

基于信息融合的网络安全态势感知模型

王选宏 肖 云¹

(西安邮电学院通信工程系, 西安 710121; 西北大学信息科学与技术学院¹, 西安 710127)

摘要 在分析已有的安全态势评估和预测方法的基础上, 提出了基于信息融合的网络安全态势感知模型。该模型采用 D-S 证据理论对多源网络安全数据进行融合, 计算漏洞、服务、主机、网络的安全态势值。同时根据历史安全态势评估结果, 利用支持向量回归理论对未来态势进行预测。相比已有的安全态势评估和预测方法, 该模型的结构更加完整, 结果更为准确有效。

关键词 态势评估 预测 D-S 证据理论 支持向量回归

中图法分类号 TP393.08; **文献标志码** A

随着计算机网络的迅速发展, 各种网络攻击事件不断发生, 网络安全问题日益成为人们关注的焦点。防火墙、入侵检测系统(IDS)等安全设备每天都会产生海量的告警信息, 这样使得网络管理员面对大量的告警信息很难了解系统的安全状况, 不能及时采取合适的响应措施。因此如何真实、准确地对网络安全态势感知已经成为网络安全领域的一个研究热点。安全态势感知(security situational awareness)是指通过技术手段从时间和空间纬度来感知并获取安全相关元素, 通过数据信息的整合分析来判断安全状况并预测其未来的发展趋势。近年来, 态势评估技术开始在计算机网络领域得到应用, 国内外的研究人员依据不同的技术思路, 设计并实现了大量针对计算机网络的安全态势感知方法。

Tim Bass 提出了基于多传感器数据融合的入侵检测框架, 并将该框架应用于下一代入侵检测系统和网络态势感知系统(Network Situation Awareness System, NSAS), 以实现入侵行为检测、入侵率计算、入侵者身份和入侵者行为识别、态势评估以及威胁

评估等功能^[1]。Stephen Lau 开发了“The Spinning Cube of Potential Doom”^[2]系统, 该系统在三维空间中用点来表示网络流量信息, 极大地提高了网络态势感知能力。SIFT^[3](Security Incident Fusion Tool)项目组设计开发一个安全事件融合工具的集成框架, 已开发的 Internet 安全态势感知软件有:NVisionIP, VisFlowConnect-IP, UCLog+ 等。这些软件工具评估指标较为单一, 对管理员经验水平要求也很高。

目前国内的相关研究主要是围绕网络安全态势评估、大规模网络预警等来开展的。西安交通大学实现了基于 IDS 和防火墙的集成化网络安全监控平台^[4,5], 该系统实现了态势评估。北京理工大学信息安全与对抗技术研究中心研制了一套基于局域网络的网络安全态势评估系统^[6], 由网络安全风险状态评估和网络威胁发展趋势预测两部分组成, 用于评估网络设备及结构的脆弱性、安全威胁水平等。胡华平等^[7]人提出了面向大规模网络的入侵检测与预警系统的基本框架及其关键技术与难点问题。上海交通大学和哈尔滨工程大学分别以 RBF^[8] 和 GA-BPNN^[9] 神经网络的方式来实现态势值的预测, 给出了态势预测问题的一个初步解决方案的探讨。韦勇等给出了基于信息融合的网络安全评估模型^[10], D-S 证据理论和时间序列分析法进行态势评估与预测。

2010 年 7 月 23 日收到 陕西省教育厅自然科学专项(08JK449)、

西安邮电学院中青年科研基金项目(ZL2008-19)资助
第一作者简介: 王选宏(1977—), 陕西武功人, 工程师, 硕士, 研究方向: 信号处理与信息安全。

以上方法为网络安全态势评估工作提供了可行的解决思路,为评估模型及算法的研究奠定了良好的基础,但也普遍存在着一些技术缺陷。例如,缺乏对网络安全因素的全面考虑,评估数据源单一,使得评估结果缺乏全面性;忽略了数据源之间的互补性和冗余性等内在联系,使得评估结果不够准确;所采用的量化算法存在一定缺陷,导致量化结果与实际结果出现偏差;另外,这些方法无法对安全状况的发展趋势进行准确的预测分析。

