논문 17-42-04-11 The Journal of Korean Institute of Communications and Information Sciences '17-04 Vol.42 No.04
https://doi.org/10.7840/kics.2017.42.4.785

광범위한 단말 정보 식별을 위한 스캔 모델링 및 성능 분석

임 선 영◆, 신 승 훈*, 노 병 희°, 이 정 태**

Scan Modeling and Performance Analysis for Extensive Terminal Information Identification

Sun-young Im◆, Seung-hun Shin*, Byeong-hee Roh°, Jung-tae Lee**

요 약

네트워크 기반 단말 정보 식별 도구는 일반적으로 포트 스캔을 사용해 네트워크 단말들의 정보를 탈취하고 취약점을 찾아낸다. 특히 Shodan과 Censys는 네트워크 기반 단말 정보 식별 도구를 이용하여 광범위한 단말 정보를 탈취하여 데이터베이스에 저장한 후, 이를 사용자에게 제공한다. 이 정보는 누구나 확인할 수 있기 때문에 사이버 공격에 악용될 수 있다. 따라서 네트워크 단말 정보 탈취 방지가 필요하며, 이를 위해서는 스캔 도구가 사용하는 스캐닝 방법을 알아야한다. 하지만 Shodan과 Censys가 사용하는 스캐닝 방법은 잘 알려져 있지 않다. 따라서 본 논문에서는 Shodan과 Censys의 스캐닝 방법을 추정해 모델링하고 성능을 분석하였다.

Key Words : Scan Model, Shodan, Censys, Network Scanning Tool, Nmap, Zmap

ABSTRACT

Network scanning tools typically use port scans to steal information from network terminals and identify vulnerabilities. In particular, Shodan and Censys use a network scanning tool to gather a wide range of terminal information, store it in their database and provide it to the users. In order to prevent such information gathering, it is required to know the scanning methods of Shodan and Censys. However, the scanning model used by Shodan and Censys is not known exactly. Therefore, this paper estimates scanning models of Shodan and Censys and analyzes the performance of each models.

Ⅰ. 서 론

네트워크 기반 단말 정보 식별 도구는 일반적으로 네트워크 관리자가 시스템의 취약점을 찾아 보완하기 위한 목적으로 사용된다. 대표적인 네트워크 기반 단말 정보 식별 도구이자 단말 정보 검색 엔진인 Shodan[1]과 Censys[2]는 자신들이 수집한 광범위한 네트워크 기반 단말 정보를 사용자에게 제공한다. 이들은 Nmap[3], ZMap[4]과 같은 네트워크 기반 단말 정보 식별 도구를 이용하거나 이들이 사용하는 방법과 유사한 형태로 정보 수집 대상 호스트들의 포트를 스캔하고 열린 포트를 찾아낸다. 그리고 열린 포트를 대상

※ 본 연구는 2016년도 정부(미래창조과학부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임 (No. 2015R1A2A2A01005577)
◆ First Author : LIG Nex1 Co., Ltd., sunyoung.im@lignex1.com, 정회원
° Corresponding Author : Department of Computer Engineering, Ajou University, bhroh@ajou.ac.kr, 종신회원
* Dasan University College, Ajou University, sihnsh@ajou.ac.kr, 정회원
** Department of Software and Computer Engineering, Ajou University, jungtae@ajou.ac.kr, 정회원
논문번호 : KICS201612-393, Received December 20, 2016; Revised April 10, 2017; Accepted April 10, 2017

으로 TCP 연결을 바탕으로 한 배너 그랩을 수행해 대상 호스트의 정보를 수집하고, 수집된 정보를 웹 사이트에 게시한다. Shodan과 Censys가 제공하는 광범위한 네트워크 기반 단말 정보는 누구나 이용할 수 있어 해킹이나 공격에 악용될 수 있다[5].

