DNS-based Network Anomaly Detection and Eradicating Scheme†

Chang-Shang Chen(陳昌盛)[1] **,** Shang-Rung Wang[2] (王向榮)**,** Ta-Chung Liu[1] (劉大川)
[1]Computer and Network Center,
[2]Campus Computer Communication Assoication
National Chiao Tung University, Hsinchu, Taiwan
Corresponding E-mail: cschen@mail.nctu.edu.tw

摘要

本文主要研究內容，在於設計並實作完成一套以結合網域名稱系統(DNS)查詢分析以及 IEEE 802.1x 認證系統為本的網路異常偵測與根除入侵者的網路管理輔助系統。

關鍵詞：網路異常偵測、網域名稱系統、IEEE 802.1x 認證、入侵移除

Abstract

In this paper, we design and implement a DNS-based network anomalous detection and intrusion eradication scheme, combining the DNS-based anomaly detection and IEEE 802.1x-based authentication scheme for supporting the intrusion eradicating process.

Keywords: DNS, IEEE 802.1x, intrusion detection, intrusion eradication

† This work was partially supported by National Science Council of the Republic of China under Grant No. NSC94-2213-E-009-111.

1. Introduction

Currently, even though there are many security tools (e.g., anti-SPAM, anti-virus, anti-spyware, and firewall systems) available to help protect from and/or filter out unwanted traffic; however, from time to time, administrators on many sites could receive complaint messages from users or administrators of local and/or remote sites. Experiences show that most of them are about SPAM mails or compromised messages, mentioning about illegal intrusion attempts.

Moreover, due to the distributed nature of network users' hosts and servers, different user environments and topologies will lead to different results. Especially, ordinary users usually do not know the theoretical and practical knowledge of the related network systems very well and some even do not know about (and do not install) these protection systems. Thus their hosts are usually weakly protected and often exploited, with the owner unknown, to conduct network abusing activities.

Nowadays, most Internet services are based on the working model that there will be some Domain Name System (DNS) [1][7] queries before the communication activities. In practice, there are anomalies or network abusing attempts located in the midst of typical DNS activities from time to time. For example, a typical scenario will be as follows. Before launching intrusion attempts, some intruders will initiate series of domain name (and/or IP) probing and port-scanning for locating targets. After the victims are found, further system exploitations are followed...

On the other hand, in some critical cases (e.g., by policy or equipment issues), it is hard to identify the very people responsible for the compromised hosts in a timely manner. For example, in some campus networks, many users do not register their computers with fixed IP addresses for Internet access (e.g., not knowing the site's policy, or just choosing to ignore it). Sometimes, one newly booted host will get address conflicts with another connected host and the user just randomly chooses another available each time. If compromised to abuse the network, these hosts often severely defer the intrusion eradication process (i.e., the administrators first have to make a lot of checks to avoid finding the incorrect ones responsible for the compromised hosts).

Therefore, in this paper, we describe our work on the design and implementation of a DNS-based, network anomaly detection and intrusion eradication scheme by combining the DNS-based query log analysis of access patterns and IEEE 802.1x-based [5][6] authentication process to help administrators identify the network anomalous activities in the early phase, locate the suspected problem sources and fix them as soon as possible to reduce the impact of the abusing hosts.

The organization of the rest of the paper is as follows. Section 2 gives background overviews for highlighting the research directions. In Section 3 we describe the system architecture and the design principles. Section 4 contains some implementation details and discussions issues. Finally, in Section 5, a concluding remark and some points are highlighted for future research.

2. Motivation Scenarios

According to domain expertise, many kinds of problems might severely affect the operation of network applications and the DNS.

2.1 Plausible DNS-based Anomaly Detection

Network anomaly detection is a particular challenging

area of data analysis, not only because it is hard to build a training set of known anomaly cases, but also because anomaly may take many forms. For example, virus or spam mail injections are typical notorious examples. Moreover, in distributed computing environment, intrusions are rarely limited to a single network application domain. Thus, recognizing security breaches (e.g., e-mail, www, p2p [9]) can require analyzing a large amount of different information sources (e.g., including the analysis of the router traffic, the analysis of the DNS traffic).

