논문 15-40-02-11 The Journal of Korean Institute of Communications and Information Sciences '15-02 Vol.40 No.02
http://dx.doi.org/10.7840/kics.2015.40.2.334

다중 요소를 가지는 SVM을 이용한 이블 트윈 탐지 방법

강 성 배◆, 양 대 헌*, 이 경 희°

Evil-Twin Detection Scheme Using SVM with Multi-Factors

SungBae Kang◆, DaeHun Nyang*, KyungHee Lee°

요 약

최근 스마트기기가 널리 보급되면서 무선망이 가능한 AP(Access Point)의 사용 또한 증가하였다. AP를 사용하여 무선망에 접속할 때, 적절한 보안이 제공되지 않는다면, 로그 AP(Rogue AP)에 의해 다양한 보안 문제가 발생될 수 있다. 이 연구에서는 로그 AP의 유형 중 하나인 이블 트윈(Evil Twin)에 대한 위협에 대해서 살펴본다. 최근 대부분의 이블 트윈을 탐지하기 위한 연구에서는 RTT(Round Trip Time)와 같이 인가된 AP와 이블 트윈 사이에서 측정될 수 있는 시간 차이를 이용하는 방법이 주로 이용되고 있다. 그러나 이와 같이 이블 트윈을 탐지하는 방법은 채널이 혼잡한 상태일 때 탐지율이 떨어지는 단점이 있다. 이러한 이유에서 이 연구에서는 이블 트윈을 탐지하는 기준으로 RTT와 함께 추가로 PIAT(Packet Inter-Arrival Time)을 측정한다. 또한 측정된 값을 SVM(Support Vector Machine)의 학습 요소로 사용함으로써, 이블 트윈 분류를 위한 비선형적 기준을 정한다. 결과적으로 채널이 혼잡한 상황에서도 최대 96.5% 최소 89.75%의 높은 확률로 이블 트윈을 성공적으로 탐지하였다.

Key Words : fingerprint, evil twin, rogue AP, network security, WLAN security, Wi-Fi

ABSTRACT

Widespread use of smart devices accompanies increase of use of access point (AP), which enables the connection to the wireless network. If the appropriate security is not served when a user tries to connect the wireless network through an AP, various security problems can arise due to the rogue APs. In this paper, we are going to examine the threat by evil-twin, which is a kind of rogue APs. Most of recent researches for detecting rogue APs utilize the measured time difference, such as round trip time (RTT), between the evil-twin and authorized APs. These methods, however, suffer from the low detection rate in the network congestion. Due to these reasons, in this paper, we suggest a new factor, packet inter-arrival time (PIAT), in order to detect evil-twins. By using both RTT and PIAT as the learning factors for the support vector machine (SVM), we determine the non-linear metric to classify evil-twins and authorized APs. As a result, we can detect evil-twins with the probability of up to 96.5% and at least 89.75% even when the network is congested.

Ⅰ. 서 론

스마트 기기의 사용이 급증하면서 무선망에 접속하여 인터넷을 사용하는 사용자가 많아졌으며 이러한 사용자 요구사항에 맞추어 광범위한 지역에 무선망을 제공하는 AP(Access Point)가 설치되고 있다. 무선망

※ 본 연구는 인하대학교의 지원에 의하여 수행되었습니다.
◆ First Author : Department of Computer and Information Engineering, Inha University, sbkang87@isrl.kr, 학생회원
° Corresponding Author : Department of Electronic Engineering, SuwonUniversity, khlee@sowon.ac.kr, 정회원
* Department of Computer and Information Engineering, Inha University, nyang@inha.ac.kr, 정회원
논문번호 : KICS2015-01-011, Received January 16, 2015; Revised February 11, 2015; Accepted February 11, 2015

은 해당 장소의 관리자에 의해 설치된 AP에 의해 제공되거나 통신사의 통신망에 연결된 스마트폰에서의 테더링 기능을 통해 핫-스팟(Hot Spot)을 설정하여 제공될 수 있다. 무선망을 제공하는 AP는 다양한 인증 방법과 암호화 방법을 사용하여 사용자에게 제공된다. 다양한 무선망 환경에서 사용자는 AP에 연결할 때마다 AP의 신원을 자세하게 확인하고 연결하기는 어렵다. 또한, 암호화되어 있지 않은 무선망이나 미리 설정되어 있어 자동으로 AP에 연결할 때는 별도로 장치의 신원을 확인하지 않는다.

이러한 사용자의 무선망 접속 패턴과 다양한 종류의 무선망 설치의 용이성 때문에 사용자는 로그 AP의 위협에 처할 수 있다. 로그 AP는 사용자를 자신의 무선망에 접속하도록 만들기 위해 인가된 AP보다 더 강한 RSSI(Received Siginal Strength Indication)를 제공하고 빠른 속도와 혼잡하지 않은 쾌적한 상태의 인터넷 환경을 제공한다. 스마트기기 사용자의 대부분은 자신의 기기를 자동으로 신호 세기가 강한 AP에 우선으로 접속하도록 설정해놓고 있기 때문에 사용자의 스마트 기기는 인가된 AP 보다 더 강한 신호 세기를 제공하는 로그 AP에 자동으로 접속할 확률이 높다. 일단 사용자가 로그 AP에 접속하여 인터넷을 사용하면 악의적인 공격자는 로그 AP를 이용하여 사용자이름, 비밀번호, 메시지, 신용카드 정보 등 사용자의 민감한 정보를 가로채어 취득한 정보들을 조합하여 사용자에게 피해를 입힐 수 있다. 또한, 공격자는 로그 AP를 이용하여 중간에서 위조 패킷을 보내거나 피싱 사이트로 우회시키는 등의 중간자 공격을 수행할 수도 있다.

이 연구에서는 로그 AP의 공격 유형 중 하나인 이블 트윈(Evil-Twin)과 그에 대한 탐지 방법을 제시한다. 그림 1에서 이블 트윈을 정의하고 있다. 이블 트윈은 악의적인 공격자가 무선망을 통해 인터넷에 접속하며 사용자에게 이 무선망을 이용하여 인터넷 연결을 중계해서 무선망을 제공하는 로그 AP 방식이다. 이블 트윈은 별도로 유선망에 연결하기 위한 노력이 필요하지 않기 때문에 공격을 수행하기가 용이하다. 즉, 이블 트윈은 무선망이 제공되는 장소라면 언제, 어디서나 쉽게 설치할 수 있고, 공격자가 자신의 공격 목적을 달성한 이후에 빠르게 로그 AP를 제거할 수 있다. 이러한 이블 트윈 공격의 용이성 때문에 사용자가 이블 트윈에 연결하여 인터넷을 사용하기 이전에 탐지하지 못한다면, 사용자의 민감한 정보는 이미 가로채어진 이후가 될 수밖에 없으므로 매우 위협적인 공격이다.

이블 트윈은 인가된 AP의 MAC 주소, IP 주소, 판

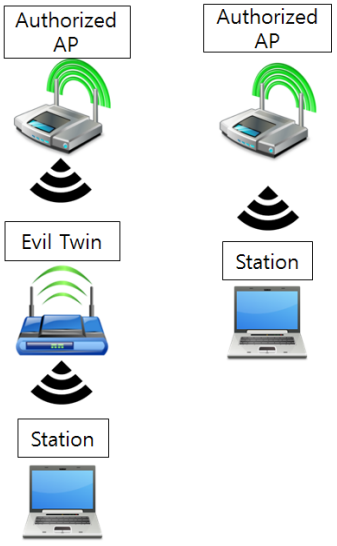

그림 1. (좌)이블 트윈, (우)인가된 AP
Fig. 1. (left)Evil-Twin, (Right)Authorized AP

매자 이름, SSID 등을 위조하고 있기 때문에 이블 트윈이 인가된 AP와 함께 같은 네트워크 안에 있을 때 화이트리스트만으로는 이블 트윈을 탐지할 수 없다. 또 다른 탐지 방법으로 네트워크 패킷들을 분석하여 악의적인 패턴을 찾아내는 방법이 있는데 이러한 방법은 네트워크 관리자에게 도움을 받아야 하고 수많은 패킷 중에서 악의적인 패턴을 보이는 AP만을 골라내기 위한 분석에 많은 시간이 걸릴 수 있다. 수많은 패킷을 분석하는데 시간이 많이 들기 때문에 패킷을 분석하여 이블 트윈을 성공적으로 탐지해냈을 때는 이미 이블 트윈의 공격자는 공격을 마친 후일 수 있다. 따라서 이블 트윈을 탐지하는 방법으로 별도의 네트워크 관리자의 도움을 필요로 하지 않으며, 화이트리스트를 사용하지 않고, 사용자 측면에서 빠른 시간 안에 이블 트윈을 탐지할 수 있는 방법이 필요하다.

최근 이블 트윈을 탐지하기 위한 방법으로 RTT(Round Trip Time)을 측정하는 연구들이 많이 수행되고 있다. 최근 연구에서는 이블 트윈 방식에서 발생할 수밖에 없는 추가적인 1홉의 무선구간에 따른 지연시간에 따라 RTT를 측정하여 인가된 AP와 이블 트윈을 분류하였다. RTT을 측정하여 이블 트윈을 탐지하는 방법은 사용자 측면에서 쉽게 사용이 가능하며 탐지하는 데 오랜 시간이 걸리지 않는다는 이점이 있다. H. Han 등은 RTT와 무선망의 혼잡정도를 RTT의 표준편차로 나타내어 선형적인 기준을 정하여 이블 트윈을 탐지하는 방법을 제시하고 있다. 이 방법은 RTT을 선형적인 기준을 정하여 이블 트윈을 탐지하기

때문에 동적으로 변하는 채널의 혼잡상태나 다양한 무선망 환경에 적절하게 적용되지 않는다. 이전 연구에서 무선망이 혼잡한 상태일 때의 오탐지율이 H. Han 등이 제시한 오탐지율보다 더 높은 것을 증명하였다.

