Tudumi: Information Visualization System for Monitoring and Auditing Computer Logs

Tetsuji Takada
Satellite Venture Business Lab.
University of Electro-Communications
zetaka@computer.org

Hideki Koike
Graduate School of Information Systems
University of Electro-Communications
koike@acm.org

Abstract

Computer security breaches are already a major problem in using computers. The most basic defense against it is to monitor and audit the computer logs. Computer logs, however, have a huge amount of textual data. It is, therefore, almost impossible to inspect them manually using current systems. We propose a log visualization system called "Tudumi". Tudumi consists of several functions which assist system administrators to perform such tasks manually. These functions are information visualization, log summarization and reflecting known rules into the visualization method. Tudumi makes it easier to detect anomalous user activities, such as intrusion, from a huge amount of computer logs.

1. Introduction

Today, computer security is already a major problem. To keep your computer secure, it is an indispensable task for system administrators to monitor and audit computer logs periodically.

There are, however, inherent problems in performing such tasks by humans. First problem is that each log message is recorded as text. Administrators, therefore, must read each message to understand its meaning. This undesirable characteristic makes it a time-consuming task. It is, moreover, almost impossible to detect really intrusive behavior from just one kind of log, because each log message provides only fragmented information of user behavior. In order to judge whether a log message is really an intrusive record or not, administrators must obtain other related information from another kinds of computer logs.

Next problem is the large size of computer log. This feature makes it infeasible to monitor and audit computer logs by humans. This is the main reason why administrators avoid performing this task, even if they know the importance of doing it.

The last problem is that it is difficult to apply rules to log monitoring and auditing practically. Many administrators generally have some monitoring rules for categorizing each log message. These are, for example, based on network structure, site policy and prohibited user activities. These rules are used to judge whether each message is a normal behavior or not in a monitored computer. There is, however, no method to apply the rules to such tasks flexibly. Moreover, many existing security systems have only the ability to divide each log message into two categories, namely, a problem or not a problem. It is impossible to divide them into additional categories such as "suspicious".

In this research, we propose a log information visualization system named "Tudumi". The purpose of Tudumi is to help administrators to monitor and audit computer logs. We also propose some solutions to address the above problems.

We use an information visualization technique to address the problem that log messages are recorded as text. Using information visualization, computer log information is visualized as a graphical image. This visual translation changes the method of recognition of computer log data by humans. It namely changes the method from reading them to looking at them. and therefore reduces the recognition load. This technique, moreover, can build a meaningful model from fragmented information in the logs. In other words, it can represent the relationships between fragmented information.

We use an information summarization technique in order to improve the problems caused from a huge amount of log messages. This technique deletes repeated messages from the log in a specific time period. Summarized log messages are, namely, unique events recorded in logs. The most important point in monitoring and auditing a computer log is to recognize various kinds of occurred events in the monitored computer. Administrators can not make any decisions without recognizing them. To assist them in recognizing all events in the log more quickly, we summarize log messages before visualizing them.

The last problem mentioned above is a difficulty to make good use of the known rules for categorizing log messages. To address this problem, we change the visual representation depending on the rules. As a result, system administrators can easily decide whether each log message is a problem or not using their own rules. Tudumi, moreover, enables the ability to apply other rules to the log repeatedly and add a new rule. This feature is important in auditing the log in order to extract intrusive behaviors in various points of view.

Tudumi does not filter out any log information. This means that Tudumi represents all kinds of events in the log at anytime. Our system just changes the visual representation of them depending on the applied rules. Tudumi, therefore, can classify log messages into more than two categories based on the degree of suspicion such as "not a problem if it is a specific user's behavior". This feature helps administrators to analyze computer logs, because they can easily start detailed log inspection from the most suspicious category. Administrators can, of course, look at all kinds of event visually, even if some events belong to the "not a problem" category.

In this paper, we will describe system modules and the details of the visualization method of Tudumi in section 2, and show some examples of detecting suspicious activities using Tudumi in section 3. In section 4, we describe about related works and future works.