통상적으로 네트워크 공격에 대한 대응 방법은 격리 네트워크 구성[6], 트래픽의 시그니처를 활용한 독립된 침입 탐지 장치의 활용[7] 및 네트워크 행위 패턴을 바탕으로 한 악성 코드 분류 기법[8] 등을 고려할 수 있다. 하지만 네트워크 기반 단말 정보 식별 도구들이 활용하는 방법은 정상적인 TCP 세션을 사용하기 때문에 이를 쉽게 식별하기 어렵고 정형적인 시그니처를 가지지 않는다. 따라서 이를 차단하거나 회피하는 것이 쉽지 않다. 하지만 네트워크 기반 단말 정보 식별 도구들이 정보를 수집하도록 방치하는 경우, 이들이 획득한 정보가 공격자에 의해 사이버 공격 시도 전 목표 정보 파악에 활용될 소지가 있으므로 이에 대한 대응 방법이 요구된다. 실제로 네트워크 공격은 사전에 분석된 취약성을 집중 공략하는 형태로 이루어지는 것이 일반적이고, 이를 위해 공격자는 공격 대상에 대한 초기 정보를 얻기 위해 네트워크 기반 단말 정보 식별 도구를 사용한다.

네트워크 기반 단말 정보 식별 도구는 인터넷에 연결되어 있는 네트워크 장치들을 원격으로 탐지하고 UDP[9], TCP SYN[10,11], TCP 연결[12]등을 활용하여 포트 스캐닝을 수행하여 스캐닝 과정에서 획득된 정보를 이용해 네트워크 단말의 운영체제 종류와 버전, 열린 포트 등의 단말 정보를 식별한 뒤, 이를 바탕으로 해당 네트워크 단말이 가지는 취약점을 수집한다[13-15].

Shodan과 Censys는 인터넷에 연결된 모든 네트워크 단말을 대상으로 스캐닝을 수행하므로, 네트워크 단말에 대한 광범위하고 민감한 정보를 노출시킬 수 있으며, 이는 해킹이나 사이버 공격과 같은 2차 범죄로 이어질 수 있다. 따라서 자신의 정보 노출을 방지하기 위해서는 Shodan과 Censys가 수행하는 포트 스캔을 탐지하고 방어할 수 있어야 한다. 이를 위해서는 Shodan과 Censys의 스캔 방법에 대해 알고 있어야 하나, 이들이 수행하는 세부적인 스캔 방법은 명확하게 알려져 있지 않다. 이에 따라 이들이 수행하는 스캔 방법과 유사한 스캔 모델을 구성하고, 각 모델이 보이는 특성을 확인함으로써 네트워크 기반 단말 정보 식별 도구를 이해하는 것이 요구된다.

본 논문에서는 기존 연구에서 공개된 Shodan과 Censys의 스캔 방법[1,2]을 바탕으로 세 가지의 스캔 모델을 구성하고 이들의 특성을 파악한다. 다만, 이들의 세부적인 구성은 공개되어 있지 않으므로, 공개된 스캔 모델을 정형화 및 변형하여 모델을 구성한다. 또한 본 논문에서 구성한 모델에서는 다수의 스캐너가 지역적으로 분산된 네트워크에 존재하는 것을 가정한다. 논문의 구성은 다음과 같다. 2장에서는 Shodan과 Censys에 대해 살펴보고, 3장에서는 스캔 모델을 제안한다. 4장에서는 제안한 각 스캔 모델의 성능을 분석하고, 5장에서 결론을 맺는다.

Ⅱ. 관련 연구

2.1 Shodan

Shodan은 여러 가지 유형의 필터를 사용하여 인터넷에 연결된 네트워크 단말의 특징을 수집해 이를 사용자에게 제공하는 네트워크 단말 검색 엔진이다. 매달 인터넷에 연결된 5억 개의 장치와 서비스들에 관한 정보를 수집하고, 수집된 정보를 일반인을 대상으로 제공한다.

Shodan은 TCP SYN 스캔을 수행하여 네트워크 단말들의 포트 개방여부를 확인한 후, 열려있는 포트를 대상으로 배너 그랩을 수행하여 단말의 정보와 단말에서 운용중인 서비스의 정보를 획득하는 것으로 알려져 있다[16]. Shodan은 스캔 범위가 개별 네트워크로 국한되는 것을 막기 위해 랜덤하게 연속적인 네트워크 스캔을 수행한다.

