

分层网络伪装入侵实时高速检测方法研究

李军华

(江苏理工学院,江苏 常州 213001)

摘 要: 提出一种分层网络伪装入侵实时高速检测方法,该方法首先对分层网络数据进行预处理,采用蚁群算法对分层网络伪装入侵实时高速检测进行特征选择,然后结合特征选择的分析进行网络伪装入侵实时高速分层聚类方法完成对分层网络伪装入侵实时高速检测方法的研究.仿真实验表明,利用分层网络入侵检测方法能有效地提高网络入侵实时高速检测准确性.

关键词: 分层网络;伪装入侵;实时高速检测

中图分类号: TP393

文献标识码: A

文章编号: 1000—7180(2017)10—0123—04

The Layered Network Invasion of High-Speed Real-Time Detection Method Research of Disguise

LI Jun-hua

(Jiangsu University of Technology, Changzhou 213001, China)

Abstract: Put forward a kind of hierarchical network intrusion real-time high-speed detection method of disguise. On hierarchical network data preprocessing in the first place, the method using ant colony algorithm of hierarchical network invasion of high-speed real-time detection feature selection of disguise, camouflage and combining with the analysis of feature selection for network invasion of high-speed real-time hierarchical clustering method for layered network mask invasion of high-speed real-time detection method research. Simulation experiments show that using the hierarchical network intrusion detection method can effectively improve the accuracy of the real-time high-speed network intrusion detection.

Key words: hierarchical network; camouflage invasion; real-time high-speed detection

1 引言

在国内,随着计算机网络技术的不断发展,网络恶意入侵也在不断的增加,计算机网络的安全问题已逐渐影响到人们的日常生活等诸多方面^[1]. 计算机网络具有开放性与共享性的特点,假设计算机网络受到恶意的攻击,计算机网络安全将会受到极大的破坏,有可能造成数据丢失以及用户的权限被篡改和网络瘫痪等一些恶劣的后果^[2]. 在这种情况下,如何实时高速的对网络进行检测成为了该领域当前急需解决的主要问题^[3]. 而有效的对分层网络进行入侵实时高速检测是解决上述问题的有效途径^[4-5], 引起了许多专家关注和重视. 由于分层网络伪装入侵实时高速检测方法具有广阔的发展空间^[6], 因此,

已成为网络入侵检测研究的核心,拥有较大的发展潜力,同时也取得了一定的成果^[7].

文献[8]提出了一种神经网络入侵检测进化方法,首先利用遗传算法对神经网络结构进行优化,结合优化后的神经网络进行网络入侵检测预警,利用神经网络进化方法的不断演变,寻找最优的神经网络结构,当进化神经网络系统进行正常的工作模式后,可对偏离正常工作的事情做出反应,但该方法存在安全性较低的问题. 文献[9]提出了一种基于混合入侵检测技术的网络入侵检测模型的方法. 对于同一种行为,异常检测的结果与误用检测的结果是不同的,利用跟踪算法能有效地解决异常检测的结果与误用检测结果不相同的问题,利用数据挖掘方法建立网络正常的行为轮廓库,最后采用全序列比较法实现网络异

常引擎,但该方法存在耗时较长的问题.文献[10]提出了基于聚类指导的网络入侵检测方法.在带有标记的训练集上对数据进行聚类,以聚类的结果作为网络分类模型对数据进行分类,该方法对参数和数据的输入顺序具有稳定性,可对分类模型进行增量更新,具有对未知入侵有检测能力的检测方法,但该方法存在不能快速进行检测的问题.

针对上述的问题,提出一种分层网络伪装入侵实时高速检测方法.该方法首先对分层网络进行特征选择,结合特征选择的分析进行网络入侵分层聚类.仿真实验表明,利用分层网络入侵检测方法能有效地提高网络入侵实时高速检测的准确性.

2 分层网络伪装入侵实时高速检测方法研究

2.1 利用蚁群算法对网络伪装入侵进行特征选择

在进行分层网络伪装入侵实时高速检测特征选择时,将分层网络伪装入侵实时高速检测特征作为蚂蚁访问一个地点,将特征优化问题转化为路径搜索问题.

网络入侵检测特征是蚂蚁必须经过的节点,当每完成一次循环,蚂蚁就要经历全部网络入侵检测特征点.每个特征都有可选择的概率,蚂蚁所经过的特征节点越多,网络特征的选择概率就越大,相对被选择的可能性就大.蚂蚁从网络特征 i 转移到 j 概率为

$$p_{ij}^k = \frac{\tau_{ij}^\alpha(t) \eta_{ij}^\beta(t)}{\sum_s \tau_{is}^\alpha(t) \eta_{is}^\beta(t)} \quad (1)$$

式中: η_{ij} 为启发因子,主要取决于网络入侵检测率, η_{ij} 因子越大,蚂蚁转移到网络特征 j 的概率就越大, $\tau_{ij}(t)$ 为时刻 t 网络特征 i 到网络特征 j 的信息素.

