doi:10.3969/j.issn.1000-436x.2013.07.008

面向网络状态的自适应用户行为评估方法

陆悠[1,2]，罗军舟[1]，李伟[1]，于枫[1]，夏怒[1]
（1. 东南大学 计算机科学与工程学院，江苏 南京 210096；2. 苏州科技学院 电子与信息工程学院，江苏 苏州 215000）

摘 要：用户复杂的行为往往会导致网络状态出现波动，破坏网络平稳运行，由于指标及权重的主观性和静态性，传统评估方法难以准确衡量用户行为对网络状态变化的影响程度。因此引入粗糙集理论，构建一种面向网络状态的自适应用户行为评估方法，使用属性约简和属性重要度方法对用户行为和网络状态数据进行挖掘，分析用户行为与网络状态变化的关联程度，以此自适应构建评估指标及权重，并随用户行为变化而动态调整，从而准确地量化用户行为对网络状态变化的影响程度。实验结果表明，该评估方法有助于准确发现造成网络状态变化的用户及其行为，能够为加强对用户行为的管控提供有效支持。

关键词：用户行为评估；网络状态；影响力；粗糙集
中图分类号：TP301 **文献标识码**：A **文章编号**：1000-436X(2013)07-0071-10

Adaptive user behavior’s evaluation method based on network status

LU You[1,2], LUO Jun-zhou[1], LI Wei[1], YU Feng[1], XIA Nu[1]
(1. School of Computer Science and Engineering, Southeast University, Nanjing 210096, China；
2. School of Electronical and Information Engineering, Suzhou University of Scince and Technology, Suzhou 215000, China)

Abstract: With the development of network, user’s complex and dynamic behaviors often lead to unexpected fluctuations of network status, and these fluctuations take great challenge to the stability of network. It is difficult to evaluate user behavior’s affection on network status fluctuations for traditional methods because of their subjectivity and static drawbacks. An evaluation method was proposed, which analyses the correlation between the user’s behavior and network status fluctuations based on actual data of behavior and network status, by means of rough set attribute reduction and attribute importance. This method can construct and dynamically adjust the evaluation indexes and their weight adaptive the actual data. The evaluation results show this method can help to detect and manage users who affect the stability of network, and provide effective support to control users and their beha ior.

Key words: user behavior evaluation; network status; behavior effection; rough set

1 引言

随着网络用户规模的不断扩大，其行为也日趋复杂，一些用户行为会对网络运行状态造成很大的影响，例如用户的突发(busting)访问、滥用资源(P2P下载等)甚至恶意攻击(attack)等行为会造成网络负载突增或失衡，导致网络拥塞等，若这种非预期的网络状态变化不能及时得到处理则会造成更为严

收稿日期：2012-10-30；修回日期：2013-01-21
基金项目：国家重点基础研究发展计划(“973”计划)基金资助项目(2010CB328104)；国家自然科学基金资助项目(61070158, 61003257, 61070161, 61070210)；国家高技术研究发展计划（“863”计划）基金资助项目(2013AA013503)；高等学校博士点学科专项科研基金资助项目(20110092130002)；江苏省网络与信息安全重点实验室基金资助项目(BM2003201)；教育部计算机网络与信息集成重点实验室基金资助项目(93K-9)
Foundation Items: The National Basic Research Program of China (973 Program) (2010CB328104); The National Natural Science Foundation of China (61070158, 61003257, 61070161, 61070210); The National High Technology Research and Development Program of China (863 Program) (2013AA013503); China Specialized Research Fund for the Doctoral Program of Higher Education (20110092130002); Jiangsu Provincial Key Laboratory of Network and Information Security (BM2003201); Key Laboratory of Computer Network and Information Integration of Ministry of Education of China (93K-9)

重的如网络节点异常甚至网络瘫痪等后果。因此为维护网络的平稳运行，有必要在网络出现非预期波动后及时分析和量化用户行为对网络状态变化的影响程度，从而有助于发现造成网络状态变化的用户并对其进行控制及追责，避免出现更为严重的后果，实现对用户行为的有效管控和网络的精细化管理[1,2]。

量化评估用户行为在面向应用的电子商务、网络安全等领域中已得到广泛应用，如电子商务领域基于评价反馈数据对用户行为进行的可信评估[3~6]方法，使用幂次法则[3]、模糊逻辑[4]、半环(semi-ring)代数[5]以及概率论[6]等工具对用户的信誉值进行量化评估，然而这些评价反馈数据仅反映应用对用户行为的评价而无法反映网络状态，故不能直接用于量化分析用户行为对网络状态变化的影响程度；又如在网络安全领域的用户攻击行为效果评估，典型的如基于 AHP 层次分析评估方法，其基于公认的评估准则(如 ISO 7498-2 标准等)构建评估指标并使用 AHP 层次分析方法确定权重，对用户的行为攻击效果进行评估量化。但由于确定权重的 AHP 层次分析方法容易引入模糊性而造成权重失真[7]，因此也有研究人员引入知识挖掘技术来提高权重的合理性，如模糊逻辑与 AHP 层次分析方法相结合[8]，灰色理论与 AHP 层次分析法相结合[9]，基于信息熵[10,11]和粗糙集[12]方法等。这些方法虽然在量化用户攻击效果方面取得不错的效果，但将这些方法直接用于评估用户行为对网络状态变化的影响程度仍面临评估指标主观性和静态性问题，主要原因在于：一方面，采用基于经验、专家建议或者评估准则等方法构建评估指标将不可避免地带来主观臆断，从而影响评估的准确性[7]；另一方面，攻击效果评估方法的指标和权重设置后是静态的，而网络和用户行为则处于动态变化中，引起网络状态变化的用户行为可能各不相同，而同样的用户行为(如 P2P 下载行为)在不同的网络环境(如在网络空闲与繁忙的不同时段)下对网络状态变化的影响程度亦有所差异，因此欲准确评估用户行为对网络状态变化的影响程度，需进一步深化对知识挖掘技术的应用，使用户行为评估方法的指标及权重构建都能够建立在对用户行为和网络状态实际数据进行分析和挖掘基础之上，且指标及权重能随网络环境和用户行为的变化而动态调整，从而克服已有评估方法主观性和静态性弊端。