그림 1은 Shodan의 스캔 방법을 나타낸다[15]. Shodan은 IP주소를 랜덤하게 선택하여 조사 대상 단말을 결정하고, SYN 스캔을 보낼 하나의 서비스 포트를 랜덤하게 결정한다. 만약 선택된 IP 주소의 서비스 포트에서 SYN 스캔 메시지에 대한 응답을 보내오면,

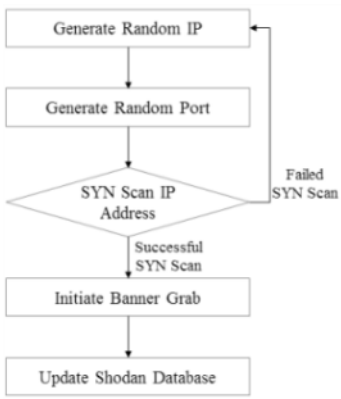


그림 1. Shodan 스캔 방법
Fig. 1. Shodan Scan Method

배너 그래빙(banner grabbing)을 수행한다. 배너 그래빙이란 네트워크 단말의 운영체제 커널, 시스템 혹은 서비스 정보 등의 수집을 위해 네트워크 단말에 서비스 요청을 보내고, 단말이 해당 요청에 대한 응답 메시지를 제공하면 이를 수집하는 행위를 의미한다. Shodan은 이를 통해 획득한 네트워크 단말에 대한 정보를 자신의 데이터베이스에 저장한다. 만약 SYN 스캔이 실패하는 경우에는 새로운 랜덤 IP 주소와 서비스 포트를 선택하여 스캐닝을 지속한다.

2.2 Censys

Censys는 인터넷에 연결된 단말 전반에 대한 스캔을 통해 수집한 데이터를 기반으로 하는 공개된 네트워크 단말 정보 검색엔진이다. Censys는 광범위한 단말 정보 수집을 목적으로 개발된 ZMap을 활용하여 정보를 수집한다. 수집된 정보는 구글 검색 엔진의 인프라를 활용해 서비스되며, 그림 2는 Censys의 스캔 방법을 나타낸다[4].

Censys의 스캔 작업은 다음과 같은 절차로 진행된다. 우선 ZMap은 SYN 스캔을 통해 단말의 IPv4 활성화 여부와 포트 개방여부를 확인한다. 만약 스캔 대상 단말로부터 응답이 존재하면, 플러그 애플리케이션 스캐너인 ZGrab이 배너 그래빙을 위한 핸드셰이크를 진행하여, 해당 포트의 애플리케이션에 대한 정보를 수집한다. 그 이후 Censys는 수집된 배너 정보 가운데 유의미한 필드를 추출하고, 여기에 추가 메타데이터를 주석으로 추가하여 데이터베이스에 저장한다. 이 과정에서 프로토콜 핸드셰이크의 정보는 네트워크 단말의 특정 프로토콜에 대한 데이터 구조로 변환된다.

Ⅲ. 스캔 모델링

Shodan과 Censys는 전 세계 네트워크 단말의 정보를 수집하므로 효율적으로 스캔을 수행해야 하는데,

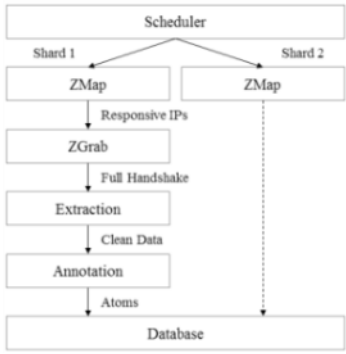


그림 2. Censys 스캔 방법
Fig. 2. Censys Scan Method

이들이 수행하는 스캐닝을 탐지하여 방어 혹은 회피하기 위해서는 이들의 스캐닝 방법의 이해가 필요하다. 다만 Shodan의 경우에는 사용하는 스캐닝 정책이 공개되어 있지 않으나, Censys는 기초적인 내용이 공개되어 있어 본 논문에서는 이를 기초로 스캔 모델을 구성한다.