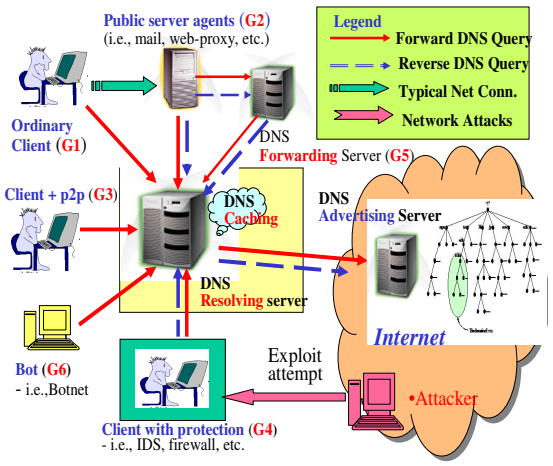


Fig. 1: Typical DNS clients and servers.

In general, DNS traffic consists of independent queries from different sources and of different types (A, MX, and PTR, etc.). In principle, as shown in Fig.1, a typical site might have several independent advertising and/or recursive DNS servers for serving incoming and outgoing queries (e.g., two for the former and another three for the latter) about the forward and corresponding domain zones. As shown in Table 1, there are typical example cases (i.e., network abusing or intrusion attempts collected from our campus network) that could be identified via the analysis of the DNS system query logs.

Table 1: DNS-based anomaly cases detection

1.	Botnet [2] probing: ♦ Repeatedly checking for currently unknown host (e.g., A-RR, MX-RR)
2.	SPAM or virus – open mail proxy and/or virus engine (e.g., MX-RR)
3.	Remote Login exploits - SSHd, Telnetd, Ftpd, etc. (e.g., PTR-RR)
4.	DNS Zone Transfer attacks by Abusing the Network
5.	DNS resolving/forwarding storm ♦ DoS attack

In practice, as shown in Table 2, most DNS queries are conducted on some major hosts. For example, the DNS clients listed in categories 1, 2, 4, and 5 are usually recognized and acceptable. On the other hand, the traffic introduced by hosts in categories 3 and 6 are usually not welcome. Often, they are either malicious programs, or underground client/server processes. All of these might consume lots of network and system resources.

Table 2: Typical users/programs of an ordinary DNS resolving server

Category	Examples (refer to Figure 1)
1. Ordinary clients (G1)	Ordinary clients without specialized protection mechanism
2. Normal server (G2, G4)	• Mail, web proxy, etc. (G2), • Personal firewall systems (G4)
3. p2p clients (G3)	BitTorrent, eDonkey, etc.
4. DNS server (G5)	Downstream DNS forwarding servers
5. Malicious program (G6)	• Botnet, network virus/worm (e.g., mail, web), etc. • intrusion attempts (SSH/Telnet/Ftp exploits, etc.), etc.

2.2 IEEE 802.1x-based Authentication System with Radius [8]

As mentioned in Sec.1, in some critical cases (e.g., by policy or equipment issues), there is still a real-world problem that it is hard to identify the very people responsible for many compromised hosts in a timely manner. If these hosts were compromised to abuse the network, these hosts often severely defer the intrusion eradication process (e.g., since you have to make a lot of checks to avoid finding the incorrect ones responsible for the compromised hosts).

Therefore, to attack the problem mentioned above, one newly developed approach is to incorporate a layer-2 authentication scheme (e.g. IEEE 802.1x, MAC address). That is, on each new connection to the Internet, any user has to pass the layer-2 authentication scheme. In this paper, we mainly adopt the IEEE 802.1x-based authentication scheme. Before enabling the 802.1x-based authentication capability and getting one's host connected to Internet, each user has to do the registration to authentication database of the Radius server.

In addition, to deal with the issues for helping identify the problem sources in the dormitory part of our campus network, we refine the DormNet IP registration system, mainly by incorporating an IEEE 802.1x-based authentication scheme. It could be used for registering the dormitory network users in our university to help identify the appropriate people responsible for the compromised hosts. For hosts unable to enable the IEEE 802.1x based authentication

scheme (e.g., missing IEEE 802.1x capabilities), the approach to keep MAC addresses (e.g., registering the MAC and corresponding IP address) will be used instead.

