一种基于日志和SNMP信息融合的网络安全态势感知分析方法

申请号：201410120989.X
申请日：2014-03-28

申请(专利权)人　哈尔滨工程大学

地址　150001 黑龙江省哈尔滨市南岗区南通大街145号哈尔滨工程大学科技处知识产权办公室

发明(设计)人　王慧强　梁晓　郭方方　吕宏武

主分类号　H04L29/06(2006.01)I

分类号　H04L29/06(2006.01)I

公开(公告)号　103905440A

公开(公告)日　2014-07-02

专利代理机构

代理人

(19)中华人民共和国国家知识产权局

(12)发明专利

(10)授权公告号 CN 103905440 B
(45)授权公告日 2017.02.22

(21)申请号 201410120989.X

(22)申请日 2014.03.28

(65)同一申请的已公布的文献号
申请公布号 CN 103905440 A

(43)申请公布日 2014.07.02

(73)专利权人 哈尔滨工程大学
地址 150001 黑龙江省哈尔滨市南岗区南通大街145号哈尔滨工程大学科技处知识产权办公室

(72)发明人 王慧强 梁晓 郭方方 吕宏武

(51)Int.Cl.
H04L 29/06(2006.01)

(56)对比文件
CN 102123149 A,2011.07.13,
CN 102340485 A,2012.02.01,
CN 101951329 A,2011.01.19,
CN 101075915 A,2007.11.21,
US 8365246 B2,2013.01.29,
CN 102624696 A,2012.08.01,
纪乃丹.基于事件场景关联的多源安全信息融合研究.《中国优秀硕士学位论文全文数据库信息科技辑》.2013,(第2期),
孙德衡.基于指标融合的网络安全态势评估模型研究.《中国优秀硕士学位论文全文数据库信息科技辑》.2013,(第1期),

审查员 张春洁

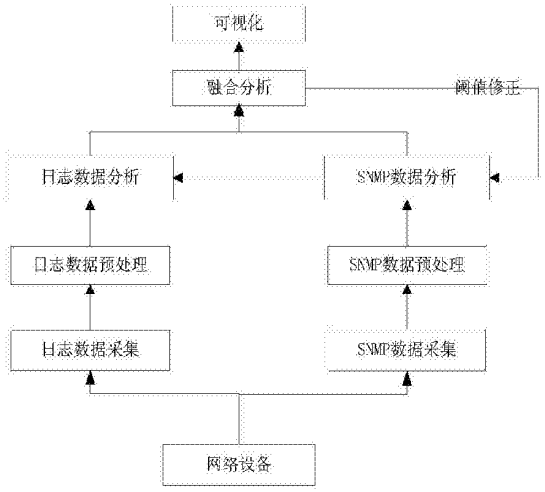
权利要求书3页 说明书10页 附图3页

(54)发明名称

一种基于日志和SNMP信息融合的网络安全态势感知分析方法

(57)摘要

本发明属于网络安全态势感知领域,具体涉及一种基于日志和SNMP信息融合的网络安全态势感知分析方法。本发明包括:基于日志和SNMP数据融合的数据采集;基于日志和SNMP数据融合的预处理;进行日志数据分析和SNMP数据分析;进行日志和SNMP数据的数据融合;进行日志和SNMP数据融合的可视化。本发明与单一分析日志数据或者SNMP数据源相比,这两种数据的结合,能较好的分析网络状态整体的运行趋势,两种数据结合分析更全面、精确;本系统的数据处理根据用户需求选择重要度高的进行处理,减轻了对大量数据处理的负担;本系统的采用阈值自动修正的方法,使用户自定义的阈值参数更加的精确,提高数据融合的准确性。



CN 103905440 B

1.一种基于日志和SNMP信息融合的网络安全态势感知分析方法，其特征在于：

(1)基于日志和SNMP数据融合的数据采集：

(1.1)进行日志数据采集：

(1.1.1)从网络设备中获取日志数据信息；

(1.1.2)设置日志采集代理的采集日志格式：日志记录时间，源主机地址，目的地址，源端口号，目的端口号，SYN标志，服务类型；

(1.1.3)启动采集代理，将采集到的日志数据存入源日志数据库；

(1.2)进行SNMP数据采集：

(1.2.1)使用时间片轮询的方式定时采集数据，轮询时间设为固定值；

(1.2.2)读取采集配置文件，设置传感器ID、时间粒度、存储路径、服务器IP；

(1.2.3)设置SNMP采集代理的采集SNMP格式：标识符id、信息产生时间time、源主机地址IP、CPU使用率UsedCPU、内存使用率UsedMem、接口利用率UsedPort、流量Flux、丢包率PacketLoss、接口信息错误率PortErrorRate、响应时间ResponseTime；

(1.2.4)启动SNMP采集代理，将采集到的SNMP数据存入源SNMP数据库；

(2)基于日志和SNMP数据融合的预处理：

(2.1)日志数据预处理：

(2.1.1)从源日志数据库中获取数据；

(2.1.2)归一化处理，转化为统一的格式，时间time、源主机地址IP，目标主机与当前连接相同的连接次数count1、出现针对主机的SYN错误的连接百分比ser1、目标端口与当前连接相同的连接次数count2、出现针对服务的SYN错误的连接百分比ser2；