Censys는 ZMap을 활용하여 대상 단말의 IP 주소를 결정하는 어드레싱 프로브(addressing probe)를 수행한다. ZMap이 단순히 숫자순으로 IPv4 주소를 탐색할 경우, 스캔 트래픽으로 인한 대상 네트워크 과부하 위험이 있고, 지역적으로 멀어서 발생 가능한 일시적인 네트워크 실패 상황도 고려가 되어야 한다. 이러한 문제를 방지하기 위해 ZMap은 주소 공간 내에서 랜덤 순열을 생성하는 규칙을 작성하여 주소를 결정하고 스캔한다[4].

본 논문에서는 ZMap이 사용하는 주소 선택 방법을 기초로 세 가지 스캔 모델을 구성한다. 첫 번째는 ZMap의 스캔 모델을 단순화 한 순수 랜덤 스캔 모델이고, 두 번째는 ZMap와 유사하게 스캐너별 스캔 범위를 지정하고, 범위 내에서 랜덤하게 주소를 선택하는 지정 범위 스캔 모델, 마지막은 스캐너 별로 서로

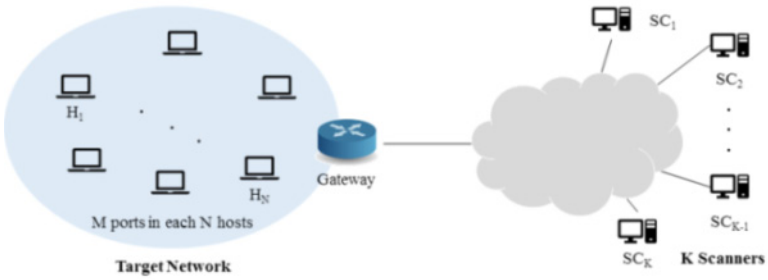


그림 3. 스캔 모델 환경
Fig. 3. Scan Model Environment

다른 균등한 범위를 담당하도록 사전 배정하는 배타 범위 스캔 모델이다.

스캐너가 운용되는 환경은 대단위 단말에 대한 스캔이 수행되는 것을 고려하여 다중 스캐너가 여러 네트워크에 분산되어 있는 것으로 가정한다. 본 논문에서 가정한 스캔 환경은 그림 3과 같다. 이 환경에서는 K개의 지역적으로 분산된 스캐너가 존재하고 스캔 대상 네트워크에는 N개의 호스트가 있으며 각 호스트의 M개의 포트가 스캐너에 의해 스캔된다. 즉, K개의 분산 스캐너가 스캔해야 할 공간은 $N \times M$개가 된다. 이 때, 스캐너는 평균이 T인 지수 분포를 따르는 T 주기로 스캔을 시도한다.

3.1 랜덤 스캔 모델

랜덤 스캔 모델은 그림 4와 같다. K개의 분산 스캐너가 스캔 대상 네트워크에 존재하는 IP와 포트를 임의로 선택해 스캔한다. 이 때 IP와 포트는 랜덤으로 선택되어 중복이 발생될 수 있으므로 각 스캐너의 스캔 시도 수는 $N \times M/K$ 보다 클 수 있다.

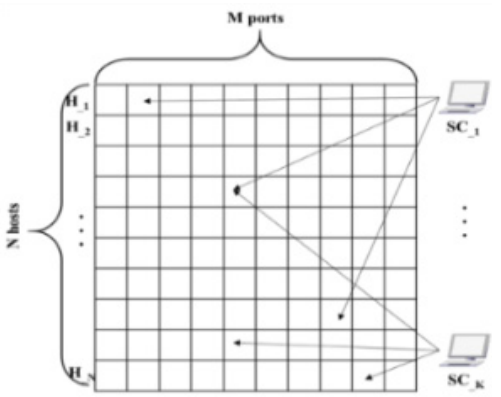


그림 4. 랜덤 스캔 모델
Fig. 4. Random Scan Model

3.2 지정 범위 스캔 모델

지정 범위 스캔 모델은 그림 5와 같다. 지정 범위 스캔 모델에서는 K개의 스캐너가 $N \times M$개의 스캔 대상을 나누어 각자 할당된 범위 내에 있는 IP와 포트를 랜덤으로 선택하여 스캔한다. 이 때 이미 스캔한 IP와 포트는 제외하여 중복을 방지한다. 범위는 균등하게 지정되므로 K개로 나뉘고, 이에 따라 각 스캐너는 $N \times M/K$번의 스캔을 시도하게 된다.

