취약성 분석 데이터를 이용한 부정확한 경고 축약 알고리즘

양진석*, 김태균*, 홍순좌*, 정태명**
*국가보안기술연구소, **성균관대학교
e-mail:*{jsyang, tgkim, hongsj}@etri.re.kr,
**tmchung@ece.skku.ac.kr

Imprecise Alert Reduction Algorithm Using Vulnerability Assessment Data

Jin-Seok Yang*, Tae-Gyoun Kim*,
Soon-Joa Hong*, and Tai-Myung Chung**
*National Security Research Institute,
**Dept. of Computer Engineering, Sungkyunkwan University

요 약

침입탐지시스템과 같은 기존의 보안 시스템의 가장 큰 단점 중에 하나는 많은 경고로 인하여 정확한 공격 탐지 분석이 매우 힘들다는데 있다. 이는 관리자의 지식과 경험으로 극복될 수 있지만 관리자의 지식과 경험이 많다 할지라도 관리 네트워크의 범위가 클수록 경고 메시지 분석에 대한 어려움이 존재한다. 또한 관리자에게 보고하는 많은 경고 중에 정확하지 않은 경고가 다수 포함되어 있기 때문에 필요하지 않은 경고까지 분석해야하는 부담을 갖는다. 본 논문에서 제시하는 알고리즘은 보안 시스템의 경고와 취약성 분석 데이터의 상관관계를 분석하여 취약성이 존재하지 않는 공격에 대한 경고 메시지를 필터링한다.

1. 서론

최근 사이버 공격은 지능화, 분산화, 자동화되고 있으며 이러한 공격을 방어하기 위한 보안 장비들도 계속해서 발전하고 있다. 그러나 이러한 보안 장비들은 몇 가지 문제를 가지고 있으며 이러한 문제로 인해 탐지 및 방어를 위한 한계가 발생한다. 예를 들면 침입탐지시스템 및 침입방지시스템은 false positive 등의 오탐지율이 높은 단점을 갖고 있다. 관리자는 오탐지를 포함하는 많은 경고 중에 올바른 경고와 그렇지 않은 경고를 선별 및 분석하고 이에 대응해야하는 부담을 안고 있다. 또한 이기종 간의 침입탐지시스템의 경우 경고의 형식과 내용이 틀려 이를 분석하는 것이 전문가들에게 조차도 쉬운 일은 아니다[1].

보안 장비 및 관리 대상은 늘어만 가고 이러한 장비들이 발생하는 많은 경고들에 대해서 관리자가 분석하고 조치하는데 한계가 있으며, 이로 인해 자동화된 보안 관리 및 정확한 탐지 및 차단 결과 보고에 대한 요구 사항이 발생하였다. 여러 보안 장비들의 이러한 한계를 해결하기 위해서 경고 메시지의 데이터마이닝, 클러스터링, 머징(merging) 등의 연구가 다각도로 진행되었다[2,3,4].

본 논문은 취약성 평가 데이터와 보안 장비의 경고와의 상관관계를 이용한 경고 메시지 축약 알고리즘을 제안한다. 취약성 평가 데이터는 취약성이 존재하는 노드의 IP 주소, 취약성 구별자, 취약성 이름, 포트 번호, 취약성에 영향을 받는 운영체제, 운영체제 비율 등의 데이터를 이용하고, 보안 장비의 경고 데이터는 원본 IP 주소, 목적지 IP 주소, 공격 이름, 서비스 포트 번호의 4개의 속성들의 상관관계를 이용하여 경고 메시지 축약 알고리즘을 기술한다.

본 논문의 구성은 다음과 같다. 2장에서는 관련 연구를 기술하고 3장에서는 경고 축약 알고리즘에 쓰이는 취약성 평가 데이터와 경고 데이터에 대해서 분석한다. 4장에서는 분석된 데이터를 기반으로 경고 축약 알고리즘을 제안하며 5장에서는 결론을 기술한다.

2. 관련 연구

경고 메시지를 감소시키기 위한 연구는 다양한

접근 방법들이 존재한다. 그 중에 대표적인 방법은 데이터 마이닝, 클러스터링, 머징이 있다.

