一种MANET入侵检测系统模型研究

何 利[1],谢 中[2]
(1.重庆邮电大学 计算机学院,重庆 400065;2.西南大学 计算机学院,重庆 400078)

摘 要:用模糊集概念统计min-sup和minconf,并加入了第三约束要素:兴趣度,使min-sup和minconf通过数据信息本身的特性计算得到,规则可信度更高,避免了这两个值设置过高会异常漏检,设置过低无法检测异常的问题。根据这种思想设计了一种新的移动自组织网络入侵检测模型,把这个模型在网络仿真软件中对基于主机的数据进行了挖掘分析,用AODV协议实现了对模型的3种典型攻击。实验结果表明该模型对这些攻击的检测率平均达到90%以上。

关键词:数据挖掘;MANET;入侵检测
中图分类号:TP393.08 **文献标识码**:A **文章编号**:1673-629X(2008)07-0135-04

Research of One Intrusion Detection Model for Mobile Ad-hoc Networks

HE Li[1], XIE Zhong[2]
(1.College of Computer, Chongqing University of Posts and Telecommunications, Chongqing 400065, China;
2.College of Computer, Southwest University, Chongqing 400078, China)

Abstract: The statistic of min-sup and minconf can be done by fuzzy set concept, at the same time the third key factor of confidence can be considered into it and min-sup and minconf can be gained by information characteristic itself. By this way, it avoids either the possibility of missing detection caused by too high value of min-sup and minconf or the possible disability to detect abnormality caused by too low value of the two variables. A new model of mobile self-constructed network intrusion checking was designed according to the thinking pointed above. The model was used in network-imitate software to do data mining based on the data of host computer. Furthermore, finished the testing of 3 representative attacks aimed to the model. The experiment results show that the average detection ratio of the model to these intrusion reached above 90%.

Key words: data mining; mobile ad-hoc networks; intrusion detection

0 引 言

移动自组织网络是由可能的多跳无线信道为移动节点提供了网络连通性,链路的开放性导致了信息容易受到侦听和泄漏,因为其不依赖于任何固定的网络设施,而是通过节点间的相互协作来进行网络互联,无法构建明确的安全防线。当前移动自组织网络受到的安全挑战[1]主要有:

1) 无线链路受通信频带的限制,无线节点则受到包括存储空间、CPU以及电池使用寿命的限制,这就决定了那些依赖于节点位置和依赖于大量I/O计算的复杂安全措施的使用受到限制。

2) MANET无中心特性以及拓扑结构的不稳定性,注定了MANET安全的安全策略必须是分布的和以个体为主的。

入侵检测技术[2]能够有效地检测到网络攻击并做出响应,它的实质是对安全审计数据的处理,安全事件审计系统作为保障信息系统安全的基础部件,已经越来越多地被应用到操作系统和网络安全管理工具中,但是在海量的数据信息中提取出具有代表性的系统特征模式,现在最有效的手段就是数据挖掘。数据挖掘本身是一项通用的知识发现技术,其目的就是要从海量数据中提取出人们所感兴趣的数据信息。目前对挖掘算法的研究已经比较成熟,但是要真正从海量数据中提取出外面有用的"特定应用数据",这个算法就必须建立在特定应用的基础之上,需要有足够的先验知识。实验表明:系统安全信息的获取需要建立在对原始数据有价值的信息的选择上,尤其是MANET这种特殊网络中,采用一种特殊的分析过程,把入侵检测本身作为一种数据分析过程,着眼于对海量的安全审计

收稿日期:2007-10-14
基金项目:重庆市教育科研资助项目(040503)
作者简介:何 利(1977-),女,硕士,讲师,主要研究方向为网络安全;谢 中,硕士,讲师,主要研究方向为数据挖掘和电子商务。

