一种基于贝叶斯判决的先进入侵检测模型*)

罗光春 卢显良 李 炯 张 骏
(电子科技大学信息中心 成都610054)

A Novel IDS Evaluation Model Based by Bayes Estimation

LUO Guang-Chun LU Xian-Liang LI Jiong ZHANG Jun
(Information Centre of UEST of China, Chengdu 610054)

Abstract One key problem for intrusion detection system is the correctness and efficiency of detection algorithm. This paper presents a revised detection algorithm through the use of Bayes decision. Bayes decision is a random pattern classified recognition method of the pattern recognition theory. The algorithm in this paper is designed refer to the lest-risk Bayes decision. Experiments show that this algorithm has better performance. In the paper, we firstly introduce the Bayes algorithm and threshold selection algorithm. Then depending on the decision, the detection algorithm of intrusion detection system is designed. In the end, the experiment results are provided.

Keywords IDS, Bayes estimation, Evaluation

0 引言

入侵检测系统(IDS)是实时地根据网络数据,检测是否存在网络攻击的一种软件系统[1]。入侵检测技术经过多年的研究,已经取得了长足的发展。但是,仍然存在一些桎梏,检测算法的误判率和漏判率较高,使得入侵检测系统在实际运用中总是不能很好地满足对网络安全有较高要求的用户的需要[2]。本文提出一个采用贝叶斯算法降低入侵检测的误判率和漏判率的思路,并据此进行讨论和实验。

1 入侵的评估

从本质上看,入侵检测算法的任务是:根据网络数据包的内容自动地确定数据包的类别。从数学角度来看,入侵检测是一个映射的过程,它将待处理的数据包映射到确定的类别中,该映射可以是确定的一一映射,分为入侵和非入侵两种情况;也可以是不确定的一对多映射,模糊地判断出存在入侵行为的可能性,其结果表现为概率。用数学公式表示如下:

$f: A \to B$

其中,A 为待处理的网络数据包,B 为检测结果的集合[3]。

因为入侵检测从根本上说是一个映射过程,所以评估入侵检测系统的标志就是映射的准确程度和映射的速度。映射的速度取决于映射规则的复杂程度,而评估映射准确程度的参照物则是通过专家根据经验对入侵行为进行检测的结果,这里假设人工检测完全正确并且排除个人思维差异的因素。那么,与人工检测结果越相近,检测的准确程度就越高,这里包含了评估入侵检测系统的两个指标:误判率和漏判率。

误判率是所有判断的入侵中与实际分类结果不吻合的文本所占的比率。其数学公式表示如下:

$$\text{误判率}=\frac{\text{检测的错误入侵数}}{\text{实际检测的入侵数}}$$

类似的,漏判率的数学公式可表示如下:

$$\text{漏判率}=\frac{\text{实际检测的入侵数}-\text{检测出的入侵数}}{\text{实际检测的入侵数}}$$

误判率和漏判率也可以等价地使用准确率和测全率来描述。误判率和漏判率反映了检测质量的两个不同方面,两者必须综合考虑,不可偏废,因此,存在一种新的评估指标,F1测试值,其数学公式如下:

$$F1\text{测试值}=\frac{\text{准确率}\times\text{测全率}\times 2}{\text{准确率}+\text{测全率}}=\frac{(1-\text{误判率})\times(1-\text{漏判率})\times 2}{(1-\text{误判率})+(1-\text{漏判率})}$$

2 贝叶斯算法

基于贝叶斯算法的训练方法和分类算法是分类系统的核心部分。首先计算特征词属于每个类别的几率向量,入侵检测系统应该首先取得$(w_1, w_2, w_3, \cdots, w_n)$,

$$w_k=P(W_k|C_j)=\frac{1+\sum_{i=1}^{|D|}N(W_k,d_i)}{|V|+\sum_{r=1}^{|V|}\sum_{i=1}^{|D|}N(W_r,d_i)}$$

