제 19 회 한국정보처리학회 춘계학술발표대회 논문집 제 10 권 제 1 호 (2003. 5)

확률적 패킷 마킹에 기반한 확장된 IP 역추적 기법

곽미라, 조동섭
이화여자대학교 과학기술대학원 컴퓨터학과
e-mail : mirakwak@ieee.org

Extended IP Traceback Scheme Based on Probabilistic Packet Marking

Mira Kwak, Dong-Sub Cho
Dept. of Computer Science and Engineering, EIST, Ewha Womans University

요 약

인터넷의 사용이 증가함에 따라 여러 가지 유형의 인터넷 공격이 자주 발생하고 있다. 특히 다양한 서비스 거부 공격이나 분산 서비스 거부 공격들은 최근의 여러 공격 사례에서 발견되고 있어 그 위험성이 크게 나타나고 있다. 이러한 공격들에 대해 대처하기 어려운 이유들 중 하나는, 공격자가 IP 패킷을 조작하여 자신의 IP 주소를 속임으로써 공격의 근원지 파악을 어렵게 한다는 것이다. 이에, 조작된 IP 패킷을 사용한 공격에 대해 그 근원지를 파악할 수 있도록 하는, IP 역추적 연구의 필요성이 대두되었다. 본 논문에서는 이러한 IP 역추적 연구의 시도들 중 그 유용성이 인정되어 많은 연구자들에 의해 지속적인 개선이 이루어지고 있는, 확률적 패킷 마킹 기법을 기반으로 한 확장된 기법을 제안한다.

1. 서론

여러 종류의 서비스 거부 공격, 분산 서비스 거부 공격 등의 네트웍 공격이 최근 급증하고 있다[1]. 이러한 공격을 행하는 공격자들은 그들의 주소를 속여 공격하므로 추적되기 어렵다는 문제를 가져 그 심각성이 크다. 현재 널리 사용되고 있는 방법은 사람의 손이 필요한 수동적인 방법이고, 공격이 이루어지는 동안이 아니면 공격자를 추적하기 어려워, 이러한 유형의 공격에 대해 근원지를 추적하는데 대한 연구의 필요성이 대두되었다. 이러한 근원지 추적 기법을 IP 역추적이라 한다.

많은 IP 역추적 기법들이 현재까지 제안되었지만, 각각 단점과 한계점들을 가지고 있다. 최근의 Savage 등이 제안한 확률적 패킷 마킹 기법은 관련 연구들 중 가장 효율적이고 실용적이어서 그를 기반으로한 추가적인 연구가 활발히 이루어지고 있다. 본 연구에서는 이를 바탕으로 하여 그 단점을 개선하고 추가적인 아이디어를 적용하여, 사람의 노력을 줄이고, 전체적인 계산의 효율성을 높이고자 하였다.

2 장에서는 관련 분야의 연구를, 3 장과 4 장에서는 제안하는 기본적인 아이디어를 소개하고, 5 장에서 논문을 마무리하겠다.

※ 이 논문은 2003 년도 두뇌한국 21 사업에 의하여 지원되었음

2. 확률적 패킷 마킹과 IP 역추적

IP 패킷 마킹 접근방법의 기본적인 개념은 라우터들로 하여금 어떠한 확률을 기준으로 자신에게 도착하는 패킷을 다음 라우터로 전달하는 동안 전체 라우팅 경로의 일부를 그 패킷에 기록하도록 하는 것이다. 그 기본적인 기법인 edge sampling 알고리즘은 edge 정보를 패킷에 기록한다. 이 기법은 시작 IP 주소와 종료 IP 주소, 거리 정보를 저장하는 필드를 패킷 안에 확보하고, 각 라우터는 이 필드들을 다음과 같이 업데이트한다.

