情報保護學會論文誌
第14券 第1號, 2004. 2

Support Vector Machine 기반 TCP/IP 헤더의 은닉채널 탐지에 관한 연구*

손 태 식,[a)†‡] 서 정 우,[a)] 서 정 택,[b)] 문 종 섭,[a)] 최 홍 민[c)]
고려대학교,[a)] 국가보안기술연구소,[b)] (주)씨큐브[c)]

A Study on the Covert Channel Detection in the TCP/IP Header based on the Support Vector Machine

Tae-Shik Shon,[a)†‡] Jung-Woo Seo,[a)] Jung-Taek Seo,[b)] Jong-Sub Moon,[a)] Hong-Min Choi[c)]
Korea University,[a)] National Security Research Institute,[b)] Secuve[c)]

요 약

폭발적으로 증가하는 인터넷 환경에서 정보보호는 가장 중요한 고려사항 중의 하나이다. 현재 이에 대한 대응방안으로 IDS, 방화벽, VPN 등 여러 보안 솔루션들이 사용되고 있지만 TCP/IP를 근간으로 하는 인터넷 환경은 기본적으로 프로토콜 자체의 취약성을 가지고 있다. 그 중에서도, TCP/IP 헤더 중 ICMP Payload, Identification(ID), Sequence Number(SEQ), Acknowledge(ACK), Timestamp의 필드 내용을 조작함으로써 특정 정보를 전송할 수 있는 은닉채널이 가능하다고 이미 알려져 있다.[1~3] 특히 본 논문에서는 TCP/IP 헤더의 여러 필드들 중에서도 IP 헤더의 ID 필드, TCP 헤더의 SEQ 필드를 이용한 은닉채널 탐지에 초점을 맞추었으며, 이러한 은닉채널의 탐지를 위하여, 패턴분류 문제 있어서 우수한 성능을 보이는 것으로 알려져 있는 Support Vector Machine(SVM)을 사용하였다. 본 논문의 실험결과에서는 제안된 탐지방안이 정상 TCP/IP 트래픽으로부터 은닉채널이 포함된 TCP/IP 패킷을 구분할 수 있음을 보여주었다.

ABSTRACT

In explosively increasing internet environments, information security is one of the most important consideration. Nowadays, various security solutions are used as such problems' countermeasure; IDS, Firewall and VPN. However, basically internet has much vulnerability of protocol itself. Specially, it is possible to establish a covert channel using TCP/IP header fields such as identification, sequence number, acknowledge number, timestamp and so on. In this paper, we focus on the covert channels using identification field of IP header and the sequence number field of TCP header. To detect such covert channels, we used Support Vector Machine which has excellent performance in pattern classification problems. Our experiments showed that proposed method could discern the abnormal cases(including covert channels) from normal TCP/IP traffic using Support Vector Machine.

Keywords: *covert channel, Support Vector Machine, TCP/IP*

1. 서 론

최근의 인터넷 환경은 급속한 네트워크의 보급과 함께 그에 따른 정보보호 문제를 발생시키고 있으며, 이러한 네트워크 보안 문제에 대한 해결책으로 IDS, 방화벽, VPN 등 여러 보안솔루션들이 대두되고 있다. 하지만 이러한 보안솔루션의 사용이 현재의 네트워크 보안에 대한 해결책으로 널리 사용될지라도 보안 솔루션들은 네트워크에 사용되는 프로토콜 자체의 문제점이나 보안 솔루션의 내부 결함으로 인한 다양

접수일: 2003년 9월 16일; 채택일: 2004년 1월 26일
* 본 연구는 대학 IT연구센터 육성 지원 사업에 의해 수행되었습니다.
† 주저자, 743zh2k@korea.ac.kr
‡ 교신저자, 743zh2k@korea.ac.kr

한 취약성을 가질 수 있다. 이때 여러 보안 솔루션들이 가지고 있는 취약성 중의 하나가, 이러한 보안 솔루션들이 작동하는 근간이 되는 TCP/IP 프로토콜 상의 은닉채널 생성의 가능성이다. 여기서 은닉채널이란 시스템의 보안 정책을 위반하는 어떤 방법을 사용하여, 정보를 전송하는 프로세스에 의해 이용될 수 있는 임의의 통신 채널로서 정의할 수 있다. 근본적으로 이와 같은 통신 채널은 일반적인 컴퓨터 설계상의 통신 수단이 아니며, 보통 특정 정보에 접근하는 것이 허락되지 않는 프로세스나 사용자들에게 정보를 전송하기 위한 수단으로서 사용된다.[4]

그러므로 본 논문에서는 TCP/IP 프로토콜에서 발생할 수 있는 여러 은닉채널 기법[1~3] 중 TCP/IP 헤더의 Identification(ID), Sequence Number (SEQ) 필드를 이용하여 은닉 데이터를 전송하는 기법을 분석하고, 분석된 TCP/IP 헤더의 ID, SEQ 필드를 이용한 은닉채널 생성 기법의 탐지를 위해서 Support Vector Machine(SVM)을 이용한다.

본 논문의 구성은 다음과 같다. 2장에서는 은닉채널에 대한 정의와 기본의 연구결과를 설명하고, 3장에서는 SVM의 개요 및 특징에 대해서 알아보고, 4장에서는 IP 헤더의 ID 필드, TCP 헤더의 SEQ 필드를 이용한 은닉채널 생성방안을 분석하며, 5장에서는 이러한 은닉채널을 탐지하기 위한 탐지 방안을 제안하고, 6장에서 SVM을 통한 실험을 수행한다. 그리고 마지막으로 7장에서 본 논문의 결론 및 향후 연구 방향을 제시한다.

II. 은닉채널 관련 연구

2.1 은닉채널

은닉채널은 그림 1과 같이 공개채널(Overt Channel)과 구분되어 정의된다.[8] 먼저, 은닉채널에서는 데이터를 전송하기 위해 일반적인 통신채널 즉, 공개채널을 사용하지만 이때 일반 통신채널 속에 포함된 특정 데이터에 대한 은닉화 기법은 은닉채널의 생성자 외에는 알 수 없다. 반면에, 공개채널에서는 은닉채널과 마찬가지로 일반적인 통신채널을 이용하지만, 데이터의 기밀성을 유지하기 위한 기법이 공개되거나 쉽게 구분이 될 수 있는 차이점을 가지고 있다.

