Towards trapping wily intruders in the large

Glenn Mansfield, Kohei Ohta, Y. Takei, N. Kato, Y.Nemoto
{glenn,kohei}@cysols.com,
{takei,kato,nemoto}@nemoto.ecei.tohoku.ac.jp
Cyber Solutions Inc.6-6-3, Minami Yoshinari, Aoba-ku, Sendai, Japan
Graduate School of Information Sciences, Tohoku University, Sendai, Japan

Abstract

The rapid increase in network bandwidth from mega bits per second to giga bits per second and potentially to tera bits per second, is making it increasingly difficult to carry out in a timely and accurate manner, the analysis required to detect network abusers. The problem is made even more difficult with the devious techniques (e.g. spoofing) used by the hackers.

Intrusions are in general preceded by some noise or indication of the intruder groping for a door, trying (unsuccessfully) a key etc. In the network context these signals may be seen in the TCP-RESET packets and the ICMP echo-response or destination/port unreachable packets. But neither all TCP-RESETS nor all ICMP packets are indicative of attempted intrusions. Analysis of network traffic has shown that the profiles of such TCP-RESETs due to intrusion attempts are distinctly different from those due unintentional mistakes. The same profiling can be carried out for ICMP unreachable packets to detect attempts at intrusion. By monitoring such suspicious signals in a distributed information collection framework intrusions or attempts thereof can be effectively detected. To counter the threats posed by spoofing, a new technique based on packet flow monitoring is introduced. The flow patterns can be traced across networks to track an intruder. For this purpose we have developed an SNMP based messaging system which allows “friendly” networks to collaborate in tracking down the intruder. Results using prototype implementations on a medium size operational network are presented.

1 Introduction

The rapid increase in network bandwidth from mega bits per second to giga bits per second and potentially to tera bits per second, is making it increasingly difficult to carry out in a timely and accurate manner, the analysis required to detect network abusers. The problem is made even more difficult with the devious techniques (e.g. spoofing) used by the hackers.

Intrusions are in general preceded by some noise or indication of the intruder groping for a door, trying (unsuccessfully) a key etc. In the network context these signals may be seen in the TCP-RESET packets and the ICMP echo-response or destination/port unreachable packets. But neither all TCP-RESETS nor all ICMP packets are indicative of attempted intrusions. Analysis of network traffic has shown that the profiles of such TCP-RESETs due to intrusion attempts are distinctly different from those due unintentional mistakes [Kato99]. The same profiling can be carried out for ICMP unreachable packets to detect attempted intrusions. By monitoring such suspicious signals in a distributed information collection framework intrusions or attempts thereof can be effectively detected. A major challenge in intrusion detection and tracing their sources are spoofed packets where the packet header contents cannot be relied on. In such cases packet flows need to be monitored. The flow patterns can be traced across networks to track an intruder. In this work we present our results at extracting and tracing flow characteristics across a network.

2 Characteristics of Network Intrusions

Normal network usage patterns in general differ from the pattern when an intrusion is attempted. The general network usage pattern is – a client accesses a server/host to avail of a small number of services. The services maybe HTTP, TELNET, FTP, SMTP, NFS, SNMP, NTP, etc. An ill intentioned user tries to scan the server to discover the services that are being offered probably to explore the possibility of exploiting the security holes in those services. Naturally such accesses are characterized by a large number of services being accessed in a short period of time. The potential intruder may also be scanning all hosts in a network for a particular service, which may be exploited. In this particular case the number of hosts/server accessed from a host will be unusually high. Invariably these ships-in-the-night approaches give rise to a large number of ICMP host/port unreachable packets. Though ICMP host/port unreachable packets are literally omnipresent in a medium size network the profile of such packet trains show a significant pattern, when intrusions are being attempted.

2.1 TCP RESET Characteristics

We studied the number of TCP-connection requests in a medium sized campus network. The TCP connection requests were analyzed for source-destination characteristics.

Fig.1 shows the number of services accessed by clients during a one-day period. It is clear that most of the clients access a very small (<3) number of services and populate the left-hand end of the graph. Nevertheless on the right-hand end there are come instances of clients which have accessed a very large number (>500) of services.

