Mathematical Model on Wireshark Operation Skill Evaluation

By Matsuda T.* Sonoda M. Etou M. Satoh H. Hanada T.
Kanahama N. Ishikawa H.

Department of Information Security, University of Nagasaki, Japan*
National Institute of Information and Communications Technology

Abstract

Wireshark is free software used for problem solving in the information security domain. This paper focuses attention on mouse operation of a Wireshark operator, and proposes a method of estimation of operator skills using data from UWSC. UWSC is a tool for recording mouse and keyboard input and automating the input operation.

1. Introduction

The technique of cyber attack continues to progress with the progress of ICT technology. To prepare for threats of cyber security like this, it can be said that development of teaching materials for security education is important. Most of studies related to security education are of the type that explains how to use tools in moving pictures and proposes a method of exercising using a virtual environment. It is thought that such content might be easy to understand for the first scholar, but it would be difficult for beginners to find content that fits their skills. This study had analyzed how the operator with skill uses the operation screen of Wireshark (Wireshark [online]), and proposed a method to judge the presence or absence of skill from data based on the concept of a latent curve analysis.

Our proposed algorithm is characterized by introducing latent variables in the parameters of the mathematical model, therefore the Potential properties of data can be estimated. This study had generated UWSC (UWSC(online)) data by reproducing the contents uploaded to YouTube. By applying the proposed method to our generated data, it is possible to automatically classify data having similar properties. Furthermore, we had confirmed that skill evaluation of the instructor is possible by applying the skill factor of the instructor to the latent variable.

2. Wireshark and UWSC

Wireshark is the most famous free software that can capture network packets. FIG. 1 shows the operation screen of Wireshark.By connecting a PC with Wireshark installed to a network device, it is possible to capture packets flowing on the network.Generally, a large amount of data is recorded in the pcap file where data is stored.Wireshark is implemented many functions such as a filtering function for finding necessary information from such a large amount of data and a function for visualizing statistical information of data (Waqar Ali,& Jun Sang,& Hamad Naeem,& Rashid Naeem,& Ali Raza. 2015).

Wireshark is also an important tool in security related problem solving, and a large number of contents explaining how to use Wireshark is shared. For example, a lot of commentary content has been uploaded to YouTube, and useful information are provided regardless of user's skill. Although these are useful as educational contents, it is not easy to judge the difference in the skills of the instructors who explain the contents. One of the reasons is that it takes time to analyze a large amount of video data.

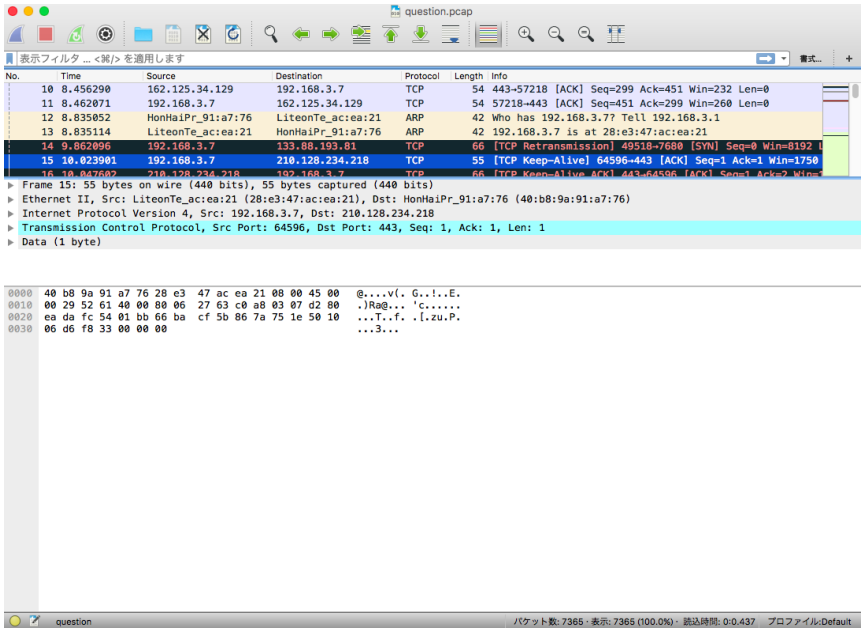

FIGURE 1. Wireshark UI

Therefore, this study investigated to estimate skills of technicians by extracting operation log of wireshark using UWSC. This study investigated the difference of mouse operation in GUI of Wireshark between operators with skill and operators without skill. In order to automatically record the state of the mouse operation, mouse operation data was acquired using UWSC as FIG. 2.

