文章编号:1006-2467(2016)09-1407-08+1421 DOI: 10.16183/j.cnki.jsjtu.2016.09.011

基于聚类分析的网络安全态势评估方法

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, 湖南 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, 湖南 株洲 412001)

摘 要: 针对现有网络安全态势评估方法具有信息来源单一、评估范围有限、模型训练与参数获取不易和时空开销较大等问题,提出了一种基于聚类分析评估网络安全态势的方法.首先构建主机上若干二级广义立方体,快速高效聚类融合主机多源异构非确定性信息源,生成主机的安全态势,并利用Dirichlet先验分布平滑后验概率,优化聚类分析结果;然后构建网络上的广义立方体,聚类融合网络上各主机的安全态势,逐步量化评估并生成网络当前安全态势.通过真实网络环境的实验,验证了所提出的方法在网络安全态势评估中的可行性和有效性.

关键词: 聚类分析; 网络安全态势; 评估方法; 信息融合; 广义立方体
中图分类号: TP 311 **文献标志码**: A

Network Security Assessment Method Based on Cluster Analysis

WEN Zhicheng[1,2], *CHEN Zhigang*[2], *TANG Jun*[3]
(1. School of Computer and Communication, Hunan University of Technology, Zhuzhou 412007, Hunan, China; 2. School of Information Science and Engineering, Central South University, Changsha 410083, China;
3. CRRC Zhuzhou Institute Co. Ltd., Zhuzhou 412001, Hunan, China)

Abstract: A new network security assessment method was proposed based on cluster analysis. First, a number of second-level general cubes were constructed, into which multi-source and heterogeneous information were gathered, using the principle of Dirichlet's prior distribution optimizing the result of posterior probability. Then, a network general cube was built to fuse the security of each host, assessing the security of network quantitatively step by step. The feasibility and effectiveness of the proposed method in the network security assessment were verified by the experiments of real network environment.

Key words: cluster analysis; network security; assessment method; information fusion; general cubes

BASS[1]于1999年提出了将态势评估与网络安全技术[2]相结合,从而将态势评估引入到网络安全领域,并指出下一代网络入侵检测系统应该综合融合大量的多源异构分布式网络数据,实现对网络态势评估.网络安全态势评估(NSSA)是指在获取海量网络安全数据信息的基础上,通过分析理解它们

收稿日期:2015-09-18
基金项目:国家自然科学基金(61379057;61309027),湖南省自然科学基金(2016JJ5034)资助项目
作者简介:文志诚(1972-),男,湖南省东安县人,博士后,教授,研究方向为网络安全与可信软件.
电话(Tel.): 0731-22183345; E-mail:zcwen@mail.shu.edu.cn.

之间的关联性，进行信息融合，获取宏观的网络安全状况，为网管人员决策与分析提供可靠依据，将风险与损失降到最低限度.

目前已在网络安全态势评估方法[3-4]与预测方法[5-6]上开展了许多可贵的研究工作，主要采用 D-S 理论[7]、粗糙集理论、模式挖掘[8]、贝叶斯[9]、神经网络、模糊逻辑[10]、熵理论和专家系统等方面，并取得了一定的成果，对以后研究工作具有指导作用. 网络安全态势评估的研究对于提高网络的监控能力、应急响应能力和预防干预能力都具有重要的意义.

文献[11]中采用改进的证据理论对网络设备服务及漏洞信息进行融合，计算了全局整体的网络态势值，利用时间序列分析实现了态势趋势预测. 文献[12]中提出了一种基于神经网络的网络安全态势感知方法，利用 RBF 神经网络找出网络态势值的非线性映射关系，采用自适应遗传算法对网络参数进行优化并感知网络安全态势. 文献[13]中在基于隐马尔可夫模型的网络安全态势评估中，改进了观测序列的获取和状态转移矩阵的确立，使得改进算法生成的风险值对网络安全态势的量化更加合理. 文献[14]中提出一种基于 Markov 博弈分析的网络安全态势感知方法，为了分析威胁传播对网络系统的影响，准确、全面地评估系统的安全性，并给出相应的加固方案. 上述这些方法对于本文工作具有重要借鉴意义.

网络安全态势评估并不只对单一数据源评估，应将来自各类异构非确定性信息源进行融合，对发生在不同时空和不同层次上相关联事件进行整体宏观分析. 针对目前网络安全态势评估方法中的大多信息来源单一、评估范围有限、模型训练与参数获取不易、时空开销较大且可信度不高等问题，本文综合考虑了影响网络安全态势的各方面因素，通过构建广义立方体聚类分析，快速高效地融合主机多源异构非确定性信息源，直观方便，逐步从下而上量化评估并生成网络当前安全态势，对当前网络安全态势有一个整体宏观的认识.

1 网络安全态势

1.1 网络安全态势定义

定义 1 网络安全态势 SA，由网络安全指数 ***ISA*** 经计算得到一个综合值，取 1～5 等级. 而 ***ISA*** 是由网络上 N 台主机安全态势 3 个维度融合而成，即由 N 台主机的基础运行维 Run、主机脆弱维 Vul 和主机威胁维 $Threat$ 经广义立方体融合而成.

定义 2 Run 由主机基础运行维指数 ***IRun*** 经计算得到一个综合值，取 1～5 等级；其他 2 个 Vul 与 $Threat$ 类似构成. Run、Vul 与 $Threat$ 组成主机安全态势的 3 维向量 $\boldsymbol{HSA}=(Run, Vul, Threat)$. 本文不打算计算主机安全态势，只标明它由这 3 维组成的向量.