C. Yang 등은 서버까지의 RTT와 접속한 AP까지의 RTT의 비율을 이용하여 이블 트윈을 탐지하는 방법을 제시하였다. 이들은 다양한 신호세기에서의 실험하여 대부분의 상황에서 100%에 가까운 이블 트윈 탐지율을 제시하였다. 하지만 이들의 연구는 채널이 유휴한 상태에서만 진행되었으며 채널이 혼잡한 상태일 경우 자신들의 알고리즘이 제대로 동작하지 않을 수 있다는 한계점을 명확하게 명시하고 있다.

우리는 이전 연구에서 H. Han 등과 같은 요소를 사용하여 SVM에 입력해서 이블 트윈을 탐지해 보았다[1]. 이 연구에서는 이전 연구를 향상시켜 사용자 측면에서 간단하게 RTT과 추가적으로 PIAT(Packet Inter-Arrival Time)을 측정하여 SVM의 요소로 입력하여 높은 확률로 이블 트윈을 탐지할 수 있는 탐지 방법을 제시한다. RTT을 측정하여 인가된 AP와 이블 트윈을 분류하는 데 있어서 H. Han 등이 이용한 선형적인 기준을 정하는 것이 아니라 SVM을 이용하여 동적으로 기준을 정하는 비선형 분류를 한다. 또한, C. Yang 등이 제시한 탐지 방법은 채널이 혼잡한 상황에서 제대로 동작할 수 없는 한계가 있다고 스스로 밝히고 있고 H. Han 등은 채널이 혼잡한 상태에서 탐지율이 떨어진다. 따라서 우리는 채널이 혼잡한 상태에서도 높은 확률로 이블 트윈을 탐지하는 방법을 제시한다. 먼저, 이 논문의 2장에서는 일반적인 로그 AP와 이블 트윈 탐지 방법에 관련된 연구들을 조사하여 각 탐지 방법을 고찰한다. 그 중 H. Han 등의 탐지 방법과 C. Yang 등의 탐지 방법을 자세하게 조사하여 두 방법에서의 장점과 한계점을 명확하게 하고, 기존의 연구가 이 연구에 어떻게 영향을 미칠 것인지 설명한다. 3장에서는 이 논문에서 주장하는 아이디어와 알고리즘을 제시하고 설명한다. 4장에서는 실험이 진행되기 위한 세부적인 실험환경을 설명한다. 5장에서는 실험결과를 분석하고 우리의 탐지 방법을 적용한 탐지율과 기존 연구와의 탐지율을 비교한다. 6장에서는 이 연구의 결론과 앞으로 수행해야 할 추가적인 연구가 무엇인지 살펴본다.

Ⅱ. 관련 연구

2.1 로그 AP 탐지 기법

2006년 P. Bahl 등은 SSID와 MAC 주소를 이용하여 화이트리스트를 유지하여 로그 AP를 탐지하는 방법을 제시하였다[2]. 이 방법은 별도의 장치를 이용하여 모든 무선망에 전송되는 패킷 데이터를 모니터링하여 수집된 패킷 정보에서 화이트리스트에 존재하지 않는 AP의 정보가 나올 때 해당 패킷을 전송한 장치를 로그 AP로 판단하는 기법이다. SSID와 MAC 주소를 위조하는 경우에도 MAC 주소의 순서번호를 분석하거나 주변에 별도의 장치를 두어 AP들의 신호 세기를 분석하는 방법을 통하여 인가된 AP의 위치를 확인하고 목록에 부합하는지를 확인하여 로그 AP를 탐지해낼 수 있다. 그러나 이 방법은 내부 네트워크로 사용이 제한되고, 별도의 추가적인 장치를 필요로 하며, AP의 이동이 잦은 곳이나 무선 AP가 많은 곳에는 적용하기가 어렵다. 2007년 D. Schweitzer 등은 모니터링 시스템에서 AP의 신호 세기를 분석하여 프로파일 맵을 생성하고 신호 세기를 분석해서 나온 AP의 위치가 인가된 AP의 위치인지 확인하는 기법을 제시하였다[3]. 이 방법은 신호 세기를 분석하기 위해 별도의 장치를 필요로 하며 인가된 AP에 대한 위치 정보를 화이트리스트로 가지고 있어야 한다. 2010년 S. Jana 등은 클락 스큐를 이용한 로그 AP 탐지 방법으로, 비콘과 프로브 요청/응답 메시지를 수집하여 모든 AP의 클락 스큐를 계산하여 현재 데이터베이스에서 가지고 있는 클락 스큐와 일치하지 않으면 로그 AP로 판단하는 방법을 제시하였다[4]. 이 기법은 핑거프린트 기술을 필요로 하며 인가된 AP의 화이트리스트를 필요로 한다. 2007년 L. Watkins 등은 유, 무선 네트워크에 전송되고 있는 패킷들의 RTT을 수집하여 분석하는 방법으로 로그 AP를 탐지하는 기법을 제시 하였다[5]. 이 방법은 데이터를 수집, 분석하는데 많은 시간이 걸릴 수 있으며 네트워크 관리자의 도움을 필요로 한다. 2007년 W. Wei 등은 TCP/ACK을 이용하여 RTT을 측정하여 로그 AP를 탐지하는 기법을 제시하였다[6]. 이 방법은 공격자가 ACK을 대신하여 응답하면 간단하게 회피가 가능하다. 2012년 3g 망을 사용하는 로그 AP를 탐지하는 방법으로 TTL 값이 1과 2로 설정된 ICMP패킷을 이용하여 RTT을 측정하고, 측정된 RTT을 K-NN(K-Nearest Neighbor Classifier)를 이용하여 로그 AP를 분류하는 기법을 제시하였다[7]. 이 방법은 공격자가 TTL값이 2인 패킷에 대하여 대신 응답함으로써 RTT을 인가된 AP처럼 보이게 할 수 있다. 또한, 이 방법은 혼잡한 상태에서도 100%의 탐지율을 보여주고 있지만, 이러한 결과는 3g 망 자체가 지연시간이 크므로 현재의 스마트한 로그 AP에 비해 로그 AP 자체의 성능이 너무 낮기 때문에 도출된

당연한 결과로 보인다.

그 외에 신호 세기 수치를 수집하거나[8], AP의 정보를 가지는 화이트리스트를 이용한 방법[9-11], 무선 주파수 변화를 이용하여 로그 AP를 탐지하는 방법이 있다[12]. 유선과 무선망을 모두 활용한 로그 AP 탐지 방법들도 있으나[13-15], 이러한 방법들은 모두 추가적인 장치를 필요로 하거나 화이트 리스트, 또는 네트워크 관리자의 도움을 필요로 하기 때문에 사용자가 직접 로그 AP를 탐지하기는 힘들다.

이 장의 나머지에서는 사용자 측면에서 효율적으로 실행해 볼 수 있으며 네트워크 관리자의 도움이나 인가된 AP들의 목록인 화이트리스트를 필요로 하지 않고, 별도의 추가적인 장치를 사용하지 않는 기존의 연구인 H. Han 등과 C. Yang 등의 논문을 자세히 살펴보고 그들의 연구의 한계점을 알아본다.

2.2 H. Han. 등의 탐지 방법

H. Han 등은 사용자가 이블 트윈에 접속하는 경우 이블 트윈과 인가된 AP 사이에 반드시 추가적인 무선 구간이 생기는 것을 이용하여 RTT를 측정함으로써 이블 트윈을 탐지하는 방법을 제시하였다[16]. 그림 2에서 H. Han 등은 이블 트윈에 의해 생기는 무선구간에 대한 지연시간을 구하기 위하여 단순히 Local DNS까지 RTT만을 측정하는 것이 아니라 사용자가 접속한 AP까지 RTT을 측정하여, Local DNSe구할 수 있으며, 홉 수를 많이 거치지 않기 때문에, RTT을 측정하였을 때 유선망이 미치는 영향을 최소화하여 추가적인 무선구간을 찾아내기 수월하게 한다. 사용자가 접속한 AP까지의 거리(사용자-AP)를 측정한 RTT을 100번 측정하여 평균을 구한 값을 RTT_{probe}라고 하고, 사용자로부터 Local DNS까지의 거리를 측정한 RTT을 100번 측정하여 평균을 구한 값을 RTT_{dns}라고 한다. 이렇게 구해진 RTT의 평균값을 이용하여 DNS까지의 거리에서 사용자가 접속한 AP 까지의 거리를 빼면 $\Delta T = RTT_{dns} - RTT_{probe}$ 값을 구할 수 있다. Local DNS까지의 거리를 측정함으로써 유선망의 지연시간을 최소화한 것과 함께 AP까지 거리를 빼는 것으로, 이블 트윈의 경우 대략적으로 추가적인 무선 구간만을 측정하는 것이 가능해진다. 따라서 ΔT 값을 이블 트윈을 탐지하는 기준으로 하여 이블 트윈을 탐지한다.

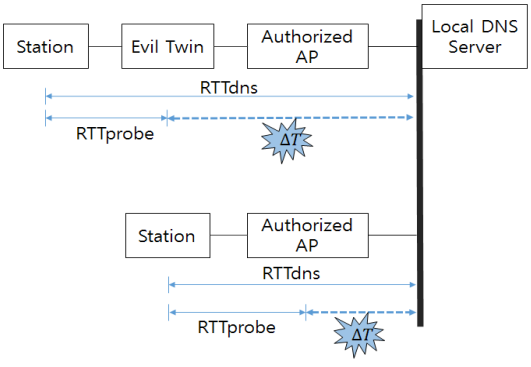

그림 2. H. Han. 등의 이블 트윈 탐지 방법.
Fig. 2. H. Han et al's Evil-Twin Detection Scheme

ΔT 값을 이용하여 이블 트윈을 분류하는 데 있어서 한계치 θ를 구하기 위하여 H. Han 등은 RTT_{probe}의 표준편차인 σ_{probe}와 RTT_{dns}의 표준편차인 σ_{dns}를 이용한다. $\theta = \alpha^*(\sigma_{probe} + \sigma_{dns})/2 + \beta$를 이용하여 한계치를 계산할 때, 변수 $\alpha = 0.49$, $\beta = 1.3$ 으로 고정되는데 이 값은 실험에 의해 도출된 값임을 밝히고 있다. 이 한계치는 RTT의 표준편차인 σ 값이 채널의 혼잡 상황을 어느 정도 나타낸다고 판단하여 표준편차가 증가할 경우 채널이 혼잡하다는 것을 나타내어 채널의 혼잡상황에 따라 한계치 값이 일정하게 증가되어 채널 상황에 맞게 분류기준이 달라지게 만든 것이다.

