트래픽 분석을 통한 IRC 봇넷과 P2P 봇넷의 특성 비교

김유승*, 최현상*, 정현철**, 이희조*

Comparison of Characteristics between IRC Botnet and P2P Botnet by Traffic Analysis

Yu-seung Kim*, Hyunsang Choi*, HyunCheol Jeong** and Heejo Lee*

본 연구는 지식경제부 및 정보통신연구진흥원의 IT핵심기술개발사업의 일환으로 수행하였음.
[2008-S-026-01, 신종 봇넷 능동형 탐지 및 대응 기술]
본 연구는 지식경제부 및 정보통신연구진흥원의 대학IT연구센터 지원사업의 연구결과로 수행되었음.
(IITA-2008-C1090-0801-0016)
본 연구는 교육부 BK21 사업의 지원을 받아 수행되었음.

요 약

봇넷은 악성 코드에 의해 감염된 봇 호스트들로 이루어진 네트워크로서 봇 마스터라고 불리는 공격자에 의해 원격 조종되며 DDoS 공격을 비롯한 각종 공격을 수행하는데 이용되는 위협적인 존재이다. 본 연구에서는 봇넷의 행위로 추정되는 네트워크 트래픽을 수집하여 이를 분석함으로써 실질적인 관점에서 봇넷의 특성을 밝혀내고자 한다. 수집된 네트워크 트래픽은 IRC 봇넷과 P2P 봇넷으로써 전자의 경우 상대적으로 C&C 서버의 위치 탐지가 용이한 반면, 후자에서는 여러 가지 회피 기술로 인해 트래픽 분석만으로는 봇넷의 구조를 파악하기 어렵다는 점을 확인하였다.

Abstract

Botnet is the significant threat in the internet, which is the network of bot hosts infected by malware and is controlled by remote attacker called bot master. In this paper, we reveal the characteristic of botnet by gathering and analyzing the suspected network traffic. It is estimated that the observed traffic is from IRC botnet and P2P botnet We reveal that it is easy to detect the C&C server in the former case, while it is difficult to grasp the structure of the latter botnet with only traffic analysis.

Key words
Botnet, DDos Attack, Traffic Analysis

* 고려대학교 정보통신대학 컴퓨터·전파통신공학과
** 한국정보보호진흥원
· 제1저자 (First Author) : 김유승, 교신저자 (Corresponding Author) : 이희조
· 접수일자 : 2008년 10월 09일, 수정일 : 1차 - 2008년 12월 09일, 게재확정일 : 2009년 02월 17일

Ⅰ. 서 론

최초 IRC 기반의 네트워크에서 동작하는 봇넷이 발견된 이후, 10년이 채 안되는 기간 동안 봇넷은 HTTP를 사용하여 정상 트래픽과의 구분을 어렵게 하거나 peer-to-peer (P2P) 기반의 네트워크 상에서 동작하여 기존의 중앙 집중형 명령 전달 구조를 분산형 구조로 바꿔 생존력을 높이는 등의 기술을 이용하는 데까지 발전하고 있다. 이외에도 실행압축기술, 코드자가변경, 명령채널의 암호화 등의 다양한 기법을 사용하여 탐지 및 회피를 어렵도록 그 공격 기술이 교묘해지고 있는 실정이다. 한편, 최근 넷봇(Netbot)과 같이 전문 해커 그룹에 의해 제공되는 유저 인터페이스를 통해 전문적인 지식이나 기술이 없는 일반인들도 쉽게 봇 코드를 생성하거나 제어할 수 있어 위험성이 더욱 부각되고 있다. 2008년에는 ASPROX 봇넷에 의해 Mass SQL Injection 공격이 소개가 되었는데 이는 MS-SQL의 보안 취약성을 이용해 해당 제품을 사용하는 서버를 통해 악성 코드를 확산되도록 하여 최근 이슈가 되고 있다. 그 외 보안 전문 그룹에 의해 발표된 보고서들에서도 봇넷의 규모, 피해 양상에 대해 뚜렷한 증가 추세를 나타내고 있다[1]-[2].

이러한 일련의 흐름들에 발맞추어 봇넷을 탐지하거나 회피하기 위한 연구들이 활발하게 진행되고 있다. 국내에서도 규모를 가리지 않고 대형 웹사이트에서부터 중소 웹사이트에까지 점차 봇넷을 이용한 DDoS 공격을 시도하고, 공격 중단의 대가로 금품을 요구하는 사례들이 보고되고 있어 심층적이고도 실제 적용 가능한 관련 연구의 필요성이 시급히 대두되고 있다.