은닉채널이란 용어는 Lampson의 논문 [7]에서 처음으로 소개되었으며, 그 개념에 대한 정의는 1985년의 미국 DOD의 기술문서 [4]에 잘 나타나 있다. 또한 유사한 의미로 "steganography," "information hiding," "subliminal channel" 등이 사용된다. 은닉채널은 보통 은닉저장채널(Covert Storage Channel)과 은닉시간채널(Covert Timing Channel)로 구분한다. 은닉저장채널은 임의의 프로세스가 특정 오브젝트의 값을 변경하게 되면 다른 프로세스는 그러한 오브젝트 값 변경의 결과를 관찰하여 특정 정보를 알아낼 수 있는 임의의 통신 채널을 지칭하며, 은닉시간채널은 임의의 프로세스가 CPU, I/O 등의 시스템 자원에 대한 어떤 효과를 유발하게 되면 그 결과가 상대 프로세스에 의해 관찰되고 이렇게 관찰된 시간을 바탕으로 특정 정보를 알아낼 수 있는 채널을 의미한다.[4,9] 은닉채널 자체에 대한 자세한 소개는 J. McHugh의 [9]를 참고할 수 있다.

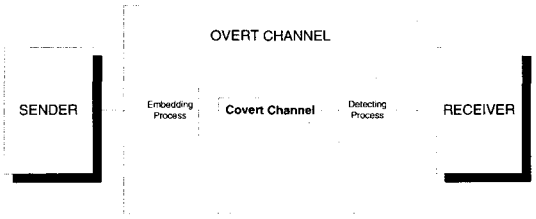

그림 1. 공개채널과 은닉채널

2.2 기존 연구 동향

은닉채널에 관한 연구는 1970~80년대에 [4] [7][9]와 같이 은닉채널의 개념 형성 및 정의에 관한 연구를 시작으로, 최근에는 멀티미디어 데이터에 대한 정보 은닉(Information Hiding)분야에서 활발하게 그 연구가 진행되고 있다.[10,11] 하지만, 네트워크 트래픽에서의 은닉채널 형성에 관한 연구는 "Covert Channels in the TCP/IP Protocol Suite"[1]에서 TCP/IP 프로토콜의 헤더 필드 중, IP identification 필드, TCP Acknowledge number 필드 그리고 Initial Sequence Number를 사용한 은닉채널 기법에 대해서 서술하고 있다. 또한 논문 [2]의 "Covert Messaging Through TCP Timestamps"에서는 TCP 프로토콜에서 Timestamp 필드를 이용한 은닉채널 방법에 대하여 서술하고 있다. 그외 논문 [3]의 "Loki: ICMP Tunneling"에서는 IP계층에서 ICMP 프로토콜 Payload를 이용한 방법을 기술하고 있다.

SVM은 1995년 Vapnik에 의해 제시된 Universal Feed Forward 네트워크의 한 종류로서 복잡한 패턴인식과 분류 문제에 효율적이며, 특히 이진분류 문제에 있어서 최적 솔루션으로 알려져 있다.[5,6]

은닉채널을 통한 공격과 같이 네트워크나 시스템에 대한 공격 탐지에 관한 연구는 주로 IDS를 통하여 이루어졌으며, 이러한 IDS에 관한 연구는 침입 모델에 따라서 오용탐지 기법과 비정상행위 탐지 기법으로 분리된다. 신경망, SVM 등을 이용한 비정상행위탐지 기법은 오용탐지 기법에 비해 변형된 공격기법이나 알려지지 않은 공격 기법 탐지에 있어 장점을 가진다.[12,13]

Ⅲ. Support Vector Machine(SVM)

3.1 Support Vector Machine 개요

SVM 이론에 따르면, 패턴 인식을 위한 전통적인 기법들이 경험적인 위험을 최소화하는데 기초한 반면, SVM은 구조적인 위험을 최소화하는 것에 기초하고 있다. 여기서 경험적 위험의 최소화는 훈련 집단의 수행도를 최적화하려는 노력을 말하고, 구조적 위험의 최소화는 고정되어 있지만 알려지지 않은 확률분포를 갖는 데이터에 대해 잘못 분류하는 확률을 최소화하는 것을 말한다. SVM의 장점은 우선 훈련 집단에 포함된 정보를 모으는 능력이 있다는 것과 상대적으로 낮은 공간의 결정 평면 집단을 사용한다는 것이다. 패턴 집단이 선형이고 분리 가능한 경우에 있어 SVM의 주요 아이디어는 간단히 설명될 수 있다. 기본적으로 SVM는 입력패턴들을 교사학습방법을 통하여, +1과 -1의 두 클래스로 패턴을 분류한다. 훈련 집단 S는 두 클래스로 분류되면, 각 클래스에 포함된 훈련 패턴들을 분리하는 초월면(Hyperplane)이 결정된다. 여기서 초월면이란 각 집단을 분리하는 절단 평면을 일컫는다. 이때, 초월면을 결정하는 입력 패턴들을 Support Vector라 한다. 패턴 집단이 분리 가능한 경우, 이 초월면은 면으로부터 Support Vector까지의 거리(마진)를 최대화하며, 모든 Support Vector는 초월면으로부터 같은 최소 거리에 위치해 있다. 그러나 실제로 패턴집단이 선형으로 분리되는 경우는 거의 드물고, 따라서 두 클래스는 선형적으로 분리가 불가능한 경우가 많을 것이다. 이 때의 초월면과 Support Vector는 제약식을 갖는 최적 문제의 해로부터 얻어진다. 최적해는 마진(각 클래스의 Support Vector사이의 거리)을 가장 크게 하는 것과 에러의 수를 최소화는 것 사이의 trade-off를 가지고 있고, 이는 정규화 된 파라미터에 의해 조정된다. 훈련 과정은 제약식을 갖는 이차 최적 문제를 풀기 위한 것과 기본적으로 같다.[5,6,14]

