김일성종합대학학보
(자연과학)
주체103(2014)년 제60권 제6호

JOURNAL OF **KIM IL SUNG** UNIVERSITY
(NATURAL SCIENCE)
Vol. 60 No. 6 JUCHE103(2014).

이상경보처리의 한가지 방법

공 혜 옥

콤퓨터체계를 파괴하고 정보를 훔쳐내는 해커들의 공격은 콤퓨터사용자들에게 커다란 피해를 주며 이러한 침입으로부터 체계나 정보를 보호하고 안전하게 관리하는 문제는 대단히 중요한 문제로 나선다.

지난 시기 이러한 침입을 막기 위한 침입검출체계(Intrusion Detection System: IDS)에 대한 연구가 많이 진행되였다.

침입검출은 콤퓨터체계나 그것이 리용하는 자원을 목표로 하는 의도적인 행위를 식별하고 처리하는 과정이며 침입검출체계는 이러한 과정을 실현하게 하는 체계이다.

IDS는 보통 망접속장치를 리용하여 망상에서 전송되는 모든 통신정보들을 실시간적으로 분석한다. 이때 패턴대조, 빈도수 및 턱값비교, 통계적이상검출 등의 수법들을 리용하여 공격을 검출하고 그 순간 경보를 울리며 대응책을 취한다. 대응책으로서는 관리자경보, 련결해제, 공격자료의 기록 및 분석 등이 속한다.

망보안체계에서는 분석자가 모든 경보를 검토하게 함으로써 의도적인 침입을 정확히 검출하려는 시도가 많아지고있다. 그런데 하루에 수감부로부터 생성되는 경보가 약 85만개정도이고 그중 심중한 경보가 약 1만8천개정도인 조건에서 홍수처럼 많은 경보를 분석자가 수동적으로 검토하기는 곤난하며 따라서 경보자료분석에 자료발굴의 수법들을 효과적으로 적용하여 분석자의 부담을 덜어주고있다.[1]

선행연구[1]에서는 사건들을 공격부류별로 분류하기 위하여 침입검출문제에 분류화수법을, 선행연구[2]에서는 망봉사의 정상적리용을 특징짓는데 분류화수법을 적용하였다.

론문에서는 현재 존재하는 침입검출의 방법을 교체하는것이 아니라 자료발굴수법을 추가적으로 적용하여 원래의 망방어체계의 성능을 향상시킬것을 목적으로 한다. 이로부터 선행연구[1, 2]에서처럼 직접적인 망접속자료가 아니라 IDS가 내보내는 망경보자료를 리용하여 이미 존재하는 특성량들을 보안하는데 자료발굴의 수법을 적용함으로써 사람의 검토를 요구하는 허위경보의 량을 감소시키게 한다.

수감부의 직접적인 자료가 아니라 일종의 간접적자료를 대상으로 하는 선행연구[3]에서는 분류화알고리듬을 리용하여 실시간－망감시도구에서 리용하는 규칙들을 갱신하였고 선행연구[4]에서는 련관규칙들로 수감부에서 생성되는 허위경보를 줄이였으며 선행연구[5]에서는 허위경보에 대한 통계적해석을 주었다.

론문에서는 IDS의 이상경보들을 일정한 기준에 따라 대응하는 공격들로 추가적으로 분류하도록 함으로써 공격과 관련되는 경보자료들은 분석자의 시야에서 사라지게 하여 분석자의 작업부담을 덜어준다. 또한 대응사건들로 분류된 경보들에 대해 점수를 주는 계산모형을 제기함으로써 비정상적인 행위의 실마리를 찾기 위한 분석자의 작업을 지원한다.

1. 경보분류의 절차

IDS의 이상경보들을 대응하는 공격들로 추가적으로 분류하기 위한 망보안체계의 모형은 그림 1과 같다.

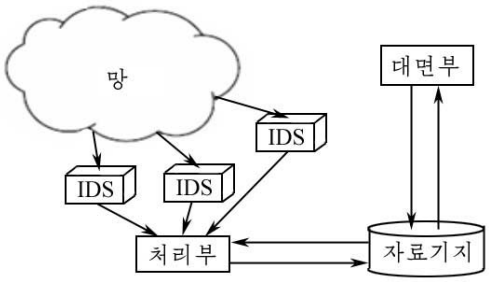

그림 1. 망보안체계의 구조

망통신은 여러 침입검출수감부에 의하여 분석되며 수감부의 자료는 모두 중심봉사기에서 처리되도록 관계형자료기지에 주기적으로 수집적재된다. 분석자들은 자료기지에 대한 Web봉사기말단의 대면부를 통하여 우연사건렬자료와 경보들을 조사한다.

