트래픽 특성에 따른 보안장비 성능 분석

정연서° 최양서 김환국 서동일
한국전자통신연구원 정보보호연구단

Performance Analysis of Security Gateway by Traffic

Youn-Seo Jeong°, Yang-Seo Choi, Hwan-Kuk Kim, Dong-Il Seo
ETRI
jys847@etri.re.kr

요 약

본 논문에서는 네트워크 보안장비들의 성능 시험에 관하여 다루고 있다. 표준화 동향과 연구기관들을 조사하고 네트워킹 장비의 성능 수행 권고안인 RFC 2544에 준하여 트래픽의 종류와 크기를 변경하여 시험을 하고, 이에 대한 결과 분석과 보안장비 시험시의 추가적인 고려 사항들에 대하여 기술하였다.

1. 서 론

현재 정보보호 장비의 성능을 측정하는 분야와 관련해서는 이렇다 할 표준이 확립되어 있지 못하다. 정보보호에 대한 인식이 확산되면서 트래픽이 직접 통과되는 방화벽이나 침입방지시스템 등의 정보보호 제품 수요가 늘고 있으나 객관적인 성능 평가 체계가 정립돼 있지 않아, 업체의 설명과 매번 정형화되지 못한 BMT를 실시하여 제품을 선정하고 있는 현실이다. 현재 대기업들은 자체적으로 정보보호 제품의 시스템 적합성 등을 테스트하기 위한 '벤치마크테스트(BMT)'를 통해 제품을 구입하고 있지만 중소기업이나 일반 소비자들은 장비와 인력의 부족으로 정보보호 업체들이 작성한 자료에 의존하고 있는 실정이다. 게다가 정보보호 제품에 대한 '성능'의 개념조차 정립되지 않아 수요자들은 성능과는 무관한 보안성 평가나 네트워크 처리속도 등 일부 항목에 대한 시험결과만을 기준으로 제품을 선택하고 있다.

그동안 정보보호 제품에 대한 평가는 국가정보원이 주관하는 보안성 평가가 주를 이뤄왔다. 하지만 국정원의 보안성 평가와 성능평가는 개념이 다르다. 오히려 성능과 보안성은 반비례한다고 볼 수 있다. 따라서 보안성 평가와는 별도로 제품의 성능을 객관적으로 평가할 수 있는 평가 수행 방법들이 요구되고 있다.

2. 연구동향

2.1 표준화동향

수 많은 인터넷 표준을 개발하고 있는 IETF에서 시험평가 관련한 연구그룹으로는 BMWG(Benchmarking Methodology Working Group)으로 손꼽을 수 있다. BMWG 작업반에서는 다양하게 연동된 기술에 대한 특성을 측정하기 위해 필요한 기술의 개발을 주요 이슈로 삼고 있으며 주요 관련 표준으로는 다음과 같은 RFC들이 있다.

- RFC 1242: Benchmarking Terminology for Network Interconnection Devices
- RFC 2285: Benchmarking Terminology for LAN Switching Devices
- RFC 2544: Benchmarking Methodology for Network Interconnect Devices
- RFC 2647: Benchmarking Terminology for Firewall Performance

RFC 1242에서는 이더넷 통신 장비의 성능에 관련된 용어를 정의하고 있으며, 2285에서는 LAN장비의 성능요소와 측정을 위한 구성을, 2544에서는 실제 통신장비의 성능 측정에 필요한 성능 측정방법을 다루고 있다. 그리고, 2647에서는 Firewall 성능 관련된 용어 정의를 기술하고 있다.

오랫동안 전기통신망에 대한 국제 표준을 다루어온 ITU에서도 성능 시험등에 대한 작업을 수행하고 있다. 통신망 측면의 성능을 연구하고 있는 SG13에서 성능과 관련 할당한 권고안 번호체계는 다음과 같다.

- Y.1530: Signalling, call and connection processing performance
- Y.1540: User information transfer performance
- Y.1550: Timing and synchronization

performance
- Y.1570: Performance of network components
- Y.1580: Performance monitoring and measurement

IETF에서는 테스팅 방법론을 주로 다루고 평가의 기준은 제시하지 않고 있으나 ITU에서는 통신망이 갖추어야 할 몇 가지 파라미터에 대해서 기준 수치를 제공하고 있다. IETF의 경우 이용자와 서비스측면에서 ITU에서는 사업자와 통신망 측면의 방법론을 다루고 있음을 알 수 있다. 본 논문에서는 IETF의 RFC 2544에 준한 방법으로 성능을 측정하고 그 결과를 분석하였다.

