

网络防火墙内部数据过滤保护算法

彭兆军

(郑州财经学院 网络中心,河南 郑州 450005)

摘 要: 提出一种基于群体信任的网络防火墙内部数据过滤保护算法,首先把网络防火墙内部数据库模型化为一棵层次树,网络防火墙内部危险数据获取问题就转化为树的遍历问题.通过对树中的属性排序,缩小查询空间.根据计算树中属性值相关度提高网络防火墙内部危险数据提取的准确度和提取效率.然后通过计算数据特征,确定流经网络防火墙的数据信任度,完成对目标数据的信任评价,进而实现网络防火墙内部数据过滤保护.通过仿真实验证明,所提方法能够有效提高网络防火墙过滤保护的准确性,提高网络防火墙内部数据过滤保护效率,具有较强的实用性.

关键词: 网络防火墙;数据过滤保护;算法

中图分类号: TP393

文献标识码: A

文章编号: 1000—7180(2017)11—0131—04

Network Firewall Internal Data Filtering Protection Algorithm

PENG Zhao-jun

(Network Centre,Zhengzhou Institute of Finance and Economics,Zhengzhou 450005,China)

Abstract: In this paper, an internal data filtering and protection algorithm based on community trust is proposed. Firstly, the internal database of network firewall is modeled as a hierarchical tree, and the problem of dangerous data acquisition in network firewall is transformed into tree traversal problem. By comparing the attribute of the tree, the query space is reduced. The accuracy and efficiency of the extraction of the dangerous data within the network firewall are improved by calculating the correlation of the attribute values. Then, the data trust of the network firewall is determined by calculating the data characteristics. The trust of the target data evaluation, and then to achieve the network firewall internal data filtering protection. The simulation results show that the proposed method can improve the accuracy of network firewall filtering protection effectively, improve the efficiency of filtering and protecting the data within the network firewall, and has strong practicability.

Key words: network firewall; data filtering protection; algorithm

1 引言

随着 Internet 网络的广泛应用及 Internet 技术的普及,人们通过把内部网连接到 Internet 网络中,从而获取需要服务,然而网络安全问题随之而来,受到人们的广泛关注.网络防火墙作为 Internet 网络 and 用户之间的连接通道,在过滤危险数据、保护用户网络安全方面存在重要作用,人们对防火墙要求不断增加^[1].由于网络防火墙具有的实时性、准确性,对此提出了一种基于 Filter-Hook 的防火墙技术,

其通过对数据包中的关键部分与预先设定的数据库进行对比,使通过防火墙数据能够根据规定的特性进行监控,把防火墙数据与特性信息进行比较,如果能够匹配则允许数据通过,反之就丢弃,从而提高防火墙过滤范围和准确度^[2],这一方法是网络防火墙进行数据过滤保护的重要手段,引起了专家学者的广泛关注,这一课题也成了业内人士研究的重要课题,产生了许多研究成果.

文献[3]提出了一种基于自学习 K 近邻的网络防火墙数据过滤算法.应用 KNN 方法对未知数据

样本进行精确匹配,以排除安全数据的误判结果,同时结合用户对危险数据的处理,自动调整过滤集合.该方法对危险数据处理方面存在优点,但也有可能把安全数据误判为危险数据.文献[4]提出一种以贝叶斯原理和人工免疫的仿生机理为基础的网络防火墙内部数据过滤保护方法,通过将贝叶斯与人工免疫相结合,设计和实现了一种基于贝叶斯和人工免疫的网络防火墙内部数据过滤算法,并利用现有的网络数据库进行实验,该方法不能及时地对数据库进行更新,导致其过滤危险数据时不能把全部的危险数据过滤出来.文献[5]提出了一个基于案例推理的防火墙网络内部数据过滤算法,将用户对数据的评价定义为案例,采用归一化的 Euclidean 距离,计算数据之间的相似度.在一个网络上进行了不同过滤算法的 recall 值对比.

针对上述问题,本文提出基于群体信任的网络防火墙内部数据过滤保护算法,并进行实验对比分析.通过仿真实验证明,所提基于群体信任的网络防火墙内部数据过滤保护算法能够有效提高网络防火墙过滤保护的准确性,提高网络防火墙内部数据过滤保护效率.

2 网络防火墙内部数据过滤保护算法

网络防火墙通过对流经它的数据进行扫描,从而过滤掉危险数据,使计算机免遭危险病毒或攻击的侵害从而保证计算机安全.

