UEC-IS-TR-2000-08

Tudumi: Log Information Visualization System for Intrusion Detection

Tetsuji Takada **Hideki Koike**

September 2000

Tudumi: Log Information Visualization System for Intrusion Detection

Tetsuji Takada
Graduate School of Information Systems
University of Electro-Communications
Chofu, 182-8585 Tokyo, Japan
zetaka@vogue.is.uec.ac.jp

Hideki Koike
Graduate School of Information Systems
University of Electro-Communications
Chofu, 182-8585 Tokyo, Japan
koike@vogue.is.uec.ac.jp

Abstract

This paper describes a novel information visualization system, which is called Tudumi, for intrusion detection.

Although much more computer users become to know the crisis of invasion to the computer, most of them have not done an action as countermeasure against an intrusion. The major reason is the difficulty of investigating log information.

In this research, we propose a log information visualization system for intrusion detection. We visualize three kinds of log information and build one figure. If the information is visualized, it makes easier for users to understand its contents ever than before. Visualizing log information, moreover, makes it possible to clear a suspicious information.

As a result, Tudumi can support log information investigation task. The task equals to intrusion detection.
keywords: *log information visualization, intrusion detection,*

1. Introduction

The numbers of intrusion to a computer are increasing more and more. Countermeasure against intrusion is urgent issue. In order to protect the computer from such intrusion, many computer sites are mainly using firewall. It, however, is clear that firewall is not enough to keep computer safe from various intrusions perfectly. Intrusion Detection System(IDS), therefore, attracts system administrator's attention currently for next generation countermeasure system. This system investigates log information, looks for suspicious activity as an intrusion and notifies it to system administrator.

The current IDS only notifies suspicious activity to the system administrator. When the system administrator has received a notification, he/she has to investigate log information manually in order to judge whether it is malicious intrusion or not.

In addition, the same task has to be done when it convinced that the intrusion has really occurred. It is, however, almost impossible for a system administrator to investigate log information and judge whether intrusion occurred or not, because log information has some characteristics that are not desirable in investigation. These characteristics are taken up in the next chapter.

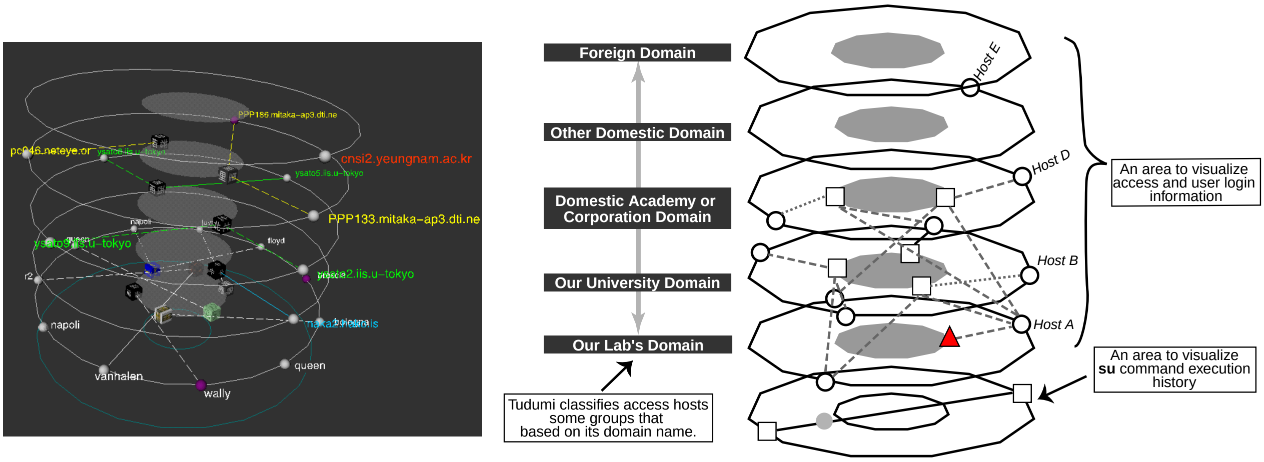
In this research, we propose a log information visualization system, which is called "Tudumi", for intrusion detection. Tudumi visualizes log information as a figure. It is, therefore, easy to understand these contents. It reduces information recognition load. It, moreover, makes suspicious activities clear than represented them by text.

These features make it possible to reduce technical skills to be necessary for intrusion detection. We, therefore, expect that many users can engage in intrusion detection task.

In this paper, we describe the problem about log information investigation in intrusion detection in chapter 2, and explain how to visualize some log information in chapter 3. We illustrate an example of intrusion detection in chapter 4 and describe related works and future works in chapter 5.

