

用语义网技术建模特征识别和攻击模型

黄俊,张雷

(中国计量学院 计算机科学系,浙江 杭州 310018)

(leizhang4662@sina.com)

摘要:安全特征识别和攻击的预测是网络安全领域内必不可少的功能部分,而攻击模型和其他安全特征的描述和定义需要专门的语言。然而,目前此类语言存在诸多问题,如语言功能单一,适用性差;缺乏开放性,语义不一致和缺乏可重用性等。为了改善这种情况,利用本体建模方法,通过一个典型攻击行为的建模,证明本体语言具有的特点支持其作为安全特征描述语言。

关键词:本体;网络安全;攻击语言;特征识别

中图分类号: TP182; TP393.08 **文献标识码:** A

Using semantic Web technology to build signatures identification and attack model

HUANG Jun, ZHANG Lei

(Department of Computer Science, China Jiliang University, Hangzhou Zhejiang 310018, China)

Abstract: Signatures identification of security and attack forecast are integrant function parts of network security field, and the description and definition of attack models and other security signatures request special language. But there exist many questions on the current such languages, such as solitary function of language and weak adaptability; lack of openness and semantic coherence, and absent reuse ability. In order to improve this state, the ontology's modeling means were used. It was demonstrated that ontology's language is fit for the description and definition of attack models and other security signatures by a representative attack.

Key words: ontology; network security; attack language; signatures identification

0 引言

防火墙和入侵检测作为网络与信息安全领域的两项重要技术,是构筑安全防护体系的重要组成部分。在防火墙中,过滤规则设计以及其形式化定义需要一种简洁高效的途径,以保障防火墙的工作效率。而IDS(Intrusion Detected System)系统,更需要专门的语言来建模攻击的特征,定义入侵检测系统的事件以及它们之间的通信,描述一个攻击进行的步骤、报告和响应攻击等。目前,已经开发了若干此类语言,但依然存在诸多问题,如语言功能单一,适用性差;缺乏标准性的语言;语法复杂,难于理解和使用;缺乏开放性,语义不一致和缺乏可重用性等。

语义网^[1]技术的发展,尤其是本体(Ontology)^[1]技术的应用为解决这些问题提供了一种技术支持。在语义网中,本体扮演关键角色,它为机器提供了可读的词汇,以便应用程序理解共享的含义。本体是共享概念的规范、精确的描述。本体本身并不是静态的模型,它具有映像和进化的功能。本体映射就是重用已存在的本体,通过一定的方法对它们进行展开和组合,集成不同领域的本体以实现一个更大的信息和知识池,从而支持新的交流和使用。同样,本体进化就是在获得新的信息和知识时适当地维护和扩充已有本体。本体支持基于领域知识的推理功能,可以有效实现自动化,这也为提供智能防火墙架构和IDS架构提供了可能。OWL(Web Ontology

Language)是一种基于描述逻辑的本体语言,由W3C专门设计用于语义网,目前已被作为推荐标准。本文尝试利用本体语言作为安全系统中的建模语言,描述安全系统中策略、规则以及复杂的关系模式,从而为解决上述问题提供新的思路。

1 相关内容研究

1.1 安全领域内的一些建模语言

入侵检测系统的攻击模型用“签名”(Signatures)来表达,这些签名的说明和描述语言称之为攻击语言(Attack Language)。攻击语言可分为6种,分别是事件语言、响应语言、报告语言、关联语言、漏洞利用语言和检测语言^[2-4]。

根据表1可以看出,每种攻击语言只描述攻击的某一种信息或者片面的功能,没有一种语言能够全面地描述所有的攻击特征,从而可以作为一种通用的入侵检测语言。理想的检测语言应提供一种抽象的、IDS独立的描述攻击签名的方式,还要包括在IDS之间的一体化攻击描述,而这样的检测语言需要具备如下的条件^[2]:

简洁 语言应该只提供描述攻击场景的特征。

表达能力 语言应该能够表达任何可检测到的攻击签名。

精确 语言应该有严格定义的、与实现无关的语法和语义,任何攻击的描述都是无二义性。

可扩展性 新的领域需要新的事件类型,并需要相关的谓词和函数。语言应具备定义良好的、简洁的扩展方式。

收稿日期:2005-01-04;修订日期:2005-03-15

作者简介:黄俊(1961-),男,湖北武汉人,高级工程师,主要研究方向:网络应用;张雷(1979-),男,河南周口人,讲师,主要研究方向:模式识别、网络应用。

可执行性和可译性 应该可以轻松地将攻击描述整合到 IDS 应用中,描述应该可执行,或者能够有效地实现。

移植性 语言处理工具应该适应不同的环境。

异质 语言能够使用多个域的概念来描述攻击,例如 IP

包和主机审核记录。

满足上述条件就需要一种能够精确地描述和表达攻击签名,具备良好可移植性和可扩展性,提供异构环境下领域知识建模的语言。

表1 攻击语言简单分类表

项目	功能	典型实例
事件语言	事件语言用于描述事件,重点是对数据格式的描述	BSM 审计记录描述, TcpDump 包格式
响应语言	用来描述系统反应行为	C, Java
报告语言	描述报警格式;报警一般包含攻击的一些基本信息	CISL, IDMEF
关联语言	通过分析几个 IDS 提供的报警,明确攻击之间的关系,达到识别协同攻击的目的;是对已有入侵检测技术的更高层次的抽象	贝叶斯网络, Honeywel 的 ARGUS 基于时间的推理, UCSB 的 STATL 基于规则的推理, SRI 的 P-Best
漏洞利用语言	用于描述攻击实施的步骤	常用的可执行语言; CASL/NASL
检测语言	用于描述攻击特征,提供了发现攻击的机制和抽象方法	P-Best, STATL, Snort, SNP-L, N-Code, Kumar, BRO

1.2 语义网和本体理论

语义网技术是人工智能技术和 Web 技术发展结合的产物,它是对现有万维网的扩展,网上信息具有良好的定义,以一种不仅可读而且是机器可理解的形式表达,从而使机器和人类能够更好地彼此合作。本体技术在语义网中扮演了关键角色。Ontology 是一种明确的共享概念化的形式说明。概念化是指对现实世界中一些事务进行抽象建模,所建立的模型确定了该事物一些相关的概念,明确意味着所使用概念的类型以及它们使用上的约束都有显式的定义。形式说明则是指 Ontology 应该是形式化及其可理解的。共享则是本体表达双方都可认可的知识,可以在人和应用系统之间达成对术语含义的共享和共同理解。

目前产生了一系列本体描述语言,如 RDF, DAML, OIL, OWL^[5]等。OWL 是基于 RDF 和 RDF Schema 的本体语言,它是在 DAML 和 OIL 的基础上产生的,基于一类描述逻辑 SHIQ (D)^[5],可以用类和属性描述领域结构,具有同描述逻辑相同的语义表达能力,因此可以利用描述逻辑推理器做基于 OWL 描述的知识推理,这为 IDS 中基于不同的攻击特征做攻击推理提供了技术基础。

2 建模过程

用本体建模系统,信息分类是至关重要的。本体建模首先要抽象信息模型内部的关键概念,定义类来表达概念,定义属性来表示概念之间的关系,定义领域知识的公理作为推理规则。本体的开发类似于面向对象设计方法学,但本体开发本身属于知识工程方法论,目前已经有一些方法论用于本体的开发^[6]。信息分类中需要定义本体类和类层次关系。

目前主要有三类层次开发方法:

自上而下方法 开发过程始于定义最一般和最抽象的领域概念,随后对概念进行说明。

自下而上方法 开发过程始于定义最特殊的类,即定义类层次中的叶子层次,然后对这些分散的类进行分组,抽象出更一般的概念。

综合方法 其开发方式是前两种开发方式的结合。例如可以先定义最突出的概念,然后进行专门归纳。

可以随意采用任何一种方法对安全领域内的信息进行本体概念的分类。这里主要集中在对入侵检测攻击,防火墙过滤以及病毒识别方面。主要关注 IDS 攻击模型的建模,因为 IDS 包含内容很多,最适合于本体建模,而且通常的 IDS 包含有病毒识别的功能。

目前黑客攻击手段主要利用开放共享、薄弱密码、编程漏洞或者系统陷门、拒绝服务、缓冲区溢出、社会工程和恶意代码等。开放共享途径包括传统的网络文件系统、远程信任登录和当前日益流行的 P2P 共享。薄弱密码是指设定密码组合简单、密码强度小,黑客可以通过强力猜测手段来破解。因此开放共享途径和薄弱密码可归为薄弱点利用这个大类;编程漏洞是编程过程中留下的漏洞或者是隐藏的 Bug,系统陷门则是指系统在开发时为了调试方便而设定的特殊功能或程序,在系统开发完毕提交的时候,如果陷门没有及时关闭就可能成为黑客攻击的途径。缓冲区溢出攻击主要是由于字符串操作函数对缓冲区操作未作边界检查导致的,攻击者通常通过此方式获取到很高的系统运作权限。编程漏洞/系统陷门与缓冲区溢出可以归为漏洞利用这个大类。社会工程则是攻击者通过非技术手段来辅助实施攻击。

