고성능 침입방지 시스템을 위해 개선한 시그니처 해싱 기반 패턴 매칭 기법

이영실* · 김낙현** · 이훈재***

*동서대학교 디자인&IT 전문대학원, **동서대학교 일반대학원

***동서대학교 컴퓨터정보공학부

An Improved Signature Hashing-based Pattern Matching for High Performance IPS

Young Sil Lee* · Nack Hyun Kim · Hoon Jae Lee**

*Graduate School of Design and IT, **Graduate School of General, Dongseo University

***Div. Computer and Information Engineering, Dongseo University

E-mail : youngsil.lee0113, nackhyunkim@gmail.com, hjlee@dongseo.ac.kr

요 약

시그니처 기반 필터링(Signature based filtering)은 이미 알려진 공격으로부터 방어하는 방법으로, 침입방지 시스템을 통과하는 패킷의 페이로드와 시그니처라 불리는 공격 패턴들과 비교하여 같으면 그 패킷을 폐기한다. 그러나 시그니처의 개수가 증가함에 따라 하나의 들어온 패킷에 대하여 요구되는 패턴 매칭 시간은 증가하게 되어 패킷의 지연현상이 발생한다. 고성능 침입방지 시스템을 위해서는 보다 효율적인 패턴 매칭 알고리즘이 필요하며, 패턴 매칭의 수행 성능 향상을 위해 가장 중요한 부분은 처리해야 하는 패킷이 도착했을 때, 해당 패킷의 데이터를 룰의 시그니처와 비교하는 횟수를 줄이는데 있다. 이에 본 논문에서는 고성능 침입방지 시스템의 개발을 위해 기존의 제안된 시그니처 해싱 기반의 침입방지 시스템에 패킷 분류를 위한 다차원 검색을 튜플 공간이라는 이차원 공간을 이용하여 검색하는 튜플 공간 패킷 분류 알고리즘과 블룸 필터를 적용한 패턴 매칭 방법을 제안한다.

ABSTRACT

NIPS(Network Intrusion Prevention System) is in line at the end of the external and internal networks which performed two kinds of action: Signature-based filtering and anomaly detection and prevention-based on self-learning. Among them, a signature-based filtering is well known to defend against attacks. By using signature-based filtering, intrusion prevention system passing a payload of packets is compared with attack patterns which are signature. If match, the packet is discard. However, when there is packet delay, it will increase the required pattern matching time as the number of signature is increasing whenever there is delay occur. Therefore, to ensure the performance of IPS, we needed more efficient pattern matching algorithm for high-performance ISP. To improve the performance of pattern matching the most important part is to reduce the number of comparisons signature rules and the packet whenever the packets arrive. In this paper, we propose an improve signature hashing-based pattern matching method. We use tuple pruning algorithm with Bloom filters, which effectively remove unnecessary tuples. Unlike other existing signature hashing-based IPS, our proposed method to improve the performance of IPS.

키워드

Bloom filter, Signature hashing, Pattern matching, Tuple pruning, NIPS

Ⅰ. 서 론

인터넷의 발전과 더불어 네트워크상에서의 침입 시도가 갈수록 증가되고 다변화됨으로써, 이에 대한 대응으로 많은 침입탐지 시스템들이 개발되었다. 그러나 현재의 대다수 침입탐지 시스템들은 갈수록 증가하는 트래픽양을 처리하는 데 어려움이 있다. 네트워크 트래픽은 이미 수기가비트 급으로 확장되었고, 이러한 고속 대용량화된 대규모 네트워크 환경에서 다양한 침입을 보다 빠르고 정확하게 탐지하고 대응하기 위한 보안 방법이 필요하다[1].

침입방지 시스템(IPS: Intrusion Prevention System)은 인라인모드(in-line mode)로 네트워크에 설치되어, 네트워크를 지나는 패킷 또는 세션을 검사하여 만일 그 패킷에서 공격이 감지되면 해당 패킷을 폐기하거나 세션을 종료시킴으로써 외부의 침입으로부터 네트워크를 보호하는 시스템을 의미한다.

