第37卷 第14期
Vol.37 No.14

计 算 机 工 程
Computer Engineering

2011年7月
July 2011

• 安全技术 •

文章编号：1000—3428(2011)14—0131—03 文献标识码：A 中图分类号：TP309

基于猴群算法的入侵检测技术

张佳佳，张亚平，孙济洲

(天津大学计算机科学与技术学院，天津 300072)

摘 要：针对入侵检测系统存在高漏报率的问题，提出一种基于猴群算法的入侵检测技术。利用猴群算法从网络审计数据 KDD99 数据集中生成一个分类的规则集合，采用支持度-置信度模型实现猴群算法的目标函数，以控制生成规则的质量，将动态生成的规则应用于基于规则的的入侵检测系统中。实验结果表明，基于猴群算法的入侵检测技术可改进生成规则的质量，提高入侵检测系统的检测率。

关键词：入侵检测；猴群算法；支持度-置信度；分类规则

Intrusion Detection Technology Based on Monkey Algorithm

ZHANG Jia-jia, ZHANG Ya-ping, SUN Ji-zhou

(School of Computer Science and Techonology, Tianjin University, Tianjin 300072, China)

【Abstract】For the current Intrusion Detection System(IDS) has high false negative rate, this paper presents an intrusion detection technology based on Monkey Algorithm(MA). It uses the MA to derive a set of classification rules from network data, KDD99 data set, and the support-confidence framework is utilized as fitness function to judge the quality of each rule. The generated rules are used to detect or classify network intrusions in a real-time environment. Experimental results show that the MA-based technology can improve the quality of generating rules, so that it can improve the performance of IDS.

【Key words】intrusion detection; Monkey Algorithm(MA); support-confidence; classification rule

DOI: 10.3969/j.issn.1000-3428.2011.14.043

1 概述

传统的信息安全技术采用严格的访问控制机制和数据加密策略，但随着攻击技术的日益成熟、攻击工具和手段的复杂多样化，这些被动的安全技术无法满足对安全高度敏感的网络服务的要求。入侵检测系统(Intrusion Detection System, IDS)通过分析网络数据和审计记录，识别系统中的攻击活动，采取相应的措施。IDS 作为主动的安全技术，已成为网络安全领域的研究热点。本文提出一种基于猴群算法(Monkey Algorithm, MA)的入侵检测技术，利用 MA 全局搜索速度快、精度高的特性，将分类规则应用到基于规则的入侵检测系统中，用以检测和分类网络入侵，达到入侵检测防御的目的。

2 相关技术

2.1 入侵检测技术

入侵检测通过对系统数据的分析发现非授权的访问和攻击行为。它通过从计算机系统中的日志文件、计算机网络中的若干关键节点等处收集相关信息，并分析这些信息，检查系统或网络中是否有违反安全策略的行为和遭到袭击的迹象。入侵检测系统是一种计算机软件硬件结合的系统，用于自动检测上述入侵行为，并收集入侵证据，为数据恢复和事故处理提供依据。

目前应用于入侵检测的先进技术包括神经网络、数据挖掘、计算机免疫学和遗传算法等。文献[1-2]将遗传算法等智能算法应用于 IDS 中，生成匹配异常连接的规则。基于专家系统的误用检测通过将安全专家的知识表示成 if-then 结构的规则，形成专家知识库，然后运用推理的算法检测入侵。它需要解决的主要问题是处理序列数据和知识库的维护，主要缺点是：难以科学地从各种入侵手段抽象出全面的规则化知识，需要处理的数据量过大，效率较低。

2.2 猴群算法综述

猴群算法[3]是一种模拟猴群爬山的优化种群智能算法，用于解决带有连续变量的全局数值最优化问题，这个算法主要包括攀爬过程、眺望过程、翻筋斗过程。猴群算法流程如图 1 所示。

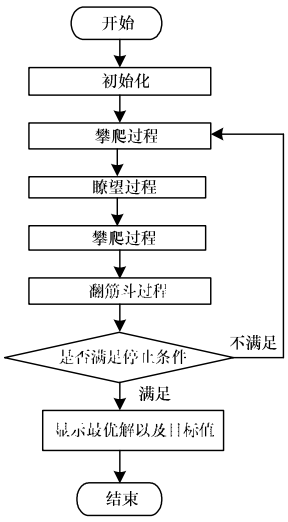

图 1 猴群算法流程

攀爬过程用于实现寻找局部最优解，通过步步迭代使猴子的位置从初始值向着能解决目标函数的新位置转移。在攀

基金项目：国家自然科学基金资助项目(60776807)；天津市应用基础及前沿技术研究计划基金资助重点项目(09JCZDJC16800)

