말단 호스트의 관점에서 본 Pi 필터의 효율적인 필터링을 위한 최적의 표기 비트 크기에 관한 연구

김순동° 홍만표 김동규
아주대학교 정보통신전문대학원, 정보통신 및 시큐리티 연구실
{sdkim°, mphong, dkkim}@ajou.ac.kr

A Study on Marking Bit Size for Path Identification on the End Host Point of View

Soon-Dong Kim° Man-Pyo Hong Dong-Kyoo Kim
Information Communication & Security Lab. Graduate School of Information Communication, Ajou Univ.

요 약

최근의 인터넷의 DDoS문제는 점점 더 빈번하고 심각해지고 있다. 많은 전문가들이 DDoS에 대한 방어를 위해 많은 연구를 하고 있다. 위장된(spoofed) IP 주소로 더욱 복잡한 DDoS 공격들에 대한 방어를 위한 일련의 연구들의 일환으로서 Pi 가 제안 되었다. Pi는 패킷 표기 방식의 접근을 함으로써 위장된 근원지 IP 주소와 관계없이 victim은 동일한 경로를 통해서 도착한 패킷인지를 패킷 단위로 구별할 수 있게 한다. 표기 비트의 크기는 Pi 방식에서 성능에 영향을 주는 중요한 요소 중 하나이다. 최적의 표기 비트의 크기는 인터넷 환경과 토폴로지에 영향을 받는다. 기존의 Pi 는 인터넷 서비스(ISP)의 가장자리에 설치함을 전제로 하였다. 이 논문에서는 인터넷 서비스의 말단 호스트에 설치되는 Pi 필터에서 최적의 효율을 위한 표기 비트의 크기를 찾기 위해 노력 하였다.

1. 서 론

최근 인터넷의 DDoS 문제는 더욱더 심각해지고 있다. DDoS 공격으로 Yahoo!, eBay 등의 대형 인터넷 사이트들도 다운되었으며 최근 2003년 1월 25일에는 Sapphire/Slammer 웜이 인터넷에 급속도로 확산되어 네트웍이 마비되기도 하였다 [4]. 인터넷의 규모가 커져 갈수록 DDoS의 위협도 커져만 가고 있다. 또한 현재의 인터넷 환경에서는 IP프로토콜 자체의 취약점으로 인해 근원지 IP 위장 등으로 인해 공격자의 패킷을 차단하는데 어려움을 겪고 있다. 이러한 취약점의 대안으로서 등장한 것이 IP traceback 이다 [3]. 그러나 이 방식은 victim이 정확한 공격 패킷의 경로를 재구성하기 위해서는 일정량 이상의 패킷을 수집해야 한다는 결점이 있다 [1]. 이러한 결점은 attacker에게 이용될 수 있는 소지가 있으며, 패킷을 수집하고 재구성 하는 프로세스 자체가 DoS를 유발할 수 있다. 다른 대안으로서 Pi는 IP 패킷의 식별 필드(identification field, ID)에 경유 라우터의 IP의 해시 값을 표기(marking)하는 방식이다. 이러한 방식은 패킷 단위로 동일 경로에 대한 식별이 가능해진다. 만일 돌어온 표기 값들이 learning 단계에서 수집된 attacker의 표기 값과 같으면 차단된다. [1].

Pi 방식에서 표기 비트 크기는 성능을 결정하는 가장 중요한 요소이다. 이 논문에서는 말단 호스트(end host)의 관점에서 최고의 성능을 위한 가장 적합한 표기 비트 크기를 찾기 위해 노력하였다.

2. 가 정

본 논문에서는 Pi 필터가 자율시스템(Autonomous System, AS)상의 말단 호스트에 설치된다고 가정한다. 이러한 가정은 서버 자원 고갈 공격(server resource attack)에 대한 일종의 대안일 수 있다. 서버 자원 고갈 공격은 다음 두 가지로 나뉜다.

- -서버 프로세스 공격(server processing attack): 이 공격의 일례로는 필요 없는 패킷을 victim으로 보내는 것이다. 서버의 프로세서는 처리 한도를 넘게 되고, 패킷은 처리되지 않은 상태로 폐기된다.
- -서버 메모리 공격(server memory attack): TCP_SYN 공격[5]이 대표적인 예로서 서버의 특정 메모리를 모두 낭비하도록 하는 공격이다.

Pi 는 패킷 단위로 차단이 가능하기 때문에 이러한 공격에 대해서 이점을 가진다. 마지막으로 인터넷의 모든 라우터가 Pi 지원한다고 가정 한다

3. 표기 비트 크기 결정을 위한 고려 사항들

3.1 자율 시스템(AS)의 토폴로지

자율시스템들은 트리 구조 혹은 매시(mash)망 구조 등 여러 형태를 가질 수 있다. 그러나 말단 호스트의 관점에서

보면 자율시스템 내부 통신을 제외하면 몇몇의 경로만이 자율 시스템 외부와의 통신에 이용된다. 말단 호스트의 입장에서 이것은 트리 구조와 유사하다. 표 1의 Skitter Map[2]의 traceroute 자료에서 보이듯이 말단 호스트에서 3 홉까지의 라우터의 IP 주소가 거의 같음을 알 수 있다.