침입방지 시스템은 크게 두 가지 종류의 동작을 수행한다. 하나는 이미 알려진 공격으로부터 방어하는 시그니처 기반 필터링(Signature based filtering)이고 다른 하나는 알려지지 않은 공격이나 비정상 세션으로부터 방어하는 자기 학습 기반의 변칙 탐지 및 방지(anomaly detection and prevention based on self-learning)이다[2].

시그니처 기반 필터링에서는 네트워크를 통과하는 패킷의 페이로드와 시그니처(signature)라고 불리는 공격 패턴을 표현한 룰들과 비교 검사하여 해당 패킷의 감염 및 이상 여부를 판단하고, 패킷을 폐기하거나 세션을 종료함으로써 네트워크를 공격으로부터 보호한다. 하지만 시그니처의 개수가 증가함에 따라 하나의 들어온 패킷에 대하여 요구되는 패턴 매칭 시간은 증가하게 되어 패킷지연 없이 동작하는 고성능 침입탐지 시스템을 개발하는 것이 어렵다.

본 논문에서는 2007년 Wang., Kwon., Jung., Kwak., Chung.,에 의해 제안[3]된 시그니처 해싱 기반의 침입방지 시스템의 방식에 2010년 Kim., Lim.,에 의해서 제안[4]된 패킷 분류를 위한 다차원 검색을 튜플 공간이라는 이차원 공간을 이용하여 검색하는 튜플 공간 패킷 분류 알고리즘과 블룸 필터를 적용한 튜플 블름 패턴 매칭 알고리즘을 사용한 기법을 제안한다.

Ⅱ. 개선한 패턴 매칭 알고리즘

패턴 매칭의 수행 성능을 향상시키기 위해서 필요한 가장 중요한 부분은 처리해야 하는 패킷이 도착했을 때, 해당 패킷의 데이터를 룰의 시그니처와 비교하는 횟수를 줄이는데 있다. 즉, 비교해야 하는 룰의 개수를 줄이거나, 비교해야 하는 경우의 수를 줄임으로써 매칭 수행 성능을 향상시킬 수 있다. 하지만 단순히 분류를 잘해서 매칭을 수행해야 하는 룰의 수를 줄이는 것은 룰의 개수가 점점 많아지고 있는 현실을 감안하면 그다지 효과적이라고 할 수 없다. 이를 보장하기 위해서는 항상 예측 가능하고, 성능이 보장되는 새로운 알고리즘이 필요하며, 시그니처 간의 상관관계에서도 자유로워야 한다.

시그니처 해싱이란 적용되는 모든 룰의 내부에 존재하는 시그니처에 대해 이 중 일부(2 byte)를 정해 해싱 값을 만들고, 이 해싱 값을 이용하여 검색하고자 하는 패킷 데이터와의 비교를 통해 실제 확인해볼 필요가 있는 시그니처만을 비교하도록 경우의 수를 줄이는 방식으로, 이를 통해 매칭에 사용될 불필요한 동작을 획기적으로 줄임으로써 매칭 속도를 향상시키고, 항상 안정된 성능을 낼 수 있도록 하기 위한 알고리즘이다.

제안된 방식은 크게 패턴 매칭을 수행하기 전에 적용할 룰을 정하여 해당 룰을 해싱 기법을 이용하여 포트 및 시그니처 해시 테이블로 정리하는 준비 단계와 실제 패킷이 들어왔을 때 이 패킷 데이터를 조사하여 시그니처가 정확히 존재하는지를 검사하는 부분으로 나뉜다.

1. 포트/ 시그니처 해시 테이블

패턴 매칭 수행에 앞서 적용할 룰에 대해 포트 및 시그니처를 기준으로 정리하는 단계를 거치게 되는데, 이를 위해 룰에서 해당되는 서비스 포트와 시그니처 중 일부(2 byte)를 추출하여 포트 해시 테이블과 시그니처 해시 테이블을 구성하는 과정이 필요하다.