定义 3 ***IRun*** 由与运行信息相关的指标融合而成，即存在融合函数 h，有：$\boldsymbol{IRun}=h(x_1, x_2, \cdots, x_n)$，函数 h 通过广义立方体实现；其他 2 维指数如主机脆弱维指数 ***IVul*** 与主机威胁维指数 ***IThreat*** 构成类似.

N 表示网络中的主机数，n 表示与安全态势相关的指标数. ***ISA*** 与主机 3 个一级指数及 ***HSA*** 均取 1～5 等的概率矩阵或向量；而 SA、主机二级指数、主机安全态势向量 3 个维度(Run, Vul, $Threat$)中每个维度均为标量，如图 1 所示. 所谓指数，与指标数不同，是对本属性定量描述，本文用概率来描述之，前面加字母 I 以示区别.

网络系统结构中存在着大量的计算机、服务器、路由器、防火墙和入侵检测系统(IDS)等称为主机. 就主机而言，***IRun*** 是指动态描述主机目前运行情况，由工作性能和服务性能等构成，称为主机安全态势的外在特征；***IVul*** 是指主机中存在的可能被威胁利用造成损害的薄弱环节，脆弱性一旦被威胁成功利用就可能对主机造成损害；***IThreat*** 是指主机包括外部和内部威胁，若成功利用脆弱性会对主机造成损害.

本文采用广义立方体作为聚类分析工具，在线量化评估网络安全态势.

1.2 网络安全指标体系

评估数据源一般来自：系统配置信息、系统运行信息以及网络的流量信息等. 配置信息是指网络设计和配置状况，如拓扑信息、防病毒软件安装与否、服务软件的安装与设置情况、系统的漏洞缺陷数量等；运行信息是指目前网络系统所受到的攻击与运行情况以及各主机当前运行状况等，主要来自于各类运行日志库.

网络安全态势评估指标是评估的基础，需要建立一套合理、科学的评估指标体系，能全面评价当前网络整体安全性能. 网络安全态势指标体系及融合方法见图 1.

1.3 网络安全等级

根据国家突发公共事件总体应急预案文稿[15]，并结合网络威胁与漏洞等要素特点，把网络安全态势等级划分为 5 个等级，用 0～1 区间的小数定量描述，如表 1 所示.

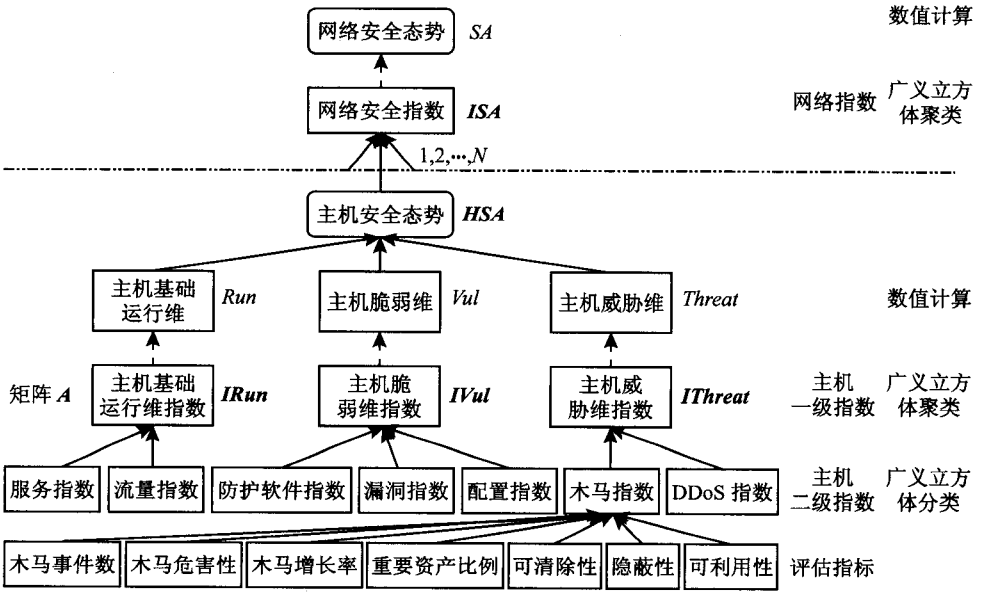


图 1　网络安全态势指标体系

Fig. 1　Index system of network security

表 1　网络安全等级参照表

Tab. 1　Reference table of network security

安全指数	安全等级	网络运行情况
0～0.20	安全(1)	正常
0.20～0.40	轻度危险(2)	受到轻微影响
0.40～0.75	一般危险(3)	受到较大影响
0.75～0.90	中度危险(4)	受到严重破坏
0.90～1.00	高度危险(5)	存在大量的严重的攻击行为

网络安全等级参照表是本文的工作基础，也是构建广义立方体及各类评估结果等级给定的有力参考依据.

2　理论基础

2.1　数据源离散化

评估指标可取离散型和连续型 2 种值，为了便于原始数据在广义立方体中的应用，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”或用“1、2、3、4、5”表示 5 个等级值. 若评估指标本来就是 5 等离散型取值，则不需离散化，直接应用；若多于 5 等离散化取值，可参照连续型离散化为 5 等值. 简捷地，本文只对取连续型评估指标做出阐述.