3.3 배타 범위 스캔 모델

배타 범위 스캔 모델은 그림 6과 같다. 배타 범위

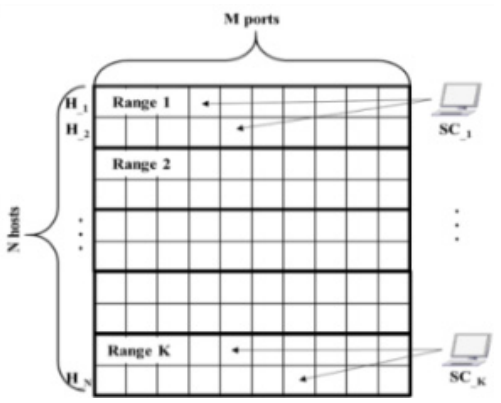


그림 5. 지정 범위 스캔 모델
Fig. 5. Specific Range Scan Model

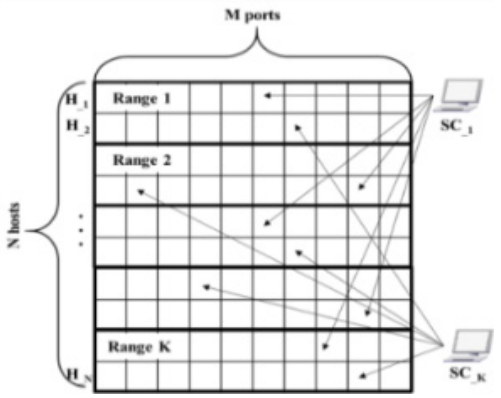


그림 6. 배타 범위 스캔 모델
Fig. 6. Exclusive Range Scan Model

스캔 모델에서는 전체 스캔 대상의 범위를 지정 범위 스캔 모델과 같이 각 스캐너에 균등하게 나누어 K개의 범위가 지정된다. 지정 범위 스캔 모델과 다른 점은 각 범위를 하나의 스캐너가 담당하지 않는다는 것이다. 각 스캐너는 $N \times M/K$개의 스캔을 시도하게 되며 이 때 배타적으로 범위를 선택하여 해당 범위 내에 있는 IP와 포트를 랜덤으로 선택해서 스캔을 시도 한다. 그리고 이미 스캔된 IP와 포트는 제외하여 중복을 제거한다.

Ⅳ. 성능 분석

세 가지 스캔 모델을 프로그램화해서 공통된 환경에서 스캔을 할 때 걸리는 총 스캔 시간을 측정하여 세 모델의 성능을 분석하였다. 소규모와 대규모 네트워크에서의 세 가지 모델의 차이를 보기 위해 호스트 수(N)를 100과 1000으로 변화시키고, 각 호스트에 존

재하는 포트 수(M)는 100으로 고정하였다. 스캐너 수(K)는 1~100개, 스캔 주기(T)는 1~100분으로 변화시켰다.

그림 7은 N이 100일 때 스캔 대상 네트워크를 모두 스캔하는데 걸리는 시간을 나타낸다. T와 K가 증가할 때 랜덤 스캔 모델이 다른 두 모델보다 확연하게 오랜 스캔 시간을 필요로 한다는 것을 알 수 있다. 지정 범위 스캔 모델과 배타 범위 스캔 모델은 총 스캔 시간이 동일하게 나타났다. 그림 8은 N이 1000일 때 총 스캔 시간을 측정한 결과를 나타낸다. 총 스캔 시간이 N이 100일 때 보다 늘어났지만 비슷한 양상을 보였다.