망보안체계의 처리부는 자료의 적재, 수집, 려과, 분류, 자료분석을 지원하는 등급매기기 등을 진행하도록 그림 2와 같이 구성한다.

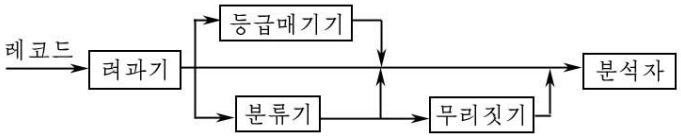

그림 2. 처리부의 구조

IDS로부터 수십만개의 경보가 처리부에 들어오면 처리부에서는 동일한 원천IP주소를 가지는 경보들끼리 묶어서 원천IP주소별로 분류된 수천개의 대응사건들을 수집한다. 려과기는 목적지IP주소들의 개수가 어떤 턱값보다 크면 이런 대응사건들을 공격으로 분류하며 등급매기기에서는 대응사건들에 등급을 매긴다. 대응사건의 부분으로 고찰되지 않은 경보들은 허위경보를 려과하는 분류기로 보내며 분류기에서 악의가 없는것으로 식별된 레코드들은 추가적검사를 위한 무리짓기로 보낸다.(그림 3)

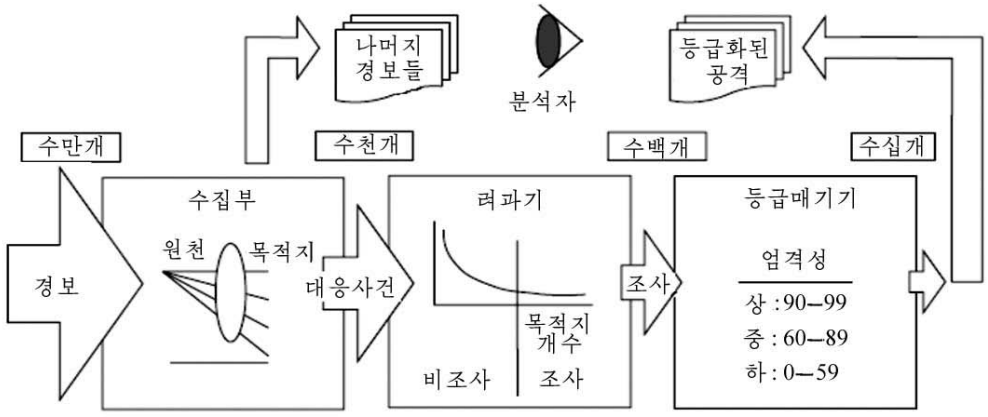

그림 3. 수집, 려과, 등급매기기의 과정

수집과정은 여러개의 수감부들로부터의 경보레코드묶음을 자료기지에 적재하는 과정이며 수집은 공통적인 원천지IP주소와 같은 단순한 수집기준에 따라 시간창문에서 발생하는 레코드들에 대하여 진행한다.

려과기를 통과한 수집자료들은 대응사건들로 표식이 붙으며 이와 관련한 레코드들은 분석자의 시야에서 사라진다. 즉 려과기는 명백한 대응사건들을 표식붙여 분석자의 작업부담을 덜어준다.

려과기에서는 수집사건의 목적지주소들의 개수가 령역전문가에 의해 설정된 일정한 턱값(실례로 99.5%)을 넘으면 그것을 하나의 대응사건으로 분류한다. 즉 대응사건이란 목적지주소들의 개수가 일정한 턱값을 넘는 동일한 원천IP주소를 가지는 경보들의 묶음이다.

이것은 하나의 원천IP가 짧은 시간동안에 비정상적으로 많은 목적지기계들과 통신한다면 그것과 관련된 레코드들을 종합하여 하나의 사건으로 간주한다는것이다.

등급매기기에서는 려과기에 의하여 대응사건으로 표식이 붙은 망수감부경보들에 대해 등급(점수)을 매기게 된다. 등급은 대응사건이 악의있는가 아니면 없는가를 잠정적으로 지적하며 분석자들이 본질적인 경보들에 더 주의를 돌릴수 있게 도와준다.

이 과정에 대응사건에서 비정상적인 행위의 실마리를 찾게 된다.

2. 경보분류를 위한 계산모형

앞에서 고찰한 경보분류절차의 등급매기기단계에서는 대응사건들로 분류된 망수감부경보들에 대해 점수를 생성하여 비정상적인 행위의 실마리를 찾게 된다.

여기서는 이러한 점수매기기를 위한 계산모형을 고찰한다.

