行政院國家科學委員會專題研究計畫　成果報告

以 DNS 系統為核心的網路異常偵測系統之研究

計畫類別：個別型計畫
計畫編號：NSC94-2213-E-009-111-
執行期間：94 年 08 月 01 日至 95 年 07 月 31 日
執行單位：國立交通大學計算機與網路中心

計畫主持人：陳昌盛

報告類型：精簡報告

處理方式：本計畫可公開查詢

中　華　民　國 95 年 10 月 30 日

行政院國家科學委員會補助專題研究計畫 ■ 成 果 報 告
□期中進度報告

以 DNS 系統為核心的網路異常偵測系統之研究

計畫類別：■ 個別型計畫　　□ 整合型計畫
計畫編號：94-2213-E-009-111
執行期間：94 年 8 月 1 日至　95 年 7 月 31 日

計畫主持人：陳昌盛助理教授
計畫參與人員：陳政國、鄭中樑、王向榮

成果報告類型(依經費核定清單規定繳交)：■精簡報告　□完整報告

本成果報告包括以下應繳交之附件：
□赴國外出差或研習心得報告一份
□赴大陸地區出差或研習心得報告一份
□出席國際學術會議心得報告及發表之論文各一份
□國際合作研究計畫國外研究報告書一份

處理方式：除產學合作研究計畫、提升產業技術及人才培育研究計畫、列管計畫及下列情形者外，得立即公開查詢
□涉及專利或其他智慧財產權，□一年□二年後可公開查詢

執行單位：國立交通大學計算機與網路中心

中　華　民　國　95　年　10　月　28　日

以 DNS 系統為核心的網路異常偵測系統之研究

A study of DNS-based network anomalous detection scheme

計畫編號：94-2213-E-009-111
執行期限：94 年 8 月 1 日至 95 年 7 月 31 日
主持人：陳昌盛助理教授 國立交通大學計算機與網路中心
計畫參與人員：陳政國、鄭中樑、王向榮

一、中文摘要

在目前多數的網路應用基本模式(例如, e-mail, 網頁瀏覽等), 通常會先進行網址查詢(DNS[3]查詢), 再連上網站。因此, 從觀察網路應用連結到 DNS 流量分佈的關聯, 可以發掘許多非常有趣且實用的資訊[1][2] (例如，突然出現大量病毒信件散佈，通常也會伴隨造成大量相關的 DNS 查詢),作為判讀網路管理趨勢的重要參考依據。本計畫中，主要的構想是希望建構一套以網域名稱系統(DNS)查詢流量為檢測核心, 兩段式網路異常偵測系統，並結合 IEEE 802.1x 認證系統[8][10], 進一步建構成為網路異常偵測與入侵者根除系統，來輔助系統管理者，早期發現疑似網路異常者，介入調查，儘快確認異常狀況，並加以排除，有效降低許多因為系統漏洞被利用的機器而不自知者, 所造成對單位網路的不良衝擊, 以免事態進一步擴大。

二、英文摘要(Abstract)

Nowadays, most Internet services are based on the working model that there will be some Domain Name System (DNS) [3][11] queries before the communication activities. Thus, for supporting DNS-based anomaly detection, the key problem is how to identify the clusters (sequences) of inappropriate DNS queries form the DNS traffic mixture that are directly generated or indirectly induced by internetworking hosts that are abnormal (i.e., including compromised and/or the original abusers) [1][2]. In this project, we propose an offline DNS-based, Two-phase Network Anomaly Detection Scheme. Based on the analysis of DNS query logs and followed by a field study to assert the identification of these threats, we design and implement a DNS-based network anomalous detection and intrusion eradication scheme, combining the DNS-based anomaly detection [6][7] and IEEE 802.1x-based[8][10] authentication scheme 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.

