# An Internet Threat Evaluation Method based on Access Graph of Malicious Packets

Masaki Ishiguro*[1] Hironobu Suzuki*[2] Yoichi Shinoda*[3] Ichiro Murase*[1] Shigeki Goto*[2]

**Abstract.** Malicious packets generated by Internet worms or port scans can be captured by monitoring ports of IP addresses where any network service is provided. Several methods have been proposed for detecting threats over the Internet by monitoring malicious packets. Most of these methods apply statistical methods to time-series frequencies of malicious packets captured at each port.

This paper proposes a new method for evaluating threats in the Internet based on access graph defined by the relation between sources and destinations of malicious packets. This method represents access relation between sources and destinations of malicious packets by bipartite graph and defines relation of threat and vulnerability between sources and destinations of malicious packets. In order to evaluate threats on the Internet, we apply a new method to this relation. This method evaluates threats by using spacial structure of access graph which has not been used by traditional methods. We applied our method to working examples monitored during the period of worm outbreaks to show the effectiveness of our method.

## 1 Introduction

Malicious packets such as Internet worm activities, DDoS attacks, port scans are monitored on the Internet. Internet monitoring system has been developed in order to detect threats over the Internet by monitoring these malicious packets. While Intrusion Detection Systems (IDS) monitor within the local network to detect intrusion or misuses, Internet Monitoring Systems monitor several IP addresses outside local network in the Internet. Several threat detection methods such as statistical discrimination method on time-series frequencies of malicious packets or extraction of characteristic access patterns have been proposed.

In this paper, we propose a new threat evaluation method based on spacial structure of access graph formed by source and destination relations of monitored packets in the Internet. In order to quantify threat in the Internet, we apply an eigen equation method to the access graph of malicious packets.

The rest of this paper is organized as follows: In section 2, related works are described. In section 3, a new threat evaluation method is presented. In section 4, experimental evaluation of our method and consideration are presented. Finally we conclude in section 5 with some remarks and future works.

## 2 Related Work

There are two types of Internet Monitoring Systems which aims at detecting threats on the Internet. The first one monitors every packets without making any response which is called passive monitoring, while the other monitors packets and sends back some response packets in some degree in order to observe actions of packets senders which are called active monitoring. The former includes CAIDA telescope[MSVS04], Internet Storm Center[SAN], Internet Motion Sensor[Uni], JPCERT/CC, ISDAS[JPC],

[1] Information Security Research Group, Mitsubishi Research Institute, Inc

[2] Faculty of Science and Engineering, Waseda University

[3] Japan Advanced Institute of Science and Technology

WCLSCAN[ISMO04], DShield[DSh]. The latter includes the work by Princeton University[PYB$^+$04] and Honeypot[Pro] by Honeynet Project.

Most of threat detection methods are based on statistical analysis of time-series frequencies of monitored packets of individual network port. Thottan proposed auto-regression model method which computationally learns and predict change of time-series frequencies of packets and make statistical test to detect threats in the Internet[TJ03]. Ishiguro proposed detection method based on Bayesian estimation to the difference between time-series frequencies and their trends[ISMO04]. Zou proposed a method for detecting evolution of Internet worm activities based on virus infection model in epidemics and Kalman filter[ZGGT03]. Telecom-ISAC/Japan is working on extracting characteristic access patterns based on correlation of source and destination information of monitored packets. In terms of active monitoring, evaluation of likelihood of Internet worm infection by monitoring failure or success of TCP connection[SJB04]. Kompella proposed the number of differences between monitored FIN packets and SYN packets[KSV04].

All of them focus on the number of packets monitored in stead of structure of access graph formed by monitored packets. This paper proposes a new method which takes a structure of access graph into account.

## 3 A Threat Evaluation Method

In this section, we present a threat evaluation method which takes advantage of structural information of access graph of monitored packets. First, we summarize structure of our Internet monitoring system and data specification for analysis. Then, comparing with the traditional method for threat detection, we describe the way to evaluate threat in the Internet and calculation method.

### 3.1 Internet Monitoring System and Target Data

Our threat evaluation method uses packet information such as access time, packet source, packet destination monitored by passive Internet monitoring system. Packets monitored at IP addresses where any network services are given are considered to be malicious packets, because anyone would access to such IP address for requesting normal network services. These malicious packets include worms' infection activities, DDoS back-scatters, port-scans etc.

