http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2015.0561

基于信息融合的网络安全态势量化评估方法

文志诚[1,2], 陈志刚[2,*], 唐军[3]

(1. 湖南工业大学 计算机与通信学院, 株洲 412007; 2. 中南大学 信息科学与工程学院, 长沙 410083;
3. 中车株洲电力机车研究所有限公司, 株洲 412001)

摘 要: 针对目前网络安全态势评估大多存在信息来源单一、评估范围有限、模型不易构建、时空开销大且可信度较低等问题,提出了一种多源异构信息融合量化评估网络安全态势的方法。首先,构建分级朴素贝叶斯分类器,快速高效地融合主机上各多源异构非确定性信息源。然后,利用拉普拉斯原理平滑参数学习,优化分类与推理结果。使用数理统计的方法融合网络上各主机的安全指数,量化评估网络安全态势,对当前网络安全状况有一个宏观整体的认识。最后,通过真实网络环境的实验,验证了所提方法在网络安全态势评估中的可行性和有效性。

关 键 词: 多源异构; 信息融合; 网络安全态势; 量化评估; 朴素贝叶斯
中图分类号: TP311
文献标识码: A **文章编号**: 1001-5965(2016)08-1593-10

随着 Internet 技术的迅速发展,网络规模也逐渐增加且复杂化,所遭受攻击多元化,非确定性问题与日俱增,安全事件大幅度上涨,安全问题变得日益突出与迫切,传统单一安防措施对待安全问题明显感觉无能为力,且各种措施之间的相互关联性欠充分考虑,不能很好地表达网络安全的重要性。网络安全态势评估(Network Security Situation Assessment, NSSA)在此背景下应运而生,逐渐成为下一代网络安防技术的研究重点,主要研究在一定的时空环境下对网络安全相关的要素信息融合、综合分析与理解,把握网络安全状况与预测发展趋势,对评估与预测结果实时决策,将风险与损失降到最低限度[1]。

国内外学者已在网络安全态势评估方法[2-4]与预测方法[5-6]上开展了许多探索性的研究,对以后研究工作具有重要的借鉴作用。文献[7]提出了一种利用神经网络来感知网络安全态势的方法,通过 RBF 神经网络找出有关非线性网络态势值的映射关系,对网络参数进行优化并采用自适应遗传算法感知网络安全态势。文献[8]使用隐马尔可夫模型评估网络安全态势,使观测序列的获取和状态转移矩阵的确立得到了改进,所得风险值能更加合理地量化网络安全态势。文献[9]提出了利用马尔可夫博弈分析的网络安全态势感知方法,通过分析威胁传播的影响,准确、全面地对网络系统评估安全性,并给出相应的加固方案。

上述安全态势评估研究主要使用了证据理论、贝叶斯、神经网络和隐马尔可夫[10-11]等方法,对当前网络的安全态势能较好地评估,给安全策略制定提供了比较可靠的理论依据,对于本文工作具有一定的指导意义,但主要存在的问题在于:评估信息源单一,准确度不高且操作不便,在模型训练与参数获取上也存在瓶颈。网络安全态势评估并不仅仅只对单一数据源评估,还应该将来自

收稿日期: 2015-08-31; **录用日期**: 2015-09-25; **网络出版时间**: 2015-11-09 09:10
网络出版地址: www.cnki.net/kcms/detail/11.2625.V.20151109.0910.001.html
基金项目: 国家自然科学基金(61379057,61309027,61073186);湖南省自然科学基金(2016JJ5034)
* **通讯作者**: Tel.: 13387480797 E-mail: czg@mail.csu.edu.cn

引用格式: 文志诚, 陈志刚, 唐军. 基于信息融合的网络安全态势量化评估方法[J]. 北京航空航天大学学报, 2016, 42(8): 1593-1602. WEN Z C, CHEN Z G, TANG J. Assessing network security situation quantitatively based on information fusion [J]. Journal of Beijing University of Aeronautics and Astronautics, 2016, 42(8): 1593-1602 (in Chinese).

各类多源异构非确定性信息融合，整体宏观分析发生在不同时空和不同层次上的相关联事件。

针对目前安全态势评估中普遍存在信息来源单一、评估范围有限、时空开销较大且可信度较低等问题，本文提出了一种便于处理非确定性信息源的基于朴素贝叶斯(Naive Bayesian, NB)量化评估方法。多方面综合考虑影响网络安全态势各因素，通过构建分级朴素贝叶斯分类器，快速高效地融合各主机多源异构非确定性信息源，再使用数理统计的方法融合网络上各主机的安全指数，逐步量化评估网络安全态势，对当前网络安全态势有一个整体宏观的认识。

1 朴素贝叶斯分类器

朴素贝叶斯分类器是一种非常实用的贝叶斯方法[12]，具有智能统计和学习的能力，已在很多方面得到了成功应用[13]，它是基于贝叶斯理论和各特征条件独立假设的分类方法，利用概率表示各种事件的非确定性。

样本集假设有 m 个属性和 k 个类别，记样本集为 $F=\{F_1,F_2,\cdots,F_m\}$，类别集为 $C=\{c_1,c_2,\cdots,c_k\}$。对于样本集的一个具体实例 $\boldsymbol{X}=(x_1,x_2,\cdots,x_m)$，$x_i\in F_i$，则属于 $Y=c_i$ 的后验概率为

$$P(c_i\mid \boldsymbol{X})=\frac{P(\boldsymbol{X}\mid c_i)\cdot P(c_i)}{\sum_{j=1}^{m}P(\boldsymbol{X}\mid c_j)\cdot P(c_j)}$$

式中：$P(\boldsymbol{X}\mid c_j)$ 为在类别 c_j 出现条件下实例 $\boldsymbol{X}$ 出现的条件概率。

朴素贝叶斯假设属性之间条件相互独立，条件独立是区别于贝叶斯网的主要条件，也是参数容易获得的原因，所以有

$$P(\boldsymbol{X}\mid c_j)=\prod_{i=1}^{m}P(x_i\mid c_j)$$

分类公式如下：

$$C(\boldsymbol{X})=\arg\max_{c_j}\{P(\boldsymbol{X}\mid c_j)\cdot P(c_j)\}=\arg\max\{\prod_{i=1}^{m}P(x_i\mid c_j)\cdot P(c_j)\}$$

在朴素贝叶斯分类器中，判定未知样本 $\boldsymbol{X}$ 的类属，若样本 $\boldsymbol{X}$ 被指派到类 c_i，当且仅当 $P(c_i\mid\boldsymbol{X})>P(c_j\mid\boldsymbol{X})$，$j=1,2,\cdots,m$ 且 $j\neq i$。

朴素贝叶斯分类器具有分类与推理两大功能，本文中只限用于主机的安全指数的推理与分类上，为网络的安全指数服务。在主机二级指数上用作样本分类，而在主机一级指数上用作概率推理，作为推理时，起着与普通贝叶斯网同等的功能，如图 1 所示。