把数据的取值约束在区间[0, 1]之间，有：

$$Ratio_{Data} = \frac{Data_i - Data_{\min}}{Data_{\max} - Data_{\min}} \qquad (1)$$

式中：$Data_i$ 表示原始数据值；$Data_{\max}$ 与 $Data_{\min}$ 表示数值上、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 之间值，再对照表 1 可离散化为相应的 5 等离散取值，是构建广义立方体的理论基础.

2.2　广义立方体

一般地，立方体由 3 个维度构成，而广义立方体则由 m 维构成，可在立方体格中聚集数据，聚集的数据向上融合成某方面特征，通过该聚合模型能从不同的角度考察数据集的特性. 在网络安全态势评估中，各评估指标构成广义立方体不同的维度，5 个离散值构成不同的抽象层次. 分析各个维度、层次上的数据，从不同角度反映当前网络安全状况.

定义 4　给定一个数据集 D、一组 m 个维度 $C=\{C_1, C_2, \cdots, C_m \mid m>0\}$ 和各个维度相应的层次集合 $H=\{H_1, H_2, \cdots, H_m\}$（$H_i$ 表示对应于维度 C_i 层次数目，H_i^j 表示维度 C_i 的第 j 个层次），在层次集合 $H=\{H_1, H_2, \cdots, H_m\}$ 上各取 1 个层次进行组合，所有这样的组合 $\{H_1^{j1}, H_2^{j2}, \cdots, H_m^{jm}\}$ 可在数据集 D 上执行分组聚合操作，则构成一个广义立方体 (C, H).

定义 5　一个广义立方体单元格为二元组 (C_H, M)，其中：① $C_H=\{H_1^{j1}, H_2^{j2}, \cdots, H_m^{jm}\}$，$H_i^{ji} \in H_i$ 是在维度 C_i 上一个具体层次 j_i；② M 是对 C_H 数据集 D 经离散化后进行聚集操作，落在本单元格中的数量. 广义立方体由广义立方体格组成：$(C, H)=\{(C_H, M)\}$.

3 主机级评估

SA 评估分 2 步进行:① 对主机级相关的指数或指标进行评估,作为网络级评估所需输入数据,是本文关键与重点;② 对网络级评估,将对主机级输出的数据作进一步信息融合,最后得出 *SA*(见图 1).

3.1 主机二级指数评估

IRun、***IVul*** 与 ***IThreat*** 中的每维所包含的指标数较多,若每维一级指数直接构建一个广义立方体,势必具有很高维度,造成空间爆炸等问题. 因此,有必要对类似的指标组合在一起,构成一个子广义立方体,分而治之. 于是主机的一级指数可由二级指数复合构成,而每个主机二级指数单独构建一个广义立方体.

如图 1 所示,一级指数"主机威胁维指数"包含一个二级指数"木马指数",而二级指数"木马指数"由木马事件数、木马危害性、木马增长率、重要资产比例、可清除性、隐蔽性和可利用性 7 个指标组成. 因此,构建"木马指数"的子广义立方体时具有 7 个维度,每个维度都具有 5 个层次,一个立方体格可表示为 $C(L_1, L_2, L_3, L_4, L_5, L_6, L_7)$,相应的立方格中样本数为 $M(L_1, L_2, L_3, L_4, L_5, L_6, L_7)$,其中 $1 \leqslant L_1, L_2, L_3, L_4, L_5, L_6, L_7 \leqslant 5$.

由于 *SA* 评估具有时段性,在一定时间段内所采集的数据才有效. 对于一组 m 个评估指标$(x_1, x_2, \cdots, x_m)$,在同一时间段内采集到 n 个 m 元组原始数据 D(整体上看成 n 个数据,n 足够大得可以聚类),每个分量 x_i 按照上述方法离散化后,则可构成对应的 n 个 m 元组离散化数据. 若主机二级指数有 k 个广义立方体,则这 m 元组分为 k 类,每类数据对应于一个广义立方体,给予形式化定义:

$$|X_1|+|X_2|+\cdots+|X_k|=m$$

$$X_i \cap X_j = \varphi,\ 1 \leqslant i,j \leqslant k,\ i \neq j$$

$$X_j = \{x_{j1}, x_{j2}, \cdots, x_{js}\},\ 1 \leqslant j_s \leqslant m$$

$$x_{js} \in \{x_1, x_2, \cdots, x_m\}$$

其中:X_i 为集合,彼此互不相交;$|X_i|$ 表示集合 X_i 中元素个数. 如上述木马指数广义立方体一共 7 个维度,假设集合 X_1 是对应的指标集,具有 7 个元素,则 X_1 指标集元素可聚集在此广义立方体中,而这一组样本其他指标集 $X_2 \sim X_k$ 共 $m-7$ 个指标分别聚集到其他 $k-1$ 个二级指数广义立方体中. 注意,集合 X_i 中的元素,作为广义立方体中的坐标(维度),而聚在其中只是 1 个点. 一组(个)m 个评估指标,聚集在 k 二级指数广义立方体中,只有 k 个点.

此处广义立方体起着分类作用,把一组样本数据分为 k 个"安全(1)、轻度危险(2)、一般危险(3)、中度危险(4)、高度危险(5)"等级. 例如 7 个指标数据聚集在"木马指数"广义立方体中,输出为 5 等中的 1 个. 因此,一个 m 元组数据输入,输出为 k 个等级点.