2. Tudumi: Overview and Its Visualization Method

We developed a visual log monitoring and auditing system called "Tudumi". Tudumi is built with the concepts mentioned in the previous section. The main goal of Tudumi is to monitor and audit user behaviors on a server computer used by a small group. Tudumi focuses on the following three kinds of user activities: accessing the server from other computers, logging in to the server and substituting a user to another user. There are two reasons why we focus on these activities.

One is that these activities are the most basic processes in using a server through the network. Even if you were a malicious user, you would always have to perform these behaviors more or less, and user substitution is frequently used in a sequence of intrusive attempts. The other is that administrators would inspect these closely related user behaviors at a time. Log-files for each behavior, however, exists separately. Administrators must inspect them respectively and make relations between the logs to determine whether such messages really show intrusive behavior or not.

We, therefore, visualized the log information resulting from these user activities and integrate their information into one visual image. Tudumi also has interactive functions that enable administrators to inspect the log through the visualized figure directly. Our system becomes a new type of user-interface for monitoring and auditing computer logs.

In this section, we describe an overview and the visualization method of Tudumi.

2.1. System Overview of Tudumi

Tudumi is a server-client model (figure 1). Thus, it is possible to monitor and audit more than one server from a single Tudumi client.

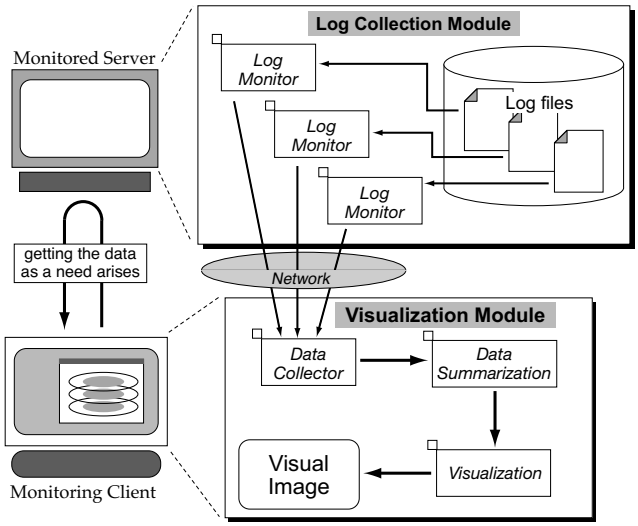


Figure 1. System Overview of Tudumi

The log collection module runs on the server to monitoring each log-file. The monitored log-files are as follows: "**Syslog**" log-file is recorded the messages about a network access from other computers. We used "TCP_Wrapper" to generate this type of messages. "**Wtmp**" log-file is recorded the messages about the time when an user logs in to and logs out from the server. "**Sulog**" log-file is recorded the messages about user substitutions.

The client has mainly two modules. One is a log collection module for collecting and summarizing them. The other is a visualization module for visualizing them and enabling the viewer to interact with the visualized image.

2.2. Visualization Method of Tudumi

we explain how Tudumi visualizes the log information. Figure 2 shows a visualization example of Tudumi.

You can see layered concentric disks. These concentric disks are divided into two groups: the bottom disk and the others. Tudumi represents user substitution information in

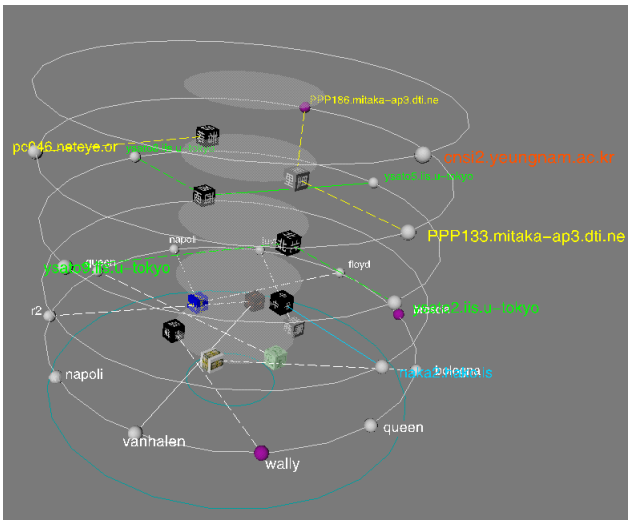


Figure 2. A Visualization Example of Tudumi

the bottom disk. The rest of the upper disks represents network access and user log-in information. We will next explain these two kinds of visualization methods.

