베이시안 기법을 이용한 다중 공격판단 시스템

안재호
고려대학교 컴퓨터정보통신대학원
axlno5@daum.net

Intrusion Detection System using Baysian Approach

JAE-HO AHN

Korea Graduate School of Computer Information and Communications
Dept of Computer Engineering, Korea University
axlno5@daum.net

요 약

보안위협은 날로 정교해지고 증가하고 있다. 이에 대응하는 인력과 정보보호인프라는 여러가지 한계점이 있다. 사람이 모든 걸 분석하기에는 그 양이, 보안인프라를 맹목적으로 신뢰하기에는 그 정확도가 문제가 된다. 이에 베이시안 기법을 이용하여 단편적인 분석이 아닌 여러 보안인프라의 오탐율과 상관관계를 고려한 공격판단 시스템을 구현하여 각 보안현상에 대한 공격여부를 판단함으로써 방대한 양과 정확도를 높이는 공격판단 시스템을 제안한다.

1. 서론

대부분의 기관이나 회사등 네트워크 환경을 운영하는 곳은 이기종의 보안인프라가 도입되어 있는 경우가 많다 대표적으로 방화벽(Firewall), IDS(Intrusion Detection System), IPS(Intrusion Prevention System) 등 이기종 보안장비들이 도입되어 운영중에 있으며 ESM (Enterprise Security Management) 등을 통하여 이기종 보안장비 실시간 로그를 상관분석(Correlation Analysis)하여 대응하는 곳도 적지 않다 이러한 일반화된 보안환경은 크게 두 가지의 문제점이 존재한다. 첫째, 정확도이다 각 보안장비의 공격탐지로그를 얼마나 신뢰할 수 있는가? 이다 실례로 대부분의 기관이나 회사에 도입되어 있는 IDS 같은 경우 오탐로그가 대부분인 경우가 많다. 이러한 오탐(False Detection) 및 오경보(False Alarm)는 보안전문가에 의하여 분석되고 바로잡아져야 하지만 현실적으로 그러한 공격분석에 능통한 보안전문가가 많은 양의 로그를 일일이 분석하기란 거의 불가능에 가까우며, 대부분의 기관 및 회사의 보안담당자의 경우 전문적인 보안지식이 없는 경우가 다반사다. 둘째 그 방대한 양이다. 매일 보안장비 별로 수만 개, 수십만 개가 발생하는 방대한 양을 어떻게 효과적으로 분석하고 대응할 것인가? 이다 단 이경우도 보안전문가를 보유한 기관이나 회사의 경우에 한한다. ESM 등의 통합보안관제솔루션을 도입하여 상관분석 하는 경우도 있으나 역시 그 양이 문제가 된다. 관제요원들이 일일이 분석하여 대응하기에는 그 양이 너무 많은 실정이다. 이러한 문제점을 해결하기위해 본 논문은 베이시안(Bayesian) 기법을 이용한 종합적이고 공격판단 시스템을 제안한다. 기존 이기종 보안로그의 통합 및 상관분석에 관한 논문 들은 대부분 그 방법에 초점이 맞추어져 있다. 결국에 이는 사람이 분석을 수행해야한다는 결론에 도달하게 된다. 이는 결국 방대한 양에 발목을 잡힌다. 또한 여러가지 방법을 이용한 침입탐지에 관한 논문은 대부분 단편적인 IDS에 대한 오탐율을 줄이는데 초점이 맞추어져 있다 그러나 현실적으로 최근의 공격방법은 단편적인 한 개의 보안장비를 가지고 판단하기에는 무리가 있다. 이러한 문제점들을 본 논문에서는 사람의 노력을 최소화하고 보다 정확하고, 포괄적으로 공격현상을 분석할 수 있는 시스템으로 해결하고자 한다.

