*Journal of The Korea Institute of Information Security & Cryptology* ISSN 1598-3986(Print)
VOL.26, NO.3, Jun. 2016 ISSN 2288-2715(Online)
http://dx.doi.org/10.13089/JKIISC.2016.26.3.667

# 대용량 네트워크 환경에서 익명 네트워크 탐지 및 효과적 대응전략에 관한 연구

서 정 우,† 이 상 진‡
고려대학교 정보보호대학원

# A Study on Detecting of an Anonymity Network and an Effective Counterstrategy in the Massive Network Environment

Jung-woo Seo,† Sang-jin Lee‡
Graduate School of Information Security, Korea University

## 요 약

유·무선 네트워크 인프라의 발전으로 과거와 비교할 수 없을 정도의 대용량 트래픽이 인터넷을 통해 서비스되고 있으며, 사물인터넷과 같은 네트워크 패러다임의 변화에 따라 트래픽은 매년 증가하여 2018년에는 약 1.6제타바이트의 트래픽이 네트워크를 통해 유통될 것으로 예상하고 있다. 네트워크 트래픽이 증가함에 따라 보안 인프라의 성능도 함께 발전하여 대용량의 트래픽을 보안장비에서 처리하고 있으며, 해킹 시도 및 악성코드 등 매일 수 십 만건의 보안이벤트를 처리하고 있다. 다양한 종류의 보안인프라에서 탐지하는 공격 시도에 대한 이벤트를 어떻게 효율적으로 분석하고 대응하느냐 하는 것은 안정된 인터넷 서비스를 제공하기 위해 매우 중요한 과제 중 하나이다. 하지만 현재의 보안관제 환경은 실시간으로 발생하는 대량의 보안이벤트를 분석하는 것에 어려움을 가지고 있으며, 다양한 환경적 요인에 의해 보안인프라에서 탐지하는 모든 이벤트를 분석하고 대응하는데 한계가 있다.

본 연구에서는 보안인프라에서 탐지된 이벤트에 대해 제안된 알고리즘을 사용하여 익명 네트워크를 분류하고 유해트래픽을 탐지함으로써 기존의 Low-Latency를 활용한 Tor 네트워크 트래픽 탐지와 같은 연구의 한계를 극복하고자 한다.

## ABSTRACT

Due to a development of the cable/wireless network infra, the traffic as big as unable to compare with the past is being served through the internet, the traffic is increasing every year following the change of the network paradigm such as the object internet, especially the traffic of about 1.6 zettabyte is expected to be distributed through the network in 2018. As the network traffic increases, the performance of the security infra is developing together to deal with the bulk terabyte traffic in the security equipment, and is generating hundreds of thousands of security events every day such as hacking attempt and the malignant code. Efficiently analyzing and responding to an event on the attack attempt detected by various kinds of security equipment of company is one of very important assignments for providing a stable internet service.

This study attempts to overcome the limit of study such as the detection of Tor network traffic using the existing low-latency by classifying the anonymous network by means of the suggested algorithm about the event detected in the security infra.

**Keywords:** Anonymity network, Tor, VPN

Received(02. 05. 2016), Modified(1st: 04. 18. 2016, 2nd: 06. 08. 2016), Accepted(06. 08. 2016)

† 주저자, korea002@korea.ac.kr
‡ 교신저자, sangjin@korea.ac.kr(Corresponding author)

## I. 서 론

사이버 공격의 발생 빈도가 증가함에 따라 경제적, 시간적, 물리적 피해 규모는 과거와 비교할 수 없을 정도로 커지고 있으며, 인터넷 서비스를 위협하는 형태로 발전 하고 있다. 이와 같은 사이버 공격의 증가 원인은 웹 서핑을 통한 손쉬운 해킹 툴의 획득과 익명 네트워크를 통한 공격 수행이 주요 요인이다.

익명 네트워크는 사용자 프라이버시 강화를 통해 인터넷 서비스 환경을 개선하기 위해 제안되었으며, 암호화 터널링 통신으로 익명의 관찰자로부터 웹 사이트의 데이터와 IP 주소를 숨길 수 있도록 한다. 싱글 홉(single-hop) 시스템, OpenSSH, SSL Proxy, VPN(Virtual Private Networks)은 암호화된 터널의 예제 들이며, 익명 네트워크는 기업의 네트워크 보안 환경과 보안 정책에 대해 많은 과제를 제시한다.

악의적인 의도를 가진 공격자가 자신의 아이피를 숨기려고 소스 IP 주소를 바꾼다면 라우터에서 패킷이 차단되거나 TCP 통신에서 연결 자체가 성립되지 않을 수 있어 아이피 주소를 바꿔 통신하는 것은 쉽지 않다. 그래서 공격자들은 Proxy 서버나 가상사설망(VPN), Tor 등의 익명 네트워크를 사용하여 자신의 아이피를 숨기는 방법을 많이 사용하고 있다[15][16].

실제로 최근 사이버 범죄는 중국이나 유럽 등에서 국내 가상사설망 제공 업체를 통해 공격을 수행함으로써 침해사고 발생 시 공격자 추적을 어렵게 하며, 국내 아이피 대역을 공격에 활용함으로써 보안 인프라에서 탐지되지 않도록 하고 있다.

기존의 익명 네트워크에 대한 트래픽 지연 및 Tor 네트워크 탐지에 대한 연구는 'How Much Anonymity does Network Latency Leak?, 2010, Nicholas hopper'와 'Traffic Analysis against Low-Latency Anonymity Networks Using Available Bandwidth Estimation, 2010, Sambuddho chakravarty', 'Detecting and Preventing Anonymous Proxy Usage, 2008, John brozycki' 등에서 연구되었으며, Tor의 Exit node에서 서버까지의 Latency를 분석하거나 화이트리스트 기반의 탐지패턴을 생성하는 방안에 대한 연구들이 수행되었다[1][3][4]. 하지만 이러한 연구는 Tor와 같은 제한된 방법에서 적용 가능하며, 탐지패턴을 지속적으로 관리해야 하는 한계가 존재한다. 또한 신규로 추가되는 Proxy 서버나 Tor 라우터에 대한 적절한 대응이 어려우며, 탐지 패턴을 통한 방식도 정확도가 많이 떨어진다[1][2].

