3 Traceable Network Technology

3-1 Research and Development of Traceable Network Technology

KADOBAYASHI Youki

Open networks mandate improved security of connected devices and their users, as well as improved security in both hardware and software of networking nodes. Situation awareness is thus an essential capability, although scale and speed of networks keep continuing growth. It is thus imperative to develop technologies to complement labor-intensive monitoring of networks. This article describes the research and development of traceable networking technology that enables accurate and accountable situation awareness across large-scale, high-speed networks.

Keywords
Network security, Traceable network

1 Introduction

Recently, the global penetration of open network technology as represented by the Internet has caused various tensions as well as global economic development. While the value of the open networks themselves are increasing significantly as their users increase, interactions fraught with a variety of problems such as conflict of values, knowledge gap and economic gap could occur at any time. Although there may be criticism against such traits of open networks, the communication infrastructure that transcends the conflict of values, knowledge gap and economic gap has become the driving force of the recent global development, and the capacity for safely using this infrastructure is a great source of competitiveness.

When the project was launched, we did not have basic tools for improving the security of such open networks. In other words, we did not have sufficient means to observe what had happened in the networks, and thus it was also difficult to develop preventive measures.

1.1 The need for situation awareness in networks

Efforts have been made to account for what is actually happening inside the networks, and the preceding study on the visualization and monitoring of the networks is one of them. Visualization and monitoring are very effective methods of situation awareness, and the Security Advancement Group within NICT also worked on these methods in the previous medium-term plan. Based on these efforts and the experience of actual network operation, we recognized the following four limitations when the project was started, and we thought that the challenge to these limitations would create the next research trends:

1) Limitation of scale: situation awareness by visual confirmation is difficult in networks beyond certain scale.
2) Limitation of precision: conventional simple pattern matching or visual detection drastically deteriorates precision.

3) Limitation of speed: visual monitoring cannot keep up with increasing speed of networks.
4) Limitation of time: it is difficult to assess what is happening in real time.

While it is quite difficult for our group alone to solve these issues at once, this was a perfect opportunity for the group to set such a grand challenge and work together with research groups in Japan and abroad.

1.2 Traceable network

The four limitations as described in **1.1** can be integrated into one grand challenge. In short, it can be described as "highly precise and traceable situation awareness in large-scale, high speed networks." We call the network that can meet this requirement the "traceable network," which we set as our research target. We assumed 100Gbps network as high-speed network, as this was the projected bandwidth of the backbone technology at the end of the project. We also determined that high precision would mean a false detection rate of 1% or less. With false detection rate being around 10% in the conventional detection method, we made attempts to significantly improve the detection rate.

We started by exploring the approach to this grand target, drawing on all the latest research results in the field of information science.

2 Research efforts on traceable networks

Operator's logs or confirmation by visualization systems are no longer sufficient to realize traceable networks. Therefore, it is necessary to lower the operator's burden by having computers take care of the major part of the recognition function instead.

In network security, discerning false positives placed a burden on operators. This burden can be lowered by improving detection precision and reducing false detections, which will enable operators to more efficiently respond to the growing scale of networks. However, it had often been the case that efforts for improving detection precision have been made through the application of conventional machine learning algorithm by security experts. Little effort had been made to understand the target area by machine learning experts, and to optimize the algorithm for the target area.

Internet and other large scale networks are not limited to a single service provider. In such an environment, coping with problems without disclosing detailed information to other service providers is required from the perspective of the privacy of communications and other legal requirements. In other words, it is required to strike a balance between privacy and the capacity to address problems.

On the other hand, network applications are becoming increasingly important, and there is a need to assess the situation when application-related problems occur, such as junk mails, and to cope with them. In short, it is necessary to increase the coverage to applications, in addition to the core of the network.

With these considerations in mind, our group made multilateral efforts toward realizing traceable networks. A many-faceted approach involving algorithms, systems and protocols was necessary to achieve this goal. Here we introduce these efforts along with four pillars; streamlining incident response, extension of coverage, enhancement of analytical capacity, and securing privacy.