针对上述问题,本文提出了基于信息融合的网络安全态势感知模型,利用 D-S 证据理论将多数据源态势信息进行融合,获得多粒度的网络安全态势评估结果,评估对象为漏洞—服务—主机—网络,并绘制出各级对象的态势曲线图;根据获得历史网络安全态势值,使用支持向量回归理论对未来单位时间的安全态势值进行有效预测,从而实现网络安全态势的评估和趋势预测。

1 基于信息融合的网络安全态势感知模型

本文以多源网络安全信息为基本的数据对象,以信息融合的知识和方法为理论指导,建立如图 1 所示的层次化的网络安全态势感知模型。该模型从功能上自下而上分为三个层次,分别为多源信息层,网络安全态势评估层,和网络安全态势预测层。

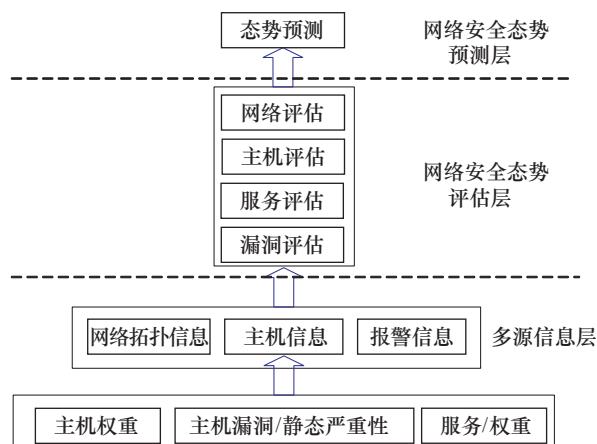


图 1 基于信息融合的网络安全态势感知模型

在多源信息层,可通过不同的数据接入方式,如专门的代理接口等,获取多源网络信息,包括:

- (1) 网络拓扑信息 I_T ,即网络所有物理链接关系集合;
- (2) 主机信息 I_H ,即包括主机权重 w_H ,主机漏洞 V_H 和漏洞静态严重性 V_s 组成的二元组 (V_H, V_s) ,服务 S_H 和服务权重 w_S 组成的二元组 (S_H, w_S) ;
- (3) 报警信息 I_A ,即对多个入侵检测系统产生的原始报警信息进行预处理,包括剔除无效时间戳报警;合并多个检测器对同一攻击的重复报警;主机报警按照是否存在相应漏洞的原则进行剔除或保留。

在网络态势评估层,首先使用 D-S 证据理论对多源信息进行融合处理,得到漏洞评估结果;其次考虑主机上存在的某项服务及其对应的漏洞集,采用求和法获得服务评估结果;再次考虑主机上存在的各项服务及其权重,采用加权求和法获得主机评估结果;最后考虑网络内各主机的权重,采用加权求和法得到网络安全评估结果。

在网络态势预测层,根据历史态势评估结果,采用支持向量回归理论得到未来单位时间内的态势预测结果。

2 基于 D-S 证据理论的网络安全态势评估算法

证据理论是由 Dempster 在 1967 年首先提出的,1976 年 Shafer 对 Dempster 的理论进行了扩充,在此基础上形成了一种处理不确定性问题的工具。因此,证据理论又称为 Dempster-Shafer 理论,简称 D-S 证据理论或证据理论^[11]。本文以 CERT 的漏洞静态严重性分值和入侵检测系统的报警统计数据作为证据,利用 D-S 证据理论融合后获得漏洞态势值,然后考虑主机存在的服务以及服务的权重获得服务安全态势和主机安全态势值,最后根据主机的安全态势值及权重获得网络安全态势值。

针对主机 H_i 的某一漏洞 V_i ,先对漏洞静态严重性证据和报警统计信息证据进行归一化处理。从

漏洞数据库查询到漏洞 V_i 的静态严重性分值,得到漏洞静态严重性证据 S_{Vi} :

$$S_{Vi} = \frac{SS_i}{180} \times 100\% \quad (1)$$

式(1)中, SS_i 表示漏洞 V_i 的静态严重性分值, $0 \leq SS_i \leq 180$ 。

从报警数据库中获得主机 H_i 使用漏洞 V_i 的报警统计信息,获得报警统计信息证据 A_{Vi} :

$$A_{Vi} = \frac{NA_i}{NA_a} \times 100\% \quad (2)$$

式(2)中, NA_i 表示单位时间内主机 H_i 上被利用漏洞 V_i 进行攻击后产生的报警数目,其中的单位时间可根据需要取小时、天、月等; NA_a 表示单位时间内主机 H_i 上被利用漏洞进行攻击后产生的报警总数目。