2. 관련분야 연구

서론에 언급했듯이 머신런닝(Machine Learning) 기법을 이용한 침입탐지 방법은 단편적인 하나의 침입탐지시스템의 오탐율을 줄이는데 초점이 맞추어져 있다. 이는 본 논문의 주제와는 다소 차이가 있으나 그 원리는 비슷한 부분이 있어 기존 연구에 포함하였다. 본 논문의 공격판단 시스템은 단편적인 하나의 보안장비의 로그를 분석하여 탐지 하는 것이 아니라 대규모 네트워크의 구현되어 있는 여러 이기종 보안장비의 공격로그의 종합적인 분석을 통하여 사람의 개입을 최소화 하는데 그 목적이 있다.

2.1 N-gram 기법

프로그램 행위 기반 침입 탐지 기법의 전제는 대부분의 공격은 프로그램 결함이나 버그로 인하여 발생할 수 있으며 프로그램의 정상적인 사용과는 그 행위가 다르다는데 있다. 그러므로, 프로그램의 행위가 적합하게 표현될 수 있다면 침입 탐지를 위한 행위 특성으로 활용될 수 있다.

N-gram 기법은 프로그램의 정상행위를 자동적으로 추출하고 정의하기 위한 대표적인 기법이다. 대부분의 경우 모든 감사 로그들은 각 어플리케이션으로부터 요구되어진 객체나 시스템호출(System Call)에 있어 사전 진단이 필요로 한다. N-gram 기법은 프로그램에 의해 발생되는 시스템 호출들을 순차적으로 고정 길이로 분할하고 정상행위로 간주하여 프로파일을 구축한다. 만약, 임의의 순차적인 시스템 호출이 프로파일에 존재하지 않는다면 이상 행위로 간주한다. N-gram 기법은 단순한 알고리즘과 높은 탐지율을 보이지만, 프로파일 데이터의 크기 및 오버헤드가 매우 크다는 단점을 갖고 있다.

2.2 Bayesian Network 기반의 변형된 침입패턴 분류 기법

다중 서열정렬(Multiple sequence alignment)에 의해서 확장된 Bayesian Network를 구축함으로써, 각각 시스템호출의 매칭에 의한 이상탐지에서 탈피하여, 시스템호출간의 관련성에 의한 프로그램 행위 수준의 이상탐지가 가능하게 되며, 어플리케이션 행위를 프로파일링하여 변형된 이상침입을 분류하는 장점을 같고 있다.

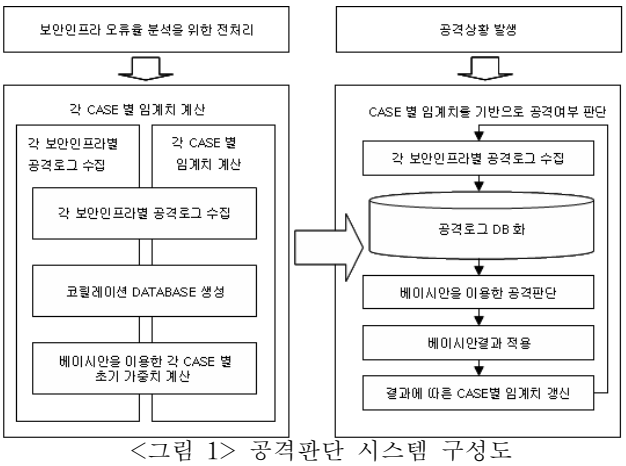
여기에 소개한 몇가지 방법들은 대부분 하나의 단편적인 침입탐지시스템(IDS) 에 적용되어 지며 종합적인 분석을 요구하는 근래의 여러 보안인프라로 구성되어지는 광범위한 네트워크 시스템에 적용하기에는 한계점이 있다.

3. 공격판단 시스템

앞에 2장에서 설명한 Bayesian 정리를 토대로 여기서는 학습을 통한 공격판단 시스템의 설계와 구현과정을 설명한다. 본 논문에서는 이기종 로그의 분석을 통한 일련의 행위가 공격인지 아닌지를 판단하는 시스템으로 한정하였다. Data-Mining을 통한 이기종 보안인프라의 공격 로그를 기준으로 Bayesian 알고리즘을 적용하여 일련의 행위에 대해 각 보안인프라의 공격탐지결과를 기준으로 Bayesian을 적용하여 공격여부를 판단한다.