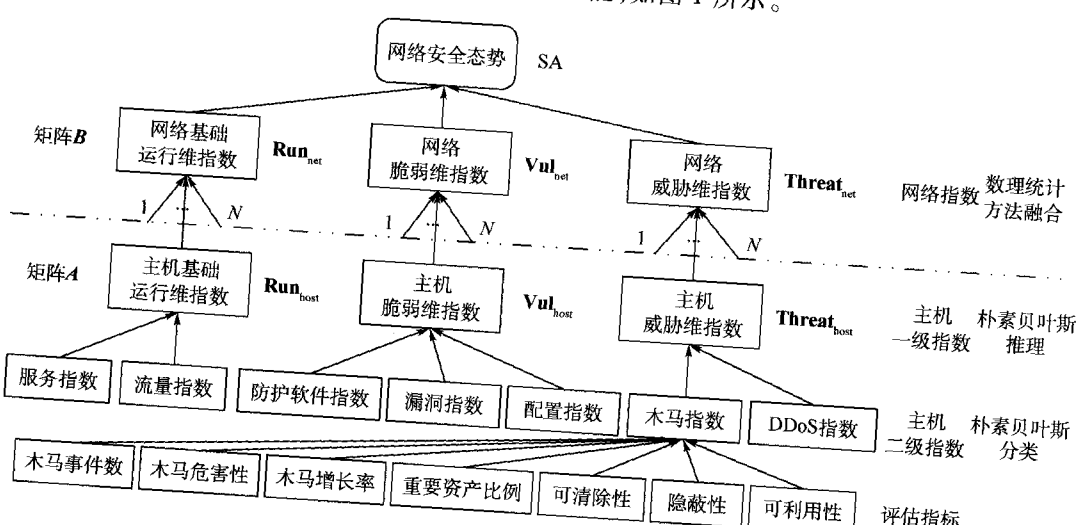

图 1 层次化指标体系

Fig. 1 Hierarchy index system

2 网络安全态势

本文中，网络安全态势可定义为由网络基础运行维指数、网络脆弱维指数和网络威胁维指数三维信息有机融合而成。

2.1 定 义

定义 1 网络安全态势 SA 由网络基础运行维指数($\mathbf{Run}_{net}$)、网络脆弱维指数($\mathbf{Vul}_{net}$)和网络威胁维指数($\mathbf{Threat}_{net}$)三维融合而成，即存在一个融合函数 f，有：$SA=f(\mathbf{Run}_{net},\mathbf{Vul}_{net},\mathbf{Threat}_{net})$。

定义 2 网络的基础运行维指数 $\mathbf{Run}_{net}$ 由网络上所有主机的基础运行维指数融合而成，存在一个融合函数 g，有：$\mathbf{Run}_{net}=g(\mathbf{Run}_{host1},\mathbf{Run}_{host2},\cdots,\mathbf{Run}_{hostN})$，本文中的函数 g 通过求算术平均实现；其他 2 个指数如网络脆弱维指数 $\mathbf{Vul}_{net}$ 与网络威胁维指数 $\mathbf{Threat}_{net}$ 的构成相类似。

定义 3　主机的基础运行维指数 $\mathbf{Run}_{host}$ 由与之运行信息相关的指标融合而成，即存在一个融合函数 h，有：$\mathbf{Run}_{host}=h(x_1,x_2,\cdots,x_n)$，本文中的函数 h 通过朴素贝叶斯分类器来实现；其他 2 个指数如主机脆弱维指数 $\mathbf{Vul}_{host}$ 与主机威胁维指数 $\mathbf{Threat}_{host}$ 的构成相类似。

与指标数不同的指数，本文用一个概率表示，是对本属性的定量描述。所定义的融合函数 f、g 和 h 将在本文中详细阐述。参数 N 表示网络中各节点的主机数目，而 n 表示与安全态势相关的指标数目。注意，网络与主机三维指数都为五等概率矩阵或概率向量，而主机上的二级指数与网络安全态势 SA 为标量，如图 1 所示。

网络拓扑结构中存在着大量的节点，可称之为主机，如网络上的计算机、各类服务器、路由器、防火墙和 IDS 等硬件设施。就众多主机而言，能动态描述主机目前运行情况，由它们的工作性能与服务性能等构成，称为主机安全态势的外在表现特性——基础运行维指数；网络节点主机中存在的可能被威胁利用造成损害的薄弱环节，一旦被威胁成功利用的脆弱性就可能对组件造成损害——脆弱维指数；一般地，主机还包括外部和内部威胁，若成功利用它们的脆弱性会对主机造成损害——威胁维指数。

每个指数有主机指数和网络指数之分，如基础运行维指数，有主机基础运行维指数和网络基础运行维指数 2 种，而网络基础运行维指数又由 N 个主机基础运行维指数按统计方法融合而成，为了有效区别网络指数与主机指数，相应的标识符加以下标 net 和 host 作为区别。

2.2　指标体系

评估数据源主要来自三大类：基于系统配置信息、基于系统运行信息和基于网络流量信息。第一类数据源是指网络设计和配置状况，如网络拓扑结构、服务软件的安装与设置以及系统的漏洞缺陷等；第二类数据源是指网络系统遭受攻击时的系统运行情况，主要来自于系统运行日志库；第三类数据源主要是指网络即时通信各种流量情况，可通过专用软件监测获取。

网络安全态势评估指标在本文中是网络安全态势评估的基础，需要建立一整套符合一定规范和原则的合理、科学评估指标体系，以便全面量化评价当前网络整体安全性能。如图 1 所示，由下而上构成网络安全态势指标体系分级，多源异构信息逐步融合成网络安全态势。

作为安全态势评估数据源，必须选取那些具有代表性、信息量较丰富、可靠度较高、实时性强以及冗余性低的数据，它们主要来自于入侵检测系统 IDS[14] 以及各种检测设备或扫描工具。本文中，每个观测指标可对应于一个随机变量 x_i，具有离散型或连续型 2 种可能观测取值。

2.3　分　级

根据 2006 年发布的《国家突发公共事件总体应急预案》[15]，把网络安全态势等级一般划分为 5 个等级，用 0 ~ 1 小数定量描述，如表 1 所示。

表 1　网络安全等级参照表

Table 1　Network security level reference table

安全指数	安全等级	网络运行情况
0 ~ 0.2	安全(1)	网络运行正常
0.2 ~ 0.4	轻度危险(2)	网络运行受到轻微影响
0.4 ~ 0.75	一般危险(3)	网络运行受到较大影响
0.75 ~ 0.9	中度危险(4)	网络运行受到严重破坏
0.9 ~ 1	高度危险(5)	网络中存在大量的严重攻击行为

网络安全等级参照表是本文的工作基础，也是构建朴素贝叶斯分类器及各类评估结果等级给定的有力参考依据。

2.4　数据源离散化

评估指标可取离散型和连续型 2 种观测值，为了便于原始数据在朴素贝叶斯分类器中的应用，把连续型取值离散化，可取“安全、轻度危险、一般危险、中度危险、高度危险”或“1、2、3、4、5”5 个等级值。在数据源离散化前，首先要获取并计算出相应的数据变化率，取 0 ~ 1 之间的实值，把数据的取值约束在区间[0,1]之间，有

$$\mathrm{Ratio}_{Data}=\frac{\mathrm{Data}_i-\mathrm{Data}_{min}}{\mathrm{Data}_{max}-\mathrm{Data}_{min}} \tag{1}$$