3.2 분류를 위한 Support Vector Machine

본 절에서는 SVM을 통한 분류를 위한 기본 개념을 알아본다.[10] 만약 훈련 데이터 $\{(x_i, d_i), i = 1, \ldots, N\}$가 주어졌을 때, x_i는 두 클래스 중 하나에 속하며, $d_i \in \{-1, 1\}$는 해당 클래스를 표시하는 라벨의 역할을 한다. SVM은 각 클래스를 구분하는 최적의 분리 경계면을 구하기 위해 분리 경계면과 가장 분리 경계면에 인접한 점과의 거리를 최대화한다. 최적의 선형 분리 경계면을 $f(x) = w^t x + b$로 놓으면, Support Vector와 $f(x)$의 거리를 $1/||w||$로 나타낼 수 있다. SVM은 $||w||^2$를 최소화하여 분리 간격을 최대화하도록 하여 최적 분리면을 찾아낸다. 이 문제는 다음과 같은 블록 최적화 문제가 된다.

$$\begin{cases} \text{minimize} & \frac{1}{2}\|w\|^2 \\ \text{subject to} & d_i(w^T x_i + b) \geq 1 \quad \text{for } i = 1, \cdots, N \end{cases} \tag{1}$$

이 문제를 라그랑제(Legendra) 배수로써 쌍대화(Dual Problem) 시키면 아래의 Quadratic 문제가 된다.

$$\Theta(\alpha) = \sum_1^N \alpha_i - \frac{1}{2}\sum_1^N \sum_1^N \alpha_i \alpha_j d_i d_j \langle x_i, x_j \rangle$$
$$\text{s.t. } \alpha_i \geq 0,\ i = 1, \cdots, N \text{ and } \sum_1^N \alpha_i d_i = 0 \tag{2}$$

선형 분리경계면으로 완전히 구분할 수 없는 서로 겹쳐져 있는 패턴의 경우에는 slack variable(ξ)을 사용한다. 식(1)로부터 아래의 모델과 같이 표현된다.

$$\text{minmize } \tau(w, \xi) = \frac{1}{2}\|w\|^2 + C\sum_1^N \xi_i \tag{3}$$
$$\text{s.t. } d_i(\langle w, x_i \rangle + b) \geq 1 - \xi_i,\ \xi_i \geq 0,\ i = 1, \cdots, N.$$

위 식(3)의 $d_i(w^t x_i + b) \geq 1 - \xi$에서 $\xi_i = 0\ (\forall_i)$이면 모든 패턴을 완전하게 분리할 수 있다는 것을 의

미한다. 그러나 대부분의 패턴은 선형적으로 분리가 가능하지 않다. 따라서 비선형 패턴을 분리하기 위하여 비선형 패턴의 입력 공간을 선형 패턴의 특징 공간으로 전환한다.

$$\theta(a) = \sum_i a_i^N - \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} a_i a_j d_i d_j K(x_i, x_j)$$
$$\text{s.t.} \sum_i^N a_i d_i = 0,\ 0 \le a_i \le C,\ \forall i. \quad (4)$$

즉, $x_i \Rightarrow \Phi(x_i)$에서 커널 함수 $K(x_i, x_j) = \Phi(xi)\Phi(x_j)$를 정의하면 비선형 패턴을 분리하기 위한 모델은 식(1), (2), (3)으로부터 아래와 같이 표현된다.

여기서 C는 식 (3)에서의 Penalty parameter이다. 위의 모델에서 라그랑제 배수 i를 구하면 특징 공간에서 가장 평평한 함수인 아래의 (5)를 구할 수 있다.

$$f(x) = \text{sgn}(\langle w, \phi(x)\rangle + b)$$
$$= \text{sgn}\left(\sum_{i=1}^{N} a_i d_i K(x_i, x) + b\right) \quad (5)$$

커널 함수로는 다음과 같은 함수 중 선택될 수 있다.

Dot kernel: x와 y의 내적
- $k(x, y) = x * y$

Polynomial kernel: d의 degree를 가짐
- $k(x, y) = (x * y + 1)^d$

Radial kernel: 파라미터 Υ를 가짐
- $k(x, y) = \exp(-\Upsilon||x - y||^2)$

Neural kernel: 파라미터 a, b를 가짐(MLP)
- $k(x, y) = \tanh(ax * y + b)$

Ⅳ. TCP/IP 헤더의 은닉채널

4.1 TCP/IP 헤더와 은닉채널

TCP/IP 프로토콜은 현재 인터넷에 사용되는 기본적인 프로토콜이며, 다양한 기능을 가지는 여러 계층으로 구성되어 있다. 본 논문에서는 그 중에서도 IP 프로토콜과 TCP 프로토콜의 헤더 부분에 은닉채널을 생성할 수 있는 방안을 분석하며, 이러한 은닉채널 분석을 통하여 탐지 방안을 연구하였다. 아래의 그림 2, 3에 IP 헤더와 TCP 헤더의 포맷이 나타나있다. IP 프로토콜과 TCP 프로토콜의 헤더에는 은밀한 방식으로 원격지에 있는 호스트에게 정보를 저장하거나 전송할 수 있는 많은 필드들을 가지고 있다는 것이 기존의 연구를 통해 이미 알려져 있다.[1~3] 그림 2, 3의 각 프로토콜의 헤더 내에는 Option 필드처럼 최대 세그먼트 사이즈, 윈도우 스케일 팩터 그리고 타임스탬프 등 여러 가지 전송기능을 설정하기 위해서 사용될 뿐 통신과정에는 사용되지 않거나 또는 송신자의 필요로 인해서만 설정되는 필드들과 ID, SEQ, Flags, Control bit 등 통신 과정에 필수적으로 사용되는 필드들을 가지고 있다. 이때 위와 같은 구분에 의해 분류된 필드들에 모두 은닉채널을 생성하는 것이 가능하지만 이렇게 은닉하여 전송할 수 있는 여러 공간들이 있을지라도, 옵션 필드처럼 사용 빈도가 적은 필드들보다는 반드시 전송과정에서 사용되는 필드들에 데이터를 포함시키게 되면 매번 통신과정에 이용되기 때문에 은닉채널을 이용한 공격의 성공률을 높일 수 있다. 또한 이러한 전송에 필수적으로 사용되는 필드들은 통신과정에서 다양한 값들을 가지기 때문에 탐지의 복잡성을 높일 수 있다. 패킷 필터링이나 재조립 등의 과정에서도 역시 옵션 필드들은 재조합 후 그 값이 변화되거나 패킷 필터링을 통해 데이터가 드러날 수 있다는 문제점을 가진 것에 비해, 전송에 필수적으로 사용되는 필드들은 패킷 재조합 과정에 의해 변화되지도 않을뿐더러, 패킷 필터링에 의해 노출될지라도 드러난 값은 통신과정에 쓰이는 비트들의 조합일 뿐 특별한 의미를 부여하지 않는다.

