1 1 1 1 1 ...
 $ SourcePort     : int  474 474 46 474 7 7 474 474 474 474 ...
 $ DestinationPort: int  460 460 15 460 49 49 460 460 460 460 ...
 $ MessagePhase   : Factor w/ 2 levels "NULL","yes": 2 2 1 1 1 1 1
1 2 1 ...
 $ Ping           : Factor w/ 2 levels "NULL","yes": 1 1 1 1 1 1 1
1 1 1 ...
 $ BandWidth      : num  9253 3288 220 6085 218 ...
 $ Label          : Factor w/ 9 levels "BigSize","Nmap",..: 1 1 6
1 1 1 1 1 1 1 ...
```

**Figure 5.** Test data set from 75% of the remaining Log

We carried the log analysis based on the characteristic of APT and early phase of attacking, put in into R-Studio integrated python library, then we apply the Deep & Wide Learning algorithm. The experimental results on R-studio are as follows

```
> BandWidth_bukets <- tf$feature_column$bucketized_column(
    BandWidth, boundaries = c(5,10,20,45,60,100,150,200,500,1000)
)

base_columns <- c(Protocol, Type, Label, Source, Destination, Band
Width_bukets)
```

**Figure 6.** BandWidth division and Creating the basic columns

```
crossed_columns <- c(
Protocol, Type, Label, Source, Destination, BandWidth_bukets,
tf$feature_column$crossed_column(c("Protocol", "Type"), hash_bucke
t_size = 10000),
tf$feature_column$crossed_column(c("Type", "Label"), hash_bucket_s
ize = 10000),
```

**Figure 7.** Creating influential data columns and dimensions.

```
> deep_columns <- c(
tf$feature_column$embedding_column(Protocol, dimension = 8),
tf$feature_column$embedding_column(Type, dimension = 8),
tf$feature_column$embedding_column(Label, dimension = 8),
tf$feature_column$embedding_column(Ping, dimension = 8),
tf$feature_column$embedding_column(MessagePhase, dimension = 8),
Duration,
Length,
BandWidth,
DestinationPort,
SourcePort
)
```

**Figure 8.** Creating columns with Deep Features.

```
> model <- dnn_linear_combined_classifier(
      linear_feature_columns = crossed_columns,
      dnn_feature_columns = deep_columns,
      dnn_hidden_units = c(100, 50)
  )
```

**Figure 9.** Model building.

```
> trainData$label <- ifelse(trainData$Label != "Nmap", 1, 0)
> testData$label <- ifelse(testData$Label != "Nmap", 1, 0)
> constructed_input_fn <- function(dataset) {
      input_fn(dataset, features = -label, response = label)
  }
> train_input_fn <- constructed_input_fn(trainData)
> eval_input_fn <- constructed_input_fn(testData)
> train(model, input_fn = train_input_fn, steps = 4)
Training 4/4 [==================================] - ETA: 0s -
loss: 6851
```

**Figure 10.** For machine learning and training

```
evaluate(model, input_fn = eval_input_fn, steps = 4)
Evaluating 2/2 [==================================] - ETA:  0s -
loss: 1191