2.1 网络防火墙内部数据提取

网络内部数据提取是在数据经过网络防火墙时,对网络中的异常数据进行提取,其具体流程如图 1 所示.

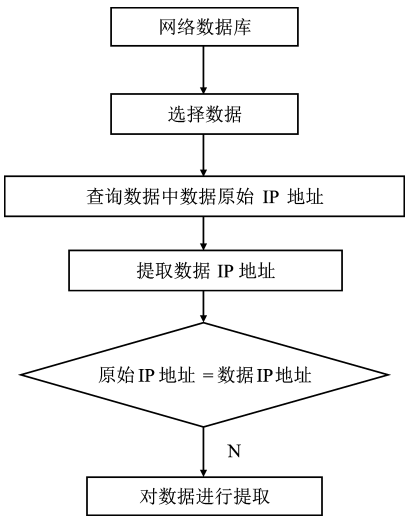


图 1 网络防火墙内部数据提取流程

网络防火墙内部数据提取时,先选择流经网络防火墙的数据,查询数据的原始 IP 地址和数据 IP 地址,将地址进行对比,如果地址不一致,则说明数据存在危险.

一个网络防火墙内部数据库 T 被划分成若干个非空子集 $\{P_1, P_2, \dots, P_m\}$, 网络防火墙内部数据库 T 中每个元组属于其中的某个子集 P_i , 这些非空子集就构成了网络防火墙内部数据库 T 的覆盖 $P(T)$, 如果满足式(1), 则称覆盖 $P(T)$ 是网络防火墙内部数据库 T 的划分, 每个子集 P_i 称为一个划分块.

$$P_1 \cup P_2 \cup \dots \cup P_m = T, P_i \cap P_j \neq \varphi \quad (1)$$

假设 k 为查询接口的常量, 只有满足式(2) 时, 每个划分块被称为有效块.

$$P_i \in P(T) (1 \leq i \leq m), |P_i| \leq k \quad (2)$$

设定一个网络防火墙内部数据库层次树如图 2 所示, 其中 a, b, c 分别代表网络防火墙数据库的查询属性. $[a, b, c]$ 表示网络防火墙数据库查询属性的序列, 首先, 利用 va_1 作为数据库数据查询条件, 对网络防火墙内部数据进行查询, 得到 $3k$ 个满足网络防火墙数据库条件的查询元组, 说明这种查询方式为上溢查询, 继续沿着层次树往下查询, 通过加入新的查询条件 vb_1 和 vc_1 , 组成新的查询序列 (va_1, vb_1, vc_1) , 利用该查询序列, 得到网络防火墙数据库查询节点 P_1 , 继续利用该查询序列进行遍历查询, 还可以得到网络防火墙数据库查询节点 P_2 , 甚至可以找到节点 P_3 和 P_4 .

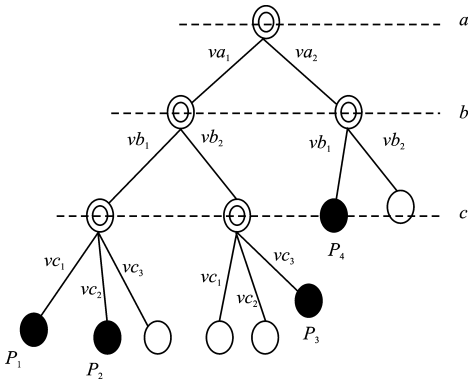


图 2 网络防火墙内部数据库层次树

数据的提取代价表示数据提取时需要与数据库服务器交互的次数, 则数据库的提取效率表达式为 $efficiency(A_q) =$

$$\frac{|PV(T)|}{\sum_{i=1}^m |num(q_i)/\tau| + |QO(A_q)| + |QI(A_q)|} \quad (3)$$

式中, A_q 代表查询属性集; $PV(T)$ 代表数据库表 T

的有效划分; $QO(A_q)$ 代表网络防火墙内部数据库上溢查询; $QI(A_q)$ 代表下溢查询; $PV(T)$ 代表有效网络防火墙数据节点查询; $q_i \in QV(A_q), num(q_i)$ 为满足查询 q_i 的元组数。

从上述论证可以看出,网络防火墙数据库查询空间与网络防火墙数据库数据提取效率成反比,与网络防火墙数据库数据提取代价成正比。所以想要提高网络防火墙数据库查询效率,从而降低网络防火墙数据库数据提取成本,就需要缩小网络防火墙数据库查询空间。

