동적이고 지능적인 방화벽을 위한 Extended Information Filtering Engine

김주한*, 이민수*, 권성구*, 한윤택*, 송오영*, 박세현*
*중앙대학교 전자전기공학부

eXtended Information Filtering Engine for Dynamic and Intelligent Firewall

Joohan Kim*, Minsoo Lee*, Sunggu Kwon*, YounTaek Han*, Ohyoung Song*, Sehyun Park*

*School of Electronical & Electrics Engineering, Chung Ang University.

요 약

윈도우 자체에서 제공하는 방화벽이나 기존의 솔루션으로 제공되고 있는 방화벽에 대한 한계점을 보완하고자 기존의 패킷 필터링 방화벽과 어플리케이션 필터링 방화벽의 장점을 모두 가지고 있는 하이브리드형의 방화벽을 제안하고, 사용자들이 방화벽 시스템의 부하를 줄이기 위해서 각종의 서비스들을 제공할 때 사용자 프로파일의 context에 기반을 둔 확장된 필터링 엔진을 제안한다.

I. 서론

기존 방화벽의 개념은 단지 네트워크상에서 패킷의 헤더만을 분석하여 그에 따라 차단 또는 통과시키는데 그 목적을 두고 있다. 최근에는 단지 외부에 대해서 공격을 막기 위한 네트워크 레이어에서의 패킷 필터링 방법과 더불어 사용자가 원하지 않는 정보에 대한 차단을 위하여 어플리케이션 레이어에서의 필터링에 대해서도 중요성이 대두 되고 있다. 현재 가장 많이 사용되는 필터링 방법으로는 URL Filtering과 Keyword Filtering 방법이 있다. 하지만 이러한 두 가지 방법에도 한계점이 있고, 사용자에게 보안성과 자율성의 trade-off를 강요하고 있기 때문에 편의성을 제공하지 못한다. 예를 들어 윈도우에서 기본적으로 제공하는 URL Filtering은 보안 수준을 높여 놓으면, 일반적인 웹 서핑을 하는데 불편한 점이 많다. 따라서 본 논문에서는 보안성과 자율성을 모두 제공할 수 있는 동적인 필터링 기능을 탑재하고 또한 사용자별 상황 적응적 지능적인 필터링 시스템을 제안한다.

본 연구는 정보통신부 및 정보통신연구진흥원의 대학 IT연구센터(중앙대학교 홈네트워크 연구센터) 지원 사업의 연구결과로 수행되었음

II. Firewall

2.1 Concept of filtering

네트워크 게이트웨이 서버에 위치하고 있는 일련의 연관된 프로그램들로서, 다른 네트워크의 사용자들로부터 사설 네트워크 또는 개인 컴퓨터의 자원들을 보호해준다. 기본적으로 패킷 필터링은 라우터 프로그램과 밀접하게 동작함으로써, 모든 네트워크 패킷들을 모든 네트워

크의 수신 처로 전달할 것인지를 결정하기 위해 검사하고, 여과한다.

2.2 Packet Filtering and Application Gateway Filtering

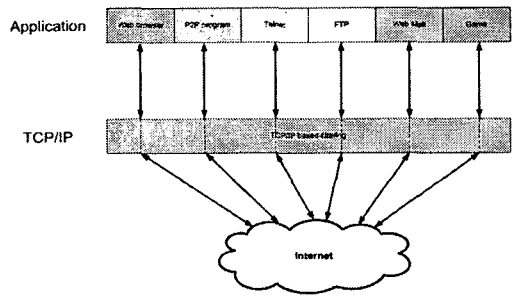

그림 1 packet filtering

패킷 필터링 방화벽은 커널 내부에 별도의 코드(Network Policy)를 가지고 있어서 네트워크에서 패킷이 도착했을 경우 Packet header의 내용을 비교하여 패킷을 전달한다. 보통 패킷의 IP, Port Number에 따라 정책을 적용하며 단순하고 처리 속도가 빠른 대신, 구조가 단순하여 사용자 별, 서비스 별 제어가 불가능하다.

앞의 방화벽은 네트워크 레이어에서의 방화벽, 즉 패킷의 헤더를 통한 방화벽의 구현을 이루었지만, 그 데이터의 Information을 분석하여 필터링할 수 있는 Application Gateway 방화벽은 각 서비스별 프록시를 이용하여 패킷 필터링 방식처럼 IP 주소 및 포트를 이용하여 네트워크 접근 제어를 할 수 있으며 강력한 인증 및 부가적인 서비스도 제공한다. 그렇지만 트래픽이 방화벽의 Application 레이어에서 처리되므로 패킷 필터링에 비해서 성능이 떨어진다.

2.3. Application Filtering Method

URL Blocking은 현재 가장 많이 사용되는 콘텐츠 필터링 방법이다. 이와 같은 방법은 블랙리스트를 생성하여 그 리스트에 등록된 URL의 웹 페이지를 막는다. 그렇지만 이 방법은 매일 새로운 웹 콘텐츠가 폭발적으로 늘어나는 것을 업데이트 하는데 많은 비용과 시간이 소모된다.