(2.1.3)将预处理后的日志数据存入日志数据库；

(2.2)SNMP数据融合预处理：

(2.2.1)从源SNMP数据库中获取数据；

(2.2.2)归一化处理：将获得的源数据转换成[0,1]之间的数据，即除了标识符、信息产生时间、源主机地址外的每个属性除以各属性的最大允许值，获得相应的百分比；

(2.2.3)将预处理后的SNMP数据存入SNMP数据库；

(3)进行日志数据分析和SNMP数据分析：

(3.1)进行SNMP数据分析：

(3.1.1)从数据预处理中的SNMP数据库中获取数据；

(3.1.2)计算事件基于SNMP的重要度，并与阈值进行比较；

(3.1.3)将重要度高的事件存入安全事件数据库；

(3.1.4)根据融合结果进行阈值修正；

(4)进行日志和SNMP数据的数据融合：

(4.1)采用五层模糊神经网络对日志信息和SNMP信息进行融合分析；

(4.2)对五层模糊神经网络进行学习训练获得每层之间的权值；

(4.3)将日志的每个字段和SNMP事件的重要度作为模糊神经网络输入层的输入；

(4.4)将实际输出与期望输出值进行比较，如果输出层的实际输出不等于期望输出，则进入后向传播过程；

(4.5)后向传播时，把误差信号按原来前向传播的通路反向传回，逐层递归的计算实际

输出与期望输出的差值，根据误差的均方差调节权值，对隐含层的每个神经元的权系数进行修改，使误差趋于最小；

(5)进行日志和SNMP数据融合的可视化：显示检测对象的网络安全态势和检测结果。

2.根据权利要求1所述的一种基于日志和SNMP信息融合的网络安全态势感知分析方法，其特征在于：所述模糊神经网络输入层的输入包括目标主机与当前连接相同的连接次数count1、出现针对主机的SYN错误的连接百分比ser1、目标端口与当前连接相同的连接次数count2、出现针对服务的SYN错误的连接百分比ser2、基于SNMP的重要度。

3.根据权利要求1或2所述的一种基于日志和SNMP信息融合的网络安全态势感知分析方法，其特征在于：所述基于SNMP重要度的计算方法包括：

(1)利用SNMP采集代理获得SNMP数据信息，进行归一化处理；

(2)确定除了标识符、信息产生时间、源主机地址外每个属性的最大允许值，用$a_i^{\max}$(i=1,2,…7)表示；

(3)根据SNMP数据库中的信息求除了标识符、信息产生时间、源主机地址外每个属性的平均值，用$\bar{a}_i$ (i=1,2,…7)表示；

(4)计算事件基于SNMP的重要度P

$$P=\sum_i \eta_i a_i^*,\text{其中}\eta_i=\frac{\Delta a_i^*}{\sum_i \Delta a_i},\qquad \Delta a_i=\left|a_i^*-\bar{a}_i\right|,\qquad a_i^*=a_i/a_i^{\max};$$

(5)SNMP事件的重要度P与阈值进行比较，大于该阈值的事件发送到日志数据分析模块和数据融合模块；获得日志信息和融合结果；

所述阈值修正方法包括：

(1)根据先验知识设定初始的阈值q；

(2)将SNMP中重要度大于阈值的事件放入安全事件数据库中，并计算其重要度的平均值$\bar{q}$；

(3)将重要度高的SNMP事件与相关联的日志事件进行融合分析，分析判断正常、异常和攻击事件，并更新安全事件数据库；

(4)使用变化量$\Delta q=q^*-\bar{q}$，其中q*为实时SNMP事件的重要度，阈值修正为：

$$q=\begin{cases}q-\eta^*\Delta q, & \text{该事件为攻击，异常}\\ q+\eta^*\Delta q, & \text{该事件为正常}\end{cases}$$

其中η是修正系数。

4.根据权利要求3所述的一种基于日志和SNMP信息融合的网络安全态势感知分析方法，其特征在于：

所述的五层模糊神经网络为：

第一层：输入层，从前述数据分析阶段中获得日志和SNMP数据，包含5个节点，分别记为x_i，i=1,2…5，用O_i^1表示第一层第i个节点的输出结果，把输入值直接传送给下一层

$O_i^1=x_i,1\le i\le 5$；

第二层：隶属度函数层，实现输入变量的模糊化，输入的特征被分别映射到模糊集，每个特征的模糊集数为3，节点个数为输入变量的模糊集合数之和，用O_{ij}^2、μ_{ij}表示第二层每个

节点的输出结果，用单一节点计算简单的隶属函数：

$$O_{ij}^{2}=\mu_{ij}=\exp(-\frac{(\theta_{i}^{1}-m_{ij})^{2}}{\sigma_{ij}^{2}}),1\le i\le 5,1\le j\le 3$$

第三层：推理层，其中$O_{\Pi j}^{3}$表示第三层上半部分的输出结果，$O_{\Sigma j}^{3}$表示第三层下半部分的输出结果，

上半部分使用Π标记所有节点，

$$O_{\Pi j}^{3}=\mu_{1j}\cdot\mu_{2j}\cdots\cdot\mu_{nj}=\prod_{i=1}^{5}\mu_{ij},(1\le j\le 3)$$

下半部分用Σ标记所有节点，

$$O_{\Sigma j}^{3}=\mu_{1j}+\mu_{2j}+\ldots+\mu_{nj}=\sum_{i=1}^{5}\mu_{ij},(1\le j\le 3)；$$

第四层：求和层，将第三层的上半部分乘以系数a(0＜a＜1)与下半部分乘以系数(1-a)对应相加，包含3个节点，用O_{j}^{4}表示第四层的输出结果，

$$O_{j}^{4}=f_{j}=aO_{\Pi j}^{3}+(1-a)O_{\Sigma j}^{3}=a\prod_{i=1}^{5}\mu_{ij}+(1-a)\sum_{i=1}^{5}\mu_{ij},(1\le j\le 3)；$$

第五层：解模糊化层，用于模糊神经网络最后清晰量的输出，用O_{j}^{5}、y_j(j＝0,1,…,3)表示第五层的输出结果，

$$O_{0}^{5}=y_{0}=\sum_{j=1}^{m}\omega_{j}f_{j}=\sum_{j=1}^{m}\omega_{j}(a\prod_{i=1}^{n}\mu_{ij}+(1-a)\sum_{i=1}^{n}\mu_{ij})$$

$$O_{j}^{5}=y_{j}=\theta(f_{j}),\quad 1\le j\le 3,$$

其中θ(•)是阶跃函数，O_{0}^{5}用于反向传播，$O_{j}^{5}(1\le j\le 3)$表示融合结果，

第一层到第三层的网络连接权值均为1，隶属度函数层(第二层)中高斯函数中的平均值m_{ij}和均方差σ_{ij}，以及第五层中的权值ω_{ij}。

一种基于日志和SNMP信息融合的网络安全态势感知分析方法

技术领域

[0001] 本发明属于网络安全态势感知领域，具体涉及一种基于日志和SNMP信息融合的网络安全态势感知分析方法。

技术背景

[0002] 随着计算机及网络技术的发展，攻击手段日趋专业化，网络安全事件层出不穷，单一的防火墙和入侵检测系统等被动防御技术已经不能确保网络的安全，因此提高网络的主动防御能力是当今网络安全研究领域的主要方向，而对网络安全态势感知领域的研究尤为突出。