2. Background: A Problem in Intrusion Detection

A lot of products for intrusion detection already exist. There is necessity of investigating log information even if intrusion detection system has widely spread. One of the reasons is that the purpose of intrusion de-



cnsi2.yeungnam.ac.kr
PPP133.mitaka-ap3.dti.ne
napoli
vanhalen
wally
queen
Foreign Domain
Other Domestic Domain
Domestic Academy or Corporation Domain
Our University Domain
Our Lab's Domain
Tudumi classifies access hosts some groups that based on its domain name.
Host E
Host D
Host B
Host A
An area to visualize access and user login information
An area to visualize su command execution history

Figure 1. An example and visualization framework of Tudumi

tection system is only to notify the suspicious activities. Many IDS are impossible to help log information investigation work. The system administrator, however, has to perform this task whenever he accepted the notification from IDS or convinced the occurrence of malicious intrusion. Thus, system manager investigates log information manually at present.

There are some problems in investigating log information.

First, it is a hard work even if a system administrator performs it. Since log information is recorded as textual information, he/she has to read log text so as to understand all of them. In other words, log information has high information recognition load.

Secondly, this task needs much more time to do it, because the enormous amount of log information. This work is tedious, but it requires time and worker's attention. It is extremely difficult to perform it correctly even if the worker is a skillful administrator. As a result, many system administrators tend to neglect this task.

Last problem is that a system administrator has to investigate more than one log information in many cases. In general, if you have only single log information, you can not judge whether intrusion is occurred or not. Because it is impossible for system administrator to convince the occurrence of malicious intrusion without various information. In order to judge a malicious intrusion precisely, many kinds of log information are needed.

We, therefore, visualize the log information to improve these problems. We consider that Tudumi provides some advantages listed below.

- To reduce the recognition load of log information
- To clarify the suspicious activities as intrusion
- To reduce the investigation load and time by integration of various log information into figure

3. Tudumi: Its Visualization Method

A lot of products for intrusion detection have already exist. Many kinds of information sources exist for intrusion detection. We visualized log information that mainly focuses on access from remote computer in network. In addition, we integrated user's login status and substitution status between users into it. We used **tcp_wrapper** [11] to get the access log information. And we also used wtmpx log file and "sulog" log file respectively to get user's login status and the substitution status.

We explain the framework of visualization about Tudumi in this section. Figure 1 shows an actual visualization example and its explanation figure.

An overview of Tudumi is a stratified concentric disk. Tudumi classifies access host information into several groups according to the rule that based on its domain name and displays them on each layer. A system administrator has to define the classification rule. We recommend that you should define the rule based on the probability of access from each host and the rule that has high access possibility is assigned to higher layer.

Then, you can extract suspicious accesses from enormous access log information. And these information is visualized on higher layer.

For example, the rule that is used in our laboratory is as follows. The accesses from the computer in our laboratory are assigned to second lowest layer. And the accesses from the computer with foreign country's domain are assigned to highest layer.

The reason is that no one will access from the computer with foreign country. Using both visualization and access host classification rule, a system administrator can recognize suspicious accesses more easily (see figure 1).

Moreover, login users, access hosts and user substitution status are visualized on the disk in each layer. Figure 2 illustrates an explanation of its visualization method.

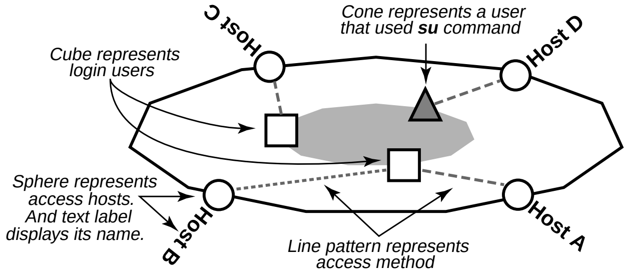

Host C
Cone represents a user that used **su** command
Host D
Cube represents login users
Sphere represents access hosts. And text label displays its name.
Host B
Line pattern represents access method
Host A

Figure 2. Visual method of access hosts and login users on each layer

Access host information is visualized as a sphere on the outer circle of the concentric disk. And login user is visualized as a cube with texture on the inner circle. The relation of them is represented by a line and the line pattern represents the access method. There are three kinds of line. The access for terminal use is represented as coarse dashed line and the ftp use is represented as fine dashed line. The solid line represents access for terminal and ftp use at a time (see figure 3).