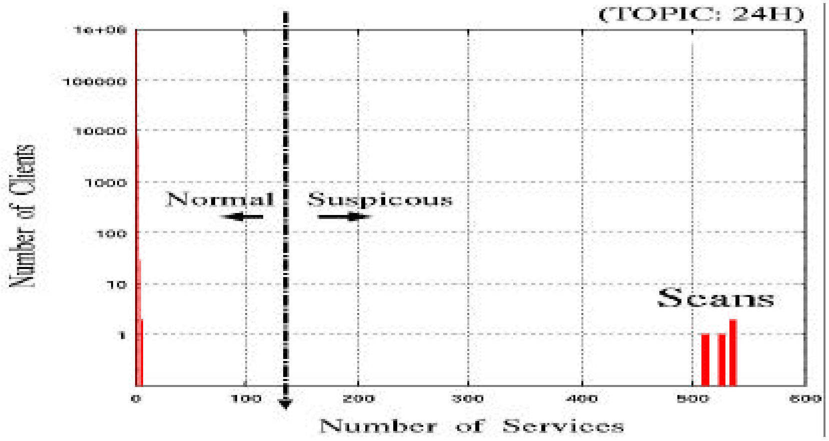

Fig. 1: Normal signals and suspicious signals

. 2.2 Characteristics of ICMP destination/port unreachable packets.

We collected and analyzed the ICMP port/destination unreachable (ICMP-UR) messages in a medium-sized campus network. We focussed on the ICMP-UR messages that originated from SNMP requests. SNMP is designed for Internet management purposes, but it is also a target for hackers and attackers. SNMP responds with a rich information about the system e.g. providing OS type, system type, currently running applications, interface information including the IP address, and other connecting nodes. An improperly configured (default community setting) SNMP agent readily volunteers information to any client. Naturally, SNMP agents are often the primary SCAN targets. Now, if an SNMP agent is running on the target host, no ICMP message is generated. But since the scanner is scanning all the hosts – there will be non-existent hosts and/or hosts not running SNMP agents resulting in the generation of ICMP destination unreachable and ICMP port unreachable messages. The ICMP messages contain the IP-header of the original packet as data. By looking at the data part of the ICMP packets we can find out the source address/port and destination address and port of the original packet.

The ICMP-UR messages generated by SNMP queries collected over a 24-hour interval are shown in **Fig.2**. It is clear that there are few ICMP-UR's that are generated by SNMP. Nevertheless there are two peaks – one at midnight and the other at 3am.

We examined the contents of the packets that constituted the peaks and listed out the timestamp, source address, destination address, source port and destination port. Table 1 shows a part of the list. It clearly shows that from one particular source (IP address) to several destinations (different IP addresses) in a network, SNMP queries were attempted unsuccessfully. The sequential nature of the destination addresses is fairly obvious. It is a clear case of an SNMP scan.

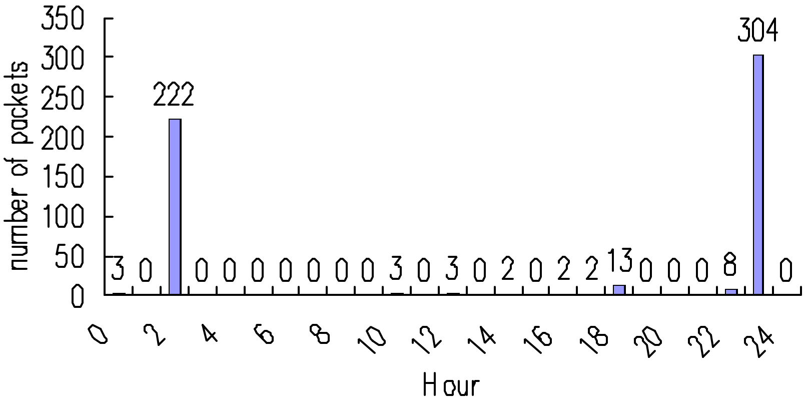

Fig. 2 Number of ICMP-UR packets (port SNMP(161))

Table 1 ICMP destination port unreachable messages for SNMP port (under scan)

Timestamp	Source IP	Destination IP	Src port	Dest Port
928256855	nnn.101.0.20	nnn.211.2.63	1026	SNMP(161)
928256855	nnn.101.0.20	nnn.211.2.62	1026	SNMP(161)
928256855	nnn.101.0.20	nnn.211.2.61	1026	SNMP(161)
928256855	nnn.101.0.20	nnn.211.2.60	1026	SNMP(161)
928256855	nnn.101.0.20	nnn.211.2.59	1026	SNMP(161)
928256856	nnn.101.0.20	nnn.211.2.25	1026	SNMP(161)
928256856	nnn.101.0.20	nnn.211.2.24	1026	SNMP(161)
928256856	nnn.101.0.20	nnn.211.2.23	1026	SNMP(161)

Fig.3 shows another characteristic of such scans viz. the inter-message interval distribution of the ICMP-UR messages generated by SNMP queries. Scans give rise to a large number of ICMP-UR messages, which have a small inter-message interval.