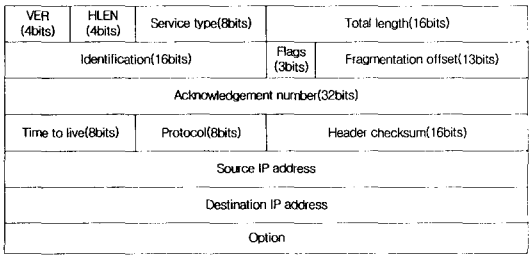

그림 2. IP 헤더 구조

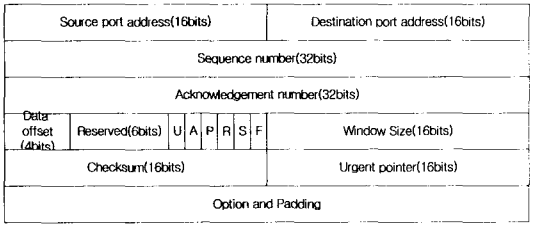

그림 3. TCP 헤더 구조

그러므로 본 논문에서는 TCP/IP 통신에 필수적으로 사용되는 ID, SEQ와 같은 필드의 은닉채널을 분석하여 TCP/IP 헤더를 이용한 보다 높은 수준의 은닉채널 탐지방안을 제안한다.

4.2 TCP/IP 헤더 은닉채널 분석

본 논문에서는 TCP/IP 헤더의 여러 필드 중에서도 IP 헤더의 ID 필드와 TCP 헤더의 SEQ 필드를 이용한 은닉채널을 탐지하는 방안에 대해서 연구하였다. 이러한 필드를 이용한 은닉채널 형성기법은 [1]에 나타나있으며, IP 헤더의 ID 필드의 경우 ID 필드를 임의의 숫자로 생성하여 특정 ASCII 값의 배가 되도록 인코딩하는 방법을 사용한다. 이러한 방법은 원격의 호스트에게 단순히 ID 필드를 읽는 것만으로 쉽게 특정 정보를 전달할 수 있게 해준다. 예를 들면, ID 필드의 경우 'H' 값을 은닉하기 위해서는 'H'의 ASCII 값인 72를 공격 이전에 일반적인 패킷에 포함되어 있는 ID 값들의 범주와 비슷한 숫자를 생성할 수 있도록 임의의 숫자 256배를 하여 계산한 값인 18432(72*256)를 ID 필드로 생성하는 것이다. 이후 이러한 패킷을 특정 포트로 수신하는 은닉채널 서버를 구성하고, 은닉채널 서버는 특정 포트로 수신된 패킷의 ID 필드 값을 256으로 나누어 원하는 데이터를 일반적인 TCP/IP 헤더의 ID 필드를 이용하여 얻을 수 있게 되는 것이다.

TCP 헤더의 SEQ 필드 역시 ID 필드의 은닉화와 같은 방법을 사용하나, SEQ 필드(32bits)에 올 수 있는 값의 범위는 ID 필드(16bits)의 값보다 훨씬 크기 때문에 'H'의 경우 72*256*65536의 값으로 ASCII 값을 생성한다. 이렇게 특정 값으로 ASCII 값을 인코딩하는 것은 앞서 설명한 바와 같이 일반적으로 TCP/IP를 이용한 통신과정에서 쓰이는 패킷들이 포함하는 ID 또는 SEQ 필드 값과 유사하게 만들기 위한 것이다.

하지만 이렇게 조작된 ID, SEQ 필드를 가지고 있는 TCP/IP 헤더는 정상적인 TCP/IP 패킷의 헤더와는 다음과 같은 차이점을 가진다. 먼저, 조작된 ID, SEQ 필드의 값이 정상적인 패킷의 ID, SEQ 필드 값이 가지는 값과는 비슷할지라도 많은 패킷들을 패턴화 시킬 경우 일반적인 패킷들과는 상이한 패턴을 가지게 된다. 또한 각 패킷은 TCP 연결시도와 같은 형태로 은닉 패킷을 전달하기 위하여 syn 플래그와 같이 특정 제어 플래그가 설정되어 있거나, IP fragment 필드 그리고 offset 설정 필드에 있어 일반적인 TCP/IP 패킷과는 다른 점을 가지게 된다. 비록 이러한 차이점이 있다고 할지라도 IDS의 시그너쳐 또는 패킷 트래픽의 관찰자의 직관만으로는 구분하는 것이 어려운 것이 현실이다.

따라서 본 논문에서는 TCP/IP 헤더의 ID, SEQ 필드를 이용한 은닉채널 탐지를 위해 다음 장에서 위조 패킷 자체의 특성 및 패킷간 시간 연관성을 이용한 탐지 방안을 제시한다. 또한 앞서 설명한 ID, SEQ 필드를 이용한 은닉채널을 형성해주는 도구로 [15]에 소개된 covert_tcp라는 도구를 사용하였다.

V. 제안하는 탐지 방안

본 장에서는 IP와 TCP 헤더 필드에 존재하는 은닉채널을 탐지하기 위해서 SVM 학습 방안을 제안한다. 제안하는 학습 방안은 SVM이 주어진 데이터 집합에 대하여 이진 분류기의 역할을 한다는 3장에서 설명된 기본 개념에 착안한 것으로서 정상 패킷들과 비정상 패킷(은닉 채널을 포함하는 패킷)들의 집합들을 구분하기 위해 여러 SVM 커널 함수들을 적용하여 최적의 분리 경계면을 찾아내는 것이다. 이러한 최적 분리 경계면을 찾는 것은 결국 TCP/IP 패킷들에 포함된 은닉채널을 탐지하는 것과 같은 결과를 가져오게 된다. 그러므로 앞장에서 분석된 IP와 TCP 헤더에 대한 은닉채널 생성 특성을 바탕으로 보다 명확하게 정상적인 TCP/IP 패킷들과 은닉채널을 포함한 TCP/IP 패킷들을 구분할 수 있는 SVM 학습 방안의 제안이 필요하다.