2.1 Streamlining incident response

As described above, development of detection algorithms with higher precision had been desirable, because high-precision detection is directly connected to streamlining incident response. However, the characteristics of pattern recognition issues in network security had not been sufficiently considered in machine learning algorithms at the beginning of the project. These characteristics can be summarized into four points as follows:

1) Class imbalance: there is a significant deviation in frequency distribution between classes.
2) Multiple classes: Rather than the issue of

binary classification such as normal/abnormal, it is necessary to consider the issue of classification into three or more classes.

3) Incremental: data that arrive continuously, such as packets, need to be recognized.
4) Online: classification and recognition need to be conducted in real time.

In order to develop highly precise machine learning algorithms that have these characteristics, we conducted a joint research with Auckland University of Technology of New Zealand, and worked on the enhancement of support vector machines into multiple classes, and high-performance implementation using Graphics Processing Unit (GPU). As a result, we were able to obtain high performance, which is 50 times faster than conventional CPU, as well as false detection rate of 1% or less in some applications such as detection of junk mails.

Furthermore, naming conventions, numbering conventions, message formats and others need to be standardized for accurate and prompt transmission of detected problems to neighboring businesses or networks. The group started to work on standardization at the International Telecommunication Union Telecommunication Standardization Sector (ITU-T) from the third year of the project, and as a result, it evolved into an international effort as the X.1500 recommendation series (Cybersecurity Information Exchange (CYBEX) Techniques). The X.1500 recommendation series consist of highly modular draft recommendations that provide respective functions such as identification of cybersecurity information, structuring, assurance and exchange. By combining these modules, cybersecurity information can be exchanged. It is expected that the adoption of CYBEX in security devices and by organizations in charge of security measures will enable the standardization of naming convention, numbering conventions, message formats and others, thus more efficient transmission of the details of detected problems and responses to these problems will become possible.

2.2 Extension of coverage

At the beginning of the project, the biggest concerns as factors impeding application-level situation awareness were file sharing applications and virtual machines. Therefore, we worked on the research and development of their monitoring technology. While virtual machine can be considered a type of operating system technology, it can also be regarded as application container that can execute all kinds of applications, when seen from the perspective of the network. We therefore worked on the enhancement of the hypervisor functions (virtual machine execution environment) so that the virtual machine can be monitored from the hypervisor, and then developed technology for monitoring the behavior of virtual machines. With information leakage being the largest threat in file sharing applications, we also worked to construct a system for observing the range of information leakage, and succeeded in establishing monitoring capability in multiple applications.

In practical use of these component technologies for the Internet, they need to be used in combination with technologies for securing privacy, which will be described later in this paper.

2.3 Enhancement of analytical capacity

In order to assess the situation in a traceable manner in the network, it is certainly important to capture the symptoms of problems precisely, but it is also necessary that the symptoms are associated with the consequent change of the situation. In order to achieve this, the ability to analyze the symptoms is necessary.

Such symptoms, for example, include character strings that look like data on the surface. They are, however, interpreted as a program when given to certain systems, leading to a system halt or information leakage. This is called an attack vector.

Although apparent failures such as a system halt can be easily observed, in order to identify which data has caused the problem, it

is necessary to reproduce a small-scale system and provide it with candidate data.

Given such a situation, we first worked on the research and development of a network reproduction technology that can reproduce local area networks and Internet backbones, and then constructed a small-scale attack reproduction system. By using network reproduction technology, a local area network or Internet backbone can be reproduced on a certain scale based on the network design, in which the system in question can be operated. Based on this technology, we created a small-scale attack reproduction system by isolating it from the Internet, putting attack vectors into the system, and equipping it with functions for remote operation/observation. Universities and research organizations were actually able to conduct analysis using this system, by putting attack vectors into the system from remote locations.

2.4 Securing privacy

High observation/detection/analysis capacity as described above may risk privacy when misused in general-purpose networks such as the Internet. For this reason, we have also conducted research on how to strike a balance between securing privacy and these capacities.

One such example is the secure two-party computation protocol. This enables a matching process between two parties without disclosing the content, using the characteristics of homomorphic encryption that enables computing multiplication and addition with the content still being encrypted (in other words, without looking at the content). In particular, we focused on the issue of secure set-intersection protocols that enable the two parties to learn whether they have observed the same event without disclosing the content to each other, and we were able to decrease the amount of computation significantly. This is an issue that must be dealt with to strike a balance between traceability and privacy.