3.1 시스템 구조

공격판단 시스템은 기본적으로 침입탐지시스템, 침입방지시스템, 바이러스월, 웹방화벽등 이기종 보안인프라에서 제공하는 차단로그를 기반으로 동작한다. 이러한 로그를 종합하여 하나의 현상에 주목하며 이 현상이 실제로 악의(malice)적인 공격인지를 판별하게 된다. 1차적으로 공격판단 시스템은 이기종 보안장비의 차단로그를 수집하여 정규화 한다. 이를 기반으로 하나의 Correlation 데이터베이스를 생성하게 된다. 이 데이터베이스는 일련의 현상에 대한 각 보안장비의 차단(Deny) 및 허용(Accept) 여부를 기록하게 된다. 2차적으로 전처리기를 통하여 각 보안인프라별 초기 수동분석결과에 기초하여 각 case 별 임계치 값을 계산한다. 이렇게 계산된 초기 임계치를 기준으로 공격여부를 판단하며 이 결과값을 다시 공격판단시스템에 적용함으로서 시스템의 신뢰성을 높인다. Correlation 데이터베이스를 기반으로 위와 같은 공격판단 알고리즘을 적용하여 특정행위에 대한 악의적인 공격시도를 판단하여 공격자를 네트워크에서 차단한다.



<그림 1> 공격판단 시스템 구성도

3.2 전처리기

전처리기 에서는 이기종 보안인프라의 공격로그로 구성된 Correlation 데이터베이스를 기반으로 하여 각 case 별 초기 임계치를 계산한다. 이때 Bayesian 정리를 사용한다.
Pr(A|B)=Pr(B|A) Pr(A) / Pr(B) (Conditional Probability)
P(B|A) = P(B^A) / P(A)
위의 식에서 임계치 값을 얻어 Bayesian 정리를 사용하여 공격유무를 결정할 수 있다.위 식을 이용하여 각 case 별 초기 임계치를 계산한다. 본 논문에서는 7개의 보안인프라로 구성되어 있는 환경을 기반으로 하였다.
먼저 각 보안인프라별 오탐율을 반영하기 위하여 수동분석이 필요하다. 각 보안인프라가 공격이라고 리포팅한 정보가 실제공격인지에 대한 수동분석으로 통하여 [표1]과 같은 결과를 얻었고 이를 기반으로 Bayesian 정리를 적용하여 각 case별 임계치 값을 계산한다.

	보안인프라S1	보안인프라S2	보안인프라S3	보안인프라S4	보안인프라S5	보안인프라S6	보안인프라S7
공격	102	91	234	195	306	143	254
공격의심	212	143	297	396	378	312	412

<표1 초기 수동분석 결과>

각 보안인프라에서 리포팅한 일정수의 공격 로그를 수동분석하여 Bayesian에 반영함으로써 각 보안인프라의 오탐에 대한 본 공격판단 시스템의 내성을 만든다.
각 case별 초기임계치는 P(A|P1,P2,P3,P4,P5,P6,P7) 과 P(NA|P1,P2,P3,P4,P5,P6,P7) 을 비교하여 얻을 수 있으며, 가령 P1을 제외한 모든 보안인프라가 공격을 탐지(detect) 못했다면 아래와 같은 식으로 그 값을 얻을 수 있다.