式中：Data_i 为原始数据值；Data_{max} 和 Data_{min} 为数值上下限。在实际应用中，应该去掉数据 Data 一定比例数量的极大值与极小值，以免陷于极端情况。去掉一定量的数据后，当计算出的值大于 1，应按 1 处理；当计算出的值为负数，应按 0 处理。

对于任何一个连续型原始采样数据，可通过式(1)化为 0 ~ 1 之间的值，再对照表 1 可离散化为相应的五等离散取值，是构建朴素贝叶斯分类器的理论基础。针对任何多源异构数据源，先按需求把它映射到相应的实数 Data_i 上，再按此方法离散化，可转化为同构数据，作为信息融合的输入。

2.5　指标遴选

在网络安全态势评估中，有必要遴选出适当的具有一定意义代表性的评估指标。由熵理论可知，当 2 个评估指标之间相互依赖，则互信息量大，反之互信息量就小。由此特性，可以用来遴选评估指标。2 个观测指标 x_i 和 x_j 的互信息可以

定义为:$I(x_i,x_j)=H(x_i)+H(x_j)-H(x_i,x_j)$,$H(x_i,x_j)$为联合熵(joint entropy)。有

$$I(x_i,x_j)=H(x_i)-H(x_i\mid x_j)=H(x_i)+H(x_j)-H(x_i,x_j)=\sum_{x_i}p(x_i)1\mathrm{b}\frac{1}{p(x_i)}+\sum_{x_j}p(x_j)1\mathrm{b}\frac{1}{p(x_j)}+\sum_{x_i,x_j}p(x_i,x_j)1\mathrm{b}\,p(x_i,x_j)=\sum_{x_i,x_j}p(x_i,x_j)1\mathrm{b}\frac{p(x_i,x_j)}{p(x_i)p(x_j)}\tag{2}$$

式中:$p(x_i)$为观测指标x_i出现的概率;$p(x_i,x_j)$为观测指标x_i和x_j同时发生的联合概率。根据第2.4节离散化方法,每个观测指标可以离散化为"1、2、3、4、5"。在某一时间段监测到数据大样本,以出现频率近似它们的概率$p(x)$,代入式(2)中计算它们的互信息量,如果$I(x_i,x_j)$大于一个指定的阈值,则认为很相关,可剔除一个冗余指标,按此方法可遴选出一些具有代表性的评估指标。

3 评估模型

3.1 朴素贝叶斯评估模型

经过互信息方法评估指标的遴选,选出与主机基础运行维指数$\mathbf{Run}_{\mathrm{host}}$、脆弱维指数$\mathbf{Vul}_{\mathrm{host}}$与威胁维指数$\mathbf{Threat}_{\mathrm{host}}$相关的评估指标。根据概率论知识,所有的评估指标构成一个向量$\boldsymbol{X}=(x_1,x_2,\cdots,x_n)$,每个分量$x_i$是一个评估指标,对应于朴素贝叶斯分类器一个具体叶子节点,看成一个随机变量,可取离散型或连续型2种观测值。

本文可逐级构建用于主机信息融合的朴素贝叶斯分类器,从上而下,分而治之,下层的输出作为上一层的输入。从图1可知,就每台主机而言,对于主机基础运行维指数$\mathbf{Run}_{\mathrm{host}}$、主机脆弱维指数$\mathbf{Vul}_{\mathrm{host}}$与主机威胁维指数$\mathbf{Threat}_{\mathrm{host}}$建立3个不同的朴素贝叶斯分类器,图2所示为主机威胁维指数的朴素贝叶斯分类器模型,起推理作用;如果分类器评估指标数太多,可把同一个类型的指标整合在一起,构建一个子朴素贝叶斯分类器,分而治之,图3所示为木马指数子分类器模型,起分类作用。

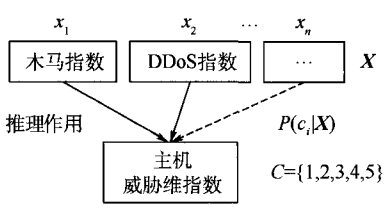

图2 主机威胁维指数朴素贝叶斯分类器模型

Fig. 2 Host threat dimension index naive Bayesian classifier model

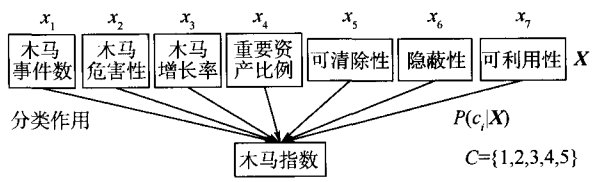

图3 木马指数朴素贝叶斯分类器模型

Fig. 3 Trojan horse index naive Bayesian classifier model

所构建的朴素贝叶斯分类器只限于主机级别,对于网络级别的安全态势,3个维指数信息融合时将采用数理统计的方法,不需要构建朴素贝叶斯分类器,图1有标示。

3.2 分类方法改进

为了更好使朴素贝叶斯分类器用于网络安全态势评估上,本文将改进分类器参数学习,主要是为了避免在分类与推理上陷入极值,采用拉普拉斯原理平滑参数学习。

朴素贝叶斯的参数确定一般使用极大似然估计方法,使用样本出现的频率估计它们的先验概率和条件概率。然而,这样可能会出现极值条件概率的情况,因样本量过少出现条件概率为0而使推理结果也为0,从而影响后验概率的估计。一般采用贝叶斯估计,加上参数λ的拉普拉斯平滑方法解决:

$$P_\lambda(X^{(j)}=x_j\mid Y=c_k)=\frac{\sum_{i=1}^{n'}I(X^{(j)}=x_j,Y=c_k)+\lambda}{\sum_{i=1}^{n'}I(Y=c_k)+s_j\lambda}\tag{3}$$

式中:n'为样本个数;I为指示函数;s_j为x_j出现个数;Y为一个取类属C上的变量。贝叶斯估计等价于在各个取值的频数上加上一个适当的正数λ,避免条件概率为0。

3.3 确定参数

经第3.2节构建主机上的朴素贝叶斯分类器,若要能在实际上应用,必须要获取各节点相应条件概率$P(\boldsymbol{X}|Y)$及其先验概率$P(Y)$,一般通过大样本的参数学习得到。设大样本训练集为

$$T=\{(x_1,y_1),(x_2,y_2),\cdots,(x_n,y_n)\}$$

如图2和图3所示的朴素贝叶斯分类器,需要训练估计节点参数$P(Y=c_k)$与参数$P(X^{(j)}=x_j\mid Y=c_k)$ $(1\leqslant j\leqslant n,1\leqslant k\leqslant m)$的概率值,通过大样本参数学习,从而可对评估指标$\boldsymbol{X}$分配为$Y$类:

$$P(Y=c_k\mid X^{(j)}=x_j)=\frac{P(X^{(j)}=x_j\mid Y=c_k)P(Y=c_k)}{\sum_{k=1}^{m}P(X^{(j)}=x_j\mid Y=c_k)P(Y=c_k)}$$