기업 네트워크에서 보안 관제를 수행하는 경우에 수많은 보안장비나 네트워크 장비에서 발생하는 모든 탐지 이벤트를 분석하는 것은 비효율적이며, 네트워크 환경에 따라 분석 자체가 불가능할 수 있다. 그러므로 익명 네트워크 기반의 탐지 이벤트와 같은 위험도가 높은 이벤트에 대해 집중적인 침해대응을 수행한다면 업무 수행에 있어 효율성이 높을 것이다.

본 연구에서는 익명네트워크 여부를 확인하기 위해 군집화 기반의 AN-map 모듈을 구성하고, 실시간 보안 이벤트에서 탐지된 트래픽을 비교 분석하여 익명 네트워크 여부를 탐지하도록 한다. 그리고 익명 네트워크로 탐지된 경우 악의적인 행동을 수행하는지 여부를 보안 관제를 통해 모니터링을 수행한다. 제안 알고리즘의 업무 효율성을 측정하기 위해 DEA(Data Envelopment Analysis, 자료포괄분석) 모형을 활용하였으며, DEA 모형은 수치적으로 계량화가 어려운 부분에 효율성 측정을 위해 사용될 수 있다.

본 논문의 구성은 2장에서 관련 연구에 대해 소개하고, 3장은 익명 네트워크 탐지 방안에 대한 알고리즘을 설명하며, 4장은 공격 패턴, 5장은 보안 관제 업무의 효율성 측정 방안, 6장은 실험결과 그리고 7장은 결론으로 구성한다.

## II. 관련 연구

### 2.1 Tor 개념 및 동작원리

Tor 네트워크를 사용하는 목적은 커뮤니케이션 파트너들과 통신에 있어 사용자를 숨기는데 있다. 예를 들어 사용자 모니터링을 통해 어느 서버에 접속하는지를 찾아내는 것은 어려우며, 서버에서 Tor를 사용한 클라이언트의 특징을 발견하는 것도 불가능하다. Tor의 근본적인 목표는 프라이버시를 강화하는데 있으며, 익명성을 요구하는 사용자에게는 유용한 서비스이다.

Tor 사용자들은 Tor 클라이언트 소프트웨어를 다운로드하고 설치한 후 Tor 네트워크에 의해 SOCKS 프락시로서 실행하며, 소프트웨어는 인증 디렉토리의 하나에 접속한다. 소프트웨어는 접속 가능한 Tor 노드들의 리스트를 다운로드하고, 3개의

노드를 선택한 후 암호화된 통신 채널을 구성 한다. 첫 번째 노드를 entry node, 중간을 middle node, 세 번째 노드를 exit node라고 정의한다.

Tor의 동작은 Mixnet이라는 개념과 관련 있으며, Mixnet은 암호 연구가였던 David Chaum에 의해 구체화된 개념으로 이메일을 익명으로 수발신하기 위한 목적을 가지고 있다. Mixnet은 Mix라고 불리는 프록시 서버들을 이용한 일종의 라우팅 프로토콜인데 발송하려는 다수의 메시지를 임의의 순서로 목적지까지 보낸다. 메시지들은 공개키 암호화 알고리즘을 사용해 암호화하며 A가 B에게 메시지를 전달하는 경우 다음과 같은 형태가 된다[2][5][12][16].

$$K_m(R1, K_b, R0, message, K_m(S1, A), K_x), B) \rightarrow K_b(R0, message, K_m(S1, A), K_x)$$

Tor는 Mixnet의 프록시를 이용한 라우팅과 암호화를 일반적인 통신을 위해 구현한 실체라고 할 수 있는데 정리해 보면 다음과 같다. 라우터에 번호가 있다고 가정하고 A에서 B로 4번→3번→5번 라우터를 거쳐 전송되는 경우, 아래와 같은 형태를 나타낸다[12][16].

$$K_m(R, K_4(R, K_3(R, K_5(R, data, K_m(S, A), K_x), 5's IP address), 3's IP address), 4's IP address)$$

라우터 하나를 통과 할 때 마다 마치 양파 껍질을 한 꺼풀씩 벗겨내는 것과 같은 모양이 되는 것이다.

Tor 네트워크에서 이용하게 되는 목적지까지의 경로는 OP, 즉 프록시가 주기적으로 설정하고 해제하는데, 각 경로(노드)에 대한 정보는 디렉토리 서버(Directory Authorities Server)로부터 얻는다. Tor를 사용하는 중계서버(노드)들은 모두 디렉토리 서버에게 자신의 정보를 공유하며, 이 과정도 암호화된 통신으로 수행된다[2][14][16].

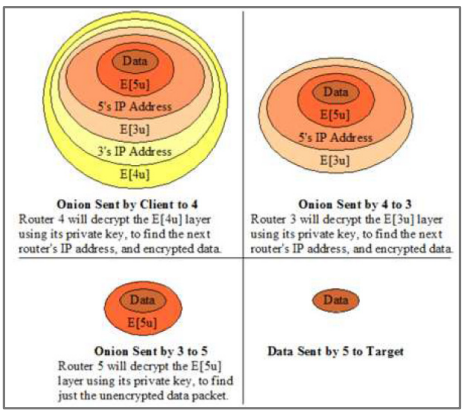


Fig. 1. Fundamentals of tor(The Onion Routing)[2]

Tor 네트워크의 장점에도 불구하고 공격자를 숨길 수 있는 장점으로 사이버침해 공격에 악용되고 있다. 이에 대한 사례를 살펴보면, '13년 6월 국가기관 홈페이지 화면이 Tor를 사용한 공격자에 의해 변조되는 사건이 발생 하였는데, Tor 네트워크는 감염시킬 PC 정보를 수집하고 추가 악성코드를 설치하는데 사용되었다.