본 논문에서는 교내망에서 IRC 기반의 봇넷이 발생하는 트래픽을 수집하고 이를 통해 봇넷의 특성을 파악한 기존의 연구[3]를 확장하여 허니넷(Honeynet)을 이용하여 P2P 기반의 네트워크에서 동작하는 봇넷으로 유명한 스톰봇(Storm Bot)의 트래픽을 수집하였고, 이에 대한 분석을 통해 봇넷들의 특성을 파악하고자 시도하였다. 통상적으로 봇넷을 분석하는 기법에는 감염 대상 호스트 상에 잠입한 봇 실행코드를 역공학(Reverse engineering) 기법을 통해 그 동작행위를 밝히는 방법과 네트워크 차원에서 봇넷이 유발하는 트래픽을 면밀히 살펴보는 방법이 있다. 본 논문에서는 후자의 방법에 초점을 맞추어 트래픽의 전송 패턴과 패킷의 페이로드에서 봇넷의 행위를 유추해보는 방식을 사용하였다. 본 연구를 통하여 실질적인 봇넷의 트래픽 특성을 파악하고 이에 대한 탐지 및 회피 기법을 개발하는데 기반이 되는 연구 자료로서 이용될 수 있을 것이다.

본 논문의 2장에서는 IRC 봇넷의 특성에 대하여 분석하고, 3장에서는 P2P 봇넷의 특성에 대해 살펴보도록 한다. 4장에서는 두 가지 봇넷에 대하여 비교 결과를 제시하고, 마지막으로 5장에서 결론과 함께 향후 전망에 대해 언급한다.

1.1 용어 정의

봇 - 로봇 (Rotot)에서 유래된 것으로 악성 프로그램에 의해 감염되어 외부 호스트로부터 원격으로 제어를 받는다. 봇 호스트와 동일한 개념이다.

봇넷 - 봇들로 이루어진 네트워크로서 봇 마스터에 의해 조종되며 봇넷의 감염 확대, DDoS 공격, 피싱 및 스팸, 개인 정보 수집, 불법 데이터 공유와 같은 악성 행위들을 수행한다.

봇 마스터 - 봇넷을 조종하는 역할을 하는 공격자의 호스트이다.

봇넷 C&C - 봇 마스터가 봇넷을 조종하기 위해 명령 및 제어를 내리는 채널을 의미한다.

봇 코드 - 호스트를 감염시켜 봇으로 만들기 위한 실행코드 자체를 의미한다.

IRC 봇넷 - 봇넷 C&C로 IRC(Intrenet Relay Chat) 서버를 이용하는 봇넷으로서 중앙 집중형의 구조를 가진다. 구현이 용이하고 규모를 확대하기 쉽다는 장점이 있어 초기 봇넷에서부터 많이 사용되어오던 방식이다. 하지만, 봇넷 C&C 서버만 탐지해내면 봇넷을 쉽게 무력화 시킬 수 있다는 단점이 있다. GTbot, SDbot, Agobot, Netbot 등이 유명하다.

HTTP 봇넷 - IRC 봇넷이 특정 포트를 사용하기 때문에 해당 포트 필터링을 통해 쉽게 드러나는데 이를 보완하기 위해 HTTP를 통신 프로토콜로 사용

하여 정상 HTTP 트래픽과의 구분을 어렵게 한다. Bobax와 같은 봇이 해당된다.

P2P 봇넷 - P2P 프로토콜을 사용하여 봇넷 C&C 채널을 구성하는 봇넷으로 2007년 발견된 스톰봇(Storm bot)이 대표적이다. 이것은 Kademlia 알고리즘을 구현한 Overnet 상에서 동작한다.

Ⅱ. IRC 봇넷 트래픽 분석

2.1 트래픽 특성

IRC 봇넷의 C&C 서버 역할을 하는 것으로 의심되는 호스트로부터 약 50분에 걸쳐 수집된 트래픽을 대상으로 분석하였다. 이들 패킷의 전반적인 특성은 다음과 같다.

총 수집 패킷 개수 : 25,094 개
평균 초당 전송 패킷 수 : 8.794 개/sec
평균 패킷 크기 : 199.210 Bytes
평균 초당 전송 바이트 : 1751.92 Bytes/sec
평균 초당 전송 비트 : 13.69 Kbit/sec

패킷들의 평균 크기는 약 200 Bytes로 나타났는데 분포를 세밀히 살펴보면 40~159 Bytes 사이의 패킷이 거의 90% 가까이 차지하고 있었다. 프로토콜 별로는 거의 대부분이 TCP로 이루어져 있었고, 그 중의 절반 정도는 HTTP로 이루어져 있었다.