三、計畫緣由與目的

在目前多數的網路應用中(例如, e-mail, 網頁瀏覽等), 通常會先進行網址查詢(DNS 查詢), 再連上網站。我們觀察到,目前的網路環境裡, 普遍存在有一個相當大的癥結, 就是目前廣大的一般用戶, 由於系統安全知識不足, 因此所使用的機器, 通常都是處在沒有保護(例如, 沒有安裝防毒系統, 防火牆, 入侵偵測等資通安全系統)或低度保護(雖有安裝前述系統, 但不常更新系統安全防護資訊, 例如定期檢測及更新病毒碼), 往往也因此而導致電腦中毒, 或者被入侵而不自知, 以致轉而被用來散佈垃圾信件、嘗試入侵或攻擊其他網路系統, 導致相關單位的網路效能大受影響。

其次, 從實務的管理角度來看, 普通的入侵偵測系統(IDS) [13], 基本設計原理主要放在一般網路應用的漏洞蒐集以及網路流量的統計分析、比對。雖然有一些系統有進行所謂的使用行為分析 (usage pattern analysis), 但基本上是以該網路應用(如 e-mail 等)為中心主體, 沒有將 DNS 查詢流量分析以及 access pattern 與網路異常偵測, 進一步結合起來。忽略了從網路整體使用的觀點來看, 諸如此類應用所造成的大量 DNS 查詢, 實質上已經成為網路異常的徵兆。 因此, 從觀察網路應用連結數量多寡到 DNS 流量的分佈的關聯, 可以發掘許多非常有趣且實用的資訊

(例如，突然出現大量病毒信件散佈，通常也會伴隨造成大量相關的 DNS 查詢)，作為判讀網路管理趨勢的重要參考依據。

本計畫中，主要的構想是提出以網域名稱系統(DNS) 查詢流量為檢測核心，再搭配運用其他網路系統知識以及記錄分析，以期能建構一套兩段式網路異常偵測系統，並結合 IEEE 802.1x 認證系統，進一步建構成為網路異常偵測與入侵者根除系統，來輔助系統管理者，早期發現疑似網路異常者，介入調查，降低對單位網路的不良衝擊，以免事態進一步擴大。

四、想法與討論

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 Figure 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 latter) about the forward and corresponding domain zones.

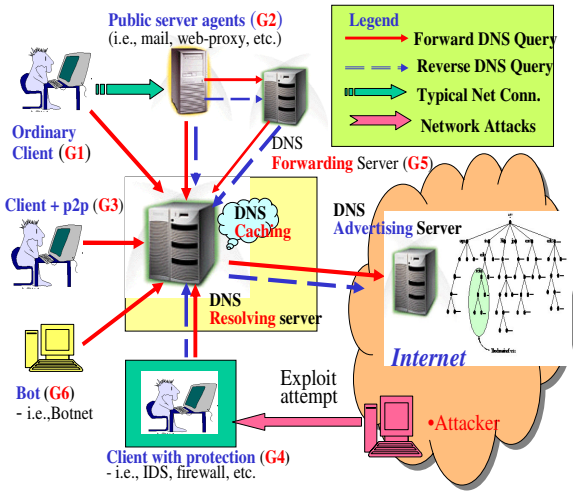


Figure 1: A simple classification scheme of DNS clients and servers.

In practice, however, there are anomalies or network abusing attempts located in the midst of normal DNS activities from time to time. 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

Case	Description
1.	Botnet [4] 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 [9]

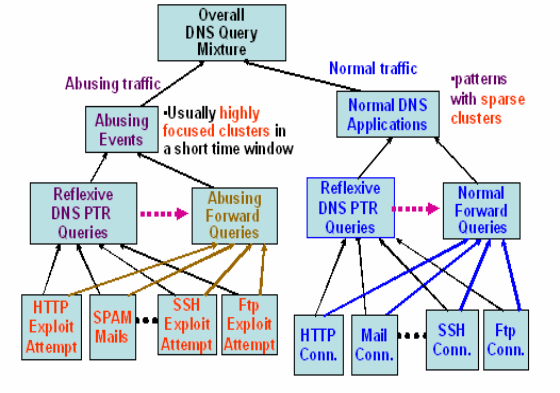


Figure 2: Mixture of DNS queries

As mentioned in [6][7], 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 Figure 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.