3. Methodology and System Architecture

As mentioned above, the first problem to be addressed is how to identify the candidate problem source in the early phase. Next, the identified information could be further used for checking with the authentication system to persuade the users of the compromised hosts to fix the problems as soon as possible.

3.1 DNS-based Anomaly Detection Scheme

In general, the intrusion detection schemes of security systems fall into one of two categories, anomaly detection or misuse detection [10]. Anomaly detectors look for behavior that deviates from normal system use. Misuse detectors look for behavior that matches a known attack scenario. Most IDS (Intrusion Detection System) or NIDS (Network Intrusion Detection System) in use are based mainly on misuse pattern matching techniques. On the other hand, in distributed computer environment, intrusions are rarely limited to a single network or domain. Thus, recognizing security breaches containing lots of different services (e.g., e-mail, www, VoIP, etc.) can require analyzing a large amount of different information sources.

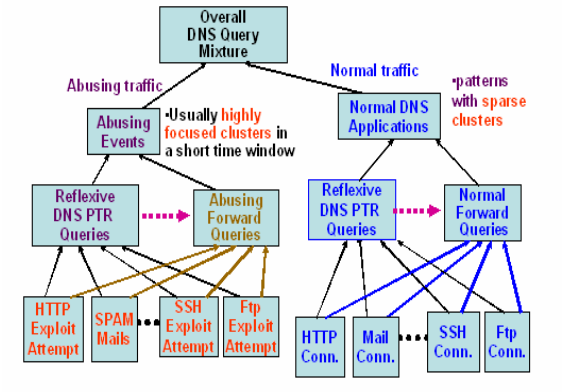


Fig. 2: Mixture of DNS queries

As mentioned in [4], since DNS servers are hierarchically distributed among different departments and organizations, the mining of the DNS traffic distribution data and comparing with their history profiles might provide a convenient and efficient way to help identify anomalous activities (as shown in Fig. 2) between the pairs (compromised/attacking hosts, victim hosts) and persuade the users of the compromised hosts, after confirmed, to eradicate the intrusion and/or vulnerability as soon as possible.

3.2 System Architecture

As shown in Fig.3, we adopt the DNS-based, two-phase network anomaly detection scheme as proposed in [3][4]. The system aims at identifying candidate sources of compromised hosts from a collection of DNS query logs and from the background knowledge provided by the domain experts.

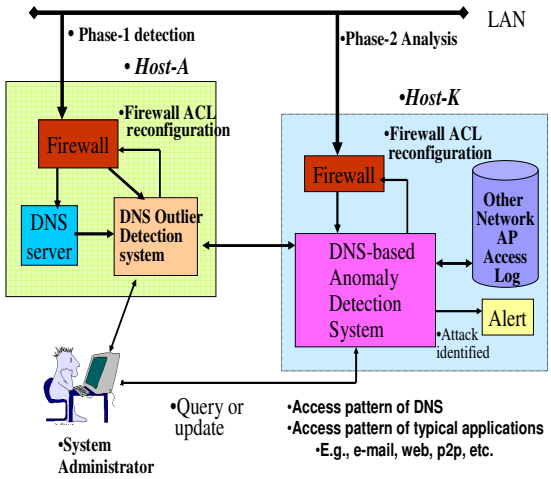


Fig.3: DNS-based two-phase network anomaly detection scheme

The general idea is as follows. As mentioned above, there are anomalies or network abusing attempts located in the midst of typical DNS activities from time to time. In phase 1, by collecting and analyzing DNS traffic from a live DNS network, the problem is detecting the outliers in the early phase and further identifying the problem types and sources (e.g., locating the virus-affected or compromised hosts) for fixing. In phase 2, by combining the data mining (e.g., machine learning, association mining) analysis of access patterns of the Domain Name System (DNS) and other network applications (e.g., e-mail, web), it is supposed that the integrated analysis could be a great tool to help the system administrators identify the network anomalous activities in the early phase, locate the suspected problem sources and fix them as soon as possible to reduce the impact of the abusing hosts.