결론적으로, ΔT 값이 한계치 θ보다 클 경우 이블 트윈으로 분류하며 ΔT값이 한계치 θ보다 작을 경우 인가된 AP로 분류한다. 채널이 유휴 상태일 때에 H. Han등의 탐지 기법은 100%의 탐지율을 보이고 있지만, 채널이 매우 혼잡한 상태일 때는 60% 이하의 탐지율을 보여준다. RTT을 측정하는 샘플링의 횟수를 300회로 늘렸을 때에 80% 정도로 탐지율을 높일 수 있음을 보여준다.

H. Han 등의 이블 트윈 탐지 방법은 사용자 측면에서 사용할 수 있고, 별도의 장치가 필요하지 않고, 네트워크 관리자의 도움을 필요로 하지 않는다는 점은 매우 훌륭한 탐지 기법이라고 볼 수 있다. 그러나 이블 트윈을 인가된 AP와 분류하기 위한 기준인 한계치를 설정하는 데 있어서 실험에 의해 도출하여 고정된 $\alpha = 0.49$, $\beta = 1.3$의 값과 함께 단순한 직선을 이용한 선형 분류 방법은 동적으로 변하는 채널상황과 다양한 네트워크 상황에 적절하게 적용되지 못하며 채널이 혼잡한 상태에서 이블 트윈 탐지율이 상당히 떨어지는 단점이 있다.

2.3 C. Yang 등의 탐지 방법

C. Yang 등은 TCP/ACK을 이용하여 서버까지 RTT을 측정하고, AP까지 RTT을 측정해서 서버-AP

RTT의 비율을 구한 값을 이용하여 적절한 한계치를 설정하여 이블 트윈을 탐지하는 방법을 제시하였다[17]. C. Yang 역시 인가된 AP는 사용자와 직접 연결하여 1홉으로 연결되어 있지만, 이블 트윈은 사용자와 인가된 AP 사이의 구간이 1홉으로 이루어져 있지 않고 추가적인 무선구간을 포함하여 2홉으로 이루어질 수밖에 없다는 점을 이용하였다.

C. Yang 등의 이블 트윈 탐지 방법은 별도로 기존의 프로토콜을 변경하거나, 새로운 프로토콜을 만들지 않아도 되며 서버와 사용자에게 전송받은 TCP 패킷에 대하여 즉각적으로 응답하는 패킷을 전송하는 프로그램을 설치하면 된다. 일반적인 TCP/ACK을 이용한 RTT의 측정은 사용자가 TCP 패킷을 전송하고 서버로부터 전송된 ACK을 사용자가 받는 시간으로 측정된다. 그러나 이러한 방법은 공격자가 서버 대신에 ACK 응답을 해주는 것으로 간단하게 회피할 수 있다. C. Yang 등의 프로그램은 서버에서 먼저 전송한 패킷에 대해 사용자 측에서 즉각 ACK을 응답하는 프로그램이다. 이렇게 함으로써 이블 트윈 공격자가 TCP/ACK 패킷을 대신 응답하여 탐지를 회피하는 방법을 무력화시킬 수 있다.

C. Yang 등의 RTT 측정 방법은 그림 3과 같다. 일단 서버에서 P_1을 사용자에게 전송하고, P_1을 받은 사용자는 즉각 응답으로 A_1을 서버로 전송한다. 이렇게 A_1을 전송받은 서버는 즉각 P_2를 전송하고, 또 다시 P_2를 받은 사용자는 A_2를 서버로 전송한다. 이러한 과정에서 P_1과 P_2를 받은 시간의 차이를 이용하여 사용자가 P_1을 받은 시각을 T_{P1}이라 하고 P_2를 받은 시각 T_{P2}라 하면 $A_{S=}T_{P2}-T_{P1}$를 이용하여 사용자로부터 서버까지의 RTT을 계산하여 IAT(Interpacket Arrival Time) 값을 계산한다.

이렇게 서버까지의 RTT을 측정한 A_S만을 이용하여 로그 AP를 탐지하는 방법은 다양한 무선망 환경에서 적절한 한계치를 찾기가 힘들기 때문에 C. Yang 등은 사용자가 접속한 AP 까지의 RTT인 A_A를 구하여, $\alpha = A_S/A_A$를 계산하여 SAIR(Server-to-AP IAT Ratio)라고 부르는 비율을 구하여 이블 트윈을 탐지한다. 이러한 비율이 802.11b 일 때에 1.31 을 넘어갈 때 이블 트윈으로 판단하고 802.11g 일 경우 1.48을 넘어갈 때 이블 트윈으로 분류한다. 이블 트윈의 탐지율을 높이기 위하여 측정한 RTT이 정해놓은 기준치 이상일 경우 폐기하는 데이터 필터링 기법과 SAIR을 한번만 구하지 않고 여러 번 측정하여 평균을 구하는 데이터 스무딩 기법을 활용한다.

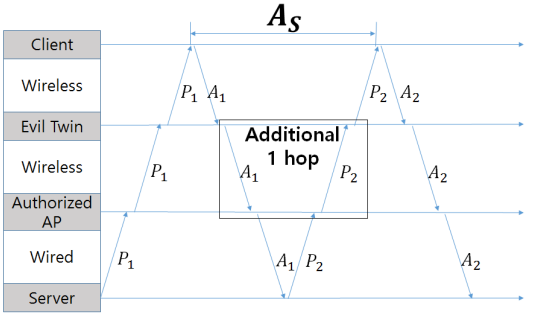

그림 3. C. Yang 등의 이블 트윈 탐지 기법
Fig. 3. C. Yang et al's Evil-Twin Detection Scheme.

C. Yang 등의 기법은 별도의 프로그램, 네트워크 관리자의 도움, 추가적인 장치 등을 필요로 하지 않으면서, 1초 이내의 빠른 시간에 사용자 측면에서 이블 트윈을 탐지해 볼 수 있다는 점에서 매우 참신한 기법이다. 그러나 이들의 기법은 채널이 유휴한 상태만을 가정하여 매우 실험적인 환경에서 실험되었기 때문에 탐지율이 100%에 가깝지만, RTT에 영향을 미치는 채널이 혼잡해지는 상황에서는 이들의 기법을 적용한 탐지 방법은 탐지율이 떨어질지도 모른다. 실제로 C. Yang 등의 논문에서는 연구의 한계점으로 자신들의 기법이 채널이 혼잡해질 경우 제대로 동작하지 않음을 인정하고 있다.

2.4 SVM(Support Vector Machine)

SVM(Support Vector Machine)은 기계 학습 분류 알고리즘의 하나로써 두 그룹을 분류하기 위하여 최적의 분리 경계면인 초평면(Hyperplane)을 제공한다. 그림 4의 (a)에서 수집된 각 데이터의 특징으로 입력된 데이터만을 가지고 두 그룹을 분류하기는 쉽지 않다. 따라서 SVM을 이용하여 입력된 데이터인 (a)를 다차원 공간인 (b)로 맵핑하여 초평면을 이용하여 분류하도록 한다. 또한, SVM은 다차원 공간에서 구해진 여러 개의 후보 초평면과 함께 각 초평면에서 각 그룹의 점들에 이르는 최소값인 여백(Margin)을 찾아준다. 여러 개의 후보 초평면 중에서 여백이 최대화되는 초평면을 골라 분류 기준으로 정하게 된다. 그림 4의 (b)에서 다차원 공간에 매핑 된 데이터는 초평면 H_1와 H_2에 의해 성공적으로 두 그룹을 분류될 수 있다. 그러나, 초평면 H_1에서 각 그룹의 점에 이르는 최소값인 여백 M_1이 초평면 H_2에서의 여백 M_2보다 적기 때문에 초평면 H_1은 추후에 들어오는 데이터를 성공적으로 분류하기가 어렵다. 따라서 SVM은 초평면

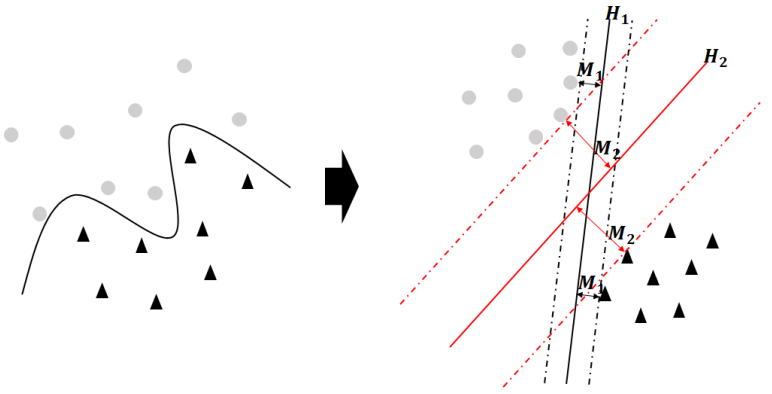

그림 4. SVM의 원리
Fig. 4. The Principle Of Support Vector Machine

H_2을 선택하여 두 그룹을 분류함으로써 분류율을 최대로 하면서 동시에 오류율을 최소화하게 된다.

SVM은 처음에 단순히 이진 분류를 위하여 개발되었으나 현재에는 다중 데이터에 대해 분류가 가능해지면서 생물정보학, 문자인식, 필기인식, 얼굴 및 물체 인식 등 다양한 분야에서 사용되고 있다. SVM은 분류율을 높이기 위하여 분류 기준을 정할 뿐만 아니라 분류 기준에 대한 여백을 최대화하는 것까지 고려함으로써 일반적으로 결정트리나 신경망 등의 분류 알고리즘보다 성능이 좋은 것으로 알려져 있다.