2.2 분석 결과

먼저 IRC 봇넷의 특성상 봇넷 C&C 서버를 중심으로 중앙 집중형의 통신 구조를 가지기 때문에 해당 서버와 빈번하게 통신이 이루어지는 호스트들을 대상으로 패킷의 분포를 조사하였다. 그러나 대량의 트래픽이 발생한 것으로 확인된 구간들 중 일부에서는 이들 호스트들과 통신한 내용이 확인되지 않았다. 그림 1에서와 같이 300초에서 1000초 이후의 구간이 이에 해당하는데 이것은 해당 구간에서 봇넷 C&C 서버가 특정한 IP를 가진 외부의 호스트와 1:1로 통신하는 것이 아니라 여러 호스트들과 1:N의 통신 형태를 가지기 때문인 것으로 해석할 수 있다. 패킷들의 페이로드를 통해 이들은 주로 8080 포트를 사용하는 TCP 패킷들로 이루어져 있음을 확인하였다. 한편, 1100초 구간에서 대량의 트래픽이 확인되고 있는데 이는 80 포트를 사용하는 TCP 패킷들로서 트래픽 혼잡에 의해 발생되는 재전송 요청 및 답신 패킷들이 주를 이루고 있었다.

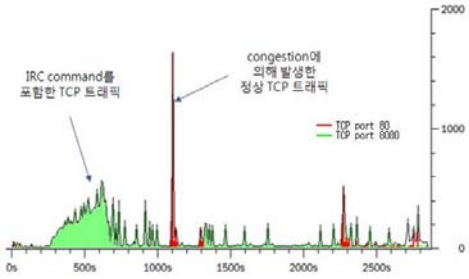

그림 1. 사용 포트별 패킷 분포
[x축 : 시간(초), y축 : 초당 전송 패킷 수(pps)]
Fig. 1. Distribution of packets by used ports
[x axis : second, y axis : packets per second]

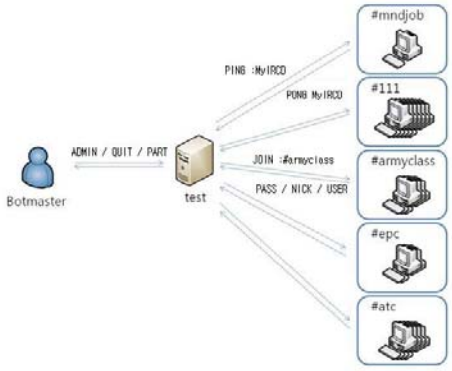

그림 2. 트래픽 분석을 통해 밝혀진 IRC 봇넷의 구조
Fig. 2. Structure of IRC botnet revealed by traffic analysis

8080 포트를 사용하는 패킷들을 살펴본 결과 IRC 명령을 다수 포함하고 있어, 8080 포트를 봇넷 C&C 채널로 사용하는 IRC 봇넷의 동작으로 추정되었다. 패킷들의 세부 내용을 분석 및 종합한 결과, 그림 2와 같이 1개의 봇 마스터와 봇넷 C&C 서버 역할을 하는 1개의 IRC 서버를 발견하였고, 5개의 채팅방과 함께 112대의 감염된 호스트가 이들 채팅방에 접속한 것을 확인하였다. 외부의 봇 마스터는 봇넷 C&C서버를 통해 채팅 서버를 관리하고 감염

된 봇들에게 시스템 폴더 및 로그 폴더 검색 등의 악성 행위를 수행하는 것으로 나타났다.

Ⅲ. P2P 봇넷 트래픽 분석

3.1 트래픽 특성

P2P 봇넷은 많은 경우 일반 P2P 프로토콜을 사용하며 패이로드를 암호화 하고 있기 때문에 일반적인 필터링 룰을 이용하여 트래픽을 수집하는 것이 쉽지 않다. 따라서 허니넷(Honeynet)을 이용하여 대표적인 P2P 봇넷으로 알려져 있는 스톰봇(Storm bot)을 내부에 유도한 후 이로부터 유발되는 트래픽을 수집하였다. 트래픽 수집은 약 2시간 50분에 걸쳐 이루어졌으며 상세한 내용은 다음과 같다.