In practice, as shown in Table 2, most DNS queries are conducted on some major

hosts. For example, as shown in Table 2, 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 [12] 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.

Moreover, as shown in Figure 1, both normal (e.g., category G2- mail transfer agents, etc.) and abnormal (e.g., category G6 - botnet, or virus/spam engine, etc.) DNS clients could usually produce huge amounts of DNS queries (resolving) in a specific time interval. Therefore, it is often hard to conduct the network anomaly detection by using a straightforward statistical-based approach on DNS queries sequences alone.

■ Methodology and System Architecture

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. The framework for our proposed DNS Knowledge-based, Two-phase Anomaly Detection Scheme is depicted in Figure 3.

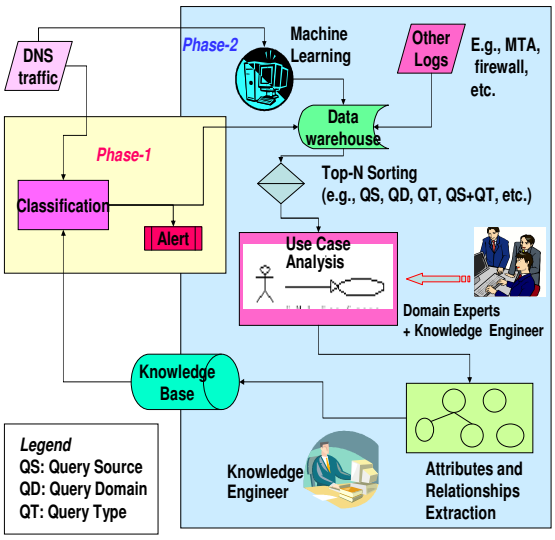


Figure 3: Proposed System Architecture

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.

The general idea is as follows. In Phase 1, we deal with the problem by trying to detect the misuse (e.g., topmost, repeated unresolved) to help identify the network anomalous activities in the early phase. Next, in Phase 2, we will further try identifying the problem types and sources (e.g., locating the virus-affected or compromised hosts) for fixing, by combining the log analysis of access patterns of the DNS and other network applications (e-mail, web, etc.). The detail of the two-phased algorithm is shown as follows:

■ DNS Knowledge-based Two-phase Anomaly Detection Algorithm

♦ Phase 1 - DNS Knowledge-based Two-phase Anomaly Detection Algorithm

- Input: DNS traffic (querylog, tcpdump trace, etc).
- Output: Network anomaly candidates

and/or alarm.

Step1. DNS query log cleansing – **Identify and prepare available data sources.**

Step2. Generate DNS Statistics (within specified intervals) on DNS resolving Servers

Step2.1: Generate *Top-N* lists of sender IP-based statistics: by number of total (a) forward queries, (b) reverse queries, (c) forward + reverse queries, etc.

Step2.2: Generate *Top-N* groups of query domain-based statistics (i.e., by sender IPs) : (a) forward queries (b) reverse queries (c) forward and reverse queries.

Step3. Misuse detection (by classification) and anomalous candidate information gathering

Step3.1: Use the stored patterns (rules) to locate possible network abusing sources.

Step3.2: Use the statistical data generated in Step2 to locate candidate sources for further processing in Phase 2 (i.e., by putting them into the data warehouse).

Step4. Send alarm messages to the responsible people if there are patterns matched.

♦ Phase 2 - DNS Knowledge-based Two-phase Anomaly Detection Algorithm

- Input: DNS querylog and the set of anomaly candidate (sender IP + rules) in Phase 1
- Output: Sources of network anomaly and rules for updating the knowledge base

Step1. Initiates Log Data Clustering

Step1.1: Group these queries by selecting and running data clustering algorithm(s) bases on the characteristics (sender IP-based, forward query-based, etc.) of the data.

Step1.2: Use the selected techniques to uncover hidden structure in the data.

Step1.3: By the help of domain experts, identify the target clusters (e.g., groups with member large than some threshold limit) for further processing.

Step1.3.1: White-list candidates: server groups of heavy-loaded DNS clients (mail server, web proxy server, etc.).

