On Detection for Scarcely Collided Super-Slow Port Scannings in IDSs' Log-Data

Kazuyoshi Furukawa[1], Satoru Shimizu[2], Masahiko Takenaka[1], and Satoru Toriil[1]
[1]Fujitsu Laboratories Ltd., Kawasaki, 211-8588, Japan
[2]Fujitsu Social Science Laboratory Limited, Kawasaki, 211-0063, Japan
Email: kaz.furukawa@jp.fujitsu.com, shimizu.satoru@jp.fujitsu.com, {ma, torii.satoru}@jp.fujitsu.com

Abstract**—**Recently, cyber-attacks against governments and enterprises more intensified, these have already taken on “Cyber-warfare.” Because the attack technics are more artful, it is too difficult to defend them perfectly. We began this research because super-slow port scannings are extracted from IDSs’ log-data placed in our managed networks for 4 months. In order to extract similar scannings from large log-data, a systematical detection method is required. In this paper, we propose a detection method of scarcely collided super-slow port scannings. This method uses only χ^2-value of number of accesses per each port without relying on time rate of traffic count. And, we report that plural kinds of scarcely collided super-slow port scannings can be detected in the IDSs’ log-data.

Index Terms**—**super-slow port scannings, detection method, χ^2-value.

I. Introduction

Recently, cyber-attacks against governments and enterprises more intensified, these have already taken on “Cyber-warfare.” U.S. government security expert described cyberspace as “the fifth domain of warfare [1].” Many entities have been attacked, and system intrusion and information leakage have been caused. Especially in “targeted cyber-attacks,” various attacks persist against specified targets. It is said that the attacks are sent sparsely mingling with normal traffics over the long term. Because the attack methods are more artful, it is too difficult to defend them perfectly.

This research was begun because quizzical traffics are extracted from IDSs’ log-data placed in our managed networks in 4 months. It has seemed that these traffics are super-slow vertical port scannings. (Here, ‘vertical port scanning’ is only described ‘port scanning’ for simplicity after this.) These scannings can be classified to into two types. One is “increment-type” we call, that is, accessed port number is used by incremental step. Another is “random-type” we call, accessed ports are randomly selected. These scannings were heuristic extracted by hand work. A heuristic extraction by hand work requires high skill and great care. Therefore, in order to be able to extract similar scannings from large log-data, a systematical detection method is required.

It is well-known that slow port scannings are an efficient technic to bypass IPSs/IDSs. Accurately, though each anomaly access, e.g. ‘half-scan’, can be detected by an IDS, the IDS cannot detect a fact that those accesses constitute port scannings. Thus, it cannot be expected that IDSs/IPSs detect sparsely port scannings for long term. There are various researches on detection for slow port scannings. The purpose of many researches, however, is to detect them in real time. It is not the target to detect sparsely port scannings for long term. On the other hand, there is research that traffics are visualized to detect slow port scannings. Although this technique can be used to detect super-slow port scannings, whether the traffics are port scannings or not must be judged from visualized them manually. This scheme is not suitable for our purpose. In order to detect these port scannings, we must analyze vast amounts of log-data for long term. It is important to extract them without relying on time rate of traffic count.

We propose a method by using χ^2-value of number of accesses per each port to detect these port scannings. This is an improvement of a method in [2] for detection of super-slow scannings. A purpose of port scannings is to find out open ports of target entity. It is most effective to scan ports of target with scarcely collided port accesses. With only χ^2-value, we detect the fact that accessed ports are scarcely collided. Thus, our method can detect them without relying on time rate of traffic count.

By using our proposal, we describe experiments to extract suspect traffics from IDSs’ log-data placed in our managed networks in 6 months (that is, the log-data of two other months was collected). And, we report that plural super-slow port scannings can be detected. As planned, our method has detected super-slow scannings which are extracted by hand work described above. That is, both increment-type and random-type scannings by individual scanners are detected.

Additionally, distributed super-slow scannings have been detected with our method, that is, these scannings are committed by plural scanners. The way of detections also is described. And the distributed super-slow scannings also are classified two types, that is, increment-type and random-type.

We analyze the behavior of scanners that commit the distributed super-slow scannings. We confirm the fact that these scanners work in cooperation obviously.