## 2.2 DEA 모형

DEA(Data Envelopment Analysis)는 다수의 투입요소와 산출요소가 있는 상황에서 평가대상의 상대적인 효율성을 산출하기 위한 방법론으로 Charnes et al.[18]에 의해 처음으로 제시되었다. 이는 평가대상이 되는 모든 의사결정 단위(DMUs, Decision Making Units)로부터 가장 효율적인 DMU를 도출한 후 이를 기반으로 선형계획법(Linear Programming)을 이용하여 개별 DMU의 상대적인 효율성을 산출한다. 효율성(efficiency)에 대한 정의는 다양하지만 일반적으로 각 의사결정단위의 효율성은 다음과 같이 총 투입요소에 대한 총 산출요소의 비율로 표시할 수 있다[17][20].

$$\theta_i = \frac{\sum_{j=1}^{m} v_j x_{ij}}{\sum_{r=1}^{n} u_r y_{ir}} \quad (1)$$

$\theta_i$ : $DMU_i$

$y_{ir}$ : $DMU_i$ 의 $r$번째 산출물

$x_{ij}$ : $DMU_i$ 의 $j$번째 투입물

$x$ : $r$번째 산출물에 대한 가중치

$v_i$ : $j$번째 투입에 대한 가중치

$\theta_i$는 효율성을 나타내는 지표로 활용될 수 있고, 0과 1사이의 값을 갖게 된다. 따라서 특정 의사결정

단위 i가 최적의 효율성을 갖기 위해서는 모든 의사결정단위의 효율이 0과 1사이의 값을 갖도록 하면서 $\theta_i$ 값이 최대가 되도록 $u_r$, $v_j$ 값을 결정하게 된다.

DEA 모형은 불변규모 수익(CRS, Constant Returns to Scale)을 가정한 모형과 가변규모 수익(VRS, Variable Returns to Scale)을 가정한 모형으로 크게 구분할 수 있다. 불변규모 수익모형은 투입과 산출의 관계가 규모에 상관없이 일정 비율로 동일하다는 가정하의 모형으로 CCR(Charnes Cooper and Rhodes)이다. 가변규모 수익모형은 CCR 모형의 불변규모 수익가정을 완화한 모형이며, BCC(Banker, Charnes and Cooper) 모형이다. DEA 모형은 투입과 산출 중 어느 쪽을 지향하는지에 따라 투입지향(input-oriented) 모형과 산출 지향(output-oriented) 모형으로도 구분할 수 있다. 투입지향 모형은 산출을 고정한 상태에서 투입을 최대로 줄이는 것이 목적이고, 산출지향 모형의 목적은 이와 반대로 투입을 고정한 상태에서 산출을 최대로 만드는 것이다.

DEA는 다음과 같은 장점을 가지고 있다. 우선 투입요소와 산출요소가 다양하여 하나의 효율성 지수로 표현하기 힘든 경우에 유용하게 사용할 수 있다. 인원수, 시간, 돈 등 투입 및 산출요소의 측정단위가 각각 다른 경우에도 적용가능하고 화폐단위로 표시할 수 없는 경우에도 이용할 수 있다. 따라서 공공부문의 경우 투입요소나 산출요소의 가격을 파악하기 어려운 경우가 많으므로 계량화하기 어렵다는 문제를 해결할 수 있다. 둘째, 투입요소와 산출요소에 대한 가중치를 직접 추정하여 평가대상 DMU의 효율성을 추정하기 때문에 사전에 투입요소와 산출요소에 대한 지식이나 규정이 불필요하다. 비용효과분석이나 비율분석 등과 같이 성과평가를 위한 항목별 가중치를 사전에 주관적으로 결정할 필요가 없다. 셋째, 모집단의 평균치를 이용하는 회귀분석과는 달리 효율적인 DMU의 개별적인 관찰에 초점을 둠으로써 개선 가능성에 대한 유용한 정보를 제공한다[17].

## III. 제안 알고리즘

익명 네트워크를 통한 사이버 공격 위협이 증가함에 따라 인터넷 서비스에서 발생하는 보안 이벤트 중 익명 네트워크를 사용하는 패킷을 탐지하고 분류하여 집중적인 보안 관제를 수행하고자 본 알고리즘을 제안한다.

제안된 알고리즘은 4개의 모듈로 구성되어 있으며, 첫째, AN-map module은 탐지 이벤트 로그 데이터 셋을 이용하여 가상사설망(VPN) 및 Tor 네트워크를 활용한 공격 트래픽을 분류하고, 군집화를 통해 AN-map을 생성한 후 Detecting module과 매칭하여 실시간 공격 트래픽을 탐지한다. 둘째, Packet capture module은 실시간 네트워크 트래픽을 네트워크 스위치(L3)에서 포트 미러링을 통해 트래픽을 수집한 후 Detection module로 전송한다. 셋째, Detecting module은 패킷에서 속성을 추출하여 해당 패킷이 익명 네트워크를 사용한 공격인지 여부를 확인하기 위해 AN-map module과 비교분석을 수행한다. 넷째, Intrusion Response module은 실시간 탐지된 보안 이벤트가 익명 네트워크 기반의 공격이면 보안 관제를 통해 집중적인 분석을 수행한다. Intrusion Response module은 Detecting module에 의해 익명 네트워크로 분류된 경우에만 수행한다.

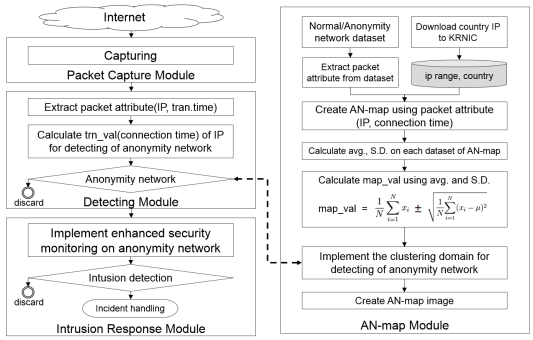


Fig. 2. Proposed Algorithm

### 3.1 AN-map 모듈

AN-map 모듈은 정상 네트워크 데이터 셋과 익명 네트워크 데이터 셋의 군집화를 통해 AN-map을 구성하기 위한 데이터 셋을 수집한다. 그리고 KRNIC에서 국가별 아이피 주소를 다운로드 받아 국가별 아이피 대역을 구성한다. 국가별 아이피 대역은 실시간 공격 트래픽에 대한 국가 정보를 파악하기 위해 활용한다. Fig. 3.은 군집화를 통한 데이터 셋의 영역을 나타낸다.