根据给定网络防火墙数据库查询属性集 $A_q = \{a_{q1}, a_{q2}, \dots, a_{qn}\}$, 设定每个属性的值域 Dom_i , 其大小关系为 $|Dom_1| \leq |Dom_2| \leq \dots \leq |Dom_m|$, 当拥有最小查询空间是, 层次树属性值域从大到小排列, 用公式表示为

$$\min(|Q(A_q^m)|) = \sum_{l=1}^m \prod_{i=1}^l |Dom_i| \quad (4)$$

$$|Dom_1| \leq |Dom_2| \leq \dots \leq |Dom_m|$$

通过上述方法, 缩小了查询空间, 然后通过同一网络防火墙的样本数据库获取属性值。

通过互信息的方法来实现网络防火墙数据库属性值和网络防火墙内部数据层次树上溢查询之间的相关度量化。设定 a_i 表示给定网络防火墙数据库属性实例, $q_i = (v_1, v_2, \dots, v_i)$ 表示该网络防火墙内部数据库层次树属性实例的一个上溢查询, $v_{(i+1),j}$ 表示该数据库属性实例 a_{i+1} 中的每一个属性值, $I(v_{(i+1),j})$ 表示查询 q_i 相关度值。最后利用 $I(v_{(i+1),j})$ 值, 通过比较 $I(v_{(i+1),j})$ 值大小, 对网络防火墙内部数据库层次树的属性值域进行从大到小排列, 并选择属性值域方便下一次查询。

设定一个网络防火墙本地数据层次树库, 假设 $q_i = (v_1, v_2, \dots, v_i)$ 表示这个数据库层次树中一次上溢查询, a_{i+1} 表示上溢查询的下一层属性节点, 则属性值 $v_{(i+1),j}$ 与查询 q_i 的相度计算公式可以表示为

$$I(v_{(i+1),j}, q_i) = \frac{P(v_i, v_{(i+1),j} | DB_{local})}{P(v_i | DB_{local}) \cdot P(v_{(i+1),j} | DB_{local})} \quad (5)$$

根据上文, 网络防火墙内部数据提取首先利用网络防火墙内部数据库属性的值域对网络防火墙数据库属性进行排序, 然后通过将网络防火墙内部数据库转化为网络防火墙内部数据库层次树的节点并利用遍历算法进行查询。根据与溢查询的相关度大小将网络防火墙内部数据库动态获取的网络防火墙数据的属性值进行从大到小排序。最后利用函数求

取属性值, 当属性值不唯一时, 选择具有最大 $I(v_{(i+1),j}, q_i)$ 的属性值与上溢查询组合, 形成层次树的下一次查询序列, 从而完成网络防火墙内部数据提取。

2.2 网络防火墙内部数据过滤保护算法

DC_i 和 DC_j 分别代表给定 2 个邻近网络防火墙内部数据节点 i 和 j 同一时段数据特征向量, $DC_i = (ex_i, en_i, he_i)$, $DC_j = (ex_j, en_j, he_j)$, TR_{ij} 代表网络防火墙内部数据节点 i 与数据节点 j 该时段相对信任, 表示为

$$TR_{ij} = \begin{cases} (1 - \frac{||ex_i| - |ex_j||}{|ex_i| + |ex_j|}) \cdot sim(DC_i, DC_j), & |ex_i| + |ex_j| \neq 0 \\ sim(DC_i, DC_j), & ex_i = 0, ex_j = 0 \end{cases} \quad (6)$$

相对信任具有阈值 $TR_{ij} \in [-1, 1]$; $TR_{ij} = TR_{ji}$; $C_i = C_j$ 时, $TR_{ij} = 1$; C_i, C_j 和代表数字特征; ex 为论域空间分布的期望; en 代表熵, 表示定性概念的可度量粒度; he 代表超熵, 表示熵的不确定性度量, 即熵的熵。

在上式中, 2 个节点相对信任的依据主要是网络防火墙内部数据特征向量中期望的值。在网络防火墙内部当 2 个节点的数据特征向量期望均为 0 时, 这 2 个节点数据特征向量的相似度则可以用传统云相似度算法进行计算, 将其计算结果作为相对信任; 网络防火墙内部数据库的 2 个节点数据特征向量中的期望不全为 0 时, 则相对信任可以表示为 $(1 - \frac{||ex_i| - |ex_j||}{|ex_i| + |ex_j|}) \cdot sim(DC_i, DC_j)$, 可以知道 $ex_i = ex_j$ 时, 即 2 个节点数据特征向量中期望相同时, 相对信任可以表示为 $sim(DC_i, DC_j)$, 当 $sim(DC_i, DC_j) = 1$ 时, 相对信任可用 $(1 - \frac{||ex_i| - |ex_j||}{|ex_i| + |ex_j|})$ 计算。

