탐지성능을 향상 위한 해시기반 패턴 매칭 시스템1)

김병훈*, 이동휘*, 김귀남*
*경기대학교 일반대학원 산업보안학과
e-mail:hackerking@kgu.ac.kr

Hash-based pattern matching system for detection performance

Byung-Hoon Kim*, Dong-Hwi Lee*, Kui-Nam Kim*
*Dept of Industrial Security System, Kyoggi-University

요 약

네트워크 대역폭과 침입 기술의 발달하는 상황에서 침입탐지 시스템의 패턴 매칭 방식으로는 대용량화된 패킷을 스니핑되는 환경에서 기존의 패턴 매치 방식은 서비스에 따른 순차적인 패턴 매칭 방법으로는 모든 패킷들의 패턴을 분석하는 것에는 한계가 발생하며, 중요한 패킷을 패턴 매칭을 통하여 분석하지 못하여 여러 취약점에 대한 공격이 발생할 수 있는 악성 봇, 정보의 유출이 발생할 수 있다. 본 연구에서는 패턴 매칭 기법을 이용하는 침입탐지 시스템의 패턴 매칭 성능 향상을 위하여 서비스별로의 순차적인 패턴 매칭 방법에서 해시 기반의 패턴 매칭을 이용하여 패턴 매칭의 시간을 단축 시켜 빠른 속도로 탐지 할 수 있는 시스템을 제시하고 검증하고자 한다.

1. 서론

최근 네트워크의 고속화에 따른 고용량화 서비스의 다양화로 인하여 매일 수많은 취약점이 발생하며 취약성을 이용한 악의적인 공격이 하루에도 수많이 발생하고 있다. 네트워크에 유입되는 많은 패킷들을 분석하여 침입, 악의적인 패킷의 유형을 인식하여 관리자에게 빠른 시간에 알리기 위하여 침입탐지시스템(Intrusion Detection System)이 반드시 필요하다. 침입탐지시스템은 탐지방법에 따라 크게 오용탐지(misuse detection) 방법과 이상탐지(anomaly detection) 방법으로 나눌 수 있다. 오용탐지방법에서는 침입, 악의적인 패킷의 유형을 가지는 패킷들의 패턴을 정의하여 패킷이 유입되었을 때 유입된 패킷의 패턴을 분류하여 정의되어 있는 악성 패킷의 패턴과의 패턴 매치 과정을 통하여 유입되는 패킷의 유형을 파악한다. 패턴의 탐지 시간에 따라서 침입탐지시스템의 성능이 좌우되며 따라서 패턴 탐지 방법을 개선하고자 하는 많은 연구들이 있어왔다. 연구배경에서는 대표적인 오픈 소스 침입탐지시스템으로 snort에서 사용되는 룰과 패턴의 처리 방식과 소요되는 시간의 관점에서 다루며, 관련연구에서는 침입탐지시스템에서 사용하는 패턴 매치 방법으로 Boyer-Moore 방법과 Aho-Corassic BM 방법을 서술한다. 패턴 탐지에 소모되는 시간을 줄이기 위한 방법으로 해시를 이용한 패턴 탐지방법을 제안한다.

1) 본 연구는 지식경제부 지역혁신센터사업인 산업기술보호특화센터 지원으로 수행되었음.

2. 연구배경

snort의 규칙은 다음과 같이 5개의 사슬 집합으로 나누어진다.

Activation->Dynamic->Alert->Pass->Log

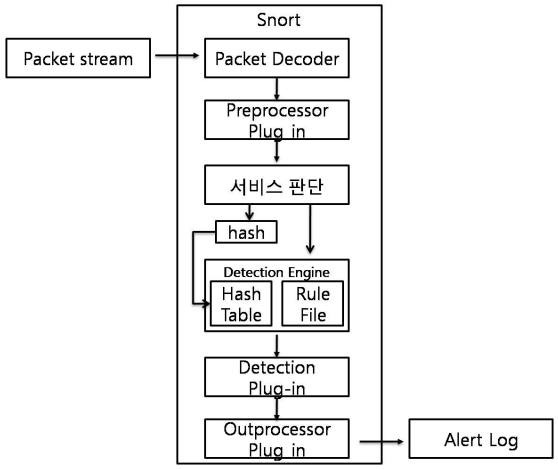
Activation은 경고를 발생시키고 또 다른 Dynamic 규칙을 활성화, Alert는 경고를 발생시키고 패킷 로그를 남긴다. Pass는 패킷을 무시하며 Log는 로그를 저장하지만 경고를 발생시키지 않는다. 5개의 규칙 사슬은 각각에 대한 프로토콜별 연결리스트가 존재하며 snort에서는 각각의 사슬에 종속되어 있는 프로토콜별의 룰 파일에 라인으로 저장되어 있는 패턴을 탐지하게 된다. snort rule 2.8에서 입수한 것으로 총 10404의 패턴이 존재하며 이 패턴들은 각각의 프로토콜별로 나눠져 저장되어 있다. 패턴 매치 과정에서 각각의 프로토콜의 서비스에 따라서 48개의 룰로 분류되어 있으며 실제 패턴 매치 과정에서 각각의 룰 파일에 있는 패턴들을 탐지하게 된다. 패턴 탐지시 라인단위로 패턴 매치를 수행하며 룰 파일 중 가장 많은 패턴을 가지고 있는 것은 630가지의 Alert 패턴을 가지고 있다. 630라인을 순차적으로 패턴 매칭 하여 마지막 라인의 패턴과 일치하는 경우 경고를 발생하는 데까지의 시간과 첫 번째 라인의 매치되는 것의 시간 차이가 많이 발생하여 탐지 성능의 지연을 느낄 수 있다. 따라서 630라인을 모두 탐색하는 것보다는 탐지성능 개선을 위하여 새로운 탐지 기법이 필요하다.