定义识别框架 $\Theta = \{\text{safe}, \text{unsafe}\}$, 其中 safe 表示在漏洞 V_i 存在的前提下计算机系统的安全状态,而 unsafe 表示在漏洞 V_i 存在的前提下计算机系统的不安全状态。相应的基本可信度分配函数为

$$\begin{cases} m_1(\text{safe}) = 1 - A_{Vi} \\ m_1(\text{unsafe}) = A_{Vi} \\ m_2(\text{safe}) = 1 - S_{Vi} \\ m_2(\text{unsafe}) = S_{Vi} \end{cases} \quad (3)$$

经过证据合成, $m(\text{unsafe})$ 计算公式如式(4)。

$$m(\text{unsafe}) = \frac{A_{Vi}S_{Vi}}{(1 - A_{Vi})(1 - S_{Vi}) + A_{Vi}S_{Vi}} \quad (4)$$

显然 $m(\text{unsafe})$ 的值越大, 系统越不安全, 漏洞 V_i 的态势值应该越高。因此漏洞 V_i 的态势值 E_{Vi} 可直接用 $m(\text{unsafe})$ 获取, 即有

$$E_{Vi} = m(\text{unsafe}) \times 100\% \quad (5)$$

基于 D-S 证据理论的网络安全态势评估方法的步骤归纳如下:

Step1 对待测主机 H_i 的某一漏洞 V_i 到漏洞数据库中查询出所对应的静态严重性分值, 并按照式(1)进行归一化处理。

Step 2 读取报警数据库中的报警统计信息, 并按照式(2)进行归一化处理。

Step 3 按照式(4)和式(5)计算漏洞 V_i 的态势值 E_{Vi} 。

Step 4 按照式(6)式获得服务 S_i 安全态势值 E_{Si} :

$$E_{Si} = \sum_{V_i \in S_i} E_{Vi} \quad (6)$$

式(6)中, $V_i \in S_i$ 表示服务 S_i 所对应的漏洞集合。

Step 5 按照式(7)获得主机 H_i 安全态势值 E_{Hi} :

$$E_{Hi} = \sum_{S_i \in H_i} w_{Si} E_{Vi} \quad (7)$$

式(7)中, $S_i \in H_i$ 表示主机 H_i 所有的服务集合。

Step 6 按照式(8)获得网络安全态势值 E :

$$E = \sum_{H_i \in N} w_{Hi} E_{Si} \quad (8)$$

式(8)中, $H_i \in N$ 表示网络 N 所有的主机集合。

3 基于支持向量机的网络安全态势预测算法

支持向量机(Support Vector Machine, SVM)是数据挖掘中的一项新技术, 是借助于最优化方法解决机器学习问题的新工具。SVM 理论最初于 20 世纪 90 年代由 Vapnik 提出, 近年的研究取得了很大进展, 在模式识别、时间序列预测、概率密度估计等领域得到了广泛的应用。

使用支持向量回归理论^[12]对安全态势进行预测, 即对于历史安全态势值所构成的样本, 考虑先前值对未来值的影响, 利用前 $n-1$ 时间单位(时间单位可选月, 天和小时)的历史值作为训练样本, 形成动态预测模型, 再用模型对下一个单位时间的态势值进行预测。模型选用支持向量回归算法 ε -SVR, 核函数选择 RBF 核函数, 实验参数中不敏感损失函数 ε 、惩罚系数 C 、核函数参数 σ 对模型的学习精度和推广能力的好坏起着决定性作用。本文参考 Melissen^[13]给出的 $\{C, \sigma, \varepsilon\}$ 的大致取值范围: $C = [1, 10^8]$, $\sigma = [0, 0.2]$, $\varepsilon = [0.01, 2.0]$, 利用试探法选择 $\{C, \sigma, \varepsilon\}$ 的值。预测结果的评价指标选用均方误差 MSE , 它是进行 n 次预测时误差平方的平均数。