먼저 첫 번째로 생각해 볼 수 있는 학습 방안으로는 개별 패킷들에 대한 학습 방안이 있다. 이 학습 방안은 개별 TCP/IP 패킷들 하나하나에 대한 전처리 과정을 수행하고, 그렇게 전처리 된 데이터들에 대한 SVM 학습을 수행하는 것으로서 그림 4와 같다(SVM 학습을 통한 탐지방안 1). 그림 4와 같은 학습 방안은 단일 TCP/IP 패킷을 SVM의 입력 데이터로서 간주하는 것이다. 하지만, 이러한 방법은 수신되는 패킷간의 관련성을 고려하지 않고 단일 패킷만의 특성을 사용하여 학습되므로 탐지된 결과가 전처리 과정에서 사용하는 feature에만 밀접한 영향을 받을 것으로 예상된다. 따라서 이렇게 하나의 패킷만을 각각 사용하여 전처리를 통한 SVM 학습에 사용하는 것이 아닌, 패킷사이의 연관 관계를 고려하

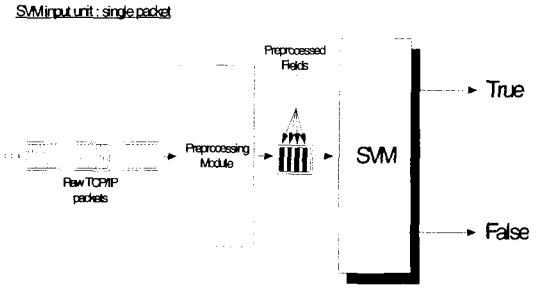

그림 4. SVM 학습을 통한 탐지방안 1

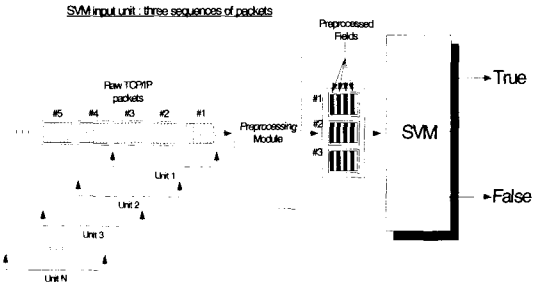

그림 5. SVM 학습을 통한 탐지 방안 2

는 탐지 방안을 제시한다. 제시하는 두 번째 탐지 방안에서는 여러 개의 TCP/IP 패킷을 특정 개수의 단위로 묶어 전처리 과정을 수행하며, 각 단위의 패킷들은 하나씩 슬라이딩되어 적용되므로 패킷간에 시간 지연이 고려된 학습방안이라고 할 수 있다. 여기서 패킷간 시간 지연을 고려한다는 것은 전후 연속한 패킷사이의 연관성이 SVM 학습에 미치는 결과를 알아보기 위하여 SVM 학습에 패킷 단위로 적용하는 것을 말한다. 이러한 패킷 슬라이딩을 이용한 학습 방안은 은닉채널을 포함하는 패킷의 은닉 데이터와 이러한 데이터를 전달하는 은닉 패킷의 전송 특성이 일반적인 통신과정에 쓰이는 TCP/IP 패킷과는 다르다는 IV. 2의 분석에 근거한 것이다. 즉, ID 필드와 SEQ 필드들을 사용하여 생성된 은닉채널을 가진 패킷들은 패킷 하나만이 의미를 가지지 못하며 또한 일반 패킷과 상이한 전송과정에 의해 패킷 전후간에 상호 연관성을 가지고 있다고 가정할 수 있다. 그러므로 연속하는 몇 개의 패킷들을 하나의 학습 입력 단위로 고려하여 SVM학습에 적용한다면 은닉채널 탐지에 보다 높은 성능을 가질 것이다. 이러한 패킷간 시간 연관성을 고려한 SVM 학습 방안은 그림 5와 같이 제안(SVM 학습을 통한 탐지방안 2)할 수 있다. 그림 5에서 제안하는 방안은 세 개의 TCP/IP 패킷을 하나의 입력 단위로 설정하고 전처리하여 SVM 학습에 사용한다. 또한 각 입력 단위의 패킷들은 매번 전처리 과정에 있어 하나씩 슬라이딩되어 세 개의 패킷중 맨 처음 수신된 패킷은 버려지고 네 번째 수신된 패킷을 다시 포함시켜 두 번째 입력 단위를 구성하게 된다.

Ⅵ. TCP/IP헤더 은닉채널 탐지실험 및 결과

5.1 TCP/IP 헤더 은닉채널 실험 환경

TCP/IP 헤더의 은닉채널 탐지 실험을 위하여 먼저 SVM을 학습하기 위한 학습 데이터 집합과 테스트 데이터 집합을 구성하였다. 이때 사용되는 정상 데이터의 경우에는 tcpdump를 사용하여 수집하며 비정상 데이터(은닉채널 패킷들)들은 covert_tcp 도구를 사용하여 수집하였다. 또한 은닉채널 탐지 실험에 있어 은닉채널을 이용한 공격의 경우를 IP 헤더의 ID 필드와 TCP 헤더의 SEQ 필드 두 가지 경우로 나누어 각각에 대하여 수행되었다. 즉, 학습 데이터 및 테스트 데이터 집합들은 ID 필드를 이용한 은닉채널 탐지와 SEQ 필드를 이용한 은닉채널 탐지 실험에 대해 별도로 구성되었다.

표 1. IP헤더의 ID 필드 features

Field	feature 개수	Feature Description
Identification	1	ID(16)
	3	ID(16) + Flag, Offset(16) + IP Header checksum(16)
	5	ID(16) + Flag, Offset(16) + IP Header checksum(16) + TCP control flag(16) + TCP Header Checksum(16)

표 2. TCP헤더의 SEQ 필드 features

Field	feature 개수	Feature Description
Sequence Number	2	Sequence(32)
	4	Sequence(32) + TCP control flag(16) + TCP Header Checksum(16)

먼저 학습 및 테스트에 사용되는 패킷들의 전처리 과정에 사용된 feature 값들은 다음의 표 1, 2와 같이 구성된다. 이때 차원 하나의 값은 패킷의 16비

트 값, 즉 16진수가 10진수로 변환하여 계산하였다. 즉, 헤더 필드의 16비트 값이 하나의 feature이며 SEQ 같은 경우 32비트로서 두 개의 feature에 해당한다. 표 1에는 IP 헤더의 ID 필드를 이용한 은닉채널 탐지를 위한 실험에 사용된 IP 헤더의 필드들을 추출된 feature의 수에 따라 1, 3, 5의 세 종류로 나누어 분류하였다. 또한 표 2에는 TCP 헤더의 Sequence Number 필드를 이용한 은닉채널 탐지를 위한 실험에 사용된 TCP 헤더의 필드들을 그 feature 수에 따라 2, 4의 두 가지로 구분하였다.