在网络状态转移概率中; α 为网络信息素的权重; β 为网络启发因子的权重, α 为一常量, β 的值为

$$\beta = \beta_0 \left(1 - \frac{n}{N_{\max}}\right) \quad (2)$$

式中: n 表示迭代的次数, β_0 表示网络启发因子权值的初值, $N_{\max}$ 表示网络最大的循环次数.

假设,蚂蚁经过 k 个重要的网络特征搜索,得出 k 个重要的特征,为避免分层网络实时高速入侵检测无关的特征或冗余特征的混淆,需在 k 个特征中搜索出最优的特征子集, u_i 表示一个特征子集, u_j 表示最优的特征子集,则第 j 个蚂蚁 S_j 适应度的函数值 $F(S_j)$ 满足:

$$F(S_j) = \min_i (S_i, \forall i) \quad (3)$$

式中, $S_i = S_m \cup u_i \cup \{f_n\}$; S_m 为经过 m 只蚂蚁; $\{f_n\}$ 为迭代 n 次剩余的特征; n 为迭代的次数.

蚁群算法利用信息来实现反馈机制,网络信息素更新机制包含两类:局部信息与整体的信息.整体信息可加快蚁群算法搜索速度,获取全局最优特征概率变大,完成搜索后,网络路径上信息浓度需重新更新,表达式为

$$\begin{cases} \tau_{ij}(n+1) = \rho \cdot \tau_{ij}(n) + \sum_k \Delta \tau_{ij}^k \\ \Delta \tau_{ij}^k = \frac{Q}{F(s_k)} \end{cases} \quad (4)$$

式中, n 表示迭代的轮数; ρ 表示网络信息素残留的因子; k 表示蚂蚁的编号; $F(s_k)$ 表示网络的适应度值; Q 表示网络信息素增长的浓度. ρ 对蚁群算法收敛性影响较明显, ρ 值越大,收敛速度就越慢,不易陷入局部最优, ρ 值越小,蚁群算法的收敛速度虽然快但容易陷入局部最优. ρ 值设定为

$$\rho = \rho_0 \left(1 - \frac{n}{3N_{\max}}\right) \quad (5)$$

式中, ρ_0 表示网络信息素残留因子的初始值.本文所适应的函数,具体如下:

$$F(S_{opt}) = \alpha \cdot \gamma_R(D) + \beta \cdot \frac{|C| + |B|}{C} \quad (6)$$

式中, $\gamma_R(D)$ 表示相对维度 D 条件属性集; R 表示分类的质量; $|R|$ 表示位置的特征子集的长度; $|C|$ 表示网络特征的总数; α 和 β 表示对应分类质量的重要性与子集长度的参数, $\alpha \in [0, 1]$ 且 $\beta \in [1 - \alpha]$, 网络随机数 R 与预定常数相关联.

从网络信息素更新的规则可知,网络适应度越小的函数特征子集,路径上的网络信息素的浓度就越大,则会吸引更多的蚂蚁对该网络路径进行搜索,为了对最优路径进行强化影响,对网络信息素进行附加奖励,即:

$$\tau_{ij}^* = \tau_{ij} + \frac{Q}{F(s_{opt})} \quad (7)$$

式中, $F(s_{opt})$ 为本轮最优特征子集适应度函数.

将特征集 F 分为不相交的两个子集 F_S 与 F_N , F_S 包括网络特征集 F 中所有符号特征,而 F_N 则包括了特征集 F 中所有数值特征.因不同的网络入侵采用不用不同的特征,因此不能判定哪些特征在网络入侵时有重要地位,在 F_S 特征子集间差异计算中采用欧氏距离,对于网络入侵事件 e_i 与 e_j 中 F_S 距离定义由式(8)表示:

$$ds(e_i^s, e_j^s) = \sqrt{\sum_{k=1}^{m_1} (f_{ik} - f_{jk})^2} \quad (8)$$

为使特征选择中包括出现的较少特征值事件能远离其他网络入侵事件,对于网络入侵事件 e_i 与 e_j , 网络入侵事件中 F_N 的数值特征子集间差异计算公

式如下:

$$dn(e_i^N, e_j^N) = \sum_{k=1}^{m_2} \frac{(n_{fik} + n_{fjk})}{n_{fik} * n_{fjk}} \tag{9}$$