It is possible for a system administrator to assign pictures to each user cube. The purpose of it is to recognize each user information more easily than textual display. Tudumi, moreover, changes visual method of login user according to the distance from viewpoint.

If distance between login user information and viewpoint is long, Tudumi represents login user as textured cube. It represents login user as 3D letters if the distance is short (see figure 4). Therefore, it is easy to get actual user account name. This feature enables the viewer to avoid visual scattering arise from displaying many a textual information.

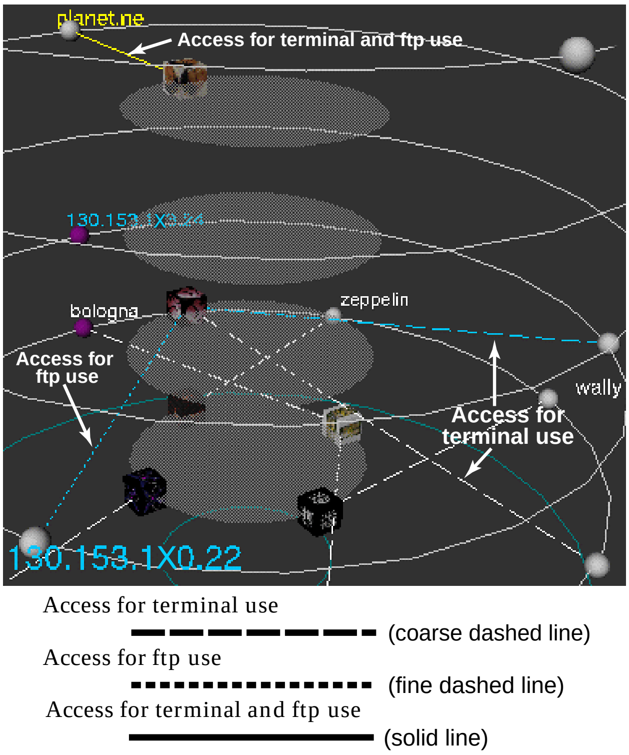


Figure 3. Line pattern represents access method

The layer where each login user is displayed is the highest layer in user used access hosts. In other words, all access hosts are displayed in the same or lower layers that displayed each user cube. This feature enables the viewer to understand an overview of access hosts on each user.

Furthermore, there is the case that login user is represented as cone instead of cube. This means that user represented as cone is trying to substitute other users (see figure 5). A system administrator can recognize a suspicious user activity. The detail information about user substitution is displayed in lowest layer.

We explain how to visualize user substitution status information. The user who is trying to substitute other user is displayed as textured cube in outer circle. Substituted user is also represented the same way. The relation of user substitution is displayed by an arrow. This arrow consists of line and sphere. The sphere represents the arrow's head. Furthermore, a color of sphere represents the probability of user substitution. If user substitution succeeds, the color of sphere will become red. If it fails, the color of it will become blue. If user substitution does more than one time, the col-

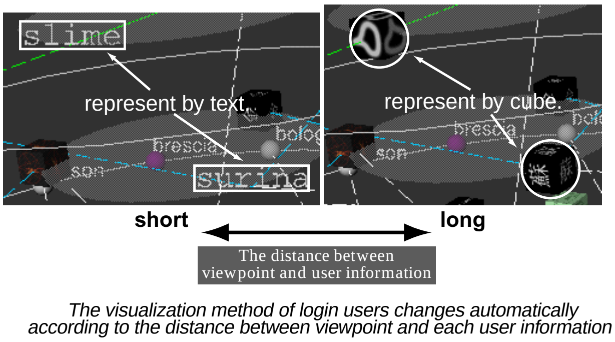

slime
represent by text.
surina
represent by cube.
short
long
The distance between viewpoint and user information
The visualization method of login users changes automatically according to the distance between viewpoint and each user information

Figure 4. Visualization method of login users

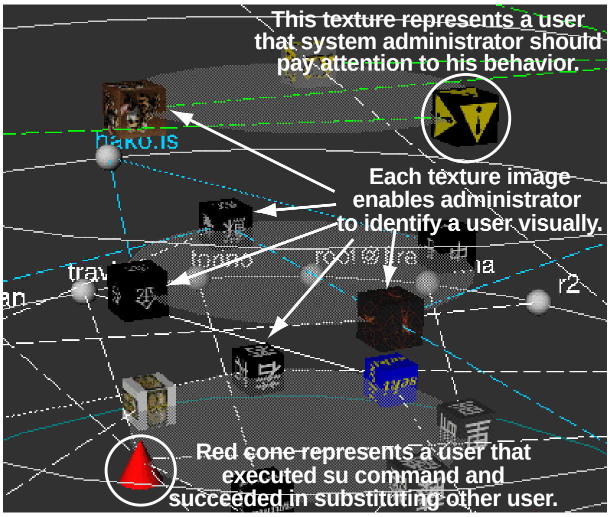

This texture represents a user that system administrator should pay attention to his behavior.
Each texture image enables administrator to identify a user visually.
Red cone represents a user that executed su command and succeeded in substituting other user.