经过大样本参数学习与拉普拉斯平滑后,根据式(3),有

$$P(Y=c_k)=s_k/s$$

$$P(X^{(j)}=x_j \mid Y=c_k)=\frac{s_{kj}+\lambda}{s_k+\lambda s_{kj}}$$

式中：s_k 为大样本参数学习集中类别为 c_k 的样本数目；s 为大样本参数学习集 T 的总数目；s_{kj} 为大样本参数学习中类别为 c_k 且属性取值 x_j 的样本数，从而可计算或推理出分类条件概率值 $P(Y|\boldsymbol{X})$。

为了防止条件概率出现极值 0 的情况，大样本参数学习的结果与普通朴素贝叶斯分类器不同，添加一个平滑因子 λ。

3.4 信息融合

本文的基本思路是：首先，通过采集网络各主机上评估指标多源异构原始数据 $\boldsymbol{X}$，对连续型指标变量按第 2.4 节及表 1 方法预处理得到相应的离散级别，再通过朴素贝叶斯分类器向上融合函数 h，计算生成各主机的基础运行维指数、脆弱维指数和威胁维指数；然后，根据数理统计上频率近似概率的方法经融合函数 g 把 N 个主机的基础运行维指数融合成网络的基础运行维指数，N 个主机的脆弱维指数融合成网络的脆弱维指数，N 个主机的威胁维指数融合成网络的威胁维指数；最后，通过融合函数 f 加权生成网络安全态势 SA。

4 量化评估方法

4.1 主机安全指数

主机安全指数由主机的三维指数（$\mathbf{Run}_{\text{host}}$，$\mathbf{Vul}_{\text{host}}$，$\mathbf{Threat}_{\text{host}}$）构成，是对主机安全状况在某时刻所处级别 i 的定量描述。通过上述方法，可分级构建融合主机多源异构信息来评估网络安全态势的朴素贝叶斯分类器，当采集到一个样本 $\boldsymbol{X}$，经多级融合成主机的 3 个指数：主机基础运行维指数 $\mathbf{Run}_{\text{host}}$、主机脆弱维指数 $\mathbf{Vul}_{\text{host}}$ 与主机威胁维指数 $\mathbf{Threat}_{\text{host}}$，这就是融合函数 h。

对于主机 3 个指数中的每维，样本 $\boldsymbol{X}$ 经朴素贝叶斯分类器推理，有 5 个概率值。对于主机基础运行维指数而言，把样本 $\boldsymbol{X}$ 分为“安全”的概率为 $P(Y=1|\boldsymbol{X})$，把样本 $\boldsymbol{X}$ 分为“轻度危险”的概率为 $P(Y=2|\boldsymbol{X})$，…，把样本 $\boldsymbol{X}$ 分为“高度危险”的概率为 $P(Y=5|\boldsymbol{X})$。其他 2 个指数一样，为了便于表达，把主机基础运行维指数的概率用 P_1 表示，主机脆弱维指数的概率用 P_2 表示，主机威胁维指数的概率用 P_3 表示，则样本 $\boldsymbol{X}$ 经朴素贝叶斯分类器推理，得到主机的三维指数的概率矩阵 $\boldsymbol{A}$，可简称为主机安全指数矩阵，有

$$\boldsymbol{A}=\begin{bmatrix}\mathbf{Run}_{\text{host}}\\ \mathbf{Vul}_{\text{host}}\\ \mathbf{Threat}_{\text{host}}\end{bmatrix}=\begin{bmatrix}P_1(1,1),P_1(1,2),\cdots,P_1(1,5)\\ P_2(2,1),P_2(2,2),\cdots,P_2(2,5)\\ P_3(3,1),P_3(3,2),\cdots,P_3(3,5)\end{bmatrix}=$$

$$\begin{bmatrix}P_1(Y=1|\boldsymbol{X}),P_1(Y=2|\boldsymbol{X}),\cdots,P_1(Y=5|\boldsymbol{X})\\ P_2(Y=1|\boldsymbol{X}),P_2(Y=2|\boldsymbol{X}),\cdots,P_2(Y=5|\boldsymbol{X})\\ P_3(Y=1|\boldsymbol{X}),P_3(Y=2|\boldsymbol{X}),\cdots,P_3(Y=5|\boldsymbol{X})\end{bmatrix}$$

式中：第 1 行表示主机基础运行维指数 $\mathbf{Run}_{\text{host}}$ 5 个概率；第 2 行表示主机脆弱维指数 $\mathbf{Vul}_{\text{host}}$ 5 个概率；第 3 行表示主机威胁维指数 $\mathbf{Threat}_{\text{host}}$ 5 个概率。网络中有 N 台主机，则每台主机经不同样本 $\boldsymbol{X}$ 推理都可以有一个 3×5 概率矩阵 $\boldsymbol{A}$，则一共有 N 个概率矩阵 $\boldsymbol{A}_i$。

物理意义上，主机安全指数矩阵 $\boldsymbol{A}$ 从 3 个方面定量描述主机的安全状况。符号约定，$\boldsymbol{A}_i$ 表示第 i 台主机概率矩阵，$A_i(j)$ 表示概率矩阵 $\boldsymbol{A}_i$ 的第 j 行，$A_i(j,k)$ 表示概率矩阵 $\boldsymbol{A}_i$ 的第 j 行第 k 列。概率矩阵 $\boldsymbol{B}$ 与 $\boldsymbol{D}$ 同理。

4.2 网络安全指数

网络安全指数由网络的三维指数（$\mathbf{Run}_{\text{net}}$，$\mathbf{Vul}_{\text{net}}$，$\mathbf{Threat}_{\text{net}}$）构成，是对网络安全状况在某时刻所处级别 i 的定量描述。经第 4.1 节计算出样本 $\boldsymbol{X}$ 的主机安全指数，网络中每台主机类似方法可得出其安全指数，概率矩阵为 $\boldsymbol{A}_i$（$i=1,2,\cdots,N$）。方便地，网络安全指数由 N 台主机的安全指数算术平均生成，当然也可以对重要的主机如服务器权重加大，其他主机权重减轻复合而成，这就是融合函数 g。例如对于网络的基础运行维指数 $\mathbf{Run}_{\text{net}}$，由 N 台主机基础运行维指数 $\mathbf{Run}_{\text{host}}$ 取算术平均值，得到网络基础运行维安全指数。三维网络安全指数构成 3×5 概率矩阵 $\boldsymbol{B}$，简称为网络安全指数矩阵，有

$$\boldsymbol{B}=\begin{bmatrix}\mathbf{Run}_{\text{net}}\\ \mathbf{Vul}_{\text{net}}\\ \mathbf{Threat}_{\text{net}}\end{bmatrix}=\begin{bmatrix}B(1,1),B(1,2),\cdots,B(1,5)\\ B(2,1),B(2,2),\cdots,B(2,5)\\ B(3,1),B(3,2),\cdots,B(3,5)\end{bmatrix}=$$

$$\begin{bmatrix}\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{A}_i(1,1),\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{A}_i(1,2),\cdots,\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{A}_i(1,5)\\ \frac{1}{N}\sum_{i=1}^{N}\boldsymbol{A}_i(2,1),\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{A}_i(2,2),\cdots,\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{A}_i(2,5)\\ \frac{1}{N}\sum_{i=1}^{N}\boldsymbol{A}_i(3,1),\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{A}_i(3,2),\cdots,\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{A}_i(3,5)\end{bmatrix}$$

