Identification of Attack Nodes from Traffic Matrix Estimation

Yuichi Ohsita
Graduate School of
Information Science and Technology,
Osaka University
1-5 Yamadaoka, Suita,
Osaka 565-0871, Japan
Phone: +81-6-6879-4542
E-mail: y-ohsita@ist.osaka-u.ac.jp

Shingo Ata
Graduate School of Engineering,
Osaka City University
3-3-138 Sugimoto, Sumiyoshi-ku,
Osaka 558-8585, Japan
Phone: +81-6-6605-2191
E-mail: ata@info.eng.osaka-cu.ac.jp

Masayuki Murata
Graduate School of
Information Science and Technology,
Osaka University
1-5 Yamadaoka, Suita,
Osaka 565-0871, Japan
Phone: +81-6-6879-4540
E-mail: murata@ist.osaka-u.ac.jp

***Abstract*— Distributed denial-of-service attacks on public servers have recently become more serious. The most effective way to prevent this type of traffic is to identify the attack nodes and detach (or block) attack nodes at egress routers of them. However, existing traceback mechanisms are currently not widely used for some reasons, such as the necessity of replacement of many routers to support traceback capability, or difficulties in distinguishing between attacks and legitimate traffic. In this paper, we propose a new scheme that enables a traceback from a victim to the attack nodes. More specifically, we identify the egress routers that attack nodes are connecting to by estimating the traffic matrix between arbitral source-destination edge pairs. We identify the edge routers that are forwarding the attack traffic, which have a sharp traffic increase to the victim, by monitoring the traffic variations obtained by the traffic matrix, we identify the edge routers forwarding attack traffic which have a sharp traffic increase to the victim. We also evaluate the effectiveness of our proposed scheme through simulation, and show that our method can identify attack sources accurately.**

***Index Terms*— Distributed Denial of Service (DDoS), Traceback, Traffic matrix, Simple Network Management Protocol (SNMP)**

I. Introduction

The recent rapid growth and the increasing utility of the Internet are making Internet security issues increasingly important. Denial-of-service (DoS) attacks are one of the most serious problems and must be resolved as soon as possible. These attacks prevent users from communicating with service providers and have damaged many major web sites all over the world.

The number of attacks has been increasing, and the techniques used to attack servers have become more complex. In the distributed denial-of-service (DDoS) attacks often seen recently, multiple distributed nodes attack a single server concurrently. A malicious user tries to hack the remote nodes by exploiting the vulnerabilities of the software running on them, installs an attack program on the hijacked nodes, and keeps them waiting for an order to attack a victim server. When the malicious user sends a signal to them, they begin to attack the same server. Even if the rate of attack for each node is small, the attack traffic can cause serious damage to the victim server when the number of hijacked nodes is large.

If we can identify the attack sources, we can effectively cut off the link to the attacker or filter attack packets by an edge router connected to the attacker. However, because attackers can easily spoof the source address fields of the attack packets, we cannot identify the attack sources by only using the source address of the attack packets.

For this reason, several methods for identifying the attack sources are proposed. In general, these methods for identifying the sources of the packets are called *IP tracebacks*. One of them is proposed in [1], [2], which uses ICMP packets. In this method, when a router forwards a packet, the router generates an ICMP traceback packet to the destination of the packet with a low probability. The victim can identify the source of the packet by using the received ICMP traceback packets. With the method, described in [3], [4], [5], a router marks forwarded IP packets with identification information instead of generating ICMP packets. The victim can identify the source of the packets using the identification information.

With another method, proposed in [6], [7], each router stores packet digests. The victim queries its upstream routers to see whether an attack packet has passed through them or not.

However, these methods have several problems. One is that they cannot work with legacy routers because they need router support. Another is that they may erroneously identify legitimate clients as attack sources. This is because these methods can only identify the source nodes of attack packets. Since there is no difference between legitimate and attack packets, identifying attack packets from the mixture of attack and legitimate traffic is difficult..

In DoS attacks, attackers send a large number of packets to exhaust the network resources. That is, when an attack starts, there is a rapid increase in the traffic from the attack sources to the victim. Therefore, we can identify the attack sources that are increasing the traffic to the victim by monitoring the traffic in the network. Identification of the attack sources by monitoring the increased traffic can distinguish the attackers from the legitimate clients, which do not sharply increase traffic. Lakhina et al propose a method for identifying the

attack sources by monitoring the traffic on each link in the network [8] . In this method, the measured traffic is separated into normal and abnormal subspaces. The normal subspace indicates the time-of-day variation of the traffic. Other variations are categorized into the abnormal subspace. Since we cannot clearly understand which traffic between the two edge nodes directly affects the abnormal subspace from measurements of the network links, we test the influence to the abnormal subspace by removing each traffic between the two edge nodes. We then identify the attack source that explains the largest amount of anomalous subspace. Although this method can identify the attack source in a single attacker case, this method has difficulty in identifying attack sources for multi source attacks like DDoS, because we need to test all cases, including changing the number of attackers. It requires a huge computation overhead.

If we can use not only the traffic data on each link, but also the traffic data between the source and destination, we can accurately identify the attack sources, even in multi-source attacks.

We propose a new method for identifying attack sources using the increase in traffic between each source and destination. The traffic transmitted between every pair of ingress and egress points is typically described as a traffic matrix. However, directly monitoring a traffic matrix is difficult because all edge routers need to hold the flow statistics of all pairs of sources and destinations. We estimate the increase in traffic between each source and destination. In our method, we modify the traffic matrix estimation method proposed by Zhang et al [9] to enable the estimation of the increases in traffic. Our method can work with existing routers because we can obtain link load data through Simple Network Management Protocol (SNMP).

In Section II, we explain an overview of our proposed method. In Section III we evaluate our method. In Section IV we conclude by briefly summarizing the paper and mentioning some of the future works we intend to do.

II. Identification of Attack Sources by Estimating Traffic Matrix

Our method identifies attack sources by estimating the increases in traffic between every pair of sources and destinations. We estimate the increases in traffic from the monitored link load. In the estimation of the traffic matrix, we don't focus on the total amount of traffic, but only focus on the amount of increase from the previous measurement. The reason why we use only the increases in traffic for the traffic estimation is discussed in the next subsection. In this section, we first show a brief overview of our proposed scheme.