Ⅲ. 다중 요소를 가지는 SVM을 이용한 이블 트윈 탐지 방법

3.1 SVM을 이용한 이블 트윈 탐지 방법

이전의 로그 AP 탐지 방법들은 화이트리스트를 이용하거나, 네트워크 관리자의 도움을 받거나 별도의 장치를 사용하여 로그 AP를 탐지하였다. 그러나 이러한 방법들은 스마트한 공격자에 의해 회피되기 쉽고, 탐지하는 데 걸리는 시간이 길다는 단점이 있었다. 최근 연구에서는 이블 트윈 공격 상황에서 추가적인 무선망 때문에 발생하는 추가적인 1홉 구간을 이용하여 RTT을 측정함으로써 사용자 측면에서 간단하게 이블 트윈을 탐지해 볼 수 있는 방법이 제시되었다. 이 논문에서도 최근 연구 동향에 맞추어 RTT을 측정하여 사용자 측면에서 빠른 시간에 높은 확률로 이블 트윈을 탐지할 수 있는 방법을 제시한다. 그뿐만 아니라 기존의 연구와 달리 RTT에 추가하여 PIAT(Packet Inter-Arrival Time)를 측정하여 이블 트윈에 의해 발생하는 추가적인 무선망 자체 또는 추가적인 AP의 업무량 증가에 의한 패킷 전송 간격의 변화를 살펴봄으로써 인가된 AP와 이블 트윈의 차이를 분석하고 이블 트윈을 탐지하기 위한 요소로 사용한다. 또한, 최근 연구에서는 인가된 AP와 이블 트윈을 분류하기 위해 설정되는 한계치를 실험적이거나 이론적으로 도출하고 선형적으로 분류하였지만, 이 연구에서는 SVM을 이용하여 두 그룹을 분류하는 데 있어 동적으로 판단 기준을 정하도록 하고 비선형적으로 분류한다.

이 연구의 이전 연구에서 H. Han 등의 실험을 재현하여 H. Han 등이 설정한 이블 트윈을 분류하는 정적이고 선형적인 기준은 동적으로 변하는 채널 상태와 다양한 무선망 환경에 적절하게 적용되지 않음을 보였다. 또한, H. Han 등에 의해 제시된 추가적인 무선망을 계산하는 $\Delta T = RTT_{dns} - RTT_{probe}$ 값과 채널의 혼잡 척도를 나타내는 $(\sigma_{probe} + \sigma_{dns})/2$를 그대로 사용하면서도 SVM의 요소로 넣음으로써 동적으로 비선형 분류를 하는 실험을 실행하여 채널이 혼잡한 상황에서도 90%에 가까운 탐지율을 보여줌을 증명하였다. 여기에서는 H. Han 등에 의해 도출된 RTT의 평균과 표준편차를 그대로 이용하지 않고, 측정한 RTT과 추가로 측정한 PIAT을 별도로 가공하지 않고 직접 SVM의 요소로 넣어줌으로써 채널이 혼잡한 상황에서도 이블 트윈을 탐지하는 확률을 높이고자 하였다. 이렇게 함으로써 이 연구에서는 채널이 혼잡한 상황에서도 탐지율을 보장해주는 방법을 제시하고 있다.

3.2 PIAT(Packet Inter-Arrival Time)

기존의 RTT을 측정하여 이블 트윈을 탐지하는 방법은 단순히 사용자로부터 서버까지의 RTT을 계산하

여 이블 트윈을 탐지하는 방법이었다. H. Han 등이나 C. Yang 등은 여기에 더하여 사용자가 접속한 AP 까지의 RTT을 계산하고, 서버까지의 RTT과 함께 그들만의 공식으로 이블 트윈의 추가적인 무선망을 찾아내는 방식을 사용하였다.

우리는 기존 방식들의 RTT을 측정하는 방법의 아이디어는 그대로 참고하면서 추가적으로 PIAT(Packet Inter-Arrival Time)을 측정하여 이블 트윈을 탐지한다. PIAT는 여러 개의 RTT을 측정하기 위해 서버에 DNS 패킷을 전송할 때에 패킷 간에 전송 간격을 일정하게 설정하여 전송한 후에, 전송한 패킷에 대해 서버에서 사용자에게 응답이 도착했을 때의 시간을 측정하여 패킷의 도착 시간에 대한 간격을 측정한 것이다. 전송된 패킷은 전송 당시에는 일정한 시간 간격으로 설정되어 전송되지만, 채널의 혼잡상황이나 AP 또는 라우터의 업무량에 따라 패킷이 전송되는 과정에서 시간 간격이 일정해지지 않게 된다. 이블 트윈은 반드시 추가적인 AP가 존재하므로 PIAT이 인가된 AP에서 측정된 PIAT보다 일정하지 않을 것임을 실험을 통하여 증명한다.

이 논문의 6장에서는 이블 트윈과 인가된 AP에서의 PIAT을 측정한 결과 값을 비교해 보고, 추가적인 AP가 존재함으로써 PIAT에 어떠한 영향을 미치는지 확인해본다.

3.3 RTT와 PIAT 측정 방법 및 SVM 분류

3.3.1 서버까지의 RTT 측정 방법

서버까지의 RTT을 측정하기 위하여 DNS 쿼리 패킷을 이용한다. DNS 쿼리를 바꾸어가며 전송하면 공격자가 제때에 적절한 응답을 할 수 없기 때문에 RTT을 감소시켜 탐지 방법을 회피하는 것을 무효화 시킬 수 있다. RTT를 측정하기 위한 Local DNS로서 교내 DNS를 사용한다. 서브네트워크 내의 DNS 서버를 사용하여 홉 수를 최소화 하는 것은 유선망으로 인한 지연시간을 줄이게 되어 더욱 쉽게 이블 트윈을 탐지할 수 있지만, 서브네트워크 내에 DNS 서버를 가지고 있는 상황은 일반적이지 않다. 또한, 홉 수가 너무 많은 DNS 서버를 사용하는 것은 DNS 서버를 신뢰할 수 없는 경우가 있고 유선망의 지연시간이 너무 길어지게 되어 이블 트윈을 탐지하는 데 방해가 될 수 있다. 이 연구에서는 교내 DNS 서버를 사용하여 서버까지는 7홉 거리이다. 일반적인 사용자는 학교나 회사 또는 통신사의 DNS 서버를 활용할 수 있고 7홉의 거리 때문에 발생하는 유선망의 지연시간은 크지 않기 때문에 실생활에 적용하기 적당한 DNS 서버라고 볼 수 있다.

3.3.2 AP까지의 RTT 측정 방법

사용자가 접속한 AP까지의 RTT을 측정하기 위해 ICMP 패킷을 이용한다. 일반적으로 이블 트윈 공격에서 ICMP 패킷을 이용한 요청/응답 메시지는 이블 트윈이 대신 응답함으로써 회피가 가능하다. 그러나 지금은 사용자가 접속한 AP가 이블 트윈인지 인가된 AP인지 여부에 상관없이 RTT을 측정하면 되기 때문에 이블 트윈이 ICMP 패킷의 요청에 대해 대신 응답하여 준다는 것은 무의미하다. 따라서 사용자가 접속한 AP까지의 RTT 값은 ICMP 패킷을 이용하여 측정한다.

3.3.3 PIAT(Packet Inter-Arrival Time)의 측정

PIAT는 일정한 간격으로 전송한 패킷 사이 간격 시간이 인가된 AP와 이블 트윈 사이에 차이가 있는지 또는 채널의 혼잡상황이나 장치의 업무량에 따라 어떠한 변화가 있는지 확인하기 위한 값이다. 인가된 AP와 이블 트윈 사이의 차이점은 추가적인 1홉의 구간 즉, 추가적인 AP가 있는 것이므로 DNS 쿼리를 전송하여 DNS 서버까지 RTT을 측정하면서 PIAT을 측정하면 이블 트윈의 경우에는 반드시 추가적인 무선망이 존재하는 이블 트윈을 거쳐서 가게 되므로 인가된 AP 보다 패킷 간격 시간의 변화가 큰 것을 관찰할 수 있게 된다. 또한, ICMP 패킷을 사용자가 접속한 AP 까지 전송하여 RTT을 측정하면서 PIAT을 측정하여 인가된 AP와 이블 트윈사이에 어떠한 차이가 있는지 관찰할 수 있다. AP 까지 PIAT을 측정, 분석함으로써 현재 접속한 AP의 업무량이나 채널의 혼잡 상황의 척도를 알 수 있다. 따라서 ICMP 패킷 또한 일정한 간격으로 전송하여 PIAT를 측정한다.

3.3.4 데이터 필터링(Data Filtering)과 데이터 스무딩(Data Smoothing)

이 연구에서는 이블 트윈 탐지의 확률을 높이고자 데이터 필터링과 데이터 스무딩 기법을 사용한다. 단 한번의 RTT의 측정은 채널의 혼잡상황이나 AP의 신호세기에 따라 변동폭이 크기 때문에 여러번의 RTT을 측정하여 SVM의 요소로 사용하도록 한다. 이 연구에서는 DNS까지의 RTT과 AP까지의 RTT을 모두 300번을 측정한다. 이렇게 측정한 300개의 RTT에서 지나치게 큰 값은 데이터 분류에 영향을 미칠 수 있으므로 RTT이 큰 10%의 데이터는 제거하고 SVM의

요소로 사용한다. PIAT는 일정한 시간 간격이 패킷이 돌아오는 순서에 따라 기존 간격을 기준으로 줄어들 수도 있고 늘어날 수도 있다. 따라서, PIAT에서는 PIAT가 비정상적으로 작은값 5%와 PIAT가 큰값 5%를 제외하고 SVM의 요소로 넣어주는 방법을 사용한다.