Manuscript received August 29, 2013; revised October 30, 2013.
Corresponding author email: torii.satoru @ jp.fujitsu.com.
doi:10.12720/jcm.8.11.788-794

In this paper, our motivation and previous works are described in Section II, and we propose a detection method in Section III. Section IV shows experiments with our proposal method. Finally, Section V concludes this paper.

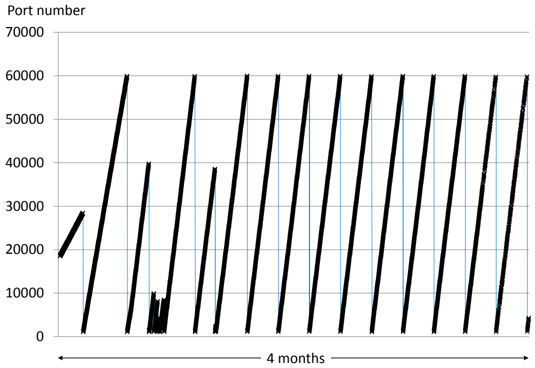

Figure 1. Increment-type port scannings

II. Super-slow port scannings

A. Motivation

We audited IDSs' log-data that is recorded in 4 months from December 2011 for our managed networks by hand work. Then, we noticed there are two set of quizzical traffics. As the results of check those in detail, external entities *scanner1* and *scanner2* commit port scannings against entities *victim1* and *victim2* in our managed networks, respectively. These are not accesses of typical port scannings but sparsely scannings for long term. Thus, we could not notice them without this auditing.

When these two set of traffics were visualized, we found that each have individual feature of scanning pattern. One is a scanning pattern from *scanner*1 to *victim1* (See Fig. 1). This scanner decides port numbers by incremental steps. From Fig. 1, sequence of accessed port number seems to be successive. In fact, the sequence is discrete, and the steps are not constant. We classify this scannings into "increment-type".

Another type of scannings is from *scanner2* to *victim2* (See Fig. 2). This scanner randomly decides port number. We classify these scannings into "random-type".

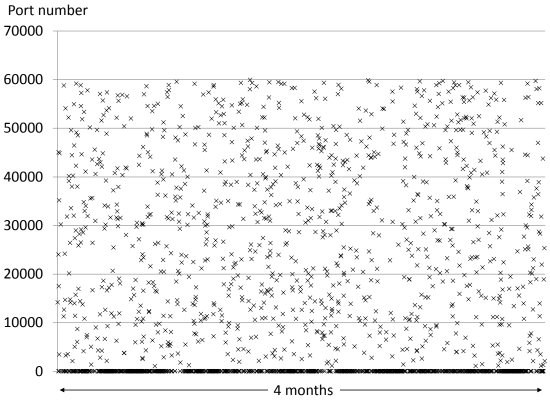

Figure 2. Random-type port scannings

Both scannings have following mutual features:

- Super-slow scannings (average several time per hour),
- Scannings against over port (1024 – 65535),
- Scanned port numbers are scarcely collided.

Such super-slow port scannings cannot be detected in existing IDS/IPS products. Mentioned above, though each anomaly access, e.g. 'half-scan', can be detected by an IDS, the IDS cannot detect a fact that those accesses constitute port scannings. In order to detect the port scannings, administrator must extract them from vast IDS's messages heuristically. A heuristic extraction by hand work requires high skill and great care. Therefore, in order to be able to extract similar scannings from vast amount of log-data, a systematical detection method is required. Our purpose of this research is that super-slow port scannings are effectively detected from vast amounts of log-data for long term.

B. Previous Works

Typical detection methods of port scannings rely on time rate of traffic count. A basic approach to detect slow port scannings is spreading time window of measurement and judgment. Snort, that is a typical open source network IDS, has a function to detect port scannings [3]. The time window is maximum 600 sec. When access count in this period is above threshold level, snort judged these accesses to be port scannings. In default setting, snort cannot detect super-slow port scannings which count is average several-time per hour. The more spread time window, the more heavy operation is required. Thus, it is difficult to detect super-slow port scannings with the basic approach.

In order to detect slow port-scannings, various approaches without relying on time window are studied.