如果网络状态的非预期变化是由用户的某些行为(如突发访问、P2P 下载或恶意攻击等)所导致的，那么这些行为的一些特定特征必然与网络状态变化之间存在较强的关联性，因此可通过对用户行为数据与网络状态数据的关联挖掘来帮助量化用户及其行为对网络状态变化造成的影响。传统的关联挖掘方法(如相关分析等)往往需要人工预置阈值(如置信度等)，一旦设置不当会影响分析结果，而粗糙集[13]则无需任何先验知识，可以自适应地发现数据之间的关联性，因此本文提出了一种基于粗糙集的自适应用户行为评估方法，在无需事先设置评估指标和权重的前提下，通过引入粗糙集属性约简和属性重要度方法对用户行为和网络状态实际数据进行分析挖掘，在此基础上自适应地设置及调整评估指标和权重，以此量化评估用户行为对网络状态变化的影响程度。实验结果表明，与传统预置指标和权重的评估方法相比，本文评估方法能更准确地量化用户行为对网络状态变化的影响程度，评估结果有助于准确发现影响网络运行的用户，从而为制定相应的用户控制策略，有效控制用户行为提供支持。

2 用户行为评估模型

2.1 用户行为影响力

评估用户行为影响力实质是从网络出发，定量地分析用户各种行为对网络状态变化的影响程度。为更好地描述用户行为影响力，本文首先给出以下定义。

定义 1 网络状态。指由各个网络节点的运行状况、资源效用等因素所构成的整个网络的工作性能和综合状态。在动态的网络环境中，可以从网络的每个节点处测量如宽带利用率、链路流量及其他资源或设备运行状况等因素(称状态因子)的测量值并按一定权重量化来获得，设有 h 项状态因子 $\{D_1,D_2,\cdots,D_h\}$，每个因子对应权重为 $\{?_1,?_2,\cdots,?_h\}$，其中，$?_i\in[0,1]$ 且 $\sum_{i=1}^{h}w_i=1$。为避免取值范围引起的误差，可参考文献[12]的标准化方法，对获取的状态因子测量值进行处理，使其取值范围统一为值域 $[0,1]$，设结果为 $\{?_1,?_2,\cdots,?_h\}$，则网络状态计算公式为

$$d=\sum_{i=1}^{h}e_iw_i \tag{1}$$

由于任意 $e_i \in [0,1]$ 且 $\sum_{i=1}^{h} w_i = 1$，于是有 $d \in [0,1]$，所以网络状态变化可用不同时刻的 d 值差异来刻画。显然，网络状态的计算将依赖于状态因子的设置，在网络态势分析领域，网络状态的描述与评估的研究较为丰富，已有的研究成果[14,15]通常采用包括吞吐率、延迟、带宽使用率以及安全警报信息等状态因子反映网络状态。

定义 2　用户行为特征。指用于描述用户网络行为且可在网络中测量获取的各项考察因素，设集合 $C_{raw}=\{C_1,C_2,\cdots,C_n\}$ 为描述用户网络行为特征的 n 项指标，$?_1,?_2,\cdots,?_n$ 为某时刻 n 项用户网络行为特征指标对应的测量值，则用户的网络行为原始数据可由向量 $x=\{?_1,?_2,\cdots,?_n\}$ 表示。

为评估的准确性，须选择对用户应用覆盖面较广的行为特征集 C_{raw}。目前也有较多研究成果可借鉴，包括流量识别、网络安全等领域都提出了候选行为特征集，如文献[16]提出高达 246 种用户行为特征等。

定义 3　用户行为评估指标。指用于评估用户行为对网络状态变化所造成影响的评价指标，即影响网络状态变化的用户行为所对应的特定特征，可表述为 $Index=\{<c_1,v_1>, <c_2,v_2>,\cdots,<c_{n'},v_{n'}>\}$，且 $n'<n$。其中，$c_i \in C_{raw}$ 为使用粗糙集属性约简方法从用户行为特征集 C_{raw} 中获取的与网络状态相关的 n' 项最小特征集中的元素，$v_i(i=1,\cdots,n')$ 为 c_i 的权重，其值即特征 c_i 使用粗糙集属性重要度方法计算所得的属性重要度，且有 $0 \le ?_i \le 1$，$\sum_{i=1}^{n'} ?_i = 1$。其中，属性约简和属性重要度方法主要用于分析用户行为特征与网络状态变化的相关性，其计算方法可参考文献[13]。