Figure 1 shows an overview of our proposed method. In our method, we introduce a control node to perform the process of identification of attack sources. We call this node a *monitoring node*, and we also define the region where the monitoring node controls as a *monitored network*. The monitoring node identifies the attack sources by periodically performing the following operations.

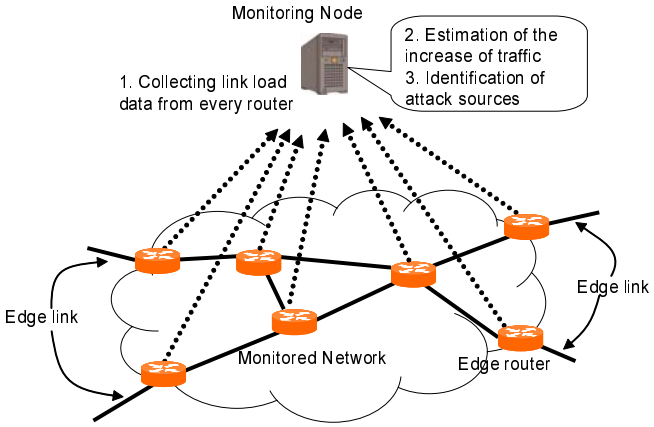

Fig. 1. Overview of proposed method

1) Obtains the statistics of the link load data from all routers in the monitored network.
2) Estimates a matrix of the increase in traffic between all arbitrary pairs of edge routers in the monitored network.
3) Identifies the attack sources from the estimated increase traffic matrix.

We can obtain link load data through SNMP. SNMP is supported by essentially every device in IP networks and is used to monitor or manage the device. That is, our method can work with existing routers.

The interval for obtaining the statistics affects the time for identifying the attack sources. If we set the interval to a larger value, the identification takes more time. On the other hand, if we set the interval to a smaller value, the loads on the routers increase though we can identify attack the sources soon after the attack starts. Thus, we should properly set this interval.

In the following sections, we describe the details about how to estimate the increase in traffic and how to identify the attack sources.

A. Increase in Traffic Estimation

1) Traffic Matrix Estimation using Gravity Model: First, we assign a set of links outside the monitored network as E. We call these links *edge links*. The router, which is connected by an edge link, is called the *edge router*. We assign a set of all the links in the monitored network, including the edge links, as L.

Traffic matrix T is defined as the $|E| \times |E|$ sized matrix, whose element $t_{i,j}$ $(i, j \in E)$ indicates the amount of traffic traversing from edge link i to edge link j. We can obtain the link loads from each router through SNMP. The link loads can be denoted by the $2|L|$-size link load matrix X

as follows:

$$X = \begin{bmatrix} x_1^{\mathrm{f}} \\ x_1^{\mathrm{b}} \\ x_2^{\mathrm{f}} \\ x_2^{\mathrm{b}} \\ \vdots \\ x_{|L|}^{\mathrm{f}} \\ x_{|L|}^{\mathrm{b}} \end{bmatrix}. \quad (1)$$

In matrix X, elements $x_l^{\mathrm{f}} \quad (l \in L)$ and $x_l^{\mathrm{b}} \quad (l \in L)$ indicate the amount of traffic on link l in the forward and backward directions respectively, because most of the network links are bidirectional. We only use the words forward/backward to distinguish the direction of the link. Therefore, there is no policy for determining the forward or backward direction of each link. However, we must distinguish between the ingress and egress traffic. To distinguish between them, we denote the ingress traffic measured on edge link i as $x_i^{\mathrm{in}} \quad (i \in E)$ and egress traffic measured on the edge link j as $x_j^{\mathrm{out}} \quad (j \in E)$.

We estimate the traffic matrix of each pair of edge links from the link loads and routing information in monitored network. [9] uses a gravity model to estimate the traffic matrix. The gravity model assumes that traffic from a source to a destination is proportional to the total traffic at the source and at the destination. Using this model, we can estimate the traffic matrix from

$$t_{i,j} = x_i^{\mathrm{in}} \frac{x_j^{\mathrm{out}}}{\sum_{\forall k \in E} x_k^{\mathrm{out}}} \quad (i,j \in E) \ , \quad (2)$$

where x_i^{in} is the element of X corresponding to the amount of ingress traffic to the monitored network measured on the edge link i and x_j^{out} is the egress traffic measured on the edge link j.

However, we cannot accurately estimate increases in traffic accurately using Eq. (2) as follows. We assume that an attack traffic whose rate is t_{attack} traverses from i to j. We also assume legitimate traffic $t_{i,j}$ can be accurately estimated by Eq. (2). Traffic from i to j, including the attack traffic is estimated from

$$t'_{i,j} = (x_i^{\mathrm{in}} + t_{\mathrm{attack}}) \frac{x_j^{\mathrm{out}} + t_{\mathrm{attack}}}{\sum_k x_k^{\mathrm{out}} + t_{\mathrm{attack}}}, \quad (3)$$

where $t'_{i,j}$ is the traffic traversing from i to j including attack traffic. Then, the increased traffic by the attack is estimated by

$$t'_{i,j} - t_{i,j} = \frac{t_{\mathrm{attack}}^2 + t_{\mathrm{attack}}(x_i^{\mathrm{in}} + x_j^{\mathrm{out}})}{\sum_{k \in E} x_k^{\mathrm{out}} + t_{\mathrm{attack}}}, \quad (4)$$

where $t_{i,j}$ is the legitimate traffic from i to j. For example, we assume the total rate of traffic in the monitored network is 20 GBytes/sec, both x_i^{in} and x_j^{out} are 2 GBytes/sec. We also assume the attack traffic from the edge link i to j has the rate of 1 GBytes/sec. From Eq. (4), the total traffic, including the attack traffic from edge link i to j is estimated as 0.23 GBytes/sec, which is quite different from the attack rate (1 GBytes/sec).