A detection method of port scannings with time independent feature set was proposed in [4]. Instead of time rate of traffic count, access log stored in IDS's memory is used, which the log is compressed data of a scanner's IP address and the port number, a victim's IP address and the port number, and access time. Their method replaces the limitation of time window with the limitation of memory. In [5], to spread time window, number of observations of scanners' IP addresses are confined. According to number of accesses, scanners' IP addresses are weighted based on fuzzy logic in 3 steps, that is 'Attack', 'Suspicious', 'Normal'. By continuously observing only scanners' IP addresses with high weight, the processing load can be reduced. There is an approach that communicative sessions instead of packets are focused in [6]. By managing scanners' IP addresses with "session window," long term observations are enabled.

Either approaches, however, is aimed at real-time detection against slow port scannings. It is difficult to be applied to super-slow port scannings because sparsely scannings continue for several months.

There is completely a different approach. A detection method by visualizing traffics was proposed in [7].

Although this method can detect slow port scannings, whether the traffics are port scannings or not must be judged from visualized them by hand work. This scheme is not suitable for our purpose.

However, this paper points an important fact. That is existence of distributed slow port scannings. In distributed scannings, amount of traffics per a scanner is reduced. Then, it is more difficult to detect the slow port scannings.

III. Super-Slow Port Scannings

Focusing a feature of super-slow port scannings, we propose a detection method. Then, we discuss about a threshold between general communications and port scannings.

A. *Scarecely Collided Scannings*

As mentioned above, it is difficult to detect super-slow port scannings by an approach relying on time rate of traffic count. Thus, a time independent statistical features should be found.

A naive approach is that number of accessed ports is counted in the log-data. If number of accessed ports is 'large', the accesses might be detected as port scannings. This approach is similar to [4]. However, to discriminate 'large' relies on time rate.

Fig. 1 has distinguishable pattern. By discriminating the pattern, such increment-type accesses might be detected. However, it cannot detect random-type accesses like as Fig. 2. For example, Snort, open source port scanning tool, has a function to port scannings with randomly selected ports [3].

In this paper, we focus another feature of port scannings. A purpose of port scannings is to find out open ports of target entity. It is most effective way of port scannings that a different port number at each access is used. Thus, sequence of port numbers that is used by port scannings is scarcely collided. Super-slow port scannings also have this feature, and we focus it. Because this feature is irrelevant to time, time independent detection method can be achieved.

B. *Detection Method by χ^2-Value of Port Accesses*

In this subsection, we study on discriminating a feature of scarcely collided port number.

A method focused on this feature is proposed in [2]. In [2], two dynamic detection methods by using χ^2-value for port scannings are proposed. The purpose of these methods is to distinguish anomaly packets from normal packets. We propose to apply one of these methods. In [2], however, target data is not IDSs' log-data but dynamically captured packet. IDSs' log-data include only alert messages. Thus, the method cannot be applied as it stands. Especially, way to decide the threshold is greatly different from [2].

We show to use χ^2-value for this index. It is assumed that accesses of port scannings are uniform distribution to all ports. Then, χ^2-value of number of port accesses is used for an index of degree of collisions. The more collisions are the more χ^2-value increases.

Let n is number of target ports, k is number of all objected accesses, x_i is number of accesses to i-th port. The χ^2-value is presented as follows:

$$\chi^2 = \sum_{i}^{n} \frac{\left(x_i - \frac{k}{n}\right)^2}{\frac{k}{n}} = \sum_{i}^{n} \frac{(nx_i - k)^2}{kn} \qquad (1)$$

C. *Threshold of χ^2-Value*

In this subsection, we study on threshold of χ^2-value between general communications and scannings.

In general χ^2-test for a hypothesis, significant level (e.g. 10% rejection) is set. If χ^2-value of observations exceeds the level, the hypothesis is rejected. In order to test a hypothesis that accesses of port scannings are uniform distribution to all ports, this threshold is adopted. However, our purpose is not testing the hypothesis, but distinguishing between normal communications and port scannings.

In [5], χ^2-values of plural sets of normal packets are calculated, and the distribution of these χ^2-values is obtained. Because χ^2-value for port scannings is outlier, the set of packets can be detected. However, IDSs' log-data includes only alert messages, and it is difficult to obtain a distribution of sets of normal packets. Therefore, new threshold value is required for detecting super-slow port scannings.