式中, n_{fik} 与 n_{fjk} 分别表示网络入侵事件集 $E = \{e_1, e_2, \dots, e_n\}$ 中相应入侵特征 $F_k (F_k \in F_s)$ 取 $f_i(e_i(F_k))$ 值与 $f_j(e_j(F_k))$ 值的个数。

依据式(8) 与式(9), 最终定义任意两个网络入侵事件数据对象间的差异的计算公式为

$$d(e_i, e_j) = \alpha ds(e_i^s, e_j^s) + (1 - \alpha) dn(e_i^N, e_j^N) \tag{10}$$

式中, α 为网络入侵权重因子, 利用 α 可更好特征子集与数值特征子集在计算网络对象间差异时有不同作用。保证了分层网络伪装入侵实时高速检测准确性。

2.2 分层网络伪装入侵实时高速检测方法

结合上述分层网络伪装入侵实时高速检测特征选择的数据, 基于分层聚类方法, 对分层网络伪装入侵实时高速检测进行优化。利用分层聚类方法的对偶原则, 将进行访问的数据的对偶为进行优化的函数:

$$Q(A) = \sum_{i=1}^N A_i - \frac{1}{2} \sum_{i,j=1}^N A_i A_j (x_i, x_j) \tag{11}$$

式中, $Q(A)$ 表示网络访问数据优化的分类函数, 依据分层网络入侵检测构建函数, 具体构建的分层网络入侵检测表示为

$$\begin{aligned} f(x) &= \text{syn}(R - A(x, x)) \\ &= \text{sgn}(R(x_i \cdot x) + b) \end{aligned} \tag{12}$$

式中, b 表示分层网络入侵检测的分类阈值。依据网络数据计算得出优化分类的函数, 对分层网络入侵检测进行优化训练:

$$A(x, x) = \sum_j Q(A_j)^2 \beta_j \tag{13}$$

$$R = A(x, x) - 2 \sum_j \beta_j A(x_j, x) \tag{14}$$

$$A(x_i, x_j) = \tan Rh(a(x_i \cdot x_j) + c) \tag{15}$$

式中, R 表示进行优化分类得出的网络数据; $A(x_i, x_j)$ 表示利用分层聚类法依据网络数据得网络入侵检测模型, 完成对分层网络伪装入侵高速检测研究。

3 实验结果与分析

为了证明分层网络伪装入侵实时高速检测方法的有效性, 需要进行一次仿真实验。实验平台采用的是 Matlab 2016b, 基于 KDD100 网络数据库提供的网络数据样本对此进行验证。选取网络数据库中 288 598 组无标签的数据, 247 016 组有标签的数据

分类器和 247 015 组数据进行测试。其中数据网络中采用固定的 40 组特征数据, 因此网络的输入节点为 40 个, 输出的节点数目则将在实验结果中进行讨论。检测率、误检率以及检测精度分别定义为

$$\begin{aligned} AN &= DB_{ae} / DB_{ai} \times 100\% \\ EBN &= DBE_{ae} / DD_{ai} \times 100\% \\ GNF &= DG_{ae} / GG_{ai} \times 100\% \end{aligned} \tag{16}$$

式中, DB_{ae} 为网络数据次数; DB_{ai} 为网络异常次数; DBE_{ae} 为网络误检次数; DD_{ai} 为网络全部异常次数; DG_{ae} 为网络入侵攻击数; GG_{ai} 为网络全部攻击数。

表 1 表示网络数据各次迭代过程中选取网络输出节点数和对应误差。由表 1 中能看出, 隐层数低的时候误差率大, 经过 3 次迭代后, 得分层网络输出的节点数。

表 1 网络隐层数和对应的误差

第一轮		第二轮		第三轮	
隐层数	误差/%	隐层数	误差/%	隐层数	误差/%
5	0.123	19	0.019	18	0.019
15	0.032	21	0.024	20	0.016
20	0.019	24	0.023	22	0.024
25	0.023	26	0.025	24	0.022
30	0.028	30	0.028	26	0.023

图 1 表示本文方法和文献[8]、文献[9]分层网络伪装入侵实时高速检测的检测率对比。由图 1 能看出, 文献[8]、文献[9]中网络节点数据在 20 以上时, 检测率趋于稳定, 而本文方法在网络节点数据在 15 以上时渐渐趋于稳定, 由此可知本文方法的检测率更准确一些。

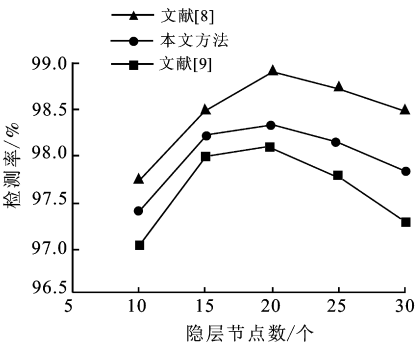


图 1 不同方法的检测率

图 2 表示本文方法和文献[8]、文献[9]分层网络伪装入侵实时高速检测的误检率对比。由图 2 可知, 文献[8]中检测的误检率浮动在网络节点数据 20 时最高, 然后慢慢降低, 文献[9]是在网络节点数据 15 时是达到最高, 然后较为平缓, 本文方法一直比较平缓、波动较小。由此可知本文方法的误检率更

好一些.