Furthermore, such a secure two-party computation protocol can also be used for locating problems. This enables service providers to collaborate with one another and mitigate the problems without disclosing detailed information to one another, even in open networks consisting of multiple service providers.

3 Proliferation of traceable network research

While we are confident of having achieved sufficient research and development of component technologies for constructing traceable networks, we are not in a position to declare the achievement of the grand challenge, since only some part of our deliverables were put to practical use.

Meanwhile, we have published more than 110 papers as a result of working on the project for five years, and have published software source codes and other deliverables in a reusable fashion. Therefore, we believe that it is possible to embody the characteristics that the traceable networks aimed to implement, by utilizing the results in the new medium-term plan that started from fiscal 2011, as well as in the forthcoming field trials.

Themes such as streamlining incident response and securing privacy, which have been tackled within our project, are distilled and pursued by the entire research center in the new medium-term plan. This may sound rather paradoxical, but the deliverables of this project can also be utilized in situation awareness through visualization.

3.1 Sharing the goals of the research

One of the goals we had in this project was to share our task with the research community to which the permanent members belong, instead of depending solely on their individual research.

As a result, the Data Mining for Cybersecurity (DMC) competition was established within the International Conference on Neural Information Processing (ICONIP) in the area of machine learning, which will induce future efforts on pattern recognition in the area of cybersecurity, as much as in traditional problem domains such as voice and image recogni-

tion.

We were also able to establish a framework for information exchange among organizations in X.1500 (CYBEX) cybersecurity standard. CYBEX has indicated a big technological trend toward the networking of security technology. International standardization allowed research organizations and security service providers to share this trend.

3.2 Proliferation of the results into education

As part of the efforts to enhance analytical capacity, we have provided the anti-Malware engineering WorkShop (MWS) with a reproduced dataset. We have also provided the reproduction environment to the IT specialist program to promote Key Engineers as securitY Specialists (IT-KEYS), a program to foster security specialists among the following four universities: Nara Institute of Science and Technology; Japan Advanced Institute of Science and Technology; Kyoto University; and Osaka University. With regard to the MWS, we provided the MWS with a reproduced dataset which was created by applying the technology of the small-scale attack reproduction environment, based on the MWS dataset created by Cyber Clean Center and the Information Processing Society of Japan. We were thus able to pave the way for young researchers at universities and other institutions without reproduction environments to develop and expand various data analysis methods. As for IT-KEYS, we provided more than 20 graduate students with small-scale incident reproduction environments annually from fiscal 2008, as part of the program to foster security experts. Through these efforts, we were able to create a unique environment for hands-on exercises to learn the mechanism of various security-related problems that can occur in corporate networks, as well as the observation methods in an environment similar to the real network, followed by consideration of countermeasures to these problems.

3.3 Ripple effect on service providers

We also received high marks among security service providers through small-scale experimental studies particularly with regard to technologies for situation awareness of information leakage, and for balancing privacy and analytical capacity. On the other hand, it is also clear that software created by research organizations cannot be immediately used by service providers as part of the service.

Furthermore, the direction toward the networking of security technology presented in the X.1500 recommendation series may significantly change the way services are constructed by security service providers.

We are continuously following up these issues even after the completion of the project.

4 Closing

When we consider to improve security in open networks, we need to challenge the four limits of scale, precision, speed and time. In this paper, we presented the outline of the research and development of traceable networks that enable highly precise and traceable situation awareness in large-scale and high speed networks, as a concrete example of the above-mentioned effort. The deliverables of the project have been developed in various ways, such as by delivery to research communities, application to educational curriculum and standardization. Ripple effects on security or network service providers are expected to be seen in the future.

Research on network security is still on its halfway mark. This is a fusion area where a two-pronged approach is required, in the networking of security technology as well as in the improved security of networking technology. However, there are still few researchers who understand both of these technologies, and therefore it is still required to tackle the issues through collaboration in the form of project research.

Acknowledgments

The author would like to thank each member of the research group who earnestly worked on the research in this project, and those concerned who have given us the opportunity to pursue this research project.

(Accepted June 15, 2011)

KADOBAYASHI Youki, *Ph.D.*

Guest Expert Researcher, Network Security Research Institute/Associate Professor, The Graduate School of Information Science, Nara Institute of Science and Technology

Cybersecurity, Internet Engineering