In typical port scannings, accesses with scarcely collided port number are occurred. When port number of all accesses is quite different, χ^2-value becomes the minimum. In this case, the χ^2-value, $\chi^2_{\min}$, is presented as follows:

$$\chi^2_{\min} = k\frac{(n-k)^2}{kn} + (\mathrm{n}-\mathrm{k})\frac{k^2}{nk} = n - k, \quad (k \le n).$$

On the other hand, when all accesses are sent to one port of target, χ^2-value becomes the maximum. In this case, the χ^2-value, $\chi^2_{\max}$, is presented as follows:

$$\chi^2_{\max} = \frac{(n-k)^2 + (n-1)k^2}{kn} = k(n-1).$$

χ^2-value of all communications exists between $\chi^2_{\min}$ and $\chi^2_{\max}$.

In normal communications, one or several ports are repeatedly accessed. It is clear that χ^2-value of almost normal communications approximates into $\chi^2_{\max}$. In this paper, to distinguish between normal communications and port scannings, we empirically adopt threshold of χ^2-value, $\chi^2_{\text{threshold}}$, shown as follows:

$$\chi^2_{\text{threshold}} = \sqrt{k}(n-1)$$

This threshold of χ^2-value is small enough compared with $\chi^2_{\max}$ to distinguish normal communications. Then, this value is large enough compared with number of freedom degree of χ^2-value, $\phi = n - 1$. Therefore, it can distinguish port scannings even if the noise is included in observed value.

In addition, it seems that false detections often occur with our method when number of observations is small. Therefore, a condition, $k \geq \sqrt{n}$, is added in this paper.

IV. Experiment Analyses with Our Proposal

In this section, we show results of experimental analyses with our proposed method. Experiments include detections of not only super-slow scannings described above but also distributed type of these scannings.

A. *Target Data Sets*

Data sets that we analyze are IDSs' log-data. These IDSs are placed in front of our managed networks. And the log-data was recorded in 6 months from December 2011. (When we analyzed it by hand work, log-data in 4 months was provided.) Terms of logs that are used in experiments are alert logs of port scannings. These are not only logs that IDS judges accesses to be port scannings but also the accesses are suspected as port scannings, which are for example "half scanning". Then, targets are accesses to over port (port number: 1024 – 65535) in the log-data.

In this regard, we omit accesses of which port numbers are 1433 and 1434. These ports are used by Microsoft SQL server [8], and there was vulnerability by using these ports in 2002. "Slammer" worm uses this vulnerability [9], a large amount of scannings to this port is still done. Because our purpose is not detecting known scannings, we omit accesses of port number 1433 and 1434 in the log-data.

B. *Detection of Super-Slow Port Scannings*

At first, we have experimented with detecting super-slow scannings from a scanner to a victim. As results, plural super-slow scannings including accesses of Fig. 1 and 2 have been detected as planned. Thus, one of our purposes, that are actualizing a systematical detection method against super-slow scannings, has been achieved.

Detected scannings other than accesses of Fig. 1 and 2 have same features as them though access frequency is different. However, these scannings are also found at detection of distributed type. Therefore, we discuss on this accesses in next subsection.

C. *Detection of Distributed Super-Slow Port Scannings*

Equation (1) consists of only terms about target (victim) port and number of access, and information about scanner is not needed in it. Thus, (1) can evaluate port scannings from not only a specified scanner but also plural scanners.

In [7], distributed slow port scannings are described. Mentioned above, we have detected port scannings by a specified scanner. When we make similar experiment without specifying scanners, it seems that distributed super-slow port scannings are detected.

Then, we have experimented with detecting super-slow scannings of distributed type, that is, scannings are committed by plural scanners to a victim. As results, two types of plural distributed super-slow scannings have been detected. One is "increment-type", and another is "random-type". Each has similar feature as the type of super-slow port scannings described in Section II.

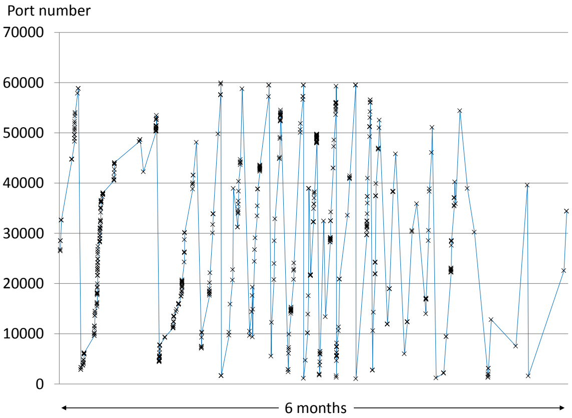

Figure 3. Increment-type distributed super-slow port scannings to *victim3* (date – port number)