점수는 피복성, 집중성, 독립성이라는 3가지 특성량들에 기초하여 이상사건들을 검출하는 수값으로 정의한다.

우리는 수감부에서 수집되여 표식이 붙은 개별적인 경보를 요소사건이라고 본다.

피복성 대응사건안에 있는 매개 경보가 목표로 하는 목적지IP들의 모임을 생각하자.

대응사건안의 서로 다른 요소사건이 그 대응사건에서 목표로 하는 목적지IP들을 모두 포함한다면 우리는 그 대응사건을 단순히 하나의 대응으로 예측한다.

그러나 어떤 특수한 요소사건이 목적지IP들의 적은 부분모임을 포함한다면 그것은 목적지의 적은 부분모임을 향한 공격에 초점을 두었다는것으로 본다.

이로부터 피복성은 $c=99\left(1-\min_{E}(n_{\mathrm{dstip}})/N_{\mathrm{dstip}}\right)$와 같이 계산한다. 여기서 E는 대응사건안의 요소사건들의 모임이며 $\min_{E}(n_{\mathrm{dstip}})$는 대응사건안의 어떤 요소사건에서의 서로 다른 목적지IP들의 개수의 최소값이고 N_{dstip}는 대응사건안의 목적지IP주소들의 총개수이다.

피복성은 대응사건의 목적지IP들의 모임의 대단히 적은 부분모임을 목표로 하는 요소사건들을 검출한다. 피복성은 대응사건안에서 요소사건에 대한 제일 적게 분포된 목적지주소들의 비률에 거꿀비례한다.

실례로 Scan Proxy로 표식된 1 000개의 요소사건들이 존재하는데 그것들은 1 000개까지의 개별적인 목적지IP들을 가진다고 하자. 또한 WEB MISC로 표식된 2개의 요소사건들이 존재하는데 그것들은 1 000개중에서 2개의 목적지IP를 가진다고 하자.

그리고 이 요소사건들이 모두 동일한 원천지IP를 가진다고 하면 $\min_{E}(n_{\mathrm{dstip}})=2$, $N_{\mathrm{dstip}}=1\,000$이고 따라서 $c=98.8$이다. WEB MISC사건들은 Scan Proxy조사에 의해 피복되는 IP공간을 피복하지 못하므로 의심스러운것으로 된다.

집중성 우리는 악의없는 대응사건에서 경보들의 개수는 목적지IP들의 전반에 대해 균등하게 분포될것이라고 본다.

만일 특수한 목표가 집중적이라면 하나의 주콤퓨터에로 집중적으로 향한 초점은 그것이 공격을 받았다는것을 의미한다.

이로부터 집중성은 $p=\begin{cases}5\nu-1, & \nu<20\\ 99, & 기타\end{cases}$, $\nu=\max_{D}(n_r)\big/\overline{n}_d$ 와 같이 계산한다. 여기서 D는 대응사건안의 목적지IP주소들의 모임이며 $\max_{D}(n_r)$ 는 대응사건안의 어떤 목적지IP에 관련되는 레코드들의 최대개수이고 $\overline{n}_d$ 는 대응사건안에서 매개 목적지IP에 관련되는 레코드들의 평균개수이다.

집중성은 불균형적으로 많은 개수의 레코드들로 목적지IP들을 목표하는 대응사건을 검출한다. 집중성은 목적지주소당 평균적인 요소사건개수에 대한 목적지주소당 최대요소사건개수의 비에 비례한다.

실례로 Scan Proxy로 표식된 1 000개의 요소사건들이 존재하는데 그것들은 1 000개까지의 개별적인 목적지IP들을 가지며 그 모임안의 어떤 목적지IP에로 여러 형태의 요소사건 100개가 추가된다고 하자. 그리고 그것들은 모두 동일한 원천지IP를 가진다고 하자.

그러면 $\max_{D}(n_r)=101$, $\overline{n}_d=1.1$ 이며 따라서 $\nu=9.18$, $p=99$ 이다.

하나의 IP에로 향한 추가적요소사건들은 하나의 특정한 주콤퓨터를 목표로 하면서 그 주콤퓨터의 원천지로부터의 보다 높은 관심을 보여주고있기때문에 의심스러운것으로 된다.

독립성 대응사건안에는 매 요소사건과 관련되는 많은 레코드들이 있다.

우리는 악의없는 대응사건에서 서로 다른 요소사건에 대한 관련레코드개수가 거의 같을것이라고 본다. 만일 대응사건안의 어떤 특수한 요소사건이 요소사건당 레코드의 평균개수에 비하여 아주 적은 개수의 관련레코드들을 가진다면 그것은 어떤 독립적인 요소사건 혹은 어떤 추가적인 행위가 대응사건에 함께 포함되여있다는것을 나타낼수 있다.