应用数据的数据挖掘算法,以一种自动和系统的手段建立一套自适应的、具备良好扩展性的入侵检测系统。

1 数据挖掘算法描述

数据挖掘是针对特定应用的数据分析处理过程,如何选择输入数据、变换数据和相应的挖掘算法,取决于具体的数据挖掘目标,按照挖掘目标的不同,数据挖掘算法分为以下几种。

1.1 关联分析算法

关联规则[3]的任务是:给定一个事务数据库 D,求出所有满足最小支持度 minsup 和最小可信度 minconf 的关联规则。该问题可以分解为两个子问题:① 求出 D 中满足最小支持度 minsup 的所有频繁项目集;② 利用频繁项目集生成满足最小可信度 minconf 的所有关联规则。问题 ① 的求解是关联规则挖掘的关键部分,问题 ② 的解决方法较为简单,对每个频繁项目集 L,计算其所有的非空子集,对每个子集 A,考察规则 $A-(I-A)$,如果该规则的可信度大于最小可信度 minconf,则输出此规则。

这种规则可以描述为:$X \Rightarrow Y$,其中 X,Y 是两组数据项,$X \subset T$,$Y \subset T$,$X \cap Y = \varnothing$

1.2 数据分类算法

数据分类[4]的目的是提取数据库中数据项的特征属性,生成分类模型,该模型可把数据库中的数据项映像到给定类别中的一个。数据分类处理步骤如下:

1) 获取训练数据集,该数据集中的数据记录具有和目标数据库中数据记录相同的数据项。

2) 训练数据集中每一条数据记录都有已知的类型标识与之相关联。

3) 分析训练数据集,提取数据记录的特征属性,为每一种类型生成精确的描述模型。

4) 使用得到的类型描述模型对目标数据库中的数据记录进行分类或生成优化的分类模型(分类规则)。

1.3 序列分析算法

序列分析[5]和关联分析具有明确的界限,关联分析是挖掘数据记录中不同数据项之间的关联性,序列分析则是发现不同数据记录之间的相关性。序列分析算法的任务是:在事务数据库中发现满足用户最小支持度和置信度的最大序列模式。序列模式算法通常有 5 个步骤:

1) 排序:以数据项发生为主体,数据项发生的时间为次体,对原始数据库进行排序,使之形成主体序列。

2) 最大序列生成:找出数据库中的所有最大数据序列。

3) 替换:用最大序列替代原始数据库。

4) 规则形成:利用最大序列模式形成规则库。

以上算法都是通用的,并不针对某一特殊的应用环境。综合分析以上各种算法的特点,对于 MANET 网络来讲,部署基于复杂数据挖掘算法的入侵检测系统一般不可能,因此要求的数据挖掘算法必须是快速、简洁的,占用的资源要求尽可能的小,能够最大限度地形成有效规则库。在文中,笔者选取的基础算法是关联分析算法,在这个算法基础上,做出了有利于 MANET 网络特性的改进。

2 MANET 中的扩展挖掘算法

Agrawal 等人[6]设计了一种高效、快速的关联规则挖掘算法,但是这个算法选取最小支持度 min-sup 和最小置信度 minconf 的方法还是人为确定,不能通过数据信息本身的特性计算得到。导致这两个值设置过高会异常漏检,设置过低无法检测异常的问题,而且在很多数据环境中,发现仅仅靠这两个值确定的规则有很多无用的甚至是错误的规则[7]。对于 MANET 网络来讲,尽量小的内存消耗和尽量快速准确的算法是必要的,因此文中采用了模糊集[8]的概念和增加兴趣度来提高挖掘算法性能。