각 라우터는 확률 q 로 패킷에 마킹한다. 라우터가 패킷에 마킹하기로 결정하면, 라우터는 그 IP 주소를 시작필드에, 그리고 0 을 거리 필드에 쓴다. 라우터가

마킹하지 않기로 결정한 경우, 거리 필드의 값이 이미 0 이면 이전 라우터에서 이미 패킷에 마킹한 것으로 간주하고 그 IP 주소를 종료 필드에 기록하여, 이전 라우터와 현 라우터 사이의 edge 정보를 나타낸다. 그렇지 않은 경우에는 거리 필드의 값을 증가시킨다. 이렇게 하여 거리 필드는 기록된 edge 로부터 공격대상 호스트까지 거리를 나타내게 된다. 공격대상 호스트는 공격 패킷들에 기록된 정보를 바탕으로 공격 경로 그래프를 구성할 수 있다. 공격 경로의 길이가 d 라 할 때, 그 재구성에는 $ln(d)/q(i-q)^{d-1}$ 개의 패킷이 요구된다.

16 비트 크기의 IP identification 필드에 위에서 설명한 것과 같은 내용을 기록하기 위해서는 특별한 기록 기법이 필요하다. 이를 위해 고안된 compressed edge fragment sampling 기법은 각 라우터들이 그 IP 주소와 중복정보를 여덟 조각 내어 그 중 한 조각과 그것이 몇번째 조각인지에 관한 정보를 확률적으로 IP 패킷에 기록하도록 한다. 이러한 기법을 설계함에 있어 중요한 문제는 IP 헤더 내의 제한된 공간만을 사용하여 효율적이고 정확하며 믿을 수 있는 기록을 가능하게 하는데 있다. 본 논문에서는 공격 대상 호스트에서 공격 경로를 재구성하는데 필요한 패킷 수를 줄이고 사람의 노력을 줄여 알고리즘의 효율성을 높이는데 초점을 맞추었다.

3. 확장된 마킹 기법

3.1. 근원지 정보를 포함한 공격 경로의 재구성

이전의 해법들은 라우터 정보만을 포함한 공격 경로의 구성을 제공하였다. 이러한 방법들을 사용하면 공격 경로를 파악할 수 있지만, 공격지 호스트를 파악하는 데에는 공격 서브넷에 도달한 후 수동적인 노력이 요구된다. 이에 본 연구에서는 공격 근원지 호스트의 정보까지 포함하도록 정보를 마킹하는 기법을 제안하고자 한다. 이 기법에서 기록하는 라우터의 정보는 지금까지의 여느 기법과 마찬가지로 기본적으로 그 IP 주소를 바탕으로 한다. 그러나 공격 근원 호스트의 IP 주소는 공격자에 의해 위조되므로 라우터와는 다른 방법으로 그 정보가 기록되어야 한다. 이를 위해 본 연구에서는 공격 근원 호스트를 식별할 내용으로 MAC 주소를 사용하고자 한다. MAC 주소는 00-50-DA-92-25-D7 와 같이, 12 개의 16 진수로 이루어져 있어 이를 표현하는 데에는 48 비트의 공간이 필요하다. 우리는 48 비트를 8 비트 크기의 여섯 조각으로 이를 표현한다. 이에 관한 자세한 설명은 4 절에서 한다.

3.2. 마킹 확률의 조정

기존의 패킷 마킹 기법들이 마킹 여부를 결정하는 기준으로 삼았던 확률값은 모든 라우터에 대해 일정했다. 이러한 고정확률 마킹은 결과적으로 공격대상 호스트로부터 먼 라우터의 정보일 수록 그보다 공격대상 호스트에 가까운 라우터들에 의해 그 정보가 엎어 쓰여 정보를 담은 패킷 수가 적어진다는 문제를 가진다. 이로 인해 공격 경로 구성에 요구되는 패킷의 개수도 늘어나게 된다. 이에, 공격 경로상의 위치에 따라 다른 확률로 패킷에 마킹하도록 하는 방법에 관한 연구가 이루어지고 있는데, Peng 등이 이에 관하여 자세하게 고찰하고 해결방법들을 제시하였다[5].