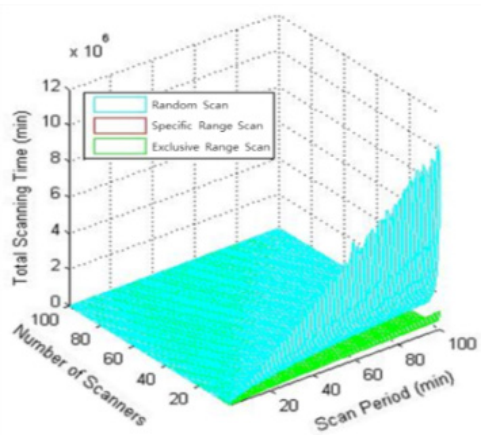


그림 7. 스캔 시간 측정 (N=100)
Fig. 7. Scan time measurement (N=100)

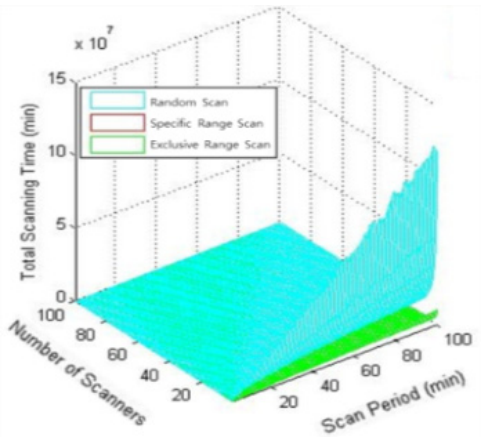


그림 8. 스캔 시간 측정 (N=1000)
Fig. 8. Scan time measurement (N=1000)

V. 결 론

본 논문에서는 네트워크 단말 스캔에 의한 정보 노출을 막기 위한 연구의 시작으로 먼저 세 가지 유형의 스캐너 모델을 구성하고, 이들의 성능을 평가하였다. 랜덤 스캔 모델에서는 스캐너가 스캔 대상 네트워크의 IP와 포트의 범위를 지정하거나 나누어서 스캔하지 않기 때문에 별도의 스캔 규칙이 필요 없다는 장점이 있다. 하지만 이미 스캔한 IP, 포트 조합을 중복 스캔하므로 총 스캔 시간은 스캔 대상 네트워크의 크기가 커질수록 가파르게 증가한다. 지정 범위 스캔 모델과 배타 범위 스캔 모델은 스캔 소요 시간이 동일하였다. 하지만 지정 범위 스캔 모델의 스캐너는 지정 범위 내에서만 스캔하므로 동일한 호스트를 빈번하게 스캔 시도하게 되어 방화벽에 탐지될 확률이 높다. 따라서 배타 범위 스캔 모델을 사용하는 것이 가장 좋다.

References

[1] Shodan, http://www.shodanhq.com/
[2] Censys, https://www.censys.io
[3] NMAP, https://nmap.org/
[4] Z. Durumeric, D. Adrian, A. Mirian, M. Bailey, and J. A. Halderman, “A search engine backed by Internet-wide scanning,” in *Proc. ACM SIGSAG 2015*, pp. 542-553, Denver, USA, Oct. 2015.
[5] G. S. Rao, P. N. Kumar, P. Swetha, and G. BhanuKiran, “Security assessment of computer networks -an ethical hacker’s perspective,” in *Proc. IEEE ICCCT 2014*, Allahabad, India, Dec. 2014.
[6] Y. Jung and M. Park, “Network defense mechanism based on Isolated Networks,” *J. KICS*, vol. 41, no. 9, pp. 1103-1107, Sept. 2016.
[7] J. Jo, H. Jang, K. Lee, and J. Kong, “SDN-based intrusion prevention system for science DNZ,” *J. KICS*, vol. 40, no. 6, pp. 1070-1080, Jun. 2015.
[8] H. Lim, W. Kim, H. Noh, and J. Lim, “Research on malware classification with network activity for classification and attack prediction of attack group,” *J. KICS*, vol. 42, no. 1, pp. 193-204, Jan. 2017

[9] S. Kumar and S. D. Sudarsan, "An innovative UDP port scanning technique," *Int. J. Future Computer and Commun.*. vol. 3, no. 6, Dec. 2014.