총 수집 패킷 개수 : 32,323 개
평균 초당 전송 패킷 수 : 3.196 개/sec
평균 패킷 크기 : 64.519 Bytes
평균 초당 전송 바이트 : 206.209 Bytes/sec
평균 초당 전송 비트 : 2.0 Kbit/sec

수집된 대부분의 패킷인 32,310개가 UDP로 이루어져 있었다.

3.2 분석 결과

수집된 전체 패킷은 모두 트래픽을 수집한 호스트의 IP 주소를 송신지 또는 수신지로 하여 UDP 포트로 3858 번을 사용하고 있었다. 또한, 33 개의 외부 호스트들이 트래픽 수집 대상 호스트와 100 개 이상의 패킷을 주고받았으며, 나머지 3,532 개의 호스트들과는 100 개 이하의 패킷만을 주고받았고 이들 중 상당수는 1개의 단방향 패킷을 주거나 받았다. 또한, 대부분은 UDP 패이로드 기준으로 25 Bytes 이하의 짧은 패킷들로 구성되어 있었다.

스톰봇은 Overnet 프로토콜 중 일부 명령어 집합을 사용하기 때문에 수집된 패킷들을 명령어 종류에 따라 분류하였다[4]. 명령어는 UDP 패이로드의 첫 바이트가 0xe3의 값을 가지며 두 번째 바이트의 내용에 따라 표 1과 같이 종류가 결정된다. 수집된 패킷들은 모두 UDP 패이로드의 첫 번째 바이트가 0xe3으로 이루어져 있었다.

표 1. Overnet 명령어 코드
Table 1. Overnet commands

명령군	명령어	명령 코드	발견된 개수	비고
제어명령군	Publicize	0x0c	13,484	접속 관련 명령어
	Publicize ACK	0x0d	9,960	
	Connect	0x0a	471	
	Connect Reply	0x0b	21	
전송명령군	Search	0x0e	6,063	데이터 검색 관련 명령어
	Search Next	0x0f	876	
	Search Info	0x10	558	
	Search Result	0x11	594	
	Search End	0x12	295	
	Publish	0x13	1	데이터 게시 관련 명령어
	Publish ACK	0x14	0	

스톰봇의 명령어 별로 패킷을 분류해 보면 트래픽을 수집한 호스트가 외부 호스트들에게 자신을 알리는 Publicize 명령과 이에 따르면 응답 메시지들이 전체 트래픽의 약 72% 정도를 차지하고 있다. 그 뒤를 이어 트래픽 수집 호스트가 외부 호스트들에 대해 검색을 시도하는 Search 명령과 이에 대한 다이얼로그 메시지들이 큰 비중을 차지하고 있다. 그림 3은 시간대별로 각 명령어 종류에 따른 트래픽 양의 추이를 보이고 있다.

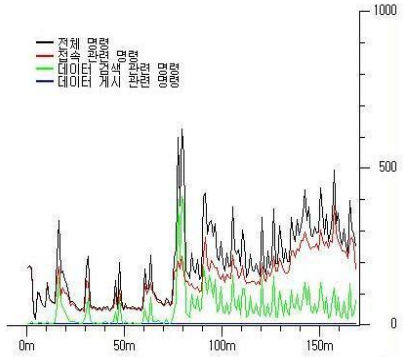

그림 3. 명령군별 트래픽 추이
Fig. 3. Traffic transition by command groups

전체적으로 접속 관련 명령이 주를 이루고 있고, 다음으로 데이터 검색 관련 명령들이 많이 관찰되는데 특히 수집 시작시간 이후 80분대에서 대량의 검색 시도가 이루어지고 있음을 알 수 있다. 데이터 게시 관련 명령들은 거의 관찰되지 않았다.

그림 4는 접속 관련 명령어들의 사용빈도가 시간에 따라 변동하는 추이를 나타내고 있다. 전체적으로 스톰봇 감염 호스트가 자신을 외부 호스트에게 알리고 있고, 수집 시간 초반에 접속 시도가 활발히 이루어지고 있음을 알 수 있다. 그러나 초반 10분간 실제 성공한 접속의 수는 6개로 극히 적었다.