3.3.5 SVM 분류

SVM은 데이터의 특징들을 요소로 넣어주었을 때 초평면을 이용하여 데이터 그룹 간의 경계면을 설정하고 분류하는 기계 학습 분류기이다. 하나의 데이터를 특징지을 수 있는 특징점이 많을수록 데이터들을 분류하는데 더 많은 도움이 된다. 따라서 우리는 인가된 AP와 이블 트윈에서 측정한 RTT과 PIAT을 가공하지 않고 그대로 SVM의 요소로 입력한다. 이렇게 함으로써 SVM은 RTT과 PIAT을 이용하여 인가된 AP와 이블 트윈의 차이점을 밝혀내고 적절한 비선형적 분류 경계면을 찾아낸다.

3.4 탐지 알고리즘

그림 5의 Algorithm 1.에서처럼 우리는 SVM의 요소로 넣어줄 RTT을 구하기 위해 서버까지의 RTT으로 DNS 쿼리를 300번 전송하여 300개의 RTT_{dns}를 구하고, AP까지의 RTT 값으로 ICMP 패킷을 300번 전송하여 300개의 RTT_{probe}를 구한다. RTT을 측정하기 위한 DNS 쿼리와 ICMP 요청 패킷은 전송할 때에 동일한 시간 간격인 30000 마이크로초로 300번 전송하면서 해당 패킷의 응답 패킷이 도착하였을 때를 측정하여 패킷 도착 시간 간격을 측정한 값을 PIAT라고 하여 300개의 PIAT를 측정한다. 측정한 RTT 값에서 비정상적으로 값이 큰 10%를 제외한 270개의 데이터를 저장하고, PIAT 값에서 비정상적인 값 상위 5%와 하위 5%를 제외한 270개의 데이터를 측정하면, 서버까지의 RTT_{dns}값 270개, $IIAT_{dns}$ 값 270개, AP까지의 RTT_{probe} 값 270개, $IIAT_{probe}$ 값 270개를 포함하여 하나의 AP에 대하여 1080개의 특징을 도출한다. 이러한 데이터들의 특징을 평균을 구하거나 표준편차를 구하는 등으로 가공하지 않고 SVM의 요소로 직접 넣어줌으로써 SVM이 비선형적으로 분류 기준을 정하고 동적으로 변하는 채널상황과 다양한 무선망 환경에서 좀 더 정확하게 이블 트윈을 탐지할 수 있게 된다. 하나의 AP에 접속하여 측정한 특징을 SVM의 요소로 입력할 때, 훈련데이터로 사용할 데이터는 훈련군에 넣으며, 이블 트윈 여부를 판단하기 위한 예측을 할 데이터는 예측군으로 넣는다. 훈련군을 사용하여 SVM 모델 파일이 생성되면, 예측군을 사용하여 이블 트윈 여부를 판단한다.

***Algorithm* 1.** *RTT Record*

1. *Connect* and *Associate with AP*
2. **for** $i = 1$ to 300 **do**
3. Send unicast *ICMPrequest* to AP
   Record Round Trip Time $RTTprobe[300]$
   Record Packet Inter-Arrival Time $PIATprobe[300]$
4. Send *DNSquery* to Local DNS server
   Record Round Trip Time $RTTdns[300]$
   Record Packet Inter-Arrival Time $PIATdns[300]$
5. **end for**
6. Except the top 10% abnormal RTT value
7. Except the top 5% & bottom 5% abnormal $PIAT$ value
8. $RTTprobe[270]$ = 270 values of the rest of the $RTTprobe$
9. $RTTdns[270]$ = 270 values of the rest of the $RTTdns$
10. $PIATprobe[270]$ = 270 values of the rest of the $PIATprobe$
11. $PIATdns[270]$ = 270 values of the rest of the $PIATdns$
12. $Feature\ of\ AP$ = ($RTTprobe[270]$, $RTTdns[270]$, $PIATprobe[270]$, $PIATdns[270]$)
13. **if** $data-set$ is a $training-set$ **then**
    Insert "$Feature\ of\ AP$" into SVM $training-set$ as Factor
14. **else if** $data-set$ is a $predict-set$ **then**
    Insert "$Feature\ of\ AP$" into SVM $predict-set$ as Factor
15. **end if**

그림 5. 다중 요소를 가지는 SVM을 이용한 이블 트윈 탐지 방법 알고리즘
Fig. 5. Evil-Twin Detection Scheme Algorithm using SVM with Multi-Factor

Ⅳ. 실험 환경

4.1 실험 환경

인가된 AP와 이블 트윈의 상황을 만들기 위하여 그림 5와 같은 실험환경을 구성하였다. 무선망 규격은 802.11g로 통일하여 54Mbps로 전송되고 사용자가 접속한 AP와의 채널은 1번, 이블 트윈이 인가된 AP에 연결하는 채널은 11번으로 설정하였다. 이것은 사용자가 접속한 AP의 채널을 혼잡 시킬 때 인가된 AP와 이블 트윈 모두에서 동일하게 1번을 혼잡 시키기 위한 것뿐만 아니라 사용자가 접속한 AP까지의 RTT 값을 측정할 때도 동일한 채널을 사용하기 위한 것이다.

일반적으로 인가된 AP는 양호한 신호 세기와 채널 사용량을 제공하여 주지만, 데이터의 전송이 많아지거나 많은 사용자가 접속하게 되면 채널이 혼잡해져 RTT가 증가하여 이블 트윈으로 판단될 수 있다. 또한, 이블 트윈의 경우 사용자가 접속한 AP까지의 채널이 혼잡해져 RTT이 증가하게 되면, 이블 트윈과 인가된 AP 사이에 발생하는 추가적인 무선구간에 따른 지연시간에 대한 RTT이 상대적으로 적어지게 되어 이블 트윈을 인가된 AP로 판단할 수도 있다. 따라서, 우리의 탐지 기법이 혼잡한 채널 상황에서도 이블 트윈과 인가된 AP를 잘 분류해내는 것을 보이기 위해 사용자가 접속한 AP의 채널을 혼잡하게 하는 상황을 만든다.

이블 트윈 공격자는 RTT을 이용한 탐지 방법을 회피하기 위하여 속도가 빠르고 채널사용량이 적은 인

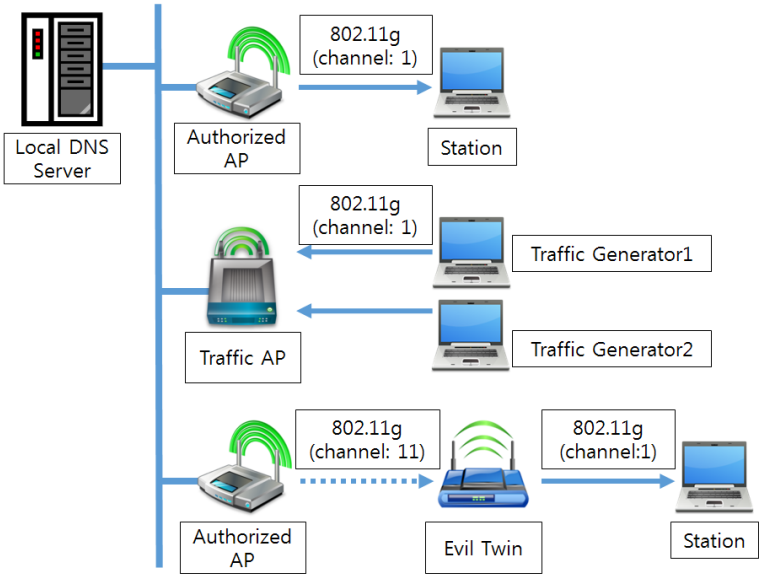

그림 6. 실험 환경
Fig. 6. Experiment environment

가된 AP를 선택하여 사용해야 추가적인 무선망에 대한 지연시간을 최소화할 수 있다. 사용자가 접속한 AP의 채널을 혼잡시킬 때 이블 트윈과 인가된 AP 사이의 채널과 사용자와 이블 트윈사이의 채널이 같거나 비슷할 경우, 이블 트윈과 인가된 AP 사이의 채널도 함께 혼잡한 상황이 되어 RTT이 증가하여 이블 트윈을 탐지하기 쉬워지므로 채널을 1번과 11번으로 나누어 설정함으로써 공격자에게 유리한 상황으로 설정한다. 무선채널은 신호가 겹쳐있기 때문에 채널이 혼잡한 경우 4개정도의 채널에 영향을 미치므로 1번과 11번으로 설정하여 사용자가 접속한 AP의 채널을 혼잡하게 하는 것이 이블 트윈과 인가된 AP 사이의 채널에 영향을 미치지 않는다.

채널을 혼잡 시키기 위해 별도의 AP를 사용하고 트래픽을 생성하는 장치도 2대를 사용한다. 채널을 혼잡 시키기 위한 데이터를 전송하는데 이블 트윈이나 인가된 AP를 사용하게 되면, 해당 AP의 업무량이 증가하기 때문에 측정하고자 하는 RTT이 증가한다. 이것은 이블 트윈과 인가된 AP 사이의 추가적인 무선망 때문에 발생한 지연시간이 아니기 때문에 인가된 AP를 이블 트윈으로 잘못 판단할 수 있게 되므로 AP의 업무량이 RTT에 영향을 주지 않도록 별도의 AP를 사용하여 트래픽을 전송한다.

서버까지의 RTT을 측정하기 위해 사용하는 Local DNS 서버는 교내 DNS 서버(7홉 이내)를 사용한다. H. Han 등은 서브네트워크 내에 있는 DNS 서버(1홉 이내)를 사용하여 RTT를 측정함으로써 유선망에 대한 지연시간을 최소화하여 이블 트윈이 가지게 되는 추가적인 무선망에 대한 지연시간을 증가시켜 이블 트윈을 탐지하는 확률을 높였다. 그러나 서브네트워크 내에 DNS 서버를 가지고 있는 경우는 일반적이지 않으며 공격자의 능력을 약화 시키는 요인이 된다. 따라서 우리는 교내 DNS 서버를 사용함으로써 일반적인 상황을 가정하며 실제로 우리의 탐지 방법을 적용하기가 쉽게 만들었다.