이러한 연구들은 기종간의 침입탐지시스템 경고의 상관관계를 분석하여 다수의 침입탐지시스템에서 발생하는 많은 경고를 감소시키고자 하는 연구이다. 경고 메시지 감소를 위한 연구들은 이기종간의 IDS 경고를 처리하기 위해서 대부분 IDMEF(Intrusion Detection Message Exchange Format)을 사용한다. 또한 경고 메시지 감소를 위해서는 경고 메시지 간의 특성을 분석하여 두 개 혹은 그 이상의 경고 메시지를 하나의 새로운 경고 메시지를 만드는 접근 방법이다[2,3,4].

상기 기술한 접근 방법은 경고 메시지들 간의 상관관계를 이용하여 경고 메시지를 감소시키는 방법들이다. 그러나 본 논문에서는 경고 메시지와 취약성 평가 데이터와의 상관관계를 분석하여 경고 메시지를 감소시키는 새로운 접근 방법에 대해서 기술한다.

3. 데이터 분석

경고 축약 알고리즘은 취약성 평가 데이터와 보안 장비의 경고 데이터의 상관관계를 통해 이루어진다. 이번 장은 경고를 축약하기 위해 사용되는 취약성 평가 데이터와 보안 장비의 경고 데이터를 각각 분석하고 두 데이터들 간의 상관관계를 기술하고 4장에서 이를 기반으로 경고 축약 알고리즘을 기술한다.

1. 경고 데이터 분석

침입탐지시스템이나 침입방지시스템은 다양한 형태의 경고 메시지를 발생한다. 본 논문에서는 이러한 경고 메시지 중 대부분의 보안 장비가 제공하는 원본 IP 주소(srcIP), 목적지 IP 주소(targetIP), 공격명(attackName), 서비스 포트(servicePort) 번호를 이용한다.

원본 IP 주소는 공격을 시도하는 곳의 IP 주소를 나타내며, 공격이 단일 호스트에서 시도하는 것이 아니라 분산된 여러 곳에서부터 시도된 경우에는 원본 IP 주소를 가지지 않는다. 목적지 IP 주소는 탐지된 공격이 단일 호스트를 공격하는 경우에 존재한다. 그러나 다수의 불특정 호스트를 대상으로 하는 경우에는 존재하지 않는다. 공격명은 탐지된 공격이 이미 잘 알려진 공격으로 CVE-ID가 부여된 경우라면 CVE-ID를 할당하게 되고, 탐지된 공격이 알려지지 않은 공격으로 CVE-ID가 존재하지 않는 경우에는 공격명을 가지지 않는다. 그리고 공격명이 존재하는 경우는 공격에 대한 명확한 정보를 알 수 있기 때문에 포트 번호에 대한 정보는 알 필요가 없다. 따라서 포트번호는 공격이 단일 포트, 즉 특정 서비스에 대한 알려지지 않은 공격을 탐지한 경우에만 존재하게 된다. 이러한 네 가지의 속성을 조합하면 모두 16가지의 조합을 만들 수 있는데, 의미 없는 6가지를 제외하면 다음 <표 1>과 같다. 여기서 의미 없는 6가지는 공격명과 포트번호를 둘 다 가지게 되는 4 가지 경우와 단지 포트번호만을 가지는 경우, 아무 정보도 없는 경우를 뜻한다.

<표 1> 경고 데이터

경고 데이터 타입	원본 IP (srcIP)	목적지IP (targetIP)	공격명 (attackName)	서비스 포트번호 (servicePort)
1	■	■	■	
2	■	■		■
3	■	■		
4		■	■	
5	■		■	
6	■			■
7		■		■
8	■			
9		■		
10			■	

각 타입의 특징을 상세히 기술하면 다음과 같다.

- 타입 1: 단일 소스로부터 단일 대상으로의 특정 공격을 나타낸다.
- 타입 2: 단일 소스로부터 단일 대상으로의 특정 서비스의 공격을 의미한다.
- 타입 3: 단일 소스로부터 단일 대상으로의 불특정한 공격을 나타내며 DoS 공격과 같은 경우가 이에 해당한다.
- 타입 4: 공격자의 무작위 소스나 특정 소스를 알 수 없는 공격이 특정 대상을 공격하는 상황을 나타내며 일반적으로 DDoS 공격과 같은 경우가 이에 해당한다.
- 타입 5: 단일 소스로부터 무작위 대상으로의 특정 공격을 수행하는 것을 나타내며 일반적으로 특정 공격을 수행하기 이전에 공격자가 수행하는 스캐닝이 해당될 수 있다.
- 타입 6: 단일 소스로부터 무작위 대상으로의 특정 서비스에 대해 공격을 수행하는 것이다.
- 타입 7: 무작위 소스로부터 특정 대상으로의 특정 서비스에 대해 공격을 수행하는 것을 나타내며 일반적으로 DDoS 공격의 특징을 나타낸다.
- 타입 8: 특정 소스로부터 무작위 대상으로의 불

특정 공격에 해당한다.