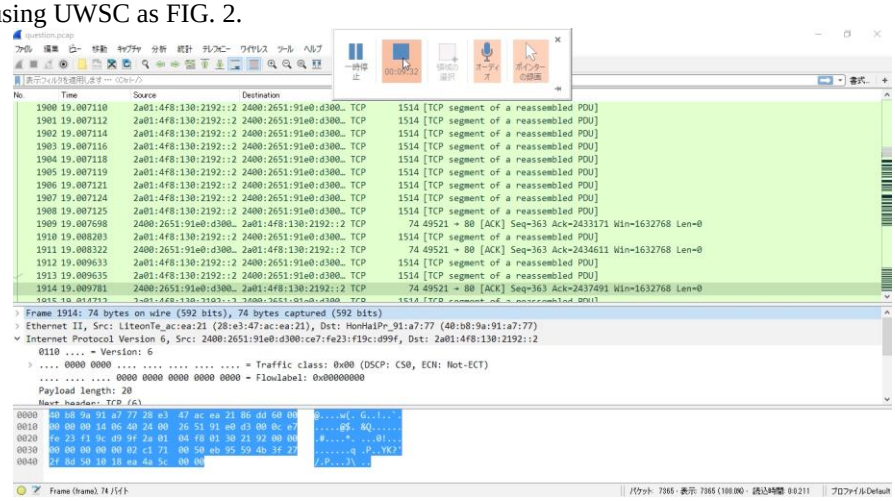

FIGURE 2. Operation log acquisition screen using UWSC

3. Proposed Method

A part of the data acquired using UWSC is shown in FIG. 3.

PROPOSED METHOD


```
CALL timer.UWS
id = GETID(GET_ACTIVE_WIN)
PRINT "^~Oc^~U^~]:" + G_SCREEN_H
PRINT "^~I¡^~U^~]:" + G_SCREEN_W
THREAD COMMONTIMER.TIMER()
MMV(1146,442,516)
MMV(1145,440,15)
MMV(1143,438,16)
MMV(1142,437,15)
MMV(1136,432,16)
MMV(1132,428,16)
MMV(1130,427,15)
MMV(1129,427,16)
MMV(1128,427,78)
MMV(1127,426,16)
MMV(1125,424,15)
MMV(1122,424,78)
MMV(1122,424,16)
MMV(1119,424,15)
MMV(1114,424,16)
MMV(1110,423,16)
MMV(1107,421,15)
MMV(1103,421,16)
MMV(1102,421,15)
MMV(1101,421,16)
MMV(1100,421,16)
MMV(1099,420,15)
MMV(1097,420,16)
MMV(1096,419,15)
MMV(1094,417,16)
```

FIGURE 3. Data of UWSC

From the value of the MMV function in FIG. 3, information on the position of the mouse on the desktop screen is obtained. The arguments of the MMV function are shown below.

MMV(x, y, [ms]),

Where x, y and [ms] indicate x coordinate, y coordinate and millisecond of mous operation, respectively.

In this study, we divide the Wireshark GUI screen into six areas, thereby acquiring and analyzing mouse operation data of operators with operation skill and operators without skills. Each area has the following roles.