Step1.3.2: Black-list candidates: other anomaly candidates of heavy-loaded clients for further processing (virus/worm, botnet, p2p, etc.).

Step2. Build and refine the DNS knowledge-based network anomaly ontology.

Step2.1: Build/refine the skeletal concept model of the ontology by following a top-down brainstorming method.

- Interview with DNS domain experts (or read the DNS books, etc.) for building and/or refining the DNS knowledge-based network anomaly ontology.

Step2.2: Conduct the attributes and relation extraction.

- Analyze and decompose the forward query and reverse cases into small components (BOTNET, open proxy, etc.).

Step2.3: Define or identify the relationships between the specified cases.

Step2.3.1: Remote login exploit (e.g., SSHd, Telnetd, Ftpd, etc.; dictionary attacks).

- Excessive PTR queries on DNS advertising server + remote login log entries from other remote network applications (e-mail, web, etc.).

Step2.3.2: BOT (members of bonet): e.g., excessive forward queries (e.g., A-RR such as "iownyourmon.info") on normal clients

Step2.3.3: Virus/open mail proxy: e.g., excessive forward queries (MX-RR and A-RR, etc.) on normal clients or excessive PTR queries on DNS advertising server and log entries from e-mail systems.

Step2.3.4: Other cases: excessive forward queries and/or reverse queries (e.g., possibly due to configuration errors) that could not be classified into any of the above cases.

Step2.4: After experts' verification, the ontology is constructed to cover DNS-based network anomaly detection knowledge.

Step3. Initiates anomaly-recognition process and keep the IP list of matched cases

Step3.1: Conduct analysis based on network anomaly types (BOTNET, open mail proxy, virus/worm, etc.).

Step3.2: Conduct analysis based on Server Types (DNS advertising server, resolving server, hybrid server, etc.).

Step4. Send alarm messages to the responsible people if there are cases matched.

■ Implementation of DNS-based Anomaly Detection System

Table 3: System Implementation Environment

Item	Description
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

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, Figure 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***").

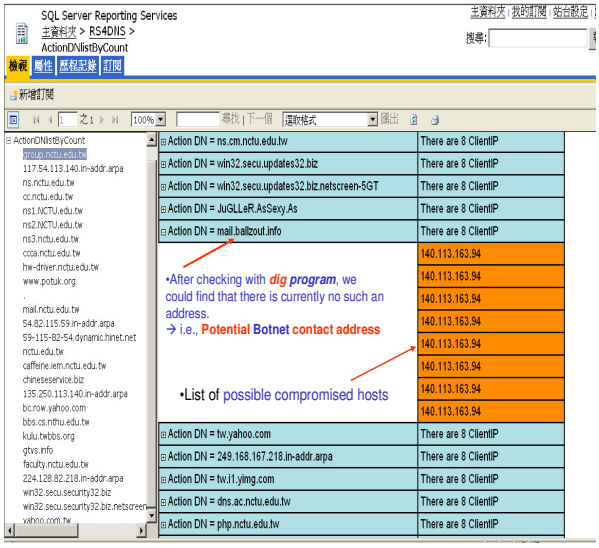


Figure 4: Identification of IP/host lists possible compromised (e.g., BOTNET)

♦ Supporting an IEEE 802.1x-based Authentication System

Next, as shown in Figure 5, 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.