여기서 IP, TCP 헤더 패킷들을 표 1,2와 같이 값을 뽑아내어 전처리를 하는 과정에 있어서 하나의 패킷을 하나의 공격 단위로 보는 경우와 각 패킷간의 연관성을 고려하여, 패킷 수신 윈도우 사이즈를 3으로 하여 3개의 패킷을 하나의 공격 단위로 보는 경우로 나눈다. 그래서 다음의 그림 6과 같이 두 가지로 각 데이터 집합으로 구분하여 전처리 한다. 이때 세 개의 패킷으로 구성되는 하나의 공격 단위는 패킷 윈도우 사이즈가 3으로서 매번 하나씩 패킷이 슬라이딩되어 다음 공격 단위를 구성한다. 그러므로 각각의 학습 데이터 및 테스트 데이터 집합은 단일 패킷으로 구성되는 집합(Train Set 1, Test Set 1)과 패킷간 시간 연관성을 고려한 3개의 패킷 단위로 구성되는 집합(Train Set 2, Test Set 2)로 구성된다.

다음의 표 3에는 IP 헤더의 ID 필드를 이용한 은닉채널 탐지를 위한 SVM 학습 데이터 집합이 구성되어 있으며, 표 4에는 TCP 헤더의 SEQ 필드를 이용한 은닉채널 탐지를 위한 SVM 학습 데이터 집합이 나타나 있다. 학습 집합은 단일 패킷으로 구성되는 학습 집합 1과 연속된 세 개의 패킷이 하나의 공격 단위로 구성되는 학습 집합 2로 나뉜다. 학습 집합 1은 모두 10000개의 패킷으로 구성되며 각각 정상 패킷 5000개와 ID 또는 SEQ 필드가 조작된 5000개씩의 패킷으로 구성된다. 학습 집합 2는 연속된 패킷 3개가 하나의 단위를 이루며 패킷이 하나씩 슬라이딩되어 구성되는 총 10000단위의 패킷으로 구성되며 각각 정상 패킷 5000단위와 ID 또는 SEQ 필드가 조작된 5000단위의 패킷으로 구성된다. 표 5와 6의 SVM 테스트 데이터 집합도 동일한 방식으로 구성되며, 다만 전체 패킷의 수는 각각 1000개, 1000단위로 구성된다.

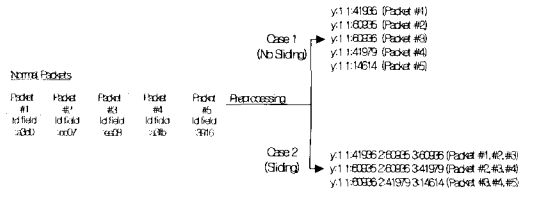

그림 6. 패킷 전처리 과정(no sliding/sliding)

본 실험에서는 [11]에서 개발된 mySVM 공개

표 3. ID 필드를 이용한 SVM 학습 데이터 집합

DataSet / 패킷종류	Training Set1 - No Sliding (10000개)	Training Set2 - Sliding (10000단위)
정상 패킷	개별 TCP/IP 패킷(5000개)	3개씩 연속된 TCP/IP 패킷(5000단위)
비정상 패킷	개별 IP 헤더의 ID 필드가 조작된 패킷(5000개)	IP 헤더의 ID 필드가 조작된 3개씩 연속된 패킷(5000단위)

표 4. SEQ 필드를 이용한 SVM 학습 데이터 집합

DataSet / 패킷종류	Training Set1 - No Sliding (10000개)	Training Set2 - Sliding (10000단위)
정상 패킷	개별 TCP/IP 패킷(5000개)	3개씩 연속된 TCP/IP 패킷(5000단위)
비정상 패킷	개별 TCP 헤더의 SEQ 필드가 조작된 패킷(5000개)	TCP 헤더의 SEQ 필드가 조작된 3개씩 연속된 패킷(5000단위)

표 5. ID 필드를 이용한 SVM 테스트 데이터 집합

DataSet / 패킷종류	Test Set1 - No Sliding (1000개)	Test Set2 - Slding (1000단위)
정상 패킷	개별 TCP/IP 패킷(500개)	3개씩 연속된 TCP/IP 패킷(500단위)
비정상 패킷	IP 헤더의 ID 필드가 조작된 개별 패킷(500개)	IP 헤더의 ID 필드가 조작된 3개씩 연속된 패킷(500단위)

표 6. SEQ 필드를 이용한 SVM 테스트 데이터집합

DataSet / 패킷종류	Test Set1 - No Sliding (1000개)	Test Set2 - Slding (1000단위)
정상 패킷	개별 TCP/IP 패킷(500개)	3개씩 연속된 TCP/IP 패킷(500단위)
비정상 패킷	TCP 헤더의 SEQ 필드가 조작된 개별 패킷(500개)	TCP 헤더의 SEQ 필드가 조작된 3개씩 연속된 패킷(500단위)

도구를 사용하여 실험하였으며 이때 SVM의 탐지 성능 비교를 위한 SVM의 커널로는 linear와 polynomial 두 개의 커널이 사용되었으며 polynomial 커널의 경우 Degree는 3으로 고정하여 실험하였다. 실험에서 사용된 mySVM 공개 도구는 SVM 분류나 회귀 예측 알고리즘을 구현한 검증된 도구 중의 하나로서 독일 Dortmund 대학에서 개발하였다. 현재 SVM을 이용한 응용 연구에 가장 널리 사용되는 도구로서 알려져있다.