3.2 主机一级指数评估

3.2.1 *数据获取* 经上述 n 个 m 元组原始监测数据经主机二级指数广义立方体聚集分类后,得到 n 个 k 元组 5 等数据,分为 3 大类聚集在 ***IRun***、***IVul*** 与 ***IThreat*** 3 个广义立方体中. n 个 k 元组 5 等数据落在每个一级指数广义立方体中都有 n 个点,以 IRun 为例,计算这 n 个点取等级"1,2,3,4,5"的概率 $P(Y=i|D)$,形成一个 1×5 的向量,3 个向量指数构成一个 3×5 概率矩阵 $\boldsymbol{A}$.

为了便于表达,把 ***IRun*** 的概率用 P_1 表示,***IVul*** 的概率用 P_2 表示,***IThreat*** 的概率用 P_3 表示,则样本 D 在 3 个一级指数广义立方体中聚类分析,得到主机的三维指数的概率矩阵:

$$\boldsymbol{A}=\begin{bmatrix}\boldsymbol{IRun}\\ \boldsymbol{IVul}\\ \boldsymbol{IThreat}\end{bmatrix}=\begin{bmatrix}P(1,1),P(1,2),\cdots,P(1,5)\\ P(2,1),P(2,2),\cdots,P(2,5)\\ P(3,1),P(3,2),\cdots,P(3,5)\end{bmatrix}=\begin{bmatrix}P_1(Y=1\mid D),P_1(Y=2\mid D),\cdots,P_1(Y=5\mid D)\\ P_2(Y=1\mid D),P_2(Y=2\mid D),\cdots,P_2(Y=5\mid D)\\ P_3(Y=1\mid D),P_3(Y=2\mid D),\cdots,P_3(Y=5\mid D)\end{bmatrix}$$

第 1 行表示 ***IRun*** 5 个概率,第 2 行表示 ***IVul*** 5 个概率,第 3 行表示 ***IThreat*** 5 个概率. 网络中有 N 台主机,则每台主机经不同样本 D 聚类都可以有一个概率矩阵 $\boldsymbol{A}$,则一共有 N 个概率矩阵 $\boldsymbol{A}$.

符号约定,$\boldsymbol{A}_i$ 表示第 i 台主机概率矩阵,$\boldsymbol{A}_i(j)$ 表示概率矩阵 $\boldsymbol{A}_i$ 的第 j 行,$\boldsymbol{A}_i(j,k)$ 表示概率矩阵 $\boldsymbol{A}_i$ 的第 j 行第 k 列.

本节最关键的工作是计算 $P(Y=i|D)$ 值,构造概率矩阵 $\boldsymbol{A}$.

定义 6 设 $M_i(i=1,2,3,4,5)$ 表示第 i 等数据点之和.

本广义立方体总共样本点的个数为 n,直观的第 i 等的频率 M_i/n 近似其概率 $P(Y=i|D)$. 注意,由于 $M_1+M_2+\cdots+M_5<n$,还有部分聚集在广义

立方体中的点没有被指派为 5 等中的任何一等，可认为它们属于不确定的一类非确定(2.5)，记 $M_6=n-(M_1+M_2+\cdots+M_5)$. 则有：

$$P(Y=i\mid D)\approx M_i/n$$

其中，$Y=i=6$ 只是作为技术处理的一个过渡等级非确定(2.5)，主要为了后面的安全态势量化计算需要. 一般地，若评估所需样本量不足，会产生较大误差，因此需 Dirichlet 先验分布平滑后验概率，优化聚类分析结果.

3.2.2 数据优化　假设每个等级 i 的先验服从 Dirichlet 分布：

$$p(\theta)=Dir(\alpha_1,\alpha_2,\cdots,\alpha_6)=\Gamma(\alpha)\prod_{i=1}^{6}\theta^{\alpha_i-1}\Big/\prod_{i=1}^{6}\Gamma(\alpha_i)$$

以 6 个参数概率 $\theta_i=P(Y=i\mid D)$ 为例，即要确定参数 $\theta=(\theta_1,\theta_2,\theta_3,\theta_4,\theta_5,\theta_6)$ 的值. 其中 $\alpha=\alpha_1+\alpha_2+\cdots+\alpha_6$ 为分布的精度，$\alpha_i(i=1,2,\cdots,6)$ 称为超参数，可由历史数据对每个参数取先验值和平滑后验概率值. 这里的 $\Gamma(\cdot)$ 称为 Gama 函数.

那么，样本 D 发生的概率为

$$P(D)=\int p(\theta)p(D\mid\theta)\mathrm{d}\theta=\Gamma(\alpha)\int\prod\theta^{\alpha_i-1}\prod_k\theta_k^{\alpha_k}\mathrm{d}\theta\Big/\prod_k\Gamma(\alpha_i)=\frac{\Gamma(\alpha)}{\Gamma(\alpha+n)}\prod_{i=1}^{6}\frac{\Gamma(\alpha_i+M_i)}{\Gamma(\alpha_k)}$$

参数 θ 的后验概率也服从 Dirichlet 分布：

$$p(\theta\mid D)=p(\theta)p(D\mid\theta)/P(D)=\Gamma(\alpha+n)\prod_k\theta_k^{\alpha_k+M_k}\Big/\prod_i\Gamma(\alpha_i+M_i)=Dir(\alpha_i+M_1,\alpha_2+M_2,\cdots,\alpha_6+M_6)$$

其中：M_i 为训练样本中第 i 等的出现的次数；n 为总的出现次数.