[0003] 网络安全态势感知系统的数据源非常丰富，其中日志和SNMP数据占据重要位置。但是单独的SNMP分析或日志分析都存在着一定的缺陷。(1)SNMP代理无法为管理站提供某一目标集的历史数据，只能提供设备的当前状态或一个很短时间段内的数据，对分析网络的整体运行趋势造成了障碍。(2)SNMP协议检测粒度粗糙、数据资料简单且无法提供网络层以上的信息。(3)日志分析的正确性很大程度上取决于计算机系统时间的准确性。如果攻击者提前对计算机时钟进行了调整，那么在运行日志分析系统时就会产生误判断，从而影响分析结果。

[0004] 而将两种数据源相结合，可以增加数据源的完整性和安全性：

[0005] (1)SNMP存在不能为分析提供历史数据这一缺点，而日志则保存较长一段时间的信息，二者结合可以弥补SNMP不能提供历史数据的缺点。

[0006] (2)SNMP检测粒度粗糙、数据资料简单且无法提供网络层以上的信息。而日志信息记录着网络系统发生的各种事件，同时可以提供网络层以上的信息。

[0007] (3)引入SNMP数据对日志进行补充，并且SNMPv2和SNMPv3中还增加了相应的安全机制，这样通过验证和访问控制等方式解决了安全隐患的问题，弥补了日志信息容易被篡改或者删除的问题。

[0008] 因此，本发明提出了一种将日志信息和SNMP数据信息进行融合分析的方法，弥补了单独日志和SNMP分析的不足之处。

[0009] 目前国内一些专家和学者已经对该领域进行了初步研究，如纪乃丹等人提出了一种基于改进的事件场景关联融合模型F-ECS，引入模糊集理论，将SNMP数据与多源日志信息进行融合分析。弱化了集合边界，降低被误报率。但其知识库的建立对专家知识的依赖性比较强，漏报率较高；且阈值固定，检测精度降低。

发明内容

[0010] 本发明的目的是提出一种面向大规模网络，对网络状态进行实时监控，并对网络中异常或攻击事件进行自动检测的基于日志和SNMP信息融合的网络安全态势感知分析方法。

[0011] 本发明的目的是这样实现的：

[0012] (1)基于日志和SNMP数据融合的数据采集：

[0013] (1.1)进行日志数据采集：

[0014] (1.1.1)从网络设备中获取日志数据信息；

[0015] (1.1.2)设置日志采集代理的采集日志格式：日志记录时间，源主机地址，目的地址，源端口号，目的端口号，SYN标志，服务类型；

[0016] (1.1.3)启动采集代理，将采集到的日志数据存入源日志数据库；

[0017] (1.2)进行SNMP数据采集：

[0018] (1.2.1)使用时间片轮询的方式定时采集数据，轮询时间设为固定值；

[0019] (1.2.2)读取采集配置文件，设置传感器ID、时间粒度、存储路径、服务器IP；

[0020] (1.2.3)设置SNMP采集代理的采集SNMP格式：标识符id、信息产生时间time、源主机地址IP、CPU使用率UsedCPU、内存使用率UsedMem、接口利用率UsedPort、流量Flux、丢包率PacketLoss、接口信息错误率PortErrorRate、响应时间ResponseTime；

[0021] (1.2.4)启动SNMP采集代理，将采集到的SNMP数据存入源SNMP数据库；

[0022] (2)基于日志和SNMP数据融合的预处理：

[0023] (2.1)日志数据预处理：

[0024] (2.1.1)从源日志数据库中获取数据。

[0025] (2.1.2)归一化处理，转化为统一的格式，时间time、源主机地址IP，目标主机与当前连接相同的连接次数count1、出现针对主机的SYN错误的连接百分比ser1、目标端口与当前连接相同的连接次数count2、出现针对服务的SYN错误的连接百分比ser2；

[0026] (2.1.3)将预处理后的日志数据存入日志数据库；

[0027] (2.2)SNMP数据融合预处理：

[0028] (2.2.1)从源SNMP数据库中获取数据；

[0029] (2.2.2)归一化处理：将获得的源数据转换成[0,1]之间的数据，即除了标识符、信息产生时间、源主机地址外的每个属性除以各属性的最大允许值，获得相应的百分比；

[0030] (2.2.3)将预处理后的SNMP数据存入SNMP数据库；

[0031] (3)进行日志数据分析和SNMP数据分析：

[0032] (3.1)进行SNMP数据分析：

[0033] (3.1.1)从数据预处理中的SNMP数据库中获取数据；

[0034] (3.1.2)计算事件基于SNMP的重要度，并与阈值进行比较；

[0035] (3.1.3)将重要度高的事件存入安全事件数据库；

[0036] (3.1.4)根据融合结果进行阈值修正；

[0037] (4)进行日志和SNMP数据的数据融合：

[0038] (4.1)采用五层模糊神经网络对日志信息和SNMP信息进行融合分析；

[0039] (4.2)对五层模糊神经网络进行学习训练获得每层之间的权值；

[0040] (4.3)将日志的每个字段和SNMP事件的重要度作为模糊神经网络输入层的输入；

[0041] (4.4)将实际输出与期望输出值进行比较，如果输出层的实际输出不等于期望输出，则进入后向传播过程；

[0042] (4.5)后向传播时，把误差信号按原来前向传播的通路反向传回，逐层递归的计算实际输出与期望输出的差值，根据误差的均方差调节权值，对隐含层的每个神经元的权系

数进行修改，使误差趋于最小；

[0043] (5)进行日志和SNMP数据融合的可视化：显示检测对象的网络安全态势和检测结果。

[0044] 模糊神经网络输入层的输入包括目标主机与当前连接相同的连接次数count1、出现针对主机的SYN错误的连接百分比ser1、目标端口与当前连接相同的连接次数count2、出现针对服务的SYN错误的连接百分比ser2、基于SNMP的重要度。