假设数据节点 i 在某段时间流经网络防火墙, 数据的群体信任 CT_i 表示在一定时间段内, 该网络防火墙内部数据库数据节点与数据库内其所有邻近数据节点相对信任的平均值, 表示为

$$CT_i = \frac{1}{M-1} \sum_{k=1}^{M-1} RT_{ki} \quad (7)$$

式中, i 表示网络防火墙内部数据库的目标节点编号; M 表示网络防火墙内部数据库目标节点 i 附近的节点总数(包括自身在内), 其中 $M \geq 2, M \in N$, RT_{ki} 是网络防火墙内部数据库节点 k 与节点 i 之间数据的相对信任值。网络防火墙内部数据库目标节

点数据特征与其他节点数据特征越接近,则说明数据越普遍,网络防火墙内部数据库群体任度越高;反之,网络防火墙内部数据库目标节点数据特征越特殊,则群体信任度越低. 通过上述算法,完成网络防火墙内部数据过滤保护.

3 实验结果与分析

为了证明本文提出的网络防火墙内部数据过滤保护算法的有效性和可行性,进行了一次仿真实验,实验是在 Windows 7 Ultimate 的操作系统下进行的,CPU 型号是 3.2 GHz 的 Intel Core I3,运行平台为 Microsoft Visual Studio. NET 2010 为运行平台完成的. 通过本文和文献[6]提出的基于多重半脆弱水印的网络防火墙内部数据过滤保护算法方法进行对比,完成本次实验.

规定一个数据库,有若干个危险数据,通过本文所提方法和文献[6]所提方法进行过滤,结果如表 1 所示.

表 1 本文所提方法和文献[6]方法过滤结果对比

序号	危险数据数	本文方法 过滤数量	文献[6]方法 过滤数量
1	5	5	4
2	15	13	11
3	30	27	24
4	40	35	31
5	50	45	40
6	60	52	45
7	80	69	65
8	100	81	75

设定过滤率表示过滤出的危险数据数量与危险数据总量的比值,对本文和文献[6]所用方法过滤率结果可表示成图 3.

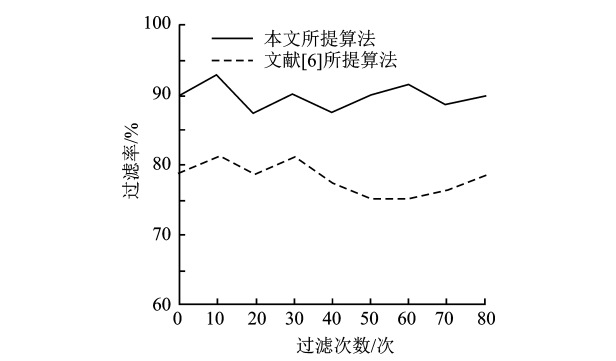


图 3 本文和文献[6]过滤结果对比

从图 3 可以看出,本文所提算法过滤率明显高于文献[6],说明本文过滤结果较好.

设定过滤准确率为误判数量与全部过滤数量的比值,表 2 表示同一段时间内本文和文献[6]过滤准确率的比较,将其转换成图表形式为图 4,通过图表可以看出,本文所提算法的准确率折线位于文献[6]所提方法折线上方,说明本文所提算法在准确率方面较文献[6]有明显提高,而且本文所提算法对比文献[6]准确率浮动较小,说明本文所提算法稳定性较好.

表 2 本文与文献[6]所提算法准确程度对比

过滤总数	本文方法 过滤正确数	文献[6]方法 过滤正确数
50	45	41
100	90	85
150	137	128
200	183	172
250	227	215
300	270	253
400	358	339
500	452	433

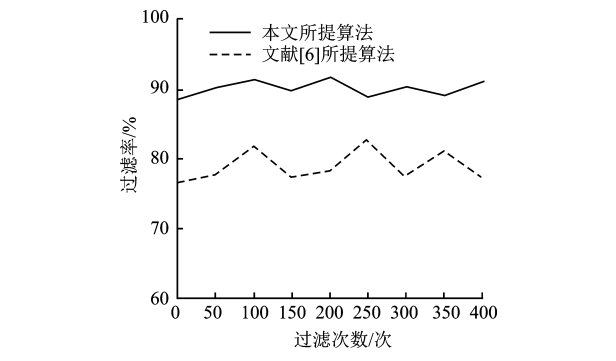


图 4 本文和文献[6]所提算法准确率对比

通过上述实验,说明本文所提算法在过滤范围,准确度等方面较其他方法有明显提高,具有可用价值.