根据概率论知识，有：

$$E(\theta_i\mid D)=\int\theta_i p(\theta_i\mid D)\mathrm{d}\theta_i=\int\theta_i\cdot Dir(\alpha_1+M_1,\alpha_2+M_2,\cdots,\alpha_6+M_6)\mathrm{d}\theta_i=(\alpha_i+M_i)\Big/\Big(\sum_{j=1}^{6}\alpha_j+\sum_{j=1}^{6}M_j\Big)=(\alpha_i+M_i)/(\alpha+n)$$

可以用 $E(\theta_i\mid D)$ 来估计参数 θ_i 的值. 因此，广义立方体中第 i 等的概率为

$$P(Y=i\mid D)=(\alpha_i+M_i)/(\alpha+n)$$

由此，概率矩阵 $\boldsymbol{A}$ 构造完毕，同时也获得作为过渡计算概率 $P(Y=6\mid D)$ 的值.

3.3 主机安全态势

主机安全态势是一个标量，与网络安全态势类似，由主机的基础运行维、主机脆弱维与主机威胁维复合构成. 本文不打算具体计算主机安全态势，但需从主机的 ***IRun***、***IVul*** 和 ***IThreat*** 分别计算生成 *Run*、*Vul* 和 *Threat*，它们是融合成网络安全指数的基础.

以 *Run* 为例，设等级向量 $\boldsymbol{E}=[1,2,3,4,5,2.5]^{\mathrm{T}}$ 为转置矩阵，通过从 ***IRun*** 计算出主机基础运行维：

$$Run'=\boldsymbol{IRun}\cdot\boldsymbol{E}=[P(1,1),P(1,2),P(1,3),P(1,4),P(1,5),P(1,6)]\cdot\boldsymbol{E}=\sum_{i=1}^{5}i\cdot P(1,i)+2.5P(1,6)$$

$$Run=Uprounding(Run')$$

Run' 取 1～5 的实数，*Run* 对其取上整，获得 1 个标量 *Run* 等级. *Vul* 和 *Threat* 等级可类似获取，因此可以得到 $\boldsymbol{HSA}=(Run,Vul,Threat)$，每维取 1～5 的一个具体等级.

注意，概率 $P(Y=6)$、$P(Y=6\mid D)$ 或 $P(1,6)$ 对应于过渡等级的非确定(2.5)，只作为计算技术处理. 下同.

4 网络级评估

4.1 网络安全指数

网络安全指数是指网络安全状况在某时刻所处级别 i 的概率 $P(Y=i)$. 在同一时刻，网络上的每台主机依照第 3 节评估方法，都可以生成主机安全态势的主机基础运行维、主机脆弱维与主机威胁维 3 个维度数据，构成一个向量 $\boldsymbol{HSA}=(Run,Vul,Threat)$. 网络上有 N 台主机，则有 N 个相应向量 ***HSA***，第 i 台主机的安全态势向量表示为 $\boldsymbol{HSA}_i(i=1,2,\cdots,N)$. 本节工作类似于 3.2 节工作，但只需构建一个广义立方体. 本广义立方体就是一个 3 维立方体，维度为 *Run*、*Vul* 与 *Threat*，层次都为 5 层，每台主机的安全态势向量 $\boldsymbol{HSA}_i$ 作为一个点聚集在此立方体中.

为了突出网络节点中服务器等主机的份量，可把这些服务器主机权重增加为普通主机的 λ 倍，一个权重大的节点等价于 λ 个普通节点. 这样网络上 N 个节点，聚集在这立方体中，有多于 N 个点，具体视网络中的服务器主机数量及其权重而定，再按3.2 节的方法进行，得到网络安全指数向量：

$$\boldsymbol{ISA}=[P(Y=1),P(Y=2),\cdots,$$

$P(Y=5),P(Y=6)]$

4.2 网络安全态势

网络安全态势 SA 是一个标量，由 ***ISA*** 经计算得到，类似于 3.3 节主机安全态势任何一维方法的计算，设等级向量 $\boldsymbol{E}=[1,2,3,4,5,2.5]^{\mathrm{T}}$ 为转置矩阵：

$$SA' = \boldsymbol{ISA} \cdot \boldsymbol{E} = [P(Y=1),P(Y=2),\cdots,P(Y=5),P(Y=6)]\boldsymbol{E} = \sum_{i=1}^{5} iP(Y=i) + 2.5P(Y=6)$$

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

SA' 取 1～5 的实数，SA 对其取上整(Uprounding)，获得整个网络安全态势等级.

5 仿真实验

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.

