文章编号:1001 - 9081(2005)11 - 2557 - 02

基于机器学习的移动自组织网络入侵检测方法

杨德明[1,2],潘 进[2],赵 爽[3]
(1. 西北工业大学 自动化学院,陕西 西安 710072; 2. 西安通信学院 网络工程教研室,陕西 西安 710106;
3. 92493 部队 通信站,辽宁 葫芦岛 125000)
(dmyang21@yahoo.com.cn)

摘 要:移动自组织网络是由无线移动节点组成的复杂分布式通信系统。研究了移动自组织网络的入侵检测问题,采用了一种新型的基于机器学习算法的异常入侵检测方法。该方法获取正常事件的内部特征的相互关系模式,并将该模式作为轮廓检测异常事件。在 Ad-hoc 按需距离向量协议上实现了该方法,并在网络仿真软件 QualNet 中对其进行了评估。

关键词:移动自组织网络;异常入侵检测;机器学习
中图分类号: TP393.08 **文献标识码**:A

Intrusion detection method for mobile ad-hoc networks based on machine learning

YANG De-ming[1, 2], PAN Jin[2], ZHAO Shuang[3]
(1. *College of Automation, Northwestern Polytechnical University, Xi'an Shaanxi* 710072, *China*;
2. *Staff Room of Network Engineering, Xi'an Institute of Communication, Xi'an Shaanxi* 710106, *China*;
3. *Communication Station*, 92493 *Unit of PLA, Huludao Liaoning* 125000, *China*)

Abstract: Mobile ad-hoc networks (MANETs) represent complex distributed communication systems comprised of wireless mobile nodes. Based on the discussion of intrusion detection problem in MANET, a novel anomaly intrusion detection method based on machine learning algorithm was proposed to detect attacks on MANET. The method captured the normal traffic's inter-feature correlation pattern which could be used as normal profiles to detect anomalies caused by attacks. The method was implemented on Ad-hoc On-Demand Distance Vector (AODV) protocol and evaluated in QualNet, a leading network simulation software.

Key words: mobile ad-hoc networks, anomaly intrusion detection, machine learning

0 引言

移动自组织网络(mobile ad-hoc networks, MANET)又称为多跳网络(multi-hop network)、无基础设施的网络(infrastructureless network)或自组织网络(self-organized network)。MANET 是一种由移动节点组成的临时性自治系统,作为一种无线移动网络,MANET 和传统的移动网络有许多不同,其中主要的区别就是 MANET 不依赖于任何固定的网络设施,而是通过移动节点间的相互协作来进行网络互联。目前 MANET 在军事和民用上,特别是对安全敏感的环境,以及一些需要紧急组网情况下均有重要应用。同时 MANET 也正逐步应用于商业环境中,比如传感器网络、虚拟教室和家庭网络等。由于这种网络的特点和广泛应用,使得 MANET 的安全问题尤为突出。传统无线网络存在的一些安全问题,在 MANET 里仍然存在。由于 MANET 的特殊性,还存在一些特有的安全挑战:

1) 由于无线链路的开放性,导致信息的被侦听和泄漏,无法构建明确的安全防线;

2) MANE 的多跳路由特性,使得路由等信息在网络中传播时,受到路由路径上所有节点的查看;

3) MANET 的无中心特性,无法在网络运行中到某一特定的认证服务器上在线认证节点的合法性;

4) MANET 的动态网络拓扑特性,使得其路由协议的传播会泄漏节点的位置信息,而且攻击者又可以及时改变自身位置,很难发现和跟踪;

5) 由于 MANET 的自组织特性,任何一个节点都可以从自身的利益出发,做出一些对于整个网络不利的行为,这就给网络的监管带来了很大的难度。

入侵检测技术[1]能有效地检测到网络攻击并作出响应,是网络安全体系中的重要组成部分。但是以往的入侵检测技术的研究基本都是建立在有线网络上的,MANET 具有全新的网络特性和特殊的安全需求,因此需要开发新的入侵检测体系结构和算法。本文采用了一种新型的基于机器学习算法的异常入侵检测方法,能够有效检测路由层的入侵。

1 MANET 中的路由

路由协议是 MANET 的关键技术之一。目前对 MANET 中的入侵检测的研究基本都建立在路由层上。本文采用的路由协议为 Ad-hoc 按需距离矢量(Ad-hoc On-demand Distance Vector, AODV)协议[2]。AODV 协议是一种按需路由协议,该类路由协议只有在需要时才请求新的路由,简单并且性能良好。

1.1 AODV 协议

收稿日期:2005 - 05 - 18;**修订日期**:2005 - 07 - 29

作者简介:杨德明(1977 -),男,辽宁兴城人,博士研究生,主要研究方向:网络与信息安全; 潘进(1960 -),男,广东广州人,教授,博士,主要研究方向:智能信号处理、信息安全; 赵爽(1976 -),男,辽宁锦州人,主要研究方向:军事通信.