通过对攻击途径分类,可以有效地建立起攻击类型划分的知识层次,同时为攻击实例提供划定的父类型。

攻击实施时伴随大量的资源状态的改变。这些资源状态包括网络资源状态和主机资源状况等,而前者又包括各种协议状态(各种连接状态)、网络设备状态(路由器、网络接入设备的负载)等;后者包括操作系统状态和运行其上的进程(线程)状态。攻击在实施过程中其特征除了其使用的途径、方法和工具所具备的内在的固有特征外,还表现为资源状态的改变,因此资源状态往往也成为攻击模型中的重要组成部分。比如蠕虫病毒在入侵网络系统的时候,除了可以根据蠕虫病毒代码指纹特征识别外,往往还可以通过检测病毒入侵的主机状态来识别,如邮件蠕虫病毒发作的时候可以通过观察网络中邮件协议包数量的异常来决定病毒发作情况。因此可以通过为入侵行为的过程建模来模拟整个攻击模式。

假使用 P 来表示攻击判定的前提条件(PreCondition), A (Attributes) 表示攻击行为的特征, S (States) 为攻击实施资源状态,其中 $S \in A$ (即资源状态为特征的一个子概念), C (Consequence) 为攻击行为的判定,则一个攻击的检测判定可以形式化表示为 (P, A, S, C) 或者 (P, A, C) 其中 $S \in A$ 。可以用本体来定义攻击模型的主要元素及其之间的关系。

图1中的“前提条件”对于入侵判定来说,可以指两个方面。当一个入侵可以通过简单的规则,比如一条规则来描述的时候,一般是攻击判定的输入条件,这些输入条件就是资源状态的实例;当一个入侵判定需要一个复合的过程时,它不仅包括简单判定的输入条件,还包括先前判定的结果作为后续

时内在继承了其父类蠕虫实例的属性。这里只是一个简单的事例,复杂的蠕虫病毒体系定义可以通过本体开发方法来完成,并相应地建立起关于蠕虫病毒的精确、简洁和可重用的特征知识库,来服务于病毒的检测识别。

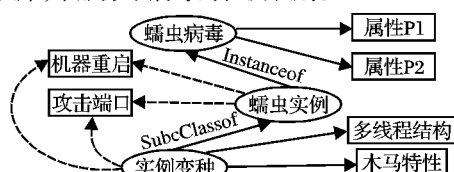


图3 本体建模蠕虫病毒示意图

4 结语

本体提供领域范围内精确、共享的形式化概念。本体具备诸多机制如本体影射、本体重用、本体进化等,文中也给出了示例。本文尝试引入本体语言作为安全系统中的建模语言,描述安全系统中的策略、规则以及复杂的关系模式,对于入侵检测系统和防火墙技术的研究会有积极的意义。

用本体描述的攻击特征,可以提供精确的定义,为安全系统提供统一的概念知识库,能有效地消除二义性。不少攻击语言用XML来描述和形式化,但是XML无法解决一词多义和多词多义的问题,此外,同XML对比,本体语言可以描述更加的概念结构和关系,具有更丰富的语义描述能力,而且W3C推荐的标准化本体描述语言如OWL,也是采用了XML的语言和语法格式,因此可以兼备XML的优势,便于交流和存储。

此外,本体本身支持基于描述逻辑的推理功能,这为智能到检测判定等方面,目前OWL设计可以用于Agent等之间的安全系统设计提供了天然的支持,其推理功能很容易用

交流,为未来将Agent等技术引入安全领域提供先天条件。

但本体开发目前还是一件困难的事情,需要领域专家的参与才能设计出高效的本体。文献[8]给出了一个IDS本体的描述,但主要是以主机入侵检测为中心构建的,我们给出了另外一种兼顾网络和主机入侵检测的本体描述。当然,这里的本体相对简单和不完善,还需要进一步的完善。

参考文献:

- [1] 邓志涛,唐世渭,张铭,等. Ontology 研究综述[J]. 北京大学学报, 2002, 38(5): 730-738.
- [2] ECKMANN S, VIGNA G, KEMMERER R. STATL: An Attack Language for State-based Intrusion Detection[J]. Journal of Computer Security, 2002, 10(1/2): 71-104.
- [3] 褚永刚,宋传恒,杨义先,等. 入侵检测系统攻击语言研究[J]. 信息安全与通信保密, 2003, 21(4): 40-42.
- [4] FEIERTAG R, KAHN C, PORRAS P, et al. A Common Intrusion Specification Language[EB/OL]. <http://www.isi.edu/~brian/cid-drafts/language.txt>, 1999-06.
- [5] MCGUINNESS DL, VAN HARMELEN F. OWL Web Ontology Language Overview, World Wide Web Consortium(W3C) recommendation[S]. www.w3.org/TR/owl-features, 2003.
- [6] NOY NF, MCGUINNESS DL. A Guide to Creating Your First Ontology[EB/OL]. <http://protege.stanford.edu/publications/ontology-development/ontology101-noy-mcguinness.html>, 2003.
- [7] ZENY G, BOLINGER D, SCHACKENBERG D. Communication in the Common Intrusion Detection Framework v 0.7[EB/OL]. <http://www.isi.edu/~brian/cid-drafts/communication.txt>, 1998-06.
- [8] UNDERCOFFER J, JOSHI A, PINKSTON J. Modeling Computer Attacks: An Ontology for Intrusion Detection[A]. Proceedings of RAID 2003, LNCS 2820[C], 2003. 113-135.

(上接第1553页)

对于DOS和Probe类攻击的识别,经过多次实验比较,BP网络决定采用41-45-35-3的双隐层的结构,隐含层使用正切特性的S型传递函数,输出层使用对数特性的S型传递函数,采用共轭梯度下降算法。训练所用时间是107.69s,达到的训练误差为 9.962×10^{-4} ,训练次数为101次。

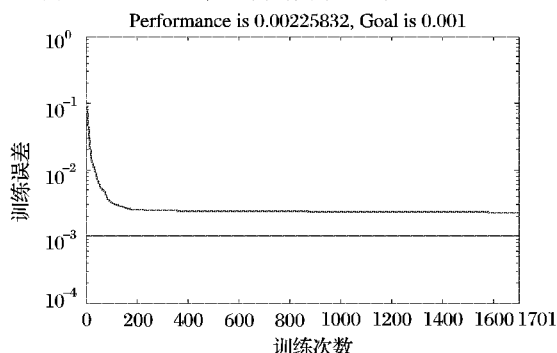


图7 BP网络对于U2R, R2L攻击的训练收敛曲线

表5 BP网络对U2R, R2L攻击的检测结果

实际类别	预测类别			
	Normal	U2R	R2L	% Correct
Normal	3990	3	7	99.75
U2R	13	40	12	61.54
R2L	1500	0	1000	40.0
训练时间	1022.2 s			

对于U2R和R2L类攻击的识别,BP网络采用41-50-40-3

的网络结构,隐含层使用正切特性的S型传递函数,输出层使用对数特性的S型传递函数,采用共轭梯度下降算法。

训练用时1022.2s,达到的训练误差为 2.258×10^{-3} ,训练次数为1701次。从图7中可以看出在训练200次左右后,误差收敛速度已趋于缓慢。从以上的实验结果可以看出,对于U2R和R2L攻击,由于它们与Normal数据的不可分性,使两种分类器的分类效果都不好;对于DOS和Probe攻击,在检测率上支持向量机略优于BP网络方法,在训练时间上支持向量机有着很大的优势。

SVM的核函数及参数的选择和神经网络的网络结构选择对于分类器的泛化能力都有很大的影响,如何选择合适的参数及网络结构,是值得继续研究的问题。

参考文献:

- [1] EUGENE S. Criss and aftermath[J]. Communications of the ACM, 1989, 32(6): 678-687.
- [2] ASAKA M, ONABUTA T, INOUE T, et al. A New Intrusion Detection Method Based on Discriminant Analysis[J]. IEEE Transactions on Information & System, 2001, E84-D(5): 570-577.
- [3] VAPNIK VN. 统计学习理论的本质[M]. 张学工,译. 北京:清华大学出版社, 2000.
- [4] SCHGLKOPF B, MIKA S. Input Space vs Feature Space in Kernel based Methods[J]. IEEE Transactions on Neural Networks, 1999, 10(9): 1000-1017.
- [5] <http://kdd.ics.uci.edu/databases/kddcup99/task.htm>[EB/OL], 2004-12.