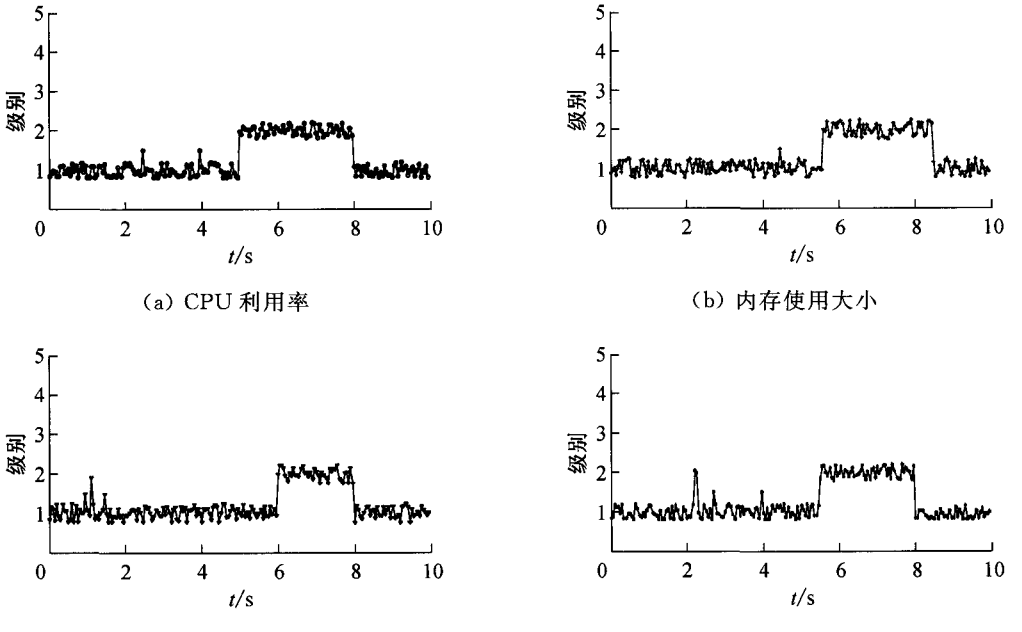
5.1 数据采样

对于连续型原始采样数据，可多次综合应用式(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 5 等离散取值. 为了便于表达，采样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 2 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.

5.2 主机级评估

(1) 样本立方体聚类图. 在网络安全态势评估前，须把所采集的样本离散化处理后，聚集到构造好的广义立方体中，以便计算 n 个样本点所处的立方体格中频率近似其概率. 广义立方体不便绘图，以 CPU 利用率、内存占用大小和子网平均数据流 3 个评估指标为例，绘出 3 维立方体聚类图，动态采集 2 000 个数据，经离散化后聚集在 3 维立方体中，如图 3 所示. 从实验可以看出，样本点聚集于等级 1(对应于立方体格 $C(1,1,1)$)的数量要占绝大多数，等级 2(对应于立方体格 $C(2,2,2)$)次之，其他 3,4,5 等级逐渐递减，当然还有少量样本点散落在其他立方格中.

(2) 主机一级指数图. 从下而上，n 个样本 D 经广义立方体聚类后，主机 3 个一级指数(***IRul***, ***IVul***, ***IThreat***)中的每维各有 5 个概率值. 就主机基础运行维指数而言，把样本 D 分为“安全”的概率为 $P(Y=1|D)$、…、把样本 D 分为“高度危险”的概率为 $P(Y=5|D)$、而样本 D 分类不确定的概率为 $P(Y=6|D)$，其他 2 个指数类似. 图 4 给出了 2 000 个样本当时主机一次广义立方体聚类后，主机每维取相应的概率值的情况. 图中：第 1 组纵列表示该 ***IRun*** 的 6 个等级概率；第 2 组纵列表示该 ***IVul*** 的 6



(a) CPU 利用率　(b) 内存使用大小

(c) 子网平均数据流　(d) 子网带宽使用率

图 2 离散化数据采样图

Fig. 2 Sampling of discrete data

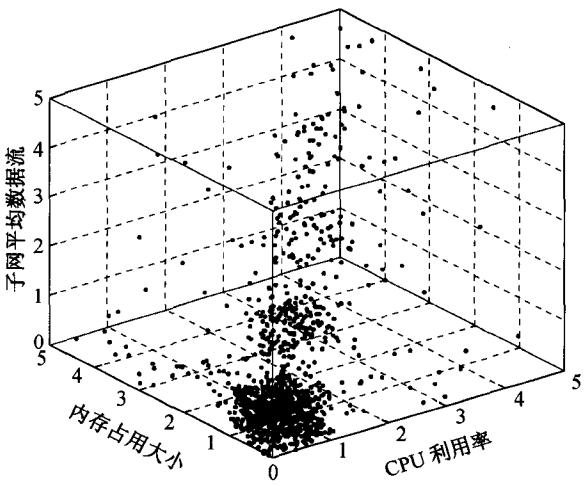


图 3　3-维立方体聚类图

Fig. 3　Three-dimensional cube clustering

个等级概率；第 3 组纵列表示该 ***IThreat*** 的 6 个等级概率. 从结论可以看出，各维取第 1 等级的概率 $P(Y=1|D)$都要高，且每维 6 个等级概率之和为 1，其中等级 6 为过渡等级.

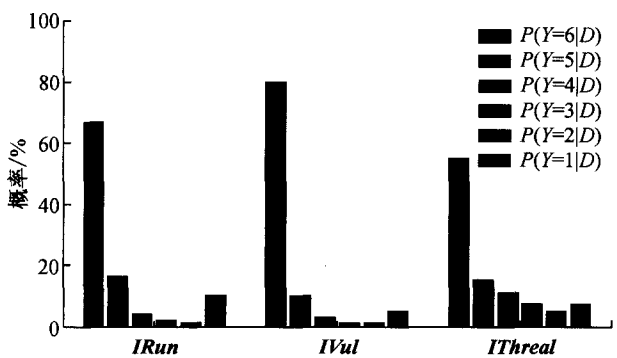


图 4　主机一级指数图

Fig. 4　First level index of host

(3) 主机安全态势向量图. 由 *Run*、*Vul* 与 *Threat* 构成主机安全态势向量 ***HAS*** $=(Run, Vul, Threat)$，从 ***IRun***、***IVul*** 和 ***IThreat*** 分别计算生成. 为了清晰表达，不打算对 3 维取整，否则在图形中变成了 3 根折线，不便区分. 如图 5 所示给出了 10 次主机安全态势向量，网络受到攻击时，相应的维度也会引起波动，但主机脆弱维相对比较稳定，因为主要涉及到较固定的配置信息.