Figure 1 shows structure of our Internet monitoring system.

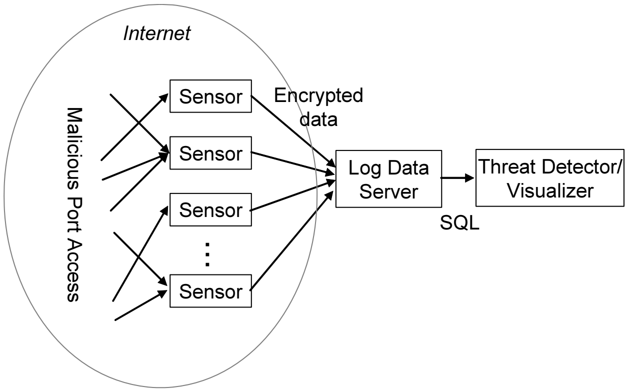

Figure 1: the Internet monitoring system

The system consists of multiple Sensors, a Log data server, and a Threat detector/visualizer. Sensors are deployed at several IP addresses and captures arriving packets. Information of packets captured at sensors is transferred to the log data server via an secure channel. The threat detector/visualizer analyse monitored packets data and detect threat in the Internet. Data to be analyzed by our threat evaluation method are summarized in Table 1

### 3.2 Relationship between Source and Destination

In this paper, we consider Internet worm which is highly contagious to be threat in the Internet. Highly contagious worms search effectively hosts

Table 1: Target data

| |
|---|
| Packet Access Time(Date,Time) |
| Protocol Type (TCP, UDP, ICMP) |
| Source IP Address |
| Source Port Number |
| Destination IP Address |
| Destination Port Number |

with vulnerable ports and this kind of vulnerable hosts exist more than other kinds in the Internet. We propose a method for evaluating threat that a port of host is posed in the Internet by those contagious worms.

Most of malicious packets monitored by Internet monitoring system are those from worms. We evaluate threat in the Internet based on access graph formed by source and destination of malicious packets.

Traditional threat detection system based on time-series frequencies of malicious packets. Figure 2 shows time-series frequencies of monitored packets for each port(Top five ports). The horizontal axis indicates time and the vertical axis indicates frequency of packets (access frequencies).

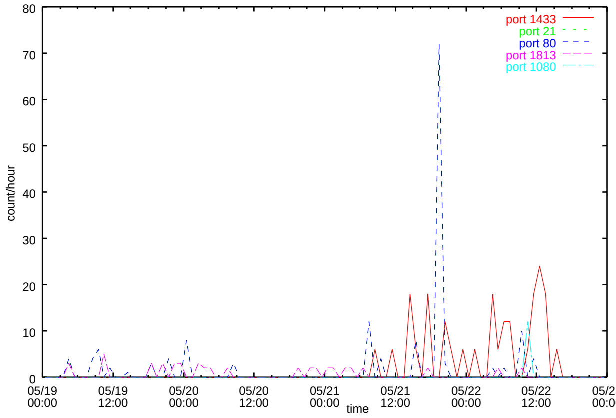

Figure 2: Time-series access frequencies by ports

Threat detection methods based on time-series frequencies of packets do not make use of spacial structure of access relations between source and destination of packets. Figure 3 shows an access graph formed by relation of source and destination of same data of packets. The left-hand side of the graph indicates source IP addresses which are renumbered for convenience and the right-hand side of the graph indicate destination port of the packets.

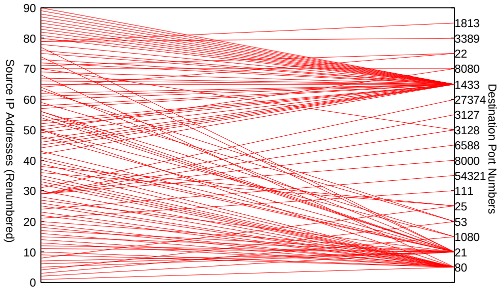

Figure 3: Access graph between sources and destinations

The data in this example was obtained during the period when SPIDA worm was active. As seen in the Figure 3, there are many access packets from many source addresses to ports 1433 (MS SQL), port 21 (ftp), port 80 (http). In order to evaluate threat based on this access graph, we consider two kinds of relationship: one is that the more vulnerable a port is, the more access packets received from highly contagious worms. The other is that the higher a contagious worm is, the more it accesses vulnerable ports[1]. These relationship can be restated as follows:

Threat relationship between source and destination of malicious packets:

**Relationship 1** Vulnerability of a destination port is high if it gets access from many different source address with high threat level.