Data Sets	1st hop	2nd hop	3rd hop	comments
Data1	4131	4131	4130	Single path only
Data2	4042	3370	1636	Two path at 3rd router
Data3	4873	4872	4337	Almost single path
Data4	4286	1344	1156	Well distributed
Data5	4806	1139	631	Well distributed
Data6	4951	1481	2481	Almost two path from 2nd router
Data7	3374	3374	3363	Almost single path

표 1. 3번째 홉까지 가장 빈번히 쓰인 IP주소의 사용빈도

이 값들은 경로를 식별 하는데 의미가 없는 값으로 작용한다. 간단히 정리하기 위해서 자율 시스템과 다른 자율 시스템간의 통신에서 경유하는 상위라우터 3개가 항상 같다고 가정할 경우, 표기 비트 크기가 1 비트 일 때와 2 비트 일 때는 다음과 같은 변별력을 가진다.

1-bit marking: Exp(2, 16-3) = 8182
2-bit marking: Exp(2, 16-6) = 1024

3 홉까지의 라우터들이 항상 같기 때문에 1 비트 표기의 경우는 3 비트가 낭비가 되며, 2 비트 표기의 경우는 6비트가 낭비가 된다.

Pi는 IP 패킷의 16 비트의 식별 필드를 이용하기 때문에 16 개의 라우터 보다 먼 거리의 라우터들의 값은 의미가 없으며, 또한 16홉 이내의(1 비트 표기의 경우 8 홉 이내의) IP 패킷의 식별 필드에는 의미 없는 값들이 채워지게 된다. 만일 14개의 라우터를 경유한다고 할 경우 2 비트는 표기가 되지 않은 채로 victim에게 전달된다. 만일 경유 라우터가 13개라고 한다면 이때의 실제적인 변별력은 다음과 같다.

1-bit marking: Exp(2, 13-3) = 1024
2-bit marking: Exp(2, 16-6) = 1024

위 식에서 보듯이 13개의 이상 16개 이하의 라우터를 경유할 경우에는 1 비트 표기가 2 비트에서 보다 더 높은 변별력을 가짐을 알 수 있다. 13개 보다 적을 경우에는 2 비트 표기 더 높은 변별력을 가진다.

3.2 서 론3.2 인터넷의 경로 길이와 표기 비트 크기

Pi의 성능은 홉 수에 민감하며, 표기된 라우터의 해시 값들에 의해 위양성(false positive)과 위음성(false negative)을 일으킨다. 1 비트 표기에서 10개의 라우터를 경유하는 경로일 경우 표기되지 않은 6비트는 위음성 혹은 위양성을 유발할 수 있다. [1]에서 보였듯이 인터넷 경로의 경유 홉 수의 평균값은 대체로 16홉에 집중되어 있다. 이 자료는 Pi가 최고의 성능을 발휘할 수 있는 환경이다.

이미 제안된 논문 [1]에서 표기 비트에 대한 상세한 고려가 있었으며, 논의의 가치가 있는 것은 1 비트와 2비트 정도임을 밝힌 바 있다. 1비트 표기는 보다 많은 홉의 라우터의 정보를 표기 하지만 한 홉의 정보가 덜 상세한 편이며, 2 비트 표기는 최대 8홉까지만 표기가 가능하지만 각각의 홉 정보는 더욱 상세한 편이다.

4. 실험 계획

본 절에서는 Pi의 성능을 가상적인 DDoS 공격 상황을 만들어서 시뮬레이션을 해본다. 실험용 견본 인터넷 데이터들과 DDoS 공격 모델을 정의하고 마지막으로 실험의 결과를 보이도록 한다.

4.1 인터넷 데이터

본 실험에서는 인터넷 토폴로지로서 CAIDA의 Skitter Map[2] 을 이용한다. 이 데이터는 하나의 호스트가 traceroute 메시지를 보내어 수집한 경로 정보이다. 이 정보 중에서는 완전한 데이터와 중간 홉들의 정보가 나타나지 않은 불완전한 정보가 있는데 실험의 정확도를 위해서 완전한 정보를 가진 데이터만을 이용하기로 한다.

4.2 DDoS 공격모델

공격 패킷을 차단 할 수 있기 위해서는 victim은 공격패킷과 일반 패킷을 구분 할 수 있는 수단을 가져야 한다. 이를 위해서 Pi는 적응단계(learning phase)와 공격단계(attack phase)를 갖는다[1]. 이 논문에서는 적응단계에서 특정 패킷이 공격인지 아닌지를 판단하는 알고리즘은 논외로 친다. victim은 적응단계에서 공격자들에 대한 표기값 목록(attack markings list)을 작성한다. 공격단계에서는 들어오는 특정 패킷의 표기값과 공격자 표기 목록의 표기값들과 비교를 해서 일치하는 표기값이 존재할 경우 공격으로 간주하게 된다.