만약 공격대상 호스트로부터의 거리가 멀수록 높은 마킹 확률을 적용할 수 있다면 최종적으로 공격 경로를 재구성하는데 필요한 패킷 수를 줄 일 수 있다. 이상적인 경우는 공격대상 호스트가 수신한 모든 패킷들 중 각 라우터로부터 마크된 패킷수가 *1/공격경로길이*로 같은 것이다. 이렇게 되도록 하기 위해, 각 라우터로 하여금 공격 경로 상 자신의 위치에 기반하여 마킹 확률을 조정하도록 해야 한다. 그러나 공격자의 위치가 알려지지 않아 라우터의 공격 경로상 위치도 알 수 없으므로, 이를 다른 취득 가능한 정보에 기반하여 추측하도록 하는 방법이 필요하다. Peng 등은 이를 위한 세가지 방법을 제안하였는데, 각 방법은 서로 다른 거리 측정법을 바탕으로 한다. 그림 1 은 이러한 서로 다른 거리 측정에 관한 정의를 보이는데, d_1 은 공격 호스트로부터 현재 라우터까지의 거리, d_2 는 지난번 마킹한 라우터로부터 현재까지의 거리, d_3 은 현 라우터로부터 공격 대상 호스트까지의 거리를 각각 나타낸다[5].

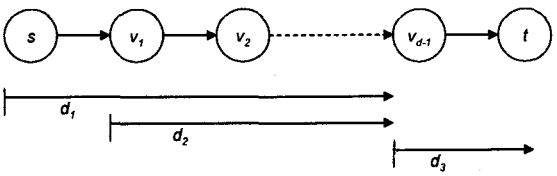

그림 1 거리 측정 기준 정의

본 논문에서는 이 중 가장 이상적인 방법인 d_1 을 사용하는 방법을 택하였다. d 를 전체 공격 경로의 길이, $i = 1, 2, \ldots, d_{i-1}$ 일 때 p_i 를 위치 i 의 라우터의 마킹 확률이라 하자. 최상의 경우는 위에서 언급한 바와 같이 각 라우터가 $1/d$ 의 확률로 패킷을 마킹하는 것이다. 따라서 다음과 같은 식을 구할 수 있다.

$$a_d = p_d = 1/d \quad (1)$$

$$a_{d-1} = p_{d-1}(1-p_d) = 1/d \quad (2)$$

$$a_{d-2} = p_{d-2}(1-p_{d-1})(1-p_d) = 1/d \quad (3)$$

식 1 로부터 $p_d = 1/d$ 을, 식 2 로부터 $p_{d-1} = 1/(d-1)$ 을, 식 3 으로부터 $p_{d-2} = 1/(d-2)$ 을 구할 수 있다. 이로부터 공격 경로상 위치 i 의 라우터의 마킹 확률을 $p_i = 1/i$ 로 정리할 수 있다. 이러한 확률을 가지도록 마킹 기법을 구현하기 위해, i 를 구할 수 있어야 한다. 우리는 라우터간 edge 의 공격대상 호스트까지의 거리 정보 저장에 5 비트를 사용했듯이, 이 정보의 저장을 위해 5 비트를 IP 헤더의

가용 공간으로부터 추가적으로 확보하려 한다. 이 필드의 기본 값은 0 이며 라우터는 패킷을 다음 라우터로 전달할때마다 그 값을 1 씩 증가시킨다. 이에 관한 자세한 내용은 4 장에서 설명한다.

4. 경로 정보의 기록

IP 역추적을 가능하게 하는 정보를 IP 패킷 내에 저장하는 방법이 필요하다. 이 장에서는 기존의 확률적 패킷 마킹에서 제안하는 기록 방법을 기본으로, 경로정보에 추가된 공격지 MAC 주소를 저장하기 위한 방법과 공격 경로상 위치에 따라 다른 마킹 확률을 구하기 위해 추가적으로 필요한 거리 정보를 저장하기 위한 방법을 논한다.