P(A|S1,S2,S3,S4,S5,S6,S7)
=P(A)P(S1|A)P(notS2|A)P(notS3|A)P(notS4|A)P(notS5|A)P(notS6|A)P(notS7|A)
=(1216/2150)*(101/1216)*(1148/1216)*(1068/1216)*(1018/1216)*(1024/1216)*(1055/1216)*(868/1216)=0.0172

P(NA|S1,S2,S3,S4,S5,S6,S7)
=P(NA)P(S1|NA)P(notS2|NA)P(notS3|NA)P(notS4|NA)P(notS5|NA)P(notS6|NA)P(notS7|NA)
=(934/2150)*(111/934)*(859/934)*(785/934)*(736/934)*(748/934)*(783/934)*(870/934)=0.0197
0.01700 < 0.01966 일 때, 해당로그 분석결과는 공격상황이라 할 수 없다. 보안인프라 7개를 기준으로 총 127가지의 case가 발생할 수 있고, 각각의 case 에 대하여 Bayesian 정리를 사용해 임계치를 계산한다. 이를 기반으로 공격판단을 수행한다.

CASE	보안인프라	Attack	Non Attack	차이값	값
1	S1	0.017006484	0.019666474	-0.00266	<
2	S2	0.011120775	0.012731262	-0.0016105	<
3	S3	0.026017076	0.027677061	-0.00166	<
4	S4	0.036516188	0.03922751	-0.0027113	<
		.			
124	S1,S2,S4,S5,S6,S7	2.24767E-06	1.62956E-06	6.1812E-07	>
125	S1,S3,S4,S5,S6,S7	5.25844E-06	3.54256E-06	1.7159E-06	>
126	S2,S3,S4,S5,S6,S7	3.43856E-06	2.29331E-06	1.1453E-06	>
127	S1,S2,S3,S4,S5,S6,S7	3.11475E-07	3.09304E-07	2.1711E-09	>

<표2 각 CASE별 초기 임계치 값>

7개의 보안인프라에서 탐지(detect)한 로그를 기반으로 공격판단을 수행하여 각 case 별로 임계치값을 비교하여 공격여부를 결정한다. 일단 결정된 결과를 다시 Bayesian에 포함시킬지 결정하고, 그 결과에 따라 임계치 값이 변경되는 시스템을 시스템1, 그렇지 않은 시스템을 시스템2라 하였다. 또한 두 시스템 모두 주기적인 수동분석 로그를 Bayesian에 적용 하였다.

4. 실험

본 실험에는 실질적인 보안인프라에서 발생되는 로그가 사용 되었으며 각 보안 인프라 현황은 표3 과 같은 보안장비들이 사용되었다. 기본적으로 코릴레이션 데이터베이스는 실험환경 자체에 구성되어 있는 ESM의 데이터베이스를 이용 하였다.

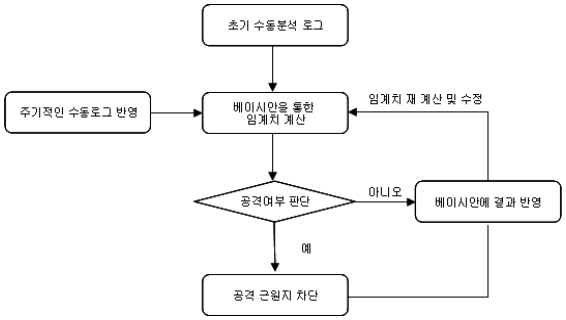
보안인프라1	보안인프라2	보안인프라3	보안인프라4	보안인프라5	보안인프라6	보안인프라7
방화벽	방화벽	IPS	바이러스월	IDS1	IDS2	웹방화벽

<표3 보안인프라 현황>

실험 네트워크 환경은 7개의 이기종 보안인프라를 기준으로 실험을 진행 하였다. 모든 보안인프라는 현재 기구축되어져 있는 ESM을 통하여 공격로그를 수집하고 있으며, ESM 데이터베이스를 코릴레이션 데이터베이스로 이용하였다. 약 3000의 공격판단을 수행하였으며, 해당 실험을 통하여 얼마나 정확한 공격판단을 수행 하는지, 공격판단 결과를 Bayesian에 다시 적용하여 계산하는 것과 그렇지 않은 경우의 정확도 및 오탐율(False-Positive) 비교에 중점을 두었다.