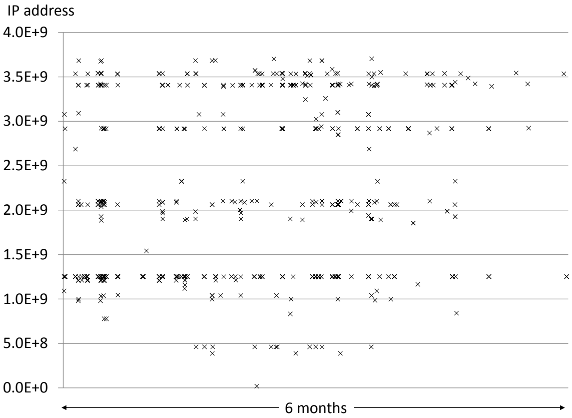

Figure 4. Increment-type distributed super-slow port scannings to *victim3* (date –scanners' IP address)

1) *Increment-type*

TABLE I. Number of Scanners' IP Addresses and Scannings to *victim3* Classified based on Assigned Countries and Regions

Country code	US	JP	SG	BG	IE	KR	RU	AU	CN	CH	NL	VG
#IP addresses	124	105	23	2	2	1	1	1	1	1	1	1
#scannings	332	285	33	2	2	2	2	1	1	1	1	1

A distributed super-slow port scannings of increment-type to *victim3* has been detected. Fig. 3 shows a port access pattern of it in 6 months, and shows a relation between time and IP addresses of scanners regarding same data a. From Fig. 3, plural scanners access ports that numbers are decided by incremental steps. These scanners must have in cooperation. These accesses are 663 distributed super-slow port scannings that 657 ports are scanned from 263 scanners. And this is a first report that distributed super-slow port scannings are committed in the real world.

Then, we have classified these scanners' IP addresses and number of scannings based on assigned countries and regions, and we have shown the statistics in Table I. Most scanners have IP addresses of US and Japan.

TABLE II. NUMBER OF SCANNERS' IP ADDRESSES AND SCANNINGS TO *VICTIM4* CLASSIFIED BASED ON ASSIGNED COUNTRIES AND REGIONS

Country code	US	CA	FR	NL	JP	BR	AU	UA	CN	CH	SE	VG	BG	CZ	TW	BR	DE	KR	RU
#IP addresses	17	2	4	8	2	7	1	4	4	1	2	1	1	1	2	1	1	1	1
#scannings	1094	53	28	26	24	23	11	9	8	6	4	3	2	2	2	1	1	1	1

2) *Random-type*

Two distributed super-slow port scannings of random-type to *victim4* and *victim5* have been detected. Fig. 5 shows a port access pattern of scannings to *victim4* in 6 months, and shows a relation between time and IP addresses of scanners regarding same data. These accesses are 1299 distributed super-slow port scannings that 1284 ports are scanned from 62 scanners.

Fig. 7 and Fig. 8 shows a port access pattern and scanners' IP addresses to *victim5*. In this case, 4189 scannings to 4030 ports from 9 scanners are found.

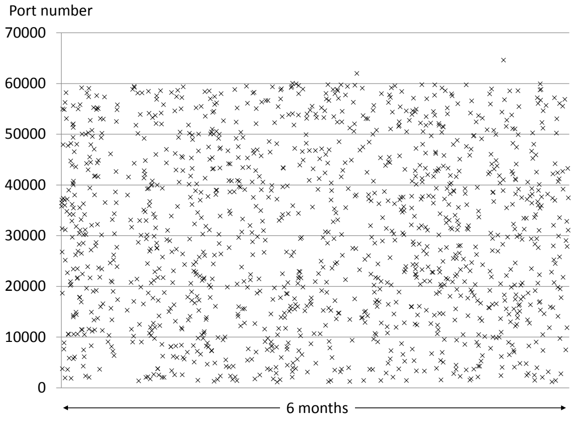

Figure 5. Random-type distributed super-slow port scannings to *victim4* (date – port number)