Thus there two significant characteristics of traffic generated by a scan viz. the spurt in the ICMP-UR messages and the thinly spread clustered or sequential nature of the destination addresses. The spurts can be readily detected by a threshold mechanism. But, of course a smart and patient scanner will use a low rate of scan packets and will distribute the destination addresses randomly in the given network range. Smarter algorithms will be necessary to detect the range, which is being targeted by the hacker. The initial threshold will need to be set reasonably low – two or three related packets should be enough to set of the alarm and get the detailed investigation ball rolling.

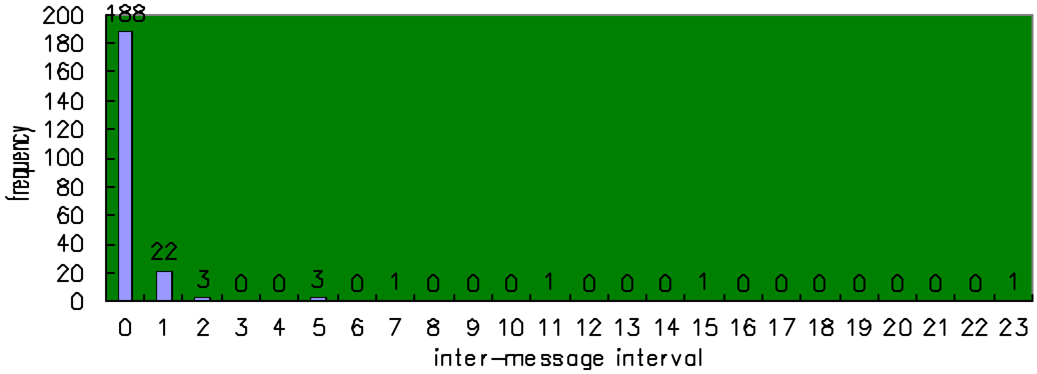

Fig. 3 Distribution of inter-message interval

3 Detection of Intrusions from traffic-flow signatures

Packet contents are conveniently examined to determine the source, destination and protocol related details to determine the traffic-profile. However, this mechanism in traffic profiling has limited applicability as

- A clever hacker will manipulate the packet contents e.g. spoof the source addresses whenever possible, particularly in the case of a Denial of Service (DoS) attack,
- The packet contents may be encrypted making it difficult if not impossible to decode and analyze the contents,
- The traffic volume itself may be very large making it a difficult task to analyze the contents of each and every packet.

Moreover in case the source address is spoofed the attack may be detected by whatever means but still the issue of tracing the attacker remains wide open.

In such cases profiling traffic-flow characteristics is a more appropriate choice. Essentially one looks for distinguishing features of the traffic profile. For example, in case there is a DoS attack in the form of a flood of TCP-SYN packets, this can be easily detected by the entity receiving the TCP-SYNS or by an RMON type device placed at an appropriate point in the network. Then the source of the malicious packets is traced by looking for the presence of a similar flow at all the inputs of the concerned network.

3.1 Traffic-flow signatures

The basic concept of signature-based traffic tracing is shown in Fig.4. The traffic monitor collects the relevant packet count information from each link, which connects the sites. The NMS compares the monitored traffic pattern, and correlates them. The correlated chain of patterns indicates the path of (probably spoofed) traffic-flow. It should be noted that the information used is packet count only, neither packet capture nor analysis is needed.

