

## 基于模糊聚类的 Linux 网络动态入侵检测

于东敏<sup>1</sup>, 史建政<sup>2</sup>

(1 河北工业大学 廊坊分校, 河北 廊坊 065000;

2 廊坊职业技术学院 计算机系, 河北 廊坊 065000)

**摘 要:** 提出基于模糊聚类的 Linux 系统异常入侵检测方式, 通过对网络动态信息进行分类检测, 能够降低入侵检测的漏检率, 动态检测出网络数据入侵程序, 避免了传统方式的缺陷. 实验证明, 利用基于模糊聚类的入侵检测方式能够快速、准确的检测出入侵程序, 保证 Linux 系统安全.

**关键词:** 模糊聚类; 相关数据集合; 入侵检测

**中图分类号:** TP393

**文献标识码:** A

**文章编号:** 1000-7180(2012)03-0136-04

## Based on the Fuzzy Clustering Linux Network Dynamic Intrusion Detection

YU Dong-min<sup>1</sup>, SHI Jian-zheng<sup>2</sup>

(1 Langfang Department, Hebei University of Technology, Langfang 065000, China;

2 Department of Computer, Langfang Polytechnic Institute, Langfang 065000, China)

**Abstract:** In order to improve the security of the system, make the fuzzy clustering Linux system anomaly intrusion detection mode, through the network information classification the dynamic test, can reduce the intrusion detection miss rate, dynamic to detect the network data flow under the condition of the invasion of the larger program, avoid the traditional way of intrusion detection. It is proved by experiment based on fuzzy clustering of intrusion detection means to be able to quickly and accurately, to detect the invasion and procedures to ensure the Linux system system security.

**Key words:** fuzzy clustering; related data sets; intrusion detection

### 1 引言

现代网络和嵌入式技术发展迅速, 已经深入到人们日常生活的诸多方面, 构成一个相互联系的信息系统. 嵌入式 Linux 系统入侵给网络安全带来了极大的威胁.

传统系统入侵检测方式存在一个弊端, 就是使用者操作信息与系统自身信息分离的难度较大<sup>[1]</sup>. 通常情况下, 入侵程序与系统的网络数据流量有一定关系, 如果将与系统本身没有关系的网络操作信息作为检测对象, 则在检测对象中包括了恶意程序相关信息<sup>[2-3]</sup>. 但是, 假设 Linux 系统入侵程序存在

于动态网络数据中, 则使用传统方式检测时, Linux 系统入侵程序不在检测对象中, 导致漏检 Linux 系统入侵程序, 降低了网络安全性. 所以, 通过对检测对象进行准确的选取, 对系统相关信息严格检测, 能够提高 Linux 系统入侵检测的性能, 这是 Linux 系统入侵检测研究的核心技术<sup>[4-6]</sup>.

为了避免上述问题, 提出一种基于模糊聚类的 Linux 入侵检测方式.

### 2 Linux 系统入侵检测理论

Linux 系统入侵检测是利用与使用者操作行为有关的信息判断是否存在 Linux 系统入侵程序. 检

收稿日期: 2011-07-22; 修回日期: 2011-08-26

基金项目: 教育部高职委资助项目(JJ-200901); 河北省教育厅教学改革立项支持项目(103004)

测对象与系统本身信息有一定关系,能够获取其正常情况下和Linux系统入侵情况下的区别。所以,通过与系统本身有关的特性对比可以检测出Linux系统入侵操作,为系统安全提供保证。下面阐述具体方法:

(1)获取在系统正常情况下,与系统本身特性有关的信息参数集合  $R(a,b)$ 。

(2)将通过上述方式获取的与系统本身特性有关的信息参数集合设置为衡量标准,将其与系统中和操作行为有关的信息进行比较,获取与系统本身特性没有关系的信息,将其作为待测对象,获取方式如下:

$$\begin{cases} R(a,b)+T>0.5, & T \text{ 是待测对象} \\ R(a,b)+T<0.5, & T \text{ 不是待测对象} \end{cases} \quad (1)$$

根据上面方式获取的待测对象是  $T(T_1, T_2, \dots, T_j)$ , 其中  $T_j$  是第  $j$  次操作行为的相关信息,  $s$  是操作次数。通过这个步骤可以获取待测对象的范围,为准确检测恶意入侵程序提供基础。