• 타입 9: 무작위 소스로부터 특정 대상으로 불특정한 공격을 나타낸다.

• 타입 10: 무작위 소스로부터 무작위 대상으로의 특정 공격을 나타내며 일반적으로 웜의 특징을 나타낸다.

2. 취약성 평가 데이터 분석

보안 평가 데이터는 취약성 점검 도구로 취득할 수 있다. 본 논문에서 사용하는 취약성 평가 데이터를 상세히 기술하면 다음과 같다.

• 노드 IP(nodeIP): 취약성이 발견된 대상 노드의 IP 주소를 나타내고 <표 1>의 목적지 IP 필드와 상관관계 분석에 이용될 수 있다.

• CVE-ID: 취약성이 발견된 대상 노드가 가지고 있는 CVE(Common Vulnerability and Exposures) ID를 의미한다. 침입탐지시스템으로부터 생성되는 경고의 공격명 필드와 상관관계 분석에 이용한다. 하지만 모든 공격에 대한 CVE-ID가 부여되는 것이 아니므로 정확히 매치가 될 수 없으며 침입탐지시스템으로부터 생성되는 경고의 공격명 필드 또한 CVE-ID 이외에 자체적으로 정의한 구별자를 가지고 있다.

• 취약성 이름(vulName): 발견된 취약성 이름. 이 정보는 추가적으로 사용될 수 있는 다른 취약성 분석 도구와 상관관계 분석에 이용될 수 있을 뿐 아니라 관리자에게 보고할 수 있는 정보이다.

• 취약성 레벨(vulLevel): 취약성 평가 결과 발견된 취약성에 따라 취약 레벨이 나누어지며 높음(high), 중간(medium), 낮음(low)으로 평가한다.

• 포트 번호(portNum): 취약성이 발견된 서비스 포트 번호를 의미. 경고 데이터의 서비스 포트 필드 정보와 상관관계 분석에 이용한다.

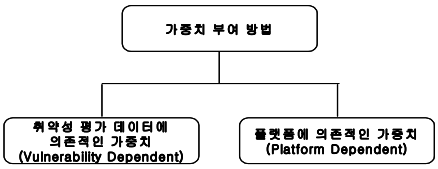
• 영향받는 운영체제(affectedPlatform): 취약성을 가지고 있는 운영체제의 종류를 나타내는 필드이다.

• 운영체제 비율(platformRatio) : 취약점 검사를 통해 해당 네트워크에 존재하는 운영체제의 종류 비율을 나타내는 필드이다. 가중치를 부여할 때 사용된다.

4. 경고 축약 알고리즘

경고 축약 알고리즘은 상기 기술한 경고 데이터와 취약점 평가 데이터의 상관관계를 이용한다. 경고 축약 알고리즘의 처리에 앞서 전처리 과정이 존재한다. 전처리 과정은 보안 장비에서 경고 메시지가 발생하였을 때 해당 경고에 대한 방화벽 룰을 검사하여 룰이 존재하면 해당 경고 메시지를 필터링하고 그렇지 않으면 경고 축약 알고리즘 처리를 수행한다. 이러한 검사는 경고에 대한 초기 대응 측면에서 아주 효율적이며 필터링 되는 경고에 대해서는 경고 축약을 수행하지 않기 때문에 시스템 구현 시 성능 측면에서 이득을 볼 수 있다.

또한 경고 축약 알고리즘을 통해 최종 경고로 결정된 경고 메시지에 대해서는 [그림 1]에서 보는 바와 같이 가중치를 부여한 후 관리자에게 보고한다.