이를 위해 먼저 각 프로토콜과 서비스 포트 번호 등에 대해 분류를 하고, 해당 포트를 해싱한 값을 이용하여 포트 해시 테이블을 구성한다. 포트 해시 테이블을 구성하면, 해당하는 포트 해시 테이블이 갖고 있는 시그니처 해시 테이블에 시그니처의 해싱 값을 등록한다. 여기서 작성한 시그니처 해시 테이블에 실제 매칭에 사용되는 룰의 정보(룰 넘버, 근원지 프리픽스, 목적지 프리픽스, 근원지 프리픽스 길이, 목적지 프리픽스 길이, 프로토타입 등)를 연결 리스트를 이용하여 등록한다.

그림 1은 룰에서 포트 해시 테이블과 시그니처 해시 테이블을 구성하는 예와 각 해시 테이블에 어떤 식으로 룰들이 맵핑되는지를 나타낸다.

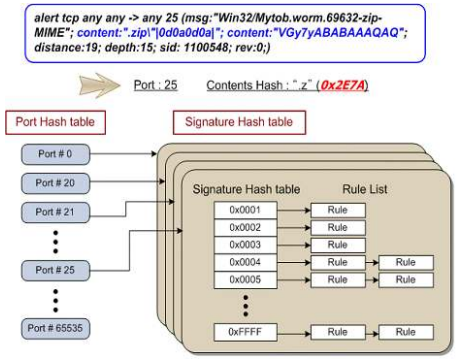

그림 1. 포트 및 시그니처 해시 테이블

2. 패턴 매칭 진행

실제 패킷 데이터에 대해 매칭을 수행함에 있어 Byte windowing과 Rule matching의 두 가지 절차를 거친다.

2.1. Phase 1 - Byte windowing

byte windowing을 통해 패킷의 데이터 부분을 windowing하며 검사하는 과정이다. 이 경우 시그니처 해싱 값에 사용되었던 기준치(2 byte)만큼씩 windowing하며 해싱 값을 구한다. 여기서 구해진 해싱 값을 가지고 이미 작성되어 있는 시그니처 해시 테이블에 같은 값이 있는지를 비교하고, 같은 값이 없다면 해당 부분은 이상이 없다고 판단, 계속 windowing을 진행한다.

만약 시그니처 해시 테이블과 비교하여 같은 값이 존재하면 해당 두 번째 Rule matching 단계를 적용하여 최종적으로 해당 패킷이 실제 룰과 일치하는지를 정밀 검사한다.

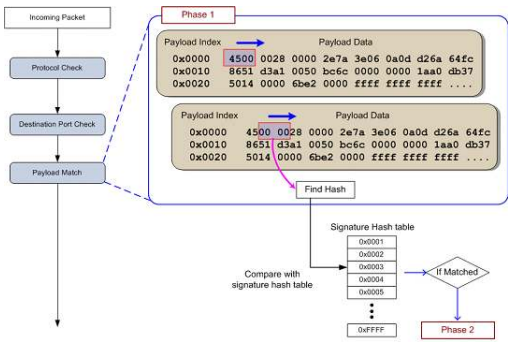

그림 2. Phase 1 – Byte windowing

2.2. Phase 2 - Rule matching

byte windowing하며 얻은 해싱 값과 시그니처 해시 테이블에 존재하는 해싱 값이 같을 경우 해당 패킷이 룰과 일치하는지를 의심하고 확인해야 한다. 이는 오탐을 막기 위해 중요한 행동이며, 룰에 기술된 다른 정보와 함께 비교하여 정확히 일치하는지를 확인하는 과정을 거쳐 최종적으로 패킷이 룰에 적용되어 적절한 조치를 취해도 되는지를 확인하는 단계이다.