2.2 관련연구기관

스위치, 라우터 등의 네트워킹 장비나 방화벽이나 IDS 등 보안제품들의 평가는 오랫동안 여러 기관들에서 연구 및 수행을 하여 왔다. IDS와 같은 보안장비의 경우 그 평가 체계와 성능을 위한 방법 마련에 대한 많은 연구와 평가가 진행되어 왔다. 대표적인 기관들을 정리하면 다음과 같다.

UCD(University of California at Davis)

첫 IDS 테스팅 플랫폼을 구축한 곳으로 Telnet, FTP, rlogin 세션을 사용해서 자동적으로 침입이 발생하도록 하는 방식을 스크립트 형태로 만들어 테스트를 진행하였다.

IBM Zurich Research Laboratory, MIT/LL (MIT Lincoln Laboratory)

FTP 서버에 대한 공격을 탐지하도록 설계된 IDS를 테스트하기 위한 자동화된 공격과 백그라운드 트래픽을 생성하도록 만들어졌다. 초기의 낮은 탐지율과 높은 false alarm을 제거하였고, 평가를 위한 실행이 완전히 자동화 작업이였다는 점이 특징적이다.

AFRL(Air Force Research Laboratory)

DARPA의 지원 아래 1998년과 1999년에 IDS에 대한 평가가 이루어졌으며, 좀더 복잡화된 계층구조의 네트워크 환경에서 실시간으로 IDS를 테스트하여 보고서를 제공하였다.

MITRE

MITRE에서는 사업자들과 정부에서 개발한 많은 IDS 시스템들의 평가를 수행하였다.

Neohapsis/Network-Computing

Network-Computing에서 지원하여 Neohapsis 랩에서 수행하였다. 최근 12개의 상업용 IDS제품과 Open source인 Snort에 대한 평가를 수행하였다.

NSS Group

2000년과 2001년에 IDS와 취약성 스캐너에 대한 평가를 수행했으며, 최근 활발한 활동을 하고 있는 기관이다. 특히 IDS의 경우에는 15개 업체의 제품과 Snort IDS를 포함하고 있다. 이 보고서에는 각 IDS에 대한 구조와 설치법, 구성, 리포팅 스타일부터 분석 방식까지 자세한 정보를 함께 제공하고 있다. Port scans, DoS, DDoS, Trojans, Web, FTP, SMTP, POP3, ICMP, Finger 공격을 모두 광범위하게 테스트한 결과를 다루고 있다. 최근 IPS 테스팅 결과를 발표하였다.

앞에 기술된 연구기관들의 경우 대개 IDS 제품의 평가 기술 연구에 많은 역할을 담당하여 왔다. 최근 실제 상업적인 제품들의 성능평가는 마이어컴(Miercom)이나 톨리(Tolly) • 익사(ICSA) • NSTL 같은 민간 기관이 맡고 있다.

국내에는 이 같은 평가인증 기관이 없어 정보보호 업체들은 주로 ICSA, Tolly 등에 의뢰해 평가받고 인증을 획득하고 있다. 따라서 앞으로 정보보호 업체들의 국제경쟁력을 높이기 위해서는 국내에도 국제적 수준의 성능평가 전담 기관이 필요하다.

정보보호 제품에 대한 성능평가는 먼저 성능에 대한 개념을 정립하여야 한다. 넓은 의미의 성능이라고 하면 표준에 맞는가를 검증하는 표준적합성 시험과 장비간 상호연동 여부를 검증하는 상호운용성시험, 그리고 네트워크상에서의 지연(latency) • 스루풋(throughput) 측정 등 네트워크 퍼포먼스를 포함한다. 좁은 의미로는 네트워크 퍼포먼스만 해당된다. 이러한 것들을 모두 포괄하는 개념의 정보보호 제품 성능평가 체계는 국내에 아직 갖춰져 있지 않고 있으며, 주로 방화벽과 가상사설망(VPN)에 대한 네트워크 성능 위주로 한국정보통신기술협회(TTA)가 시험인증서비스를 제공하고는 있다.