[그림 1] 경고 가중치 종류

취약성 평가 데이터에 의존적인 가중치(VD)는 취약성 점검 도구가 제공하는 취약성 레벨을 나타낸다. 이 가중치는 크게 높음(high), 중간(medium), 낮음(low)으로 나누져 가중치를 부여한다.

플랫폼 의존적인 가중치(PD)는 식 (1)과 같이 부여된다.

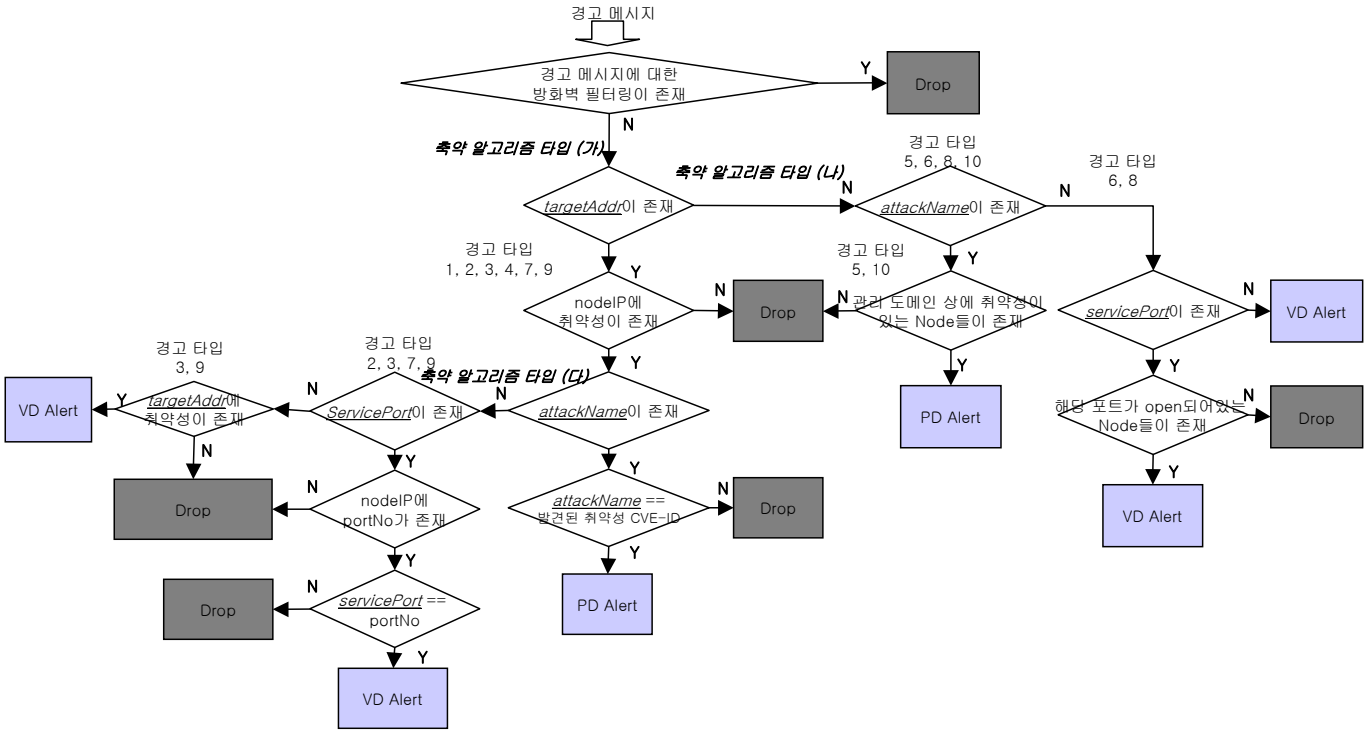
$$\frac{\text{해당공격에영향을받는운영체제의수}}{\text{전체시스템의수}} \square 100 \quad \text{식(1)}$$

식 (1)은 임의의 공격이 발생하였을 경우 관리 네트워크 내에서 해당 공격에 대해 영향을 받는 시스템의 비율을 나타낸다. 세분화된 경고는 관리자 및 대응 시스템의 세밀한 대응을 위해서 참고 데이터가 된다.

경고 축약 알고리즘은 [그림 2]에서 보는 바와 같이 편의상 크게 (가), (나), (다) 3가지 타입으로 나누었다.