式中：第 1 行第 1 列 $B(1,1)$ 表示目前网络基础运行维指数 $\mathbf{Run}_{\text{net}}$ 为“安全”的概率；第 1 行第 5 列 $B(1,5)$ 表示目前网络基础运行维指数 $\mathbf{Run}_{\text{net}}$ 为“高度危险”的概率；第 2 行表示网络脆弱维指数 $\mathbf{Vul}_{\text{net}}$ 概率；第 3 行表示网络威胁维指数 $\mathbf{Threat}_{\text{net}}$。

物理意义上,是对网络上 N 台主机的安全指数融合成一个网络安全指数,也就是从 N 个主机安全指数矩阵 $\boldsymbol{A}_i$ 融合成一个网络安全指数矩阵 $\boldsymbol{B}$。

4.3 网络安全态势评估

由第4.1节和第4.2节可知,先从最基层的众多评估指标 $\boldsymbol{X}_i$,融合成各主机的基础运行维指数 $\mathbf{Run}_{\text{host}}$、主机的脆弱维指数 $\mathbf{Vul}_{\text{host}}$ 与主机的威胁维指数 $\mathbf{Threat}_{\text{host}}$,生成主机 i 的安全指数矩阵 $\boldsymbol{A}_i$;再由各主机的安全指数矩阵 $\boldsymbol{A}_i$ 生成网络的安全指数矩阵 $\boldsymbol{B}$。

本节中,从网络安全指数中的基础运行维指数、网络的脆弱维指数与网络的威胁维指数矩阵 $\boldsymbol{B}$ 最终生成网络安全态势 SA。分两步进行:第1步,由网络安全指数的三维分别加权融合生成网络安全态势指数;第2步,由网络安全态势指数计算出网络安全态势。

注意术语,网络安全指数可由 3×5 的概率矩阵 $\boldsymbol{B}$ 描述,网络安全态势指数可由 1×5 的矩阵 $\boldsymbol{D}$ 描述,而网络安全态势是一个标量 SA。

4.3.1 生成网络安全态势指数

网络安全态势指数为五等概率矩阵 $\boldsymbol{D}$,分别描述目前网络安全状况所处五等 i 的概率。根据经验与专家推荐,给出网络基础运行维指数 $\mathbf{Run}_{\text{net}}$、网络脆弱维指数 $\mathbf{Vul}_{\text{net}}$ 与网络威胁维指数 $\mathbf{Threat}_{\text{net}}$ 的权值 $\boldsymbol{\omega}=(\omega_1,\omega_2,\omega_3)$ $(\omega_1+\omega_2+\omega_3=1.0)$,使得指数矩阵 $\boldsymbol{B}$ 生成网络安全态势指数矩阵 $\boldsymbol{D}$ 为

$$\boldsymbol{D}=\boldsymbol{\omega}\cdot\boldsymbol{B}=(\omega_1,\omega_2,\omega_3)\cdot\boldsymbol{B}=(\omega_1,\omega_2,\omega_3)\cdot\begin{bmatrix}B(1,1),B(1,2),\cdots,B(1,5)\\B(2,1),B(2,2),\cdots,B(2,5)\\B(3,1),B(3,2),\cdots,B(3,5)\end{bmatrix}=$$

$$[D(1,1),D(1,2),\cdots,D(1,5)]$$

式中:$D(1,k)=\frac{1}{N}\sum_{j=1}^{3}\omega_j\cdot\sum_{i=1}^{N}A_i(j,k),k=1,2,\cdots,5$。

矩阵 $\boldsymbol{D}$ 表示网络安全态势指数,是1行5列的矩阵,每个分量元素 $D(1,i)$ 是0~1的实数,5个数值之和为1,分别对应着安全态势等级1~5等的概率。

物理意义上,对网络安全指数矩阵 $\boldsymbol{B}$ 的三维($\mathbf{Run}_{\text{net}}$, $\mathbf{Vul}_{\text{net}}$, $\mathbf{Threat}_{\text{net}}$)进行加权综合,融合成一维网络安全态势指数矩阵 $\boldsymbol{D}$,这是对网络安全态势目前所处五等 i 的一个定量描述 $D(1,i)$。

4.3.2 生成网络安全态势

由网络安全态势指数矩阵 $\boldsymbol{D}$,与相应的等级 i 经数学综合计算得到网络安全态势 SA,设等级向量 $\boldsymbol{E}=[1,2,3,4,5]^{\mathrm{T}}$ 为转置矩阵,有

$$\mathrm{SA}'=\boldsymbol{D}\cdot\boldsymbol{E}=[D(1,1),D(1,2),D(1,3),D(1,4),D(1,5)]*[1,2,3,4,5]^{\mathrm{T}}=D(1,1)+2D(1,2)+3D(1,3)+4D(1,4)+5D(1,5)=\sum_{i=1}^{5}i\cdot D(1,i)$$

$$\mathrm{SA}=\mathrm{Rounding}(\mathrm{SA}')$$

物理意义上,等级向量 $\boldsymbol{E}$ 中的分量 i,对应的概率为 $D(1,i)$,综合计算后结果 SA′是1~5之间的实数,按照四舍五入就可得到网络安全态势 SA 所在的等级。

5 评估算法

5.1 参数学习算法

输入:评估指标 $\boldsymbol{X}$ 大样本观测数据。

输出:参数化朴素贝叶斯分类器。

(1) $s\leftarrow$评估指标 $\boldsymbol{X}$ 样本总数。

(2) let $s_k=0$, $s_{kj}=0$, $\lambda=1$

(3) for every $\boldsymbol{X}$ and x_j

(4) if $Y=c_k$ then $s_k=s_k+1$

(5) if $X^{(j)}=x_j$ then $s_{kj}=s_{kj}+1$

(6) endfor

(7) for every c_k, let $P(Y=c_k)=s_k/s$

(8) for every x_j and c_k let

(9) $P(X^{(j)}=x_j|Y=c_k)=(s_{kj}+\lambda)/(s_k+\lambda s_{kj})$

(10) endfor

(11) output NB parameter $P(Y)$ and $P(\boldsymbol{X}|Y)$

本参数学习算法的时间复杂度为 $O(skm)$,k 为类别数,m 为样本维数,s 为样本量。

5.2 安全态势生成算法

输入:评估指标 $\boldsymbol{X}$ 一次观察数据。

输出:网络安全态势 SA。

(1) 采集一组评估指标 $\boldsymbol{X}$ 实时观测值,并离散化五等。

(2) for every host i compute matrix $\boldsymbol{A}_i$

(3) from every matrix $\boldsymbol{A}_i$ compute matrix $\boldsymbol{B}$

(4) let $\boldsymbol{\omega}=(\omega_1,\omega_2,\omega_3)$

(5) compute matrix $\boldsymbol{D}=\boldsymbol{\omega}\cdot\boldsymbol{B}$

(6) let matrix $\boldsymbol{E}=[1,2,3,4,5]^{\mathrm{T}}$

(7) compute two matrix product $\boldsymbol{D}\cdot\boldsymbol{E}$

(8) SA = rounding($\boldsymbol{D}\cdot\boldsymbol{E}$)