在新的网络访问信息到达时,根据入侵特征进行分类,然后按下面的公式计算该访问 d_i 属于类 C_j 的几率:

$$P(C_j|d_i;\hat{\theta})=\frac{P(C_j|\hat{\theta})\prod_{k=1}^{n}P(W_k|C_j;\hat{\theta})^{N(W_k,d_i)}}{\sum_{r=1}^{|C|}P(C_r|\hat{\theta})\prod_{k=1}^{n}P(W_k|C_r;\hat{\theta})^{N(W_k,d_i)}}$$

其中,$P(C_j|\hat{\theta})=\frac{C_j\text{训练网络访问数}}{\text{总训练网络访问数}}$,$P(C_r|\hat{\theta})$ 为相似含义,$|C|$ 为类的总数,$N(W_k,d_i)$ 为 W_k 在 d_i 中的词频,n 为入侵特征总数。

最后,比较新的网络访问属于所有类的几率,将网络访问分到几率最大的那个类别中。经过处理之后,检测出的概念P将大于原有的概率[4]。

3 阈值的确定

上面的算法都是将网络访问归于入侵还是正常访问,而事实上,入侵检测系统中的分类不是完全互斥的,存在这样一些既属于入侵,又同时属于正常访问的网络访问,对于这种情

*)本文由国家九七三(项目号973-1-4-2)和电子科技大学青年基金支持。罗光春 博士研究生。卢显良 教授,博士生导师。李 炯 讲师。张 骏 讲师。

况需要对每种访问确定阈值，当访问在该类的阈值之上时，就将文本归于入侵或者正常访问。但是，阈值的确定是十分困难的，理论上，没有很好的解决方法，一般采用预定初始值，然后给出测试访问，比较人工测试结果后再根据分类的准确程度调整初始值，这样的方法有两个缺点，首先，初始值的确定不容易，完全是根据经验或简单的测试而定，其次，调整的幅度无法确定，当初始值过高或过低需要增减时，增减的幅度无法很好地确定，只能反复测试，反复调整，这样就大大地增加了工作量。而且，一个入侵系统的阈值由于入侵的不同也无法完全应用于另一个分类系统中[5]。

我们在系统中考虑了一种确定阈值的方法，称为百分比阈值确定法，它的基本思想是：首先依据上述训练算法和分类算法构造分类器，然后用分类器分类所有的网络访问，从而每个网络访问都得到一个相关的值。然后按递减顺序排列所有训练数据得到的值，假定入侵有 n 种网络访问文本，那么这些访问的值为 $d_1, d_2, \cdots, d_n$，那么入侵阈值 y 确定如下：

$y=d_{m\%}$

其中，s 为初始值，根据训练的访问的质量程度，可以确定为80或更高，这样就确定了入侵的初始阈值，可以想象，s 越大，该分类器的漏判率就越低，误判率就越高，相反地，s 越小，漏判率就越高，误判率就越低，然后根据测试进行调整。相应地，调整阈值可以转化为调整 s 值，如果对漏判率满意而对误判率不满意，那么可以减少 s 值，否则就增加 s 值。

4 实验结果

为了简便起见，我们用 snort 为基础，使用贝叶斯算法改写了数据匹配模块。在网络上放置了被攻击服务器、IDS 服务器、攻击模拟机。我们搜集了120个网络数据包，其中80个为入侵攻击，40个为正常访问。将这些数据分为4组，随机的选取其中的3组作为训练样本，将这些数据由攻击模拟机发送到被攻击服务器，对 IDS 服务器上的入侵检测系统的入侵进行测试分类。为了避免实验的偶然性，重组训练样本集合，对测试样本进行反复测试，取平均值作测试结果。为了比较实验结果，我们同时测试了不采用贝叶斯算法的检测结果。