Figure 5. Variation of representation method about login user

or of sphere will be mixed corresponding to the ratio of user substitution result. This visualization makes it easier to understand user relation and result status of user substitution (see figure 6).

3.1. Why does Tudumi have Such Framework?

In this section, we describe why we visualize log information with such stratified shape.

It is necessary for Tudumi to display a large quantity of information in a constant screen area, because most of the log information has enormous quantity. We visualized these information with arranging to concentric disk. This visualization method enables the viewer to display so enormous information into a fixed area.

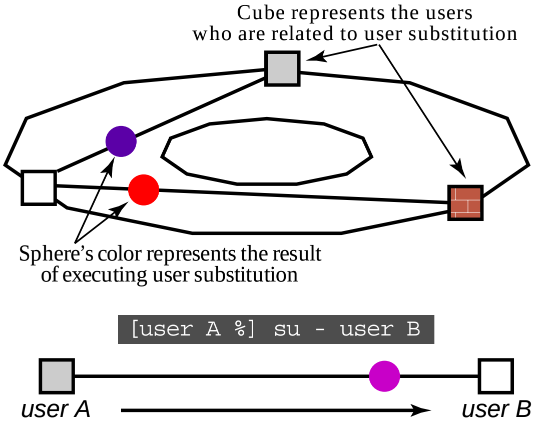

Cube represents the users who are related to user substitution
Sphere's color represents the result of executing user substitution
[user A %] su - user B
user A
user B

Figure 6. Visual method of user substitution status

Tudumi also needs to make it clear whether the suspicious activities exist or not. We built the shape that is concentric disk with stratified formation and assign access hosts classification rules to each layer. This feature enables the viewer to classify information according to the rules defined by users. It is possible to understand suspicious access more easily. And it also possible to reduce the information recognition load. The reason is that information is displayed in a fixed space based on a certain rule.

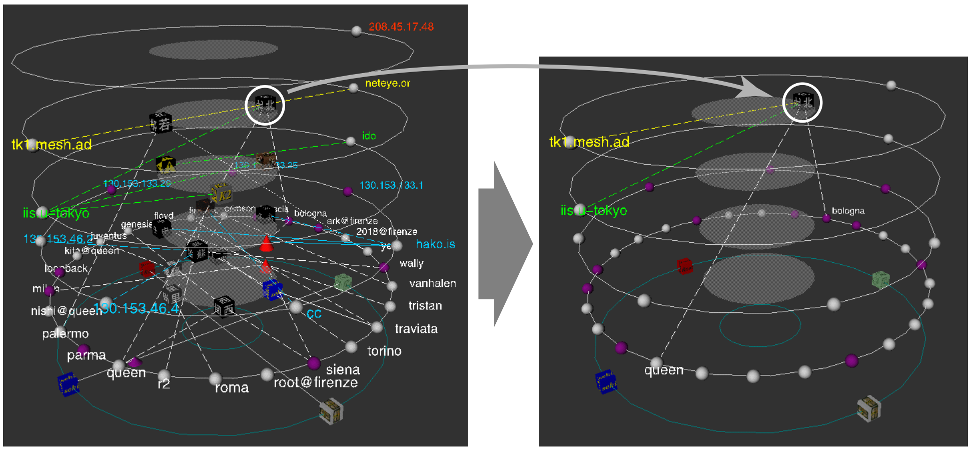
Furthermore, Tudumi displays access hosts and login user information with arranging to concentric disk. This makes it easier to grasp the overview of the relation between them. The reason is that all lines that represent the relation of them are drawn between inner and outer circle. Tudumi has the advantage of automatic browsing function by rotating whole scene because its shape is like a cylinder.

3.2. Investigation Support by Interactive Operation

Tudumi has some interactive operations and can help investigation work of log information. The purpose of these functions is to manage the displayed information and to clarify the specific information.

Extraction of specific information by an object selection

If a system administrator selects the information that he interests in, Tudumi will only display the selected



If the viewer select user cube by mouse,
Tudumi displays specified user and related information only.