**Relationship 2** Threat level of a source address is high if it sends more packets to vulnerable destination ports.

[1] We can assume that TCP access worms do not spoof IP address, because it has to create connection to that target host. Therefore, we use only TCP packets for the analysis

We show how to evaluate threats in the Internet based on these relationship by using simple examples. Figure 4 shows relationship between source and destination of monitored packets. Arrows from left to right indicates an existence of an access from a left node to a right node.

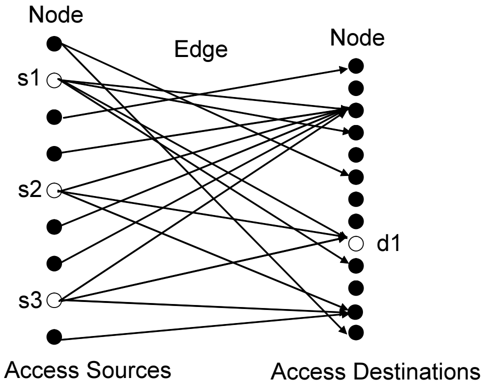

Figure 4: Relation between destination d1 and several sources

First, we define a vulnerability of a destination based on the relationship 1. We assume all source nodes are assigned tentative threat level. Vulnerability of the destination d1 in the figure is defined by a weighted sum of threat of source nodes connected by edges. Weight of edges is defined in Section 3.3.

Next in Figure 5, we define a threat level of a source based on the relationship 2. We assume destination nodes are assigned tentative vulnerability. Threat of a source node s4 in the figure is defined by a weighted sum of vulnerability of destination nodes connected by edges in the same way.

In the former relationship, threat level of source nodes are assumed to be given in order to define vulnerability of destination nodes. In the latter relationship, vulnerability of destination nodes are assumed to be given in order to define threat of source nodes. By starting arbitrary initial values of threats and vulnerability and applying above two relations interchangeably, convergent values indicate threats and vulnerabilities of source and destination nodes.

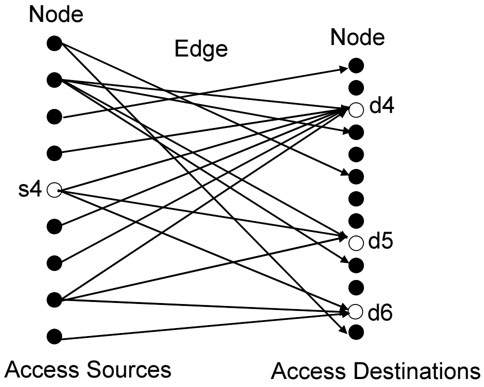

Figure 5: Relation between source s1 and several destinations

## 3.3 Threat Calculation

We apply eigen equation method to access graph we described in the previous section in order to evaluate threat in the Internet. Figure 6 shows access graph formed by relationship between source and destination of monitored packets. Source nodes represent IP addresses and destination nodes represent port numbers. Arrows represent access from source to destination of a monitored packet.

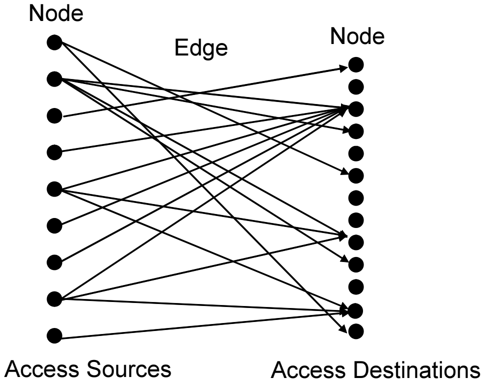

Figure 6: Graph of port accesses on the Internet

Monitored packets comes from outside the sensors to the sensors. Since nodes of source and destination does not overlap, the access graph is a bipartite graph.