2.3 RFC 2544

본 논문에서는 RFC 2544에 정의된 테스트 수행 방법을 기반으로 보안장비에 대한 성능평가를 수행하게 된다. 장비를 평가하는데 필요한 내용들을 다음에 정리하여 기술하였다.

먼저, RFC2544에서는 이더넷 시험에 사용하는 프레임 사이즈를 아래와 같이 7개로 정하고 있으며, 프리앰블 값과 패킷간의 갭은 각각 64bit, 96bit로 권고하고 있다.

64, 128, 256, 512, 1024, 1280, 1518

Preamble 64 bit, Gap 96 bits

그리고, 권고안에는 100M 이더넷 장비의 최대 프레임 처리율에 대한 참고 값을 포함하고 있는데 전송 패킷 사이즈별 처리 프레임수로 나타내 주고

있는 것을 볼 수 있다.

(bytes)	(pps)
64	14880
128	8445
256	4528
512	2349
768	1586
1024	1197
1280	961
1518	812

시험에 사용되는 트래픽은 아래 형태의 UDP 패킷(UDP echo request)을 정하고 있으며, 패킷의 발생 시간은 적어도 60초로 규정하고 이진검색의 방법으로 평가 대상 장비의 실제 패킷 처리율 값을 찾는다.

```
-- DATAGRAM HEADER
 offset data (hex)              description
 00    xx xx xx xx xx xx        set to dest MAC
address
 06    xx xx xx xx xx xx        set to source MAC
address
 12    08 00                    type

-- IP HEADER
 14    45   IP version - 4 header length(4 byte
units)-5
 15    00                       TOS
 16    00 2E                    total length*
 18    00 00                    ID
 20    00 00     flags (3 bits)-0 fragment, offset-0
 22    0A                       TTL
 23    11                       protocol - 17 (UDP)
 24    C4 8D                    header checksum*
 26    xx xx xx xx              set to source IP
address**
 30    xx xx xx xx              set to destination IP
address**

-- UDP HEADER
 34    C0 20                    source port
 36    00 07                    destination port 07 =
Echo
 38    00 1A                    UDP message length*
 40    00 00                    UDP checksum

-- UDP DATA
 42    00 01 02 03 04 05 06 07     some data***
 50    08 09 0A 0B 0C 0D 0E 0F
```

3. 트래픽 성능 시험

3.1 수행환경

본 시험에서는 공격 탐지와 패킷 차단의 기능을 가진 보안게이트웨이 시스템을 대상으로 트래픽 처리 성능을 분석하였다. 시스템의 사양은 Intel Xeon 3.06Ghz, 512 cash memory, 2G main memory가 장착되었으나 시스템의 성능보다는 트래픽 특성에 따른 시스템의 성능 변화를 위한 시험이어서 시스템의 성능을 위한 별다른 조치는 하지 않고 수행하였다. 침입탐지 정책을 150여개 설정하고, 패킷 필터링 정책의 경우는 1개만 적용하였다. 시험 환경은 아래 그림과 같이 테스터기에서 테스팅 대상 장비로 연결하고 장비를 거쳐 나온 트래픽을 다시 테스터기가 분석하여 처리 성능을 평가 수행하게 된다.

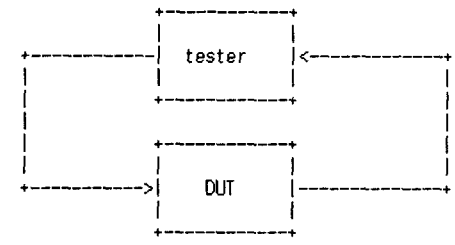

그림.1 수행환경

3.2 시험 결과 및 분석

먼저 RFC 2544에 기술된 UDP 패킷과 동일한 형태의 TCP 패킷을 발생시켜 각 패킷 사이즈별로 트래픽을 단방향으로 발생하여 실험을 수행하였다. 패킷 사이즈 중간에 보이는 370Byte의 경우는 국내에서 일반적인 성능평가나 BMT환경에서 사용되고 있는 인터넷 평균 트래픽 값이다. 국외의 평가 수행기관에서는 540Byte를 사용한 경우도 있다.