4.2 실험 장비 구성

표 1에서는 실험에 사용된 장비를 설명하고 있다. 사용자는 Windows 7 환경에서 실험 되었고 802.11b,g,n을 모두 지원하는 네트워크 카드를 사용하였다. 트래픽 생성기로 2대의 노트북이 사용되었으며, 802.11b,g,n과 802.11b,g를 제공하는 노트북이 사용되었다. 이블 트윈 공격을 위한 로그 AP는 H. Han 등의 실험과 동일하게 실제 AP를 사용하는 것이 아니라, 노트북을 이용하여 랜카드를 2장 꼽는 형태로 사용된다. 이러한 방법은 로그 AP의 성능은 약간 떨어질 수 있지만, 적극적으로 로그 AP의 위치를 변경해가며 실제 사용자에게 강한 신호를 보내어 이블 트윈에 접속하도록 유도하기가 쉽다. 성능이 좋은 로그 AP를 사용하여 공격자의 능력을 강화할 수 있지만 본 연구에서는 기존 연구의 실험 환경과 동일하게 구성함으로써 로그 AP의 성능과 상관없이 탐지 알고리즘 자체

표 1. 실험 장비 구성
Table 1. Hardwares in Experiment

Computer				
	OS	CPU	RAM	Protocol
Station	Windows 7 Ultimate 64bit	Intel i3-2100(3.1Ghz)	8G	802.11b,g,n
Traffic Generator1	Windows 7 Ultimate 64bit	Intel i5 460m(2.5Ghz)	8G	802.11b,g,n
Traffic Generator2	Windows 7 Ultimate 32bit	Intel Core 2 Duo L7100(1.2 Ghz)	4G	802.11b,g
Rogue AP	Windows 7 Ultimate 64bit	Intel i5 3317U(1.7Ghz)	8G	802.11b,g,n
Access Point				
	Device Name	Protocol		
Legitimate AP	D-Link DIR-655	802.11b,g,n		
Traffic AP	D-Link DIR-655	802.11b,g,n		

성능의 우수성을 보이고자 하였다. 실험에 사용되는 AP는 D-Link에서 만든 안테나가 3개 달린 성능 좋은 DIR-655 모델로서 802.11b,g,n을 모두 지원한다. 우리의 실험진행을 위한 장비들은 802.11g를 지원하기에 충분한 성능을 가지고 있으며 이블 트윈의 성능이나 트래픽 생성기의 성능도 뒤쳐지지 않는다. 향후 연구에서 802.11n을 위한 실험을 하기에 충분하다.

4.3 소프트웨어 및 라이브러리

4.3.1 WinPcap

이 연구에서는 패킷의 전송과 캡쳐를 위하여 WinPcap 라이브러리를 사용한다. RTT_{dns}와 RTT_{probe}를 측정하기 위하여 패킷의 헤더를 직접 계산 및 작성하여 RawPacket으로 만들어서 전송한다. WinPcap은 전송받은 패킷이 Link-Layer에 도착하였을 때 시간을 측정하기 때문에 패킷이 상위계층을 거쳐 응용계층까지 오는데 걸리는 시간을 단축할 수 있다. 전송하는 패킷 또한 네트워크 카드로 직접 전송해 줄 수 있어 상위계층에서의 지연시간을 단축할 수 있다. WinPcap을 이용함으로써 패킷을 처리하는데 걸리는 시간을 절약할 수 있으며, 무선망과 유선망에서 전송되는 시간만을 더욱 정확하게 측정한다.

4.3.2 SVM

우리는 여러 가지 형태의 SVM의 타입과 커널 펑션을 선택할 수 있는 LibSVM을 이용한다. SVM의 타입은 C-SVC를 이용하고 커널의 형태는 RBF(Radial Basis Function)을 사용한다. 데이터를 훈련시킬 때 수집한 데이터 자체를 직접 SVM의 요소로 사용하여도 되지만 이블 트윈 탐지율을 높이기 위하여 LibSVM의 스케일링을 활용하여 데이터들을 균일한 값으로 스케일링한다. 스케일링된 데이터들은 기본적으로 [-1, 1] 사이의 값으로 변환된다. 또한, 데이터를 스케일링한 다음 선택된 SVM을 사용하기 위해 데이터를 훈련시킬 때에 C-SVC를 위한 Cost(-c) 옵션과 RBF를 위한 Gamma(-g) 옵션에 최적의 옵션 값을 선택하여 주어야 한다. 이 옵션 값은 데이터에 따라 최적의 값이 다르기 때문에 실험적으로 구해서 사용자가 결정해야 하지만, LibSVM에서 제공하는 Grid.py 라는 스크립트 파일을 실행하면 현재 데이터 그룹에 대하여 최적의 Cost와 Gamma 값을 구할 수 있다. 따라서, 데이터를 스케일링한 다음 Grid.py 파일을 실행하여 Cost 값과 Gamma 값을 찾아내어 해당 값을 옵션으로 하여 데이터를 훈련한 다음, 이렇게 잘 훈련된 훈련군을 이용하여 다음에 들어올 예측군을 성공적으로 분류할 수 있다.

V. 실험 결과

이블 트윈의 분류를 위한 RTT과 PIAT의 측정은 채널이 유휴한 상황일 때 인가된 AP와 이블 트윈에서 측정되었고 채널이 혼잡한 상황일 때 인가된 AP와 이블 트윈에서 측정하였다. 패킷 전송 간격을 30000 마이크로초로 설정하여 전송하며 300개의 RTT 값과 PIAT를 구하기 위하여 9초 정도의 시간이 소요되었다. 각 시간을 측정하는데 오류를 줄이기 위해 각 데이터를 수집하는 간격을 20초로 설정하였다. 따라서, 1분 동안 3개의 데이터를 얻을 수 있으며, 1시간동안 180개의 데이터를 얻을 수 있고, 24시간 동안 4320개의 데이터를 얻을 수 있다. 이 연구에서는 유휴 상태의 인가된 AP, 유휴 상태의 이블 트윈, 혼잡 상태의 인가된 AP, 혼잡 상태의 이블 트윈 각각에서 약 24시간동안 측정된 12만개의 측정된 값으로 구성된 4000

개의 데이터를 가지고 인가된 AP와 이블 트윈의 차이점을 분석하고 측정된 데이터를 SVM의 요소로 넣어서 이블 트윈을 탐지한다.

5.1 RTT(Round Trip Time)의 분석

그림 7은 유휴 상태에서 측정된 RTT의 분포도를 나타낸 것이다. AAP(Authorize-AP)는 인가된 AP에서 측정된 RTT 값을 나타내고 EAP(Evil-Twin AP)는 이블 트윈에서 측정된 RTT 값을 나타낸다. 인가된 AP와 이블 트윈에서 측정된 RTT_{probe}은 대부분이 500마이크로초에서 1000마이크로초 사이에 있으며 두 그룹 사이에 크게 차이가 없음을 알 수 있다. RTT_{dns}는 인가된 AP에서는 대부분이 1500 마이크로초에 분포하고 있고 이블 트윈에서는 2300 마이크로초 정도에서 대부분의 값이 측정됨을 알 수 있다. 따라서 유휴상태일 때에는 RTT_{dns}에서 인가된 AP와 이블 트윈 사이에 명확한 차이가 존재하기 때문에 두 그룹을 RTT 값을 보고 분류하는 것만으로도 이블 트윈을 탐지할 수 있어 보인다.

그림 8은 혼잡 상태에서 측정된 RTT의 분포를 나타낸 것이다. 혼잡 상태에서 측정된 RTT은 유휴 상태일 때와는 다르게 RTT이 특정 구간에 많이 분포하지 않고 광범위하게 다양한 값을 가지며 분포되어있다. 혼잡 상태에서 측정된 RTT_{probe}는 많은 값이 약 1000 마이크로초에서 4000마이크로초 정도에서 분포하고 있다. 인가된 AP와 이블 트윈 사이에 RTT_{probe}의 값은 큰 차이를 보이지 않는다. 이것은 RTT_{probe} 값이 혼잡 상태를 나타내는 지표가 될 수는 있으나

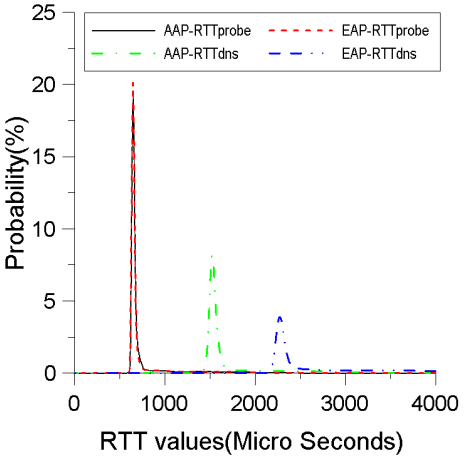

그림 7. 채널 유휴 상태일 때 측정된 RTT의 분포
Fig. 7. Distribution of Round Trip Time in Idle Congestion

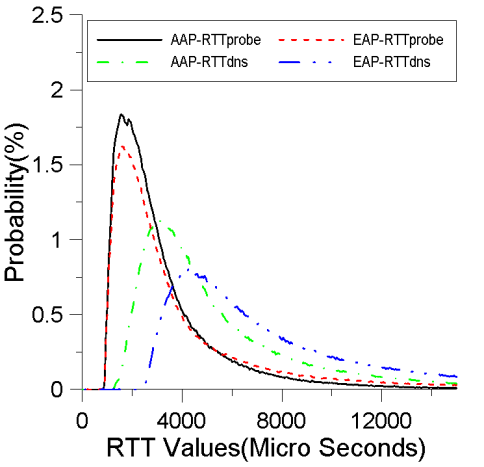

그림 8. 채널 혼잡 상태일 때 측정된 RTT의 분포
Fig. 8. Distribution of Round Trip Time in Full Congestion

RTT_{probe}만으로 인가된 AP와 이블 트윈을 구분하기는 힘들다는 것을 의미한다. 혼잡 상태에서 RTT_{dns} 또한 유휴 상태에서 측정된 RTT_{dns}와는 다르게 다양한 시간에서 측정되어 분포하고 있어서 RTT_{dns} 값만을 가지고 인가된 AP와 이블 트윈을 분류하는 것은 어려워 보인다.