경고 축약 타입 (가)는 경고 데이터에 목적지 IP 주소(targetIP)가 존재하는 경우와 그렇지 않은 경우로 나누었다. 따라서 타입 1, 2, 3, 4, 7, 9와 타입 5, 6, 8, 10으로 구분하여 단일 호스트를 대상으로 하는 공격인지 네트워크를 대상으로 하는 공격인지 판단한다. 경고 메시지 타입 1, 2, 3, 4, 7, 9에서 취약성 평가 데이터에 일치하는 노드의 IP가 존재하지 않으면 해당 경고를 필터링한다. 만약 존재할 경우는 경고 메시지에 공격 이름이 존재하는지의 여부를 검사하여 존재하면 취약성 평가 데이터의 CVE-ID와 동일 여부를 검사한 후 동일하면 관리자에게 플랫폼 의존적인 경고 메시지를 보낸다.

경고 축약 타입 (나)는 targetIP 필드가 존재하지 않는 5, 6, 8, 10을 대상으로 이루어진다. targetIP



[그림 2] 취약성 평가 데이터와의 상관관계를 이용한 경고 축약 알고리즘

필드가 없기 때문에 관리 도메인 상의 모든 호스트를 대상으로 한다. 경고 데이터 타입 6과 8은 targetIP와 attackName 필드가 존재하지 않기 때문에 취약성 평가에 포함되지 않는다. 따라서 다른 방법의 경고 축약이 필요하다. 경고 데이터에 sourceIP 필드가 없는 경고 데이터 타입 10인 경우, 방화벽에서 이러한 공격을 필터링 할 수 없기 때문에 경고를 알린다. 경고 데이터 타입 5는 sourceIP 필드가 존재하지만 방화벽에 필터링이 존재하지 않기 때문에 경고를 알린다. 경고 축약 타입 (나)에서 attackName이 존재하지 않는 경고 데이터 타입 6, 8은 serviceType 필드, 즉 포트에 대한 정보가 있는지 검사한다. 포트 정보를 얻을 수 있다면, 관리 도메인 상에 해당 포트가 개방되어 있는 노드의 존재 유무를 검사하게 된다. 만약 그러한 노드가 존재한다면 경고를 알리게 되지만, 그렇지 않다면 공격은 관리 도메인에 영향을 주지 않는 공격이므로 경고를 알리지 않는다.

경고 축약 타입 (다)는 경고 데이터에 targetIP가 존재할 경우에 대해서 상관관계를 분석한다. 이 경우 경고 데이터 타입 2, 3, 7, 9가 해당된다. 경고 데이터 3과 9는 경고 메시지의 서비스 포트 번호가 존재하지 않는 경우이므로 경고 메시지의 tergetIP에 대한 취약성이 하나라도 존재하는 경우는 경고 메시지를 관리자에 보고한다. 만약 서비스 포트 번호가 존재하면 취약성이 있는 노드의 IP 주소의 취약성 포트가 존재하는지를 검사한 후 존재하지 않으면 필터링하고 그렇지 않으면 경고 메시지와 비교하여 필터링 여부를 결정하게 된다.

5. 결론

본 논문은 기존 보안 장비의 한계점을 극복하기 위한 방법으로 취약성 평가 데이터와 보안 장비의 경고 데이터의 상관관계를 분석하여 경고를 축약하는 알고리즘을 제안하였다. 경고 축약하는 알고리즘은 취약성이 존재하지 않은 노드의 경고에 대해서는 관리자에게 보고하지 않으므로 부정확한 경고 메시지를 감소시킨다. 향후 제안한 알고리즘에 대한 성능 평가를 수행할 것이다.

참고문헌

[1] 정보홍외 2명, 『침입방지시스템 기술 현황 및 전망』, ITFIND 주간기술동향, 2003.

[2] Alfonso Valdes, Keith Skinner, "Probabilistic Alert Correlation", Workshop on Recent Advances in Intrusion Detection (RAID), Oct. 2001.

[3] Dan Andersson, Martin Fong, Alfonso Valdes, Heterogeneous Sensor Correlation: A Case Study of Live Traffic Analysis", IEEE Workshop on Information Assurance and Security, Jun. 2002.

[4] Frederic Cuppens, "Alert Correlation in a Cooperative Intrusion Detection Framework", Proceedings of the 2002 IEEE Symposium on Security and Privacy, May 2002.