维普资讯 http://www.cqvip.com

基于Minkowskey距离的模糊入侵检测方法研究

余建桥 张 林 彭喜化
(西南农业大学信息学院 重庆400716)

Research on Fuzzy Intrusion Detection Method Based on Minkowskey Distance

YU Jian-Qiao ZHANG Lin PENG Xi-Hua
(Informaion College, Southwest Agircultural University, Chongqing 400716)

Abstract After analyzing the deficiency of traditional intrusion detection system, a new kind of intrusion detection method based on Minkowskey is put forward for reducing false positive rate. The essence of it is to set up a normal behavior fuzzy sub collection A on the basis of watching the normal system transfer of the privilege process, and set up a fuzzy sub collection B with real time transfer array, then detect with the principle of minimum distance in fuzzy discern method.

Keywords Intrusion detection, Minkowskey distance, Fuzzy discern, System transfer

1 引言

由于互联网能够带来快速、便捷的服务和丰富、全面的信息,越来越多的企业、政府及个人加入到互联网。同时,互联网的安全成为个人、企业甚至国家都关注及迫切要解决的问题。统计数据表明,超过80%的入侵和攻击是从企业内部发起的,由于防火墙自身所固有的防外不防内,使得防火墙对这些入侵和攻击无能为力。因此,入侵检测技术应运而生。入侵检测系统(Intrusion Detection System)可以弥补防火墙的不足,检测来自网络内部的攻击,为网络提供实时的监控并且在发现入侵的初期采取相应的防护措施。

目前入侵检测方法分为两类:误用入侵检测(misuse intrusion detection)和异常入侵检测(anomaly intrusion detection)。误用入侵检测是先对已知的入侵方式进行入侵特征的抽取定义入侵模式,然后将检测到的数据特征和入侵模式相匹配来检测入侵。异常入侵检测假设入侵活动是异常活动的子集,利用系统或用户的正常行为模式检测入侵,它首先建立正常行为模式,当发现系统或用户行为与正常行为模式发生显著偏离即认为是入侵。

由于误用检测不能检测出未知的攻击,因此现在多数入侵检测系统都采用异常检测方法或两者结合使用,虽然能检测出未知的攻击,但误报率较高。为了降低误报率,本文将民可夫斯基距离(Minkowskey Distance)引入了入侵检测系统,提出了基于Minkowskey距离的入侵检测方法MDID(Intrusion Detection Based on Minkowskey Distance)。

2 MDID的原理与异常行为的检测

2.1 MDID的原理

MDID的实质是在监控特权进程的正常系统调用基础上建立正常行为模糊子集$\underset{\sim}{A}=[a_1,a_2,\cdots,a_n]$,用检测到的实时调用序列建立模糊子集$\underset{\sim}{B}=[b_1,b_2,\cdots,b_m]$,然后用模糊识别方法中的最小距离原则进行检测,下面给出最小距离原则的原理。

最小的距离原则是:设在论域U上有N个模糊子集$\underset{\sim}{P_1}$,$\underset{\sim}{P_2},\cdots,\underset{\sim}{P_n}$。若i={1,2,…,n}使$d(\underset{\sim}{A},\underset{\sim}{B})=\min\{d(\underset{\sim}{A},\underset{\sim}{B})\mid 1\leqslant j\leqslant n\}$,则$\underset{\sim}{B}$与$\underset{\sim}{P_j}$的距离最小,$\underset{\sim}{B}$应归入模式$\underset{\sim}{P_i}$,其中$d(\underset{\sim}{A},\underset{\sim}{B})$代表民可夫斯基(Minkowskey)距离。

我们设N个模糊子集$\underset{\sim}{P_1},\underset{\sim}{P_2},\cdots,\underset{\sim}{P_n}$包含了1个正常行为模式和(n-1)个已知的入侵模式。通过异常归类,从而很好地将异常检测和误用检测结合起来。

2.2 特征参数的提取

首先,我们跟踪由特权进程运行时所产生的所有系统调用,进行预处理并构建包含长度指定为k的系统调用序列的数据集。

特征参数的提取分3步进行:

(1)在特定的时间间隔内记录特权进程的系统调用和参数序列;

(2)对参数进行预处理;

(3)用长度为k的滑动窗口跟踪系统调用。

下面用一个例子来说明特征参数提取过程。假设我们所跟踪的系统调用和参数序列如下:

open, p_1, read, p_2, write, p_3, close, p_4, open, p_5, read, p_6, mmap, p_7

其中p_i为系统调用参数。接下来设定一个阈值T,建立比较函数对参数进行预处理:

$$f(p_i)=\begin{cases}0, p_i\leqslant T\\ 1, p_i\geqslant T\end{cases}, i=1,2,\cdots,n \quad (1)$$

用$f(p_i)$代替p_i得到预处理后系统调用和参数序列为:

open,0,read,0,write,1,close,1,open,1,read,0,mmap,0

我们采用大小为k的滑动窗口跟踪每个系统调用,并把它们记到每个窗口内。设k=4,则我们跟踪调用后得到以下长度为4的短序列:

open,0,read,0
read,0,write,1
write,1,close,1
close,1,open,1

余建桥 教授,主要研究方向:人工智能、数据库。张 林 硕士研究生,主要研究方向:网络安全,数据库。彭喜化 硕士研究生,主要研究方向:人工智能,Agent技术。

维普资讯 http://www.cqvip.com

open,1,read,0

read,0,mmap,0

我们得到了 n 个不同的序列(本例中 n 为6),如果要用数组去存储,则需要将数组的大小规定为 O(nk),所以用树形结构来存储,这样既节省存储空间又会提高整个系统的使用效率。以上的序列用树形结构表示得到如图1的森林。

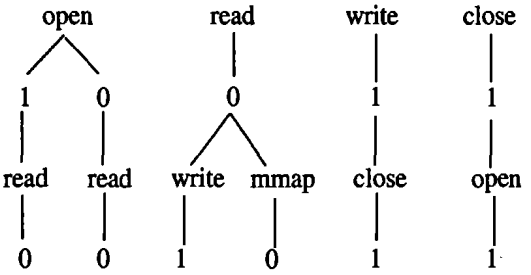

图1 系统调用和参数序列的森林

以上序列树的森林包含了20个节点,而采用数组存储则需要(6×4)个节点。

2.3 异常行为的检测

建立好正常行为的数据集后,可以依据该数据集对所跟踪的系统调用进行判断。将新跟踪的系统调用划分成以 k(本文取 k=4)为长度的调用序列,查看其是否在数据集中出现。若序列未在该数据集中,则称该序列不匹配。理想情况下,一个正常行为发生时,不匹配序列数量为零,出现异常行为时,不匹配序列数量将发生很大变化。序列不匹配的多少直接影响着行为的异常程度,因此,采用对不匹配序列计数的方法或对不匹配序列占总序列的百分比的方法来判断异常。

设$\underset{\sim}{A}=[a_1,a_2,\cdots,a_n]$和$\underset{\sim}{B}=[b_1,b_2,\cdots,b_m]$是论域 U 的两个模糊子集,其中 a_i 为正常行为数据库中的短序列,b_i 为待检测的短序列,则$\underset{\sim}{A}$和$\underset{\sim}{B}$的民可夫斯基(Minkowskey)距离为:

$$d(\underset{\sim}{A},\underset{\sim}{B})=[\sum_{i=1}^{m}(b_i-A)^q]^{1/q} \tag{2}$$

其中,如果 b_i 与 a_j 匹配则(b_i-A)的取值为"0"否则为"1",即$\underset{\sim}{B}$与$\underset{\sim}{A}$的不匹配度增加1。为了更好地表示匹配度,本文对民可夫斯基 (Minkowskey)距离作了如下处理:

$$d(\underset{\sim}{A},\underset{\sim}{B})=\frac{1}{m}[\sum_{i=1}^{m}(b_i-A)^q]^{1/q}\times 100\% \tag{3}$$