定义 4　用户行为影响力。即用户行为对网络状态变化的关联程度，用 e 表示，其值可由函数

$$e = F(x', Index) = \frac{?_1 + ?_2 + \cdots + ?_{n'}}{n'}, \quad e = \in [0,1] \qquad (2)$$

进行计算，其中 $Index=\{<c_1,v_1>,<c_2,v_2>,\cdots,<c_{n'},v_{n'}>\}$ 为用户行为评估指标，设 $x'=\{b_1, b_2,\cdots,b_{n'}\}$ 为采集并经过预处理后所得到 n' 项评估指标上的一条用户行为数据，令 $?_i = b_i \times ?_i (i \in [1,n'])$ 为用户行为在具体评估指标元组 $<c_i,v_i>$ 上所得到的评估值，则该条数据对应用户行为的影响力 e 可由 n' 项评估值的算术平均值表示，而一段时间内用户行为对网络状态变化的影响力则可由该段时间内所有行为数据影响力的平均值表示，其具体计算过程在 3.3 节进行介绍。

2.2　基于影响力的用户行为评估模型

为准确地量化用户行为对网络状态变化的影响力，基于影响力的用户行为评估方法首先需要充分采集用户行为与网络状态数据，然后基于实际数据使用粗糙集方法分析不同的用户行为特征与网络状态变化之间的关联性，继而有侧重地选择与状态变化相关的最小用户行为特征集作为评估指标，按关联程度设置其权重，最后按评估指标和权重计算用户行为影响力，对用户行为数据进行量化评估。据此，本文提出基于影响力的用户行为评估模型，包括数据采集、数据预处理、评估指标及权重设置和用户行为评估 4 个模块，如图 1 所示。

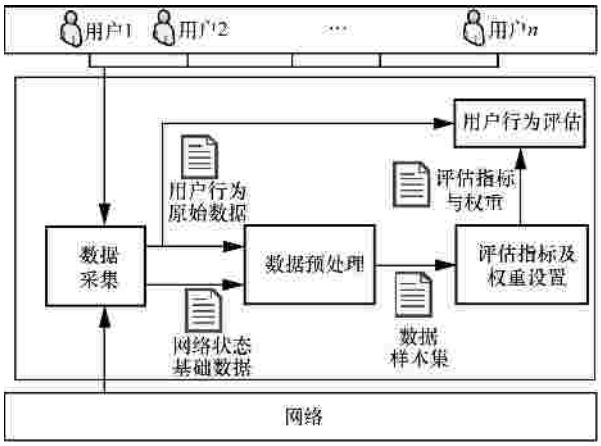

图 1　基于影响力的用户行为评估模型

1) 数据采集模块：该模块仅在逻辑上表示负责获取用户行为与网络状态变化的原始数据，而在实现上可利用分布式的第三方如 NetFlow Monitor、Bandwidth Monitor 等采集器构建，其中，部署于用户接入点(如网关、3 层交换机等)的采集器负责收集用户行为数据，而网络状态数据可由部署于网络的核心节点以及服务器等网元的采集器收集。采集器获取的数据经初步处理(如统计均值、极值、比例值等)后形成反映用户行为和网络状态的原始数据。

2) 数据预处理模块：该模块负责对采集模块生成的原始数据进一步进行标准化、离散化，构建评估指标及权重设置模块使用的数据样本集，具体处

理方法详见 3.1 节。

3) 评估指标及权重设置模块：该模块在数据样本集基础上，使用粗糙集中的属性约简、属性重要度方法来挖掘分析用户行为与网络状态变化之间的关联性，找出与当前网络状态变化相关的最小用户行为特征集作为评估指标，摒弃无关特征，并根据相关性大小来设置评估指标的权重。

4)用户行为评估模块：该模块根据评估指标和权重对用户行为数据进行量化评估，获得的评估值反映用户行为对网络状态变化的影响力。

数据预处理、评估指标及权重设置以及用户行为评估 3 个模块均可部署于网络管理的中心节点进行集中式处理，数据采集模块则呈分布式部署于网络各个节点，因此在体系结构上本文评估模型能够与现有的网络管理系统有效的融合，具有较强的可实现性。

3 用户行为评估方法

3.1 数据预处理

设置评估指标及权重是用户行为影响力评估的核心内容，其依赖于对用户行为与网络状态实际数据的挖掘和分析。为分析的准确和便利，需要对测量采集到的用户行为和网络状态数据进行预处理，预处理过程分为以下 3 步。