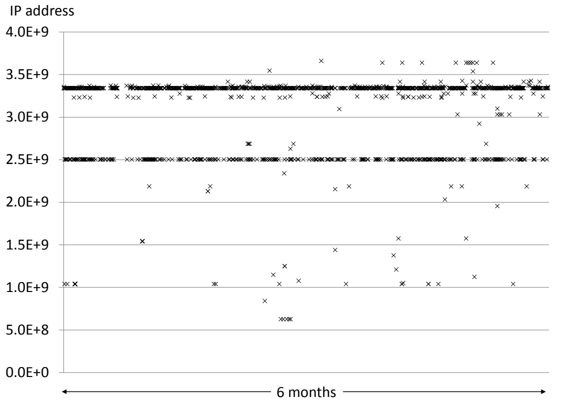

Figure 6. Random-type distributed super-slow port scannings to *victim4* (date –scanners' IP address)

Then, we have classified these scanners' IP addresses and number of scannings to *victim4* and *victim5* based on assigned countries and regions, and have shown the statistics in Table. II and Table. III, respectively. In both scannings, almost scannings are committed from IP address of US. And almost the scannings from US are committed by 3 scanners. It is clear from Fig. 6 and Fig. 8. The 3 scanners scan both *victim4* and *victim5*.

TABLE III. NUMBER OF SCANNERS' IP ADDRESSES AND SCANNINGS TO *VICTIM5* CLASSIFIED BASED ON ASSIGNED COUNTRIES AND REGIONS

Country code	US	CA	CN	FR	VG
#IP addresses	3	1	1	2	1
#scannings	4176	8	2	2	1

Two lines have clearly appeared in Fig. 6 and Fig. 8. In practice, upper line consists of accesses of adjacent 2 IP addresses. Thus, it seems that these 2 scanners clearly commit port scannings in cooperation.

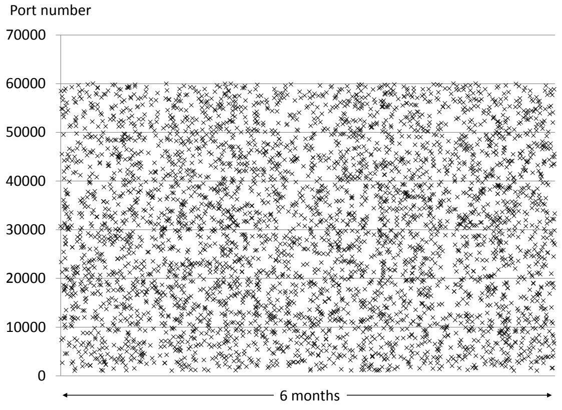

Figure 7. Random-type distributed super-slow port scannings to *victim5* (date – port number)

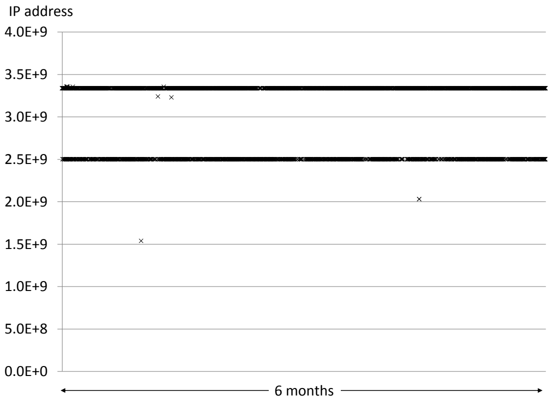

Figure 8. Random-type distributed super-slow port scannings to *victim5* (date –scanners' IP address)

Fig. 9 shows intervals of scannings from these 3 scanners from US to *victim5* in no particular order. There are distinctly more accesses at period of just one hour. If these 3 scanners individually access it, the period should become more random. Therefore, it seems that these 3 scanners are working together.

D. Discussion

From the IDSs' log-data that we have used in this paper, 8 suspicious data sets are extracted with our proposed method. All these data sets have discriminative access pattern of super-slow port scannings and distributed super-slow port scannings as mentioned above. Thus, it is thought that number of false positive matches with our method is little.

On the other hand, it is clear that false negative occurs. For example, initial scannings cannot be detected with our method. For validation of false negative, we have analyzed accesses to other hosts from detected scanners by hand work. Because almost scanner to *victim5* commits super-slow scannings to *victim4*, it seems that a similar case is found.

As results, it has been revealed that each type of scanners commit same type scannings to other hosts, respectively. And some of them are undetected super-slow scannings with our method.

In the following, we discuss these undetected super-slow scannings.