Keyword Filtering은 개념적으로는 컨텐츠에 포함된 키워드에 대한 접근을 막는다는 것으로 간단하지만, 컨텐츠 키워드의 잘못된 구문에 대한 해석으로 Over-Blocking이 자주 일어난다.

Intelligent Content Recognition은 웹 컨텐츠가 웹 브라우저를 통과하기 전에 동적으로 스스로 발견하고 판단하는 적절한 메커니즘이다. 이는 컨텐츠의 단어 수, 페이지의 구성, 배경색 등과 같은 컨텐츠의 속성들을 분석하고 분류한다.

III. Dynamic and Intelligent Filtering

기존 방화벽의 보안 레벨과 서비스 이용에 대한 자유도는 Trade-off 관계이다. 외부로부터 원하지 않게 들어오는 traffic은 모두 차단하고 내부의 인증된 사용자들만이 외부 네트워크를 이용할 수 있도록 하기 위해서는 Packet Filtering 방식이 필요하고, 외부의 특정 사용자에게 접속을 허용하거나 IP Spoofing과 같은 해킹 공격을 차단하기 위해서는 사용자 인증이 가능하고 서비스별 보안 레벨의 제어가 가능한 Application Gateway 방식의 방화벽이 필요하다. 그러므로 현재의 Firewall의 장점을 통합하여 기존 Firewall의 한계를 극복할 수 있는 Hybrid Firewall을 구성한다.

이 하이브리드 방화벽은 수신되는 information에 대해서는 패킷필터링을 통해 한 번 필터링된 정보를 어플리케이션에서 한 번 더 필터링해주고, 자신의 정보를 내보내는데 있어 즉 나가는 정보에 대해서 두 번의 필터링을 하게 하여 패킷 필터링과 어플리케이션 필터링의 이점을 모두 살릴 수 있어 좀 더 진화된 방

화벽의 모델을 제시한다.

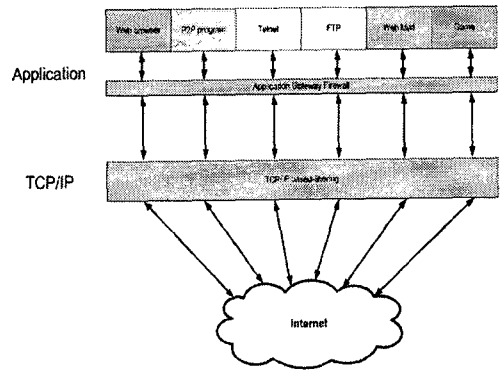

그림 2 hybrid filtering

Application Gateway Firewall에서의 Information Filtering은 Firewall을 통해 진입하는 정보의 keyword를 이용한 Filtering을 기본으로 한다. 웹에서 HTML 문서를 parsing하여 메타데이터를 이용하여 정보를 분류, Filtering하는 Information Filtering Engine을 확장하여 이용한다. 이 Information Filtering Engine의 경우에 정보를 Filtering 하기 위한 정보의 기준인 keyword list가 중요하다. 이 keyword list에 대하여 Clustering Categorization과 user-profile을 context 화하여 keyword list들에 대한 작성을 사용자에게 보다 편리하면서 보안성을 높이게 한다.

IV. Extended Information Filtering Engine

4.1 Information Filtering Engine

기본적으로 이 Information Filtering Engine은 웹에서 주로 사용하는 방식으로 HTML형식의 문서형태를 파라미터 형태로 분할하여 Group화하여 유사함을 찾아내는 방식이다. 이 파라미터들은 개인적 단어, 바탕 색, 링크 개수, 이미지 개수, 단어 개수 등의 특징들을 포함하고 있다. 이들은 Raw Data Vector로 구성되고, 이 벡터는 HTML 페이지로부터 해석된 모든 정보로 정의한다. RDV는 너무 크기 때문에 실시간으로 처리할 수 없거나, 신빙성을 제공할 수 없기 때문에 적절한 Information으로 만들어지기 위해서는 가공되어야만 한다. RDV로부터 특정한 패턴을 찾는 것에서부터 이러한 가공은 시작된다. 예를 들어, 폰트의 색과 배경색을 비교하여 구별되는 패턴을 찾는 것으로부터 시작해서 수백 가지의 RDV를 수십 가지의 RDV로 줄일 수 있다. 이러한 클러스터링 메커니즘에 의해서 특징들의 관계들과 조합들을 가져오고 미리 분류된 카테고리들에 대응하는 그룹을 형성하게 된다. 이렇게 clustering되고 Categorization된 information이 주로 application filtering의 키워드로 사용한다.