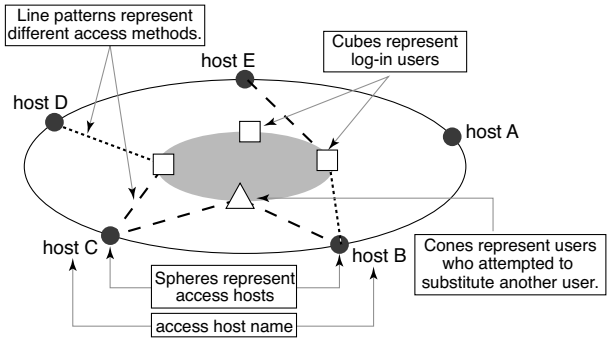


Figure 3. Visualization of Network Access and Log-in User Information

First, we explain how Tudumi visualizes network access and log-in user information into concentric disks. Figure 3 illustrates an explanation of the visualization method. Tudumi displays access hosts, log-in users, the relations of these and attempts of user substitution on the disks.

Access hosts that a user logs in from are represented as a sphere on the outer circle of the concentric disk, and log-in users are represented as cubes with texture image on the inner circle. Some users, on the other hand, are represented as a red cone. This illustrates that the user has attempted to substitute as another user. The lines are drawn between the sphere and the textured cube. These lines represents the relation between the access hosts and the log-in users. The line patterns represent the access method. The coarse dashed line represents a terminal use. It means that a user intends to log in the server and execute some commands. The fine dashed line represents a file transfer use. And the solid line represents both cases in a certain period of time.

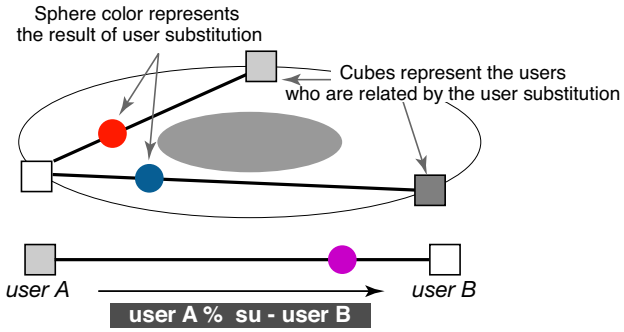


Figure 4. Visualization of User Substitution

Next, we explain how Tudumi visualizes user substitution information into the concentric disk. Figure 4 illustrates an explanation of the visualization method.

We consider user substitutions as a transition between users. Tudumi represents this situation as two users connected by an arrow. Two users related by a user substitution are represented as a textured cube on the outer circle of the concentric disk in the bottom layer. These cubes are the same with the log-in users, and Tudumi connects such users with an arrow. The head of the arrow is represented as a colored sphere. The color of it indicates the success rate of user substitution between the two users. When user substitution has succeeded, the color is red and when it has failed, the color is blue. If user substitution occurred more than one time in a visualized time period, mixed color in proportion to the rate is assigned to the cube.

2.3. Visualization of Known Rules

Administrators generally have certain criteria to judge whether each event is a normal behavior or not. An example rule is that “A user should access the server from remote computers ’*A*’ and ’*B*’ only”. Tudumi visualizes such rules as a figure.

Tudumi can currently reflect two kinds of rules into a visual representation. One is the rule about access hosts. The other is the rule about log-in users. Tudumi represents the rules about access hosts as layers. Administrators can define the rules for dividing access host information into several groups using site policies and their own knowledge. When you apply these rules to Tudumi, Tudumi visualizes access host information that belong to each rules into one concentric disk and builds the layers.