作者简介：张佳佳(1986－)，男，硕士研究生，主研方向：信息安全；张亚平，副教授、博士；孙济洲，教授、博士

收稿日期：2011-01-06 **E-mail：**studentxmu@gmail.com

爬过程之后，每个猴子到达了自己所在区域的山顶，然后左右环顾周围是否有比当前更高的山峰，即更优的解，眺望过程为了找到优于当前解决方案的并接近目标值的点从而加快猴群的搜索过程。在反复经过攀爬过程和眺望过程之后，每个猴子都找到了自己所在区域的局部最优解，为了避免陷入无休止的循环，翻筋斗过程是为让猴群更快地转移到下一个搜索区域，重新开展以上过程。经过一定次数的循环或者达到了一定的终止条件后，算法终止，最优解或者近似最优解被找到。

3 基于猴群算法的入侵检测技术

入侵检测系统可以看作是一个基于规则的系统，而猴群算法等智能算法是为这个系统来产生准确的规则，输入到 IDS 中，以此来检测入侵。

为了实现基于猴群算法的系统，本文采用 KDD99 训练和测试数据集，KDD99 数据集具有良好的结构和可用性，是入侵检测评估采用的标准数据集。KDD99 包括 500 万条训练数据和 200 万条测试数据，包括 38 种攻击。每一条数据包括 41 维 TCP、时间、容量等特征，可以根据实际算法选取合适的特征属性。

基于猴群算法的入侵检测方法包括 2 个模块[4]，每个都工作在不同的阶段。在训练阶段，本文利用猴群算法在离线的环境从网络审计数据中产生一个规则集合，在入侵检测阶段，生成的规则用来在实时的网络环境中分类网络连接数据，检测入侵。一旦规则产生，入侵检测就变得很简单有效。

3.1 数据表示

猴群算法能产生针对网络流量的简单规则，使用这些规则来区分正常行为和异常行为。这些异常行为指的是入侵事件，规则可以以下列形式表示：

IF(条件) Then(行为)

如果规则是合理的，则系统产生预期结果。条件是指当前网络数据与入侵检测内规则的配匹。应用猴群算法最后的目的就是产生配匹异常数据的规则。这些规则能测试历史数据，并且能够过滤新的数据，发现可疑的网络流量。这里仅仅考虑每个连接明显的属性。举例说明，相关规则可定义为：

if{源 IP 地址为 108.11.??.??；目的 IP 地址：111.15.128.??；源端口号：10000；目的端口号为：80；连接时间：362 s；连接被创建人中断；使用的协议是 TCP；创建人发送了 320 bit 的数据；并且接收人发送了 3 445 bit 数据}

then {入侵类型是 DoS 攻击，应该断开连接}

3.2 猴群算法参数与目标函数

本文利用支持度-置信度模型[5]来检查猴群算法生成的规则，如果规则是：IF(A) Then(B)，则可以根据以下式子来决定生成规则的合适度，即猴群算法目标函数：

$support=|A \text{ and } B|/N$

$confidence=|A \text{ and } B|/|A|$

$F= w1\times support+w2\times confidence$

其中，N 是审计数据中所有网络连接的总数；$|A|$代表网络连接中仅仅匹配条件 A 的次数；$|A \text{ and } B|$是网络连接中匹配规则 IF(A) Then(B)的数目，权重 $w1$ 和 $w2$ 是用来控制这 2 个组合的平衡，默认值是 $w1$=0.2 和 $w2$=0.8。使用这样的权重控制函数的好处是当改变 $w1$ 和 $w2$ 时，该方法可以很容易地识别网络连接并且精准地对入侵类型进行分类。比如：$w1$=1 和 $w2$=0，代表前一种情况匹配入侵；$w1$=0 和 $w2$=1 是后一种情况。

3.3 具体算法流程

输入 网络数据 KDD99，迭代次数和种群规模

输出 一个分类的规则集合

1. 初始化猴子种群
2. W1 = 0.2，W2 = 0.8，T = 0.5
3. N = 训练集合的总记录数
4. For each 种群的每一个个体
5. A = 0，AB = 0
6. For each 训练集合的每一条记录
7. If 记录和个体相匹配
8. AB = AB + 1
9. End if
10. If 只匹配条件部分
11. A = A + 1
12. End if
13. End for
14. F = W1 * AB / N + W2 * AB / A
15. If F > T
16. 猴群进行攀爬过程
17. End if
18. End for
19. For each 每个攀爬到顶峰的猴子
20. 应用眺望过程算法
21. 应用翻筋斗过程算法
22. End for
23. If 迭代次数足够，跳到第(4)步。