(3)对比分析检测对象。利用窗口表达式对待测对象相关信息数据进行运算,假设窗口距离是  $x$ , 则通过下述表达式获取测试取值  $E(l)$ :

$$E(l) = \sin r(T_1 + T_2 + \dots + T_j) \quad (2)$$

(4)将测试取值  $E(l)$  与衡量标准进行对比,判断待测对象中是否含有Linux系统入侵程序:

$$\begin{cases} E(l) > \mu, & \text{待测对象中包括Linux入侵程序} \\ E(l) \leq \mu, & \text{待测对象中不包括Linux入侵程序} \end{cases} \quad (3)$$

通过上述判断标准可以得知,假设测试取值大于衡量标准,那么代表待测对象中包括Linux系统入侵程序;假设测试取值小于或者等于衡量标准,那么代表待测对象中不包括Linux系统入侵程序。

通过上面阐述的方法可以得知,在获取待测对象的过程中,待测对象数量与判断结果的准确性成正比例关系,假设网络数据流量成动态变化,则漏检率较高。

入侵操作判断结果。通过下述表达式可以表示入侵的漏检率:

$$M = \frac{(E(l) - \omega_e)/s}{\sum_{j=2}^s T} \quad (4)$$

使用传统方法检测是否存在Linux系统入侵操作,待测对象主要是网络数据动态信息,假设Linux系统入侵程序存在于这些动态信息下,很可能检测不到,导致最终无法检测出这种Linux系统入侵操

作。如果检测对象数量比较少,则获取的检测取值准确性比较低,那么检测结果误差比较大,无法准确检测出Linux系统入侵操作,最终不能保证系统的安全。

与传统Linux系统入侵检测方法相比,基于模糊聚类的Linux系统入侵检测方式待测对象是系统中全部与操作行为相关的信息,可以对所有操作进行模糊聚类分析,然后进行检测,从而获取准确性较高的检测结果,保证了系统安全。这种检测方式主要针对存放于系统中的Linux系统入侵程序,获取结果的漏检率比传统方式低。

### 3 模糊聚类入侵检测

通过模糊聚类方式获取网络动态数据集合,需要依据获得的聚类结果取值变换到抗体中的方法,建立抗体与操作行为本身的相关系数表达式,选取复制函数。复制函数是模糊聚类方式的核心内容,分为复制行为,复制因素行为,复制结果选取,复制终结行为。可以用图1阐述。

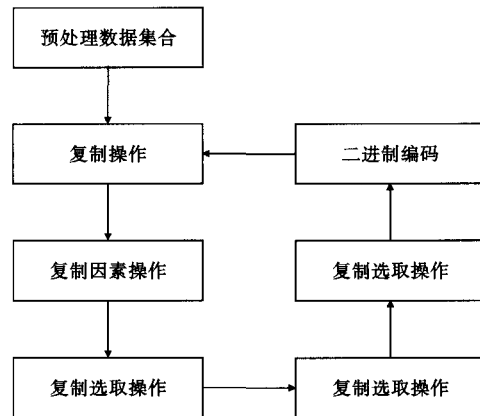


图1 模糊聚类方式结构示意图

基于模糊聚类的Linux系统检测方法分为无监督聚类方法,标类方法。下面阐述一下这两种方法的具体步骤。

#### 3.1 无监督模糊聚类方法

算法的第一步是对操作行为相关信息进行预处理,使测试取值与聚类方法相对应。聚类方法是通过相关信息获取模糊方程和聚类表达式,而模糊方程和聚类表达式密切相关,因此聚类表达式是一个数据集合,把其中的操作行为相关信息参数组成一个数据集合。

(1)获取预处理结果。选取预处理对象  $C(0) = [C_1(0), C_2(0), C_3(0), \dots, C_m(0)]$ , 每个预处理对象代表一种操作行为相关信息结合方式,用二进制编

写其序号。

(2)通过运算获取操作行为相关性系数. 每个数据集代表其相关信息组成方式,从而获取新的数据集,利用表达式  $h(C)=[E(W,R)+1]\times m^2$  获取相关性系数,在这个表达式中,  $m=2$ .

(3)确定是否符合停止迭代处理的衡量标准. 设置衡量标准是进行迭代处理的次数. 假设符合衡量标准,那么停止迭代处理,选取最优解. 假设没有符合衡量标准,则继续进行迭代处理.