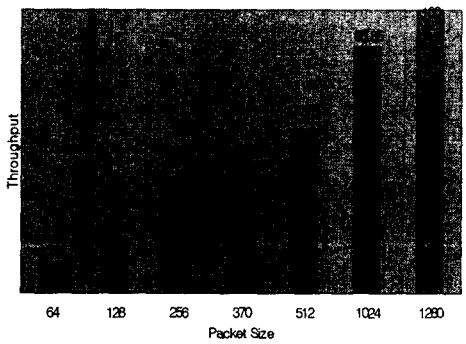

그림 2. UDP 트래픽 시험

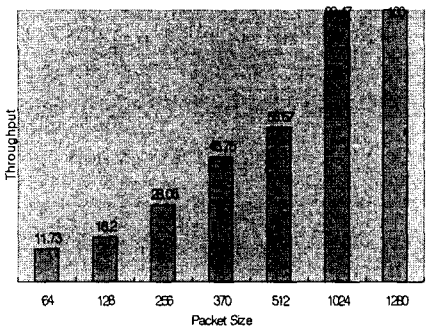

그림 3. TCP 트래픽 시험

시험결과와 같이 UDP, TCP 트래픽의 종류에 따른 큰 차이는 나지 않았으며, 패킷의 크기에 따라 작은 크기의 패킷에서 많은 차이를 보이고 있다. 평균치 값으로 측정한 경우 거의 반 이하로 처리율이 내려가 있음을 알 수 있다. RFC 2544의 경우 2, 3 Layer 네트워킹 장비를 위한 평가 방법을 기술하고 있으므로 실제 보안 장비를 위한 시험시에는 다른 여러 가지 것들이 추가적으로 고려되어야 한다. 2, 3 Layer에서 수행되는 장비의 경우는 상관없지만 연결 상태 값이 필요한 방화벽이나 IDS, IPS 장비와 같은 경우 처리 세션수도 성능 척도로 사용되고 있으므로 TCP 트래픽의 경우 ACK 트래픽을 발생시킨 후 연결 상태를 유지하면서 평가를 수행하여야 하며, 탐지 룰이나 패킷 차단 룰의 개수에 따라서도 성능의 차이를 보이며 실제 망의 트래픽 종류(TCP, UDP, ICMP)에 따라서 성능 수치가 차이 날 것으로 보인다.

실제 보안 장비를 선정하려는 경우 적용망의 특성을 고려하여 적절한 처리 성능을 가진 장비를 선정하여야 할 것이다. 특히, 게임이나 전자상거래, 인터넷 뱅킹 등의 트래픽이 주로 통과하는 경우는 일반적으로 패킷의 크기들이 작으며, FTP 서버들이 많은 곳이나 개인 P2P 사용자들의 사용이 많은 곳에서는 크기가 큰 TCP 데이터의 양이 많을 것이다.

4. 결론

본 논문에서는 트래픽의 크기와 종류에 따른 보안장비의 성능을 시험하였다. 시험 결과 패킷의 크기에 따른 처리율은 많은 차이를 나타내고 있으며 종류에는 크게 성능을 좌우하지 못한 것으로 나타났다.

본 실험에서는 TCP 트래픽의 경우 connection에 대한 고려를 하지 않고 시험을 하였기 때문에 실제 환경 적용시에는 크지 않지만 어느 정도의 차이가 있을 것으로 판단된다. 차후로 이 점을 보완하고 TCP, UDP, ICMP 패킷의 양을 인터넷 트래픽의 평균 추정치로 혼합하여 시험을 진행할 예정으로 있다.

참고문헌

[1] 전자신문 "정보보호제품 성능평가체계 마련 급하다" - 2002. 8
[2] 정보보호21 "정보보호솔루션 성능평가 활성화되나" - 2003. 7
[3] 정보보호21 "벤치마킹테스트, 얼마나 믿고 계십니까" - 2003. 11/12
[4] Peter Mell, Richard Lippmann, Josh Haines and Marc Zissman, An Overview of issues in Testing Intrusion Detection Systems, NIST and MIT/LL report
[5] Gigabit Intrusion Detection Systems group test, NSS Report, July 2003.