RTT을 측정함으로써 유휴 상태에서는 인가된 AP와 이블 트윈이 명확한 차이를 보이기 때문에 두 그룹을 분류하는 것이 어렵지 않아 보였다. 그러나 혼잡 상태에서 측정된 RTT 값은 두 그룹을 분류하는데 명확한 차이가 보이지 않기 때문에 RTT만으로 정적인 기준을 정하여 이블 트윈을 분류해서 탐지하는 것은 한계가 있음을 짐작할 수 있다.

5.2 PIAT(Packet Inter-Arrival Time)의 분석

PIAT의 측정을 위해서 일정한 간격으로 전송한 패킷들의 응답이 도착했을 때 시간 간격에 변화가 있는지 살펴보기 위해 30000 마이크로초의 일정한 간격으로 전송된 패킷들의 응답이 도착한 패킷으로부터 PIAT을 측정한다. 30000 마이크로초 이하의 간격으로 패킷을 전송하면 채널이 혼잡한 상황에서 PIAT 값의 분산정도를 나타내기에 값이 적기 때문에 유휴 상태일 때와 혼잡 상황일 때에 인가된 AP와 이블 트윈의 PIAT 값의 차이를 나타내기에 충분한 값이 30000 마이크로초의 간격으로 전송하였다.

그림 9, 그림 10는 측정된 PIAT의 분포를 보여주는 박스-플롯이다. 그림 9는 유휴 상태에서 측정된 PIAT 의 분포를 표현한 그래프이다. 유휴 상태에서

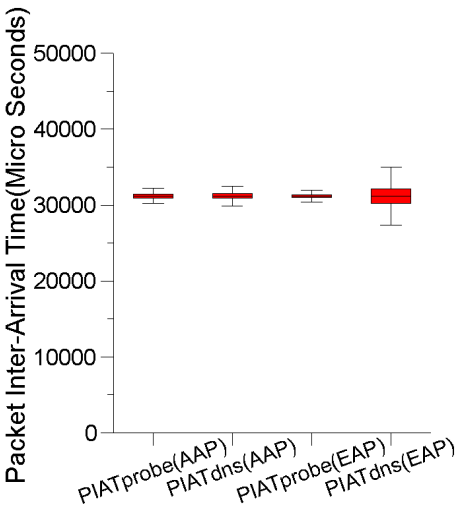

그림 9. 채널 유휴 상태일 때 PIAT의 분포
Fig. 9. Distribution of Packet Inter-Arrival Time in Idle Congestion

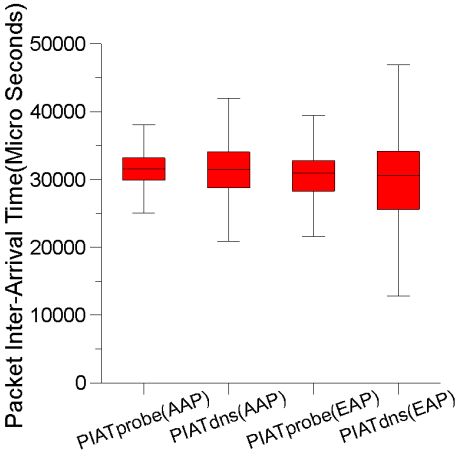

그림 10. 채널 혼잡 상태일 때 PIAT의 분포
Fig. 10. Distribution of Packet Inter-Arrival Time Distribution in Full Congestion

인가된 AP와 이블 트윈의 ΠAT_{probe}는 패킷을 전송할 때의 간격인 30000 마이크로초에서 크게 벗어나지 않고 일정한 간격으로 도착한 것을 알 수 있다. 유휴 상태에서 ΠAT_{dns}는 인가된 AP에 비하여 이블 트윈의 값이 편차가 크기 때문에 ΠAT_{dns}를 이용하여 두 그룹을 어느 정도 분류할 수 있는 것으로 보인다.

그림 10는 혼잡 상태에서 측정된 PIAT 값의 분포를 표현한 그래프이다. 혼잡 상태에서의 PIAT 값은 유휴 상태일 때와 비교하였을 때 30000 마이크로초를 기준으로 전체적으로 편차가 큰 경향을 보인다. 혼잡 상태에서 ΠAT_{probe}는 인가된 AP와 이블 트윈이 비슷한 편차를 가지기 때문에 두 그룹을 분류할 수 있는 기준은 될 수 없으나 이러한 경향을 채널의 혼잡상황이나 AP의 업무량의 지표로 볼 수 있을 것이라 예상한다. 혼잡 상태에서 ΠAT_{dns}도 유휴 상태와 비슷하게 인가된 AP 와 비교하였을 때 이블 트윈의 편차가 좀 더 크기 때문에 ΠAT_{dns}가 두 그룹을 어느 정도 분류할 수 있는 것으로 보인다.

5.3 SVM 결과

인가된 AP와 이블 트윈을 분류하기 위하여 각각 270개의 RTT_{probe}, RTT_{dns}, ΠAT_{probe}, ΠAT_{dns}를 하나로 모아서 1080개의 특징을 가지는 하나의 데이터를 SVM의 요소로 입력한다. 총 데이터는 4일 동안 유휴 상태의 인가된 AP와 이블 트윈, 혼잡 상태의 인가된 AP와 이블 트윈에서 각각 4000개의 데이터가 수집되었다. SVM의 분류를 위하여 각 회차에 200개 인가된 AP와 200개의 이블 트윈에서 측정된 400개의 데이터로 훈련된 훈련군에 200개의 인가된 AP와 200개의 이블 트윈에서 측정된 400개의 예측군으로 총 10회를 예측한다.

표 2는 유휴 상태일 때 SVM을 이용하여 이블 트윈을 탐지한 결과이다. Positive는 인가된 AP로 판단한 경우이며 Negative는 이블 트윈으로 판단한 경우이다. 측정된 값을 분석 하였을 때 유휴 상태에서는 RTT 뿐만 아니라 PIAT 값 또한 인가된 AP와 이블 트윈을 구별하기에 명확한 차이를 보여주었기 때문에

표 2. 채널 유휴 상태일 때 SVM을 이용한 이블 트윈 탐지율
Table 2. Evil-Twin Detection Rate using SVM in Idle Congestion

Number of Test	1th	2th	3th	4th	5th	6th	7th	8th	9th	10th
Detection Rate	100% 400/400	100% 400/400	100% 400/400	99% 396/400	100% 400/400	100% 400/400	99% 399/400	100% 400/400	100% 400/400	100% 400/400
False Positive	0%	0%	0%	0.5%	0%	0%	0%	0%	0%	0%
False Negative	0%	0%	0%	1.5%	0%	0%	0.5%	0%	0%	0%

표 3. 채널 혼잡 상황일 때 SVM을 이용한 이블 트윈 탐지율
Table 3. Evil-Twin Detection Rate using SVM in Full Congestion

Number of Test	1th	2th	3th	4th	5th	6th	7th	8th	9th	10th
Detection Rate	96.50% 386/400	91.25% 365/400	92.25% 369/400	92% 368/400	93.00% 372/400	93.50% 374/400	94.25% 377/400	89.75% 359/400	94.50% 378/400	95.75% 383/400
False Positive	5%	9.5%	0%	5%	3.5%	6.5%	6.5%	8.5%	2%	1.5%
False Negative	7%	8%	15.5%	11%	10.5%	6.5%	5%	12%	9%	7%

측정된 값을 SVM의 요소로 하는 분류를 통해서도 100%에 가까운 탐지율을 보여준다. 긍정 오류나 부정 오류는 낮은 확률임을 보인다.

표 3은 혼잡 상태일 때 SVM을 이용하여 이블 트윈을 탐지한 결과이다. 채널 혼잡 상태일 때는 최저 89.75%, 최고 96.5% 의 확률로 인가된 AP와 이블 트윈을 분류하여 이블 트윈을 탐지하였다. 부정 오류가 최대 15.5% 인 경우가 있었으나 대부분은 10% 이하의 확률을 보여주고 긍정 오류는 10% 이하의 확률을 보여준다.

5.4. 탐지율

그림 11은 H. Han 등의 실험에 의한 이블 트윈의 탐지율과 이 연구에서 실험한 SVM을 이용한 이블 트윈 탐지 결과를 비교하여 나타내었다. C. Yang 등의 연구도 RTT을 이용한 연구이기는 하지만 채널이 혼잡해지는 경우에 탐지 알고리즘이 제대로 동작하지 않기 때문에 비교 대상에서 제외한다. H. Han 등의 탐지 알고리즘은 RTT을 100번 측정하여 평균과 표준편차를 계산하고 실험에 의해 도출된 기준으로 인가된 AP와 이블 트윈을 분류 하였을 때 혼잡한 상황에서는 60%까지 탐지율이 떨어진다. 동일한 방법으로 RTT을 300번 측정함으로써 H. Han등은 채널이 혼잡한 상황에서 80% 까지 탐지율을 높일 수 있었다. 우리는 이전 연구에서 H. Han 등의 실험과 동일하게 RTT을 측정하여 평균과 표준편차를 이용하면서 이 값으로 인가된 AP와 이블 트윈을 분류할 때 정적인 분류 기준을 사용하지 않고 SVM을 사용함으로써 채널이 혼잡한 상태일 때에도 탐지율을 90%까지 향상시켰다. 이 연구에서는 이전 연구에 PIAT 값을 추가하여 측정하고 데이터를 가공하지 않고 SVM의 요소로 넣어줌으로써 유휴 상태일 때는 물론이며 혼잡한 상황에서도 평균 93%, 최저 89.75%, 최대 96.5% 의 탐지율을 보여준다.