Figure 7. Example of extracting focus and its related information by user's interaction

and related to its information (see figure 7).

When a system administrator discovers attractive information, he can filter out unnecessary information by selecting it by mouse. This function makes not only focused information but also related information clear. This feature added to Tudumi the advantage to reduce the investigation mistake by oversight.

Display Control of Access Hosts by Summarizing Domain Name

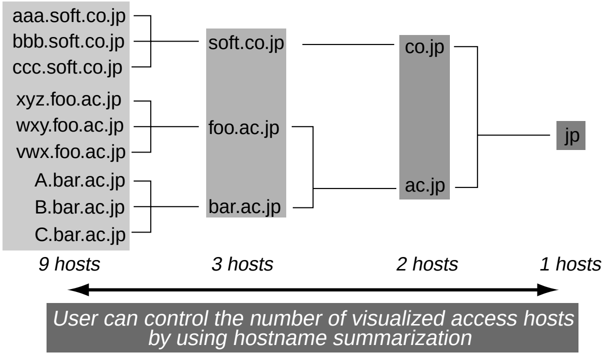


Figure 8. Host name summarization method using domain name for controlling the amount of visualized access hosts information

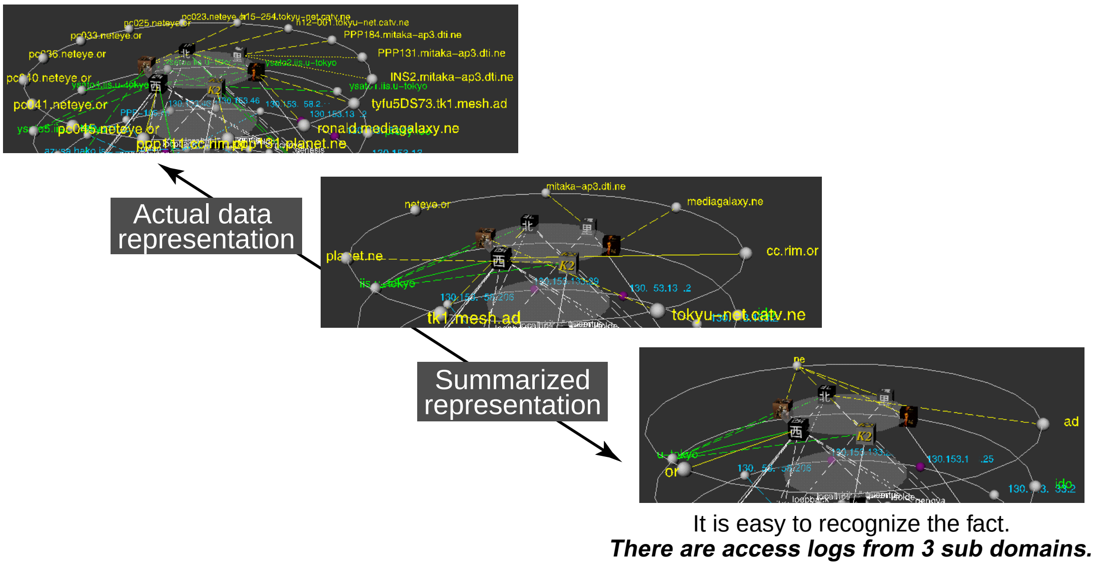
It is assumes that access host information becomes enormous quantity. We propose abstraction technique of access hosts information using domain name. This abstraction technique uses the hierarchical structure of domain name system and enables the viewer to control the amount of displaying access hosts information. As a result, a system administrator can investigate a lot of information more effectively.

This method summarizes the access host information by shortening the host name. It is clear that more than one host name can be summarized to one domain name (See figure 8). As a result, a system administrator can control visualized access host information. And the recognition load for access host information is reduced. Moreover, grasping overview of access hosts becomes more easily, because access host information is summarized.

Figure 9 is an example of display using access host summarization. You watch the highest layer in each figures. You can understand that number of displayed access host is decreased as 14,4,3,2 from the top to bottom respectively. Moreover, you can also know that it was accessed from “dti.ne”, “planet.ne” and “neteye.or” in JP domain.