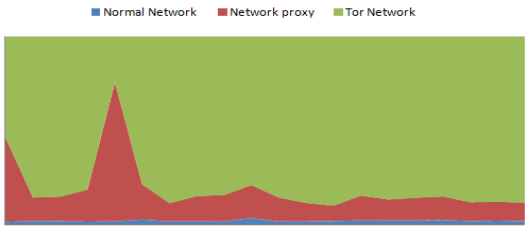


Fig. 3. Domain of dataset using clustering

AN-map은 데이터 셋의 IP 주소, connection time을 사용하여 군집화 영역을 생성한다. AN-map에서 각 데이터 셋에 해당하는 평균 값과 편차 값을 계산하고, 평균과 편차를 사용하여 map_val을 계산한다.

$$\text{map\_val} = \frac{1}{N}\sum_{i=1}^{N} x_i \pm \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i-\mu)^2} \quad (2)$$

네트워크 트래픽 유형에 따라 map_val 값을 계산한 후 군집화를 통해 클러스터링을 구성하고, Detecting module의 trn_val과 비교 분석을 통해 익명 네트워크 여부를 확인하는데 사용한다. 군집화 결과 분석을 통한 AN-map을 Fig. 4.와 같이 생성하며, 새로운 데이터 셋을 수집하여 군집화를 통한 클러스터링 작업을 반복적으로 수행한다.

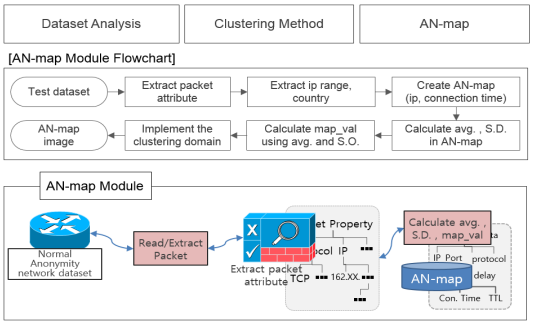


Fig. 4. Process for AN-map module

### 3.2 패킷 캡처 모듈

DMZ 영역에 위치한 서버에 접속하는 네트워크 트래픽을 분석하기 위해 스위치에 포트 미러링을 설정하고, 분석서버의 NIC(Network Interface Card)를 통해 수집된 패킷을 Detecting module로 전송한다.

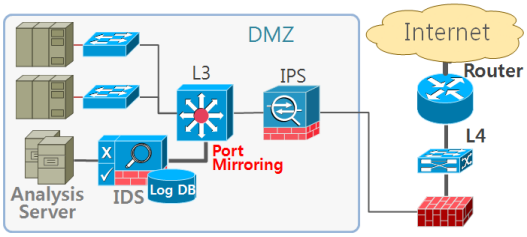


Fig. 5. Port mirroring for packet capture

### 3.3 탐지 모듈

탐지 모듈(Detecting Module)은 AN-map module에서 생성한 map_val 값과 비교분석을 통해 익명 네트워크 여부를 확인하기 위해 trn_val 값을 계산한다. trn_val 값 계산을 위해 실시간 수집된 패킷의 IP 주소와 connection time 값을 계산하여 유형별로 분류된 map_val과 비교분석 작업을 수행한다. 예를 들어 trn_val 값이 가상사설망 기반의 map_val에 포함될 경우 익명 네트워크를 사용한 네트워크 트래픽이라고 판단한다.

M = MAX{map_val}, N = {MIN(map_val}
map_val = {X | N ≤ map_val ≤ M}
trn_val = {Y | connection time}
if X⊃Y, X is Anonymity Network

trn_val 값이 map_val 값의 범주에 포함될 경우 익명 네트워크 기반의 트래픽으로 분류하고, 탐지 아이피 대역이 국내 아이피 대역이면 해당 네트워크 트래픽은 침해공격 의도를 가진 트래픽으로 분류한다.

### 3.4 침해 대응 모듈

실시간 탐지된 트래픽이 익명 네트워크 기반의 트래픽이면, 보안 관제에서 관심을 가지고 이벤트 추적을 수행해야 한다. 익명 네트워크를 이용하는 경우 공격자를 추적하는 것이 불가능 하며, 공격 시스템의 변경이 자유로워 보안 측면에서 큰 위협이 된다. Fig. 6.은 본 연구의 침해 대응 모듈(Intrusion Response module) 수행 절차를 도식화하여 표현한다.

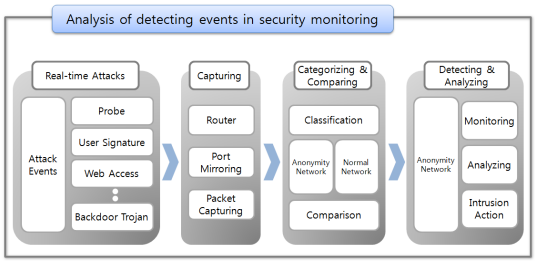


Fig. 6. Incident response life cycle

## IV. 공격 패턴

이번 장은 익명 네트워크 기반의 공격 유형을 살펴보고, 테스트 환경 설계를 통해 익명 네트워크를 탐지할 수 있는 구성 방안에 대해 살펴본다.

### 4.1 유형별 위협

**가상사설망(VPN)**은 인터넷망과 같은 공중 네트워크를 사설 네트워크처럼 사용하여 비용을 크게 절감할 수 있는 통신 서비스를 말한다. 최근 대규모의 국내 IP를 가상사설망 서버로 임대하여 인터넷을 통해 판매하는 경우가 증가하고 있으며, 임대한 가상사설망 서버를 사용하여 공격자 IP를 속이거나 다른 해킹을 위한 중간 경유지로 악용하고 있다.

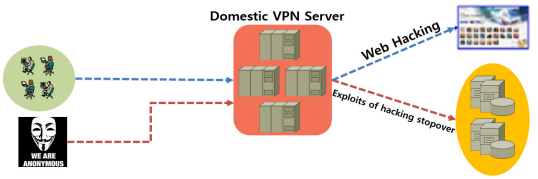


Fig. 7. Attack using domestic VPN server

Fig. 7.은 인터넷을 통해 판매되고 있는 가상사설망 아이피를 활용한 공격 흐름을 나타내고 있으며, 공격자는 ISP(Internet Service Provider)에서 판매하는 가상사설망 아이피를 구입하여 자신의 정보를 속이는데 사용한다.