设有集合 $C=\{c_1,c_2,\cdots,c_n\}$,其中 c_i 为 C 中的第 i 例,表示信息集合 C 中的一个事务序列。令 $P=\{p_1,p_2,\cdots,p_m\}$,P 表示信息集合 C 的属性集,其中每个属性 $p_k(1\leqslant k\leqslant m)$ 再按模糊规则划分为若干个模糊集 $\{p_{k1},p_{k2},\cdots,p_{kw}\}(w>0)$,相应的关系函数为 $F_m p_k=\{fp_{k1},fp_{k2},\cdots,fp_{kw}\}$。对于属性集合中的任意值 p_{jk},将全部事务对该属性的支持计数相加后除以总的事务数 n,就可以得到全部事务对该事务属性的支持度:

$$V=\frac{\sum_{i=1}^{n} d_{ijk}}{n}。$$

用项集的支持数 N 代替支持度 V,将所有事务项集的支持数相加,除以总事务数 total(C) 就得到了项集的 min-sup $=\dfrac{\sum_{c_i\in c}(\prod_{P_K\in P})}{\text{total}(C)}$,相应的其最小置信度为 $B=\dfrac{\text{min-sup}}{V_p}$。

为了进一步约束规则项集的有效生成,文献[9,10]等提出了在 min-sup 和置信度的基础上再增加一个兴趣度值。一般意义上说,兴趣度应该是在基于统计独立性假设下真正的事务发生的强度与期望的强度之

比,然而如果只是把人为设置的支持度和可信度采用统计加权,得出的兴趣度值是不可信的[11],但是采用模糊统计的方式就可以更为精确地计算兴趣度值。具体的定义如下[9]:

规则 R 的兴趣度为:

$$IN_i = \frac{B_i - V_i}{4} * (B_i + V_i)$$

$B_i - V_i$ 的计算结果可正可负。计算出来的兴趣度 Interest 可能大于 0,也可能小于 0。

对数据信息库中的每个记录的扫描可以看作是新的项目加入候选项的过程,这个过程可以通过合并项集[12] 的操作完成,因为关联规则算法表明如果一个集合通过了测试,则它的所有子集也都能通过相同的测试。算法逐个读取信息库中的记录,并考察当前是否有符合 min - sup 和 IN 的要求的频繁项,如果有,则删除该项的真子集。重复该操作直到信息库中的记录全部读完为止。项目 I 加入项集的规则为:$S_{new} = S_{old} \cup \{I\}$,同时及时地清除将要加入项集的真子集,可以提高节点的空间利用率,同时减少下一次加入操作的比较次数,加快挖掘速率。在得到所有的频繁项集后,为了得到最大频繁项集,采用的剪枝策略[13] 是:

第一步是删除 C_{k+1} 中支持度小于 min - sup 的项集;

第二步是删除 C_{k+1} 中的含有非频繁子集的项集;

第三步是删除 C_{k+1} 含有属于同一模糊属性的项集的真子集。

重复以上步骤,直到 $C_{k+1} = \emptyset$,最后得到的就是包含最多属性的最大频繁 K - 项集 L。然后由最大频繁项集 L 结合最小置信度来产生模糊关联规则,最后再由兴趣度值来产生匹配规则集。

3　入侵检测模型

由于 MANET 网络的特殊性,基于数据挖掘的入侵检测的研究根据其数据源的不同主要分为基于网络的和基于主机的数据挖掘两种。对于 MANET 网络来讲,由于网络的异构性和节点的不稳定性,因此网络数据信息的规则属性很难统计。因此本模型的审计数据主要来自于节点主机。审计数据记录了用户使用计算机系统进行所有活动的过程,因为用户对系统的任何一次操作都是由一段程序来反映的,整个程序的调用虽然具有一定的随机性,但是这些程序形成的调用序列通过分类形成的子序列却具有相当的稳定性。对这些稳定序列形成的数据库使用数据挖掘算法中的关联分析和序列挖掘,提取用户所执行命令中存在的相关性和规律,针对每个用户构建其正常行为的行为模式,对每一次用户的操作都会挖掘出其中包含的当前模式,将当前模式和历史模式进行比较,计算两者的相似度。相似度越高说明该用户的当前行为是正常的,否则,即为异常行为。典型入侵检测过程如图 1 所示。