(9) output SA

本网络安全态势生成算法时间复杂度为 $O(N)$。

6 仿真实验

本文搭建了一个网络实验环境,验证本文所提

出评估方法的合理性与正确性。在该环境下进行安全态势量化评估实验。普通用户 User 和攻击者 Attacker 可通过 Internet 访问该网络上各主机。

定期采集入侵检测系统 IDS 攻击信息、主机 Nessus 中采集漏洞扫描信息、Snort 采集日志报警信息和路由器 Netflow 采集网络流量信息,作为本次仿真实验的多源异构原始数据源。

6.1　数据采样

1) 原始数据

不断地人为对上述网络环境发起各类攻击,为了清晰绘图,在一个 10 s 时间内,动态采集了 200 个样本,以 4 个评估指标(CPU 利用率、内存占用大小、子网带宽使用率和子网平均数据流)采样为例,原始数据如图 4 所示,主机受到攻击时样本实时数据会产生一定的波动,数据之间具有相互关联性。

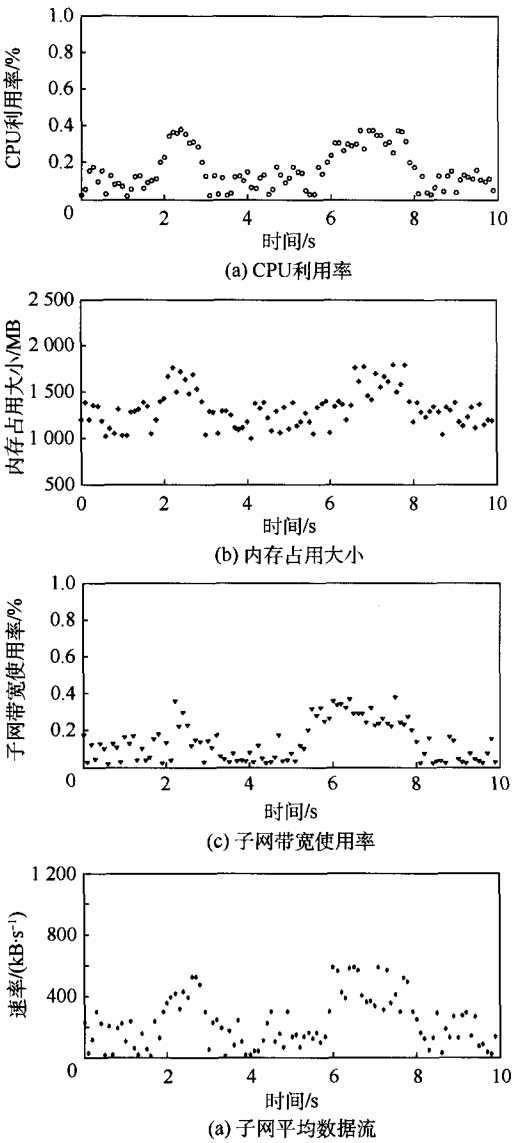

图 4　原始数据采样图

Fig. 4　Original data sampling diagram

2) 离散化后数据

对于图 4 所示连续型原始采样数据,可应用式(1)归一化处理为 0 ~ 1 之间的实数值,再对照表 1 可取相应的五等离散取值。为了便于表达,把图 4 中的数据按中间值处理后平移到相应的位置,而不是直接取离散值,否则变成一根折线,表达不了数据之间的差异性,如图 5 所示。离散化平移后,数据在相应的离散值附近上下小幅度波动。在应用时,在等级 i 附近上下波动的数据就取离散化值 i,方便且易于操作。

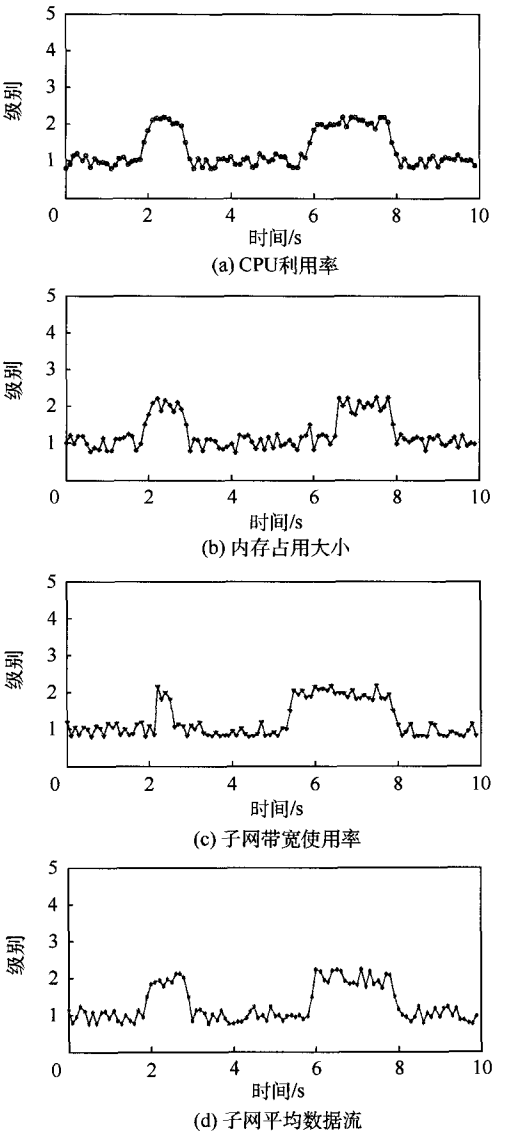

图 5　离散化数据采样图

Fig. 5　Discrete data sampling diagram

6.2　参数学习

为了让所建立的朴素贝叶斯分类器能正常使用,必须对其进行参数学习。本实验随机采集所需样本量 5 000 个,对普通贝叶斯网、朴素贝叶斯网与拉普拉斯平滑后的朴素贝叶斯网参数学习算法作了相应的比较,如图 6 所示。

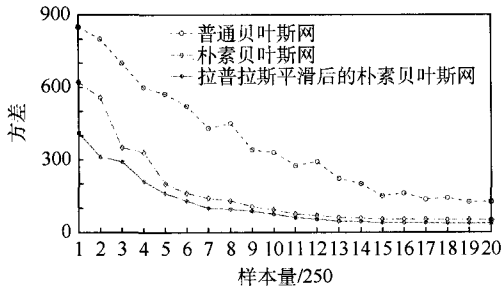

图6 朴素贝叶斯参数学习对比图

Fig. 6 Naive Bayesian parameter learning comparison chart

根据朴素贝叶斯分类器公式和 $P_{jk}=P(X^{(j)}=x_j|Y=c_k)$，横坐标代表学习样本量，纵坐标表示经过一次样本学习后，评估指标先验概率与条件概率与上一次学习结束时的距离，即根方差 $\sigma=\sqrt{[\sum_{k=1}^{m}(P_k-P'_k)^2+\sum_{j=1}^{n}\sum_{k=1}^{m}(P_{jk}-P'_{jk})^2]/(n-1)}$，对于普通贝叶斯网，可用其联合条件概率 $\theta_{ijk}=(m_{ijk}+\alpha_{ijk})/\sum_{k=1}^{r_i}(m_{ijk}+\alpha_{ijk})$ 代替 P_{jk}，为了方便计算，把所有的概率放大10倍。从图6中可知，普通贝叶斯网的学习速度较慢，且具有较大的波动性，需要样本量多；而未经平滑的朴素贝叶斯网，小样本时，不如平滑后的朴素贝叶斯网明显，但样本量趋于3 000时，具有同等效果。因此，拉普拉斯平滑后的朴素贝叶斯在小样本时具有优越性。