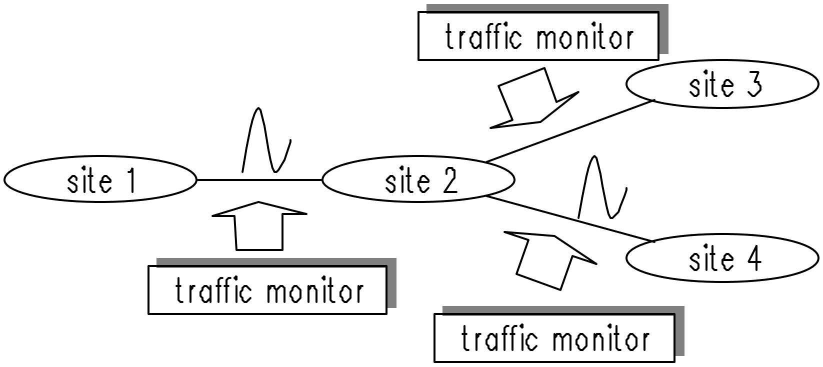

Fig. 4 Concept of pattern based traffic tracing

3.2 Definition of traffic-flow signature

Traffic-flows are measured in time slots. The signature of the flow is viewed in a window, which comprises an integral number of slots. The traffic flow signature is defined by the time slot size, window size and the measurement of the metric in each slot in the window (**Fig.5**).

Traffic flow signatures are defined by the vector.

$$\boldsymbol{A} ? (a_1, a_2, \cdots, a_n)$$

$$? a_i \, (1 ? i ? n) \textit{ is number of packets a time slot } ?$$
$$? $$
$$?? \; ? \; n ? \; ?$$

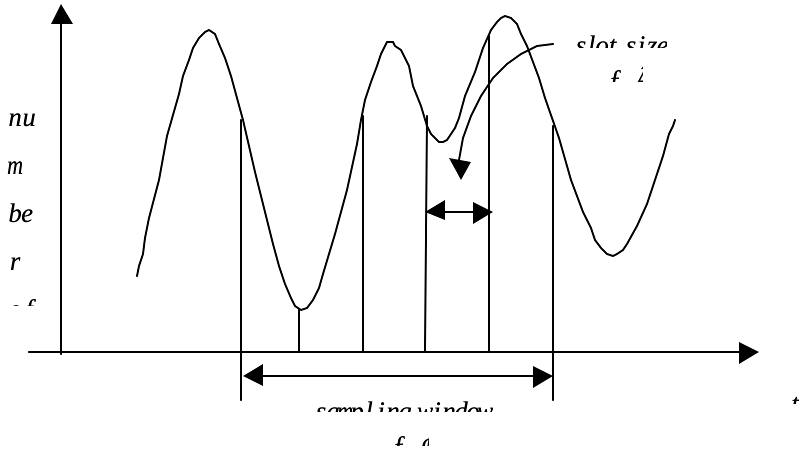

Fig. 5 Model of traffic pattern

3.3 Correlating traffic-flow signatures

Correlation of traffic pattern is based on correlation coefficient *r*, calculated as follows. (***A*** is incoming flow, and ***B*** is outgoing flow)

$$r(A,B) ? \frac{1}{ns_{in}s_{out}} \overset{n}{\underset{i?1}{?}} (a_i ? \overline{A})(b_i ? \overline{B})$$

$$\begin{array}{l} ? \\ ?s_{in} ? \sqrt{\frac{1}{n}\overset{n}{\underset{i?1}{?}} (a_i ? \overline{A})^2} \\ ? \\ ? \\ ? \\ ?s_{out} ? \sqrt{\frac{1}{n}\overset{n}{\underset{i?1}{?}} (b_i ? \overline{B})^2} \\ ? \end{array}$$

If *r(**A**,**B**)* is close to 1, the traffic-flow with signature ***A*** is the same as the traffic with signature ***B***. In other words the same flow has been detected at the two points.

3.4 Experimental evaluation

In a network configuration shown in **Fig.6**, we carried out our measurements. Two probes, *probe1* and *probe2* were used to monitor two 100Mbps FDDI loop links for ICMP echo/response packets. The measurement and correlation parameters are shown in **Table. 2.**

We have focused on ICMP Echo Request/Reply packet flows. Such packets are used in one form of DoS type attack like *Smurf*. In a typical *Smurf* attack, an ICMP echo request is sent to a broadcast address and the source address is spoofed to that of a host, which is the target of the attack or is in the targeted network. In such cases any information in the packet header cannot be reliably used to track down the perpetrator of the attack.