- area1 : Function menu
- area2 : Display filter
- area3 : Packet list
- area4 : Packet details
- area5 : Scroll
- area6 : Blank part

In order to facilitate analysis of data, we had divided the data of UWSC into 30 seconds each, and counted the number of times the mouse passed through each area. Table 1 and 2 are the data of an operator with operation skill and without operation skill, respectively. The feature of Table 1 is that many areas 1 and 2 are not 0 frequency. On the other hand, the data of table 2 does not have the above feature, except for some data. Area 1 and 2 are the part to use the functions of Wireshark. Therefore, operators with operation skill are well used area 1 and 2. Regarding the data collected this study, the frequencies on areas 1 and 2 can be said to be important information for estimating whether or not operators have the skill. Actually, the data in Table 3 are data of operators different from Tables 1 and 2. However, there is considerable

overlap in the data of both data for areas 3, 4, 5 and 6. This study had investigated whether data including such duplication can be classified by using latent curve analysis. Since the data in Tables 1 and 2 are frequency data counted every 30 seconds, the discrimination of skills should take account of the change in time series data. We propose a model that discriminates whether skill elements are included in each data or not.

area 1	area 2	area 3	area 4	area 5	area 6	skill
2	0	2	0	0	1	expert
0	1	2	2	0	1	expert
2	1	5	0	1	3	expert
0	0	2	1	0	0	expert
0	2	5	1	0	0	expert
1	1	4	0	2	1	expert
1	0	3	1	0	1	expert
0	0	0	1	0	1	expert
0	0	1	0	0	0	expert
2	1	1	0	0	2	expert

TABLE 1. Feature of an operator with operation skill

area 1	area 2	area 3	area 4	area 5	area 6	skill
0	0	2	0	2	0	beginner
0	0	3	0	3	0	beginner
0	1	2	0	2	1	beginner
0	0	1	0	1	3	beginner
0	0	1	0	0	0	beginner

TABLE 2. Feature of an operator without operation skill

area 1	area 2	area 3	area 4	area 5	area 6	skill
4	1	3	1	0	5	expert
0	1	3	0	0	2	expert
0	0	2	3	0	1	beginner
0	0	2	3	3	0	beginner
0	0	1	1	1	0	beginner

TABLE 3. Part of Test Data

We proposed the following model to extract operator skill to extract the changes of the data in Tables 1 and 2.

$$y_x = a_2 x^2 + a_1 x + a_0, \tag{3.1}$$

where

$$a_2 = p_{21} t + p_{20},$$
$$a_1 = p_{11} t + p_{10},$$
$$a_0 = p_{01} t + p_{00}.$$

Here, we assumed that

$$y_x - \left(a_2 x^2 + a_1 x + a_0\right)$$
$$a_2 - \left(p_{21} t + p_{20}\right),$$
$$a_1 - \left(p_{11} t + p_{10}\right),$$
$$a_0 - \left(p_{01} t + p_{00}\right).$$

are obeyed from the normal distribution with mean 0, $x \hat{\mathrm{I}} \{1,2,3,4,5,6\}$ and a_2, a_1, a_0, $p_{21}, p_{20}, p_{11}, p_{10}, p_{01}, p_{00}$ are real numbers. The variable $t \hat{\mathrm{I}} \{0,1\}$ indicates the latent variable that determines skill.

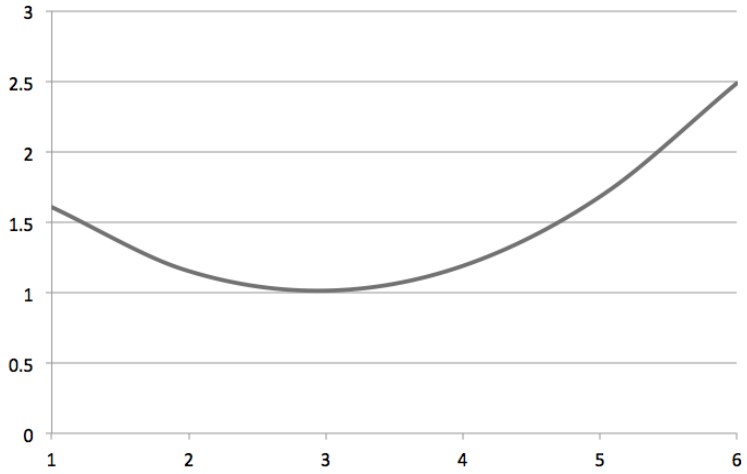

FIGURE 4. Operator with operation skill ($t = 1$)