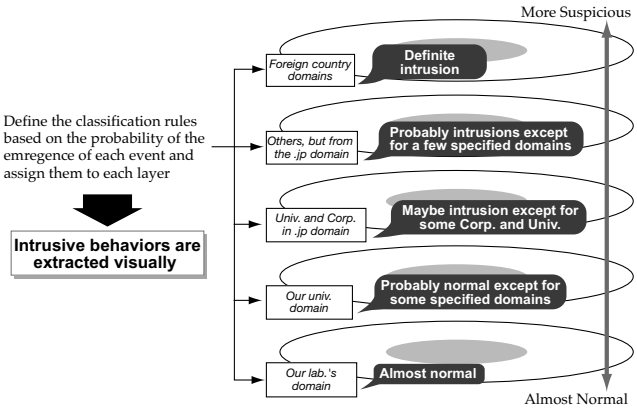


Figure 5. Layers represent the rule regarding access hosts

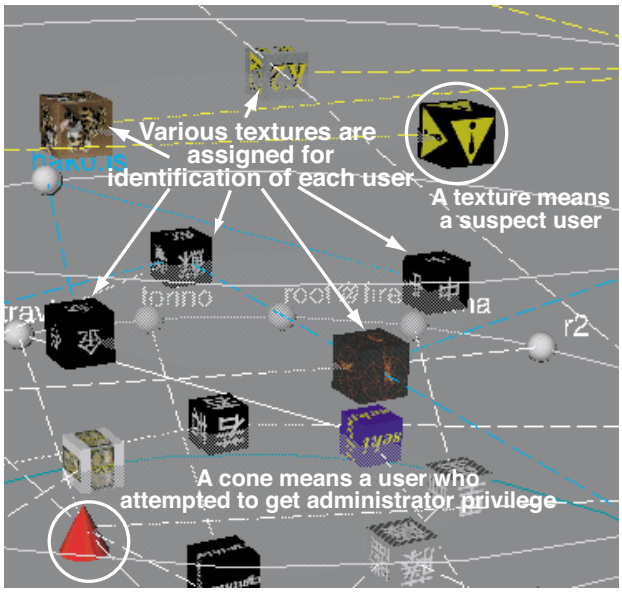


Figure 6. Various representation of Log-in Users

We recommend that you should define the rules based on the probability of access from each host or domain, and that low possibility accesses, representing an intrusive access, should be assigned to a higher layer. Figure 5, for example, shows the rules that are used in our laboratory. Accesses made from our laboratory's computer are displayed in the second lowest layer. On the other hand, accesses made from a computer with a foreign country's domain and unresolved domain name are displayed in the highest layer. The reason is that we know that none of our laboratory's members access the server from a computer that has a foreign country's domain. Thus, Tudumi enables the extraction of intrusive access from the log visually. Administrators, therefore, can recognize them more easily and quickly.

Another visualization of the rules is to assign texture images to each cube. Tudumi visualizes a log-in user as a cube with a texture image. Administrators can assign an arbitrary image to them. If you assign these texture images meaningfully when you define the rules, it is possible to represent useful information other than just the user identity.

For example, we classify the users in our laboratory into three groups. These user groups are alumni, administrators and others. We define the rules that the alumni and the administrators are each assigned to a specific image which are easily recognized at a glance. The reason is that it is practically impossible for alumni to log in to the laboratory's server. This is really a rare case in our laboratory. We, therefore, define the rules that the cubes that represent alumni and administrators are assigned to a special image that must call a viewer's attention. On the other hand, the other users are assigned normal images that a viewer can identify who each user is visually (figure 6).

These functions are very simple. It is, however, widely applicable to various environments. And it also makes recognizing already known anomalous user activities more easily than before.

2.4. Control the Amount of Visualized Data: Summarization and Interaction

The greatest problem in monitoring and auditing a log is the large amount of data. Even if we use information visualization techniques, Tudumi can only display a limited amount of information in a fixed area to keep them within visibility. We, therefore, introduce two more functions to Tudumi. One is information summarization. The other is interactive operations.

Why does the amount of log data become unmanageably large? We think the primary factor is that many log messages are generated from normal operations or events. We, therefore, introduce the log summarization in order to eliminate the excess messages (figure 7).