经过主机5 000个大样本数据朴素贝叶斯参数学习，获得后验概率参数 $P(X=x^{(j)}|Y=i)$ 近似值，以主机基础运行维指数为例，如表2所示朴素贝叶斯分类器的部分参数表。

表2 朴素贝叶斯分类器的参数表

Table 2 Naive Bayesian classifier parameter table %

基础运行维	$P(X=x^{(j)}\|Y=i)$				
	CPU利用率	内存占用率	子网平均数据流	子网带宽使用率	子网流量变化率
$Y=1$	76.23	73.27	73.17	69.46	71.33
$Y=2$	10.18	18.77	11.22	9.77	11.42
$Y=3$	4.32	9.62	10.72	4.32	3.12
$Y=4$	0.58	7.30	3.62	0.98	1.51
$Y=5$	0.26	1.08	3.62	23.71	0.40

6.3 在线评估

1）主机安全指数图

一个样本 $\boldsymbol{X}$ 经朴素贝叶斯分类器推理，对于主机3个指数（$\mathbf{Run}_{host}$，$\mathbf{Vul}_{host}$，$\mathbf{Threat}_{host}$）中的每维，各有5个概率值。

图7给出了当主机一次朴素贝叶斯推理后，主机每维取相应的概率值的情况。第1竖列表示该主机基础运行维指数 $\mathbf{Run}_{host}$ 5个等级概率，第2竖列表示该主机脆弱维指数 $\mathbf{Vul}_{host}$ 5个等级概率，第3竖列表示该主机威胁维指数 $\mathbf{Threat}_{host}$ 5个等级概率。从结论可以看出各维取第1等级的概率 $P(Y=1|\boldsymbol{X})$ 都要高，且每维5个等级概率之和为1，说明各主机“安全”的概率要大得多，“高度危险”的概率要少得多。

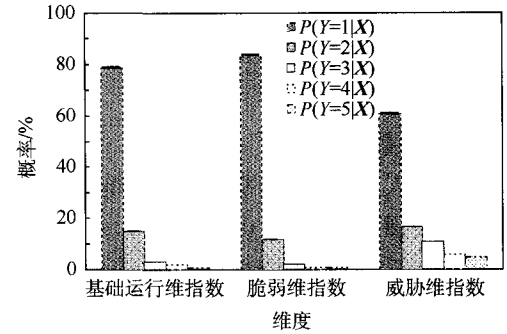

图7 主机安全指数图

Fig. 7 Host security index diagram

2）网络安全态势指数图

网络安全态势指数为5个等级概率矩阵 $\boldsymbol{D}$，分别描述目前网络安全状况所处5个等级 i 的概率。本文根据经验与专家推荐，给出网络基础运行维指数 $\mathbf{Run}_{net}$、网络脆弱维指数 $\mathbf{Vul}_{net}$ 与网络威胁维指数 $\mathbf{Threat}_{net}$ 的权值 $\boldsymbol{\omega}=(\omega_1,\omega_2,\omega_3)=(0.50,0.25,0.25)$，体现网络的基础运行维指数的重要性，从网络安全指数矩阵 $\boldsymbol{B}$ 生成网络安全态势指数矩阵 $\boldsymbol{D}$，矩阵 $\boldsymbol{D}$ 表示网络安全态势指数，是1行5列的矩阵，每个分量元素 $D(1,i)$ 是0～1的实数，5个数值之和为1。图8描绘了15个时刻的网络安全态势指数。网络上受到攻击时，相关时刻指数会产生波动。图8中每个时刻上的5个值（竖排）之和为1，分别对应着安全态势等级1～5等的概率。

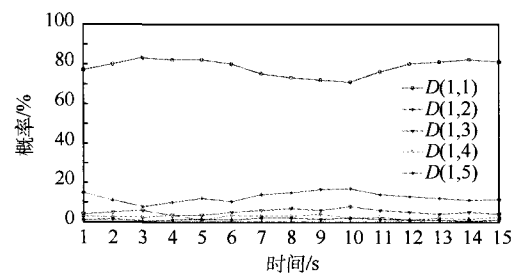

图8 网络安全态势指数图

Fig. 8 Network security situation index diagram

3）网络安全态势评估图

在网络安全态势评估前，需动态采集各个评估指标值，表3表示某个时刻 t 网络上一主机所有评估指标的离散取值。网络上有多少个主机，在这个时刻 t 时就有多少个类似参数表3，共同融合成网络的安全态势SA。

表3表示此台主机正受到网络攻击，因为威胁维指数5个等值基本上处于2、3等级，经融合

可得网络安全态势指数矩阵为 $\boldsymbol{D}=[0.71,0.18,0.07,0.03,0.01]$，与相应等级向量$[1,2,3,4,5]^{\mathrm{T}}$之积，得 $SA'=1.45$，取上整得到网络安全态势为第 1 等，近似第 2 等。

本实验动态采集了 10 次样本，网络上主机当达到一定数量受到攻击时，整个网络安全态势会产生明显波动。实验中，对普通贝叶斯评估方法、朴素贝叶斯评估方法以及本文中的拉普拉斯平滑后的朴素贝叶斯评估方法作了比较，网络安全态势评估对比如图 9 所示。根据网络安全态势等级划分，若对 SA′的值四舍五入，拉普拉斯平滑后的朴素贝叶斯评估方法具有同等的效果，能及时反映网络当前的安全状况。

表 3　主机评估指标所取离散值

Table 3　Host assessment indicators taking discrete values

$\mathbf{Run}_{\mathrm{host}}$ 可观测指标	五等值	$\mathbf{Vul}_{\mathrm{host}}$ 可观测指标	五等值	$\mathbf{Threat}_{\mathrm{host}}$ 可观测指标	五等值
CPU 利用率	1	网络漏洞数目及等级	1	蠕虫攻击	2
内存使用情况	1	系统配置	1	DDoS	2
子网平均无故障时间	2	防护软件是否安装	1	子网带宽使用率	2
子网流量变化率	1	关键设备漏洞数目及等级	1	木马和普通病毒数目	3
子网内存活关键设备数目	1	子网内安全设备数目	1	子网流入量增长率	2
子网内不同大小数据包的分布	1	子网内各关键设备开放端口	2	子网数据流入量	3
子网数据流总量	1			报警数目	2
子网内关键设备平均存活时间	2				

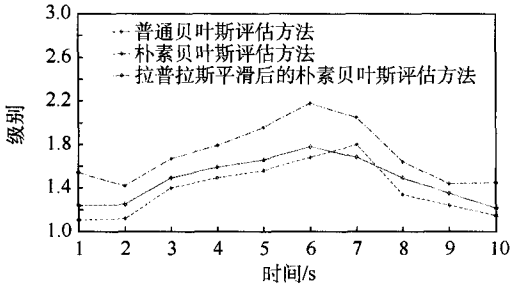

图 9　网络安全态势量化评估对比

Fig. 9　Comparison of quantitative assessment of network security situation

从上述仿真实验可以看出，本文算法具有高效性与准确性。图 6 表明当样本量达 3 000 左右基本上收敛在某个平稳状态，所需样本量比贝叶斯方法要少，参数学习效率高；图 8 表明安全态势受多个评估指标与主机影响，之间具有相互关联性，受攻击时产生波动，可详细反映当时网络安全所处各状态的情况；图 9 表明本文所采取的拉普拉斯平滑后的朴素贝叶斯评估方法具有稳定性，优于其他两个方法。