As previously mentioned, when attack traffic is injected, the estimated increase in traffic is proportional to the total rate of traffic monitored at the source. That is, the gravity model is infeasible for directly estimating the attack traffic because the impact of the attack traffic is distributed among the edge links that have legitimate traffic to the victim. As a result, the estimated attack rate is significantly lower than the rate of the attack traffic that is really generated.

2) Traffic matrix estimation focusing on increased traffic: To accurately estimate the increase in traffic, we propose a matrix estimation method focusing not on the total rate of traffic but on the increase in traffic.

First, we calculate the increases in traffic on each link from

$$G_n = X_n - \bar{X}_n, \quad (5)$$

where G_n is the $2|L|$-size vector in which the elements $g_{i,n}^{\mathrm{f}} \quad (i \in L)$ and $g_{i,n}^{\mathrm{b}} \quad (i \in L)$ indicate the increase in traffic on link i in the forward and backward directions at time n, respectively. X_n is the link load vector at time n and $\bar{X}_n$ is the $2|L|$-size vector in which $\bar{x}_{i,n}^{\mathrm{f}}$ is the average rate of legitimate traffic on the link i in the forward direction before time n and $\bar{x}_{i,n}^{\mathrm{b}}$ is one on the same link in the backward direction. We explain how to calculate $\bar{X}_n$ in Subsection II-A.4.

Then, by using G_n, we estimate the increases in traffic between every pair of sources and destinations. The increase in traffic can be shown as a $|E| \times |E|$ matrix F_n whose element $f_{i,j,n} \quad (i,j \in E)$ indicates the increase in traffic traversing from edge link i to edge link j.

Eq. (2) cannot be used to estimate the traffic increase matrix from G_n, which may include negative values, because it supports only positive values. Therefore, we modify Eq. (2) to support negative values. We define the traffic increase matrix F_n, having the traffic increase $f_{i,j,n}$, from edge link i to j between the time $n-1$ and n. The value of $f_{i,jn}$ is calculated from

$$f_{i,j,n} = \begin{cases} g_{i,n}^{\mathrm{in}} \frac{g_{j,n}^{\mathrm{out}}}{\sum_{\{k:(g_{k,n}^{\mathrm{out}} > 0)\}} g_{k,n}^{\mathrm{out}}} & (g_{i,n}^{\mathrm{in}} > 0, g_{j,n}^{\mathrm{out}} > 0) \\ -\left| g_{i,n}^{\mathrm{in}} \frac{g_{j,n}^{\mathrm{out}}}{\sum_{\{k:(g_{k,n}^{\mathrm{out}} < 0)\}} g_{k,n}^{\mathrm{out}}} \right| & (g_{i,n}^{\mathrm{in}} < 0, g_{j,n}^{\mathrm{out}} < 0) \\ 0 & (\mathrm{others}). \end{cases} \quad (6)$$

Focusing on the increase in the traffic, we can eliminate the effect of the amount of legitimate traffic and estimate the increase in the traffic more accurately. That is, we can estimate that the increase in traffic from attack sources to the victim is large by checking the increase in traffic when the attack starts. If the monitored network suffers from multiple attacks whose sources and victims are different, some traffic from different sources to different destinations concurrently increases. In this case, the estimated increase in traffic is proportional to the increase in traffic measured at the sources. That is, traffic from a source of an attack to a victim of another attack is estimated as increased. However, we can identify the attack sources that generate the attack traffic, even if we could not identify the

victim node exactly where the attack source sends the attack traffic to.

3) Modification of traffic matrix: Although F_n is a $|E|\times|E|$ matrix, F_n can be denoted as following the $|E|^2$-size vector;

$$F_n = \begin{bmatrix} f_{1,1,n} \\ f_{1,2,n} \\ \vdots \\ f_{1,|E|,n} \\ f_{2,1,n} \\ \vdots \\ f_{|E|,|E|,n} \end{bmatrix} \tag{7}$$

Due to the fact that the total amount of traffic on the link is the summation of the traffic of flows that are passing the link, F_n and G_n satisfy

$$G_n = AF_n, \tag{8}$$

where A is a $2|L| \times |E|^2$ routing matrix which is given by

$$A = \begin{bmatrix} a^{\mathrm{f}}_{1,1,1} & a^{\mathrm{f}}_{1,2,1} & \cdots & a^{\mathrm{f}}_{|E|,|E|-1,1} & a^{\mathrm{f}}_{|E|,|E|,1} \\ a^{\mathrm{b}}_{1,1,1} & a^{\mathrm{b}}_{1,2,1} & \cdots & a^{\mathrm{b}}_{|E|,|E|-1,1} & a^{\mathrm{b}}_{|E|,|E|,1} \\ a^{\mathrm{f}}_{1,1,2} & a^{\mathrm{f}}_{1,2,2} & \cdots & a^{\mathrm{f}}_{|E|,|E|-1,2} & a^{\mathrm{f}}_{|E|,|E|,2} \\ a^{\mathrm{b}}_{1,1,2} & a^{\mathrm{b}}_{1,2,2} & \cdots & a^{\mathrm{b}}_{|E|,|E|-1,2} & a^{\mathrm{b}}_{|E|,|E|,2} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ a^{\mathrm{f}}_{1,1,|L|} & a^{\mathrm{f}}_{1,2,|L|} & \cdots & a^{\mathrm{f}}_{|E|,|E|-1,|L|} & a^{\mathrm{f}}_{|E|,|E|,|L|} \\ a^{\mathrm{b}}_{1,1,|L|} & a^{\mathrm{b}}_{1,2,|L|} & \cdots & a^{\mathrm{b}}_{|E|,|E|-1,|L|} & a^{\mathrm{b}}_{|E|,|E|,|L|} \end{bmatrix}. \tag{9}$$

$a^{\mathrm{f}}_{i,j,k} \quad (i,j \in E, k \in L)$ is equal to 1 if the traffic from edge link i to edge link j traverses on link k in the forward direction, or set to zero otherwise. In a similar way, $a^{\mathrm{b}}_{i,j,k} \quad (i,j \in E, k \in L)$ is equal to 1 if the traffic from edge link i toedge link j traverses on link k in the backward direction or zero otherwise. Matrix A can be obtained by monitoring the routing messages, such as the Link State Advertisement (LSA) of OSPF [10] or by simulating routing [11].