We define a vector **t** to be a tuple of threat levels of source nodes $i$ and a vector $\mathbf{v}$ to be a tuple of threat levels of destination nodes as follows:

$$\begin{aligned}\mathbf{t} &= (t_1, t_2, \cdots, t_n) \qquad (1)\\ \mathbf{v} &= (v_1, v_2, \cdots, v_m) \qquad (2)\end{aligned}$$

We call $\mathbf{t}$ a source threat vector and $\mathbf{v}$ a destination threat vector.

First, threat level $v_j$ of destination $j$ is defined as a weighted sum of threat level $t_i$ of source $i$ , based on the relationship 1 in Section3 (Equation 3).

$$\begin{aligned}v_1 &= c_1(w_{1,1}t_1 + w_{2,1}t_2+, \cdots, w_{n,1}t_n) \quad (3)\\ &\cdots\\ v_m &= c_1(w_{1,m}t_1 + w_{2,m}t_2+, \cdots, w_{n,m}t_n)\end{aligned}$$

A coefficient $c_1$ is fixed by solving an eigen equation and described later. The weights are assigned to the edge connecting from source $i$ to destination $j$ depending on how much an access from source $i$ affects destination $j$. Since accesses from the different source suggest highly contiguous worm than repeated access from the same source, we define $w_{i,j}$ as follows: we consider two continuing observation terms, the former term and the latter term. If any access from source $i$ to destination $j$ exists in the latter term and no access in the former term, the weight is defined as 1. Otherwise the weight is define as 0.

Next, threat level $t_i$ of source $i$ is defined as a weighted sum of threat level $v_j$ of source $j$, based on the relationship 2 in Section 3 (Equation 4).

$$\begin{aligned}t_1 &= c_2(w_{1,1}v_1 + w_{1,2}v_2+, \cdots, w_{1,m}v_m) \quad (4)\\ &\cdots\\ t_n &= c_2(w_{n,1}v_1 + w_{n,2}v_2+, \cdots, w_{n,m}v_m)\end{aligned}$$

A coefficient $c_2$ is fixed by solving an eigen equation and described later.

Equation 3 defines relationship to calculate destination threat vector $\mathbf{v}$ from source threat vector $\mathbf{t}$. On the other hand, Equation 4 defines relationship to calculate source threat vector $\mathbf{t}$ from destination threat vector $\mathbf{v}$ in inverse way. Starting from an arbitrary initial vectors of $\mathbf{v}$ and $\mathbf{t}$ and applying the above two equations interchangeably, we can obtain convergent threat vector for $\mathbf{v}$ and $\mathbf{t}$.

These convergent vectors can be calculated by solving eigen equation. We define a access matrix composed of weights $w_{i,j}$ of graph edge from source $i$ to destination $j$ in Equation 5.

$$W = \begin{pmatrix} w_{1,1} & w_{1,2} & \cdots & w_{1,m}\\ w_{2,1} & w_{2,2} & \cdots & w_{2,m}\\ \vdots & & & \vdots\\ w_{n,1} & w_{n,2} & \cdots & w_{n,m}\end{pmatrix} \qquad (5)$$

Equation 3 and Equation 4 are defined by using the access matrix $W$ as follows:

$$\mathbf{v} = c_1 \underset{m\times n}{{}^tW}\,\mathbf{t} \qquad (6)$$

$$\mathbf{t} = c_2 \underset{n\times m}{W}\,\mathbf{v} \qquad (7)$$

,where the matrix ${}^tW$ is a transposed matrix of $W$. $m \times n$ under matrices indicate number of rows and columns.

By transforming above equation, we can obtain the following eigen value equations.

$$\mathbf{v} = c_1 c_2 \underset{m\times m}{{}^tWW}\mathbf{v} \qquad (8)$$

$$\mathbf{t} = c_1 c_2 \underset{n\times n}{W{}^tW}\mathbf{t} \qquad (9)$$

Equation eq:eigen1 shows that the destination threat vector $\mathbf{v}$ is an eigen vector of a square matrix $(\underset{m\times m}{{}^tWW})$ of size $m$ for an eigen value $\frac{1}{c_1c_2}$. Equation eq:eigen2 shows that the destination threat vector $\mathbf{t}$ is an eigen vector of a square matrix $(\underset{n\times n}{W{}^tW})$ of size $n$ for an eigen value $\frac{1}{c_1c_2}$.