서비스를 제공하는 서버가 국내에 존재하고, 국내 이용자만 접근 가능한 서비스를 제공한다면 침입차단 시스템은 해외 아이피 주소 접속을 차단할 것이다. 하지만 가상사설망(VPN) 서버를 통해 국내 IP 주소를 받은 후 해당 서비스에 접속하면 방화벽은 국내 IP 주소로 판단하여 서비스를 허용하게 된다. 즉, 공격자 IP를 숨기기 위해 가상사설망 서비스를 통해 국내 IP로 변조하여 웹 사이트 해킹, 개인정보 유출 등의 공격을 수행한다. 이처럼 가상사설망을 해킹 경유지로 악용할 경우 공격 근원지 추적에 어려움이 존재한다.

**Tor**는 익명 네트워크 중 하나이며, 전 세계적으로 2,000,000명 이상의 사용자가 사용하는 것으로 추정되고 있다. Tor의 목적은 인터넷 사용자의 익명성을 보호하는 것인데, 사용자가 목적 시스템에 서비스를 받기 위해서는 참여자가 제공한 라우터를 활용하여 서비스를 요청하게 된다. 이때 각 노드는 entry node, middle node, exit node로 구성되며, 암호화된 통신을 한다. 목적지 시스템에 데이터를 요청하는 아이피는 exit node의 아이피이며, 요청자의 아이피를 확인할 수 없다.

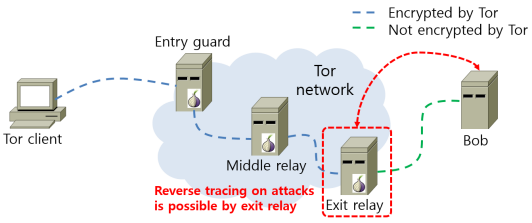


Fig. 8. Tor's tracing

Tor는 공격자에 의해 Onion Router(OR)를 활용하여 네트워크를 사용할 수 있으며, 자신의 서버를 Onion 라우터로 소유할 수 있다. 이를 통해 네트워크 트래픽을 삭제하거나 수정, 지연 등의 작업을 할 수 있다. 공격자는 Tor 네트워크를 통해 대량의 데이터를 탈취하기에 속도가 느리기 때문에 C&C 서버와 통신을 통해 공격 대상 네트워크에 대한 익명성을 유지하면서 공격을 수행하는데 활용한다.

**Proxy server**는 클라이언트와 서버 사이에서 데이터를 중계하는 역할을 하는 가상서버이며, 방화벽 기능 및 캐시 기능이 있어 네트워크 트래픽을 줄이고 데이터 전송 시간을 향상시킨다. Anonymous proxy server의 장점에도 불구하고 지역적 및 정치적 특수성으로 인해 Proxy server를 활용한 공격이 점점 증가하고 있으며, 피해 사례도 다양화 되고 있다. 이와 같이 공격자는 Proxy server를 통한 우회 공격을 통해 증거 수집 및 역 추적 등을 불가능하게 하여 자신을 숨기면서 공격 효과를 증대시키는 방법을 사용하고 있다.

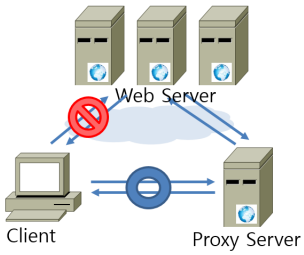


Fig. 9. Detour connection using proxy server

Proxy server를 무료로 사용할 수 있는 리스트는 인터넷에 공개되어 있어 불특정 사용자가 자유롭게 활용할 수 있으며, 공격자의 서비스 요청은 Proxy server를 통해 목적지 서버와 신뢰할 수 있는 통신망을 구성하게 되고 공격자는 Proxy server를 C&C로 활용하여 공격을 수행하게 된다. Proxy server를 활용한 공격 방법으로는 IP spoofing attack 이나 웹 취약점을 이용한 웹 해킹 등 다양한 형태로 이용되고 있다.

### 4.2 공격 분석 방법론

보안 인프라에서 발생하는 실시간 이벤트를 분석하는 것은 매우 많은 자원을 필요로 하며, 때로는 불가능할 수 있다. 그러므로 익명 네트워크와 같은 위험도가 높은 탐지 이벤트를 선별적으로 분류하여 상세분석을 수행하도록 함으로써 업무 효율성을 높이는 것이 필요하다.

제안된 알고리즘을 적용한 익명 네트워크 탐지의 테스트 환경은 Fig. 10.과 같이 구성한다.

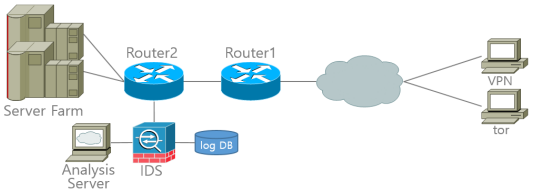


Fig. 10. Configuration of testbed

## V. 보안 관제 업무의 효율성 측정 방안

네트워크 속도 및 대역폭의 증가에 따라 보안 장비에서 처리해야 하는 데이터의 양은 이전과는 비교할 수 없이 증가하고 있으며, 보안 관제 인력이 처리할 수 있는 이벤트는 Table 1.과 같이 발생 이벤트의 10% 수준이다. 주요 침해위협에 대한 적기 대응율도 30%를 넘지 못하는 것으로 분석된다. Table 1.에서 분석한 보안 이벤트는 하루 평균 500Gbyte 자료가 유통되는 네트워크 환경에서 발생하는 보안 시스템들의 이벤트 건 수이며, 보안 관제 인력 10여 명이 탐지 이벤트에 대한 분석 및 대응을 수행하는 비율을 나타낸다. 실제로 보안장비에서 탐지한 이벤트 대부분을 보안 관제에서 효과적으로 분석하지 못하고 있으며, 탐지된 보안 이벤트도 기업의 네트워크 환경에 위협적이지 않은 오탐이 많이 포함된 것을 알 수 있다. 결과적으로 Table 1.과 같이 대량의 탐지 이벤트들에 대해 적절한 분석을 수행하지 못하고 있으며, 보안 관제 업무에 있어서도 비효율성이 증가하고 있다.