The traffic matrix estimated by the gravity model cannot satisfy Eq. (8) because Eq. (6) does not use the traffic statistics on the internal links of the monitored network, but uses only the traffic measurements of the edge links. Therefore, we adjust the traffic matrix estimated by the gravity model to satisfy Eq. (8). We can obtain the final estimation results for F_n from

$$F_n = F'_n + pinv(A)(G_n - AF'_n), \tag{10}$$

where F'_n is the $|E|^2$-size vector indicating the results estimated by Eq. (6), and $pinv(A)$ is a pseudo inverse of A. $pinv(A)$ can be obtained by using a function of Scilab [12].

4) How to estimate average of legitimate traffic: Our method for estimating the increase in traffic uses the average rate of legitimate traffic. The rate of legitimate traffic varies according to the time of day. To follow the time-of-day variation of this traffic, we assume that the average rate of legitimate traffic $\bar{X}_n$ is basically estimated by the weighted average of the monitored traffic rate from

$$\bar{X}_{n+1} = \alpha X_n + (1-\alpha)\bar{X}_n \quad (0 < \alpha < 1) \ . \tag{11}$$

However, when the traffic suddenly and rapidly increases suddenly and rapidly (we call these *spikes* throughout the rest of this paper), $\bar{X}_n$ becomes large after the spike. The large $\bar{X}_n$ value causes difficulties in the identification of the increase in traffic after the spike, because the larger $\bar{X}_n$ value makes the impact of $(X_n - \bar{X}_n)$ small, even for cases of increases in traffic. For this reason, we must estimate the average of the legitimate traffic without the effect of spikes.

We can eliminate the effect of spikes by updating only the elements of $\bar{X}_n$ corresponding to the link on which the increase in traffic is under a threshold. However, as described in the previous subsection, our method assumes the situation covered by Eq. (8). For this reason, we should update $\bar{X}_n$ by satisfying Eq. (8).

For this purpose, we update $\bar{X}_n$ using an element from estimated F_n, which is not rapidly increasing. First, we extract the element not increasing rapidly from F_n. We denote the $|E|\times|E|$ matrix of the extracted elements as $\hat{F}_n$. Each element $\hat{f}_{i,j,n} \quad (i,j \in E)$ is defined by

$$\hat{f}_{i,j,n} = \begin{cases} f_{i,j,n} & (f_{i,j,n} < \mu_{i,j} + \beta\sigma_{i,j}) \\ 0 & (\text{others}) \end{cases} . \tag{12}$$

where $\mu_{i,j}$ is the average of the last J values of $f_{i,j,k} \quad (i,j \in E, n-J < k \le n)$ and $\sigma_{i,j}$ is the variance of the last J values of $f_{i,j,k} \quad (i,j \in E, n-J < k \le n)$. β is the parameter by which we can set the threshold. By Eq. (12), when the traffic from i to j sharply increases at time n beyond the threshold, $\hat{f}_{i,j,n}$ is zero, while in other cases, $\hat{f}_{i,j,n}$ is $f_{i,j,n}$.

After that, we update $\bar{X}_{n+1}$ with the following equation.

$$\bar{X}_{n+1} = \alpha(\bar{X}_n + A\hat{F}_n) + (1-\alpha)\bar{X}_n \tag{13}$$

In Eq. (13), we calculate the increase in traffic on each link from $\hat{F}_n$ by $A\hat{F}_n$. Using the increase in traffic, we calculate the amount of traffic at time n as $\bar{X}_n + A\hat{F}_n$. Then, we update $\bar{X}_{n+1}$ as the weighted average of the monitored traffic using the amount of traffic at time n.

With the above stated equations, we can update $\bar{X}_{n+1}$ without the effect of any spikes in F_n. By deciding whether each element of F_n should be used to update, we can satisfy Eq. (8).

B. Identification of attack sources

When an attack starts, the traffic sharply increases from the attackers to the victim. Moreover, the larger the increase is, the more serious the impact on the network resources is. We identify the sources increasing the traffic on the victim as attack sources. However, when many attack sources are widely distributed, the impact of the attack is serious, even if each attack source generates a small rate of attack traffic. Thus, the identification of the attack sources, by setting a static threshold to the increase in traffic, is not sufficient. Instead of setting a threshold, we identify the attack sources by comparing the increase in traffic from each edge link to the victim. When the victim detects an attack, it is reasonable enough to assume that the source generating more traffic to

the victim has more likelihood of being considered an attack source. With this assumption, we identify attack sources from the nodes generating a lot of traffic to the victim node. We also use the total rate of traffic to detect the event of an attack. By using the total rate of attack traffic, we can identify the attack sources even in cases of DDoS. The total rate of attack traffic can be estimated from the increase of the egress traffic to the victim.

When an attack starts, the egress traffic increases with the rate of the attack traffic. However, the rate of legitimate traffic may also change according to the time-of-day. Assuming the increase of egress traffic to the victim is attack traffic may be an overestimation of the attack traffic, because an increase in egress traffic includes both legitimate and attack traffic. As a result of this overestimation, the source node sending only legitimate traffic may be mislead as an attack source. For this reason, we estimate the rate of the attack traffic $\tilde{g}^{\mathrm{out}}$ from results of traffic estimation. When an attack to edge link j starts at the time n, $\tilde{g}^{\mathrm{out}}$ is estimated from

$$\tilde{g}^{\mathrm{out}} = g^{\mathrm{out}}_{j,n} - \mu^{\mathrm{out}}_{j} - \gamma, \tag{14}$$

where $g^{\mathrm{out}}_{j,n}$ is the egress traffic on edge link j to the outside of the monitored network, μ^{out}_{j} is the average of the last J values of $g^{\mathrm{out}}_{j,k}$ $(n - J \leq k < n)$, and γ is the parameter indicating the variation in the rate of the legitimate traffic. In this equation, μ^{out}_{i} represents the effect of the time-of-day variation of the legitimate traffic and γ mitigates the effect of the other variations of the legitimate traffic. By adequately setting γ, we can estimate $\tilde{g}^{\mathrm{out}}$ as the value which may be a little smaller than the actual attack rate, but is never larger than the actual attack rate.