According to the theorem of Perron-Frobenius, if every elements of $\underset{m\times m}{{}^tWW}$, $\underset{n\times n}{W{}^tW}$

are positive, all elements of a dominant eigen vector for the largest eigen value are positive. Therefore, in this case, source and destination threat vectors can be obtained uniquely.

In the Internet, since we can assume a very little random noise packets can be monitored at all IP addresses, we can add a small quantity $\delta(\ll 1)$ to all elements of an access matrix $W$. Therefore, all elements of eigen vectors obtained by solving the eigen equation 8 are positive.

# 4 Evaluation Experiments

We evaluate our method by applies working examples obtained by Internet threat monitoring system. Since it is difficult to tell threat in the Internet, we assume the period when critical warnings were issued to be in high threat.

## 4.1 MS SQL Incident

Target data for evaluation is obtained in the period where JPCERT/CC Alert JPCERT-AT-0006 was issued regarding MS SQL vulnerability on port 1433. This incident occurred during July 9th, 2005 to 13th.

We apply our method to these 5-days monitored data for 4 times as described in Figure 7. We use a pair of 1-day data every time: one day for the former period and other day for the latter period. By using 2-day data every time, we can calculate access matrix defined in Section 3.3.

Table 2 shows top ten list of ports' threat for each day. "port" column means port numbers. "count" column means number of access during a period. "threat" column means threat level evaluated by our method.

In the Table 2, threat level of the incident port (i.e. port 1433) increases 0.132, 0.130, 0.233, 0.331 from July 10 to 13 accordingly. The rank increases as 5th, 4th, 3rd, 2nd during this period. Figure 8 shows time-series change of threat level for top 5 ports.

On July 13th, threat level of port 1433 exceeds that of port 445, even if the access count is smaller than that of port 445. On the contrary, if we look at count columns, rank increases as 4th, 4th, 3rd, 3rd which is slow compared to our threat level. From these experiments, we can say that our method responds well to the critical incident compared with the access count in the period of incident outbreak,

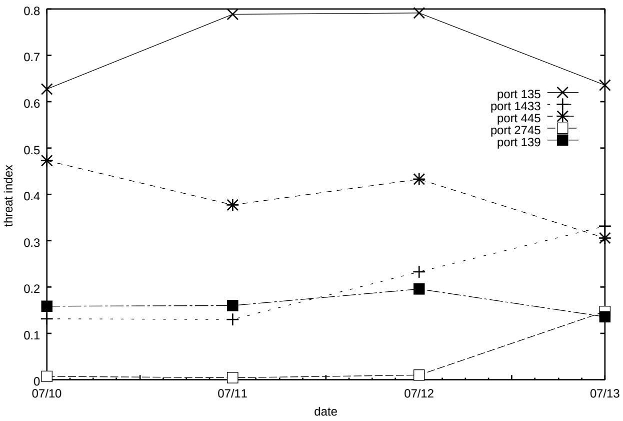

Figure 8: Time-series threat levels for the port 1433 incident

In Table 2, port 12345(Amitis.B backdoor) on July 10, port 9898 (Win32.Dabber.B worm) on July 12, port 2745 (Agobot bot worm that uses Bagle worm backdoor) on July 13 shows high threat level even if access count is small compared to other ports. This result cannot be derived by threat detection method based on access count.

## 4.2 Windows File Share Incident

The next data for experiment is those obtained in the period when IPA issued an alert on Window file share vulnerability on port 139. The period of this incident started from June 8, 2005 to June 12. In this experiment, we applied our method in the same way as the previous experiment in that we applied our method for each 2-day data.

Table 3 shows top 10 ports with highest threat levels. In this experiment, threat of the vulnerable port 139 increases as 0.029, 0.055, 0.081, 0.106 and ranks increases 20th, 33th, 4th, 3rd.