[10] X. Zhang, J. Knockel, and J. R. Crandall, "Original SYN: Finding machines hidden behind firewalls," in *Proc. IEEE INFOCOM 2015*, Hong Kong, China, May 2015.

[11] L. Markowsky and G. Markowsky, "Scanning for vulnerable devices in the internet of things," in *Proc. IEEE IDAACS 2015*, vol. 1, pp. 463-467, Warsaw, Poland, Sept. 2015.

[12] V. Kathayat and L. Ahuja, "Network security with open source firewall," *Int. Res. J. Comput. and Electron. Eng.*, vol. 1, no. 1, May 2013.

[13] S. Shin and G. Gu, "Attacking software-defined networks: A first feasibility study," in *Proc. ACM HotSDN 2013*, pp. 165-166, Hong Kong, China, Aug. 2013.

[14] S. Scott-Hayward, G. O'Callaghan, and S. Sezer, "SDN security: A survey," in *Proc. IEEE SDN4FNS 2013*, pp. 1-7, Trento, Italy, Nov. 2013.

[15] I. Ahmad, S. Namal, M. Ylianttila, and A. Gurtov, "Security in software defined networks: A survey," *J. IEEE Commun. Surveys & Tuts.*, vol. 17, no. 4, pp. 2317-2346, 2015.

[16] H. Mohammadzadeh, M. Mansoori, and I. Welch, "Evaluation of fingerprinting techniques and a windows-based dynamic honeypot," in *Proc. Australasian Info. Sec. Conf.*, vol. 138, Adelaide, Australia, Jan. 2013.

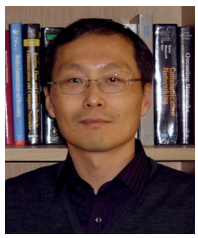
임 선 영 (Sun-young Im)

2015년 2월 : 아주대학교 컴퓨터공학과 학사
2017년 2월 : 아주대학교 컴퓨터공학과 석사
2017년 1월~현재 : LIG넥스원
<관심분야> 네트워크 보안, IoT

신 승 훈 (Seung-hun Shin)

2000년 2월 : 아주대학교 정보컴퓨터공학부 학사
2002년 2월 : 아주대학교 정보통신공학과 석사
2011년 2월 : 아주대학교 정보통신공학과 박사
2011년 9월~2016년 2월 아주대학교 소프트웨어융합학과 특임교수
2016년 3월~현재 : 아주대학교 다산학부대학 조교수
<관심분야> SW 테스트 자동화, 네트워크 보안, 멀티미디어 데이터 전송 정책

노 병 희 (Byeong-hee Roh)



1987년 2월 : 한양대학교 전자공학 학사
1989년 2월 : 한국과학기술원 전기및전자공학 석사
1998년 2월 : 한국과학기술원 전기및전자공학 박사
1989년 3월~1994년 2월 : 한국통신 통신망연구소
1998년 2월~2000년 3월 : 삼성전자
2000년 3월~현재 : 아주대학교 소프트웨어학과/대학원 컴퓨터공학과 교수
<관심분야> IoT 플랫폼 SW 및 서비스, 모바일 멀티미디어 QoS/QoE, 미래 인터넷 기술, 국방 전술통신 네트워크, 네트워크 보안

이 정 태 (Jung-tae Lee)

1979년 2월 : 서울대학교 농과대학 학사
1981년 2월 : 서울대학교 자연과학대학 계산학 석사
1988년 2월 : 서울대학교 자연과학대학 계산학 박사
1981년~1983년 울산대학교 전산학과 전임강사
1983년 3월~현재 : 아주대학교 소프트웨어학과 교수
<관심분야> 자동차, 의료기기 임베디드 SW 아키텍처