Step1 构建原始数据矩阵。即关联网络状态和用户行为的测量数据：对网络用户行为与网络状态分别进行采样，设时间段 t 内，用户行为数据采集器获取 m 条用户行为原始数据，其中每一条数据 x_i (i=1,⋯,m)都是用户在行为特征集 C_{raw} 的 n 个特征下的测量值构成的向量$\{x_{i_1}, x_{i_2}, \cdots, x_{i_n}\}$，与此同时网络状态采集器获取 m'条网络状态原始数据，其中，每条数据 e_j (j=1,⋯,m')都是网络在状态因子集 $\{D_1, D_2, \cdots, D_h\}$ 下的测量值标准化后构成的向量 $\{e_{j_1}, e_{j_2}, \cdots, e_{j_n}\}$，使用式(1)计算得到网络状态值 d_j(j=1,2,⋯,m')，对每个 x_i 根据时间戳寻找对应的网络状态数据 d_j 并令其为 y_i，将 x_i 与 y_i 联立，构成时间段 t 内 m×(n+1)原始数据矩阵 $\boldsymbol{X}$。

$$\boldsymbol{X}=\begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1n} & y_1 \\ x_{21} & x_{22} & \cdots & x_{2n} & y_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ x_{m1} & x_{m2} & \cdots & x_{mn} & y_m \end{bmatrix}_{m\times(n+1)}$$

Step2 标准化。原始数据矩阵中的用户行为数据由采样数值构成，一般是物理量纲值(dimension data)或百分比值(percentage data)，为避免取值范围引起的误差，需对其标准化(网络状态值 y_i 计算时已标准化，在此无需处理)。设矩阵 $\boldsymbol{X}$ 中第 j(j=1,⋯, m+1)列最大值为 $r_{\max}^j$，最小值为 $r_{\min}^j$。于是得规范数据矩阵 $\boldsymbol{B}$ 为

$$\boldsymbol{B}=\begin{bmatrix} b_{11} & b_{12} & \cdots & b_{1n} & y_1 \\ b_{21} & b_{22} & \cdots & b_{2n} & y_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ b_{m1} & b_{m2} & \cdots & b_{mn} & y_m \end{bmatrix}_{m\times(n+1)}$$

其中，

$$b_{ij}=\begin{cases} x_{ij}, & x_{ij}\text{正向递增的百分比值} \\ 1-x_{ij}, & x_{ij}\text{正向递减的百分比值} \\ (x_{ij}-r_{\min}^j)/(r_{\max}^j-r_{\min}^j), & x_{ij}\text{正向递增的物理量纲值} \\ (r_{\max}^j-x_{ij})/(r_{\max}^j-r_{\min}^j), & x_{ij}\text{正向递减的物理量纲值} \end{cases}$$

Step3 离散化。规范数据矩阵 $\boldsymbol{B}$ 中数值的取值范围是连续的，因此 $\boldsymbol{B}$ 中的每一列可能存在较多互相接近而不等的值，不利于粗糙集运算，因此需要将值按一定分界区间进行归并，即对数据矩阵 $\boldsymbol{B}$ 进行离散化。

对用户行为数据而言，由于行为复杂性，数据测量值范围及分布都是未知的，不适宜使用面向均匀分布的等宽、等频区间等离散化方法，本文采用基于熵的离散化方法[17]。该方法通过信息熵来分析样本数据的分布情况，通过寻求熵的损失与适度的区间数之间的最佳平衡来得到优化的区分边界，从而保留原数据分布所蕴含的内在知识信息。

对网络状态数据来说，其取值范围、分布以及评价标准通常是固定的，因此可以事先制定区间边界，本文选择 5 级制作为区间划分标准。则离散化后的网络状态值 y_i' 可以经由式(3)计算。

$$y_i'=\begin{cases} 5, & 0.8\le y_i\le 1 \\ 4, & 0.6\le y_i<0.8 \\ 3, & 0.4\le y_i<0.6 \\ 2, & 0.2\le y_i<0.4 \\ 1, & 0\le y_i<0.2 \end{cases} \tag{3}$$

经规范化和离散化，原始数据转化为可直接用于粗糙集方法的样本集。由于分析目的是发现

用户行为与网络状态变化之间的关联性，为分析便利，可首先按网络状态类别对样本分类，随后每个类别下对数据按时间排序，样本集结构如表 1 所示。

表 1　用户行为网络影响力评估的样本集

时间序列	用户行为				网络状态
	C_1	C_2	…	C_n	D
x_1	b'_{11}	b'_{12}	…	b'_{1n}	1
…	…	…	…	…	
x_t	$b'_{t'1}$	$b'_{t'2}$	…	$b'_{t'n}$	2
…	…	…	…	…	
…	…	…	…	…	…
…	…	…	…	…	5
x_m	b'_{m1}	b'_{m2}	…	b'_{mn}	

3.2　评估指标与权重的自适应设置

粗糙集[13~16]理论中，“知识”被视为集合划分的能力，通过把划分集合的知识嵌入集合本身来扩展经典集合论。如果将经过预处理的用户行为及对应网络状态数据视作一条条待分析和推理的信息，则数据预处理后所得的样本集即粗糙集的系统决策表(SDT, system decision table)，SDT=$\{U,A,V,f\}$中，其中，$U=\{x_1, x_2,\cdots,x_m\}$称为论域，$x_i$ 为用户行为网络影响力评估样本集的一条数据，$A=C\cup D$ 为条件属性和决策属性的并集，条件属性 $C=C_{\text{raw}}=\{C_1,C_2,\cdots,C_n\}$即用户行为特征集，决策属性 D 为网络状态，V 为论域元素值域，f：$U\times A\to V$ 表示元素 $x\in U$ 在属性 $a\in A$ 上取值。