Table 1. Timely handling ratio about security event

| Risk | Event number (daily) | Intrusion handling ratio |
|---|---|---|
| High | 49,655 | 10%~30% |
| Middle | 3,635 | under 10% |
| Low | 6,304 | under 5% |
| Total | 59,594 | 15% |

본 연구에서는 자료포괄분석(DEA: Data Envelopment Analysis) 방법을 사용하여 제안된 알고리즘을 적용한 경우의 업무 효율성을 측정한다. 자료포괄분석을 이용한 효율성 측정은 분석대상인 DMU를 결정하고, DEA는 분석 대상의 동질성을 전제로 하여 개별 DMU가 유사한 활동을 수행함으로써 DMU 간에 서로 비교할 수 있는 제품이나 서비스를 산출한다. 모든 DMU에는 유사한 범위 내에서 자원이 투입되며, 외부요인으로 인해 성과가 영향받지 않도록 운영 한다.

DEA를 성공적으로 적용하기 위해서는 투입·산출변수의 선정이 중요하며, 투입변수로 보안이벤트(개수/월), 투입인력(명/월) 3가지 항목을 선정한다. 투입변수와 산출변수는 투입변수가 증가하면 산출변수 역시 증가해야 하는 양의 상관관계를 가진다.

Table 2. Inputs and Outputs

| Item | Inputs | Outputs |
|---|---|---|
| Effectiveness of security control | Security event, Input manpower | Response time |

## VI. 실험 평가

실험의 목적은 공격자가 Tor나 VPN 등 익명 네트워크를 통해 국내 IP 주소로 위장하여 공격을 수행하는 경우 탐지 이벤트의 위험도를 VH(Very High)로 설정하여 상세분석을 수행할 수 있도록 한다. 위험도를 높이는 이유는 익명 네트워크를 활용한 탐지 이벤트의 경우 공격자가 특정 목적을 가지고 공격을 수행했을 가능성이 높기 때문에 해당 트래픽에 대한 상세분석을 수행하도록 한다. 실험환경 구성은 Fig. 11.과 같으며, 익명 네트워크 기반의 보안이벤트 탐지영역과 이벤트에 대한 상세분석을 위한 관제영역으로 구분된다.

### 6.1 보안 이벤트 탐지

테스트를 위한 실험환경은 실제 인터넷 서비스를 제공하고 있는 DMZ(Demilitarized zone) 영역에서 네트워크 트래픽에 대한 수집과 분석을 수행한다. 테스트 환경의 백본 통신망 및 DMZ는 1GbE 네트워크 환경으로 구성 되어있으며, 하루 평균 500Gbyte의 데이터가 송수신되며 59,000건(daily)의 보안 이벤트가 침입방지시스템(IPS, Intrusion Protection System)과 침입탐지시스템(IDS, Intrusion Detection System), 침입차단시스템(Firewall)에서 탐지되고 있다. IPS와 IDS, F/W는 1GbE 네트워크 환경의 Fiber Cable로 연결되어 있으며, 2,000Mbps throughput을 처리할 수 있다.

Analysis Server는 2.40GHz(6-Core) CPU, 6GB MEM, 2TB HDD로 구성되어 있으며, Log DB는 80TByte NAS 솔루션으로 연결되어 있다.

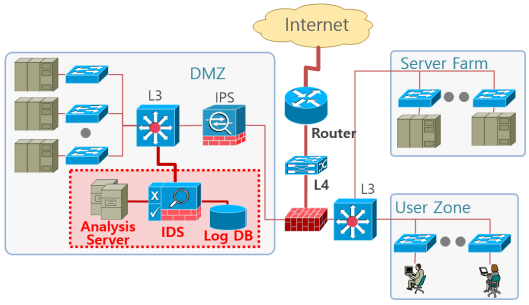


Fig. 11. Diagram of experiment environment

테스트 환경의 Log DB에서 보안 이벤트에 탐지된 로그 데이터를 추출하여 군집화를 위한 데이터 셋으로 활용한다. 데이터 셋은 AN-map module에서 map_val을 구하기 위해 사용하며, IP 주소와 connection time 등의 속성 값을 추출한다. 익명 네트워크를 사용하는 트래픽의 connection time을 추출하기 위한 방법은 Fig. 12.와 같으며, connection time은 $t = t + T_{cr} + T_{rs}$와 같다.

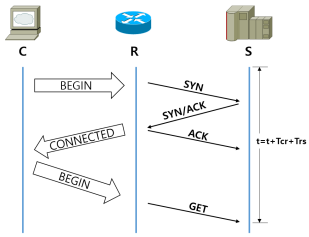


Fig. 12. Measuring method of connection time

Table 3.은 데이터 셋에서 추출한 정상 트래픽과 익명 네트워크(Tor, VPN)의 유형별 connection time을 나타낸다.

Table 3. Learning result on network traffic type

| Data | Normal | VPN | Tor |
|---|---|---|---|
| LD0001 | 0.019 | 0.599 | 0.710 |
| LD0002 | 0.020 | 0.138 | 0.940 |
| … | | | |
| LD200 | 0.017 | 0.130 | 0.840 |
| LD201 | 0.017 | 0.187 | 0.894 |
| … | | | |
| LD499 | 0.020 | 0.295 | 0.907 |
| LD500 | 0.018 | 0.129 | 0.539 |

Normal은 정상적인 인터넷 서비스 요청에 대한 connection time이며, VPN은 해외에서 국내 VPN 서비스 업체를 이용하여 국내 아이피 대역으로 변환 후 공격을 수행하는 경우의 connection time이다. Tor는 전 세계 참여자가 제공하는 Onion Router(OR)를 통해 공격자를 숨기면서 공격이 가능하고, End node의 IP 주소가 서비스 요청 IP 주소가 된다.

Fig. 13.은 데이터 셋의 유형에 따른 connection time 그래프와 AN-map 그래프를 나타낸다.

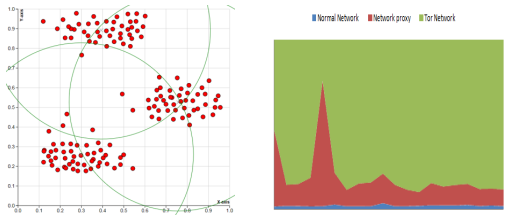


Fig. 13. Graph of connection time and AN-map

실험 환경에서 익명 네트워크를 탐지하기 위해 Table 4.와 같은 유형별 테스트 데이터를 정상적인 경우와 VPN, Tor 유형별로 구분하여 실험 데이터를 전송하였으며, 보안 장비에서 탐지된 이벤트에 대한 connection time을 측정하였다.

