计算机及其应用

属性数据融合算法在分布式入侵检测系统中的应用研究

潘晓君,张佑春

摘　要:随着网络的快速发展,网络的安全问题日益成为人们研究的一个热点,受到人们的广泛关注.该文首先介绍了分布式入侵检测及数据融合技术,然后提出了一种基于属性数据融合的分布式入侵检测技术,对网络数据攻击的防护进行性能的测试.实验结果表明,该方案在检测率、误报率、融合比等方面都表现出良好的性能,在网络安全防护中具有很好的应用前景.

关键词:分布式入侵检测;属性数据融合;网络安全

中图分类号:TP393.08　　**文献标识码**:A　　**文章编号**:1008-7974(2017)05-0078-04

DOI:10.13877/j.cnki.cn22-1284.2017.10.021

1　分布式入侵检测系统及数据融合概述

分布式入侵检测系统(Distributed Intrusion Detection System,DIDS)是一种分布于网络环境的入侵检测系统,在入侵检测基本框架上引入了分布处理、分层次过滤及管理等策略,将多个具有类似传感器的代理分布在网络中的主要关键点上,通过一定的体系结构,相互协作形成一个统一的有机整体,从而能够及时了解网络整体的运行状况,并进行相关的数据检测和报警,可以大大提高网络的安全性[1-2].

数据融合是一个多级别、多层次的数据处理过程,它能对来自多个信息源的数据和信息进行检测、关联、估计及综合处理,进而得到比较精确的属性估计和状态估计,以及完整和及时的威胁估计和态势评估[3-5].通过将数据融合技术应用于分布式入侵检测系统,可以很好地把来自多个异质分布式传感器的各种信息和数据综合输入到一个统一系统进行相关的处理,大大减少了告警,降低了误报率,提高了网络的安全性[6].

2　入侵检测系统的架构

入侵检测系统架构如图1所示,该体系主要由数据采集及处理、数据训练、属性规则关联及提取、日志系统等几个模块组成,并对这些模块的功能进行相关的分析.该系统架构不但确保了系统的实时性和可扩展性,而且也避免了网络中的单点失效的问题,提高了系统的鲁棒性[7-8].

收稿日期:2017-05-23

基金项目:安徽高校自然科学研究重点项目(KJ2017A761);安徽高校自然科学研究重点项目(KJ2016A082);安徽省高校优秀青年人才支持计划重点项目(gxyqZD2016438).

作者简介:潘晓君,江西九江人,安徽工商职业学院电子信息系讲师;张佑春,安徽工商职业学院教师(安徽 合肥 231100).

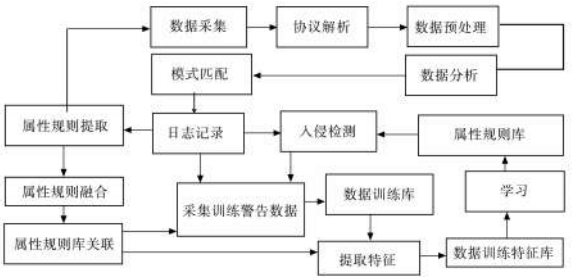

图1　入侵检测系统结构

3　属性数据融合算法的设计

3.1　属性数据报警模型

针对分布式的入侵检测系统,本文的属性数据融合约束函数在综合考虑多种相关因素的情况下,可以对评判的目标给出一个非常客观的评价,可以非常好地处理多因素中数据融合结果的不确定性,因此,利用此方法来进行报警数据的融合处理.

建立的数学模型如下:

设定约束函数评判集 $V=\{v_1, v_2, \cdots, v_n\}$,因素集 $U=\{u_1, u_2, \cdots, u_n\}$.

因素评判关系: f: $U\rightarrow I(V)$;

映射关系: $u_i \rightarrow f(u_i)=\{r_{i1}, r_{i2}, \cdots, r_{in}\}\in I(V)$;

由映射关系引出关系 $R_f \in I(U\times V)$,也就是 $R_f(U_i, V_j)=f(u_i)(V_j)=r_{ij}$;

则 R_f 用矩阵 R 表示为:

$$R=\begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1n} \\ r_{21} & r_{22} & \cdots & r_{1n} \\ \cdots & \cdots & \cdots & \cdots \\ r_{m1} & r_{m2} & \cdots & r_{mn} \end{bmatrix} \tag{1}$$

基于数据融合的报警参数主要有:报警时间(alert_time)、源 IP(Source_ip)和报警特征(alert_signature).在报警域中,当前产生的报警假定为 v_0 ,报警评判集用 V 表示, $V=\{v_1, v_2, \cdots, v_n\}$ 是当前所有数据报警中最新的报警集合.

使用约束函数来确定矩阵 R 中的参数.在确定约束函数时,有如下特性:如果两次报警时间所发生的时间间隔越短,则越有可能属于相同的一次攻击;当两个融合报警的源地址相同程度越高则越有可能属于相同的一次攻击;两次融合报警的数据特征越相同越有可能属于相同的一次攻击.