用户行为评估指标与权重构建和调整依赖于对用户行为与网络数据的关联分析，而粗糙集中的属性重要度恰好反映了条件属性与决策属性的相关性。因此可以通过计算属性重要度来进行关联分析。由于用户行为评估的数据中存在不一致性特点，同一时间戳的网络状态往往对应多名用户的不同行为，因此本文采用信息观下的粗糙集方法，参考文献[13]，以空集为初始集合，通过计算不同行为特征增减后对集合的条件熵变化程度来计算属性重要度，并以其为启发搜索行为特征中与网络状态关联度最大的最小子集，进行属性约简，于是约简结果中的用户行为特征即构成评估指标集，每个特征的属性重要度即构成评估指标的权重，算法的详细流程如图 2 所示。

输入：用户行为网络影响力评估的样本集 $B=\{U, A=C\cup D, V, f: U\times A\to V\}$

输出：评估指标及权重集合 EVA$\{\{c_1, v_1\},\{c_2, v_2\},\cdots,\{c_n, v_n\}\}$，其中，$c_i\in C$，$v_i\in[0,1]$

1) 初始化，设属性约简集 RED(U)←C；EVA←$\varnothing$

2) 对 C 中每一个属性 c_i 使用下式计算属性重要度：

$$\text{SIG}(c_i)=(H(D|RED(U)-\{c_i\})-H(D|RED(U)))/H(D)$$

其中，H 为粗糙集中条件熵，计算式[13]为

$$H(A|B)=\sum_{i=1}^{m}\sum_{j=1}^{n}\frac{|Y_j\cap X_i|}{|U|}\left|\frac{|\overline{Y_j}\cap\overline{X_i}|}{|U|}\right.$$

，Y_j 为属性集合 A 对应的等价类，X_i 为属性集合 B 对应的等价类

将集合 C 的属性按重要度递增排列，得到集合

$A^*=\{c_1, c_2,\cdots, c_n\}$，$n=|C|$；

3) While $H(D|\text{RED}(U))\neq H(D|C)$

4) 取 $a_i\in A^*$，且属性重要度最小，RED(U)←RED(U)-$\{a_i\}$，A^*←A^*-$\{a_i\}$，

5) 对 RED(U)中的属性重要度进行标准化，对每一个属性 $c_i\in$RED(U)：

EVA←EVA$\cup\{c_i, \text{SIG}(c_i)\}$

6) 返回 EVA

图 2　自适应评估指标选择与权重设置算法

3.3　用户行为评估

根据行为评估指标与权重对测量得到的用户行为数据量化计算即可得用户行为对网络状态变化的影响程度评估值，根据评估需要，在相应时间段内采集的数据中抽取评估对象用户的所有行为数据，对每一条数据按评估指标投影，获取评估指标对应的行为数值，按评估指标及权重对其量化并取所有行为数据结果的平均值，即该用户的行为影响力评估值，具体过程可描述如下。

Step1　抽取用户 user 在时间段 t 的规范数据集 B 中所有 m' 条行为数据，构造用户 user 的行为数据规范矩阵 $\boldsymbol{B}(user)$，即

$$\boldsymbol{B}(user)=\begin{bmatrix} b_{11} & b_{12} & \cdots & b_{1n} \\ b_{21} & b_{22} & \cdots & b_{2n} \\ \vdots & \vdots & \vdots & \cdots \\ b_{m'1} & b_{m'2} & \cdots & b_{m'n} \end{bmatrix}_{m'\times(n)}$$

Step2　矩阵 $\boldsymbol{B}(user)$ 中每一行都是用户在用户行为特征集合 $C_{\text{raw}}=\{C_1,C_2,\cdots,C_n\}$ 上采集到并预处理后的数据，按行为评估指标 $Index=\{<c_1,v_1>, <c_2,v_2>, \cdots, <c_{n'},v_{n'}>\}$ 对其进行投影(使用θ函数表示)，即取 $B(user)$中所有特征 $c_i(i\in[1,n'])$对应列的数

据，即可得评估所需的行为规范数据矩阵 X'：

$$X' = q\,(B(user), Index) = \begin{bmatrix} b'_{11} & b'_{12} & \cdots & b'_{1n'} \\ b'_{21} & b'_{22} & \cdots & b'_{2n'} \\ \vdots & \vdots & \vdots & \vdots \\ b'_{m'1} & b'_{m'2} & \cdots & b'_{m'n'} \end{bmatrix}_{m'\times n'}$$

Step3 X' 中的每条数据按式(2)计算对应的影响力，最后取 m' 条行为数据对应的影响力平均值作为用户行为的评估值 E

$$E = \frac{e_1 + e_2 + \cdots e_{m'}}{m'}$$

其中，

$$\begin{bmatrix} e_1 \\ e_1 \\ \vdots \\ e_{m'} \end{bmatrix} = \frac{1}{n'} \times \begin{bmatrix} b'_{11} & b'_{12} & \cdots & b'_{1n'} \\ b'_{21} & b'_{22} & \cdots & b'_{2n'} \\ \vdots & \vdots & \vdots & \vdots \\ b'_{m'1} & b'_{m'2} & \cdots & b'_{m'n'} \end{bmatrix}_{m'\times n'} \times \begin{bmatrix} ?_1 \\ ?_2 \\ \vdots \\ ?_{n'} \end{bmatrix}$$