이로부터 독립성은 $u=99\left(1-\min_{E}(n_r)\big/N_r\right)$ 와 같이 계산된다. $\min_{E}(n_r)$ 는 대응사건안에서 요소사건에 대한 레코드들의 개수의 최소값이며 N_r 는 대응사건안에 있는 레코드들의 총개수이다.

독립성은 대응사건안에 어떤 요소사건과 관련되는 레코드들이 거의나 존재하지 않을 때 그러한 대응사건을 검출한다. 독립성은 대응사건안에서 가장 적게 리용된 요소사건과 관련되는 레코드들의 비률에 거꿀비례한다.

실례로 Scan Proxy로 표식된 1 000개의 요소사건들이 존재하는데 그것들은 1 000개까지의 개별적인 목적지IP들을 가진다고 하자. 그리고 WEB MISC로 표식된 1개의 요소사건, 그것은 1개의 목적지IP를 가진다고 하자.

그러면 $\min_{E}(n_r)=1$, $N_r=1\,001$ 이며 따라서 $u=98.9$ 이다. WEB MISC요소사건은 오직 1개의 경보만을 가지므로 의심스러운것으로 된다. 반면에 보통 대응사건에서 요소사건들은 보통 대응사건안에 있는 목적지IP들의 대부분 혹은 모두를 지적한다.

등급매기기 먼저 우의 3가지 특성량들에 대한 스칼라함수 $m'=1-\alpha\ln(1+m/100)$ 을 고찰하자. 여기서 m 은 특성량(c, p, u 중의 하나)이고 m' 는 스칼라화된 특성량이며 α 는 무게인자이다. 즉 특성량의 스칼라함수는 특성량이 감소할 때 예민하게 증가하는 반전함수이다.

모의실험을 통하여 우리는 집중성과 독립성은 $\alpha \approx 0.1$로서, 피복성은 $\alpha \approx 0.2$로서 무게화하여 피복성을 강조하고있다.

등급매기기에서는 피복성, 집중성, 독립성의 관측값들을 대응사건을 평가하는 하나의 점수 s로 결합한다. 결합은 여러가지 방법으로 할수 있다.

우리는 단순한 공식 $s=100(1-c'p'u')$를 리용한다. 여기서 c', p', u'는 피복성, 집중성, 독립성에 대한 특성량들의 스칼라함수들이다.

결국 s는 특성량이 클수록 더 큰 값을 가지게 된다.

s의 값에 따라 공격의 등급을 상(90－99), 중(60－89), 하(0－59)로 구분할수 있으며 그것에 따르는 적중한 대응책을 취할수 있다. 또한 공격으로 려과되지 않은 경보들은 허위경보의 분류기로 보내여 해당한 처리를 하도록 할수 있다.

맺 는 말

론문에서는 IDS의 이상경보들을 대응하는 공격들로 추가적으로 분류하기 위한 절차적 모형과 함께 대응사건들로 분류된 망수감부경보들에 대한 점수계산모형을 제기함으로써 기존의 망수감부에서 생성되는 허위경보의 수를 감소시키고 새로운 공격에 대응되는 비정상적경보사건들을 검출하도록 하였다. 실험결과 분석자의 검토를 요구하는 5일동안의 71 094개의 심중한 경보자료가 1 011개로 감소되였다.

론문에서 고찰한 경보처리방법은 이상검출체계뿐아니라 서명검출체계에도 적용할수 있으며 망기반 및 주콤퓨터기반의 IDS에 다 적용할수 있다.

참 고 문 헌

[1] C. Elkan; Newsletter of the ACM Special Interest Group on Knowledge Discovery and Data Mining **1**, Kluwer Academic Publication, 63～64, 2000.

[2] W. Lee et al.; In Proceedings of the 7th USENIX Security Symposium, 79～94, 1998.

[3] W. Lee et al.; In Proceedings of the 1st Workshop on Intrusion Detection and Network Monitoring, 5～14, 1999.

[4] S. Manganaris et al.; Computer Networks, **34**, 571, 2000.

[5] Cheng Yuan Ho et al.; IEEE Communications Magazine, **3**, 146, 2012.

주체103(2014)년 2월 5일 원고접수

A Method Processing Anomaly Alerts

Kong Hye Ok

This paper gives a description how to use data mining methods to improve the processing of network alerts. We present a computational model for the score to detect unusual alert events that may correspond to new attacks.

Key word: network alert