表1 简单匹配的检测结果

	人工判为攻击	人工判为正常访问	总计
系统判为攻击	30	11	41
系统判为正常访问	30	19	49
总计	60	30	90

然后，我们采用贝叶斯算法进行入侵检测实验。表2是每次测试的阈值 λ，表3是对于不同阈值 λ 进行实验的结果。通过这样的实验，可以选取合适的 λ 值，达到误判率和漏判率的优选。

表2 决策表

次数	λ_{11}	λ_{12}	λ_{21}	λ_{22}
1	0	0.4	0.6	0
2	0	0.3	0.7	0
3	0	0.2	0.8	0
4	0	0.1	0.9	0
5	0	0	1.0	0

表3 最小风险贝叶斯入侵检测系统的实验结果

λ_{12}		系统判为攻击	人工判为正常访问	总计
0.4	系统判为攻击	50	12	62
	系统判为正常访问	10	18	28
	总计	60	30	90
0.3	系统判为攻击	50	10	60
	系统判为正常访问	10	20	30
	总计	60	30	90
0.2	系统判为攻击	49	8	57
	系统判为正常访问	11	22	33
	总计	60	30	90
0.1	系统判为攻击	50	11	61
	系统判为正常访问	10	19	29
	总计	60	30	90
0	系统判为攻击	51	17	68
	系统判为正常访问	9	13	22
	总计	60	30	90

在 $\lambda_{12}=0$ 时，贝叶斯算法退化成“朴素”贝叶斯算法，对这80个网络包的检测，朴素贝叶斯入侵检测算法检测出了68个攻击，其中17个是误判，而漏掉了9个入侵。朴素贝叶斯判漏判率较低，但是误判率较高。而采用适当的选择阈值，在略增大误判率的同时，漏判率有不同程度的改善。详见表4的有关统计数据。

表4 算法的误判率和漏判率比较

贝叶斯算法	误判率	漏判率
$\lambda_{12}=0.4$	16%	20%
$\lambda_{12}=0.3$	16%	16%
$\lambda_{12}=0.2$	18%	14%
$\lambda_{12}=0.1$	16%	18%
$\lambda_{12}=0$	15%	25%

应该指出的是，对于不同类型的网络入侵，存在不同的阈值的优选值，这是由不同类型的网络入侵的特征和这种类型的检测器的准确率所决定的。

结束语 在网络入侵检测系统中运用贝叶斯算法，是在已有的网络入侵检测算法的现有误判率和漏判率的基础上，通过学习，减小误判和漏判的一种有效途径。贝叶斯算法的学习需要搜集大量的网络访问的数据，并且人工确定是否为入侵，借此来训练入侵检测系统的学习模型。这是一项非常复杂的工作，但是从实验结果看来，也是一项很有价值的重要工作。如果通过某种合作的途径来共享这些学习的数据，并随时根据新的网络入侵特征更新这些数据，必将对入侵检测技术的发展起到极大的推动作用。

参考文献

1 孙即祥. 现代模式识别. 国防科技大学出版社，2002
2 杨正光，吴岷，张晓莉. 模式识别. 中国科学技术大学出版社，2002
3 唐正军. 网络入侵检测系统的设计与实现. 电子工业出版社，2002
4 Giarratano J, Riley G. Expert Systems Principles and Programming. PWS Publishing Company, 1998
5 Stevens W R. Unix Network Programming Networking APIs: Sockets and XTI. Prentice Hall PTR
6 Satyanarayananan M, Kistler J J, Kumar P, et al. Coda: A Highly Available File System for a Distributed Workstation. IEEE Transactions on Computers, 1990, 39(4)
7 Papadopoulos G A, Arbab F. Coordination Models and Languages. Advances in Computers. Academic Press, 1998
8 Russell S J, Norvig P. Artificial Intelligence. Prentice Hall, New Jersey, 1995
9 Ford K M, Coffey J W, Andrews E J. Diagnosis and explanation by a nuclear cardiologu expert system. Internatioal Journal of Expert System, 1996, 8(4)