3. 관련연구

기존 snort 에서 탐지하는 방법으로는 Boyer-Moore 알고리즘을 이용한 패턴매칭과 AC-BM(Aho-corasic Boyer-Moore) 알고리즘을 이용한 패턴매치를 이용하여 탐지해왔다. Boyer-Moore 알고리즘의 기본 방식은 문자를 뒤에서부터 비교 하여 맨 앞의 스트링 문자까지 일치하게 되면, 검색이 성공했음을 알린다. 불일치 문자 발생 시에 기본적으로 '불일치 문자 방책(Bacd Character Shift)' 이라는 방식을 쓰게 되는데, 이에 단점이 있어 일치 '접미부(Good Suffix Shift) 방식' 과 혼용하여 성능을 향상시켰다.[2] snort 2.0 이하에서 사용되던 Boyer-Moore 알고리즘의 구조를 변경하여 패킷 매칭 속도를 크게 향상시킨 방법이다. 이 방법은 패킷 내용 검사를 비교해야 할 키워드 별로 매번 진행해야 하는 번거로움이 있는 점을 개선한 방법이다 '불일치 문자 방책' 이란 불일치 문자 발생 시 불일치가 일어난 곳의 텍스트 문자를 스트링에서 검색하여 가장 최근에 발생한 위치의 다음번의 스트링 문자부터 텍스트와 검색하는 방식으로 Boyer-Moore 알고리즘에서 기본적으로 사용되는 정책이다.[3]

4. 해시기반 패턴 매치방법

패킷이 수신되어 디코더와 전처리기에서 처리된 후 예외 처리된 패킷들은 해시함수를 통하여 해시화 한다. 패킷의 해시값을 기존의 alert 탐지 패턴의 해시 테이블과 비교한다. 인덱스화된 해시테이블을 이용하여 패턴의 매치를 작업을 수행할 수 있다. 아래 그림은 해시기반 패킷 매치 과정이다.



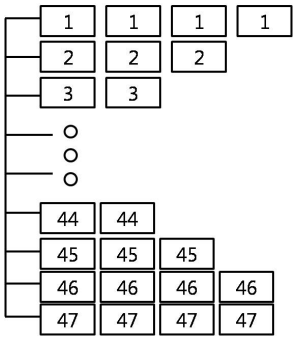
alert 패턴 중 패턴의 라인의 수가 적은 룰 파일은 해시함수를 이용하지 않고 바로 룰 파일에서 패턴을 탐지 한다. 아래 표는 snort 가 가지고 있는 alert 패턴 룰 파일중 48개의 패턴라인을 가지고 있는 룰 파일이다.

netbios.rules 630	oracle.rules 305	web-cgi.rules 358	web-misc.rules 123
deleted.rules 243	rpc.rules 128	web-iis.rules 119	exploit.rules 106
web-php.rules 127	smtp.rules 75	ftp.rules 76	misc.rules 68
web-client.rules 49	backdoor.rules 78	icmp-info.rule 93	sql.rules 48

위 표와 같은 서비스로 룰 파일에 기록된 라인들을 해시함수를 통하여 인덱스화 한다. 문자열을 포함하는 해시 함수 적용을 위하여 아래와 같은 해시 함수를 적용한다.

$$(X_1)*C^{\wedge n} + (X_2)*c^{\wedge(n-1)} + (X_3)*c^{\wedge(n-2)} + \ldots + (X_{n-1})*C + X_n$$

문자열 해시는 충돌 방지를 위하여 모든 문자열의 해시를 하도록 하며 각 문자의 아스키 코드를 더해서 해시를 생성 할 수 있지만, 문자열의 순서가 바뀌는 경우에도 충돌이 발생 할 수가 있다. 따라서 각각의 문자의 아스키코드에 상수의 거듭제곱을 곱한 후 모두 더하는 것으로 순서 변화에 따른 충돌을 막을 수 있다. 상수를 C하고 문자를 X_1 ~ X_n이라 하면 위와 같은 식으로 나타낼 수 있다. 해시테이블을 위하여 식에서 산출된 값을 인덱스로 생성할 값으로 나누어 나머지를 구하여 나머지 값을 인덱스로 사용한다. 나머지는 1부터 47까지 이며 나머지의 수는 인덱스의 수와 동일하다. 해시함수를 이용하여 인덱스 화된 테이블은 아래 그림과 같다.



5. 결론

하나의 인덱스 안에는 여러개의 패킷을 해시화 한 나머지 값을 가지고 있다. 라인단위로 매치를 하여 탐지를 수행 했을 때 최대 630라인까지 패턴 매치를 수행하여 탐지를 하는 경우 발생하는 지연을 탐지 패턴의 해시화를 이용하여 47개의 인덱스를 가진 해시기반의 탐지엔진에서는 최대 13개의 패턴의 검색을 통하여 유입되는 패킷의 패턴을 탐지 할 수 있다.

6. 참고문헌

[1] 김대영, 김선일, 이준용 "시그내처 기반의 네트워크 침입 방지에서 고속의 패킷 필터링을 위한 시스템 구조" ,정보과학회논문지, 2007,2월

[2] COLE, R., 1994, Tight bounds on the complexity of the Boyer-Moore pattern matching algorithm, SIAM Journal on Computing 23(5):1075-1091

[3] C. Coit, S. Staniford, J. McAlerney, "To wards Faster String Matching for Intrusion Detection, " Proceedings of DARPA Information Survivability Conference and Exposition(DISEC), Anaheim, California, June 12 14, 2001.