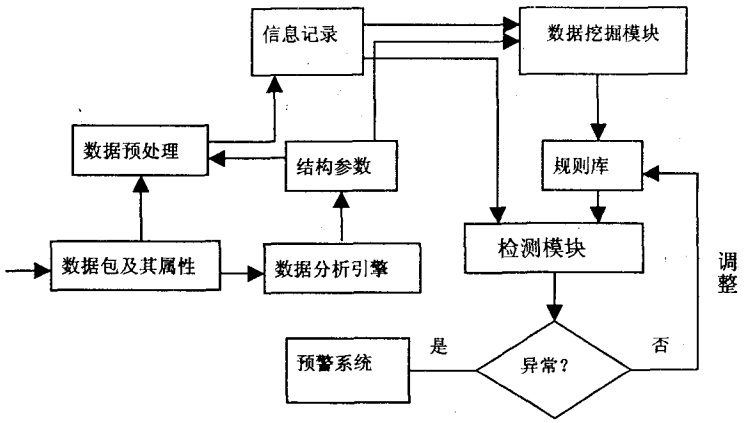

图 1　基于数据挖掘的入侵检测模型

4　仿真实验及结果分析

采用网络仿真器[14]进行 MANET 仿真,文献[14]描述的是一个无线网络仿真专用软件,可以选择仿真协议,具有大容量、高速度的特性,适于移动网络仿真。训练数据采用入侵检测离线数据集[15]。在该仿真软件中基于 AODV 协议实现了 3 种典型的攻击来研究算法的有效性和检测方法的可行性。这几种攻击类型是:①数据包丢弃的自私和拒绝服务攻击;②恶意的 IP sweep 攻击;③错误的源路由和最大序列号的黑洞攻击。

在实验中,使用正常运行的跟踪数据来训练异常检测模型,然后运行攻击并收集跟踪数据来评估入侵模型。由于设计的入侵检测模型是基于主机的,从关联分析角度来看,与系统调用有关的各个变量具有很强的相关性。因此实验设置一次系统调用表示为:[进程 ID　系统调用 ID　用户 ID　访问对象　访问权限]。选用的系统平台是 UNIX 操作系统。

实验结果发现,该模型对于攻击类型①和②的检测率都在 94.6% 以上,而对于攻击类型③检测率则只有 89.4% 左右,这种结果与攻击类型的技术复杂度密切有关,攻击类型①和②属于简单的直接攻击类型,特征明确,容易检测,而攻击类型③实现的技术相对复杂,特征相对较为隐秘,不易被发现。但是可喜的是,该模型的误报警率始终控制在 1.1% 以内,对于 MANET 网络来讲,这种检测性能算是一种理想状态[16]。

5 结束语

基于数据挖掘的 MANET 网络入侵检测技术已经成为近年来研究的一个热点。文中采用数据挖掘的算法思想,尽可能快速高效地从网络海量数据信息中挖掘出有效的规则数据,并进行规则匹配,在模拟的移动自组织网络中,初步实现了该模型思想。以后的研究方向主要有:

(1)进一步改进挖掘算法,提高挖掘效率;

(2)根据移动自组织网络特性,采用分布式方法改进入侵检测模型,把入侵检测尽可能部署到 MANET 网络的移动代理中。

参考文献:

[1] Ichlamtac, Mconti. Mobile Ad hoc Networking: Imperatives and Challenges[J]. Ad hoc Networks ,2003,1(1):13-64.

[2] 卿斯汉,蒋建春,马恒太,等.入侵检测技术研究综述[J].软件学报,2004,25(7):19-29.

[3] Agrawal R, Imielinski T, Swanmi A. Mining Association Rules Between Sets of Items in Large Database[C]//Proc. of ACM-SIGMOD on Management of Data. Washington D. C: [s. n.],1993:206-207.

[4] Agrawal R, Srikant R. Fast Algorithms for Mining Association Rules[C]//Proc. of VLDB. Santiago, Chile: [s. n.], 1994: 487-499.

[5] Han J, Cai Y, Cercone N. Data-driven Discovery of Quantitative Rules in Relational Databases[J]. IEEE Transaction on Knowledge and Data Engineering,1993,5(1):29-40.