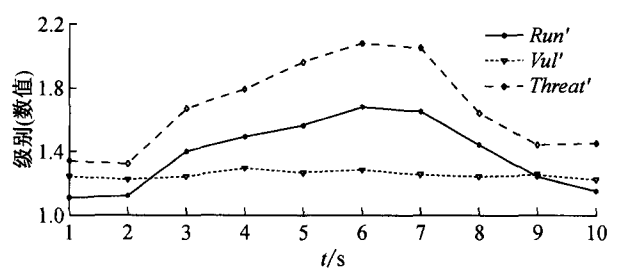


图 5　主机安全态势向量对比图

Fig. 5　Comparison of host security situation vector

5.3　在线评估

(1) 网络安全指数图. ***ISA*** 从网络上众多主机的安全指数经 3 维(Run、Vul 与 $Threat$)立方体聚类而成，类似于主机安全指数的获取. 网络实验环境上的主机共 90 台，图 6 描绘了 15 个时刻的网络安全指数，网络上受到攻击时，相关时刻指数也会产生波动. 图中，每个时刻上的 6 个值(纵向)之和为 1，分别对应着安全态势等级 1～6 等的概率.

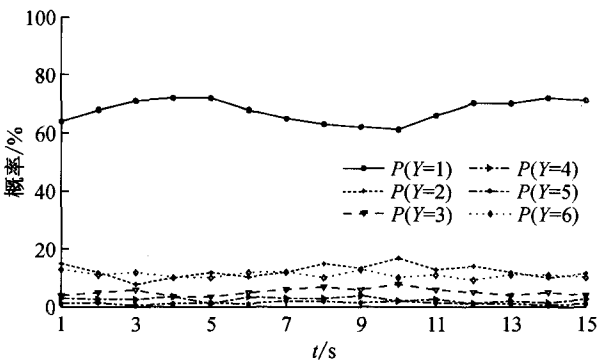


图 6　网络安全指数对比图

Fig. 6　Comparison of network security index

(2) 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 2 所示某个时刻 t 网络上一主机所有评估指标的离散取值. 网络上多少个主机，在这个时刻 t 时就有多少个类似参数表 2，共同融合成网络的安全态势 SA.

表 2　主机评估指标所取离散值

Tab. 2　Discrete values of host assessment indicators

主机评估指标					
IRun		***IVul***		***IThreat***	
指标项	5 等值	指标项	5 等值	指标项	5 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				

表 2 表示此台主机正受到网络攻击，因为威胁维指数 5 等值基本上处于 2 和 3 等级，经融合可得网络安全态势指数为[0.69,0.14,0.06,0.03,0.01,0.07]，与相应等级$[1,2,3,4,5,2.5]^{T}$之积，得 $SA'=1.495$，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. 在本文实验中，对贝叶斯网评估方法、广义立方体聚类评估方法以及本文中的 Dirichlet 平滑后的广义立方体聚类评估方法作了比较，网络安全态势评估对比如图 7 所示. 根据网络安全态势等级划分，若对 SA'的值取上整，Dirichlet 平滑后的广义立方体聚类方法有同等的效果，及时反映网络当前的安全状况.

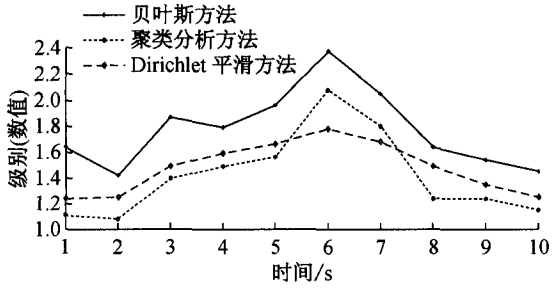


图 7 网络安全态势评估对比图

Fig. 7 Comparison of network security assessment

6 结 语

网络安全态势评估是指在一定的时空条件下提取与网络安全相关的要素并进行信息融合，对其综合分析与理解，进而把握当前网络安全状况与预测安全态势发展趋势. 针对目前网络安全态势评估方法中的大多信息来源单一、评估范围有限、模型训练与参数获取不易、时空开销较大且可信度不高等问题，本文提出了基于聚类分析量化评估网络安全态势方法. 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，通过构建二级广义立方体快速高效融合主机上的多源异构非确定性信息源，并利用 Dirichlet 先验分布平滑后验概率，优化聚类分析结果；然后构建网络上的广义立方体，聚类融合网络上各主机的安全态势，逐步量化评估并生成网络当前安全态势，具有可信性. 广义立方体在聚类分析与信息融合上，具有快速高效直观性，易于实现.

参考文献：

[1] BASS T. Intrusion detection systems and multi-sensor data fusion: Creating cyberspace situational awareness[J]. **Communications of the ACM**, 1999, 43(4): 99-105.

[2] GÖRNITZ N, KLOFT M, RIECK K, *et al*. Toward supervised anomaly detection[J]. **Journal of Artificial Intelligence Research**, 2012, 46(1): 235-262.