4.1. IP options과 추가적인 패킷의 사용

IP options 를 사용하는 것이 추가 정보의 저장을 위해서 적당한 방법으로 여겨진다. 그러나, 현재 사용되는 많은 라우터들이 하드웨어적으로 options 부분을 가진 IP 패킷들을 처리하지 못하고 있다[11]. 또한 앞으로의 라우터들이 해당 기능을 가지게 되더라도 Savage 등이 제안한 IP 역추적 접근 방법과 관련한 문제들이 여전히 남을 것이다[2]. 이러한 이유로 IP optinos 부분은 우리가 사용하기에 적당하지 않다.

우리가 기록하고자 하는 경로 정보를 IP 패킷들에 저장하기보다, 새로운 프로토콜을 사용하여 이러한 정보들을 캡슐화하여 전송하는 방법도 고려할 수 있다. 그러나 이러한 패킷에 대한 인증이 추가적으로 필요하고 구체적인 설계 방법에 따라서는 트래픽이 대량 증가하는 문제가 생길 수 있으므로 적당한 방법이 아니다.

4.2. IP 헤더: ID 필드와 TOS 필드

IP 패킷 내의 ID 필드 field 는 IP 에 의해 fragment 의 재조합에 사용되는 16 비트 크기의 필드이다. 0.25% 이하의 인터넷 트래픽만이 fragment 들이므로 [12], 이 필드를 사용하는 것은 문제되지 않는다[2]. 이러한 이유로 이전의 패킷 마킹 기법들도 이 필드를 라우터 간 edge 정보의 기록에 사용해왔다.

그러나 이제 우리는 공격지로부터 라우터까지 거리 정보도 저장해야 하므로, 추가적인 공간이 필요하다. 이를 위하여 IP 헤더의 TOS(type of service) 필드를 사용하고자 한다. TOS 필드는 데이터그램의 출발지가 우너하는 서비스의 특성을 나타내는데 사용되는 5 비트 크기의 필드이다. 과거, 이 필드는 드물게 사용되었고, Dean 등은 이 필드를 임의로 세팅하는 것이 패킷의 전달과 사용에 문제를 만들지 않음을 보였다[6].

본 연구에서는 IP 역추적을 위한 정보의 저장에 IP 헤더 내 ID 필드와 TOS 필드를 사용한다.

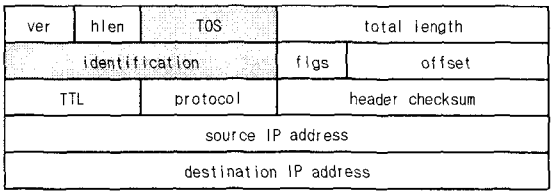

그림 2 IP 헤더 내 사용가능한 공간

4.3. 공격 근원지의 MAC 주소 저장

공격 근원지가 속한 서브넷의 라우터가 패킷에 마킹을 결정한 경우, 공격 근원지의 MAC 주소가 기록되는데 이를 알리기 위해 그림 3 의 '종류' 필드는 1 이 세팅되고, MAC 주소의 fragment 와 그것이 몇번째 fragment 인지를 알리는 내용이 각각 '출발지 MAC *fragment*', '*offset*' 부분에 기록된다.

그리고, 도착지에서 이것이 공격자가 조작한 정보가 아님을 확인하는 방법을 제공하기 위해 각 라우터를 거칠 때마다 특정 위치의 8 비트를 계속하여 XOR 하는데, 이 내용은 그림 4 의 '확인 바이트 offset', '확인 바이트'에 기록된다. 도착지에서 수집된 모든 경유 라우터의 정보를 바탕으로, 라우터들의 첫번째 여덟 비트들의 XOR 결과부터 라우터들의 마지막 여덟 비트들의 XOR 결과를 구할 수 있다. 그 값들과 전달된 패킷들의 '확인 바이트'의 값들을 비교함으로써 패킷이 조작된 것인지, 패킷에 기록된 공격 근원지 MAC 주소를 믿을 수 있는지 판단할 수 있다.