Tudumi, moreover, has another log summarization method. It is a summarization of access host information based on domain name. This method makes use of the concept of a domain name tree as shown in figure 8. This method represents some access hosts as one domain-name by cutting the lower part, namely the branch, of the domain-name. This feature suppresses an increase in the amount of visualized access hosts.

The information about the time and the number of log messages are lost in the summarization of the log. However, we believe that it is not a significant problem in moni-

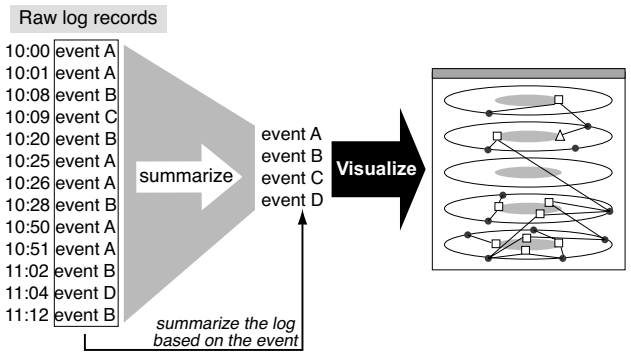
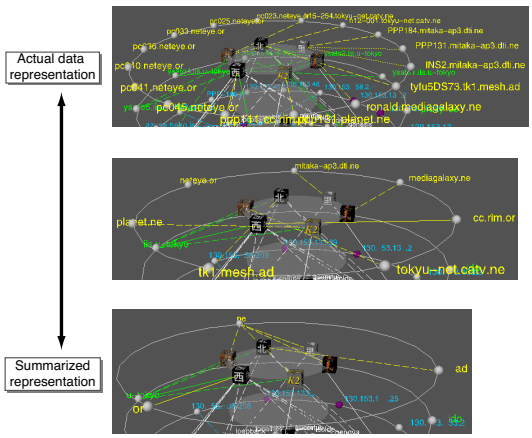


Figure 7. In illustration of log summarization



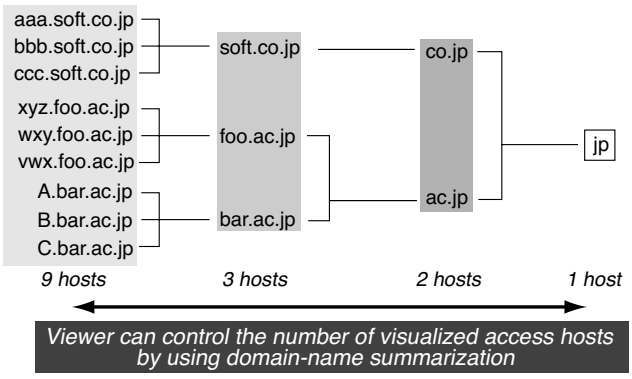


Figure 8. Log Summarization of host information based on domain name

toring and auditing because these are secondary information in such tasks. The most important point is to recognize all events that occurred in the monitored computer. Log summarization is a good method to recognize all events more quickly. It also suppresses an increase in the amount of information being visualized. As a result, Tudumi effectively helps manual log monitoring and auditing.

Tudumi also has certain interactive functions.

Figure 9 shows an example of the transition of the display using interactive filtering. If you click on a cube, you can get a modified image such as the one shown in the right image of figure 9. This figure displays only selected users and its related information. Interactive filtering enables administrators to inspect focused information more easily and intuitively.

Figure 10 shows an example of changing the number of visualized layers. Tudumi can display only a portion of the layers. This function enables the viewers to not only control the amount of information being visualized but also to reduce the visual clutter resulting from displaying too many objects.

3. Detecting Anomalous Activities with Tudumi

In this section, we give two examples of anomalous activity detection using Tudumi. We will describe how Tudumi has the ability to detect such activities.

First of all, it is clear that Tudumi has the ability to detect some of specific anomalous activities, as explained in the previous section. Tudumi classifies log messages into several groups based on the emergence probability of each event. This clarifies the degree of anomaly of each event. Tudumi, moreover, visualizes all of them in a specific manner that reflects the degree of anomaly of them in its representation. It effectively assists administrators to detect anomalous user activities. These features also make manual monitoring and auditing of the log more reliable.