Evaluation completed after 4 steps but 4 steps was specified
```

**Figure 11.** Results achieved.

To understand more about the efficiency of the proposal, we change the ratio between testing dataset and training dataset. We do the same things with case 2 (80% -20%) and case 3 (70% - 30%).

So briefly, in *case 1* (the ratio between test dataset and training dataset are 75% - 25%), the achieved results: In the test dataset, we have 6,851 loss in the total 94,725 of Log lines, reaching an error rate of 6,851/94,725. In the training dataset, we have 1,191 Log lines that do not match and exclude, the invalid rate is 1,191/31,576.

*Case 2* (the ratio between test dataset and training dataset are 80% - 20%): In the test dataset, we have 6,134 loss in the total 94,725 of Log lines, reaching an error rate of 6,134/101,041. In the training dataset, we have about 1,501 Log lines that do not match and exclude, the invalid rate is 1,501/25,260.

Case 3 ( the ratio between test dataset and training data are 70% - 30%) In the test dataset, we have about 5,652 errors in the total 94,725 of Log lines, reaching an error rate of 5,652/88,411. In the training dataset, we have about 2,410 Log lines that do not match and exclude, the invalid rate is 2,410/37,890.

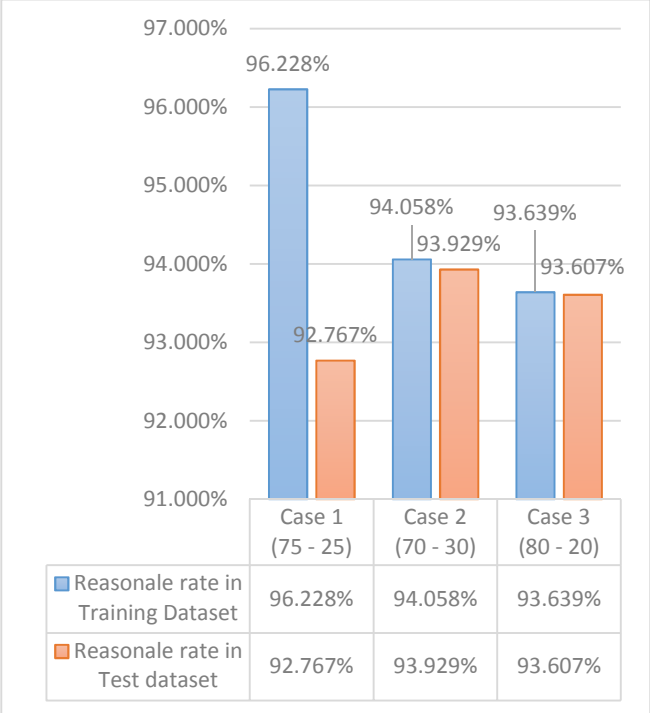


**Figure 12.** The reasonable ratio of Log lines of the Test dataset, training dataset in 3 cases.

From the results achieved of the experimental model, it is shown that the reasonable ratios of Log lines in the Test data in 3 case are very high. This shows the optimism in combining Log processing and Intelligent algorithms to build early warning models.

## V. Conclusion

The topic helps readers get an overview of the safety and security of network security, which includes basic attack techniques and especially APT attacks, in addition to Tri application. Artificial Intelligence (AI), Machine Learning - Deep & Wide Learning on log-data analysis.

However, there are still limitations in this research. Typically, it does not cover all the cases, but it is updated and supplemented to complete in the future. When it comes to practicality, the topic has come up with a solution and built an attack warning software based on Logs analysis, especially APT attack, which can help administrators realize the dangers before they happen.

Based on the results achieved and the limitations still exist, in the future we will try to study network attack models, new reconnaissance methods to better detect and warn, conduct research and develop algorithms to improve Logs collection and analysis efficiency.

## Acknowledgment

We, the authors of this paper, would like to thank VNPT Binh Duong, especially Mr. Nguyen Duc Huy working in

sales department, for supporting and assisting us in providing research, simulation environment and equipment for the implementation of this article.

## REFERENCES

[1] Beth E. Binde, Russ McRee, Terrence J. O'Connor, "Assessing Outbound Traffic to Uncover Advanced Persistent Threat", SANS Technology Institute, Joint Written Project, 5/22/2011.

[2] Nart Villeneuve and James Bennett, "Detecting APT Activity with Network Traffic Analysis", Trend Micro Incorporated Research Paper 2012.

[3] Sophos Limited, Andrew J. Thomas, "Advanced Persistent Threat Detection" ,Cross-Reference To Related US Applications Data, U.S. patent application Ser. No. 14/263,955 filed on Apr. 28, 2014, now Pat. No. 9,392,015.

[4] Sin Chun Ng and Majid Bakhtiarib, "Advanced Persistent Threat Detection Based On Network Traffic Noise Pattern and Analysis", Journal of Advanced Research in Computing and Applications, ISSN (online): 2462-1927 | Vol. 2, No. 1. Pages 1-18, 2016.

[5] Harikrishnan V N, Gireesh Kumar T, "Advanced Persistent Threat Analysis using Splunk", International Journal of Pure and Applied Mathematics, Volume 118 No. 20 2018, 3761-3768 ,ISSN: 1314-3395.

[6] Artur Rot and Boguslaw Olszewski, "Advanced Persistent Threats Attacks in Cyberspace. Threats, Vulnerabilities, Methods of Protection", Computer Science and Information Systems, pp. 113–117, DOI: 10.15439/2017F488 ISSN 2300-5963 ACSIS, Vol. 12., 2018.

[7] Marchetti, Mirco & Pierazzi, Fabio & Colajanni, Michele & Guido, Alessandro, "Analysis of high volumes of network traffic for Advanced Persistent Threat detection", Computer Networks, 109. 10.1016/j.comnet.2016.05.018

[8] Principal Author: Jason Creasey and Principal reviewer Ian Glover , "Cyber Security Monitoring and Logging Guide",Version 1, Managing Director, Jerakano Limited, Sep.2015

[9] Heng-Tze Cheng, Levent Koc, Jeremiah Harmsen, Tal Shaked, Tushar Chandra, Hrishi Aradhye, Glen Anderson, Greg Corrado, Wei Chai, Mustafa Ispir, Rohan Anil, Zakaria Haque, Lichan Hong, Vihan Jain, Xiaobing Liu, Hemal Shah," Wide & Deep Learning for Recommender Systems", https://arxiv.org/abs/1606.07792 , Submitted on 24 Jun 2016.