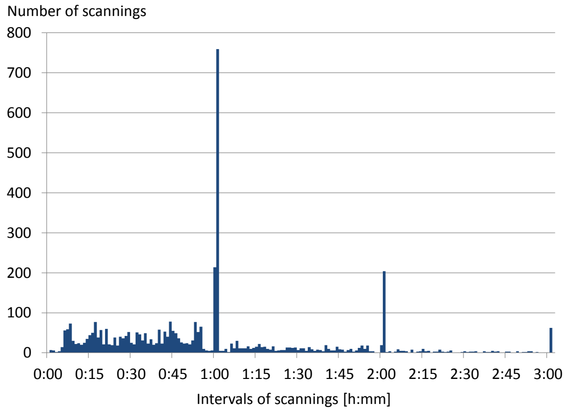

Figure 9. Interval of scannings from top 3 scanners to *victim5*

1) Scanners of increment-type

Almost scanners that frequently scan *victim3* also commit similar scannings to other hosts in our managed networks. Fig. 10 and 11 show scanning patterns to *victim6* and *victim7* respectively. It seems that these scanners construct a scanner-network and commit distributed super-slow scannings in cooperation.

These scannings could not be detected by our proposal method. Because other scanners accessed a specific port over 70000 times, scannings to *victim6* cannot be detected. Many accesses to specific ports produce an increase in χ^2-value, and it obstruct our detection. If this specific port is used by a real service of *victim6*, log-data of accesses to this port should be omitted. Thus, the super-slow port scannings become being detected with our method. If not, it seems that *victim6* has received two types of attack. We think that an attack to the specific port should be detected at first.

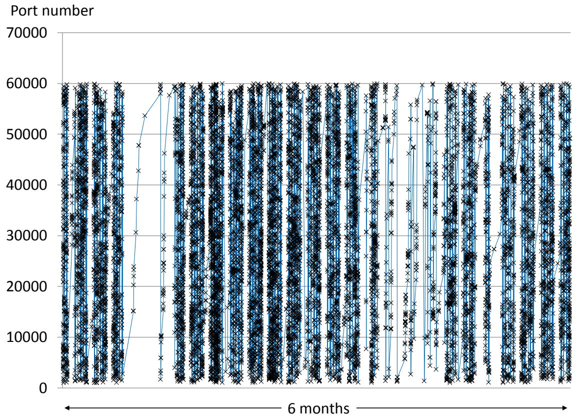

Figure 10. A scanning pattern to *victim6* from scanners of *victim3*

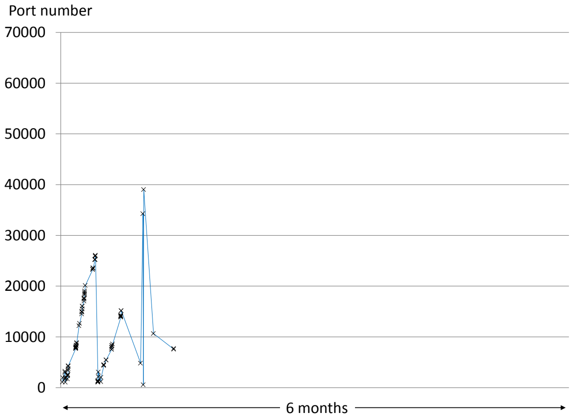

Figure 11. A scanning pattern to *victim7* from scanners of *victim3*

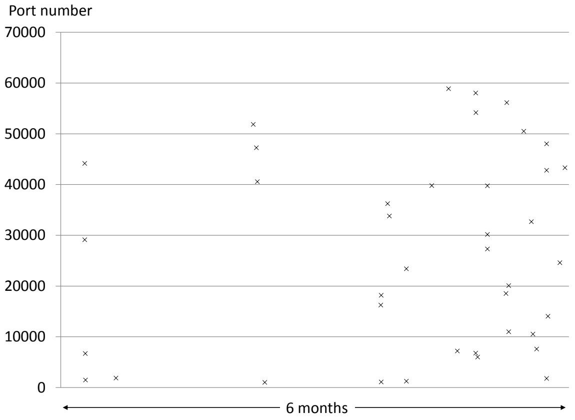

Figure 12. A scanning pattern to *victim8* from scanners of *victim4*

Regarding scannings to *victim7*, number of accesses has been small yet. Thus, it doesn't satisfy a condition $k \geq \sqrt{n}$.