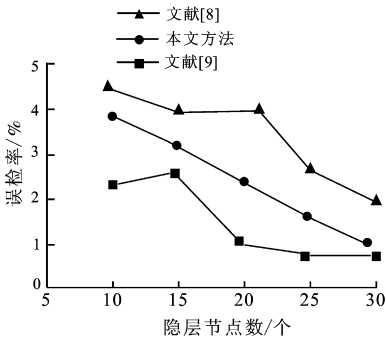


图 2 不同方法的误检率

图 3 表示本文方法和文献[8]、文献[9]分层网络伪装入侵实时高速检测的检测精度对比. 由图 3 能看出文献[8]和文献[9]随着网络节点数据的增加,网络入侵检测的精度也有所变化,而本文方法网络入侵检测的精度无论网络节点数据多还是少变化一直很均匀,由此可知,本文方法的网络入侵检测的精度较高.

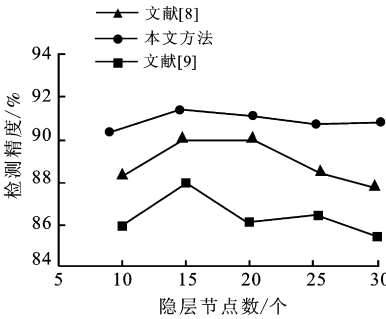


图 3 不同方法的检测精度

仿真实验证明,利用分层网络入侵检测方法能有效地提高网络入侵实时高速检测的准确性.

4 结束语

采用当前方法是利用遗传算法对神经网络结构进行优化,结合优化后的神经网络进行网络入侵检测预警,利用神经网络进化方法的不断演变,寻找最

优的神经网络结构,当进化神经网络系统进行正常的工作模式后,可对偏离正常工作的事情做出反应,但该方法存在安全性较低的问题. 为此,提出一种分层网络伪装入侵实时高速检测的研究方法. 仿真实验证明,利用分层网络入侵检测方法能有效地提高网络入侵实时高速检测的准确性,具有广泛的实用价值.

参考文献:

[1] 汪波,聂晓伟. 基于多目标数学规划的网络入侵检测方法[J]. 计算机研究与发展, 2015, 52(10): 2239-2246.

[2] 梁辰,李成海,周来恩. PCA-BP 神经网络入侵检测方法[J]. 空军工程大学学报(自然科学版), 2016, 17(6): 93-98.

[3] 杨宏宇,王信元. 基于链路覆盖的无线 Mesh 网络入侵检测方法[J]. 北京邮电大学学报, 2015, 38(6): 60-64.

[4] 张永良,张智勤,吴鸿韬,等. 基于改进卷积神经网络的周界入侵检测方法[J]. 计算机科学, 2017, 44(3): 182-186.

[5] 马勇. 模糊推理结合 Michigan 型遗传算法的网络入侵检测方案[J]. 电子设计工程, 2016, 24(11): 108-111.

[6] 钱亚冠,关晓惠. 网络入侵检测系统中的漂移检测[J]. 电信科学, 2015, 31(3): 67-73.

[7] 华辉有,陈启买,刘海,等. 一种融合 Kmeans 和 KNN 的网络入侵检测算法[J]. 计算机科学, 2016, 43(3): 158-162.

[8] 赵旭,王伟,陈亮. 网络入侵检测系统规则链表的优化研究[J]. 计算机工程与应用, 2015, 51(20): 91-96.

[9] 杜红乐,樊景博. 基于半监督模糊聚类的入侵检测[J]. 计算机工程与应用, 2016, 52(3): 96-99.

[10] 赵悦,程子傲,陈雷,等. 移动 Ad-Hoc 网络叶节点簇降低能源节约入侵检测[J]. 控制工程, 2016, 23(7): 1137-1142.

作者简介:

李军华 男, (1980-), 硕士研究生, 工程师. 研究方向为网络信息安全、网络管理与安全、虚拟化.
E-mail: ljh@jsut.edu.cn.