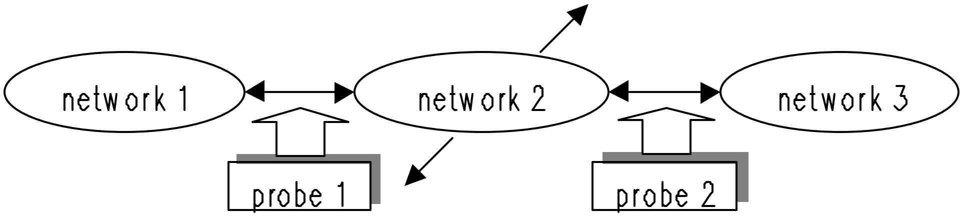

Fig. 6 Experiment environment

Table 2 experimental configuration

Size of time slot .	1 minute
Window size .	5 slots
Threshold of correlation coefficient	0.9

3.5 Relay of ICMP echo reply

A burst of ICMP echo replies may indicate that a *Smurf* attack is underway. **Fig.7** shows a sample result of the detected and correlated burst, which indicates a *Smurf*. The light bar indicates the traffic-flow from site 1 to site 2 measured at probe1 the dark bar indicates the traffic-flow from site 2 to site 3 measured at probe2. We can see that the correlation is (almost) exact.

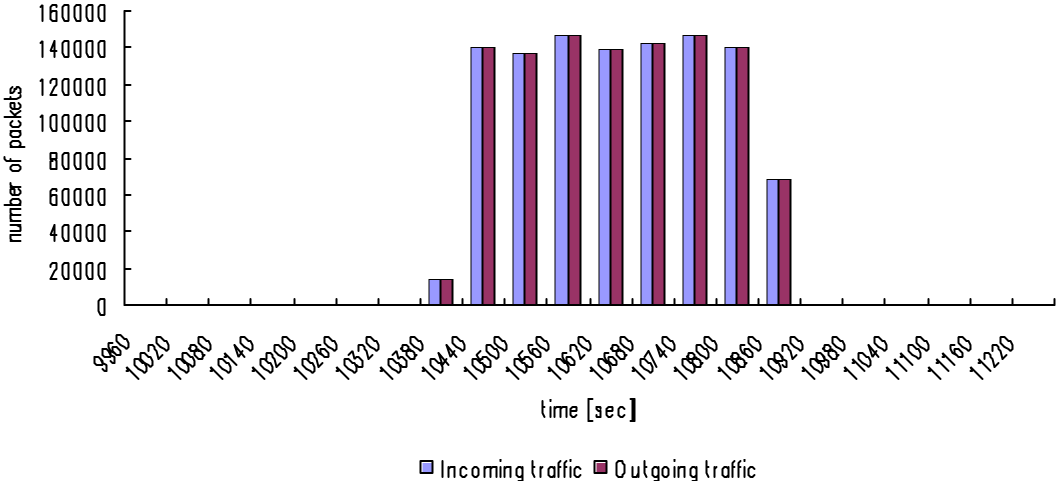

Fig. 7 ICMP echo reply relay

3.6 Relay of ICMP echo request

Fig.8 shows the correlation of ICMP echo requests measured by probe1 and probe2. The difference between **Fig.7** and **Fig.8** is order of the traffic. In case of echo request, the amount of traffic is significantly less than that in the case of echo reply traffic. But, both the traffic behaviors could be detected using the same configuration of parameters.