Table 4. Test result on type of network traffic

| Data | Normal | VPN | Tor |
|---|---|---|---|
| RD001 | 0.020 | 0.147 | 0.875 |
| RD002 | 0.025 | 0.127 | 0.572 |
| RD003 | 0.018 | 0.097 | 0.747 |
| RD004 | 0.017 | 0.100 | 0.701 |
| ... | | | |
| RD100 | 0.018 | 0.106 | 0.937 |

Fig. 14.는 Table 4.에 대한 테스트 결과를 그래프에 나타내고 있으며, 정상적인 네트워크 트래픽은 익명 네트워크 영역과 다른 영역에 위치하고 있는 것을 확인할 수 있다.

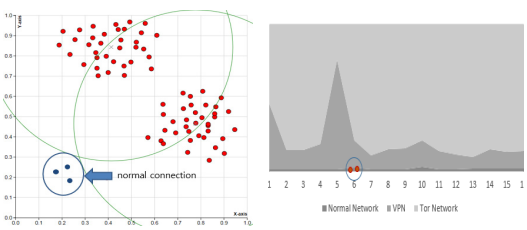


Fig. 14. Result of detection on normal network

Fig. 15.는 실험 데이터에서 VPN과 Tor 기반의 공격 트래픽에 대한 결과 값을 나타내며, 익명 네트워크를 사용한 공격 트래픽의 connection time을 그래프로 나타낸다.

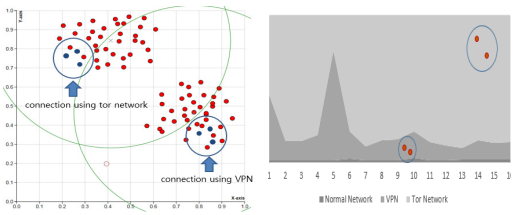


Fig. 15. Result of detection on anonymity network

실험 환경에서 탐지된 보안 이벤트가 Tor 또는 VPN 영역의 connection time에 존재할 경우 해당 트래픽은 익명 네트워크라는 것을 확인할 수 있으므로 Intrusion Response module에서 침해대응을 수행한다.

### 6.2 DEA 모형 기반의 업무 효율성 분석

DEA 모형은 규모의 효과에 대한 가정에 따라 CCR 모형과 BCC 모형으로 구분할 수 있으며, 효율성 측정의 목적에 따라 투입지향 모형과 산출지향 모형으로 구분할 수 있다. 본 연구에서는 업무 효율성 분석을 위해 CCR 모형을 적용하였으며, CARLOS P. BARROS and FERNANDO P. ALVES가 제시한 일반적인 기준에 따라 산출지향 모형을 적용하였다.

효율성을 산출한 결과는 Table 5.와 같으며, 산출지향 DEA 모형의 결과 값은 효율적인 DMU의 경우는 1.0로 나타내며 비효율적인 DMU는 1.0보다 작은 값으로 나타낸다. 예를 들어 DMU 1.0의 CCR 모형 효율성 값은 0.9로 산출되는 경우 이는 현재 대비 산출을 11% 향상시켜야 효율적인 DMU가 된다는 의미이다.

본 연구에서 설정한 투입 및 산출변수들은 특성변수들이 결합되어 있다. 즉, 투입변수는 보안장비에서 발생하는 보안 이벤트 수와 보안 관제를 수행하는 직원 수로 구분되고, 산출변수는 보안이벤트 대응 시간으로 구분한다. 따라서 본 연구에서는 2개의 투입변수(security event, staff)와 1개의 산출변수(response time)를 동시에 고려하여 보안 관제 업무 효율성을 분석한다.

보안 관제 업무 효율성을 평가하기 위해 사용한 CCR모형에 대한 DMU의 효율성 분석결과는 Table 5.와 같이 정리하였다. Table 5.의 수치는 DEA의 효율성 정도를 나타내는 것으로 상대적으로

가장 효율적인 DMU의 수치가 1.0으로 나타나며, 비효율적인 DMU일수록 수치가 1.0미만으로 점점 낮아진다.

Table 5. Effectiveness point using DEA

| Item | Effectiveness point |
|---|---|
| DMU1 | 0.385 |
| DMU2 | 0.127 |
| DMU3 | 1.0 |
| DMU4 | 0.447 |
| DMU5 | 0.239 |

Fig. 16.은 DEA에 의한 효율성을 plot 형태로 나타내며, DEA-Analyst 소프트웨어를 사용하여 DEA 분석을 수행하였다.

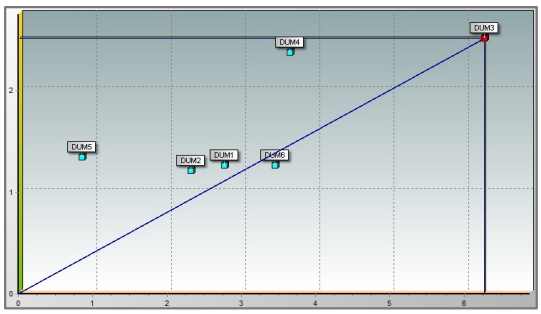


Fig. 16. Effectiveness plot using DEA

분석 결과는 2개의 투입변수와 1개의 산출별수를 모두 고려하여 보안 관제 업무 효율성을 산출한 결과 DMU3이 가장 효율적인 모형으로 선정되었으며, DMU3은 탐지된 보안 이벤트에서 제안된 알고리즘을 이용하여 익명 네트워크 트래픽을 분류 한 후 위험도를 높여 보안 관제를 수행하도록 하였다. 그 결과 익명 네트워크 기반의 사이버 위협을 집중적으로 분석한다면 필수적으로 침해대응을 수행해야하는 보안 이벤트 수를 최소화 하면서 효율적인 인력 운영으로 중요 보안위협을 탐지 및 대응할 수 있다는 결론을 도출할 수 있다.

## VII. 결 론

사이버공격은 더욱 지능화 및 고도화 되고 있으며, 기업의 영업비밀을 해킹을 통해 불법적으로 취득하거나 국가기밀을 절취하여 사회적 혼란을 발생시키고 있다. 특히 네트워크 환경의 발전에 따라 대량의 트래픽이 실시간으로 유통되면서 기존의 정보보안 체계에서 유해 트래픽을 분석하고 대응하는 것은 쉽지 않다. 이처럼 대량의 보안 이벤트를 효율적으로 처리하는 것은 기업의 정보보안 강화를 위한 핵심적 미션이 되고 있다.