Figure 9 shows time-series threat level of top 5 ports. This experiment also shows relatively

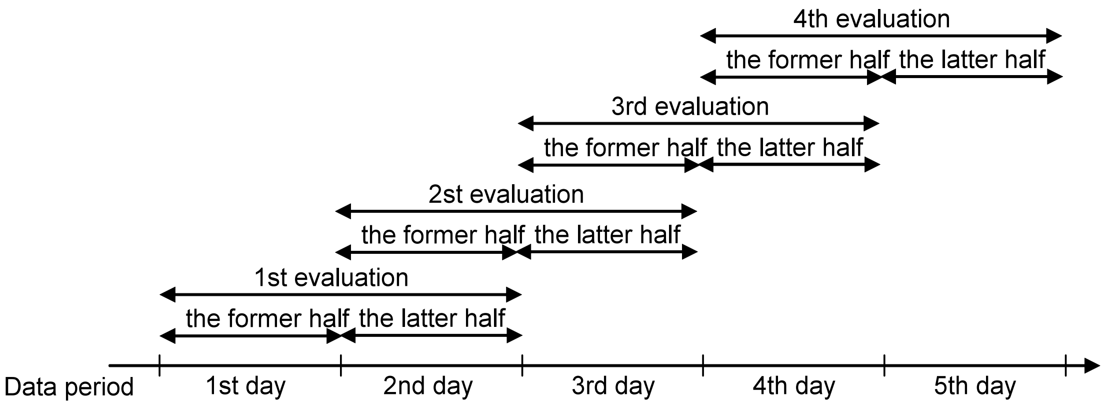

Figure 7: Data usage for experiment

Table 2: Top 10 of list by threat levels for the port 1433 incident

| July 10 | | | July 11 | | | July 12 | | | July 13 | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| port | count | threat | port | count | threat | port | count | threat | port | count | threat |
| 135 | 1031 | 0.627 | 135 | 1038 | 0.789 | 135 | 885 | 0.792 | 135 | 1057 | 0.636 |
| 445 | 1121 | 0.472 | 445 | 822 | 0.378 | 445 | 820 | 0.432 | 1433 | 346 | 0.331 |
| 12345 | 10 | 0.163 | 139 | 208 | 0.160 | 1433 | 222 | 0.233 | 445 | 739 | 0.305 |
| 139 | 232 | 0.159 | 1433 | 159 | 0.130 | 139 | 219 | 0.195 | 2745 | 6 | 0.148 |
| 1433 | 115 | 0.132 | 12345 | 13 | 0.109 | 9898 | 7 | 0.089 | 139 | 204 | 0.135 |
| 3410 | 8 | 0.123 | 901 | 14 | 0.109 | 1024 | 2 | 0.085 | 2100 | 3 | 0.111 |
| 901 | 9 | 0.123 | 3410 | 11 | 0.087 | 4899 | 64 | 0.078 | 8080 | 3 | 0.111 |
| 22 | 12 | 0.112 | 3389 | 6 | 0.087 | 3306 | 19 | 0.064 | 8535 | 3 | 0.111 |
| 3090 | 7 | 0.112 | 3306 | 18 | 0.087 | 2100 | 1 | 0.064 | 25 | 6 | 0.111 |

Table 3: Top 10 of list by threat levels for the port 139 incident

| June 9 | | | June 10 | | | June 11 | | | June 12 | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| port | count | threat | port | count | threat | port | count | threat | port | count | threat |
| 135 | 2551 | 0.954 | 135 | 2174 | 0.883 | 135 | 2834 | 0.879 | 135 | 1906 | 0.846 |
| 445 | 751 | 0.209 | 445 | 1008 | 0.227 | 445 | 1308 | 0.244 | 445 | 989 | 0.249 |
| 1433 | 140 | 0.078 | 1080 | 4 | 0.104 | 12345 | 11 | 0.085 | 139 | 242 | 0.106 |
| 4899 | 43 | 0.052 | 44599 | 8 | 0.099 | 139 | 257 | 0.081 | 42857 | 2 | 0.102 |
| 1521 | 1 | 0.052 | 10589 | 4 | 0.099 | 21 | 4 | 0.077 | 4899 | 46 | 0.076 |
| 8535 | 1 | 0.052 | 8080 | 2 | 0.070 | 1433 | 142 | 0.065 | 143 | 1 | 0.076 |
| 8536 | 1 | 0.052 | 4899 | 47 | 0.070 | 44599 | 3 | 0.064 | 3306 | 9 | 0.076 |
| 2100 | 3 | 0.052 | 22 | 23 | 0.070 | 10589 | 3 | 0.064 | 1256 | 3 | 0.076 |
| 22 | 10 | 0.052 | 25 | 10 | 0.070 | 11524 | 2 | 0.064 | 2419 | 1 | 0.076 |
| 143 | 1 | 0.052 | 3306 | 4 | 0.070 | 42857 | 2 | 0.064 | 6346 | 3 | 0.076 |

high increase of threat of vulnerable port compared with other ports.