4. Experiment and Discussion

4.1 Implementation of DNS-based Anomaly Detection System

At the time of writing, the system environment is listed as shown in Table 3. In general, our DNS-based scheme and implementation help lessen the problem to identify the network anomalous activities in the early phase and locate the suspected problem sources for fixing to reduce the impact of the abusing hosts on the overall network operation. For example, Fig. 4 shows the snapshot of an identified anomaly candidate (i.e., a possible bot of a certain Botnet) on the Phase-2 data analysis server. The listed host was repeatedly

trying to send forwarding DNS queries for a currently unresolved domain name (e.g., "***mail.ballzout.info***").

Table 3: The system operation environment

1. DNS servers	PC-based server running • FreeBSD (4.11, 5.4) • BIND DNS server (9.3.2) • Tool – dig, Dnstop
2.data warehouse server	• Windows 2003 Standard Eng • MS SQL Server 2005 Enterprise edition

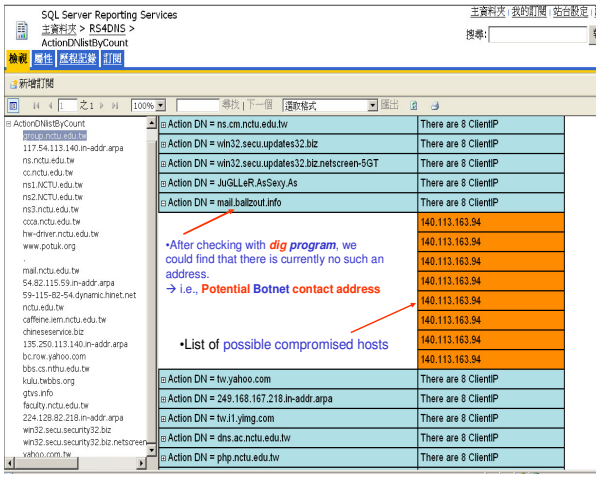
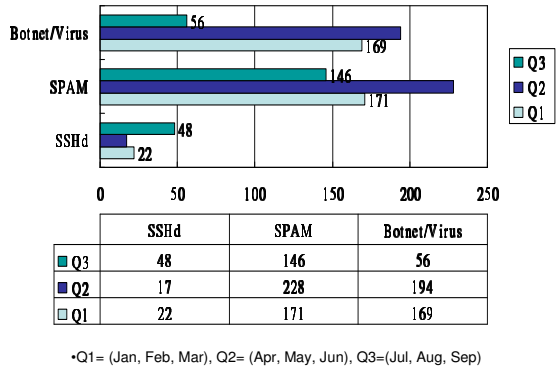


Fig. 4: Identification of IP/host lists possible compromised (e.g., BOTNET)



	SSHd	SPAM	Botnet/Virus
Q3	48	146	56
Q2	17	228	194
Q1	22	171	169

•Q1= (Jan, Feb, Mar), Q2= (Apr, May, Jun), Q3=(Jul, Aug, Sep)

Fig. 5: Accumulative Statistics of NCTU IP/hosts abusing events from 2006.01 to 2006.09.

4.2 Typical Example Cases

Fig. 5 shows the accumulative statistics of NCTU IP/hosts abusing events from 2006.01 to 2006.09. Roughly speaking, nearly 50% of the reported events are concerning SPAM activities and 40% are events about hosts being compromised by worm/virus/botnet. The rest are events about SSHd compromised.

4.2.1 **CS-1:** Virus infection or open relay/proxy

- **Short Summary:** The IP address listed below (located in the DormNet of NCTU) has either a Virus infection or is an open relay/proxy.
- **Remark**: In practice, since it is a student PC located in the DormNet, a normal client host will not issue such a huge amount of DNS queries (e.g., A and/or MX RR), with the same query port (#1046) and in such a short time interval (e.g., within seconds). That is, it strongly indicates that the specified host either has a high possibility of Virus infection or is an open relay/proxy.