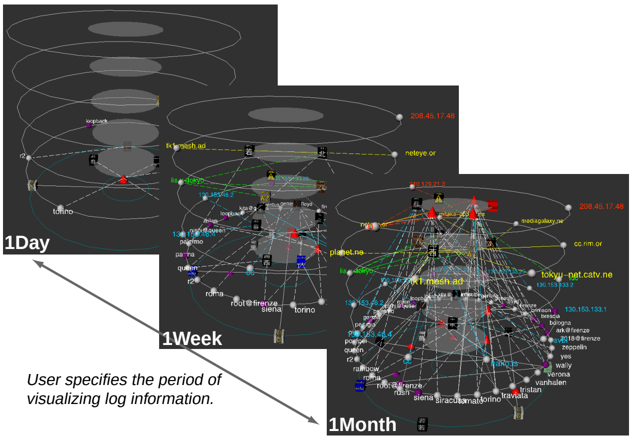
Display Information Control by Specifying Data Span

Tudumi visualizes log information that is generated in a past fixed period from current time. With changing that period, a system administrator can investigate only recent log information. They can also investigate them in past long-term period (see figure 10).



It is easy to recognize the fact.
There are access logs from 3 sub domains.

Figure 9. An example of controlling the amount of access hosts



User specifies the period of visualizing log information.

Figure 10. An example of changing the period of visualizing data

Display Information Control by Displaying a part of the Stratum

Tudumi deteriorates the readability of information in middle and lower layers, because increasing displayed information interfere with the information recognition in some case. We can improve this problem by visualizing log information partially. With this function, Tudumi visualizes only lower layer than user specified. It improves the readability of log information in the lower layers (see figure 11).

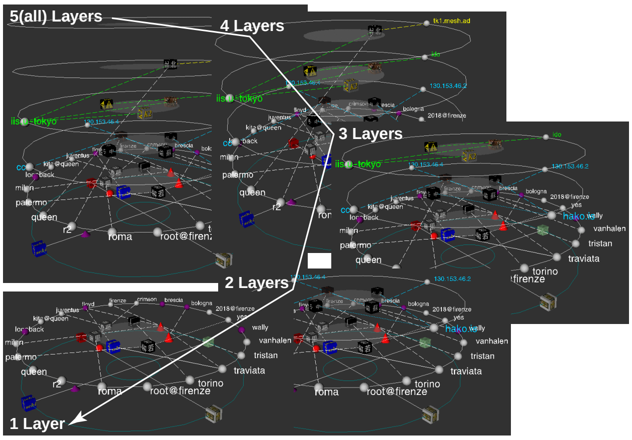


Figure 11. An example of visualize only one or some layers

3.3. The Ability to Apply Large-Scale Information

We refer to the applicability of Tudumi to large-scale log information. Figure 12 illustrates large-scale visualization example. It is displayed 120 access hosts. Even if the target log information is enormous quantity, Tudumi can display it into a fixed area. A main problem of the increasing target log information is the growth of access host information. Tudumi, however, can manage it with some interactive functions such as classifying with layers, summarizing with domain name and enable to change the period of visualizing log information. And a system administrator can translate scatter display that much information is visualized into simple display that is easily recognized by using above methods. As a result, we consider that Tudumi has the ability to operate large-scale information.

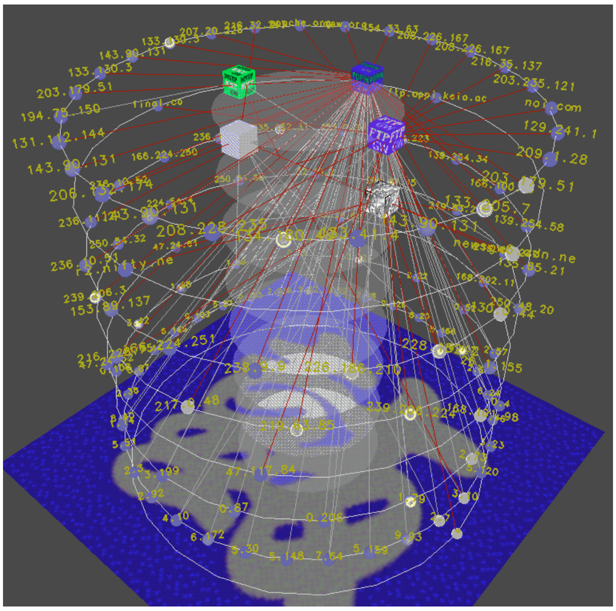


Figure 12. An example of Large quantity of log information using Tudumi

4. Intrusion Detection Examples with Information Visualization

We illustrate that Tudumi can detect suspicious activities as intrusion by watching log information.

First of all, it is clear that Tudumi has the ability to detect intrusive access through network. As mentioned above, the reason is that Tudumi extracts such accesses from log information by using access hosts classification rule that is defined by the user and visualizes them.

It, furthermore, can also recognize the activities such as that someone accesses to monitored computer but they could not login it. This case is displayed a sphere which is no connected lines.