## 5 Summary

We proposed a new threat evaluation method based on access graph formed by relationship

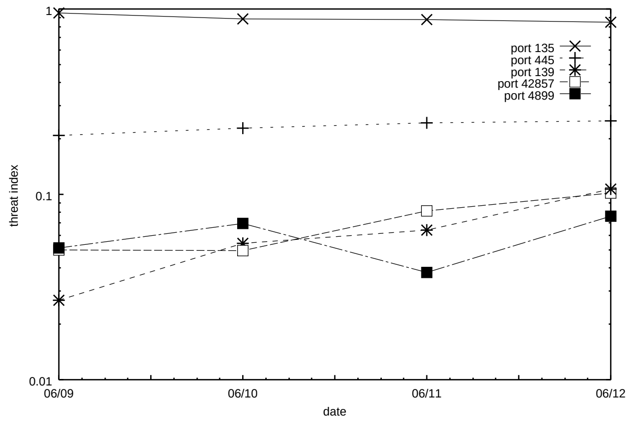

Figure 9: Time-series threat levels for the port 139 incident

between source and destination of monitored malicious packets. Traditional threat detection methods are based on time-series frequencies packets. Our method is different from traditional method in that it make use of spacial structure of access graph to evaluate threat in the Internet.

We apply eigen vector method to evaluate threat in the Internet. Our improvement enables to solve eigen equation of smaller size matrix.

By applying our method to the working example data obtained from the Internet monitoring system, threat level calculated by our method respond better to critical incident compared with frequencies of packets. As a future work, characteristics of our method to several type of incident should be clarified.

## References

[DSh] DShield.org. Distributed intrusion detection system. http://www.-dshield.org/index.html.

[ISMO04] Masaki Ishiguro, Hironobu Suzuki, Ichiro Murase, and Hiroyuki Ohno. Internet threat detection system using bayesian estimation. In *Proceedings of 16th Annual FIRST Conference on Computer Security Incident Handling*, 2004.

[JPC] JPCERT/CC. internet scan data acquisition system (isdas). http://www.jpcert.or.jp/isdas/.

[KSV04] Ramana Rao Kompella, Sumeet Singh, and George Varghese. On scalable attack detection in the network. In *4th ACM SIGCOMM conference on Internet measurement*, pages 187 – 200, 2004.

[MSVS04] David Moore, Colleen Shannon, Geoffrey M. Voelker, and Stefan Savage. Network telescopes: Technical report. Technical report, CAIDA, 2004. http://www.caida.org/-outreach/papers/2004/tr-2004-04/.

[Pro] The Honeynet Project. Tools for honeynets. http://www.lucidic.net/.

[PYB$^{+}$04] R. Pang, V. Yegneswaran, P. Barford, V. Paxson, and L. Peterson. Characteristics of internet background radiation. In *Proceedings of ACM Internet Measurement Conference*, 2004.

[SAN] SANS Institute. Internet storm center. http://isc.sans.org/.

[SJB04] Stuart Schechter, Jaeyeon Jung, and Arthur W. Berger. Fast detection of scanning worm infections. In *7th International Symposium on Recent Advances in Intrusion*, 2004.

[TJ03] Marina Thottan and Chuanyi Ji. Anomaly detection in ip networks. *IEEE TRANSACTIONS ON SIGNAL PROCESSING*, 51(8), August 2003.

[Uni] University of Michigan. Internet motion sensor (ims). http://ims.eecs.umich.-edu/index.html.

[ZGGT03] Cliff Changchun Zou, Lixin Gao, Weibo Gong, and Don Towsley. Monitoring and early warning for internet worms. In *the 10th ACM conference on Computer and communications security*, pages 190 – 199, 2003.