报警时间约束函数为:

$$r_{1j}=\begin{cases} 0, & T_j>60\,\text{min} \\ \dfrac{60-T_j}{50}, & 10\,\text{min}\le T_j\le 60\,\text{min} \\ 1, & T_j<10\,\text{min} \end{cases} \tag{2}$$

在这里, $T_j=v_{0\cdot\text{alert\_time}}-v_{j\cdot\text{alert\_time}}$,其中, $j=1,2,\cdots,n$.

报警融合数据源IP的约束函数为:

$$r_{2j}=\frac{h}{32} \tag{3}$$

这里 h 是 $v_{0\cdot\text{Source\_ip}}$ 与 $v_{j\cdot\text{Source\_ip}}$ 从前到后相互比较相同的位数, $j=1,2,\cdots,n$.

数据融合相关度为:

$$r_{3j}=\begin{cases} 1, v_o.\text{alert\_signature}=v_j.\text{alert\_signature} \\ 0, v_o.\text{alert\_signature}\ne v_j.\text{alert\_signature} \end{cases} \tag{4}$$

同理, $j=1,2,\cdots,n$.

这里取因素集 $U=\{u_1, u_2, u_3\}$,按照约束函数中的式子(2)、(3)、(4)则可以确定矩阵 R ,再运用式子(1)则可以运算出报警融合数据的关联度.依据最大值隶属原则,取 $b_L=\max\{b_1, b_2, \cdots, b_n\}$, $1\le L\le n$,将当前生成的融合报警 v_0 与报警线程中最新的报警进行对比,则同已经产生的融合报警 v_L 最有可能属于同一个数据攻击.

3.2　属性相异度的运算

在计算两个警报的相异度时,参与相异度计算的警报可以用 n 个属性变量来表示,如源IP、目的IP、端口号、警报标识、攻击时间等.警报的属性类型主要有数值型变量、枚举型变量与布尔型变量.对于警报的关键属性,这里都为其定义了一个相异函数,总的相异度由各个关键属性相异度运算得到.下面首先介绍不同属性类型相异度的运算方法,接着再给出警报相异度的运算公式.

(1)数值型变量.数值型变量是一个连续的变量如攻击时间.这里设警报 $p=(Ap1, Ap2, \cdots, Apn)$, $q=(Aq1, Aq2, \cdots, Aqn)$,其中, $Ap1, Ap2, \cdots, Apn$ 与 $Aq1, Aq2, \cdots, Aqn$ 分别是警报 p 与警报 q 中类型为数值型的属性.警报 p 与警

报 q 中关于第 m 个到第 n 个属性相异度的运算则有如下公式：

$$d_{p,q}^{(m...n)}=\sqrt{w^{(1)}\left|A_{p1}-A_{q1}\right|^2+w^{(2)}\left|A_{p2}-A_{q2}\right|^2+...+w^{(r)}\left|A_{pr}-A_{qr}\right|^2} \quad (5)$$

其中，$w(m)$ 表示第 m 个属性在相异度运算中所占的权重多少.

(2)枚举型变量.由于枚举型变量的特点，它具有多个取值.这里假定警报 p 和 q 的第 m 个属性为枚举类型，则对于这两个警报关于 m 属性上的相异度运算可以采用简单匹配的方法，其中 u 是所有枚举型变量的数目，而 v 是匹配的数目，也就是警报 p 与 q 取值相同时属性变量的数目，则有如下公式：

$$d_{p,q}^{(m)}=w^{(m)}\frac{u-v}{u} \quad (6)$$

(3)布尔型变量.与枚举型变量不同，布尔型变量有且仅有两个状态：1或0.这里假定警报 p 和 q 的第 m 个属性为布尔类型，则对于这两个警报关于 m 属性上的相异度运算可以采用简单匹配系数法，对应的公式如下：

$$d_{p,q}^{(m)}=w^{(m)}\frac{a+b}{a+b+c+e} \quad (7)$$

其中，a 是警报 p 值为1、警报 q 值为0的属性变量的数目，b 是警报 p 值为0、警报 q 值为1的属性变量的数目，而c则是警报 p 与 q 均为1的属性变量的数目，e 是警报 p 与 q 均为0的属性变量的数目.

3.3 属性数据融合算法的流程

警报轨迹所对应的事件并不一定是网络攻击或者入侵的一部分，只能认为它们是数据融合过程中非常重要的一部分.具体的警报所关联的属性数据融合算法的流程如下：

(1)开始，新警报p到来，初始相异度设为无穷大.