2.2.1. filtering

시그니처 해시 테이블의 해당 해싱 값에 연결된 룰에 대한 연결 리스트를 확인한다. 시그니처의 전체 값을 확인하기에 앞서 시그니처의 근원지 프리픽스, 목적지 프리픽스와 패킷의 해당 데이터를 비교하여 최우선 순위의 룰을 찾아내는 과정이다.

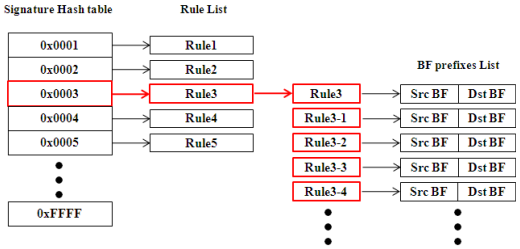

그림 3. 해시 테이블 및 Rule List 연결 구조

통합 블룸 필터를 사용하여 입력된 패킷에 대하여 근원지 주소필드와 목적지 주소 필드의 정보를 각각의 블룸 필터를 병렬적으로 통과시켜 블룸 필터에서 일치 가능성이 있는 양성 값의 길이 정보를 얻어낸 후 해시 테이블 검색을 수행하여 참-양성 길이 정보만을 얻어낸다. 각 필드의 참-양성 값을 이용하여 크로스 프로덕팅 단계를 거친 후 남은 튜플을 이용하여 해시 테이블 검색을 수행하여 최우선 순위의 룰을 찾아낸다.

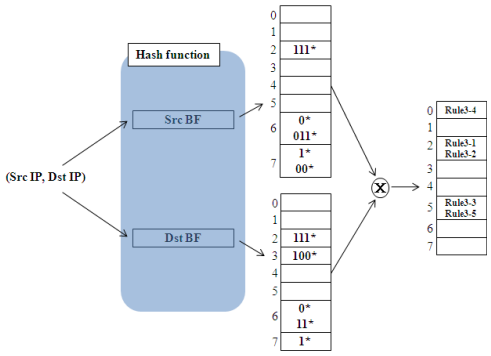

그림 4. 튜플 블름 필터

2.2.2. final matching

앞의 filtering 단계에서 찾은 최우선 순위의 룰을 시작으로 최종적으로 패킷이 룰에 정확히 일치하는지를 확인하는 과정이다.

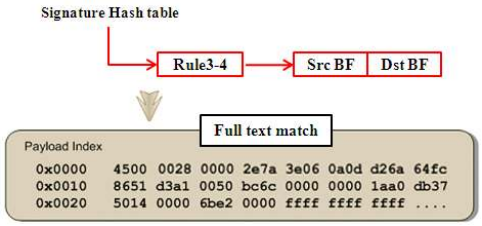

그림 5. final matching

III. 성능 및 효율성 분석

본 논문에서는 기존의 시그니처 해싱 알고리즘에 튜플 공간 패킷 분류 알고리즘과 블름 필터를 적용한 튜플 블름 알고리즘을 사용하여 개선한

패턴 매칭 기법이다.

기존의 시그니처 해싱 기법은 기존 패턴 매칭 알고리즘에 비해 룰의 개수가 많아지더라도 전체 네트워크 성능에 큰 차이 없이 고르게 향상된 성능을 보이고, 다양한 실험을 통해 패턴 감지 능력에 손상 없이 패킷 처리 성능을 향상시켜 신뢰도를 유지하도록 개선한 알고리즘임을 나타냈다. 그리고 패킷 사이즈의 변화, 즉 초당 처리 패킷 개수의 변화도 룰의 개수에 상관없이 항상 일정한 형태의 추이를 나타내고, 룰의 개수가 많아짐에 따라 그 성능의 차이는 더욱 크게 나타남을 알 수 있다.