[6] Agrawal R, Omiecinski E, Navathe S. An Efficient Algorithm for Mining Association Rules in Large Databases[C]// the 21st VLDB Conference. Zurich, Switzerland: [s. n.],1995.

[7] Inmon W H. The Operational Data Store[M]. New York: John Wiley & SonsInc,1996.

[8] Au Wai-Ho, Chan K C C. FARM: a data mining system for discovering fuzzy association rules[C]//The 1999 IEEE International Fuzzy Systems Conference. FUZZ-IEEE'99. Seoul, South Korea: [s. n.],1999:1217-1222.

[9] Orchard R. Fuzzy reasoning in jess: the fuzzy J toolkit and fuzzy jess[C]//Proceedings of ICEIS 2001,3rd International Conference on Enterprise Information Systems. Setubal, Portugal: [s. n.],2001:533-542.

[10] 朱天清,熊 平.模糊关联规则挖掘及其算法研究[J].武汉工业学院学报,2005,24(1):24-28.

[11] Kuok C, Fu A, Wong M. Mining fuzzy association rules in database[J]. The ACM SIGMOD Record,1998,27(1):41-46.

[12] 毛国君,刘椿年.基于项目序列操作的数据挖掘算法[J].计算机学报,2002,25(4):417-422.

[13] Pasquier N, Bastide Y, Taouil R, et al. Efficient Mining for Association Rules Using Closed Item set Lattices[J]. Information Systems,1999,24(1):25-46.

[14] Massachusetts Institute of Technology. Lincoln Laboratory Data Sets[EB/OL]. 1999. http://www.11.mit.edu/IST/ideval/data/.

[15] Scalable Network Technologies[EB/OL]. 2005-03-10. http://www.qualnet.com.

[16] Huang Y A, Lee W. A Cooperative Intrusion Detection System for Ad Hoc Networks[C]//Proceedings of the 1st ACM Workshop Security of Ad Hoc and Sensor Networks. Fairfax, Virginia: [s. n.],2003.

(上接第 134 页)

名的盲性。私钥只有签名者自己知道,别人伪造他的签名也是不可能的,方案的安全性依然基于圆锥曲线上离散对数的困难性。

4 结束语

对有限域上的圆锥曲线进行了分析,指出参数选择对其安全性的影响,针对圆锥曲线上的离散对数难题,提出一种数字签名方案,并实现了 Schnorr 盲签名在圆锥曲线上的模拟。为其他数字签名和盲签名方案在圆锥曲线上的模拟提供了可能性。另外,在本方案中,如何取得阶数比较大的基点 P 值得进一步研究。

参考文献:

[1] 张明志.用圆锥曲线分解整数[J].四川大学学报:自然科学版,1996,33(4):356-359.

[2] 曹珍富.基于有限域 F_p 上圆锥曲线的公钥密码系统[C]//刘木兰等编.第五届中国密码学学术会议论文集.北京:科学出版社,1998:45-49.

[3] 曹珍富.RSA 与改进的 RSA 圆锥曲线模拟[J].黑龙江大学自然科学学报,1999(4):15-18.

[4] Dai Zong-duo, Pei Ding-yi, Yang Jun-hui, et al. Cryptanalysis of a public key cryptosystem based on conic curves [C]//Proceeding of Cryp TEC'99: International Workshop on Cryptographic Techniques and E-Commerce. Hong Kong: City University of Hong Kong, 1999:259-261.

[5] 孙 琦,朱文余,王 标.环上圆锥曲线和公钥密码协议[J].四川大学学报:自然科学版,2005,42(3):471-478.

[6] 许旭东,靳岩岩,赵 磊.圆锥曲线公钥密码算法参数选择[J].计算机工程,2007,33(15):159-160.

[7] 王化群,张力军,赵君喜.基于椭圆曲线的 Schnorr 盲签名[J].计算机工程与设计,2005,26(7):1819-1822.