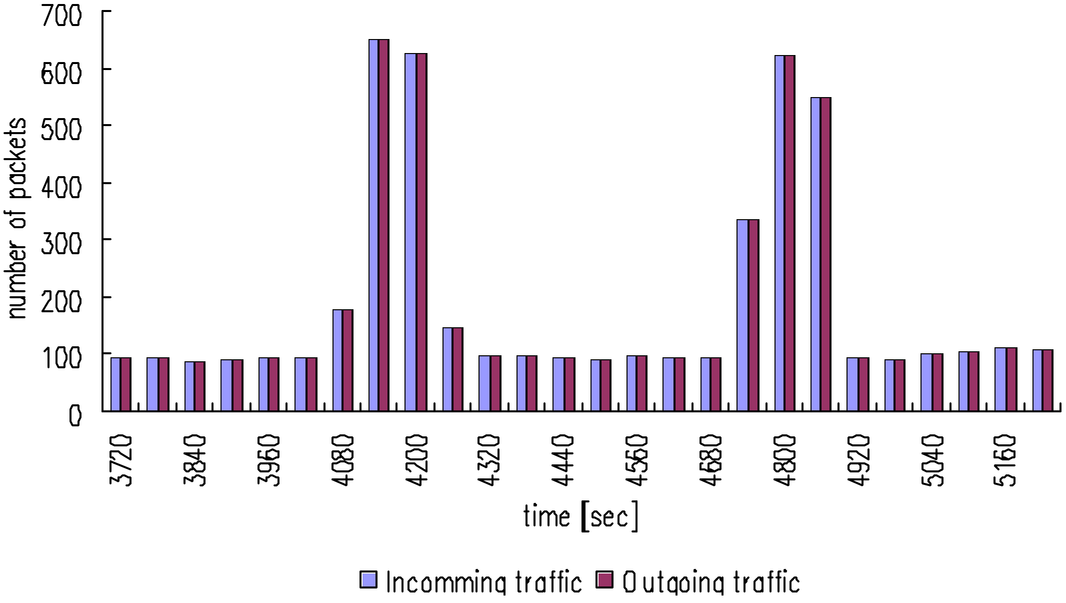

Fig. 8 ICMP echo request relay

3.7 Traffic-flow correlation- its usability

The traffic-flow correlation is reasonably successful, leading us to conclude that traffic-flows may be traced from link to link across a network. However complications can be expected in case the link speeds are different at different points in the network. The static and dynamic characteristics of the link are expected to influence the traffic-flow patterns. In such cases more involved techniques will be necessary to take these factors into account.

4 Map-based distributed Intrusion tracing

Network configuration information or network maps can be very effectively used in tracking intruders. In [Glenn99] the technique of network configuration information synthesis from information available in the

networks is discussed. Using this information network maps showing AS-level interconnectivity can be generated and visualized. Using this technology in conjunction with our traffic-flow tracing technique a powerful and effective means of tracking the intruder can be devised.
The methodology envisages inter-AS cooperation and collaboration. It works in a distributed manner. When there is a surge of TCP-SYNs or TCP-RESETs the flow pattern is matched with patterns in the traffic from the connected network links. The network configuration information is effectively used in finding the corresponding links. By following the chain of links the source of the spoofed attack can be narrowed down to a smaller part of the Internet. In Fig.9, The methodology is illustrated. A surge in TCP-SYN packets is detected in the TOPIC network. The TCP-SYN surge pattern is looked for at the two links of TOPIC. It is detected on the WIDE-TOPIC link. Then the pattern is searched for on the links to WIDE and so on.

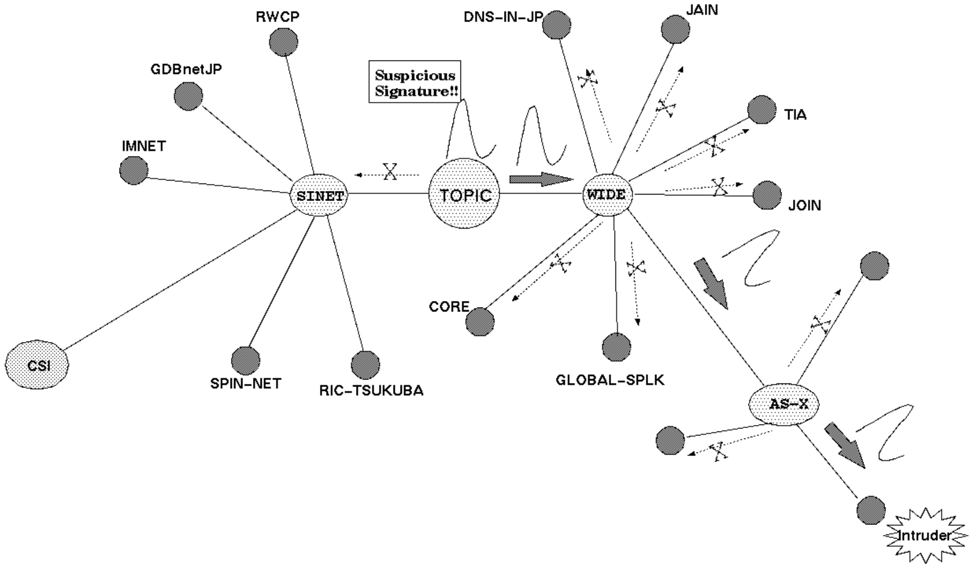

Fig.9 Map-based intruder tracking

5 Implementations and Results

The traffic monitoring is carried out using agents which watch all the traffic but process only the suspicious packets. The agents can be accessed, queried and configured using the standard SNMP management protocol. The Security Manager system is alerted on the detection of potential attempts. The Security Manager uses the network configuration information to trap and/or track-down the intruder. The communication between the different Manager's and the agents is carried out using the standard SNMP management protocol. The communication utilizes the security features provided in the SNMP framework viz. authenticity confidentiality, integrity, reliability. The asynchronous alerts are communicated using *Inform* requests. The message conveyed contains a list of managed objects describing the event that has been detected. The prototype is currently under evaluation.