由于评估方法反映了用户行为与网络状态变化的影响程度的大小，因此基于评估值有助于发现造成网络状态变化的用户，进而可对不同影响力的用户进行控制、激励或追责，降低网络可能遇到的风险。

4 实验及分析

4.1 实验环境

为评价本文用户行为评估方法的准确性和有效性，本文使用苏州科技学院 2 名学生机房中采集到的真实数据进行分析，其实验网络拓扑结构如图 3 所示。

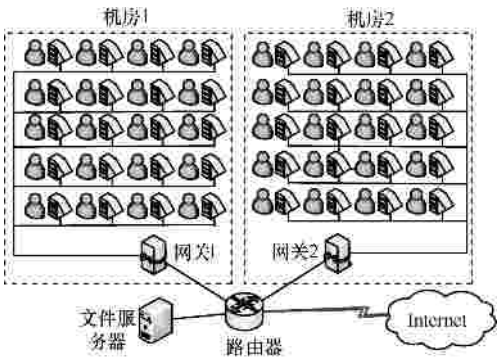

图 3 实验网络拓扑

实验环境由 2 间分别可容纳 20 名学生的机房构成，每个机房通过网关连接至路由器并接入 Internet，另外路由器还与一台 FTP 文件服务器相连接。在 2 个网关采集用户行为数据，同时在网关、服务器及路由器处采集网络状态数据。2 个机房分别按照表 2 给出的要求各安排 20 名学生各自随机进行不同的行为。

表 2 实验说明

序号	时间段	机房 1 学生行为	机房 2 学生行为
1	10：30~11：30	普通访问	10 人突发访问文件服务器，其余普通访问
2	11：30~12：30	3 人使用 P2P，其余普通访问	15 人突发访问文件服务器，其余普通访问
3	12：30~13：30	使用 P2P 增至 8 人，其余普通访问	16 人 LDDoS 攻击，其余普通访问
4	13：30~14：30	使用 P2P 增至 15 人，其余普通访问	12 人使用 P2P，其余普通访问

实验中安排学生进行的用户行为包括以下 4 类。

1) 普通访问，由学生自行选择目标，进行 Web 浏览、邮件以及即时通信等应用，在用户数量较少的情况下(2 个机房的学生总数不超过 40 人)对网络状态影响不大。

2) 突发访问文件服务器，即若干学生同时访问机房内部 FTP 服务器进行文件下载，由于 FTP 文件服务器的处理速度以及带宽限制，当有较多用户同时突发访问文件服务器时会导致网络以及服务器负载加重。

3) P2P 行为，学生自行选择应用软件及目标，进行基于 P2P 的文件下载及在线视频浏览等，由于 2 个机房的网关以及 Internet 接入采取共享带宽形式，因此甚至少数用户的 P2P 下载行为也会占据较多的网络资源而影响网络性能。

4) 攻击行为，学生自行组合并随机发起攻击(LDDoS)，该行为会严重影响网络的性能。

借鉴文献[16]的思路，根据实验的实际情况，本文选取了适合在实验环境中获取的 17 种用户行为指标作为用户行为特征 C_{raw}，如表 3 所示。根据文献[18]的研究结论，这些特征能覆盖大部分常见的应用识别所需要的特征。

表 3 用户行为特征集

特征类别	特征项
报文特征	平均报文长度、报文间隔时间、数据发送分组率、最大报文长度、最小报文长度
流特征	源端口数、目的端口数、流数、平均流长度、最大流长度、最小流长度
应用层特征	上传流量、下载流量、TCP 上传流量、TCP 下载流量、UDP 上传流量、UDP 下载流量

网络状态则通过采集网络节点的性能数据(状态因子为带宽使用率、CPU 利用率以及转发队列占

缓冲区比例，出于方便，权重则分别设为 0.33、0.33 和 0.34)，并按式(1)计算。

实验软硬件平台如下：硬件配置为 Intel Core2 Quad 2.3 GHz，4 GB 内存，操作系统为 Windows XP SP3。数据库采取 SQLServer 2005，软件环境为：使用粗糙集工具 Rosetta 进行用户行为与网络状态数据分析挖掘，使用 MATLAB 软件进行用户行为评估及评估结果分析。

4.2　实验结果分析

4.2.1　有效性分析

实验采集了整个实验的全部用户流量以及各网络节点的状态数据，在其基础上使用基于本文模型实现的评估系统共获得了大约 3 万条用户行为与网络状态的原始数据，对原始数据进行预处理后分别得到 4 个时间段中的用户行为和网络状态的样本，运行图 2 中的算法，可得每个时间段所对应的用户行为评估指标与权重，具体如表 4 所示。

表 4　评估指标与权重说明

时间段	评估指标	权重
1	TCP 下载流量	0.63
	报文间隔时间	0.37
2	TCP 下载流量	0.45
	报文间隔时间	0.28
	UDP 下载流量	0.27
3	平均报文长度	0.2
	数据分组发送率	0.32
	UDP 上传流量	0.18
	UDP 下载流量	0.3
4	目的端口数	0.26
	UDP 上传流量	0.28
	UDP 下载流量	0.46