其中，猴群算法的攀爬过程算法描述如下：

(1)随机产生向量 $\Delta \boldsymbol{x}_i=(\Delta x_{i1},\Delta x_{i2},\cdots,\Delta x_{in})$，满足：

$$\Delta x_{ij}=\begin{cases} a & \text{占}1/2\text{概率} \\ -a & \text{占}1/2\text{概率} \end{cases},\quad j=1,2,\cdots,n$$

其中，参数 $a(a>0)$为攀爬过程的步长，通常视具体情况而定，这里令 a = 0.000 01。

(2)计算：

$$f'_{ij}(\boldsymbol{x}_i)=\frac{f(\boldsymbol{x}_i+\Delta \boldsymbol{x}_i)-f(\boldsymbol{x}_i-\Delta \boldsymbol{x}_i)}{2\Delta x_{ij}},\ j=1,2,\cdots,n$$

向量 $\boldsymbol{f}'_i(\boldsymbol{x}_i)=(f'_{i1}(\boldsymbol{x}_i),f'_{i2}(\boldsymbol{x}_i),\cdots,f'_{in}(\boldsymbol{x}_i))$ 被称为目标函数在点 X_i 的伪梯度。

(3)令 $y_j=x_{ij}+a\cdot sign(f'_{ij}(\boldsymbol{x}_i)),\ j=1,2,\cdots,n, y=(y_1,y_2,\cdots,y_n)$。

(4)如果 y 是可用的，则用 y 更新 $\boldsymbol{x}_i$，否则 $\boldsymbol{x}_i$ 不变。

重复步骤(1)~步骤(4)，直到在附近迭代目标函数的值没有改变或最大允许数量的迭代已经达到。

猴群算法的眺望过程描述如下：

(1)随机从 $(x_{ij}-b,x_{ij}+b)$，$j=1,2,\cdots,n$，产生实数 y_j，分别令 $y=(y_1,y_2,\cdots,y_n)$。

(2)如果 $f(y)\geqslant f(\boldsymbol{x}_i)$，并且 y 是可用的，那么用 y 来更新 $\boldsymbol{x}_i$，否则重复步骤(1)直到一个合适的 y 点被找到。本文只替换那些函数值大于等于当前 $f(\boldsymbol{x}_i)$ 的 $\boldsymbol{x}_i$。

(3)重复采用 y 作为初始位置的攀爬的过程。

猴群算法的翻筋斗过程描述如下：

(1)随机从空翻区间$[c, d]$产生实数 α，空翻区间$[c, d]$通常视具体情况而定。

(2)令 $y_j=x_{ij}+\alpha(p_j-x_{ij})$，其中，$p_j=\frac{1}{M}\sum_{i=1}^{M}x_{ij}$，$j=1,2,\cdots,n$，其中，点 $p=(p_1,p_2,\cdots,p_n)$ 被称为空翻支点。

(3)如果 $y=(y_1,y_2,\cdots,y_n)$ 是可用的，则令 $\boldsymbol{x}_i=y$。否则重

复步骤(1)~步骤(2)直到可用的 y 被找到。

3.4　系统实现过程

基于猴群算法的入侵检测系统实现过程如图 2 所示。

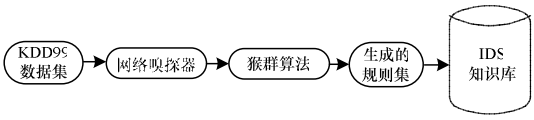

图 2　基于猴群算法的入侵检测系统实现过程

网络审计数据采用的是 KDD99 数据集，数据经过处理分析能够输入到猴群算法应用中，算法执行后规则集产生，这些规则被存储到数据库中，被 IDS 使用在在线实时网络入侵检测中。

4　实验与分析

实验主要在局域网络环境下进行，选用的样本数据是目前入侵检测领域比较权威的测试数据，KDD CUP99 数据来源于 1998 年 DARPA 入侵检测评估程序。该数据一共提供了 500 万条数据，入侵数据有 4 大类，24 小类。本文对测试集作了一些过滤，表 1 说明了训练和测试数据集中记录类型的分配情况。

表 1　记录类型分配情况

记录类型	训练数据集记录数	测试数据集记录数
Normal	49 148	19 787
DoS	280	60
R2L	128	72
U2R	64	35
PROBE	110	46
总数	50 000	20 000