(4)实现复制操作  $V_e^E$ ,从而获取  $C^{-1}(m)$ .

(5)实现复制因素操作  $V_i^E$ ,从而获取  $C^{-2}(m)$ .

(6)实现复制因素选取操作  $V_u^E$ ,从而获取相关数据集  $C(m+3)$ .

(7)实现复制终结操作  $V_f^E$ . 假设  $C(m-)$  包括  $C_k(m+2)$  和  $C_l(m+2)$ , 并且符合表达式  $h[C_k(m+2)]=h[C_l(m+2)]=\max\{C(m+2)\}$ , 能够获取一个新的相关数据集,并且获取终结表达式  $C_k(m+2)$  和  $C_l(m+2)$ .

(8)计算相关性系数. 根据其排序能够得到特性数据子集,利用相关性系数表达式获取相关数据集  $C(m+3)$  和相关性系数  $L[C(m+3)]$  集合.

(9) $k++$ ,继续执行步骤 3.

通过上述过程,可以得到测试取值与聚类方法一一对应的关系,即模糊聚类的结果.

### 3.2 分类方法

通过上面方法可以得到聚类结果,然后区分聚类结果下入侵的种类. 通常情况下,系统正常操作相关信息的聚类结果数量比 Linux 系统入侵情况下聚类结果数量多,所以能够对获取的聚类结果进行有序排列,并且设置一个衡量标准,判断聚类结果是否是正常的类别. 假设  $E_k, k=1,2,3,\dots,Q$  是初次进行聚类获取的结果,  $V\in(3,4)$  是衡量标准.

衡量标准是其中的核心数据,根据经验取值,并以此为标准进行判断. 利用这种方式能够简单、迅速地地区分其类别. 但这种方法有一个缺陷,就是每个种类中样本数量与准确率成正比例关系,种类越少,则结果准确率越低. 为了避免这种缺陷,需要增加总的样本数量,从而增加每个种类中的样本数量,最终提高分类结果的准确性,为准确检测提供保证.

### 3.3 Linux 系统检测步骤

如果  $z$  是待测操作网络动态数据集.

(1)通过操作行为相关信息初始化方式可以获得  $z\in E_k, 1\leq k\leq Q$ .

(2)根据分类方式可以判断  $E_k$  是否存在异常.

(3)假设  $E_k$  不存在异常情况,那么判断其不存在 Linux 系统入侵操作,假设  $E_k$  存在异常情况,那么判断其存在 Linux 系统入侵操作.

基于模糊聚类的 Linux 系统入侵检测方式可以快速、准确地检测出系统中的 Linux 系统入侵程序,保证了系统安全.

## 4 实验结果

对传统方式与基于模糊聚类的 Linux 系统入侵检测方式进行对比分析,判断二者获取结果的区别,需要首先构造模糊聚类模型和系统检测模型,从而得到基于聚类方式检测的有序序列和传统方式检测的有序序列,同时使用这两种方式对 Linux 系统入侵操作进行检测,并对获取结果进行对比分析,具体过程如图 2 所示.

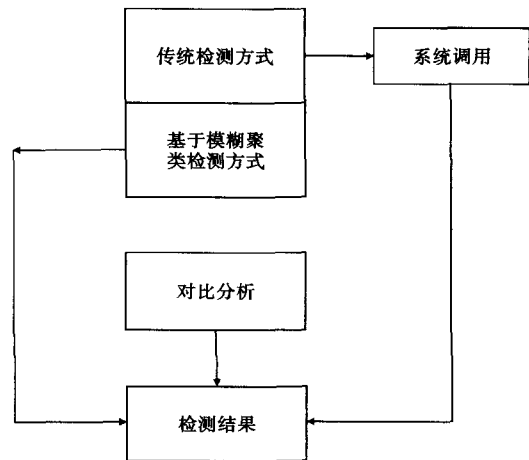


图2 两种检测方式结果示意图

让系统执行相同的操作,分别使用传统检测方式和基于模糊聚类的 Linux 系统入侵检测方式对执行的操作进行检测,将检测相关数据如表 1 描述.