6 Conclusion

We have discussed methods of profiling network traffic to distinguish normal usage from abnormal or ill-intentioned usage. By monitoring such suspicious signals in a distributed information collection framework intrusions or attempts thereof can be effectively detected. We have discussed a new technique based on packet flow monitoring to counter the threats posed by spoofing. Using network configuration information the flow patterns can be traced across networks to track an intruder. For this purpose we have developed an SNMP based messaging system which allows "friendly" networks to collaborate in tracking down the intruder.

The system does not depend on specific type of attacks or patterns and as such does not attempt to reconstruct the detailed traffic pattern. It relies on the general protocol mechanisms and is therefore simpler and effective. At the same time it will be unable to detect a well-informed intruder who finds access to his/her target with ease.

References

[Kato99] N. Kato, H. Nitou, K. Ohta, G. Mansfield, Y. Nemoto, *A Real-time Intrusion Detection System(IDS) for Large Scale Networks and Its Evaluations*, IEICE Trans. Communications, Vol. E82-B, N0.3 (1999).

[Mukherjee94] Biswanath Mukherjee, L.Todd Herberlein, and Karl N. Levitt: *Network Intrusion Detection*", IEEE Network, Vol.8,N0.3,[1994] pp. 26--41

[Postel81] J. Postel, "Internet Control Message Protocol", RFC 0759.

[Dorothy87] Dorothy E Denning, *An Intrusion Detection Model*", IEEE Trans. on Software Engineering, Vol.SE-13, N0.2[1987] pp. 222--232

[Othmar97] Othmar, *Inernet Security:Risk Analysis, Strategies and Firewalls*", International Thomson Publishing 1997.

[Iguchi99] M. Iguchi, and S. Goto: "Detecting malicious activities through port profiling", IEICE Trans. on Information and System, 1999, in press

[Glenn99] G. Mansfield, K. Jayanthi, A. Ashir, N. Shiratori, "*Network Maps: Synthesis and Applications*", Proc. APSITT'99, Mongolia, August 1999.

Authors List

Glenn Mansfield obtained his Ph.D. specializing in Logic Programming, from Tohoku University, Japan. He has worked as a research associate I the computer center of Tohoku University for a period of three years and is currently a senior visiting researcher at Tohoku University and director of Cyber Solutions, Inc. His areas of interest include network management, network cartography, high speed networks, directory systems, logic programming and expert systems. He is a member of Internet Society, the IEEE, IEEE communications Society, and the Information Processing Society, Japan.

Kohei Ohta received his M.S and Ph.D. degrees from Graduate School of Information Sciences, Tohoku University in 1995 and 1998, respectively. Now he is a senior researcher at cyber Solutions. He has been engaged in research on network Management.

Yohsuke TAKEI received his B.E from Tohoku University, Japan in 1998. He has been engaged in research on Internet management, measurement and security. He is currently working toward the M.S. degree at Graduate School of Information Sciences of Tohoku University

Nei Kato received his M.S and the Dr.Eng degrees in information engineering from Tohoku University in 1988 and 1991, respectively. He joined Computer Center, Tohoku University in 1991 and now he is an associate professor at Graduate School of Information Sciences, Tohoku University. He has been engaged in research on computer networking, pattern recognition and neural networks. Dr. Kato is a member of IEEE and the Information Processing Society of Japan.

Yoshiaki Nemoto received his B.E, M.E and Ph.D. degrees from Tohoku University, Japan in 1968, 1970, and 1973 respectively. Now he is a professor with Graduate School of Information Sciences, and the director of Computer Center, Tohoku University. He has been engaged in research work on microwave networks, communication system, computer network system, and image processing and handwritten character recognition. He was a co-recipient of the 1982 Microwave Prize from IEEE microwave theory and Techniques Society. Dr. Nemoto is a member of the IEEE and the Information Processing Society of Japan.