본 연구에서는 익명 네트워크 기반의 공격 트래픽을 탐지하는 알고리즘을 제안하며, 탐지된 공격 트래픽의 우선순위를 높여 상세분석을 수행하도록 함으로써 보안관제 업무 효율성을 증대하도록 하였다. 업무 효율성에 대한 측정 방법은 DEA(Data Envelopment Analysis) 모형을 사용하였으며, DMU(Decision Making Unit)의 개별적인 관찰에 초점을 둠으로써 개선 가능성에 대한 유용한 정보를 제공하도록 하였다.

결론적으로 본 연구를 통해 익명 네트워크를 탐지하는 효과적 알고리즘을 제안하고, 제안된 방법을 적용해 보안관제 업무의 효율성을 실험을 통해 평가하였다.

## References

[1] John Brozycki, Eric Cole, "Detecting and Prevventing Anonymous Proxy Usage," SANS Institute InfoSec Reading Room, Sep 2008.

[2] Example of Analysis about principal of Tor network and related malware, Korea Internet & Security Agency, May 2014.

[3] Nicholas Hopper, Eugene Y. Vasserman, and Eric Chan-Tin, "How Much Anonymity does Network Latency Leak?," ACM Transactions on Information and System Security, Vol. 13, Feb. 2010.

[4] Stevens Le Blond, David Choffnes, Wenxuan Zhou, Peter Druschel, Hitesh Ballani and Paul Francis, "Towards Efficient Traffic-analysis Resistant Anonymity Networks," ACM SIGCOMM '13, pp. 303-314, 2013.

[5] Norman Danner, Danny Krizanc, and Marc Liberatore, "Detecting Denial of Service Attacks in Tor," Financial

Cryptography and Data security LNCS 5628, pp. 273-284, 2009.
〔6〕 Jun-Ki Lee, Kwang-Sun Park, "Countermeasure against Cyber Crime using VPN," Journal of Korean Digital Forensics Society, Dec 2013.
〔7〕 Dingledine, R., Mathewson, N., Syverson, P., "The Second-generation onion router," Proceedings of the 13th USENIX Security Symposium, pp. 303-320, 2004.
〔8〕 Sambuddho Chakravarty, Angelos Stavrou, and Angelos D. Keromytis, "Traffic Analysis against Low-Latency Anonymity Networks Using Available Bandwith Estimation," ESORICS 2010 LNCS 6345, pp. 249-267, 2010.
〔9〕 Young-Jin Kim, Su-Yeon Lee, Hun-Yeong Kwon, and Jong-In Lim, "A Study on the Improvement of Effectiveness in National Cyber Security Monitoring and Control Services," Korea Institute of Information Security & Cryptology, Vol 19, pp. 103-111, 2009.
〔10〕 Agrawal, D., Kesdogan, D., "Measuring Anonymity: The Disclosure Attack," IEEE Security & Privacy, pp. 27 - 34, 2003.
〔11〕 Back, A., Moller, U., Stiglic, "Traffic Analysis Attacks and Trade-Offs in Anonymity Providing Systems," Information Hiding, Vol 2137, pp. 245-257, Oct 2001.
〔12〕 Bauer, K., McCoy, D., Grunwald, D., Kohno, T., Sicker, D., "Low-Resource Routing Attacks Against Tor," Proceedings of the 2007 ACM Workshop on Privacy in Electronic Society, pp. 11 - 20, 2007.
〔13〕 Borders, K., Prakash, A., "Web Tap: Detecting Covert Web Traffic," Proceedings of the 11th ACM Conference on Computer and Communications Security, pp. 110-120, Oct 2004.
〔14〕 Burch, H., Cheswick, B., "Tracing Anonymous Packets to Their Approximate Source," Proceedings of the 14th USENIX Conference on System Administration, pp. 319 - 328, Dec 2000.
〔15〕 H. AND K"OPSELL, S., "JAP: Java anonymous proxy, " http://anon.inf.tu-dresden.de/, 2006.
〔16〕 S. J. and DANEZIS, G., "Low-cost traffic analysis of Tor," In Proceedings of the 2005 IEEE Symposium on Security and Privacy, pp. 183-195, 2005.
〔17〕 Kyungwan Ko, Daecheol Kim, "The Analyses of the Operational Efficiency and Efficiency Factors of Retail Stores Using DEA Model," The Korean Operations Research and Management Science Society, Vol 31, pp. 135-150, 2014.
〔18〕 Banker, R.D., Charnes, A and Cooper, W.W, "Some models for estimating technical and scale inefficiencies in data envelopment analysis, Management Science, pp. 1078-1092, 1984.
〔19〕 Lawrence M. Seiford, Joe Zhu, "Modeling undesirable factors in efficiency evaluation," European Journal of Operational Research, Vol 142, pp. 16-20, Oct 2002.
〔20〕 A. Kleine, "A general model framework for DEA," Omega, Vol 32, pp. 17-23, Feb 2004.
〔21〕 Barros, C. P., "Efficiency measurement among hypermarket and supermarket and the identification of the efficiency drivers : a case study," International Journal of Retail and Distribution Management, Vol.34, pp.135-154, 2006.
〔22〕 Taechang Ryu, "A Study on the Efficiency Evaluation of Traditional Market using a DEA Model" Korea Planning Association, Vol 46, pp. 257-270, Oct 2011.

## 〈저자소개〉

서 정 우 (Jung-woo Seo) 학생회원
2004년 2월: 고려대학교 정보보호대학원 석사
2013년 9월~현재: 고려대학교 정보보호대학원 박사과정
2004년 2월~2012년 8월: 삼성전자 책임연구원
2012년 8월~현재: KMA
〈관심분야〉 디지털 포렌식, 네트워크 보안, 악성코드 탐지

이 상 진 (Sang-jin Lee) 종신회원
1987년 2월: 고려대학교 수학과 학사
1989년 2월: 고려대학교 수학과 석사
1994년 8월: 고려대학교 수학과 박사
1989년 10월~1999년2월: ETRI 선임연구원
1999년 3월~2001년8월: 고려대학교자연과학대학조교수
2001년 9월~현재: 고려대학교 정보보호대학원 교수
2008년 3월~현재: 고려대학교 디지털포렌식 연구센터 센터장
〈관심분야〉 디지털 포렌식, 심층 암호, 해쉬 함수