在时间段 1 中，机房 1 安排学生进行 Web 浏览为主的普通访问行为，机房 2 则安排 10 名学生访问文件服务器下载文档(其余普通访问)，网络中出现用户较高频度的 FTP 突发访问并造成网络状态的波动，网络节点与服务器的带宽利用率和 CPU 占用率都随着 8 名学生的访问而出现了变化，显然该网络状态的变化与用户突发访问服务器行为有很大的关联，而 FTP 行为较普通行为的突出特征包括下载流量（以 TCP 为主）较大、报文时间间隔较短等，与本文评估方法获得的评估指标一致。

在时间段 2 中，机房 1 安排 3 名学生进行 P2P 行为，机房 2 访问文件服务器的学生数量增加到 15，其余学生仍进行普通行为。由于实验环境中带宽是共享的，P2P 行为当出现以及访问服务器学生数量的增加都进一步加重了网络负荷，除带宽利用率和 CPU 利用率外，转发队列占缓冲区比例也开始进一步上升，此时间段的网络状态变化与用户的突发访问服务器行为都存在关联，而与普通行为和 FTP 下载相比，P2P 行为的特征包括较大的流量(以 UDP 上传和下载为主)、端口数等，本文方法产生的评估指标中则出现了 UDP 下载流量这一体现 P2P 行为的指标，与实际一致。

在时间段 3 中，机房 1 中的 P2P 用户数量增加为 8 名，而机房 2 安排了 16 名学生进行 LDDoS 攻击行为(其余学生进行普通行为)。P2P 行为增多加重了网络负担，LDDoS 则造成网络出现拥塞，使得 CPU 利用率上升，带宽利用率与转发队列占缓冲区比例出现大幅波动。更是极大加重了网络负担(访问文件服务器的行为则停止了)，LDDoS 行为的特征主要在周期性的高发送分组率以及短报文间隔时间，而 P2P 行为特征则在网络的流量方面。本文方法所得到的评估指标能覆盖两者的特征，而 TCP 下载量等不再成为评估指标，与实际一致。

在时间段 4 中，2 个机房的学生仅安排较多学生(机房 1 有 15 人，机房 2 有 12 人)进行 P2P 行为，其余为普通行为。由于机房中接入 Internet 的路由器带宽有限，因此较多的 P2P 行为使得网络陆续出现高负荷甚至拥塞现象，CPU 占用率提高，带宽利用率与转发队列占缓冲区比例出现波动，而本文方法所得出的评估指标则相应体现 P2P 行为特征(如端口数、UDP 的吞吐量等)，与实际一致。

由以上 4 个时间段的分析可以看出，本文的评估指标与权重设置算法能够准确地反映用户行为与网络状态变化的关联性。

获取如表 4 所示的评估指标与权重后，可按 3.3 节的评估步骤，对实验的 4 个时间段内用户行为进行评估，量化分析其对网络状态变化的影响程度，为验证评估方法所获取的用户行为评估值是否有效地反映了不同用户行为对网络状态的影响力，本文根据实验的安排，分别考察每个时间段内不同行为类别用户(分为普通用户和目标用户 2 类)所获取评估值的分布及其差异情况，其中，时间段 1 中的目标用户为进行突发访问的 10 名学生，时间段 2

中3名P2P行为和15名突发访问的学生构成目标用户，时间段3中8名P2P行为和16名攻击行为学生为目标用户，时间段4中27名P2P行为学生为目标用户，其余学生则视为普通用户。另外设置对照方法，将用户行为特征全集作为评估指标并平均分配权重(即不分析用户行为与网络状态变化的关联性)。评估值的分布情况使用均值(算术平均值)、中位数(数据从小到大排列后列中间位置的数据)及标准差3项指标进行考察，其中，均值与中位数可反映一组用户评估值的一般情况，而标准差则可揭示这组用户评估值的离散程度，分析结果如表5所示。

表5 评估结果分析及比较

时间段	评估方法	用户类型	分析指标		
			均值	中位数	标准差
1	本文方法	普通用户	0.12	0.10	0.12
		目标用户	0.79	0.90	0.29
		全体用户	0.28	0.12	0.34
	对照方法	普通用户	0.50	0.51	0.07
		目标用户	0.54	0.57	0.09
		全体用户	0.51	0.51	0.08
2	本文方法	普通用户	0.29	0.26	0.19
		目标用户	0.66	0.77	0.25
		全体用户	0.45	0.28	0.28
	对照方法	普通用户	0.51	0.51	0.09
		目标用户	0.53	0.54	0.07
		全体用户	0.52	0.52	0.08
3	本文方法	普通用户	0.24	0.25	0.06
		目标用户	0.49	0.50	0.11
		全体用户	0.39	0.46	0.15
	对照方法	普通用户	0.47	0.49	0.07
		目标用户	0.54	0.56	0.09
		全体用户	0.49	0.49	0.08
4	本文方法	普通用户	0.32	0.22	0.25
		目标用户	0.64	0.73	0.19
		全体用户	0.54	0.66	0.26
	对照方法	普通用户	0.45	0.44	0.06
		目标用户	0.50	0.49	0.04
		全体用户	0.46	0.46	0.06