[0045] 基于SNMP重要度的计算方法包括：

[0046] (1)利用SNMP采集代理获得SNMP数据信息，进行归一化处理；

[0047] (2)确定除了标识符、信息产生时间、源主机地址外每个属性的最大允许值，用a_i^{max}(i＝1,2,…7)表示；

[0048] (3)根据SNMP数据库中的信息求除了标识符、信息产生时间、源主机地址外每个属性的平均值，用$\bar{a}_i$ (i=1,2,…7)表示；

[0049] (4)计算事件基于SNMP的重要度P

[0050] $P=\sum_i \eta_i a_i^*$，其中$\eta_i=\frac{\Delta a_i^*}{\sum_i \Delta a_i}$，$\Delta a_i=\left|a_i^*-\bar{a}_i\right|$，$a_i^*=a_i/a_i^{\max}$；

[0051] (5)NMP事件的重要度P与阈值进行比较，大于该阈值的事件发送到日志数据分析模块和数据融合模块。获得日志信息和融合结果；

[0052] 阈值修正方法包括：

[0053] (1)根据先验知识设定初始的阈值q；

[0054] (2)将SNMP中重要度大于阈值的事件放入安全事件数据库中，并计算其重要度的平均值$\bar{q}$；

[0055] (3)将重要度高的SNMP事件与相关联的日志事件进行融合分析，分析判断正常、异常和攻击事件，并更新安全事件数据库；

[0056] (4)使用变化量$\Delta q=q^*-\bar{q}$，其中q*为实时SNMP事件的重要度，阈值修正为：

[0057] $$q=\begin{cases} q-\eta^*\Delta q, & \text{该事件为攻击，异常} \\ q+\eta^*\Delta q, & \text{该事件为正常} \end{cases}$$

[0058] 其中η是修正系数。

[0059] 五层模糊神经网络为：

[0060] 第一层：输入层，从前述数据分析阶段中获得日志和SNMP数据，包含5个节点，分别记为x_i，i＝1,2…5，用O_i^1表示第一层第i个节点的输出结果，把输入值直接传送给下一层

[0061] $O_i^1=x_i, 1\le i\le 5$；

[0062] 第二层：隶属度函数层，实现输入变量的模糊化，输入的特征被分别映射到模糊集，每个特征的模糊集数为3，节点个数为输入变量的模糊集合数之和，用O_{ij}^2、μ_{ij}表示第二层每个节点的输出结果，用单一节点计算简单的隶属函数：

[0063] $$O_{ij}^2=\mu_{ij}=\exp\left(-\frac{(O_i^1-m_{ij})^2}{\sigma_{ij}^2}\right), 1\le i\le 5, 1\le j\le 3$$

[0064] 第三层：推理层，其中$O_{\Pi j}^{3}$表示第三层上半部分的输出结果，$O_{\Sigma j}^{3}$表示第三层下半部分的输出结果，

[0065] 上半部分使用Π标记所有节点，

[0066] $$O_{\Pi j}^{3}=\mu_{1j}\cdot\mu_{2j}\cdots\cdot\mu_{nj}=\prod_{i=1}^{5}\mu_{ij},(1\le j\le 3)$$

[0067] 下半部分用Σ标记所有节点，

[0068] $$O_{\Sigma j}^{3}=\mu_{1j}+\mu_{2j}+\ldots+\mu_{nj}=\sum_{i=1}^{5}\mu_{ij},(1\le j\le 3)$$；

[0069] 第四层：求和层，将第三层的上半部分乘以系数a(0＜a＜1)与下半部分乘以系数(1-a)对应相加，包含3个节点，用O_{j}^{4}表示第四层的输出结果，

[0070] $$O_{j}^{4}=f_{j}=aO_{\Pi j}^{3}+(1-a)O_{\Sigma j}^{3}=a\prod_{i=1}^{5}\mu_{ij}+(1-a)\sum_{i=1}^{5}\mu_{ij},(1\le j\le 3)$$；

[0071] 第五层：解模糊化层，用于模糊神经网络最后清晰量的输出，用O_{j}^{5}、y_j(j＝0,1,…,3)表示第五层的输出结果，

[0072] $$O_{0}^{5}=y_{0}=\sum_{j=1}^{m}\omega_{j}f_{j}=\sum_{j=1}^{m}\omega_{j}(a\prod_{i=1}^{n}\mu_{ij}+(1-a)\sum_{i=1}^{n}\mu_{ij})$$

[0073] $$O_{j}^{5}=y_{j}=\theta(f_{j}),\quad 1\le j\le 3$$，

[0074] 其中θ(•)是阶跃函数，O_{0}^{5}用于反向传播，$O_{j}^{5}(1\le j\le 3)$表示融合结果，

[0075] 第一层到第三层的网络连接权值均为1，隶属度函数层(第二层)中高斯函数中的平均值m_{ij}和均方差σ_{ij}，以及第五层中的权值ω_{ij}。

[0076] 本发明的有益效果在于：

[0077] 本发明提供的一种基于日志和SNMP信息融合的网络安全态势感知分析方法目的在于提高网络安全态势感知的准确性，与单一分析日志数据或者SNMP数据源相比，这两种数据的结合，能较好的分析网络状态整体的运行趋势，两种数据结合分析更全面、精确；本系统的数据处理根据用户需求选择重要度高的进行处理，减轻了对大量数据处理的负担；本系统的采用阈值自动修正的方法，使用户自定义的阈值参数更加的精确，提高数据融合的准确性。

附图说明

[0078] 图1为本发明的具体实施方案的网络部署图；

[0079] 图2为本发明的总体框架图；

[0080] 图3为SNMP数据分析模块流程图；

[0081] 图4为模糊神经网络结果图。

[0082] 具体实施方法

[0083] 下面结合附图和具体实施方法对本发明作更加详细的描述：

[0084] 基于日志和SNMP数据融合的网络安全态势感知分析方法包括数据采集、预处理、数据分析、数据融合和可视化五个阶段。

[0085] 1、基于日志和SNMP数据融合的数据采集阶段包括日志数据采集和SNMP数据采集，其中

[0086] (1)日志数据采集包括以下三个步骤：

[0087] ①从网络设备中获取日志数据信息。

[0088] ②设置日志采集代理的采集日志格式：日志记录时间，源主机地址，目的地址，源端口号，目的端口号，SYN标志，服务类型。

[0089] ③启动采集代理，将采集到的日志数据存入源日志数据库。

[0090] (2)SNMP数据采集包括以下四个步骤：

[0091] ①使用时间片轮询的方式定时采集数据，轮询时间设为固定值。

[0092] ②读取采集配置文件，设置传感器ID、时间粒度、存储路径、服务器IP。

[0093] ③设置SNMP采集代理的采集SNMP格式：标识符、信息产生时间、源主机地址、CPU使用率、内存使用率、接口利用率、流量、丢包率、接口信息错误率、响应时间，分别用id、time、IP、UsedCPU，UsedMem，UsedPort，Flux，PacketLoss，PortErrorRate，ResponseTime表示。