4.1 두 개의 공격판단 시스템

본 실험에서는 두 개의 시스템을 구성하여 비교 실험을 진행 하였다. 그 이유는 Bayesian으로 공격판단을 수행한 결과를 다시 Bayesian에 적용할지에 대한 여부를 판단하기 위해서 이다. 시스템1은 공격판단 결과를 다시 Bayesian에 적용 하였고, 시스템2는 결과를 다시 적용하지 않았다. 또한 두 시스템 모두 주기적으로 일정량의 수동분석 데이터를 적용 하였다. 실험은 아래의 그림과 Flowchart로 진행 되었다.



<그림2 실험 진행 차트>

먼저 1000개의 ESM 로그를 기준으로 수동분석을 실시하였다. 그 결과 위에 표1 과 같은 결과를 얻었고 이를 기반으로 Bayesian을 적용하여 임계치 값을 계산 하였다. 그 결과를 기반으로 공격여부를 판단하며 시스템1에 경우는 해당 결과를 다시 Bayesian에 적용하였으며, 시스템2는 적용하지 않았다. 또한 두 시스템 모두 주기적인 수동로그 분석 결과를 양 시스템에 적용하였다. 최초 1000개의 수동분석 결과를 기반으로 200번의 공격판단을 수행하고 50개의 수동분석 로그를 Bayesian에 적용하고 다시 200번의 공격판단 수행.. 이런 형태로 3 set 를 진행 했으며, 시스템1 같은 경우에는 200번의 공격판단 결과를 다시 Bayesian에 적용하였고, 시스템2 는 적용하지 않았다.

4.2 실험결과

3 set 실험을 진행한 결과 아래와 같은 결과를 얻었으며, 두 시스템 모두 공격 실제공격 이였으나 탐지하지 못한 로그가 있다. 수동분석 결과 대부분 DoS(Denial of

Service) 공격종류 었으며, 현 보안인프라 대부분이 DoS(Denial of Service) 공격을 탐지하고 방어하는데 한계점을 가지고 있음으로 인하여 나온 결과로 판단된다. 추후 DoS(Denial of Service) 공격관련 보안인프라가 추가적으로 운영된다면 나아지리라 예상된다. 실험결과상 차이점을 보이는 부분은 오탐(False-Positive)로그인데, 시스템1 이 시스템2에 비해 상대적으로 오탐율이 적었다. 시스템2 같은 경우는 공격이 아닌 상황을 공격으로 판단하는 사례가 시스템1보다 매우 많았다.

	1차 분석			2차 분석			3차 분석		
	시스템1	시스템2	수동분석결과	시스템1	시스템2	수동분석결과	시스템1	시스템2	수동분석결과
탐지	39	55	41	24	49	26	29	47	32
정탐	35	35	41	22	22	26	27	27	32
미탐지	6	6	0	4	4	0	5	5	0
오탐	4	20	0	2	27	0	2	20	0

탐지로그 : 각시스템에서 공격이라고 판단한 로그
정탐로그 : 해당 로그를 수동분석한 결과 실제 공격이였던 로그
미탐지로그 : 실제공격상황이나 시스템에서 탐지하지 못한 경우
오탐로그 : 공격이 아닌 상황을 시스템에서 공격이라고 판단한 로그