또한, 기존의 패킷 분류를 위한 다차원 검색을 튜플 공간이라는 이차원 공간을 이용하여 검색하는 튜플 공간 패킷 분류 알고리즘에 블룸 필터를 적용하여 불필요한 튜플들을 제거하여 적은 메모리 공간을 차지하면서도 적은 메모리 접근 횟수로 패킷을 분류할 수 있는 알고리즘을 추가로 적용하여, 실제 룰 매칭 수행 시 부가 정보들의 내용을 먼저 비교하여 패킷 데이터 부분과 시그니처 전체를 매칭하지 않아도 공격 패킷 여부를 쉽게 파악하여 불필요한 검색을 제한하고 실제 매칭 수행의 가능성을 낮출 수 있다. 블룸 필터는 필터의 입력 값이 필터에 미리 저장된 정보인지 아닌지를 빠르게 판단하여 불필요한 검색들을 걸러내는 역할을 한다. 그리고 블룸 필터의 특징상 false negative는 존재하지 않으며, flase positive라는 오탐을 발생시킬 수 있지만 이 오탐 발생확률을 제어할 수 있으며 아래의 식을 통해 계산할 수 있다[5].

$$F = (1 - e^{-kn/m})^k \quad (1)$$

식에 근거하여 해싱 함수의 개수 k, 블룸 필터의 크기 m, 집합 원소의 개수 n을 적절히 조절하면 최적의 false positive의 비율을 구할 수 있다.

Ⅳ. 결론 및 향후 연구 방향

고성능의 침입방지 시스템을 비롯하여 패킷을 깊은 수준에서 분석하고 확인하는데 있어 매칭 알고리즘은 네트워크 성능이 증가하고 그 처리속도가 중요시되는 현실에 비추어 볼 때 매우 중요한 부분이다.

본 논문에서는 룰의 개수에 영향을 받지 않고 실제 시그니처에 수행되는 룰의 경우의 숱을 줄임으로써 룰이 적게 걸린 것과 같은 효과를 내서 결국 룰의 개수와 관계없는 예측이 가능하고 높은 성능을 낼 수 있도록 고안된 기존의 제안된 시그니처 해싱 알고리즘에 추가로 튜플 블름 알고리즘을 적용하였다. 해싱 기법을 사용하면 룰의 상관관계가 깊을수록 같은 시그니처 해싱 값을 가질 가능성이 높아진다. 이에 튜플 블름 알고리즘을 이용하여 부가정보들을 먼저 검색함으로써 패킷 데이터 부분과 시그니처 전체를 매칭하지 않아도 공격 패킷 여부를 쉽게 파악하여 불필요한 검색을 제한하고 실제 매칭 수행의 가능성을 낮출 수 있도록 기존의 알고리즘을 개선하였다.

그러나 기존의 패턴 매칭 알고리즘에 비해 포트 및 시그니처 해시 테이블을 유지하고 관리하며, 추가로 필터링을 사용함으로써 기존의 방법보다 상대적으로 큰 메모리를 차지하게 되는 측면도 존재한다. 따라서 향후 제안된 알고리즘의 구현 및 추가적인 여러 시험을 통해 메모리 사용량 측면의 문제점을 보완하고, 보다 나은 성능 향상과 오탐률을 낮춘 효율적인 패턴 매칭 기능을 제공하기 위한 기법들을 연구해 나가고자 한다.

참고문헌

[1] ETRI, "하드웨어 기반의 고성능 침입탐지 기술“, 전자통신동향분석, 제22권, 제1호, 2007.02.

[2] X.Zhang, C.Li, and W.Zheng, "Intrusion Prevention System Design", Proceedings of the Fourth International Conference on Computer and Information Technology, 2004. 09.

[3] 왕정석, 권희웅, 정윤재, 곽후근, 정규식, “시그니처 해싱에 기반한 고성능 침입방지 시스템”, 한국컴퓨터종합학술대회, Vol.34, No.1(D), 2007.

[4] 김소연, 임혜숙, “패킷 분류를 위한 블룸 필터 이용 튜플 제거 알고리즘”, 한국정보과학회 논문지: 정보통신, 제37권, 제3호, 2010.06.

[5] Wikipedia, "Bloom filter", from: http://en.wikipedia.org/wiki/Bloom_filter