[0094] ④启动SNMP采集代理，将采集到的SNMP数据存入源SNMP数据库。

[0095] 2、基于日志和SNMP数据融合的预处理阶段包括日志数据预处理和SNMP数据预处理，其中

[0096] (1)日志数据预处理包括三个步骤：

[0097] ①从上述数据采集阶段中的源日志数据库中获取数据。

[0098] ②归一化处理，转化为统一的格式，即时间time、源主机地址IP，在一个时间段内2分钟目标主机与当前连接相同的连接次数count1、出现针对主机SYN错误的连接百分比ser1、目标端口与当前连接相同的连接次数count2、出现针对服务的SYN错误的连接百分比ser2。

[0099] ③将预处理后的日志数据存入日志数据库。

[0100] (2)SNMP预处理包括三个步骤：

[0101] ①从上述数据采集阶段中的源SNMP数据库中获取数据。

[0102] ②归一化处理。将获得的源数据转换成[0,1]之间的数据，即每个属性除了标识符、信息产生时间、源主机地址外除以各自的最大允许值，由主机性能和先验知识得到，从而获得相应的百分比。

[0103] ③将预处理后的SNMP数据存入SNMP数据库。

[0104] 3、基于日志和SNMP数据融合的数据分析阶段包括日志数据分析和SNMP数据分析，其中

[0105] (1)SNMP数据分析包括以下四个步骤：

[0106] ①从上述数据预处理阶段中的SNMP数据库中获取数据。

[0107] ②计算事件基于SNMP的重要度，并与给定的阈值进行比较。具体计算方法见下文

[0108] ③将重要度高的事件存入安全事件数据库。

[0109] ④根据融合结果进行阈值修正。

[0110] 前述的事件基于SNMP重要度的计算方法具体包括：

[0111] ①利用SNMP采集代理获得SNMP数据信息，并将其进行归一化处理，使其形式统一，其格式为：标识符、信息产生时间、源主机地址、CPU使用率、内存使用率、接口利用率、流量、

丢包率、接口信息错误率、响应时间。分别用id、time、IP、UsedCPU,UsedMem,UsedPort,Flux,PacketLoss,PortErrorRate,ResponseTime表示。

[0112] ②确定每个属性的最大允许值,除了标识符、信息产生时间、源主机地址外,分别用a_i^{max} (i=1,2,…7)表示。

[0113] ③根据SNMP数据库中的信息求每个属性的平均值,除了标识符、信息产生时间、源主机地址外,分别用$\bar{a}_i$, i=1,2,…7表示。

[0114] ④计算事件基于SNMP的重要度P。

[0115] $P=\sum_i \eta_i a_i^*$,其中$\eta_i=\frac{\Delta a_i^*}{\sum_i \Delta a_i}$, $\Delta a_i=\left|a_i^*-\bar{a}_i\right|$, $a_i^*=a_i/a_i^{max}$,

[0116] ⑤将每个SNMP事件的重要度P与阈值进行比较,大于该阈值的事件发送到日志数据分析模块和数据融合模块。从而获得与之相关的日志信息和融合结果。

[0117] 前述基于SNMP重要度的阈值修正方法实现过程如下:

[0118] ①首先根据先验知识设定初始的阈值q。

[0119] ②将SNMP中重要度大于阈值的事件放入安全事件数据库中,并计算其重要度的平均值$\bar{q}$。

[0120] ③将重要度高的SNMP事件与相关联的日志事件进行融合分析,分析判断正常、异常和攻击事件,并更新安全事件数据库。

[0121] ④使用变化量$\Delta q=q^*-\bar{q}$,其中q*为实时SNMP事件的重要度,阈值修正计算公式如下:

[0122] $$q=\begin{cases} q-\eta^*\Delta q, & \text{该事件为攻击,异常} \\ q+\eta^*\Delta q, & \text{该事件为正常} \end{cases}$$

[0123] 其中η是修正系数。

[0124] (2)日志数据分析包括以下三个步骤:

[0125] ①从上述数据预处理阶段中的日志数据库中获取数据。

[0126] ②根据上述SNMP预处理中得到的重要度高的事件,获取与之相关联的日志数据,即同一时间段、源主机地址相同的日志。

[0127] ③将相关联的日志数据发送到数据融合阶段。

[0128] 4、基于日志和SNMP数据融合的数据融合阶段

[0129] 数据融合阶段包括以下五个步骤:

[0130] ①采用五层模糊神经网络对日志信息和SNMP信息进行融合分析。

[0131] ②对该五层模糊神经网络进行学习训练获得每层之间的权值。

[0132] ③将日志的每个字段和SNMP事件的重要度作为该模糊神经网络输入层的输入,包括在一个时间段2分钟内目标主机与当前连接相同的连接次数count1、出现针对主机SYN错误的连接百分比ser1、目标端口与当前连接相同的连接次数count2、出现针对服务的SYN错误的连接百分比ser2、和基于SNMP的重要度。

[0133] ④将实际输出与期望输出值进行比较,如果输出层的实际输出不等于期望输出,则进入后向传播过程。

[0134] ⑤后向传播时，把误差信号按原来前向传播的通路反向传回，逐层递归的计算实际输出与期望输出的差值，并根据误差的均方差调节权值，对隐含层的每个神经元的权系数进行修改，以使误差趋于最小。

[0135] 5、基于日志和SNMP数据融合的可视化阶段

[0136] 可视化阶段即显示检测对象的网络安全态势和检测结果的查询。

[0137] 本发明进一步细化分为：

[0138] 首先需要将本发明所提供的方法应用到具体的网络中，如图1所示：

[0139] 1、在网络中的任意一台Linux主机上安装本发明所采用的日志采集代理和SNMP采集代理。同时在Web服务器和防火墙上也安装本发明所采用的日志采集代理和SNMP采集代理。