4 结束语

网络防火墙是网络安全系统的重要组成部分,做好网络防火墙内部数据过滤保护工作对防火墙技术的发展有重要影响,本文提出的基于群体信任的网络防火墙内部数据过滤保护算法能够有效提高过内部数据的准确度和过滤范围,具有较强的有效性和实用性.

参考文献:

[1] 杨海月,朱玉婷,施化吉,等. 基于用户信任度和社会相似度的协作过滤算法[J]. 电子技术应用,2016,42(1): 100-103.

- tinuous IOP monitoring [J]. IEEE Trans. Biomed. Eng., 2013, 60(1):250-256.
- [3] Donida A, Di Dato G, Cunzolo P, et al. 21. 7A0. 036mbar circadian and cardiac intraocular pressure sensor for smart implantable lens[J]. IEEE Int. Solid-State Circuits Conf, 2015(22/26):392-394.
- [4] Boujamaa E M, Alandry B, Hacinef B. Low power interface circuit for resistive sensors with digital offset compensation[C]//IEEE Int. Symp. Circuits and Systems, IEEE, 2010:3092-3095.
- [5] 孙彤. 低功耗逐次逼近模数转换器的研究与设计[D]. 北京:清华大学, 2007.
- [6] Pang Wen-Yi, Wang Chao-Shiun, Chang You-Kuang, et al. A 10-bit 500-KS/s Low Power SAR ADC with splitting comparator for bio-medical applications[C]//IEEE Asian Solid-State Circuits Conference, IEEE, 2009:149-152.
- [7] Allen P E, Holberg D R. CMOS. 模拟集成电路设计[M]. 冯军, 李智群, 译. 北京: 电子工业出版社, 2011.
- [8] 林武平, 郭良权, 于宗光, 等. 新型高速低功耗动态比较器[J]. 半导体技术, 2008, 33(12): 1119-1122.
- [9] 侯斌, 莫亭亭. 用于 16 bit SAR ADC 的高精度比较器的设计[J]. 微电子学与计算机, 2016, 33(7):15-18.
- [10] Alan Hastings. 模拟电路版图的艺术(2 版)[M]. 北京: 电子工业出版社, 2007.

作者简介:

马志峰 男, (1973-), 博士, 讲师. 研究方向为视频图像处理、小波变换、模式识别.

李熙泽(通讯作者) 女, (1992-), 硕士研究生, 讲师. 研究方向为模拟集成电路. E-mail: 771410107@qq. com.

张 锋 男, (1977-), 博士, 研究员. 研究方向为高速低功耗存储器的设计、高速低功耗接口电路的设计.

陈铖颖 男, (1982-), 博士, 副研究员. 研究方向为高性能模拟/射频/混合信号 CMOS 集成电路设计, 低功耗、高精度、低噪声传感器接口及信号处理电路设计, 高可靠加固标准单元库设计.

(上接第 134 页)

- [2] 章志明, 李萍, 熊小勇. 一种有效的无线传感器网络虚假数据过滤方案[J]. 计算机工程与应用, 2015, 51(24):78-85.
- [3] 顾李华. 退墒曲线在墒情数据过滤及预报方面的应用[J]. 水文, 2016, 36(1):80-83.
- [4] 刘华富, 刘志雄, 黎梨苗. 适用于虚假数据过滤的最优传感器覆盖研究[J]. 计算机应用研究, 2016, 33(10): 3147-3150.
- [5] 冯锋, 吴杰. 基于 Bloom filter 的 RFID 中间件数据过滤算法研究[J]. 计算机应用研究, 2015, 32(5): 1441-1444.
- [6] 刘华富, 刘志雄, 黎梨苗. 适用于虚假数据过滤的最优传感器覆盖研究[J]. 计算机应用研究, 2016, 33(10): 3147-3150.

作者简介:

彭兆军 男, (1980-), 硕士, 讲师. 研究方向为网络安全及管理. E-mail: pengzhaojunjob@aliyun. com.