<표4 실험결과>

결론적으로 두 시스템은 비슷한 수준의 정확도를 보였으나 오탐율 부분에서 시스템1이 다소 우수한 결과를 얻었다. 시스템1과 시스템2 모두 정확도(는 거의 동일한 결과를 얻었다. 이는 DoS 공격등 보안인프라에서 탐지하지 못하는 공격상황을 제외하고 시스템1이 탐지한 공격상황과 시스템2가 탐지한 공격상황은 거의 동일했다. 하지만 오탐율의 경우는 약간의 차이를 보였는데 시스템1 같은 경우 상대적으로 시스템2에 비해 오탐이 적었다. 즉 공격이 아닌 상황을 공격으로 인식하는 부분에서 시스템2가 좀 더 많은 오류를 범했다고 볼 수 있다. 이는 기본적으로 정확한 데이터를 공급받는 시스템2가 더 오탐율이 낮아야 하지만, Bayesian의 확률적인 관점에서 볼 때 85%의 정확도가 넘는 더 많은 데이터를 공급받는 시스템1이 좀 더 낮은 오탐율을 가진 것으로 분석할 수 있다. 물론 시스템 2의 경우에도 시스템1 보다 많은 정확도 100%의 수동분석 데이터를 Bayesian에 적용한다면 당연히 시스템2가 오탐율이 낮을 것이다. 하지만 방대한 공격로그의 분석이라는 관점에서 볼 때 시스템1의 손을 들어주게 된다. 오탐율 추이를 보더라도 시스템1같은 경우 3set를 진행하는 동안 점점 줄어들어 추후 많은 분석이 반복되면 좀 더 줄어들 가능성이 보였다. 하지만 시스템2 같은 경우 오탐율이 늘었다 줄었다 를 반복하여 명확하게 줄어들 것이라는 결론을 내리기 또한 본 논문의 취지 역시 사람의 노력을 최소화 하며, 공격상황을 분석하는 것이기에 시스템1을 적용하고 부분적인 수동분석을 가미 한다면 가장 좋은 결과를 얻을 수 있으리라 본다.

	bayes 시스템1	bayes 시스템2	수동분석 결과
정확도	85%	85%	100%
오탐율	9%	44%	0%

<표5 두시스템의 정확도 및 오탐율 비교>

5. 결론

본 논문에서 구현한 공격판단 시스템을 통하여 많은 양의 이기종 보안인프라별 자동화된 공격분석(Intrusion Analysis)이 가능하며, 데이터가 많아지면 많아질수록 점점 더 신뢰성 있는 시스템으로 발전한다. 실험을 통하여 일정량의 공격 로그들을 수동 분석한 결과와 본 공격판단 시스템에서 분석한 결과를 비교하여 높은 수준의 정확도를 얻을 수 있었으며, 시간이 지나고, 데이터가 많아질수록 그 수치는 점점 더 증가한다. 또한 본 논문의 공격판단 시스템은 초기 수동분석 단계에서 각 보안인프라별 오탐율을 반영한 시스템이기 때문에 각각의 보안인프라에 대한 오탐율과 오동작에 대한 내성이 있다. 이를 기구축된 보안관제등 각종 보안업무에 적용한다면, 그동안 수작업으로 수행하던 많은 부분을 시스템화 할 수 있다. 즉 사람의 노력을 최소화 하며 순도 높은 공격판단을 통하여 사이버 침해에 능동적으로 대응하고 많은 업무부하를 줄일 수 있으리라 본다. 하지만 역시 초기 수동분석 시에 분석 로그양과 사람의 분석능력에 의존도가 높은 게 사실이다. 또한 DoS공격 등 현존하는 보안인프라로 탐지하지 못하는 공격에 대한 탐지 율이 저조한 부분도 문제점으로 보인다. 이는 향후 전문적인 서비스 및 솔루션 개발 등으로 풀어야할 숙제로 남는다.

참고문헌

[1] E. Gyftodimos and P. A. Flach, "Hierarchical Bayesian network: A probabilistic reasoning model for structured domains,"

[2] G.N. Wang, "An Adaptive Hybrid Neural Network Approach for Learning Non-stationary Manufacturing Processes", Ph. D. Dissertation, Texas A&M Univ., 1993.

[3] Sreven L. Scott, "A Bayesian Paradigm for Designing Intrusion Detection Systems and Data Analysis", June 20, 2002.

[4] Marco Pagni, "Introduction to Patterns, Profiles and Gidden Markov Models", Swiss Institute of Bioinformatics(SIB), August 30, 2002.

[5] Mehdi Nassehi, "Characterizing Masqueraders for Intrusion Detection, Computer Science", Mathematics, 1998.