[0140] 2、将Linux主机、数据库服务器、Web服务器和监控平台都连接到华为交换机。而内网和外网之间使用华为硬件防火墙来保护内网的安全性。

[0141] 3、设置数据库服务器，负责存储由各个传感器上传的原始数据、预处理、分析和融合的数据。监控平台通过华为交换机从数据库服务器中获得相关数据。并通过本发明的方法，如图2所示的总体框架图，对数据库中的数据进行分析，将分析结果以界面的形式显示出来，并存入数据库服务器中。

[0142] 结合图2所示，本发明的基于日志和SNMP信息融合的网路安全态势感知分析方法主要包括5个阶段：数据采集阶段、数据预处理阶段、数据分析阶段、数据融合阶段和可视化阶段。其中数据采集阶段包括日志数据采集和SNMP数据采集；数据预处理阶段包括日志数据预处理和SNMP数据预处理；数据分析阶段包括日志数据分析和SNMP数据分析；而SNMP数据分析阶段中又根据融合分析结果对阈值进行修正。

[0143] 1、基于日志和SNMP数据融合的数据采集阶段中的日志数据采集包括以下三个步骤：

[0144] (1)从网络设备中，包括Linux主机、Web服务器和防火墙，获取日志数据信息。

[0145] (2)设置日志采集代理的采集日志格式：日志记录时间，源主机地址，目地地址，源端口号，目的端口号，SYN标志，服务类型。

[0146] (3)启动采集代理，将采集到的日志数据存入数据库服务器中的源日志数据库。

[0147] 2、基于日志和SNMP数据融合的数据采集阶段中的SNMP数据采集包括以下五个步骤：

[0148] (1)从网络设备中，包括Linux主机、Web服务器和防火墙，获取SNMP数据信息。

[0149] (2)使用时间片轮询的方式定时采集数据，轮询时间设定为120s。

[0150] (3)读取采集配置文件，设置传感器ID、时间粒度、存储路径、服务器IP。

[0151] (4)设置SNMP采集代理的采集SNMP格式：标识符、信息产生时间、源主机地址、CPU使用率、内存使用率、接口利用率、流量、丢包率、接口信息错误率、响应时间，分别用id、time、IP、UsedCPU,UsedMem,UsedPort,Flux,PacketLoss,PortErrorRate,ResponseTime表示。

[0152] (5)启用SNMP采集代理，将采集到的SNMP数据存入数据库服务器中的源SNMP数据库。

[0153] 3、基于日志和SNMP数据融合的预处理阶段中的日志数据预处理包括以下三个步

骤：

[0154] (1)从上述数据采集阶段中的源日志数据库中获取数据。

[0155] (2)进行归一化处理，转化为统一的格式，即时间time、源主机地址IP，在一个时间段内，2分钟，目标主机与当前连接相同的连接次数count1、出现针对主机的SYN错误的连接百分比ser1、目标端口与当前连接相同的连接次数count2、出现针对服务得到SYN错误的连接百分比ser2。

[0156] (3)存入数据库服务器中的日志数据库。

[0157] 4、基于日志和SNMP数据融合的预处理阶段中的SNMP数据预处理包括以下三个步骤：

[0158] (1)从上述数据采集阶段中的源SNMP数据库中获取数据。

[0159] (2)归一化处理。将获得的源数据转换成[0,1]之间的数据，即每个属性除了标识符、信息产生时间、源主机地址外除以各自的最大允许值，从而获得相应的百分比。

[0160] (3)存入数据库服务器中的SNMP数据库。

[0161] 5、基于日志和SNMP数据融合的数据分析阶段中的SNMP数据分析包括以下四个步骤(如图3所示)：

[0162] (1)从上述数据预处理阶段中的SNMP数据库中获取数据，包括标识符、信息产生时间、源主机地址、CPU使用率、内存使用率、接口利用率、流量、丢包率、接口信息错误率、响应时间。并分别用id、time、IP、UsedCPU，UsedMem，UsedPort，Flux，PacketLoss，PortErrorRate，ResponseTime表示。

[0163] (2)计算事件基于SNMP的重要度。

[0164] ①确定每个属性的最大允许值，除标识符、信息产生时间、源主机地址外，分别用$a_i^{\max}$ (i＝1，2，…7)表示。

[0165] ②求每个属性的平均值，除标识符、信息产生时间、源主机地址外，分别用$\bar{a}_i$，i=1,2,…7 表示。

[0166] ③计算事件基于SNMP的重要度P：

[0167] $$P=\sum_i \eta_i a_i^*, \text{其中} \eta_i=\frac{\Delta a_i^*}{\sum_i \Delta a_i}, \qquad \Delta a_i=\left|a_i^*-\bar{a}_i\right|, \qquad a_i^*=a_i \Big/ a_i^{\max},$$

[0168] (3)阈值比较，将每个SNMP时间的重要度P与阈值q进行比较，将重要度大于阈值q的事件存入数据库服务器中的安全事件数据库。

[0169] (4)根据融合结果进行阈值修正。

[0170] ①首先根据先验知识设定初始阈值q。

[0171] ②计算安全事件数据库中SNMP事件的重要度的平均值$\bar{q}$。

[0172] ③根据融合分析结果，判断正常、异常和攻击事件，并更新安全事件数据库。

[0173] ④阈值修正，其中q*为实时SNMP事件的重要度，Δq为重要度变化量，η为修正系数。

[0174] $$q=\begin{cases} q-\eta^*\Delta q, & \text{该事件为攻击，异常} \\ q+\eta^*\Delta q, & \text{该事件为正常} \end{cases}, \quad \Delta q=q^*-\bar{q}$$

[0175] 6、基于日志和SNMP数据融合的数据分析阶段中的日志数据分析包括以下四个步骤：

[0176] (1)从上述数据预处理阶段中的日志数据库中获取数据。

[0177] (2)根据上述SNMP预处理中得到的重要度高的事件，获取与之相关联的日志数据，即同一时间段、源主机地址相同的日志。

[0178] (3)将相关联的日志数据发送到数据融合阶段。

[0179] 7、基于日志和SNMP数据融合的融合分析阶段

[0180] (1)采用五层模糊神经网络对日志信息和SNMP信息进行融合分析。

[0181] (2)对该五层模糊神经网络进行学习训练获得每层之间的权值。

[0182] (3)将日志的每个字段和SNMP事件的重要度作为该模糊神经网络输入层的输入(包括在一个时间段内目标主机与当前连接相同的连接次数count1、出现针对主机的SYN错误的连接百分比ser1、目标端口与当前连接相同的连接次数count2、出现针对服务的SYN错误的连接百分比ser2、和基于SNMP的重要度。