4.3 실험 설계, 공격 시나리오, 성능

먼저 본 논문에서는 일반 호스트로 CAIDA의 Skitter Map에서 임의의 200,000개의 경로를 선정했다. 이 중에서 어느 정도의 공격자를 임의(randomly)로 선정하여 그들의 경로에 대한 표기를 victim의 공격자 표기값 목록에 추가하였다. 만일에 어떤 호스트가 공격자로 선택되었다면 그 호스트가 보낸 패킷은 항상 공격패킷으로 간주되어 차단된다. 모든 패킷의 식별 필드는 임의의 쓰레기 값들(garbage values)로 채워져서 보내지며 경로상의 각각의 라우터를 지날 때 마다 라우터의 해시 값이 표기가 된다. 이때 각각의 라우터가 표기하는 위치는 패킷의 TTL값을 기준으로 하게 된다[1]. 따라서 1 비트 크기의 표기일 경우, 전체 경유 라우터가 8개라면 8비트가 모두 쓰레기 값으로 채워진다. 성능에 대한 평가를 위한 지수로서는 오판률(false rate)를 사용하도록 하겠다. 오판률은 위양성과 위음성의 합으로 정의하기로 한다.

5. 결과

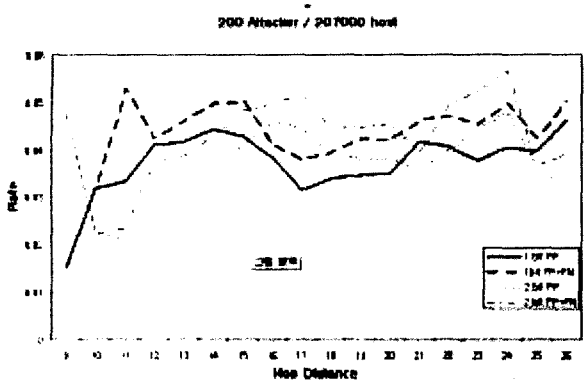

그림 1. 각각의 경유 거리별 오판률

그림 1에서는 각각의 경로의 경유 거리에 따른 오판률을 나타내었다. 대략 207,000의 호스트 중에서 200개의 공격자(attacker)를 생성했으며, 공격자와 일반 호스트는 같은 비율로 패킷을 생성하여 총 100만 패킷을 victim으로 전송하였다. CAIDA의 Skitter Map에서 이미 보았듯이 9홉 보다 작은 경로는 매우 적었으며 나타내지 실험에 크게 영향을 주지 않으므로 나타내지 않았다.

그림 1에서 보듯이, 대부분의 오판률은 위양성에 기인한다. 이것은 대부분의 공격자가 victim으로부터 일정한 거리 이상에 위치하기 때문이다. 공격자 대부분이 말단 호스트이기 때문이다.

2 비트 크기의 표기는 13홉 보다 적은 라우터를 경유 할 경우 더 좋은 효과를 나타내고 있다. 1 비트 표기는 13홉 이상 20홉 미만에서 더 좋은 성능을 나타낸다. 인터넷의 대부분의 경유 길이가 16홉 근처에 집중되어 있다는 지난 결과를 보면 그림 1에서의 약간의 차이는 좀더 큰 의미로 해석이 된다. 이 결과는 본 논문에서 예견했던 것과 잘 일치한다.

6. 결론

본 논문에서는 말단 호스트의 관점에서 Pi 표기 비트 크기에 대한 고려를 하였으며, 인터넷 경유 라우터 수를 중심으로 살펴보았다. 1 비트와 2 비트 크기의 표기는 약간의 서로 다른 장단점을 가진다. 본 논문에서는 Pi의 성능에 대한 지표로서 위양성률과 위음성률의 합인 오판률을 정의 하였다. 2 비트 크기의 표기는 13개 이하의 라우터를 경유하는 경우에 좀더 좋은 성능을 보였으며, 1 비트 표기는 인터넷 경유 길이의 대부분을 차지하는 13~20 여개의 경우에 좀더 좋은 성능을 보였다. 따라서 말단 호스트 관점에서의 Pi 표기 비트의 크기는 1 비트가 좀더 나은 성능을 보임을 알 수 있다.

참고문헌

[1] Abraham Yaar, Adrian Perrig, Dawn Song, "Pi: A Path Identification Mechanism to Defend against DDoS Attacks", Security and Privacy, 2003. Proceedings. 2003 Symposium on , May 11-14, 2003, Pages:93 107

[2] CAIDA. Skitter. http://www.caida.org/tools/measurement/skitter/, 2000

[3] Zhaole Chen; Moon-Chuen Lee, "An IP traceback technique against denial-of-service attacks", Computer Security Applications Conference, 2003. Proceedings. 19th Annual , 8-12 Dec. 2003, Pages:96 104

[4] Berkeley University, "The Spread of the Sapphire/Slammer Worm", http://www.cs.berkeley.edu/~nweaver/sapphire/

[5] Computer Emergency Response Team(CERT). TCP SYN flooding and IP spoofing attacks. Technical Report CA-96:21, Carnegie Mellon University Pittsburgh, PA, Sept. 1996.