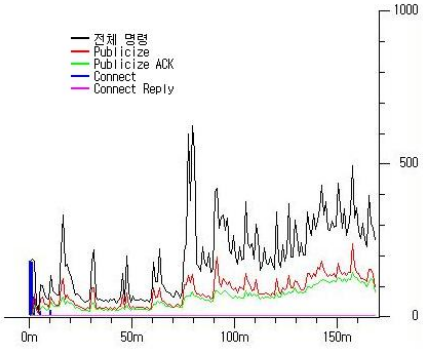

그림 4. 접속 관련 명령어들의 트래픽 추이
Fig. 4. Traffic transition by commands related to connection

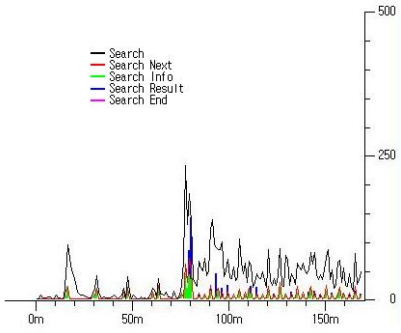

그림 5. 데이터 검색 관련 명령어들의 트래픽 추이
Fig. 5. Traffic transition by commands related to search

데이터 검색 관련 명령어들의 사용 추이를 자세히 살펴보면 앞에서 언급한 바와 같이 80분 경과 전후 시점에서 그림 5와 같이 활발한 검색 명령들이 관찰된다. Search Info의 경우 약 250여개의 외부 호스트들에게 트래픽을 수집한 감염 호스트가 명령을 송신하고 있으며, Search Result의 경우에는 42개의 호스트들이 감염 호스트에게 명령을 송신하고 있었다.

여기에서 검색의 결과에는 “bcp://ipaddr:port”와 같이 위치 정보를 가진 메타 태크가 포함되어 있다.. 스톰봇은 이 위치 정보를 현재 날짜, 체크섬과 같이 해쉬하여 암호화한다[5].

한편, 수집된 트래픽에서는 봇넷의 악성 행위가 직접적으로 드러나지 않고 있기 때문에 세부적인 봇넷의 구조와 C&C 채널을 밝히기는 어렵다. 대신 봇 호스트들은 Publicize 명령을 사용해 Overnet 상에서 자신의 존재를 지속적으로 알리기 때문에 이를 통해 대략적인 봇넷의 크기를 유추해 볼 수 있다. 수집된 트래픽에서 관찰된 13,448 개의 Publicize 명령들이 총 3,565 개의 호스트에서 송신되었는데, 이들 중 1시간 이상 해당 명령을 지속적으로 송신하면서 누적 송신 개수가 60개 이상인 호스트들만을 대상으로 한다고 가정하면 약 47개의 호스트들이 봇넷에 연관되어 있는 것으로 추정해 볼 수 있다. 물론, 수집된 트래픽이 전체 봇넷의 일부라는 점, 패킷의 내용만으로 봇 호스트와 정상 Overnet 호스트를 구분하기 어렵다는 점에서 이러한 추정은 실제 값과 상당한 차이를 보일 수 있다.

Ⅳ. 비교 결과

수집된 IRC 봇넷과 P2P 봇넷의 트래픽에서 C&C 채널 상에서 봇넷이 송수신하는 메시지는 급격히 증가하는 트래픽의 양으로는 파악하기 어렵다. 이는 공격이 시작된 후의 DDoS 트래픽과는 달리 명령 및 제어에 관련된 트래픽의 양이 상대적으로 적기 때문에 일반 응용 프로그램이 유발하는 트래픽과 구분하기 어렵기 때문이다.

표 2에서 IRC 봇넷과 P2P 봇넷을 분석한 결과를 요약하였다. 표에서 보이고 있는 바와 같이 수집한 P2P 봇넷의 경우에는 봇이 재정의한 UDP 포트를 통해 일반 P2P 프로토콜을 사용하여 통신이 이루어진다. 이때, P2P 프로토콜은 봇넷의 위치정보를 암호화하여 봇 호스트들 간에 교환하는데 사용된다. 따라서 봇 호스트 내부에서 동작하는 봇 코드의 암

호화 단계에서부터 정적인 분석이 이루어지지 않으면 네트워크 트래픽 분석만으로는 그 구조를 밝히기 어려운 한계를 지니고 있다. 또한, 명령 및 제어를 내리는 C&C 서버 자체가 여러 개의 노드들로 분산되어 있고 이것이 프로토콜에 의하여 쉽게 이동할 수 있어 그 위치를 파악하는 것이 상당히 어렵다.