其中 m 为待检测的短序列个数,根据检测的要求不同 q 可取不同的值,本文为了在误报率和漏报率之间达到平衡取 q=2,如果要求漏报率低,则 q 取1,如果要求误报率低,则 q 去3,本文对 $d(\underset{\sim}{A},\underset{\sim}{B})$设定检测阈值 T_1,则待检测序列异常的判断条件是:

$$d(\underset{\sim}{A},\underset{\sim}{B})\geqslant T_1,\ 0\leqslant T_1\leqslant 1 \tag{4}$$

计算匹配度的算法如下:

算法1

```
Main()
{
   count=0;
   i=1;
   j=1;
   for(i=1;i<=m;i++){
    for(j=1;j<=n;j++){
     if(b[i]==a[j])
       continue;//如果 b[i]为正常调用序列,则不匹配度 count 值不变
     else
       count+=1;//如果 b[i]为异常调用序列,则不匹配度 count 值加1
     }
   }
}
```

现在,我们在跟踪系统调用时得到如下序列:

open,0,read,1,mmap,1,close,1,open,1,mmap,0,mmap,0

用 k=4的滑动窗口对序列进行处理得到如下短序列:

open,0,read,1

read,1,mmap,1

mmap,1,close,1

close,1,open,1

open,1,mmap,0

mmap,0,mmap,0

通过算法比较发现五个不匹配序列:

open,0,read,1

read,1,mmap,1

mmap,1,close,1

open,1,mmap,0

mmap,0,mmap,0

如果我们设异常调用占全部系统调用的最大忍受度即 T_1为20%,那么因为

$$d(\underset{\sim}{A},\underset{\sim}{B})=\sqrt{5}/6\times 100\%=37.3\%>20\%$$

从而得出该行为异常的结论。同时如果检测到它与异常模式 $\underset{\sim}{P_i}$ 的匹配度为"0",则可以将检测到的异常归类到$\underset{\sim}{P_i}$,从而检测出入侵的种类,使防火墙系统或管理员能够更好地处理入侵事务。

结束语 本文提出了一种基于系统调用的入侵检测方法——MDID。该方法由于采用了 Minkowskey 距离降低了系统的误报率,能很好地将误用检测和异常检测结合起来,由于采用了树形存储结构,从而降低了存储空间,提高了检索速度。由于网络的流量、时间限制及系统调用间的相互关系等因素,这些因素会直接影响到行为是否异常,因此在今后的研究中将引入行为间的相关系数和基于实例的学习方法,提高系统的学习能力,降低误报率和漏报率。

参 考 文 献

1 Dickerson J E. Fuzzy network profiling for intrusion detection; IEEE 2000 [S]

2 Bace R G. Intrusion Detection [M]. Macmillan Technical Publishing. Indianapolis. in 462900 USA, 2000

3 Anderson D, Frivold T, Valdes A. Next-generation intrusion detection expert system (NIDES) [R]. Computer Science Laboratory (SRI International, Mello Park. CA); [technical Report SRI-CSL-95-07]

4 Lunt T F, Tamaru A, Giham F. A real-time intrusion detection expert system (IDES)-Final technical report [R]; Computer Science Laboratory

5 Vigna G, Kemmerer R A. NeSTAT: A Network Based Intrusion Detection System [J]. Journal of Computer Security, 1999(7)

6 Bass T. Intrusion detection systems and multi-sensor data fusion [J]. Communications of ACM, 2000, 43(4): 99～105

7 Northcutt S. Network Intrusion Detection: Analyst's Handbook. Authorized translation from the English language edition [M] published by New Riders an imprint of Macmillan Computer Publishing U. S. A

8 陈鸣钊,张志烈,樊宝康. 模糊数学及其应用[M]. 河海大学出版社

9 李之棠,杨红云. 模糊入侵检测模型[J]. 计算机工程与科学

10 曾志峰. 网络安全测防体系的研究与实现[D]:[北京邮电大学博十学位论文]. 2001

11 李鸿培,王新梅. 入侵检测技术的研究[J]. 通信保密,1999(3)