4.2 User-profile based Context-awareness

사용자의 상황인지를 위한 사용자 프로파일을 생성하고 관리하여, 이 프로파일에 기반을 두어 사용자의 선호 application service를 관리하기 위해서 4.1절의 가공된 Information과 사용자의 선호 프로그램을 위한 packet을 관리하여 사용자에게 맞춰진 서비스를 제공하도록 한다.

프로파일 기반 사용자별 패킷을 관리하여 사용자의 서비스들을 계층별로 모듈화하고 사용하고자 하는 서비스를 미리 인지하여 그에 따른 패킷과 프로세스를 예측하여 시스템을 보다 능동적으로 동작하게 하고 시스템의 부하를 감소시킨다.

사용자들마다 서비스들에 대한 프로파일을 layer 접근법에 의하여 프로파일을 작성하고 layer에 따라 서비스들을 User-profile에 의하여 그림 3과 같이 모듈화 하여 제공한다. 이렇게 모듈화된 서비스는 각각의 사용자의 환경에 맞게 적용하는 것이 좀 더 편의성을 가질 수 있다. 사용자들이 선호하는 서비스가 다르지만, 서비스에 접근하는 환경이 동일하다면, 프로세스에 우선순위를 제공하여 서비스를 모듈화함으로써 보다 빠르고 효율적인 시스템을 구성한다.

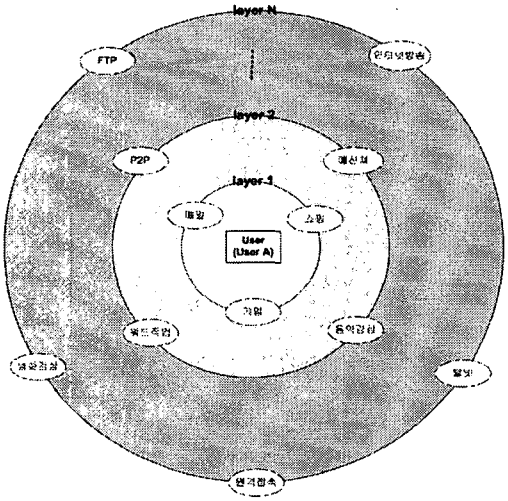

그림 3 모듈화된 서비스

각 사용자마다 생성된 프로파일을 context 화 하면, 그림 4와 같이 만들 수 있다. 그림4에서 사용자A가 즐겨 이용하는 게임의 경우는 사용 시간과 Process counter가 높아 시스템의 ready queue의 높은 순위에서 대기하게 되고 Firewall 에서 의도하지 않은 block을 해제시켜 사용자의 자율성을 높이되, Information Filtering Engine 에서 내부의 사용자의 선호하지 않는 내용을 필터링하여 보안성을 높인다.

User A	사용한 port	사용한 IP address	사용시간		Process Counter
게임	7777	211.218.152.218	1시간	------	1
뉴스	80	218.145.68.236	20분		1
쇼핑	80	211.44.62.230	10분		1
게임	7777	211.218.152.218	30분	------	2
동작감상	-	-	43분		1
쇼핑	80	211.233.17.11	11분		2
	⋮		⋮	------	⋮
게임	7777	211.218.152.218	20분		3

그림 4 profile의 context화

V. 결론

본 논문에서는 기존 방화벽의 보안 수준을 높이면 자율성이 낮아지고, 자율성을 높이면 보안성이 약해지는 문제점이 있던 서비스 이용의 한계에 대해서 보완하였다. 그리고 여러 서비스를 이용할 때, 프로세스들에 대한 부하들을 줄이기 위해서 사용자의 프로파일에 기반을 둔 방화벽에 대한 Filtering 엔진에 대해서 연구하였다.

본 논문에서 제시한 동적이고 지능적인 방화벽 엔진은 기존의 패킷에서 주소에 의한 필터링, 혹은 어플리케이션에서 keyword 중심의 필터링 기능을 가진 방화벽보다 자율성과 보안성의 trade-off를 좀 더 완화시키며 더 강력한 기능을 제공할 수 있을 것으로 기대된다.

[참고문헌]

[1] N. Churcharoenkrung, Y. S. Kim and B. H. Kang, Dynamic Web Content Filtering based on User's Knowledge, Proceedings of the International Conference on Information Technology: Coding and Computing(ITCC '01) 0-7695-2315-3/05

[2] Vijoy Pandey, Dipak Ghosal and Biswanath Mukherjee, Exploiting User Profiles to Support differentiated Services in Next-Generation Wireless Network, IEEE Network, 2004.09, pp. 40-48

[3] P. Y. Lee, S. C. Hui, A. C. M. Fong, An Intelligent Categorization Engine for Bilingual Web Content Filtering, IEEE Transactions on multimedia, Vol 7, No 6, Dec 2005, pp. 1183-1190

[4] A. Belaid, L. Pierron and N. Valverde, Part-of-Speech Tagging for Table of Contents Recognition, IEEE 2000, pp. 451-454

[5] Daniela Godoy, Analia Amandi, User Profiling for Web Page Filtering, 2005 IEEE Internet computing, pp. 56-64