表1 基于模糊聚类检测和传统方式检测数量

| 检测过程 | 传统方式<br>检测对象数目 | 基于模糊聚类的<br>入侵检测对象数目 |
|------|----------------|---------------------|
| 上传文件 | 2656           | 4526                |
| 下载文件 | 4895           | 6235                |
| 获取结果 | 59             | 652                 |

根据表 1 得知,基于模糊聚类的检测方式检测样本数量大于传统方式检测样本数量,而且其内部存储器能够识别操作者的操作权限,其优势在于操作过程简单、便捷.

通过公式(4)可以获取两种检测方式的漏检率,

并得到检测次数.如表2所示.

表2 检测次数对比

| 检测方式        | 检测次数 | 漏检率(%) |
|-------------|------|--------|
| 传统方式        | 87   | 5      |
| 基于模糊聚类的入侵检测 | 195  | 0.5    |

通过实验能够得知,基于模糊聚类的Linux系统入侵检测方式检测次数大于传统方式检测方式的检测次数,能够准确检测到位于系统文件中的Linux系统入侵程序,保证了系统安全.

## 5 结束语

本文提出了基于模糊聚类的Linux系统入侵检测方式,将动态网络信息作为对象,能够检测出传统检测方式无法检测到的动态网络信息中的Linux系统入侵程序,降低了Linux系统入侵检测的漏检率.实验证明,基于模糊聚类的Linux系统入侵检测方式检测内容比较完整,漏检率比较低,能够准确检测出Linux系统入侵操作,保证了系统安全.

## 参考文献:

- [1] 王玉峰,范明钰,王绪本,等.入侵检测系统的研究[J].计算机仿真,2004(5):66-68.
- [2] Ee C T, Ratnasamy S, Shenker S. Practical data-cen-

tric storage[C]//Proceedings of the 3rd Conference on Networked Systems Design & Implementation. San Jose, CA: IEEE, 2006:137-142.

- [3] 李权,周兴社.基于KPCA的多变量时间序列数据异常检测方法研究[J].计算机测量与控制,2011(4):77-78.
- [4] Kim S, Fonseca R, Dutta P, et al. Flush: a reliable bulk transport protocol for multihop wireless networks [C]//Proceedings of the 5th international Conference on Embedded networked Sensor Systems. New York, NY, USA: ACM, 2007.
- [5] Mao Y, Wang F, Qiu L, et al. S4: small state and small stretch routing protocol for large wireless sensor networks[C]//Proc of the USENIX NSDI Conf, Cambridge, MA: IEEE, 2007.
- [6] Musaloiu-E R, Liang C-J, Terzis A. Koala: Ultra-low power data retrieval in wireless sensor networks[C]//Proc of the International Conference on Information Processing in Sensor Networks (IPSN 2008). St Louis, Mo: IEEE, 2008:421-432.

## 作者简介:

于东敏 女,(1972—),硕士,讲师.研究方向为计算机控制技术、网络安全.

史建政 男,(1971—),硕士,讲师.研究方向为物联网、网络安全.

(上接第135页)

## 参考文献:

- [1] Chandra L, Roy S. Estimation of energy consumed by software in processor caches [C]//Proceedings of the IEEE International Symposium on VLSI Design, Automation and Test (VLSI-DAT'08). Hsinchu, Taiwan: IEEE, 2008: 21-24.
- [2] Aaraj Najwa, Raghunathan time analysis of a software-based trusted platform module [C]//Proceedings of the Design, Automation and Test in Europe Conference and Exhibition (DATE'07). Nice, France: IEEE, 2007: 1-6.
- [3] 苟小龙,贺熙颖,肖育劲. PLC在培训仿真机中的应用[J].计算机仿真,2003,11(60):222-225.

- [4] Konstantakos V, Chatzigeorgiou A, Nikolaidis S, et al. Energy consumption estimation in embedded systems [J]. IEEE Transactions on Instrumentation and Measurement, 2008, 57(4): 797-804.
- [5] Mukherjee T, Tang Qinghui, Ziesman C, et al. Software architecture for dynamic thermal management in datacenters [C]//Proceedings of the 2nd International Conference on Communication Systems Software and Middleware. Bangalore, India: IEEE, 2007: 1-11.

## 作者简介:

李吉彪 男,(1963—),副教授.研究方向为计算机网络.

王果 男,(1977—),硕士研究生,讲师.研究方向为计算机应用.