AODV 的路由发现是一个按需的反应式的过程。AODV 协议中每个节点都维护一个路由表,路由表记录了网络中所有可达节点的下一跳、距离和序列号。路由表用于响应 ROUTE REQUEST,然后决定本节点接收的数据包所要转发的下一跳。

1.2 针对 MANET 路由协议的攻击

目前已知的针对 MANET 路由协议的攻击有很多种。在各种路由协议中,路由器一般具有以下两个功能:为进一步的数据包转发建立路径和基于建立好的路由转发数据包。据此,将针对 MANET 的路由协议的攻击分为两类:

路由逻辑破坏攻击 这类攻击在网络中插入不正确的路由信息来破坏网络的结构或者使网络瘫痪,这类攻击也可以进一步分为外部攻击和内部攻击。例如,黑洞攻击就是一种典型的路由逻辑攻击,恶意节点宣称其拥有到所有节点的最短路径。黑洞攻击可以造成拒绝服务。

通信量破坏攻击 这类攻击可以侦听网络通信量,控制或者破坏包头或者包的内容,或者出于恶意的目的阻塞特定类型的通信。典型的例子有身份伪装攻击和分组丢弃攻击等。

2 基于机器学习算法的 MANET 入侵检测方法

在正常的行为当中存在很强的特征关联,这些关联可以用来检测由异常行为造成的背离。我们采用交叉特征分析的方法[3]来研究每个特征与其他所有特征之间的关联。交叉特征分析的基本思想是解决分类问题 $\{f_1, f_2, \cdots, f_{i-1}, f_i, f_{i+1}, \cdots, f_L\}$,这里 $\{f_1, f_2, \cdots, f_L\}$ 是特征向量。在入侵检测领域,我们将通过已知的系统信息,也就是特征,来研究系统的安全状况,即系统的分类。

入侵检测技术一般分为误用入侵检测(misuse detection)和异常入侵检测(anomaly detection)两种。误用入侵检测主要根据已知的特征来检测入侵行为,该方法的有效性取决于特征库。而异常入侵检测是基于入侵的行为进行检测,可以检测到未知的入侵。MANET 是一种新型的无线网络,正处于快速的发展阶段,新型的针对 MANET 的入侵和攻击可能层出不穷,建立完备的入侵特征库将是非常困难的工作。因此,我们采用异常入侵检测算法来检测针对 MANET 的入侵行为。虽然异常入侵检测容易存在误报警率较高的问题,但是在 MANET 中,通过合理设置阈值,可以将误报警率控制在可容忍的范围。

异常入侵检测的一个基本假设是正常事件和异常事件能够根据相应的特征向量将彼此区分开。也就是,给定一个特征向量,就可以精确地判断出其相关的事件是否正常。基于这样的假设,称与正常事件相关的特征向量为正常向量;类似地,称不与任何正常事件相关的特征向量为异常向量。为了便于分析和处理,这里假设所有的特征值都是离散的。

对于所有的正常向量,选择一个特征作为目标来进行分类,然后采用所有正常向量计算出一个模型,根据其余的特征值来预测所选目标的特征值。也就是,训练一个分类模型 C_i: $\{f_1, \cdots, f_{i-1}, f_{i+1}, \cdots, f_L\} \rightarrow f_i$。对于正常事件,由 C_i 产生的预测值很可能与特征的真实值相同;然而,对于异常事件,预测值与特征的真实值很可能不同。这是因为 C_i 是由正常数据训练成的,它们的特征分布和模式与异常行为不同。这意味着当采用 C_i 测试正常向量的时候,f_i 的预测值和真实值相匹配的可能性很大,而对于异常向量,这种可能性会很低。因此,通过研究结果匹配的程度,很容易发现正常模式和异常模式之间的不同。我们将上面的模型命名为关于 f_i 的子模型。显然,依靠一个这样的子模型是不够的,因为没有考虑该特征和其他特征的相互关系。因此,对于每个特征都要建立模型,直至 L 个子模型都训练完。这就完成了交叉特征分析方法的第一步,也就是训练过程,该过程可以用算法 1 表示。

算法 1:

$\forall i$, train C_i: $\{f_1, \ldots, f_{i-1}, f_{i+1}, \ldots, f_L\} \rightarrow f_i$

return $C_1, \ldots, C_L$

当所有子模型训练完毕后,采用如下方法分析跟踪日志。当分析一个事件时,在所有子模型中应用特征向量。当预测值与目标特征的真实值相匹配时就计数,然后将计数值除以 L,将输出称为平均匹配数。实际上,并不需要所有的子模型都匹配,只需要一个合适的决策阈值。当且仅当平均匹配数低于该阈值时,认为事件是异常的。开发一个理想的解决方案来决定异常检测的决策阈值是十分困难的。对于异常入侵检测,是允许一个小的误报警率的。我们采用如下方法来决定阈值:对于所有的正常事件计算平均匹配数,用结果的低边带值减去误报警率就是决策阈值。算法 2 列出了测试的过程,为简便起见,$f_i(x)$ 表示事件 x 的特征 f_i 的值,当断言 π 为真时$[\![\pi]\!]$返回 1。

算法 2:

$\sum_i [\![C_i(x) = f_i(x)]\!]/L \rightarrow AvgMatchCount$;

if $AvgMatchCount \geqslant \theta$ then return "normal";

else return "anomaly";

算法中,$C_1, \cdots, C_L$ 为分类器,事件 $x = (f_1, \cdots, f_L)$,决策阈值为 θ,$AvgMatchCount$ 为平均匹配数。

对算法 2 的改进是采用概率代替 0、1 计数值,每个可能的类的概率值由归纳学习器得出,例如决策树等,这个方法可以提高检测的准确性。设 $p(f_i(x) \mid x)$ 是目标特征属于正确的类的概率。定义所有分类器中概率值的平均输出为平均概率。算法 2 的改进版本如算法 3 所示。

算法 3:

$\sum_i p(f_i(x) \mid x)/L \rightarrow AvgMatchCount$;

if $AvgMatchCount \geqslant \theta$ then return "normal";

else return "anomaly";

算法 2 可以看作是算法 3 的特例。

下面讨论如何采用分类算法计算概率函数。本文选择决策树学习器[4]计算概率函数,选用 C4.5[5]算法。C4.5 算法是机器学习中的经典算法,在归纳学习中,它通过对一组训练数据的学习,构造出决策树形式的知识表示,在决策树的内部结点进行属性值的比较并根据不同的属性值判断从该结点向下的分支,在决策树叶节点得到结论。所以从根到叶结点的一条路径就对应着一条规则。设 n 是一个叶节点中实例的总数,n_i 是相同的叶节点中具有类标识 l_i 的实例的数目,那么 $p(l_i \mid x) = \frac{n_i}{n}$ 是 x 属于类 l_i 的概率。

3 仿真实验及结果分析

采用网络仿真软件 QualNet3.8[6] 进行 MANET 仿真。QualNet 是无线网络仿真软件 GloMoSim 的商业化版本,具有大容量,高速度的特性,适于移动网络仿真。

根据攻击的分类,将特征也相应分为与通信量无关的特征和与通信量有关的特征两类。与通信量无关的特征主要描述网络的拓扑和周期更新的路由结构。网络中的数据包来自不同的网络层和数据源,基于这个原因,构建与通信量相关的

(下转第 2576 页)

3 安全性分析

SAML 构建在需要公钥基础结构(PKI)的基础之上,以提供数字签名和 SAML 断言的加密。在交换安全信息的过程中,必须对发送端和接收端的身份以及 XML 文档的创建者进行身份认证。SAML 规范本身没有定义新的认证方法,而是采用已有的认证方式进行认证,如:Password,Kerberos,SSL/TLS 和 XML Digital Signature。

在 SAML 辅件方式下,除了保证断言信息的安全性,还要保证辅件的保密性和完整性。下面是基于 SAML 辅件单点登录系统可能面临的威胁及其相应的策略。

3.1 窃取辅件

威胁:偷听者窃取用户的辅件并企图利用该辅件冒充真实用户获取目标站点资源。

应对策略:

1)站点和用户浏览器之间传送辅件时,采用 HTTP/SSL 来保证辅件的保密性,防止偷听者窃取辅件。

2)源站点设置从发送出辅件到收到目标站点传来辅件的最大时间差,若超过此值,则认证该辅件无效,不向目标站点提供该辅件对应的断言。

3)目标站点限定允许的 IP 地址范围,不接收在该范围以外的浏览器发送的辅件信息。

3.2 攻击 SAML 协议的信息交换

威胁:在目标站点和源站点交换断言信息的过程中,可能遭受多种方式的攻击,包括 SAML 断言的窃取、报文的篡改、重放攻击以及中间人攻击(man-in-the-middle attack)。