7　结　论

针对目前安全态势评估中的信息来源单一、评估范围有限、时空开销较大且可信度低等问题，本文提出了基于多源异构信息融合的量化评估网络安全态势方法。

1）在评估过程中，分层处理，综合考虑了影响网络安全态势的各方面因素。

2）首先通过构建分级朴素贝叶斯分类器融合主机的多源异构非确定性信息源，改进了朴素贝叶斯分类器参数学习，给予了拉普拉斯平滑方法，优化分类与推理结果；然后使用数理统计的方法融合网络上各主机的安全指数，量化评估网络安全态势，具有可信性；最后通过真实网络环境的实验，验证了所提方法在网络安全态势评估中的可行性和有效性。

3）朴素贝叶斯分类器在参数学习、分类与推理，具有快速高效性。

参考文献（References）

[1] BASS T. Intrusion detection systems and multisensory data fusion[J]. Communications of the ACM, 2000, 43(4): 99-105.

[2] JANSEN A, MELCHERS K G, LIEVENS F, et al. Situation assessment as an ignored factor in the behavioral consistency paradigm underlying the validity of personnel selection procedures [J]. Journal of Applied Psychology, 2013, 98(2): 326-341.

[3] SHARMA C, KATE V. ICARFAD: A novel framework for improved network security situation awareness[J]. International Journal of Computer Applications, 2014, 87(19): 26-31.

[4] BECHTSOUDIS A, SKLAVOS N. Aiming at higher network security through extensive penetration tests[J]. IEEE Latin America Transactions, 2012, 10(3): 1752-1756.

[5] 黄同庆，庄毅. 一种实时网络安全态势预测方法[J]. 小型微型计算机系统，2014，35(2)：303-306.
HUANG T Q, ZHUANG Y. An approach to real-time network security situation prediction[J]. Journal of Chinese Computer Systems, 2014, 35(2): 303-306(in Chinese).

[6] 刘玉岭，冯登国，连一峰，等. 基于时空维度分析的网络安全态势预测方法[J]. 计算机研究与发展，2014，51(8)：1681-1694.
LIU Y L, FENG D G, LIAN Y F, et al. Network situation prediction method based on spatial-time dimension analysis[J]. Journal of Computer Research and Development, 2014, 51(8): 1681-1694(in Chinese).

[7] 谢丽霞，王亚超，于巾博. 基于神经网络的网络安全态势感知[J]. 清华大学学报(自然科学版)，2013，53(12)：1750-1760.
XIE L X, WANG Y C, YU J B. Network security situation

awareness based on neural networks[J]. Journal of Tsinghua University(Science and Technology),2013,53(12):1750-1760(in Chinese).

[8] 席荣荣,云晓春,张永铮,等.一种改进的网络安全态势量化评估方法[J].计算机学报,2015,38(4):749-758.
XI R R,YUN X C,ZHANG Y Z,et al. An improved quantitative evaluation method for network security[J]. Chinese Journal of Computers,2015,38(4):749-758(in Chinese).

[9] 张勇,谭小彬,崔孝林,等.基于Markov博弈模型的网络安全态势感知方法[J].软件学报,2011,22(3):495-508.
ZHANG Y,TAN X B,CUI X L,et al. Network security situation awareness approach based on Markov game model[J]. Journal of Software,2011,22(3):495-508(in Chinese).

[10] KHREICH W,GRANGER E,MIRI A,et al. Adaptive ROC-based ensembles of HMMs applied to anomaly detection[J]. Pattern Recognition,2012,45(1):208-230.

[11] SENDI A S,DAGENAIS M,JABBARIFAR M,et al. Real time intrusion prediction based on optimized alerts with hidden Markov model[J]. Journal of Networks,2012,7(2):311-321.

[12] LAMINE F B,KALTI K,MAHJOUB M A. The threshold EM algorithm for parameter learning in Bayesian network with incomplete data[J]. International Journal of Advanced Computer Science and Applications,2011,2(7):86-91.

[13] 张轮,杨文臣,刘拓,等.基于朴素贝叶斯分类的高速公路交通事件检测[J].同济大学学报(自然科学版),2014,42(4):558-563.
ZHANG L,YANG W C,LIU T,et al. A naive Bayesian classifier-based algorithm for freeway traffic incident detection[J]. Journal of Tongji University(Natural Science),2014,42(4):558-563(in Chinese).

[14] PANDA M,ABRAHAM A,PATRA M R. A hybrid intelligent approach for network intrusion detection[C]// International Conference on Communication Technology and System Design 2011. Amsterdam:Elsevier,2012,30:1-9.

[15] 国务院.国家突发公共事件总体应急预案[M].北京:中国法制出版社,2006:1-2.
The State Council of the People's Republic of China. A overall emergency plans of national public event[M]. Beijing:China Legal Press,2006:1-2(in Chinese).

作者简介:

文志诚 男,博士,教授,硕士生导师。主要研究方向:网络安全与软件工程。

E-mail: zcwen@mail.shu.edu.cn

陈志刚 男,博士,教授,博士生导师。主要研究方向:分布式处理。

Tel.: 13387480797

E-mail: czg@mail.csu.edu.cn

Assessing network security situation quantitatively based on information fusion

WEN Zhicheng[1,2], CHEN Zhigang[2,*], TANG Jun[3]

(1. School of Computer and Communication, Hunan University of Technology, Zhuzhou 412007, China;
2. School of Information Science and Engineering, Central South University, Changsha 410083, China;
3. CRRC Zhuzhou Institute Co., Ltd., Zhuzhou 412001, China)

Abstract: Concerning the problem that current network security situation assessment has the characteristics of single information source, limited assessment scope, not easy to build model, high time and space complexity and not high credibility, a new method of network security situation assessment is proposed based on multi-source and heterogeneous information fusion. A hierarchical naive Bayesian classifier was constructed based on the theory of Laplace's principle for smoothing parameter learning in order to optimize the result of classification and inference. The quantization for the network security situation was assessed using the method of mathematical statistics, which can generate every host security index through information fusion. The current network security situation should be understood overall and macroscopically. The feasibility and effectiveness of the proposed method for network security situation assessment are verified by the experiments in real network environment.

Key words: multi-source and heterogeneous; information fusion; network security situation; quantitative assessment; naive Bayesian

Received: 2015-08-31; **Accepted**: 2015-09-25; **Published online**: 2015-11-09 09:10

URL: www.cnki.net/kcms/detail/11.2625.V.20151109.0910.001.html

Foundation items: National Natural Science Foundation of China (61379057,61309027,61073186); Natural Science Foundation of Hunan Province(2016JJ5034)

* **Corresponding author**. Tel.: 13387480797 E-mail: czg@mail.csu.edu.cn