5.2 실험 결과

SVM을 이용한 TCP/IP 헤더 은닉채널 탐지 실험은 5장에서의 제안방안처럼 패킷간의 시간 연관성을 고려한 두 개의 학습 셋으로 먼저 분류 한 후 각 학습 셋에 대하여 SVM에 적용한 커널과 feature 수 변화에 따른 탐지결과를 분석하였다.

표 7은 ID 필드와 SEQ 필드를 이용한 은닉채널 탐지 실험에 대한 전체 실험 결과이며 그림 7, 8에서는 ID, SEQ 필드 각각을 이용한 은닉채널 탐지에 대한 실험결과를 보여준다. 실험에 대한 전체 결과는 표 7에 나타나 있으며 표 8, 9을 통하여 ID, SEQ 넘버 필드 각각에 대한 학습 셋, 커널의 종류, features 수에 따른 실험 결과를 알 수 있다. 학습 패턴에 따른 결과를 분석하면 단일 패킷보다는 시간 연관성을 가지고 패킷을 슬라이딩하는 경우에 더 우수한 탐지 효율을 보였다. 또한 일반적으로 feature의 수가 많아질수록 탐지율이 높았으며, 본 논문의 실험결과표에 나타내지 않았지만, features가 ID의 경우 5, SEQ 필드의 경우는 4를 넘어서는 경우에는 거의 99%의 정확성을 가지고 패킷을 분류하였다. 그리고 kernel의 경우 degree값 3을 가지는 polynomial 커널의 경우에 linear에 비해 우수한 탐지 성능을 보였다.

표 7. TCP/IP 헤더 은닉채널 탐지 결과

field	Kernel	Features	Test Set1 (No Sliding)			Test Set2 (Sliding)		
			FP	FN	TC	FP	FN	TC
ID	Linear	1	31.5	7.0	61.5	31.4	14.6	54.0
		3	9.0	31.7	67.4	0.2	14.2	85.6
		5	0.4	0.5	99.1	0.0	0.1	99.9
	Poly-nomial	1	16.9	43.0	40.1	3.0	9.2	87.8
		3	29.3	0.0	70.7	1.2	5.2	93.6
		5	0.5	0.0	99.5	0.1	0.1	99.8
SEQ	Linear	2	1.0	33.2	65.8	11.1	1.0	87.9
		4	0.0	1.3	98.7	9.0	0.0	99.1
	Poly-nomial	2	2.5	28.0	69.5	5.0	7.5	92.0
		4	0.0	0.1	99.9	0.0	1.0	99.9

* Polynomial Kernel의 Degree = 3, FP: False Positive(%), FN: Fasle Negative(%), TC: Total Correctness(%)

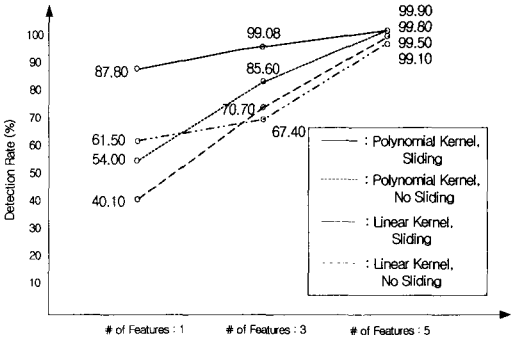

그림 7. IP 헤더의 ID필드 결과 그래프

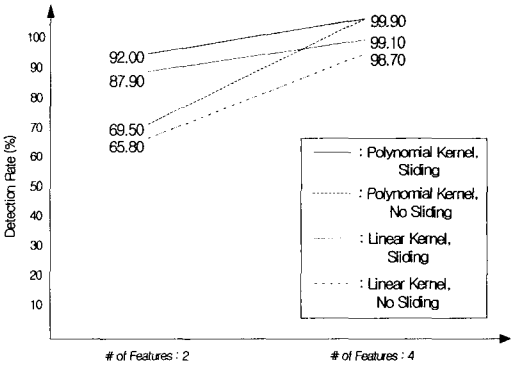

그림 8. TCP 헤더의 SEQ필드 결과 그래프

표 8. IP 헤더의 ID 필드 실험 결과(%)

	ID field Covert Channel						
인자	TS1	TS2	KR1	KR2	F1	F3	F5
탐지율	67.68	86.78	77.92	81.92	60.85	79.33	99.56

* TS1: Test Set1(No Sliding), TS2: Test Set2(Sliding), KR1: Linear, KR2: Polynomial(Degree=3), F1: 1 Feature, F3: 3 Features, F5: 5 Features

표 9. TCP 헤더의 SEQ 필드 실험 결과(%)

	SEQ field Covert Channel					
인자	TS1	TS2	KR1	KR2	F2	F4
탐지율	73.05	94.73	87.88	90.33	78.80	99.40

* TS1: Test Set1(No Sliding), TS2: Test Set2(Sliding), KR1: Linear, KR2: Polynomial(Degree=3), F2: 2 Features, F4: 4 Features

Ⅷ. 결론 및 향후 연구 방향

본 논문에서는 TCP/IP 헤더에 존재할 수 있는 많은 은닉채널 가운데에서도 전송 과정에 필수적으로 사용되는 IP 헤더의 ID 필드와 TCP 헤더의 SEQ 필드를 이용한 은닉채널 기법을 분석하였다. 분석 결과 각 필드에 대한 ASCII 인코딩 기법이 사용되고 있음을 알 수 있었고, 이러한 기법으로 생성된 은닉채널은 일반 통신 채널과 매우 흡사하다는 결과를 얻었다. 따라서 본 논문은 패턴분류에 있어 우수한 성능을 보이는 SVM을 사용하여 IP헤더의 ID 필드, TCP 헤더의 SEQ 넘버 필드를 이용한 은닉채널 학습기법을 제안하였다. 또한 SVM 학습에 있어서는 단순히 패킷간 시간적 연관성을 무시한 단순 패킷 학습방안과 패킷간 시간 요소를 고려한 패킷 슬라이딩을 통한 학습방안을 제시하고 실험을 수행하였다. 앞서 설명된 실험 결과에서처럼 SVM을 통한 두 가지 학습의 결과 90%이상의 높은 탐지 정확성을 가짐을 알 수 있었다.