[0183] (4)将实际输出与期望输出值进行比较，如果输出层的实际输出不等于期望输出，则进入后向传播过程。

[0184] (5)后向传播时，把误差信号按原来前向传播的通路反向传回，逐层递归的计算实际输出与期望输出的差值，并根据误差的均方差调节权值，对隐含层的每个神经元的权系数进行修改，以使误差趋于最小。

[0185] 上述采用的五层模糊神经网络具体结构如下：

[0186] 第一层：输入层。从前述数据分析阶段中获得日志和SNMP数据，作为第一层的输入。该层包含5个节点，分别记为x_i，i＝1，2…5。用O_i^1表示第一层第i个节点的输出结果。该层只是把输入值直接传送给下一层，即

[0187] $$O_i^1 = x_i, 1 \le i \le 5$$

[0188] 第二层：隶属度函数层。实现输入变量的模糊化，输入的各个特征在该层被分别映射到相应模糊集，每个特征的模糊集数为3。该层的节点个数为各个输入变量的模糊集合数之和即3*5＝15。用O_{ij}^2、μ_{ij}表示第二层每个节点的输出结果，用单一节点计算简单的隶属函数高斯函数，因而有：

[0189] $$O_{ij}^2 = \mu_{ij} = \exp\left(-\frac{(O_i^1 - m_{ij})^2}{\sigma_{ij}^2}\right), 1 \le i \le 5, 1 \le j \le 3$$

[0190] 第三层：推理层。该层分为两部分，即加法推理和乘积推理。其中$O_{\Pi j}^3$表示第三层上半部分加法推理的输出结果，$O_{\Sigma j}^3$表示第三层下半部分乘积推理的输出结果。

[0191] 上半部分使用Π标记所有节点，对应为隶属度函数的乘积推理，即

[0192] $$O_{\Pi j}^3 = \mu_{1j} \cdot \mu_{2j} \cdots \mu_{nj} = \prod_{i=1}^{5} \mu_{ij}, (1 \le j \le 3)$$

[0193] 下半部分用Σ标记所有节点，对应隶属度函数的加法推理，即

[0194] $$O_{\Sigma j}^3 = \mu_{1j} + \mu_{2j} + \ldots + \mu_{nj} = \sum_{i=1}^{5} \mu_{ij}, (1 \le j \le 3)$$

[0195] 第四层：求和层。该层是将第三层的上半部分乘以系数a(0＜a＜1)与下半部分乘以系数(1-a)对应相加。该层共包含3个节点，用O_j^4表示第四层的输出结果。即

[0196] $$O_j^4 = f_j = aO_{\Pi j}^3 + (1-a)O_{\Sigma j}^3 = a\prod_{i=1}^{5}\mu_{ij} + (1-a)\sum_{i=1}^{5}\mu_{ij}, (1 \le j \le 3)$$

[0197] 第五层：解模糊化层输出层。经过上面四层的计算得到融合值的模糊量，该层实现了模糊神经网络最后清晰量的输出。用O_j^5、y_j(j＝0,1,…,3)表示第五层的输出结果。

[0198] $$O_0^5 = y_0 = \sum_{j=1}^{m}\omega_j f_j = \sum_{j=1}^{m}\omega_j (a\prod_{i=1}^{n}\mu_{ij} + (1-a)\sum_{i=1}^{n}\mu_{ij})$$

[0199] $$O_j^5 = y_j = \theta(f_j), \quad 1 \le j \le 3,$$

[0200] 其中θ(•)是阶跃函数，O_0^5用于反向传播，从而使实际输出值与期望值得误差为最小，而$O_j^5(1 \le j \le 3)$示融合结果。

[0201] 在该神经网络结果模型中，第一层到第三层的网络连接权值均为1。而其中需要辨识的参数有三类：隶属度函数层第二层中高斯函数中的平均值m_{ij}和均方差σ_{ij}，以及第五层中的权值ω_{ij}。他们都是通过BP神经网络学习算法训练得来的。

[0202] 本发明提供的一种基于日志和SNMP信息融合的网络安全态势感知分析方法目的在于提高网络安全态势感知的准确性，与单一分析日志数据或者SNMP数据源相比，这两种数据的结合，能较好的分析网络状态整体的运行趋势，两种数据结合分析更全面、精确；本系统的数据处理根据用户需求选择重要度高的进行处理，减轻了对大量数据处理的负担；本系统的采用阈值自动修正的方法，使用户自定义的阈值参数更加的精确，提高数据融合的准确性。