基于支持向量机的网络安全态势预测算法归纳如下:

Step 1 选择预测对象: 服务、主机或网络, 根据

第2节所获得安全态势值,构造相应的安全态势值样本 $\{S_1, S_2, \dots, S_n\}$ 。

Step 2 利用前 $n-1$ 个态势值作为训练样本,并选择合适的实验参数获得训练模型。

Step 3 利用预测模型得到第 n 个态势预测值 \hat{S}_n 。

Step 4 计算多次预测结果的均方误差,评估预测结果的准确性。

4 结论

本文在分析网络安全态势感知研究现状的基础上,提出了一个基于信息融合的网络安全态势感知模型,该模型综合考虑多源网络安全信息,利用D-S证据理论获得网络安全态势评估结果,评估对象为漏洞、服务、主机、网络;利用支持向量回归理论获取未来一个单位时间内网络安全态势值。

本文所提出的基于信息融合的网络安全态势感知模型各部分的功能已基本实现,并且在实验室测试环境中可进行准确的评估和预测,取得了预期的效果。下一步的工作包括大流量环境下全面的功能测试,算法的进一步的完善等。

参 考 文 献

- 1 Bass T. Intrusion detection systems and multisensor data fusion; creating cyberspace situational awareness. Communications of the ACM,

- 2000;43(4):99—105
 2 Lau S. The spinning cube of potential doom. Communications of the ACM, 2004; 47(6):25—26
 3 Yurcik W. Visualizing NetFlows for security at line speed; the SIFT tool suite. Proceedings of the 19th Usenix Large Installation System Administration Conference (LISA), San Diego, CA USA, Dec, 2005; 169—176
 4 张慧敏,钱亦萍,郑庆华,等.集成化网络安全监控平台的研究与实现.通信学报,2003; 24(7):155—163
 5 陈秀真,郑庆华,管晓宏,等.层次化网络安全威胁态势量化评估方法.软件学报,2006;17(4):885—897
 6 北京理工大学信息安全与对抗技术研究中心.网络安全态势评估系统技术白皮书. <http://www.thinkor.com/product/download/>
 网络安全态势评估系统技术白皮书2.doc, 2005
 7 胡华平,张 怡,陈海涛,等.面向大规模网络的入侵检测与预警系统研究.国防科技大学学报, 2003;25(1):21—25
 8 任 伟,蒋兴浩,钱 锋.基于RBF神经网络的网络安全态势预测方法.计算机工程与应用,2006;31:136—138
 9 胡明明,王慧强,赖积保.一种基于GA-BPNN的网络安全态势预测方法. http://www.paper.edu.cn/paper.php?serial_number=200707-437, 2007-7-24
 10 韦 勇,连一峰,冯登国. 基于信息融合的网络安全态势评估模型. 计算机研究与发展,2009;46(3):353—362
 11 段新生. 证据理论与决策、人工智能. 北京:中国人民大学出版社,1993
 12 Vapnik V N. 统计学习理论的本质. 张学工,译. 北京:清华大学出版社,2000
 13 Melssen W J. Determination of optimal support vector regression parameters by genetic algorithms and simplex optimization. Analytical Chimica Acta, 2005;544(1):292—305

Network Security Situational Awareness Model Based on Information Fusion

WANG Xuan-hong, XIAO Yun¹

(Dept. of Communicate Engineering, Xi'an Institute of Post & Telecommunications, Xi'an 710121, P. R. China)

Dept. of Information Science & Technology, Northwest University¹, Xi'an 710127, P. R. China)

[Abstract] Based on analyzing the existing security evaluation and forecast methods, a network security situational awareness model based on information fusion is proposed. Using D-S evidence theory to fuse multi-source network security information, this model gets the security situational values of vulnerabilities, services, hosts and network. It uses support vector regression theory to forecast future security situation in view of historical security situational values. Compared with the existing security evaluation and forecast methods, the proposed model is more integrated, effective and accurate.

[Key words] situational evaluation forecast D-S evidence theory support vector regression