Then, we identify source i as attack source when source i satisfies

$$\sum_{(k:f_{k,j,n} > f_{i,j,n})} f_{k,j,n} \leq \tilde{g}^{\mathrm{out}}, \tag{15}$$

where $f_{i,j,n}$ is the element of the estimated increase traffic matrix F_n corresponding to the traffic from edge link i to victim edge link j. Before using Eq. (15), we must first sort out the set of $f_{k,j,n}(1 \leq k \leq N)$ by descending order based on their values. We then calculate the total of the top m traffic to the victim node. We compare the total top m traffic with the estimated egress traffic g^{out}. We increment m by one and calculate the total top m traffic until the total traffic exceeds g^{out}. Finally, we identify these m nodes as the attack sources.

Let us denote the actual rate of attack traffic as t^{attack} and that the sum of the top m increases of the egress traffic to the victim as $t^{\mathrm{top}(m)}$. If $t^{\mathrm{top}(m)}$ is smaller than $\tilde{g}^{\mathrm{out}}$ and $t^{\mathrm{top}(m+1)}$ is larger than $\tilde{g}^{\mathrm{out}}$, then we can identify $m+1$ attack sources. In this case, the total rate of attack traffic from the identified attack sources is $t^{\mathrm{top}(m+1)}$, which is larger than g^{out}. That is, the rate of the attack traffic from the unidentified attack sources is at most $t^{\mathrm{attack}} - \tilde{g}^{\mathrm{out}}$, which is calculated from $\gamma + \mu - f^{\mathrm{normal}}$ where f^{normal} is the increase in legitimate traffic. Therefore, by adequately setting γ adequately, we can identify most of the attack sources and limit the rate of attack traffic from the unidentified attack sources.

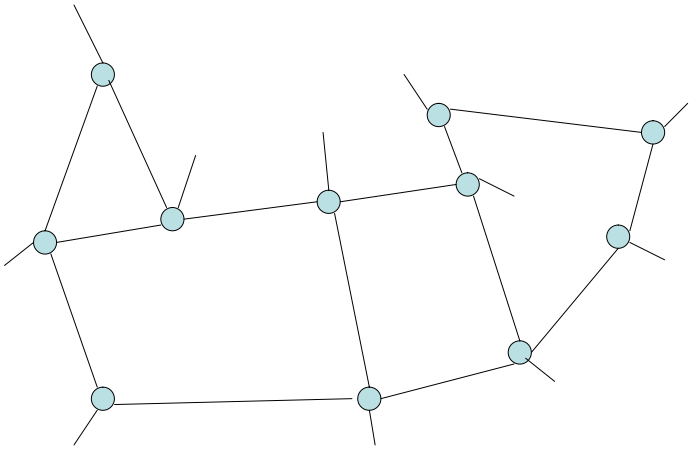

Fig. 2. Backbone Topology of the Abilene

III. Evaluation

We evaluated our method using simulations. In all our simulations, we used the topology shown in Figure 2 for the monitored network. We used statistics of traffic monitoring from the gateway of Osaka University for the legitimate traffic in the simulation. More specifically, we captured all of the IP headers that passed the gateway of Osaka University. We then made a group of packets based on a 16 bit prefix of the source address, so that the number of the kinds of 16 bit prefixes in each group was equal. We then calculated the aggregated traffic rate for each group with a 60 seconds interval. The topology, shown in Figure 2, has 11 edge nodes. There are $11 \times 10 = 110$ source-destination pairs. For each pair, we assigned the above-mentioned aggregated traffic rate as the legitimate traffic. In our simulations, we set α to 0.1 and β to 3, which allows a time-of-day variation of the traffic.

A. Accuracy in estimating the increase of traffic

First, we validated that our method can accurately estimate the increase in traffic. Figure 3 shows the time-dependent variation of the arrival rate of each packet between a source and a destination. Figure 4 compares the actual time-dependent variation of the increase in arrival traffic with its estimated rate. Comparing Figures 3 and 4, we can see that by monitoring the increase in traffic, we can eliminate the time-of-day variation of the traffic. That is, by monitoring the increase in traffic, we can identify the attack sources without the affect of a time-of-day variation in the traffic. From Figure 4, we also see that in the cases where a rapid increase in traffic occurs, our method can accurately estimate it.

We performed another simulation to evaluate accuracy when attacks from several sources start. We injected attack traffic from four sources to a single destination. Figures 5 and 6compare the results of the estimations with actual values. The horizontal axis is the actual rate of traffic and the vertical axis is the estimated value. In Figure 5, the attack rate from each source is 200 packets/sec. In Figure 6, the attack rates from the four sources are 1000 packets/sec, 830 packets/sec,

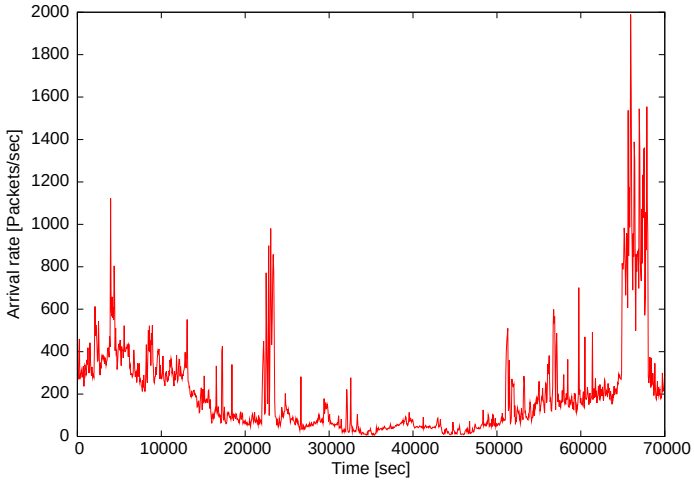

Fig. 3. Time-dependent variation of arrival rate of packets

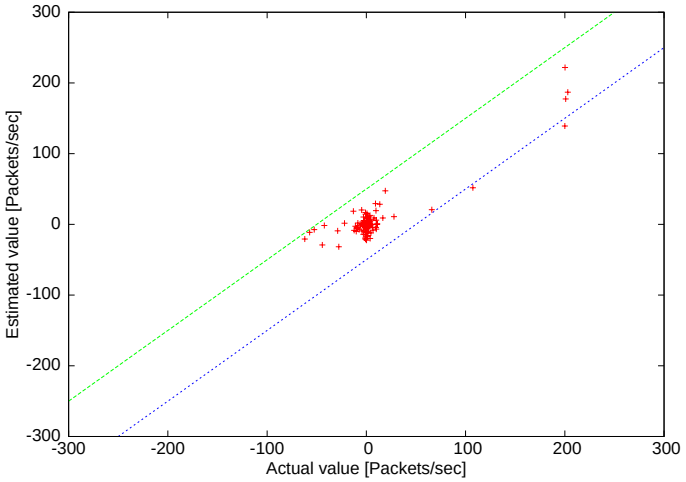

Fig. 5. Estimated value vs. Actual value (200 packets/sec attack injected)