- Nov 22 08:11:41.974 info: client 140.113.92.244#1046: query: mx.siemens.com IN A
- Nov 22 08:11:42.286 info: client 140.113.92.244#1046: query: mx-pa-1.pobox.com IN A
- Nov 22 08:11:44.417 info: client 140.113.92.244#1046: query: mail.siemens.com IN A
- Nov 22 08:11:44.481 info: client 140.113.92.244#1046: query: smtp.siemens.com IN A
- Nov 22 08:11:44.545 info: client 140.113.92.244#1046: query: mx1.siemens.com IN A
- Nov 22 08:11:44.753 info: client 140.113.92.244#1046: query: mx-pa-2.pobox.com IN A
- Nov 22 08:11:46.859 info: client 140.113.92.244#1046: query: mxs.siemens.com IN A
- Nov 22 08:11:48.513 info: client 140.113.92.244#1046: query: mx-pa-7.pobox.com IN A
- Nov 22 08:11:49.173 info: client 140.113.92.244#1046: query: mail1.siemens.com IN A
- Nov 22 08:11:49.238 info: client 140.113.92.244#1046: query: relay.siemens.com IN A

[deleted]

4.2.2 **CS-2: Recognizing Possible** SSHd Attacks

- **Short Summary:** The queried IP address (e.g., PTR-RR, 140.121.175.138) by the following NCTU IP clients indicate a highly suspected source of SSH attacking/attacked host.
- **Remark:** In practice, systematic exploits (attacks) to the SSHd is one of most common exploits on Unix-like platforms.

- Nov 21 08:58:34.735 info: client 140.113.23.2#4269: query: **138.175.121.140**.in-addr.arpa IN PTR
- Nov 21 08:58:35.293 info: client 140.113.22.134#1163: query: 138.175.121.140.in-addr.arpa IN PTR
- Nov 21 08:58:49.675 info: client 140.113.28.88#48974: query: 138.175.121.140.in-addr.arpa IN PTR
- Nov 21 08:58:49.689 info: client 140.113.28.88#48974: query: 138.175.121.140.in-addr.arpa IN PTR
- Nov 21 08:59:15.857 info: client 140.113.36.26#1024: query: 138.175.121.140.in-addr.arpa IN PTR
- Nov 21 08:59:15.911 info: client 140.113.39.238#1024: query: 138.175.121.140.in-addr.arpa IN PTR
- Nov 21 08:59:16.017 info: client 140.113.36.190#1552: query: 138.175.121.140.in-addr.arpa IN PTR

- Nov 21 08:59:16.030 info: client 140.113.36.183#1032: query: 138.175.121.140.in-addr.arpa IN PTR

[Deleted]

4.3 Supporting an IEEE 802.1x-based Authentication System

As mentioned above, before enabling IEEE 802.1x-based authentication process in the DormNet Blocks of IP addresses, each user has to do the registration to the Radius server. Fig.6 shows a snapshot of the Radius server for supporting IEEE 802.1x-based authentication scheme on NCTU DormNet (URL, https://140.113.27.27/).

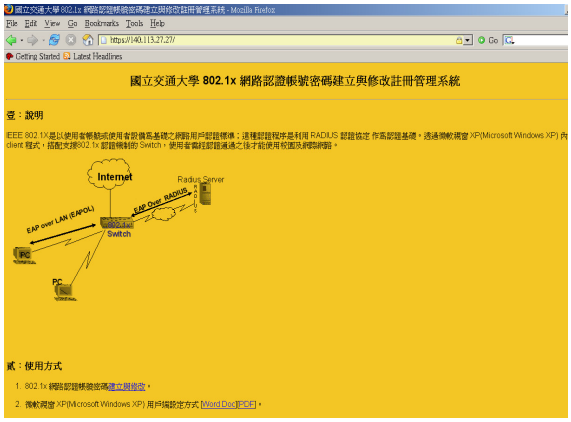


Fig. 6: NCTU IEEE 802.1x-based Authentication System

Next, as shown in Fig.7, we have refined the DormNet IP registration system, mainly by incorporating an IEEE 802.1x-based authentication scheme, for registering the dormitory network users in our university to help identify the appropriate people responsible for the compromised hosts. For hosts unable to enable the IEEE 802.1x based authentication scheme (e.g., missing IEEE 802.1x capabilities), the approach to keep MAC addresses (e.g., registering the MAC and corresponding IP address) will be used instead.