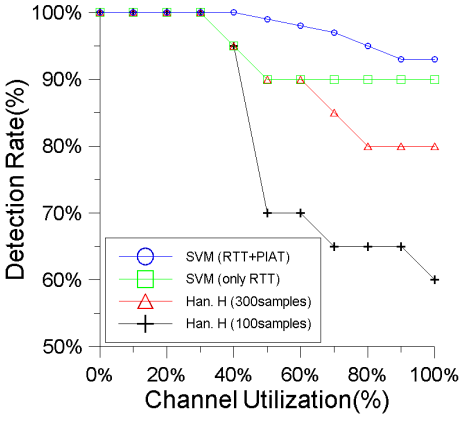

그림 11. 이블 트윈 탐지율 비교
Fig. 11. Comparison of Evil-Twin Detection Rate

Ⅵ. 결론 및 향후연구

이 연구의 목표는 사용자가 직접 RTT과 PIAT을 측정함으로써 채널이 혼잡한 상태에서도 이블 트윈을 높은 확률로 탐지해내는 것이다. 기존의 연구에서 C. Yang 등은 AP 까지의 RTT과 서버까지의 RT을 측정하고 서버-AP 비율을 계산하여 인가된 AP와 이블 트윈을 분류함으로써 100%에 가까운 확률로 이블 트윈을 탐지하였다. 그러나, C. Yang 등의 연구에서는 사용자가 접속한 무선망의 채널이 혼잡하지 않은 상황에서만 성공적으로 이블 트윈을 탐지할 수 있으며 채널이 혼잡할 때는 자신들의 탐지 방법이 적절하게 적용되지 않을 수 있음을 연구의 한계점으로 밝혔다. H. Han 등은 DNS 서버를 사용하여 서버까지의 RTT을 측정하고 사용자가 접속한 AP까지의 RTT을 측정하여 서버까지의 RTT에서 빼는 방법을 이용하여 이블 트윈을 탐지하였다. 이블 트윈을 탐지하는 기준으로 추가적으로 RTT의 표준편차를 활용하였다. 이러한 H. Han. 등의 연구는 RTT과 표준편차를 활용한 이블 트윈 탐지 방법은 사전 실험을 통하여 도출된 특정한 값을 기준으로 하여 이블 트윈과 인가된 AP를 분류하

기 때문에, 동적으로 변하는 무선망환경에서 실질적으로 적용되기에 한계가 있음을 이전 연구에서 증명하였다. 또한, H. Han. 등의 방법을 이용하면서 동적으로 기준을 정한다 하더라도 채널이 혼잡한 상황에서 인가된 AP와 이블 트윈의 RTT 값이 비슷해질 때에는 이러한 단순 선형 분류방법으로는 한계가 있음을 이전 연구에서 증명하였다.

이 연구에서는 기존의 연구의 한계점들을 분석하여 RTT을 활용하지만 특정한 값으로 가공하지 않으며 추가적으로 PIAT를 계산함으로써 인가된 AP와 이블 트윈을 더욱 정확하게 분류하였다. 측정된 RTT과 PIAT을 가지고 인가된 AP와 이블 트윈을 분류하는데 있어서 정적인 기준을 가지고 선형 분류를 하지 않고 SVM을 이용하여 비선형 분류를 함으로써 동적으로 분류기준을 정하여 두 그룹을 효과적으로 분류하였다.

H. Han 등의 연구에서는 채널이 혼잡한 상황에서 샘플링의 개수를 300개로 늘림으로써 80%의 확률로 이블 트윈의 탐지에 성공하였다. 이 연구에서는 RTT과 PIAT를 측정하는데 H. Han 등과 같이 300개의 샘플링을 사용하면서 측정시간을 크게 늘리지 않고 최고 96.5% 최저 89.75% 정도의 탐지율을 보여주었다. PIAT없이 RTT을 사용하여 H. Han. 등과 동일한 요소를 사용하여 SVM으로 분류했던 이전 연구에서는 채널이 혼잡한 상황에서 90%의 탐지율을 보였던 것에 비하여 탐지율이 향상되었음을 알 수 있다.

앞으로의 연구에서는 RTT이나 PIAT을 측정하는데 시간을 단축시키면서 탐지율을 떨어뜨리지 않는 연구가 필요해 보인다. 또한 이블 트윈의 기능 자체가 노트북에서 소프트웨어로 구성되어 있다는 점으로 볼 때 성능 저하의 요인이 될 수 있으므로 이블 트윈의 성능을 높이면서 높은 탐지율을 유지할 수 있는 탐지 기법을 개발하는 것 또한 향후 연구문제로 남겨놓는다.

References

[1] S. Kang, D. Nyang, J. Choi, and S. Lee, “Relaying rogue AP detection scheme using SVM,” *J. KIISC*, vol. 23, no. 3, pp. 431-444, Jun. 2013.

[2] P. Bahl, R. Chandra, J. Padhye, L. Ravindranath, M. Singh, A. Wolman, and B. Zill, “Enhancing the security of corporate Wi-Fi networks using DAIR,” *MobiSys*, pp. 1-14, Jun. 2006.

[3] D. Schweitzer, W. Brown, and J. Boleng, “Using visualization to locate rogue access points,” *J. Computing Sci. in Colleges*, vol. 23, no. 1, pp. 134-140, Oct. 2007.

[4] S. Jana and S. K. Kasera, “On fast and accurate detection of unauthorized wireless access points using clock skews,” *IEEE Trans. Mob. Computing,* vol. 9, no. 3, pp. 449-462, Mar. 2010.

[5] L. Watkins, R. Beyah, and C. Corbett, “A Passive approach to rogue access point detection,” *IEEE Global Telecommun. Conf. (GLOBECOM ’07)*, pp. 355-360, Washington DC, USA, Nov. 2007.

[6] W. Wei, K. Suh, B. Wang, Y. Gu, J. Kurose, and D. Towsley, “Passive online rogue access point detection using sequential hypothesis testing with TCP ACK-pairs,” in *Proc. 7th ACM SIGCOMM Conf. Internet Measurement (IMC ’07)*, pp. 365-378, NY, USA, Oct. 2007.

[7] I. Kim, J. Cho, T. Shon, and J. Moon, “A method for detecting unauthorized access point over 3G network,” *J. KIISC*, vol. 22, no. 2, pp. 259-266, Apr. 2012.

[8] Y. Sheng, K. Tan, G. Chen, D. Kotz, and A. Campbell, “Detecting 802.11 MAC layer spoofing using received signal strength,” *The 27th Conf. Comput. Commun. IEEE*, (*INFOCOM 2008),* Phoenix, AZ, USA, Apr. 2008.

[9] J. Park, M. Park, and S. Jung, “A whitelist-based scheme for detecting and preventing unauthorized AP access using mobile device,” *J. KICS*, vol. 38, no. 8, pp. 632-640, Aug. 2013.

[10] J. Mun and S. Jung, “A scheme for detecting and preventing an unauthorized device using context awareness and mobile device management,” *J. KICS*, vol. 39, no. 1, pp. 1-8, Jan. 2014.

[11] D. Shin, J. Kang, D. Nyang, S. Lee, and K. Lee, “A method of authenticating WLAN APs for smartphones,” *J. KICS*, vol. 39, no. 1, pp. 17-28, Jan. 2014.

[12] V. Brik, S. Banerjee, M. Gruteser, and S. Oh, “Wireless device identification with radiometric signatures,” *14th ACM Int. Conf. Mob. Comput. Netw. (Mobicom ’08)*, pp. 116-127, San Francisco,

CA, USA, Sept. 2008.

[13] L. Ma, A. Y. Teymorian, and X. Cheng, "A hybrid rogue access point protection framework for commodity Wi-Fi networks," *The 27th Conf. Comput. Commun. IEEE, (INFOCOM 2008),* Phoenix, AZ, USA, Apr. 2008.

[14] H. Yin, G. Chen, and J. Wang, "Detecting protected layer-3 rogue APs," 4*th Int. Conf. Broadband Commun. Netw. Syst. (BROADNETS 2007),* pp. 449-458, Raleigh, NC, USA, Sept. 2007.

[15] A. Adya, P. Bahl, R. Chandra, and L. Qiu, "Architecture and techniques for diagnosing faults in IEEE 802.11 infrastructure networks," *The 10th Annu. Int. Conf. Mob. Comput. Netw. (MobiCom '04)*, pp. 30-44, Philadelphia, USA, Sept. 2004.

[16] H. Han, B. Sheng, C. C. Tan, Q. Li, and S. Lu, "A timing-based scheme for rogue AP detection," *IEEE Trans. Parallel Distrib. Syst.*, vol. 22, no. 11, pp. 1912-1925, Nov. 2011.

[17] C. Yang, Y. Song, and G. Gu, "Active user-side evil twin access point detection using statistical techniques," *IEEE Trans. Inf. Forensics and Security,* vol. 7, no. 5, pp. 1638-1651, Oct. 2012.

강 성 배 (SungBae Kang)

2012년 2월 : 인하대학교 컴퓨터 공학과 졸업
2014년 2월 : 인하대학교 컴퓨터 공학과 석사
2014년 3월~현재 : 인하대학교 컴퓨터 공학과 박사과정
<관심분야>정보보호, 무선 인터넷 보안, 네트워크 보안

양 대 헌 (DaeHun Nyang)

1994년 2월 : 한국과학기술원 과학기술 대학 전기 및 전자 공학과 졸업
1996년 2월 : 연세대학교 컴퓨터과학과 석사
2000년 8월 : 연세대학교 컴퓨터과학과 박사
2000년 9월~2003년 2월 : 한국전자통신연구원 정보보호연구본부 선임연구원
2003년 2월~현재 : 인하대학교 컴퓨터정보공학부 부교수
<관심분야> 암호 이론, 암호 프로토콜, 인증 프로토콜, 무선 인터넷 보안

이 경 희 (KyungHee Lee)

1993년 2월 : 연세대학교 컴퓨터과학과 학사
1998년 8월 : 연세대학교 컴퓨터과학과 석사
2004년 2월 : 연세대학교 컴퓨터과학과 박사
1993년 1월~1996년 5월 : LG 소프트(주) 연구원
2000년 12월~2005년 2월 : 한국전자통신연구원 선임연구원
2005년 3월~현재 : 수원대학교 전기공학과 부교수
<관심분야> 바이오인식, 정보보호, 컴퓨터비전, 인공지능, 패턴인식