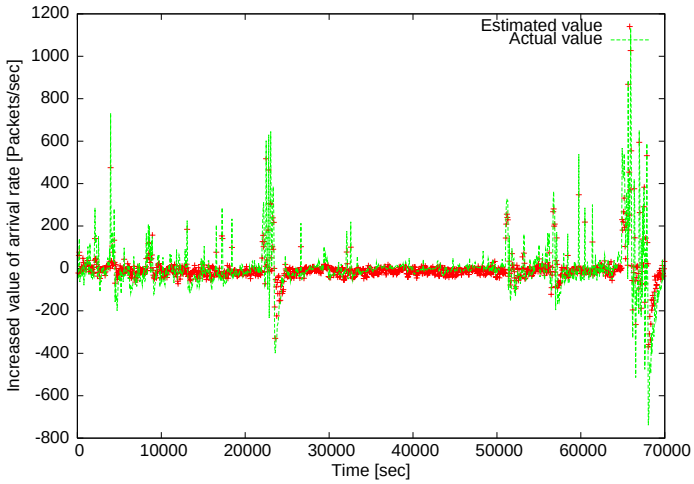

Fig. 4. Time-dependent variation of increase of arrival rate of packets between source and destination.

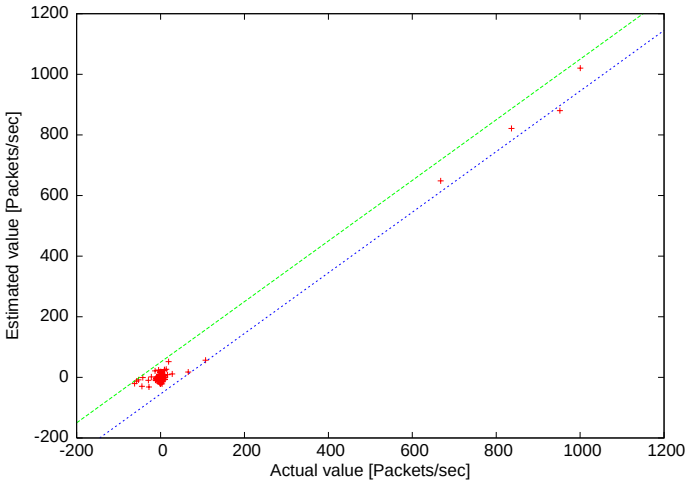

Fig. 6. Estimated value vs. Actual value (1000 packets/sec, 830 packets/sec, 660 packets/sec and 500 packets/sec attacks injected)

660 packets/sec and 500 packets/sec. The lines in both figures show $x \pm 50$. These figures show we can accurately estimate the increase in traffic. Even for large attacks, we can estimate the increase in traffic with an error rate of less than 50 packets/sec.

B. Accuracy of identification of attack sources

1) Definition of false-positive and false-negative: The accuracy of our method for identifying attack sources is evaluated by two metrics, false-positive and false-negative. We define false-positive as a case where a source not generating attack traffic is erroneously identified as an attack sources. We define false-negative for cases where an attack source cannot be identified. That is, the number of false-positives indicates the number of sources erroneously identified as attack sources and the number of false-negatives indicates the number of attack sources that cannot be identified. We also define the false-negative and false-positive rates as follows:

$$\text{false-negative rate} = \frac{\text{\# of false-negative}}{\text{\# of attack sources}}$$

$$\text{false-positive rate} = \frac{\text{\# of false-positive}}{\text{\# of sources not generating attack traffic}}$$

2) Number of attack sources vs. false-positives and false-negatives: We simulated our method to identify attack sources, changing the number of attack sources. We injected attack packets at 16 different times. We changed the number of attack sources from one to five. In this simulation, the total rate of attack traffic is 1000 packets/sec irrespective of the number of attack sources and the attack rate from each attack source is equal. For example, for one attack source, the attack rate from the attack source is 1000 packets/sec and for five attack sources, the attack rate from an attack source is 200 Packets. In this simulation, we set γ to 200 packets/sec.

TABLE I

NUMBER OF ATTACK SOURCES VS. FALSE-POSITIVES AND FALSE-NEGATIVES

# of attack sources	# of false-negatives (false-negative rate)	# of false-positives (false-positive rate)
1	0 (0.00)	2 (0.01)
2	0 (0.00)	0 (0.00)
3	0 (0.00)	3 (0.02)
4	3 (0.04)	4 (0.04)
5	12 (0.15)	4 (0.05)

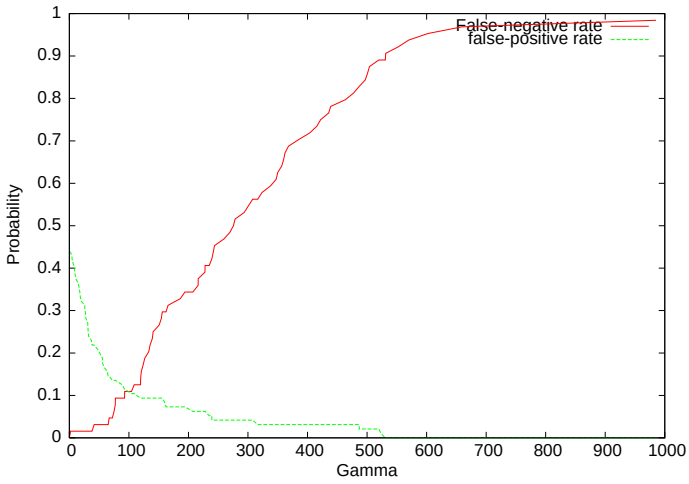

Fig. 7. γ vs. false-negative and false-positive (attack rate = 500 packets/sec)

Table I shows the total number of false-positives and false-negatives of 16 attacks and their rates. From these results, we can accurately identify the attack sources regardless of the number of attack sources. Although there are a few false-positives, these false-positives are caused by the rapid increase in traffic traversing to the link that is near the link to the victim. In these cases, the rapid increases cause errors because most of the path of the increased traffic is common with the path from the source of the increased traffic to the victim.