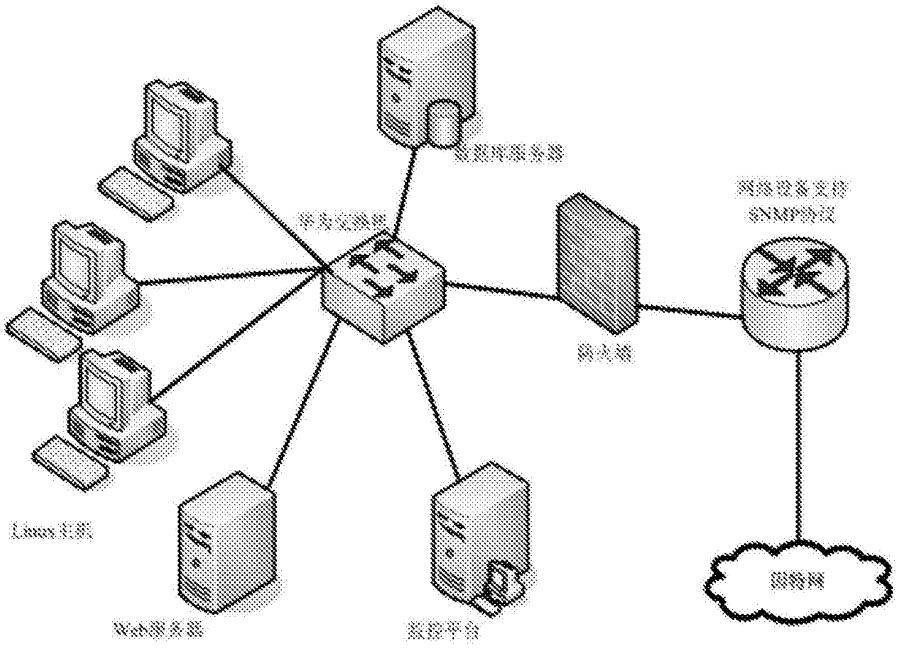


图1

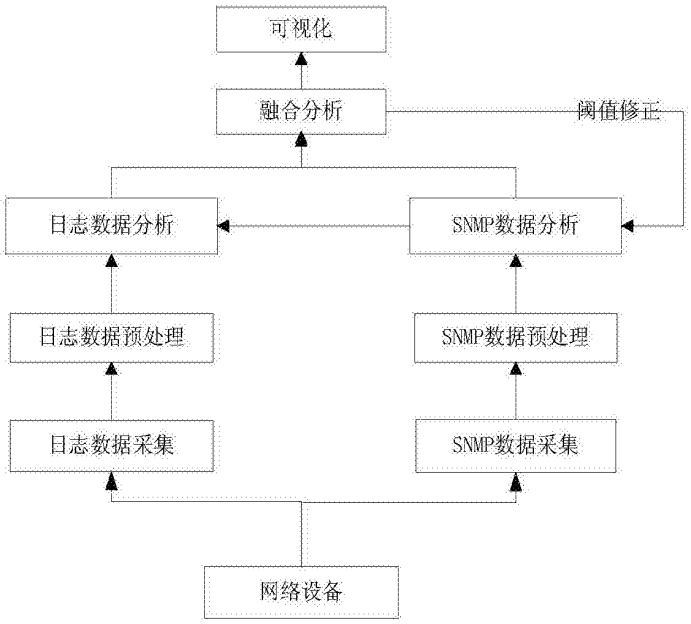

可视化
融合分析
阈值修正
日志数据分析
SNMP数据分析
日志数据预处理
SNMP数据预处理
日志数据采集
SNMP数据采集
网络设备

图2

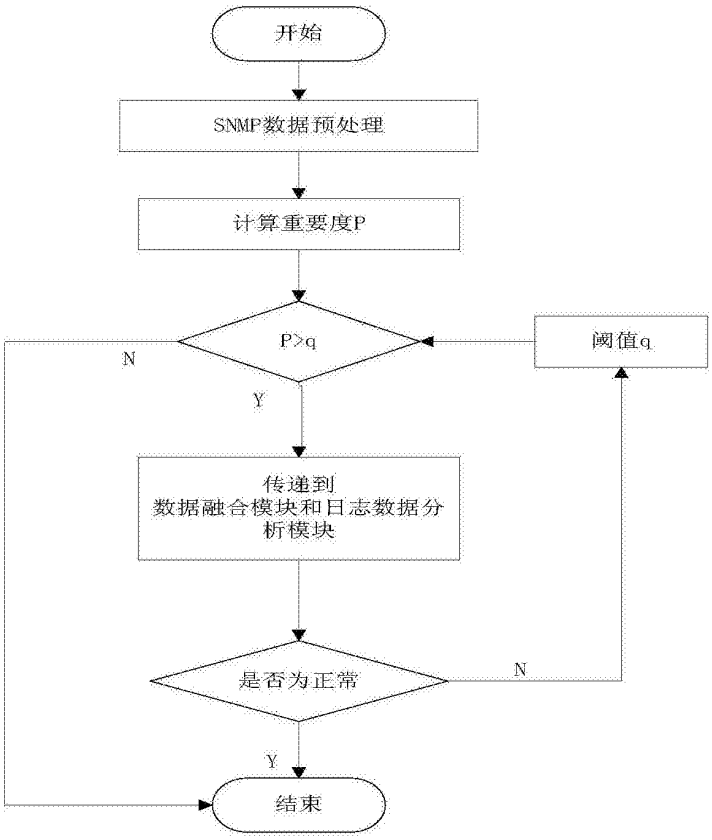

开始
SNMP数据预处理
计算重要度P
P>q
阈值q
N
Y
传递到
数据融合模块和日志数据分析模块
是否为正常
N
Y
结束

图3

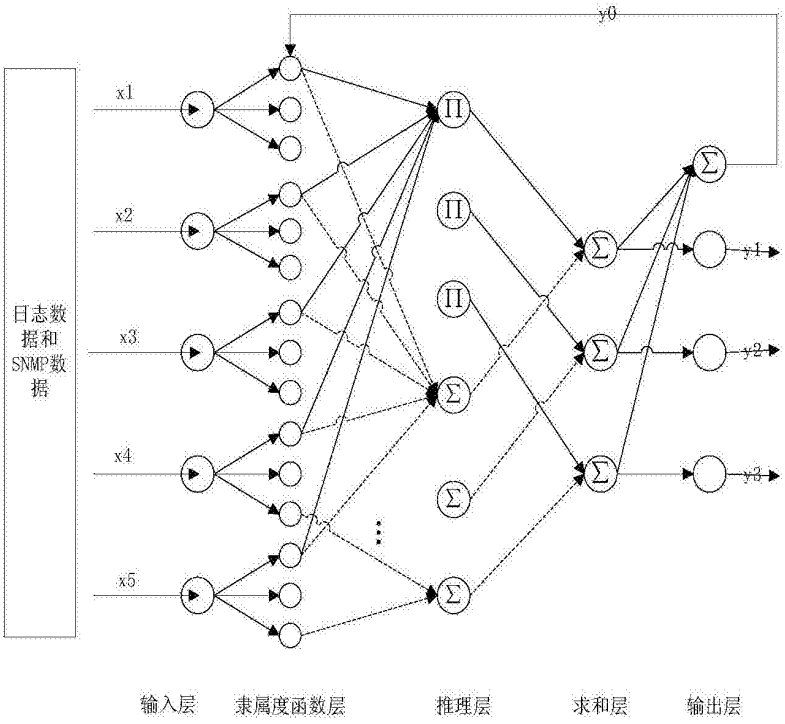


图4