应对策略:在源站点和目标站点的信息交换过程中采用双向身份认证,以保证报文的机密性和完整性。

3.3 SAML 辅件的重用

威胁:用户浏览器从源站点中得到辅件信息,该信息可长期保存于用户系统中,该辅件在以后可能会被重用。

应对策略:对辅件实行“一次性使用”原则,即一旦源站点从目标站点处获得辅件并将其对应的断言发送至目标站点后,源站点将些辅件信息予以删除,从而防止此辅件被重用。

4 结语

SAML 为用户实现单点登录提供了安全、方便的实现方式,而 SAML 辅件作为 SAML 规范中非常重要的组成部分,很巧妙地解决了浏览器无法使用 GET 方式传送 SAML 断言的问题。SAML 作为安全性服务语言,已被广泛地应用于电子商务、电子政务等方面,许多大公司,如 SUN、IBM 等都推出了基于 SAML 的单点登录产品,相信随着 SAML2.0 新标准的推出,将会进一步推动 Web 服务的发展。

参考文献:

[1] HUGHES J, MALER E. Technical Overview of the OASIS Security Assertion Markup Language(SAML)[EB/OL]. http://www.oasis-open.org/committees/documents.php?wg_abbrev=security, 2004.

[2] MALER E, PHILPOTT R. Assertion and Protocol for the OASIS Security Assertion Markup Language(SAML)[EB/OL]. http://www.oasis-open.org/committees/documents.php?wg_abbrev=security, 2003.

[3] MALER E, PHILPOTT R. Binding and Profiles for the OASIS Security Assertion Markup Language(SAML)[EB/OL]. http://www.oasis-open.org/committees/documents.php?wg_abbrev=security, 2003.

[4] MALER E, PHILPOTT R. Security and Privacy Considerations for the OASIS Security Assertion Markup Language(SAML)[EB/OL]. http://www.oasis-open.org/committees/documents.php?wg_abbrev=security, 2003.

[5] GROB T. Security Analysis of the SAML Single Sign-on Browser/Artifact Profile[A]. Proceedings of the 19th Annual Computer Security Applications Conference(ACSAC 2003)[C], 2003.

[6] PFITZMANN B, WAIDNER M. Analysis of liberty single-sign-on with enabled clients[J]. Internet Computing, IEEE, 2003, 7(6).

(上接第 2558 页)

特征,该特征可以定义为向量 < 数据包类型, 流向, 采样周期, 统计尺度 > 。

表 1 与通信量相关的特征

维	取值
数据包类型	数据、路由、ROUTE REQUEST、ROUTE REPLY、ROUTE RRROR 和 HELLO 消息
流向	接收、发送、转发和丢弃
采样周期/s	3,60 和 900
统计尺度	计数和内部数据包的标准差

例如,用于计算每 5s 接收到的 ROUTE REQUEST 数据包的内部包间隔的标准差的特征,可以编码为 <2, 0, 0, 1 > 。与通信量相关的特征的总数为 6×4×3×2=144 个。

在 QualNet 中基于 AODV 路由协议实现了 4 种典型的攻击,来研究算法的有效性。这 4 种攻击分别为:采用错误的源路由和最大序列号的黑洞和睡眠剥夺攻击(攻击类型 1)、采用数据包丢弃的自私和拒绝服务攻击(攻击类型 2)、采用恶意泛洪的睡眠剥夺攻击(攻击类型 3)和采用欺骗的路由环路(攻击类型 4)。

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模型。例如,如果一个仿真的整个运行时间是 1000s,采样速率是 5s,那么跟踪数据就具有 200 个数据点或者事件。当评估异常检测模型时,统计被正确检测到的异常事件数量和被错误判定为异常事件的正常事件的数量,然后计算出检测率和误报警率。经过实验,模型对于攻击类型 2 和 3 具有很好的性能,检测率都在 97% 以上,这是因为这些攻击所采用的技术简单而且造成的后果很明显。然而对于攻击类型 1 和 4,由于采用的技术复杂精巧而且造成的后果与前两种类型相比也不够明显,模型的性能稍差,检测率约为 90%。实验中,误报警率始终低于 0.98%,属于可以容忍的范围。

参考文献:

[1] 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.

[2] PERKINS CE, ROYER EM. Ad-hoc on demand distance victor routing[A]. The IEEE Workshop on Mobile Computing Systems and Applications(WMCSA)[C]. New Orleans, LA, 1999.

[3] HUANG Y-A, LEE W. A Cooperative Intrusion Detection System for Ad Hoc Networks[A]. Proceedings of the 1st ACM Workshop Security of Ad Hoc and Sensor Networks[C]. Fairfax, Virginia, 2003.

[4] 杨善林, 倪志伟. 机器学习与智能决策支持系统[M]. 北京: 科学出版社, 2004.

[5] QUINLAN JR. Induction of Decision Trees[J]. Machine Learning, 1986, (1): 81-106.

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