3) γ vs. false-positive and false-negative: We evaluate the relationship between γ and the false-positives or false-negatives in our method by using a simulation with various values of γ. In addition, we injected attack packets from four sources at 16 different times. Figures 7 and 8 show the results. In Figure 7, the total rate of attack traffic is 500 packets/sec. The total rate of attack traffic is 1000 packets/sec in Figure 8. From these figures, we can see that the proposed method can reduce the number of false-positives by setting γ to a larger value. However, a large γ causes many false-negatives. In addition, when comparing these figures, we can also see that if we set γ to the same value, we have less false-negatives in cases of larger attacks than in smaller attacks.

Figure 9 compares the false-positives for attacks at various rates. In this figure, the total rates of attack traffic are 200 packets/sec, 400 packets/sec, 600 packets/sec, 800 packets/sec and 1000 packets/sec. When we set γ to a larger value than

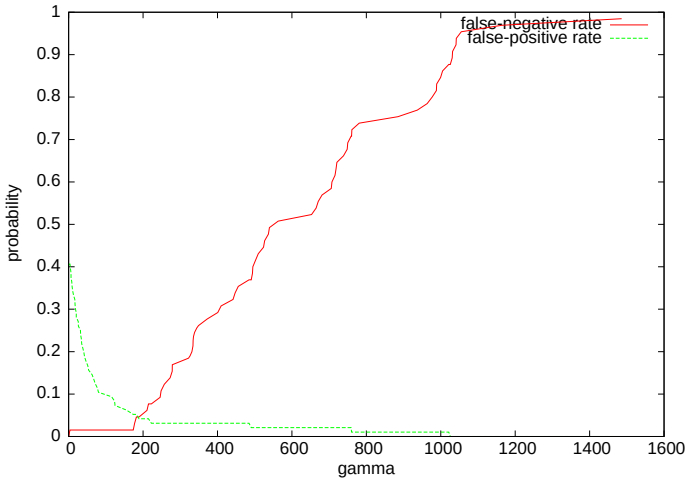

Fig. 8. γ vs. false-negative and false-positive (attack rate = 1000 packets/sec)

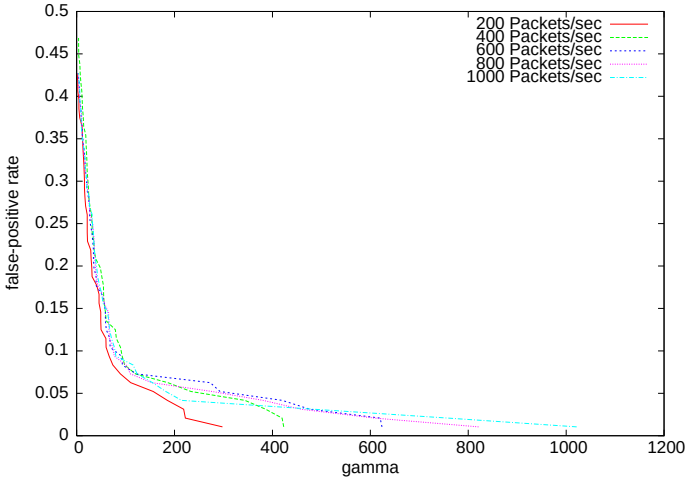

Fig. 9. Impact of γ against false-negative rate (various attack rates)

200 packets/sec, the number of false-positives is larger for larger attacks than smaller ones. However, these false-positives are caused by the rapid increase, as mentioned in the previous section. From this figure, we can also see that the number of false-positives is almost the same, regardless of the injected attack rate, when we set γ to the same value of less than 200 packets/sec. That is, the attack rate does not affect the number of false-positives.

4) γ vs. attack rate from unidentified attack sources: To evaluate the relationship between γ and the total rate of attacks from unidentified attack sources, we simulated our method to identify attack sources, changing the attack rate. In this simulation, we injected attack packets from four sources at 16 different times and the attack rate from each source is equal.

In Figure 10, the horizontal axis is the total rate of the attack traffic. Each line shows γ, which can identify one of the four attack sources, two of the four attack sources, three of the attack sources or all of the attack sources at all time. From