[3] BRADSHAW J M, CARVALHO M, BUNCH L, *et al*. Sol: An agent-based framework for cyber situation awareness[J]. **Künstl Intell Springer**, 2012, 26(2):127-140.

[4] LIU Sunjun, LU Le, YANG Jin. Research on network security situation awareness technology based on AIS[J]. **International Journal of Knowledge and Language Processing**, 2011, 2(2): 23-34.

[5] GUO Chunxiao, SU Yang. A new optimized algorithm based on quantum evolutionary strategy for network security situation prediction[J]. **Journal of Chinese Computer Systems**, 2014, 35(6):1248-1252.

[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 Yuling, FENG Dengguo, LIAN Yifeng, *et al*. Network situation prediction method based on spatial-time dimension analysis[J]. **Journal of Computer Research and Development**, 2014, 51(8):1681-1694.

[7] DIGIOIA G, FOGLIETTA C, OLIVA G, *et al*. Aware online interdependency modeling via evidence theory[J]. **International Journal of Critical Infrastructures**, 2013, 9(1/2): 74-92.

[8] JANSEN A, MELCHERS K B, 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): 326-328.

[9] QIN Biao, XIA Yuni, WANG Shan, *et al*. A novel Bayesian classification for uncertain data[J]. **Knowledge-Based Systems**, 2011,24(8): 1151-1158.

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

[11] 韦勇，连一峰，冯登国，等. 基于信息融合的网络安全态势评估模型[J]. **计算机研究与发展**,2009, 46(3): 353-362.
WEI Yong, LIAN Yifeng, FENG Dengguo. A network security situational awareness model based on information fusion[J]. **Journal of Computer Research and Development**, 2009, 46(3): 353-362.

(下转第 1421 页)

部特征的舰船型号识别的方法. 其中，SIFT 为主要的识别特征，识别时进行分区匹配，Harris 角点多分布于舰身边缘，作为补充特征加入. 实验证明，该算法可以有效地实现舰船型号的识别. 本文遥感图像分辨率为 0.5 m，当分辨率增加时，舰船细节信息丰富，采用本文方法的型号识别率会进一步提高. 但是成像系统受其固有的传感器阵列排列密度的限制，遥感影像的分辨率不能很高，那么超分辨率重建技术便能发挥作用. 由此可见，研究对光学遥感影像切实有效的超分辨率重建技术意义重大，这将是今后的一个研究方向.

参考文献：

[1] 张风丽，张磊，吴炳方. 欧盟船舶遥感探测技术与系统研究的进展[J]. **遥感学报**，2007，11(4)：552-562.
ZHANG Fengli，ZHANG Lei，WU Bingfang. Progress of ship detection technology and system based on remote sensing technology in European Union[J]. **Journal of Remote Sensing**，2007，11(4)：552-562.

[2] 李毅，徐守时. 基于支持向量机的遥感图像舰船目标识别方法[J]. **计算机仿真**，2006，23(6)：180-183.
LI Yi，XU Shoushi. A new method for ship target recognition based on support vector machine[J]. **Computer Simulation**，2006，23(6)：180-183.

[3] 杜春，孙即祥，李智勇，等. 光学遥感舰船目标识别方法[J]. **中国图象图形学报**，2012，17(4)：589-595.
DU Chun，SUN Jixiang，LI Zhiyong，*et al*. Method for ship recognition using optical remote Sensing data[J]. **Journal of Image and Graphics**，2012，17(4)：589-595.

[4] LOWE D G. Distinctive image features from scale-invariant key points[J]. **International Journal of Computer Vision**，2004，60(2)：91-110.

[5] BICEGO M，LAGORIO A，GROSSO E，*et al*. On the use of SIFT features for face authentication[C]// **2006 Conference on Computer Vision and Pattern Recognition Workshop**. New York：IEEE Press，2006：35.

[6] KIM N P，PARK T W. Assessing the performance of corner detectors for point feature tracking applications[J]. **Image and Vision Computing**，2004，22(8)：663-679.

[7] AZAD P，ASFOUR T，DILLMANN R. Combining Harris interest points and the SIFT descriptor for fast scaleinvariant object recognition[C]// **2009 IEEE/RSJ International Conference on Intelligent Robots and Systems**. St Louis：IEEE Press，2009：4275-4280.

[8] SMITH S M，BRADY J M. SUSAN—A new approach to low level image processing[J]. **International Journal of Computer Vision**，1997，23(1)：45-78.

[9] 徐玮，王炜，张茂军，等. 一种基于角点匹配的视图合成方法[J]. **系统仿真学报**，2007，19(14)：3263-3265.
XU Wei，WANG Wei，ZHANG Maojun，*et al*. Corner matching-based approach of view synthesis[J]. **Journal of System Simulation**，2007，19(14)：3263-3265.

(上接第 1414 页)

[12] 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 Lixia，WANG Yachao，YU Jinbo. Network security situation awareness based on neural networks[J]. **Journal of Tsinghua University(Science and Technology)**，2013，53(12)：1750-1760.

[13] 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 Rongrong，YUN Xiaochun，ZHANG Yongzheng，*et al*. An improved quantitative evaluation method for network security[J]. **Chinese Journal of Computers**，2015，38(4)：749-758.

[14] 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 Yong，TAN Xiaobin，CUI Xiaolin，*et al*. Network security situation awareness approach based on Markov game model[J]. **Journal of Software**，2011，22(3)：495-508.

[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.