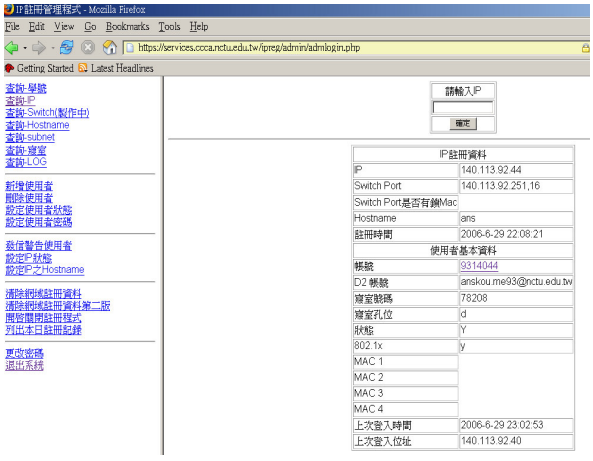


Fig.7: DormNet IP Registration/Query System

According to the statistics, currently, we have 6827 students applying for the Dorm (2006). From the DormNet IP Registration System, we could find that there are 5907 IP/hosts with 802.1x-based capability enabled and another 2002 IP/hosts with MAC-based authentication scheme (i.e., 5907:2202, or 77%:23%). Fig. 8 shows the Layer-2 authentication statistics of the NCTU DormNet by student identities (i.e., Enrolling School Year).

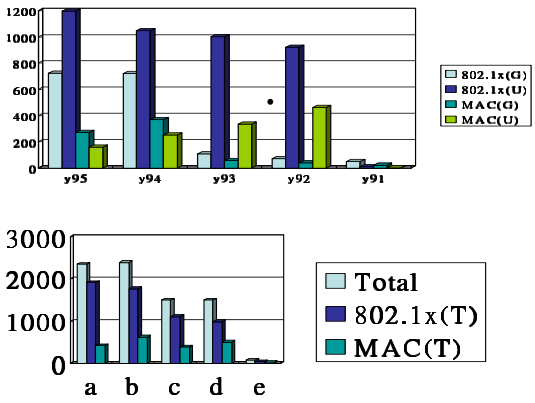


Fig. 8: Layer-2 Authentication Statistics of the NCTU DormNet by Student Identities (i.e., Enrolling School Year)

Legend 1

a:y95(i.e., 2006), b:y94(i.e., 2005), c:y93(i.e., 2004), d:y92(i.e., 2003), e:y91(i.e., 2002)

Legend 2

- **802.1x (T) = 802.1x(U) + 802.1x (G)**
- **MAC (T) = MAC (U) + MAC (G)**
- **U: undergraduate, G: graduate**

Note: The result of the addition of the two numbers is not equal to number of the applicants of the Dorm (i.e., 6827) since some people might have more than one computer installed and many did not have their computers registered (but with 802.1x-based authentication capability enabled).

4.4 Discussions

There are many things that need improving. First, DNS works in a hierarchical and distributed (e.g., temporal and spatial) manner. A typical site might have several independent advertising and/or recursive DNS servers for serving incoming and outgoing queries (e.g., two for the former and another three for the latter) about the forward and their corresponding reverse domain zones independently. Hence, by nature, a federated anomaly detection scheme will collect information and generate the eradicating responses in a more effective and timely manner. Second, as shown in Table 4, many design and deployment issues should be addressed from both the performance and anomaly detection considerations.

Moreover, as shown in Fig.9, in some critical cases,

it is often hard to identify many real compromised hosts. For example, in the campus network, many people (e.g., using a portable notebook PC) do not register his/her computer with a fixed IP address for Internet access (e.g., using DHCP+NAT+VPN). Through such a series of address mapping, it is often very hard to identify many real compromised hosts in a timely manner.