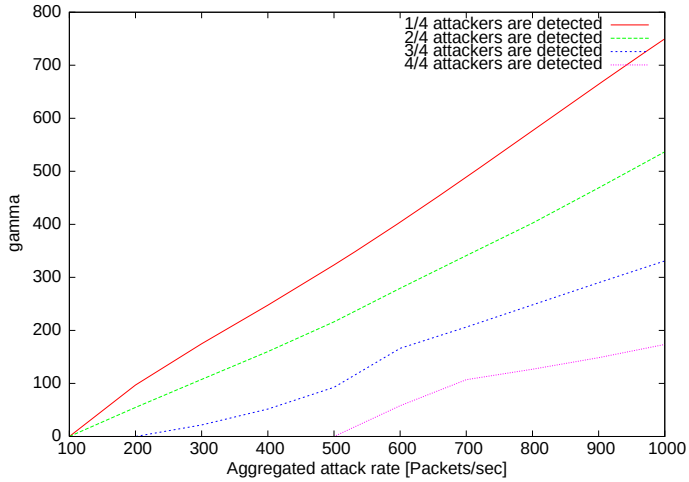

Fig. 10. Relationship between attack rate and γ to identify attack sources

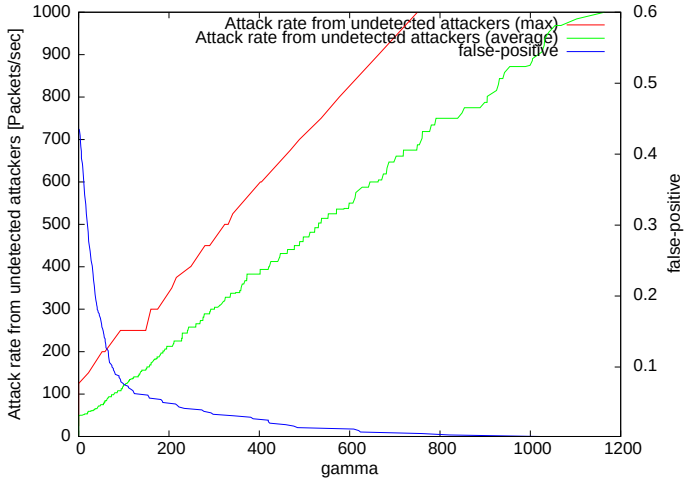

Fig. 11. γ vs. total rate of traffic from unidentified attack sources

this figure, we can see that a smaller γ is needed to identify attack sources for smaller attacks or to identify more attack sources. This figure also shows that even when we set γ to the same value, we can identify more attack sources for large attacks. For example, by setting γ to 100 packets/sec, we can identify only one attack source when the total rate of attacks is 200 packets/sec. However, by setting γ to the same value, we can identify three attack sources when the total rate of the attacks is 600 packets/sec.

Figure 11 shows the relationship between γ and the total rate of attack traffic from unidentified attack sources. In this figure, the three lines indicate the false-positive rate and the maximum and average of the total rate of attack traffic from the unidentified attack sources. From this figure, we can see that by setting γ to a smaller value, the attack rate from unidentified attack sources can be small while a smaller γ causes more false-positives. We can also see that the average of the total rate of attack traffic from unidentified attack sources is near γ. That is, the total rate of attack traffic from unidentified attack sources is closely related to γ.

When we set γ to a value less than 300 packets/sec, the maximum total rate of attack traffic from unidentified attack sources is near $\gamma + 100$. This is because in this simulation the increased value of legitimate traffic varies within ± 100 and the minimum increase in legitimate traffic on the link to the victim is -100 packets/sec. When we set γ to a value larger than 300 packets/sec, the maximum total rate of the attacks from unidentified attackers is near $\gamma + 200$. This is caused by errors in our method for estimating the increases in traffic. Our method for estimation has errors in the range of ± 50 packets/sec. That is, the estimated increase in traffic from an attack source may be 50 packets/sec less than the actual increase, while the difference from one to another attack source may be 50 packets/sec larger than the actual increase. In this case, this error causes 100 packets/sec attack traffic from unidentified attack sources. However, we can accurately identify attack sources sending attack traffic whose estimated rate is larger than $\gamma + \mu - f^{\mathrm{normal}}$. That is, by adequately setting γ, we can identify attack sources even when the estimated increases have several errors.

As previously mentioned, the total rate of attack traffic from unidentified attack sources is closely related to γ. That is, by defining the maximum attack rate that does not affects the network resources, we can adequately set γ to limit the total attack rate from unidentified attack sources to the defined maximum attack rate.

IV. Conclusion and Future Works

In this paper, we have proposed a new method for identifying attack sources by estimating traffic matrices. Our method periodically collects link load data from each router through SNMP and estimates the increase in traffic between each source and destination. When attacks start, our method identifies the sources of the attack using the estimated increase. We have also shown that our method can accurately identify attack sources without any false-positives by setting the adequate parameters of γ.

One of our future projects will be to simulate our method using real traffic data.

References

[1] S. Wu, L. Zhang, D. Massey, and A. Mankin., "On design and evaluation of intention-driven ICMP traceback," in *Proceedings of IEEE International Conference on Computer Communications and Networks*, Apr. 2001.

[2] B.-T. Wang and H. Schulzrinne, "A denial-of-service-resistant IP traceback approach," in *Proceedings of IEEE Symposium on Computers and Communications*, June 2004.

[3] S. Savage, D. Wetherall, A. Karlin, and T. Anderson, "Practical network support for IP traceback," in *Proceedings of ACM SIGCOMM 2000*, Aug. 2000.

[4] D. X. Song and A. Perrig, "Advanced and authenticated marking schemes for IP traceback," in *Proceedings of IEEE INFOCOM 2001*, Apr. 2001.

[5] K. Law, J. C. Lui, and D. K. Yau, "You can run, but you can't hide: An effective methodology to traceback DDoS attackers," in *Proceedings of International Symposium on Modeling, Analysis, and Simulation of Computer and Telecommunications Systems*, Oct. 2002.

[6] C. Partridge, L. A. Sanchez, C. E. Jones, F. Tchakountio, B. Schwartz, S. T. Kent, and W. T. Strayer, "Single-packet IP traceback," in *Proceedings of the ACM SIGCOMM 2001*, Aug. 2001.
[7] T.-H. Lee, W.-K. Wu, and T.-Y. W. Huang, "Scalable packet digesting schemes for IP traceback," in *Proceedings of IEEE International Conference on Communications 2004*, June 2004.
[8] A. Lakhina, M. Crovella, and C. D. February, "Diagnosing network-wide traffic anomalies," in *Proceedings of ACM SIGCOMM 2004*, Aug. 2004.
[9] Y. Zhang, M. Roughan, N. Duffield, and A. Greenberg, "Fast accurate computation of large-scale ip traffic matrices from link loads," in *Proceedings of ACM SIGMETRICS*, June 2003.
[10] D. Watson and C. Labovitz, "Experiences with monitoring OSPF on a regional service provider," in *Proceedings of IEEE International Conference on Distributed Computing Systems (ICDCS)*, Mar. 2003.
[11] A. Fedmann, A. Greenberg, C. Lund, N. Reingold, and J. Rexford, "NetScope: Traffic engineering for IP networks," pp. 11–19, Apr. 2000.
[12] "Scilab development team." available at `http://www-rocq.inria.fr/scilab/`.