Figure 13 illustrates just such kind of the case. There is no line that connected to the below-left sphere. This case and the kind of the example mentioned before lead to a conclusion that the sphere which is drawn on higher layer without connected line must represent intrusive access.

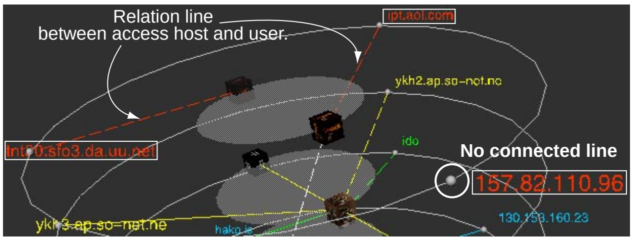


Figure 13. An example of visualized intrusive access

The next example is a suspicious user activity. We previously mentioned that Tudumi has the ability to visualize user substitution status. Most of the intruders try to get the administrator's authority. They, namely, try to masquerade other users. If, therefore, there is login user visualized as a cone, you need to investigate various log information about such user's activity in detail.

Moreover, the user is able to assign image or color to each user's information freely. This feature makes it possible to classify user information visually. As a result, extracting suspicious user activities have become easier.

We introduce the assignment of image and color for each user in our laboratory. We assign red color to administrative users. We also assign image represents warning to some specified users. One example of such users has no probability to use the computer login. These visual representations let a system administrator notice some anomalous activities. The example of such user is the student who had already graduated in university.

These examples indicate that Tudumi is able to extract suspicious user activities and recognize them visually (see figure 5).

As last, We give the example about accessing from hosts that have more than one domain. The system administrator should doubt that access records from the computer with multiple domains in a short term are as intrusion. If the viewer has some knowledge about each user, (e.g., there is possibility of accessing specified domain in a user.), you can estimate them as intrusion or not more accurately. We give an example of such rule. *No one can accesses monitored host from the computer with foreign country's domain although*

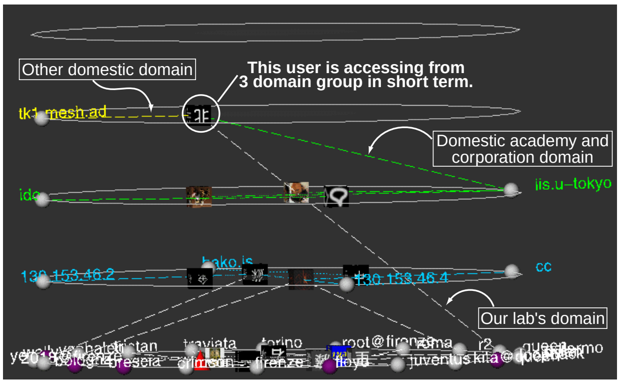


Figure 14. An example of accessing from multiple domains in short term

he/she was in our laboratory until a few minutes before.

Figure 14 is the view of Tudumi from the side. There are lines that are drawn across the layers in this figure. These lines represent that a user used the computer with accessing from multiple domains in a fixed period.

We explain what suspicious activity is illustrated in figure 14. This figure displays log information that has recorded in past 24 hours. There is a cube that illustrates login user on second top layer. We can know that he/she accesses from three domain groups, namely our laboratory's domain, domestic university domain and domestic internet service provider domain from this figure.

A system administrator must judge such activities as intrusive activity and got more detailed information from him about it. As a result, these activities were not malicious intrusion in this case, because he/she said that these activities had done by himself. There is, however, possibility that such figure illustrates intrusion. Thus, Tudumi can display such suspicious activities as figure and makes it easier for the system administrator to recognize them.

5. Related Works and Discussion

We consider a difference with former study, related work and future work in this section.

5.1. Consideration of Information Visualization System and Related Work

Visual Audit Browser[1] is an information visualization system with the aim of helping log information investigation task. This system helps to analyze audit log information. It, however, can only visualize a part of them. It, therefore, is impossible to recognize the overview of them. And a system administrator needs to know which log information has to be examined. Furthermore, audit log information has not been recorded generally in most of computer system. We consider that this system has less generality than Tudumi.

We also mention SeeLog[2] as log information visualization system. This system represents log information as colored line in order to discover anomalous information. This system has much generality, but we consider that the visual support of log investigation is insufficient if its purpose is intrusion detection.

On the other hand, another log information visualization system[3] exists. This system visualizes system accounting information. It can be used as intrusion detection system but it is difficult to do it only by itself. The reason is that system accounting information is useful for intrusion detection, but it is impossible to detect intrusion with this information only.