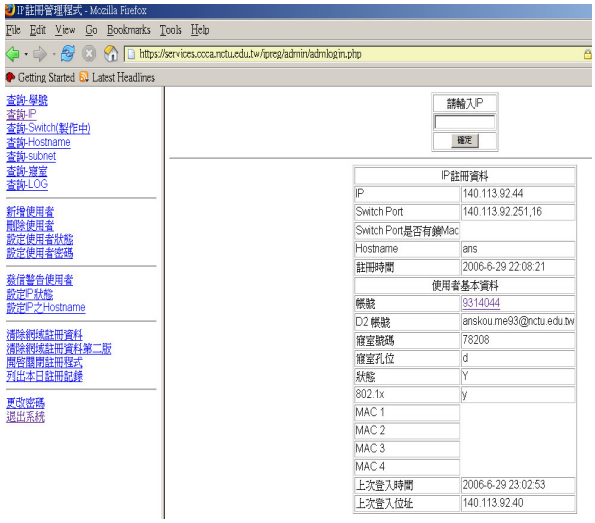
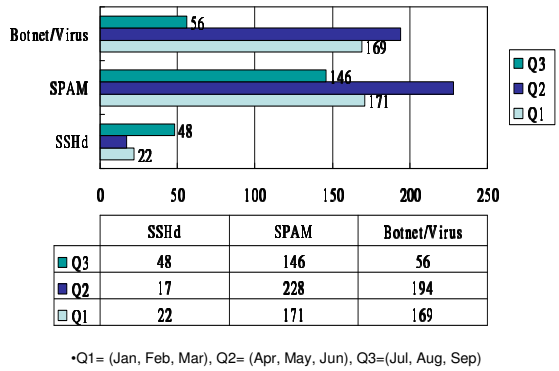


Figure 5: DormNet IP Registration/Query System

■ Typical Anomalous Example Cases

Figure 6 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.



	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)

Figure 6: Accumulative Statistics of NCTU IP/hosts abusing events from 2006.01 to 2006.09.

五、初步計畫成果自評

本計畫中，我們提出以 DNS 查詢流量分析為檢測核心，根據所建構出來的 DNS-based Network Anomaly Model (Figure 3) 以及收集常見的 DNS 問題分類之後(如 Table 1)，搭配相關網路應用記錄開始設計並建構一套 DNS-based 兩段式異常偵測系統為本，並結合 IEEE 802.1x 認證系統，進一步建構成為網路異常偵測與入侵者根除系統， 來輔助系統管理者，早期發現疑似網路異常者，介入調查，儘快確認異常狀況，並加以排除。基本上，我們已經達成當初計畫的預期目標，這個系統的確有助於降低受創系統 (也就是，有系統漏洞被利用的機器而不自知者) 對單位網路所可能造成不良衝擊，以免事態進一步擴大。

目前本計畫的研究成果已經有相當斬獲，共計發表了一篇期刊論文[5]，以及兩篇會議論文[6][7]，後續我們還將進一步進行 DNS 與其他應用系統流量的統整分析，以期發展成為更完整的 DNS-based 網路異常偵測與入侵者根除系統。最後，感謝此計劃的推動及補助，能讓該研究相關領域有更進一步的探討及進展。

五、參考文獻

[1]. 陳昌盛, “DNS 系統的異常流量偵測、管理以及除錯”, TANet 2001 會議論文輯

[2]. 陳嘉玫, 黃世昆, 鄭進興, “DNS 伺服器安全檢測與防護體系之建置與運作”, 台灣網路資訊中心 (TWNIC) 91 年度專案委託研究計劃

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

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

[5]. Chien-Liang Liu, **Chang-Sheng Chen**, Shian-Shyong Tseng, “**DNS Ontology-based Search Service**”, Journal of Internet Technology, Vol.7 No.3 July 2006 , **NSC94-2213-E-009-111**.

[6]. Chang-Sheng Chen (陳昌盛), Zheng-Guo Chen (陳政國), Chin-Shiuan Chang(張晉瑋), 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.

[7]. **Chang-Shang Chen(陳昌盛),** Shang-Rung Wang (王向榮)**,** Ta-Chung Liu (劉大川), "DNS-based Network Anomaly Detection and Eradicating Scheme", accepted for publish, to appear in TANet 2006, **NSC94-2213-E-009-111.**

[8]. IEEE, “802.1X - Port Based Network Access Control”, http://www.ieee802.org/1/pages/802.1x.ht

ml

[9]. Koh, J.L. (2001). Recent Developments and Emerging Defenses to D/DoS: The Microsoft Attacks and Distributed Network Security. SANS Institute, URL: http://www.sans.org/infosecFAQ/DNS/developments.htm.

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

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

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

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