(2)计算新警报p与警报轨迹数据库一个警报的相异度d'，主要的伪代码为：

```
IF  d'<d  THEN
BEGIN  修改d值  d=d'
END
```

(3)IF没有遍历完警报轨迹数据库 THEN.

```
BEGIN  q=q+1(从警报轨迹数据库取下一条警报)
GOTO (2)
END
```

(4)IF d>属性最大相异度 THEN.

```
BEGIN  则向警报轨迹数据库中添加一个新的警报轨迹
END
ELSE
BEGIN  修改警报轨迹数据库，把警报p存入已有的警报轨迹中
END
```

(5)停止.

4 实验测试与性能分析

4.1 实验环境的搭建

本实验的计算机配置为Pentium(R) 4 CPU 2.40GHz，4G内存、500G硬盘.操作系统为Windows 7，编程语言Visual Basic，测试工具为Snort.

4.2 入侵检测系统的性能测试

本文所用的测试数据是DPRAP中的数据，实验主要进行的是原始警报数、数据过滤融合报警数与未过滤仅融合警报数的比较.具体实验数据结果如表1、表2所示，比较分析如图2所示.本实验中涉及到的报警缩减比率和报警精准率的参数定义如下：

$$AR=\frac{FA-BA}{FA}*100\% \quad (8)$$

$$AQ=\frac{FTA-BTA}{FTA}*100\% \quad (9)$$

表1 数据融合前后的报警数

入侵检测工具	报警区域	原始报警数		数据融合后报警数	
		警报数	真实报警数	报警数	真实报警数
Snort	DMZ	906	73	273	111
	Inside	939	61		
	总计	1845	134		

表2　数据过滤融合前后的报警数

入侵检测工具	报警区域	原始报警数		数据过滤后报警数		数据融合后报警数	
		报警数	真实报警数	报警数	真实报警数	报警数	真实报警数
Snort	DMZ	906	73	497	64	138	103
	Inside	939	61	474	52		
	总计	1845	134	971	116		

其中，AR 为报警缩减比率，FA 为处理前的总警报数，BA 为处理后的总报警数，AQ 为报警精准率，FTA 为处理前的真报警数，BTA 为处理后的真报警数.

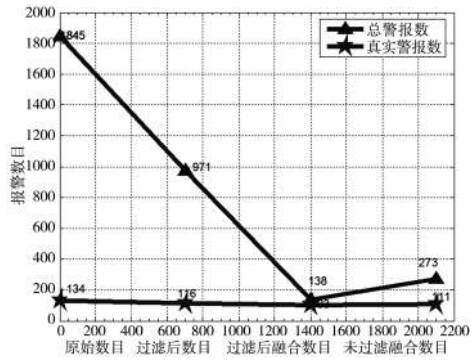

图2　数据过滤融合前后的报警数目比较

DPRAP中的数据集是从DMZ(非军事区)和内网两个区域内获取的网络数据流.它是外部网路与内部网路之间的一段网路,可以得到整个防火墙系统的相关保护.Inside和DMZ的数据集是分开单独重播的.实验结果表明,本系统可以有效地处理来自异构网络的报警信息,能够过滤掉大部分虚警,压缩了报警的数量,融合的效率得到大幅提升,报警的精准率明显提高,极大地增强了整个系统的检测效率.

5　结束语

伴随着网络的不断发展,网络上的安全问题也日益成为人们研究的热点.针对此问题,本文提出了一种基于属性数据融合的分布式入侵检测系统,对模拟攻击数据进行检测率、误报率、融合比等方面的实验.实验结果表明,本文的方案与传统的入侵检测系统相比,在网络安全防护中具有很好的应用前景.

参考文献:

[1]边婧,彭新光,王颖,等.入侵检测不平衡样本子群发现数据简化策略[J].计算机应用研究,2014,31(7):2123-2126.

[2]杨雅辉,黄海珍,沈晴霓.基于增量式GHSOM神经网络模型的入侵检测研究[J].计算机学报,2014,37(5):1216-1221.

[3]Bin Luo, Jingbo Xia. A novel intrusion detection system based on feature generation with visualization strategy [J].Expert Systems With Applications, 2014, 41 (9): 4139-4147.

[4]Feng Jiang, Yuefei Sui, Cungen Cao. An incremental decision tree algorithm based on rough sets and its application in intrusion detection [J].Artificial Intelligence Review,2013,40(4):517-530.

[5]Georgios P Spathoulas, Sokratis K Katsikas. Enhancing IDS performance through comprehensive alert post-processing [J]. Computers & Security, 2013, 37 (9): 176-196.

[6]阳时来,杨雅辉,沈晴霓.一种基于半监督GHSOM的入侵检测方法[J].计算机研究与发展,2013,50(11):2375-2382.

[7]孙文静,钱华.改进BM算法及其在网络入侵检测中的应用[J].计算机科学,2013,40(12):174-176.

[8]江颉,王卓芳,陈铁明.不平衡数据分类方法及其在入侵检测中的应用研究[J].计算机科学,2013,40(4):131-135.

(责任编辑:王前)