Fig. 11 shows an interesting property. The distributed super-slow scannings have terminated. According to frequency of accesses, it takes several years though these scanners access to all ports. From only this log-data, we could not know the scanning period.

2) Scanners of random-type

Scanners of random-type port scannings to *victim4* are classified into two classes. One commits similar scannings to *victim5*, and 3 scanners commit almost scannings as mentioned above.

Another commits more distributed port scannings. These scanners also commit similar scannings to *victim8* in our managed networks. Fig. 12 shows scanning pattern to *victim8*. Because number of accesses to *victim8* has been small yet, the scannings could not be detected by our proposed method. However, it seems that scannings of *victim8* become possible detection in a little while.

In order to detect earlier, it is possible to loosen the condition $k \geq \sqrt{n}$. On the other hand, loosening the condition causes false positive. It is a future work.

V. Concluding Remarks

In this paper, it have been firstly reported the fact that super-slow port scannings are committed in the real world, and a detection method based on χ^2-value has been proposed. Using our method, we have experimented to analyze IDSs' log-data of our managed networks in 6 months. Then, not only typical super-slow port scannings but also distributed super-slow port scannings have been detected. The access patterns of detected super-slow port scannings have been classified into two types, increment-type and random-type. And we have clarified that plural scanners commit port scannings in cooperation and make up scanner-networks to scan many hosts. Our results have contributed that it is signaled that super-slow port scannings are committed in the real world and that these can be detected.

References

[1] R. Clarke, *Cyber War: The Next Threat to National Security and What to Do About It*, 1st ed. New York, USA: Ecco, 2010.

[2] H. Kim, S. Kim, M. A. Kouritzin, and W. Sun, "Detecting network portscans through anomaly detection," in *Proc. Signal Processing, Sensor Fusion and Target Recognition XIII,* 2004, pp. 254–263.

[3] Snort: Open Source Network Intrusion Prevention and Detection System. [Online]. Available: http://www.snort.org/

[4] H. Baig and F. Kamran, "Detection of port and network scan using time independent feature set," in *Proc. Intelligence and Security Informatics*, New Brunswick, 2007, pp. 180–185.

[5] J. Kim and J. Lee, "A slow port scan attack detection mechanism based on fuzzy logic and a stepwise policy," in *Proc. 4th International Conference on Intelligent Environments*, Seattle, 2008, pp. 1-5.

[6] M. Dabbagh, A. Ghandour, K. Fawaz, W. Hajj, and H. Hajj, "Slow port scanning detection," in *Proc. 7th International Conference on Information Assurance and Security*, Melaka, 2011, pp. 228-233.

[7] C. Muelder, K. Ma, and T. Bartoletti, "Interactive visualization for network and port scan detection," in *Proc. 8th international conference on Recent Advances in Intrusion Detection*, 2005, pp. 265–283.

[8] Microsoft, SQL Server. [Online]. Available: http://www.microsoft.com/en-us/sqlserver/

[9] CERT Advisory CA-2003-04, MS-SQL server worm. [Online]. Available: http://www.cert.org/advisories/CA-2003-04.html

Kazuyoshi Furukawa received his B.E. degree in information science in 2002 from Hiroshima City University. He received his M.E. degree in science and engineering in 2004 from Tokyo Institute of Technology. Since 2004, he has been engaged in research and development on cryptography and mobile security at Fujitsu Laboratories Ltd.

Satoru Shimizu has been engaged in research and development on security and network at Fujitsu Social Science Laboratory Limited. He has licentiates of network specialist and security specialist in JITEC.

Masahiko Takenaka received his B.E. and M.E. degrees in electronic engineering in 1990, 1992 respectively from Osaka University. He received his Ph.D. in engineering in 2009 from Tsukuba University. Since 1992, he has been engaged in research and development on cryptography, side channel analysis and network security at Fujitsu Laboratories Ltd. He is currently a research manager. He was awarded Computer Security Symposium (CSS) Paper Prize in 2002 and 2004, and the OHM Technology Award in 2005. He is a member of IEICE.

Satoru Torii received his B.E. degrees in information science in 1985 from Tokyo University of Science. Since 1985, he has been engaged in research and development on network security at Fujitsu Laboratories Ltd. He is currently a research manager. He was awarded Computer Security Symposium (CSS) Paper Prize in 2004, He is a preceding director of IPSJ.