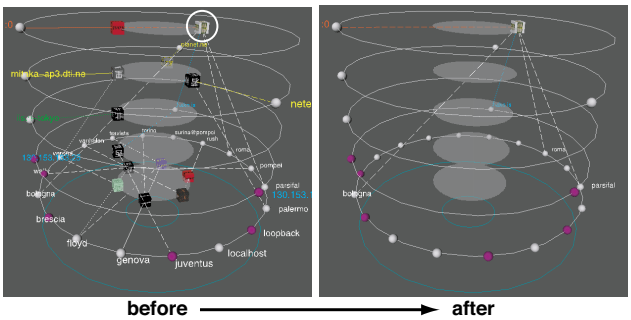
We, furthermore, introduce two examples of detecting the anomalous events using Tudumi. These examples show the usefulness of visual integration of related logs.

Figure 11 illustrates the first example.

A sphere that has no connected line is shown in the bottom right of the figure. This sphere represents the existence of an anomalous access. According to our visualization rule, this means that someone accessed the monitored computer through the network, but he/she could not log in. Moreover, this sphere is displayed on the highest layered disk, meaning that the host that this user has gained accesses from is one that is seldom used to access the monitored computer by legitimate members. For the reasons described above, the viewer should be convinced that the instance of this visual example must be an intrusive attempt.

Figure 12 illustrates the second example.

This visual example shows a side view of Tudumi. A cube with three connected lines is shown. It should be noted in this figure that the origin of lines are in different layers.



If the viewer selects an object of interest with the mouse, Tudumi displays only the selected object and its related information.

Figure 9. An Example of Focused Log Extraction

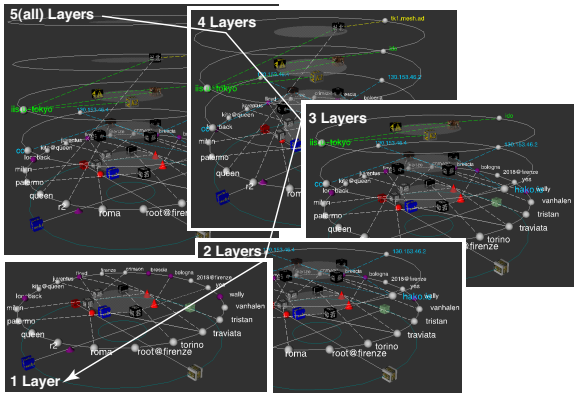


Figure 10. An example of display control of layers

This means that a user accessed the computer from three different domain groups in a specific period of time. If this figure resulted from the log messages generated in a short time frame, the viewer should suspect that someone broke into the monitored computer using that user's account. In this case, the three domain groups that the lines originated from are defined as the following: our laboratory, a company or educational facility in Japan and others in Japan. We consider such a situation as a rare occurrence in daily use, because a user normally does not access from the computers belonging to such different domains in a short term.

If the viewer has some knowledge about the particular user, (e.g., the possibility that this user might access the computer from each domain group.), he/she can evaluate the events more accurately as intrusive attempts or not. As an example, a rule that would be applied in our case: *No user would access to our computer from a foreign country's computer, if that user was in our laboratory just a few minutes before.*