在训练过程中，猴群算法应用到入侵检测中，最大进化代数为 2 000 代，初始种群规模为 1 000，参数设置为 $w1$=0.2，$w2$=0.8，a=0.000 01，[c, d]=[-1, 1]。当训练过程完成时，选取 200 个性质最好的规则作为最后的分类规则，用这些规则分别检测训练数据和测试数据，结果如图 3 所示。

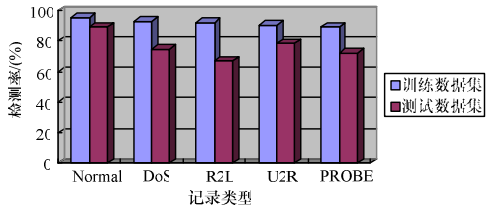

图 3　数据集检测率测试结果

实验结果显示了本文方法的测试结果符合预期效果。如果选择更有效的目标函数和猴群算法参数，测试结果会更好。当用最终生成的规则分类测试数据时，网络攻击的检测率降低约 15%。因此，结果表明产生的规则更适合检测训练数据，因为规则毕竟是来自于测试数据。这个问题可用 2 个方法解决：(1)当训练系统时，使用较少的迭代次数；(2)当分类新的网络数据时，应用更少的最优规则。无论哪种方法，都是通过具体实验找到一个适当的阈值。

5　结束语

本文提出了一种基于猴群算法的入侵检测技术，依据网络审计记录推导出分类规则，用支持度-置信度函数作为目标函数进行规则评估，产生的规则用于实时环境中的检测或分类网络入侵。在面对攻击类型不断增加、攻击技术不断复杂化的现实世界，该技术可以为系统增加和更新规则，获知新的攻击模式。但随着规则的增加，生成规则库的规模越来越大，相应的检测时间变长，而且不同的攻击类型与网络数据中各个特征的相关度也不相同，因此，下一步将对规则集的简约以及规则的最优覆盖方式做更深入的研究。

参考文献

[1] Li Wei. Using Genetic Algorithm for Network Intrusion Detection[C]//Proc. of the United States Department of Energy Cyber Security Group Training Conference. Kansas City, USA: [s. n.], 2004.

[2] Gong Renhui, Zulkernine M, Abolmaesumi P. A Software Implementation of a Genetic Algorithm Based Approach to Network Intrusion Detection[C]//Proc. of SNPD/SAWN'05. [S. l.]: IEEE Press, 2005.

[3] Zhao Ruiqing, Tang Wansheng. Monkey Algorithm for Global Numerical Optimization[J]. Journal of Uncertain Systems, 2008, 2(3): 164-175.

[4] 朱　凯，孟相如，马志强. 一种用于入侵检测的改进人工免疫算法[J]. 计算机工程, 2009, 35(18): 145-147.

[5] Owais S. Using Genetic Algorithm Approach in Intrusion Detection Systems Techniques[C]//Proc. of CISIM'09. [S. l.]: IEEE Press, 2009.

编辑　顾姣健

(上接第 130 页)

采用最优无偏估计方法计算推荐信任度，使计算出的推荐信任度更精确。后验分布能及时更新信誉，信任模型就可及时获得新的评价。在层次过滤算法中引入时间衰减因子和惩罚因子，能更有效地过滤各种恶意 Agent 发布的虚假、不公正和误导的反馈。仿真结果表明，本文模型克服了已有模型的部分局限性，能有效处理简单恶意 Agent、不诚实推荐 Agent 和策略恶意 Agent 的恶意攻击方式，具有广泛的应用前景。

参考文献

[1] Jøsang A, Ismail R. The Beta Reputation System[C]//Proc. of the 15th Bled Conference on Electronic Commerce. Bled, Slovenia: [s. n.], 2002.

[2] Whitby A, Jøsang A, Indulska J. Filtering Out Unfair Ratings in Bayesian Reputation Systems[J]. The ICFAIN Journal of Management Research, 2005, 4(2): 48-64.

[3] Teacy W T L, Patel J, Jennings N R, et al. Travos: Trust and Reputation in the Context of Inaccurate Information Sources[J]. Autonomous Agents and Multi-Agent Systems, 2006, 12(2): 183-198.

[4] Jøsang A, Haller J. Dirichlet Reputation Systems[C]//Proc. of the 2nd International Conference on Availability, Reliability and Security. [S. l.]: IEEE Computer Society, 2007.

[5] 郭俊香，宋俊昌. 信任模型中虚假推荐过滤算法的改进[J]. 计算机工程, 2010, 36(15): 162-163.

编辑　顾姣健