표 2. 분석한 IRC 봇넷과 P2P 봇넷의 트래픽 비교
Table 2. Comparison of traffic between IRC botnet and P2P botnet

	IRC 봇넷	P2P 봇넷
사용 프로토콜	TCP 기반의 IRC 프로토콜	UDP 기반의 Overnet 프로토콜
사용 포트	IRC 포트 또는 HTTP 포트로 고정되어 있음.	유동적임.
패킷 내용	직접적인 악성 행위 (호스트의 중요 폴더 검색 등)	봇넷의 위치 정보 교환
암호화	암호화 되지 않음.	필요에 따라 일부 암호화함.
C&C 서버 위치	한 곳으로 집중되어 있음. (탐지 쉬움.)	여러 노드에 분산되어 있음. (탐지 어려움.)

V. 결 론

봇넷은 중앙 집중형 구조에서 분산형 구조로 진화하며 다양한 회피 기술을 통해 탐지 및 대응을 더 어렵게 하고 있다. 본 논문에서는 실제 봇넷이 유발하는 네트워크 트래픽을 수집하여 중앙 집중형 구조인 IRC 봇넷과 분산형 구조를 지닌 P2P 봇넷을 분석하고 그 특성을 비교하였다. 특히, P2P 봇넷의 경우 네트워크 차원에서의 분석만으로는 세부 구조 및 메시지 내용을 밝히기 어렵기 때문에 봇 호스트 차원에서의 정적 분석이 병행되어야 할 필요가 있다.

추후 연구에서는 P2P 봇넷의 명령 및 제어 채널과 함께 직접적인 악성 행위를 수집하고, 감염 호스트 내부에서 봇 코드의 행위 분석을 통해 전체 봇넷의 구조와 이에 대한 대응책을 마련하고자 한다.

감사의 글

본 연구는 2008년도 지식경제부 및 정보통신연구진흥원, 한국과학재단, 교육부 BK21 사업의 지원에 의하여 이루어진 연구로서, 관계부처와 본 논문의 심사위원님들께 감사드립니다.

참 고 문 헌

[1] D. Turner, M. Fossl, E. Johnson, T. Mack, J. Blackbird, S. Entwisle, M. K. Low, D. McKinney, C. Wueest. Symantec Global Internet Security Threat Report, 2008

[2] Arbor Networks. “Worldwide Infrastructure Security Report”, 2007

[3] 김유승, 최현상, 김인환, 권종훈, 이희조, 봇넷 트래픽 특성 분석: 사례 연구, 제 30회 한국정보처리학회 추계학술대회, 제 15권, 제 2호, 2008.11, pp. 1429-1432

[4] P. Porras, H. Saidi, V. Yegneswaran, A Multi-perspective Analysis of the Storm (Peacomm) Worm, SRI International CSL Technical Note, Oct. 2007

[5] J. Stewart, Inside the Storm: Protocols and Encryption of the Storm Botnet, http://www.blackhat.com/presentations/bh-usa-08/Stewart/BH_US_08_Stewart_Protocols_of_the_Storm.pdf

저자소개

김 유 승 (Yu-seung Kim)

2002년 8월 : 고려대학교 컴퓨터학과 학사
2002년 8월 ~ 2008년 2월 : 삼성전자 정보통신총괄 선임연구원
2008년 3월 ~ 현재 : 고려대학교 컴퓨터·전파통신공학과 석사과정
관심분야 : DDoS attack, 무선 보안

최 현 상 (Hyunsang Choi)

2000년 3월 ~ 2004년 8월 :
고려대학교 컴퓨터학과 학사
2004년 9월 ~ 2006년 8월 :
고려대학교 컴퓨터학과 석사
2007년 3월 ~ 현재 : 고려대학교
컴퓨터·전파통신공학과 박사과정
관심분야 : 네트워크 보안,
공격시각화, 봇넷 탐지

정 현 철 (HyunCheol Jeong)

1989년 3월 ~ 1996년 2월 :
서울시립대학교 전산통계 학사
1997년 3월 ~ 1999년 8월 :
광운대학교 전자계산 석사
1996년 7월 ~ 현재 :
한국정보보호진흥원 팀장
관심분야 : 네트워크 보안

이 희 조 (Heejo Lee)

1989년 3월 ~ 2001년 2월 :
포항공대 컴퓨터공학과
학사/석사/박사
2000년 3월 ~ 2001년 2월 : Purdue
University 박사후연구원
2001년 3월 ~ 2003년 10월 : 안철수
연구소 CTO
2004년 3월 ~ 현재 : 고려대학교
컴퓨터·전파통신공학과 부교수
관심분야 : 네트워크 보안, 인터넷웜/DDos 공격 대응기술,
고가용성 시스템 설계