For coping with these to support the intrusion eradicating process, in the future, we would try developing a federated model of network anomalous detection and intrusion eradication scheme, combining the DNS-based anomaly detection and IEEE 802.1x-based authentication scheme.

Table 4: Suggestions to improve DNS and related network performance

1.	Separation of DNS resolving servers from DNS advertising servers (e.g., two for the former and another three for latter). • Avoid outsiders' Dos attacks	Security and anomaly detection considerations
2.	Build special DNS resolving servers for the separation of heavy-loaded clients (listed below) from ordinary clients. • Mail servers, proxy servers, etc. • Experimental learning systems (e.g., for network programming courses)	Availability and anomaly detection considerations
3.	Disable DNS forwarding services on normal DNS resolving servers with well-connected Internet environments.	DNS forwarding • only for inside firewalls
4.	Implement DNS filtering to facilitate anti-spam • chineseservice.biz, newshome.info, etc.	Block Spam with domain names on Free DNS services

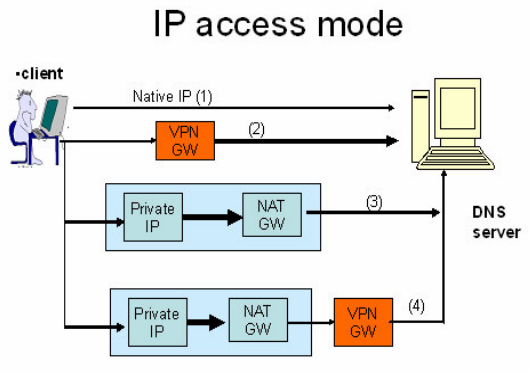


Fig.9: Complexity of the identification of suspected problem source

5 Concluding Remarks

In this paper, we design and implement a DNS-based network anomalous detection and intrusion eradication scheme, combining the DNS-based anomaly detection and IEEE 802.1x-based authentication scheme for supporting the intrusion eradicating process. Future research will focus on extending the design and implementation to be a federated model of network anomalous detection and intrusion eradication scheme for supporting the intrusion eradicating process in a timely manner.

References

[1] Albitz, P. and Liu, C. (2001). DNS and BIND 4th edition, O'Reilly & Associates, Inc., Sebastopol, CA, 2001

[2] Botnet, from the free encyclopedia Wikipedia - http://en.wikipedia.org/wiki/Botnet, Accessed Jul 10, 2006.

[3] Chen, C.S.,Tseng, S.S., Liu, C.L.(1999), "A Distributed Intrusion Detection Model for the Domain Name System", in Special Issue on Parallel and Distributed Systems, Journal of Information Science and Engineering, Vol.18, pp.999-1009

[4] Chang-Sheng Chen (陳昌盛), Zheng-Guo Chen(陳政國), Chin-Shiuan Chang(張晉璿), Shian-Shyong Tseng(曾憲雄), "DNS Knowledge-based Two-Phase Network Anomaly Detection Scheme", in Proceedings of the 16th information security conference (ISC2006), June 8-9, 2006, Taichung, Taiwan, NSC94-2213-E-009-111.

[5] *IEEE,* "802.1X - Port Based Network Access Control", http://www.ieee802.org/1/pages/802.1x.html

[6] Microsoft, "*IEEE 802.1x Authentication Client in Microsoft Windows for Wireless and Wired Networks*", http://www.microsoft.com/wifi/.

[7] Mockapetris, P., "Domain Names - Concepts and Facilities," RFCs 1034, November 1987.

[8] Rigney, C., Willens, S., Rubens, A. and W. Simpson, "Remote Authentication Dial In User Service (RADIUS)", RFC 2865, June 2000.

[9] Roger Clarke, "*Peer-to-Peer (P2P) - An Overview*", accessed on March 2, 2006, http://www.anu.edu.au/people/Roger.Clarke/EC/P2POview.html.

[10] T.Lunt, "A Survey if Intrusion Detection Techniques", Computer and Security, vol. 12, no.4, June 1993, pp.405-418.