5.2. Comparison with Current IDS

The current IDS extracts suspicious log information as an intrusion and notifies it to a system administrator. System administrator, however, has to investigate log information manually in order to judge it as intrusion or not.

Tudumi does not notify suspicious information to system administrator actively. We could say that Tudumi has the ability of passive notification as picture. It, therefore, is possible to help investigation of log information and makes such work easier although it has no function to notify intrusive activities actively.

Tudumi only functions these abilities whenever the system administrator starts running and looks them carefully. We, however, consider that Tudumi is quite easy to use. It is difficult to introduce / operate / manage in most current IDS even if there is a skillful system administrator. These are the reason why IDS is not used widely. Tudumi can improve these problems. We consider this system has ability of making it possible to investigate log information by not only the system administrator but also other users.

5.3. Future Work

First, we will try to visualize various log information and integrate these visualization tools into one intrusion detection system. There is other useful log information for intrusion detection. We will visualize them and integrate them into Tudumi. Tudumi will become more effective tools for intrusion detection.

In addition, the current intrusion detection systems have various methods to extract suspicious activities from log information. We give the example of such methods as expert system, machine learning and data mining. We consider it useful that we integrate these extraction techniques into Tudumi as back-end process and reflect these result into visualization.

6. Conclusions

Although the crisis of malicious invasion to computer increases more and more, IDS has not been widely used yet. In this paper, we proposed a log information visualization system for intrusion detection which is called "Tudumi".

Tudumi mainly visualizes the access log information and integrates three kinds of log information into one visual image. Tudumi can reduce the recognition load of log information, because Tudumi displays them as figure. Furthermore, the visual representation in Tudumi and its interactive display control function make suspicious activities clear than investigating log information in text.

The stratified disk representation with access hosts classification rule clarifies suspicious access hosts information. Displaying login user information as textured or colored cube makes unexpected user activities clear. The user can control a quantity of visualized information interactively. These features reduce the load of investigating log information and make it easier to detect intrusion.

IDS is now highly demanded. There is, however, no IDS that can use easily. From this point of view, we consider that Tudumi will become new IDS that can be easily used for a system administrator. If the user can take part in the intrusion detection work, we expect that many users will perform the task by themselves on each computer. As a result, we can build the robust monitoring framework against malicious intrusion.

We will continue to visualize various log information and integrate them.

References

[1] J.A. Hoagland: Audit Log Analysis Using the Visual Audit Browser Toolkit, Computer Science Department U.C.Davis, Technical Report (CSE-95-11), (1995).

[2] S.G. Eick and M.C. Nelson and J.D Schmidt: Graphical Analysis of Computer Log Files, COMMUNICATION OF THE ACM, Vol.37, No.12, pp.50-56, (1994).

[3] S.G. Eick and P.J. Lucas: Displaying trace files, Software Practice and Experience, Vol.26, No.4, pp.399-409, (1996).

[4] Becker, R.A. and Eick, S.G. and Wilks, A.R.: Visualizing Network Data, *IEEE Trans. Visualization and Computer Graphics*, Vol. 1, No. 1, pp.16-28, (1995).

[5] Tamara Munzner, H3: Laying Out Large Directed Graphs in 3D Hyperbolic Space, IEEE Symposium on InfoVis'97, pp 2-10, (1997).

[6] Kenneth C. Cox, Stephen G. Eick, Graham J. Wills and Ronald J. Brachman, Visual Data Mining: Recognizing Telephone Calling Fraud, Journal of Data Mining and Knowledge Discovery, Vol.1, No.2, pp.225-231, (1997).

[7] Debra, A. and Thane, F. and Alfonso, V.: Next-generation Intrusion Detection Expert System(NIDES) A summary, *SRI-CSL-95-07*, SRI International, Menlo Park CA, May (1995).

[8] Illgun K., Kemmerer P. A. and Porras PH. A., State Transition Analysis: A Rule-Based Intrusion Detection Approach, IEEE Transactions on Software Engineering, Vol.21, No.3, pp.181–199, Mar, (1995).

[9] Giovanni, V. and Richard, A.K.: NetSTAT: A Network-based Intrusion Detection Approach, *14th Annual Computer Security Applications Conference*, pp.25-34, (1998).

[10] PaxsonV., Bro: A System for Detecting Network Intruders in Real-Time, Proceedings of the 7th USENIX Security Symposium, Jan, (1998).

[11] Wietse Zweitze Venema, TCP Wrapper, ftp://ftp.porcupine.org/pub/security/index.html, (1992).