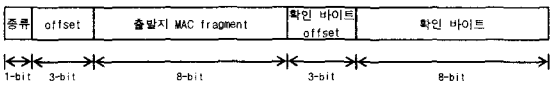

그림 3 공격 근원지 MAC 주소 기록

4.4. 라우터 간 edge 정보의 저장

라우터가 edge 정보를 저장하는 경우, IP 헤더의 ID 필드 사용 방법은 기존의 패킷 마킹 기법들이 사용하는 방법과 같다. TOS 필드에는 이것이 라우터 간 edge 정보의 저장임을 나타내기 위해 그림 4 의 '종류' 부부분이 0 으로 세팅되고, '출발지로부터의 거리' 부분에는 3.2 절에서 설명한 것과 같이 값이 기록된다.

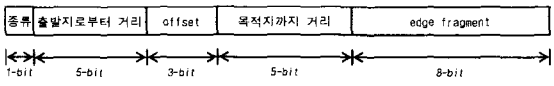

그림 4 라우터 간 edge 정보 기록

5. 결론

본 논문에서는 확률적 패킷 마킹 기법의 개선을 위한 두 가지 아이디어를 제시하였다. 하나는 Peng 등이 제안한 가변적 확률 적용 방법 중 하나를 수정하여 적용함으로써 공격 경로 재구성에 필요한 패킷 수를 감소시켜 전체적인 계산의 효율성을 높인 것이다. 다른 하나는 공격 근원지의 MAC 주소를 전체 경로에 포함하도록 함으로써, 사람의 노력을 줄이도록 한 것이다. 그러나 MAC 주소가 위조되는 경우에 관하여는 아직 고려되지 않았다. 향후, 공격 근원지가 속한 라우터에서 MAC 주소를 확인 후 마킹하도록 하여 제안한 기법의 정확성을 높이고자 한다.

참고문헌

[1] J. Ioannidis, S. M. Bellovin, *Implementing pushback: router-based defense against DDoS attacks*, NDSS, February 2002.

[2] S. Savage, D. Wetherall, A. Karlin, T. Anderson, *Practical network support for IP traceback*, Proceedings of the 2000 ACM SIGCOMM Conference, pp. 295-306, Stockholm, Sweden, August 2000.

[3] M. Waldvogel, *GOSSIB vs. IP traceback rumors*, Proceedings of 18th Annual Computer Security Applications Conference, 2002.

[4] D. X. Song, A. Perrig, *Advanced and authenticated marking schemes for IP traceback*, Proceedings IEEE Infocomm, 2001.

[5] T. Peng, C. Leckie, K. Ramamohanarao, *Adjusted probabilistic packet marking for IP traceback*, Proceedings of the Second International IFIP-TC6 Networking Conference, p. 697, Pisa, Italy, May 19-24, 2002..

[6] D. Dean, M. Franklin, A. Stubblefield, *An algebraic approach to IP traceback*, Network and Distributed System Security Symposium Conference Proceedings, 2001.

[7] T. Dunigan, *Backtracing spoofed packets*, 2000.

[8] K. Park, H. Lee, *On the effectiveness of probabilistic packet marking for IP traceback under denial of service attack*, Proceedings of IEEE INFOCOM'2001 (20th), Page(s): 338-347 vol.1, 2001.

[9] A. C. Snoeren , C. Partridge, L. A. Sanchez , C. E. Jones, F. Tchakountio, S. T. Kent, W. T. Strayer, *Hash-Based IP Traceback*, Proceedings of ACM SIGCOMM'2001, August 2001.

[10] T. Baba , S. Matsuda, *Tracing Network Attacks to Their Sources*, IEEE Internet Computing, Volume 6 Issue 2, March 2002.

[11] S. M. Bellovin, *Personal communications*, May 2000.

[12] I. Stoica, H. Zhang, *Providing guaranteed services without per flow management*, Proceedings of ACM SIGCOMM '99, pp. 81-94, Cambridge, MA, 1999.

[13] K. Nichols, S. Blake, F. Baker, D. Black, *Definition of the differentiated services field (DS field) in the IPv4 and IPv6 headers*, RFC2474, Dec. 1998.