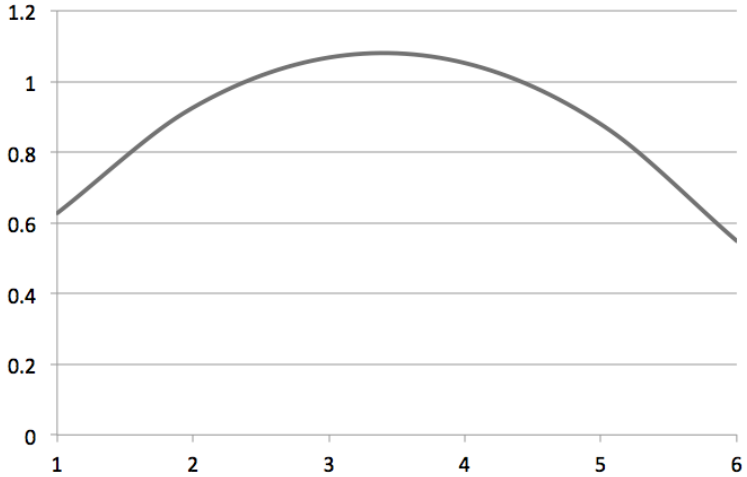

FIGURE 5. Operator with operation skill ($t = 0$)

FIG. 4 and 5 show the feature of operators with operation skill and operators without skills, respectively.
The equation in the graph of FIG. 4 is

$$y_x = 0.1576x^2 - 0.9270x + 2.3759$$

and the equation in the graph of FIG. 5 is

$$y_x = -0.0787x^2 - 0.5350x + 0.1709,$$

respectively. Curves learn by rearranging frequencies in descending order as Table 4.

area 3	area 6	area 1	area 2	area 4	area 5	skill (estimation)	skill (true)
3	5	4	1	1	0	expert	expert
3	2	0	1	0	0	expert	expert
4	1	0	0	2	1	expert	expert
3	1	1	0	1	1	expert	expert
1	0	0	0	0	0	beginner	expert
3	1	1	0	1	2	expert	expert
1	1	0	0	1	0	beginner	expert
2	0	0	2	1	0	expert	expert
1	0	0	1	0	0	beginner	expert
3	0	0	2	2	0	expert	expert
1	0	0	1	1	0	beginner	expert
3	0	0	1	0	1	expert	expert
1	0	0	1	0	0	beginner	expert
2	0	0	2	0	0	expert	expert
2	1	0	0	3	0	expert	beginner
2	0	0	0	3	3	expert	beginner
1	0	0	0	1	1	beginner	beginner
1	0	0	0	0	0	beginner	beginner
1	0	0	0	1	0	beginner	beginner
0	0	0	0	1	0	beginner	beginner
2	0	0	0	0	0	expert	beginner
2	1	0	0	1	0	expert	beginner
2	0	0	0	0	0	expert	beginner
1	1	0	0	0	0	beginner	beginner

TABLE 4. Feature of an operator with operation skill

4. Result and Summary

We prepared different test data (Table 4) from the data used for learning. Table 4 summarizes the results. Here, we define an expert as an operator with operation skill, and an operator without skills as beginner. From the results in Figs. 3 and 4, our proposed method seems to capture the trends of the data in Tables 1 (expert) and 2 (beginner). However, the estimation results in Table 4 are not very good. This result was obtained by calculating the residual of the test data and the curve, and estimated using the curve with the smaller residual. By adding frequency information on areas 1 and 2 to the model, improvement of the accuracy of the model is expected, so we would like to consider it as a future task.

REFERENCES

Wireshark(online) https://www.wireshark.org/download.html (2017.11.09).
Automated testing tool by using free software – UWSC(online)
http://ftp.forum8.co.jp/forum8lib/pdf/JaSST2008.pdf (2017.11.09)

Waqar Ali,& Jun Sang,& Hamad Naeem,& Rashid Naeem,& Ali Raza. 2015 Wireshark window authentication based packet captureing scheme to pervent DDoS related security issues in cloud network nodes 2015 6th IEEE International Conference on Software Engi-
neering and Service Science (ICSESS), 114–118.