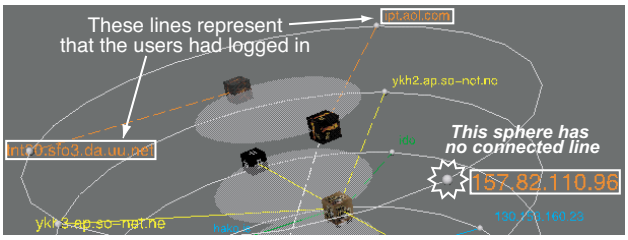


Figure 11. A visual example of intrusive access through the network

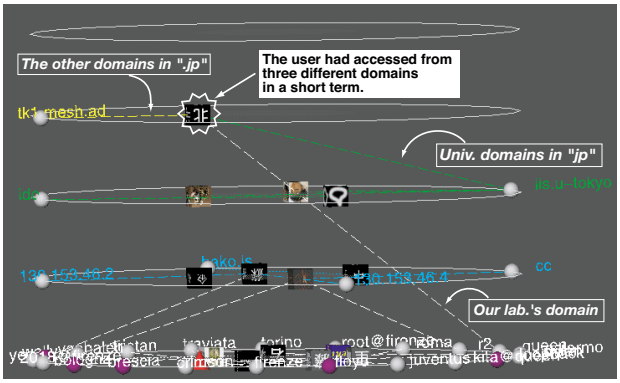


Figure 12. A visual example of user access from multiple domain groups

We have explained two visual examples of detecting anomalous activities. We emphasize that the ability to detect these anomalies result from integrating multiple log information and converting them into a visual image. It would not have been possible to detect them from only one of these logs. In Tudumi, we integrate log information from both network access and log-in user data.

4. Related Works and Discussion

In this section, we discuss Tudumi from two perspectives. One is in comparison with other visualization systems and intrusion detection systems. The other is regarding future work of Tudumi.

There are already some log visualization systems [1, 2, 3, 4, 5, 6]. According to these researches, there is no doubt about the effectiveness of visualization. We, however, question whether these systems are effective or not for detecting intrusive activities. The reasons are as follows: some of them visualize only one kind of log, have no way to handle a

huge amount of log messages, or they do not clearly distinguish between intrusive and non intrusive behavior visually. We consider that visualization system for detecting intrusive behavior[1] also have similar problems. We propose log summarization, reflecting the known rules into the visual representation and visualizing them in three-dimensional space for resolving these problems.

Intrusion detection systems(IDS)[7][8] also have problems. One is that it is difficult to introduce and manage them. This is a major reason why IDS is not used widely. The other is that it is still necessary to monitor and audit the log, even if you introduced IDS into your site.

As a result of the above discussion, we consider that a desirable tool for resolving these problems is a human interface for assisting administrators to monitor and audit computer logs. We think an interactive visualization system is the best solution.

The following lists our proposed future work. We will try to visualize other computer activities and integrate them into Tudumi. Tudumi presently visualizes only one aspect of various activities on a computer. Tudumi is helpful for detecting some kinds of anomalous activities. There are, however, other kinds of anomalous activities that administrators can not detect using only Tudumi. Another proposed feature is to present the degree of doubt as an intrusive symbol. Their extraction is currently based on only the rules defined by administrators. We will add analysis functions such as machine learning or data mining[9] for extracting them from the log.

5. Conclusion

We have described a log information visualization system called “Tudumi”. This system monitors and audits computer logs. Tudumi integrates three kinds of computer logs and visualizes them into one visual image. Our system assists administrators to monitor and audit the logs and makes detecting anomalous activities from them easier. Tudumi has useful functions to improve certain problems whenever administrators perform such tasks. These are log summarization, interactive operations and reflecting known rules into the visualization method.

We agree that IDS is an essential tool. We, however, also understand that it is difficult to make use of the advantage of IDS practically. We, therefore, propose another approach to deal with this problem. This approach is to monitor and audit the logs that you can easily obtain from your computer. In order to realize this approach, a strong human-interface is required for performing these tasks. Tudumi is an example of a system for that purpose. Our system complements the functions of IDS. We also aim to present Tudumi as a human-interface for novice system administrators.

References

[1] Robert F. Erbacher and Deborah Frincke: Visualization in Detection of Intrusions and Misuse in Large Scale Networks, Intl. Conf. of Information Visualization, pp.294-299, July, (2000).

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

[3] S.G. Eick, M.C. Nelson and J.D. Schmidt: Graphical Analysis of Computer Log Files, COMMUNICATION OF THE ACM, Vol.37, No.12, pp.50-66, (1994).

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

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

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

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

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

[9] W.Lee and S.Stolfo: Data Mining Approaches for Intrusion Detection, In Proc. of 7th USENIX Security Symposium, Jan (1998).