비록 본 논문에서는 특정 도구에 의해서 생성된 은닉채널 패킷을 사용하였지만, 이것은 단지 탐지 실험을 위한 은닉채널 생성의 한 예일 뿐이며 본문에서 언급하였다시피 SVM은 두 개의 클래스로 입력된 데이터 집합을 구분하기 위한 최적화 분리면을 찾아낸다는 것이 이미 증명된 SVM이 가진 특징이다. 따라서 기존의 학습 기법들에서 나타났던 특정도구에 의존적인 학습에 의한 실험 오류는 방지할 수 있다. 하지만, 추후 연구를 통해 특정 도구로 생성된 비정상 데이터와 정상데이터 사이의 학습 중요도를 비교할 수 있는 One-Class SVM 등과 같은 알고리즘을 적용한 실험과 SVM 학습을 통한 분류에 있어 정상 패킷들에 대한 분류가 은닉 채널 패킷 외에 비정상 패킷들에 의해서도 일어날 수 있으므로 결과 분석을 통하여 은닉채널만을 정확하게 분리할 수 있는 실험 방안을 찾는 것도 필요하다. 마지막으로 TCP /IP 트래픽에 생성될 수 있는 보다 많은 은닉채널 기법에 대한 분석 및 기계 학습 분야의 SVM 이외에 신경망과 같은 기법에 대한 비교 검증 실험과 학습 셋과 테스트 셋을 좀 더 확장하고, 성능 향상 변수로 사용될 수 있는 커널의 종류와 제약 조건 인자에 대한 고려 역시 보다 정확한 실험 결과의 도출을 위해 필요할 것으로 사료된다.

참 고 문 헌

[1] C. H. Rowland, Covert channels in the TCP/IP protocol suite, Tech. Rep. 5, First Monday, *Peer Reviewed Journal on the Internet*, July 1997.

[2] John Giffin, Covert Messaging Through TCP Timestamps, *PET2002*, pp.194-208, Apr 2002.

[3] Daemon9, "Loki: ICMP Tunneling", *Pharack Magazine*, Vol.6, Issue 49, article 6 of 16

[4] Department of defence trusted computer system evaluation criteria, *Tech. Rep. DOD 5200.28-ST, Department of Defence*, Dec 1985. Supersedes CSC-STD -001-83.

[5] Vapnik V., The Nature of Statistical Learning Theory, *Springer-Verlag*, New -York, 1995.

[6] C. Campbell and N. Cristianini, Simple Learning Algorithms for Training Support Vector Machines, *Technical report, University of Bristol*, 1998

[7] B. W. Lampson, A note on the confinment problem, *in Proc. of the Communications of the ACM*, no. 16:10, pp. 613-615, Oct 1973.

[8] K. Ahsan and D. Kundur, "Practical Data Hiding in TCP/IP," *Proc. Workshop on Multimedia Security at ACM Multimedia '02*, 7 pages, French Riviera, Dec 2002.

[9] J. McHugh, Covert Channel Analysis, Technical Memorundum 5540:080A, Naval Research Laboratory, Washington D.C., 1995. *A Chapter of the Handbook for the Computer Security Certification of Trusted Systems.*

[10] Neil F. Johnson et al, Information Hiding: Steganography and Watermarking - Attacks and Countermeasures, *Kluwer Academic Publishers*, 2000.

[11] Fabien A. P. Petitcolas, editor, Infor-

mation hiding. *Proceedings of the 5th international workshop on information hiding*, LNCS 2578, Noordwijkerhout, The Netherlands, Oct 2002.

[12] S. Mukkamala, G. Janowski, A. H. Sung, Intrusion Detection Using Neural Networks and Support Vector Machines, *Proceedings of IEEE IJCNN*, pp.1702-1707, May 2002.

[13] Dorothy E Denning, An Intrusion Detection Model, *In IEEE Transactions on SE*, Number 2, pp. 222-250, Feb. 1997

[14] Pontil, M. and Verri, A., "Properties of Support Vect or Machines," A.I. Memo No. 1612; CBCL paper No. 152, Massachusetts Institute of Technology, Cambridge, 1997.

[15] Joachmims T, mySVM - a Support Vector Machine, *Univerity Dortmund*, 1999.

〈著者紹介〉

손 태 식 (Tae-Shik Shon)
2000년 2월: 아주대학교 정보 및 컴퓨터 공학부(학사)
2002년 2월: 아주대학교 정보통신전문대학원 정보통신공학과(석사)
2004년 2월: 고려대학교 정보보호대학원 정보보호학과(박사 수료)
2003년 9월~12월: 서경대학교 정보통신공학과 강사
2003년 5월~12월: 한국정보보호교육센터 강사, ICU 부설 정보통신교육원 강사
2002년 8월~현재: 고려대학교 정보보호기술연구센터 연구원
2004년 2월~현재: Dept. of Computer Science, University of Minnesota 객원 연구원
〈관심분야〉 네트워크 보안, 패턴인식, 신경망, 리눅스 보안

서 정 우 (Jung-Woo Seo)
2002년 2월: 호남대학교 정보통신 공학부 졸업 (공학사)
2004년 2월~현재: 고려대학교 정보보호대학원 졸업(공학석사)
2004년~현재: 삼성전자
〈관심분야〉 시스템/네트워크 보안, 생체인식

문 종 섭 (Jong-sub Moon)
1981년 2월: 서울대학교 계산통계학과 학사
1983년 2월: 서울대학교 계산통계학과 석사
1992년 2월: Illinois Institute of Technology 박사
1993년~현재: 고려대학교 전자 및 정보공학부 교수, 고려대학교 정보보호대학원 겸임 교수
〈관심분야〉 IDS, 신경망, 생체인식, 운영체제

서 정 택 (Jung-Taek Seo)
1999년 2월: 충주대학교 컴퓨터공학과 졸업(공학사)
2001년 2월: 아주대학교 대학원 컴퓨터공학과 졸업(공학석사)
2000년 11월~현재: ETRI 부설 국가보안기술연구소 선임연구원
〈관심분야〉 정보전, 시스템/네트워크 보안, 취약점 분석·평가

최 홍 민 (Hong-Min Choi)
2000년 2월: 아주대학교 정보 및 컴퓨터 공학부 졸업(공학사)
2002년 2월: 아주대학교 정보통신공학과 졸업(공학석사)
2001년 7월~현재: (주) Secuve 연구원
〈관심분야〉 침입탐지시스템, 리눅스 보안