从表5可以看出，在4个时间段内，目标用户与普通用户使用本文评估方法所得的评估值在均值和中位数上都有明显差异且目标用户的评估值更高(例如4个时间段目标用户均值较普通用户均值分别要高0.67、0.37、0.25和0.32)，另外，目标用户和普通用户的均值、中位数与全部用户相比也有明显差异，而对照方法所得的评估值均值和中位数则非常接近（差距为0.04、0.02、0.07和0.05），与全体用户相比也没有太大差距，说明本文方法产生的评估值能有效区分不同用户行为对网络状态变化的影响程度，目标用户的影响力明显高于普通用户。再考察标准差，目标用户和普通用户使用本文评估方法所得的评估值标准差都低于全部用户的标准差，说明就分布而言，目标用户与普通用户各自的内聚性要高于全体用户，即从整体来看，用户评估值明显被分为2个不同的“小团体”，因此在事先未知用户行为类别的情况下，本文方法对用户行为的评估结果有利于使用聚类方法来发现影响网络运行的用户，而对照方法所得的目标用户、普通用户以及全体用户评估值的标准差则十分接近，说明在分布上所有用户的评估值内聚性很高，普通用户与目标用户的评估结果较为接近且互相混同，在事先未知用户行为类别的情况下，难以使用聚类方法对其进行准确区分。

根据以上分析可看出，与不进行关联分析的对照方法相比，由于本文评估方法所使用的评估指标与权重能准确地反映用户行为与网络状态变化的关联性，因此评估值能有效量化用户行为对网络状态变化的影响力，并为发现影响网络运行的用户提供有效支持。

4.2.2 准确性分析

基于本文方法的评估值可区分影响网络运行的目标用户。本文使用层次聚类AGENES算法进行区分，将用户评估值划分为2个簇，取评估值高的簇为目标用户，然后将区分结果与实验安排情况比较，使用如下指标衡量区分的准确性，设实验安排的目标用户数量为N，普通用户数量为T，在每个时间段内根据评估值区分用户，设该时间段内被识别为目标用户的数量为TN，而其中实际为目标用户的数量设为N'，评价指标如下。

1) 准确率(precision)：$P=\frac{N'}{TN}$，即被识别为目标用户中真实的目标用户所占的比例。

2) 漏报率(false negative)：$FN=\frac{N-N'}{N}$，即目标用户中未被识别出的比例。

3) **误报率**(false positive)：$FP = \dfrac{TN - N'}{T}$，普通用户中被误识别为目标用户的比例。

基于 4 个时间段的用户行为数据评估值，使用聚类方法对用户进行区分，对照实验预设的安排，区分结果如表 6 所示。

表 6　区分结果

时间段	目标用户/人	识别结果/人	正确被识别/人	误被识别/人
1	10	10	9	1
2	18	16	14	2
3	24	23	22	1
4	27	28	25	3

区分结果准确率、漏报率和误报率如表 7 所示。

表 7　区分结果的准确率、漏报率和误报率

时间段	准确率/%	漏报率/%	误报率/%
1	90.0	10.0	3.3
2	87.5	6.7	8.0
3	95.7	8.3	6.3
4	89.3	7.4	23.1

从表 7 可知，依靠评估方法能够有效地区分出目标用户，正确率在实验中最高超过 90%，最低也能达 87%左右，由于评估方法主要为发现造成网络状态变化的用户提供支持，在其基础上还可进一步采取应用识别等方法来辨识和细分用户，因此，漏报率指标也很重要，实验中区分结果的漏报率最低能达到 6%，最高也可在 10%及以下。

比较本文评估与传统评估方法在准确性方面的差别，本文采样基于 AHP 层次分析法和文献[11]的基于信息熵的评估方法作为对照，使用用户行为特征全集作为对照方法的评估指标，同样采用以上 4 个时间段的数据进行用户行为评估，并根据评估值聚类结果对用户进行区分，结果如图 4~图 6 所示。

从比较结果可知，本文的评估方法在区分造成网络状态变化目标用户的准确率、漏报率和误报率大都较 AHP 层次分析法和基于熵的评估方法要好，原因在于本文方法可以通过对用户行为和网络状态数据的分析动态构建并调整评估指标和权重，从而更准确地衡量用户行为的影响力，而 AHP 层次分析法由于评估指标和权重无法随用户行为变化而动态调整，其有效性最差，基于熵的评估方法其权重能动态调整，因此有效性要比 AHP 层次分析法要高，但其指标仍是固定的，因此有效性较本文评估方法要差。

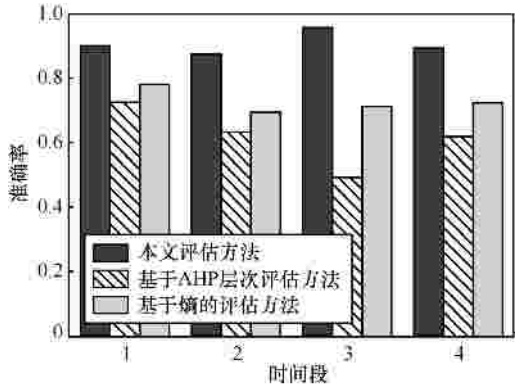

图 4　准确率比较结果

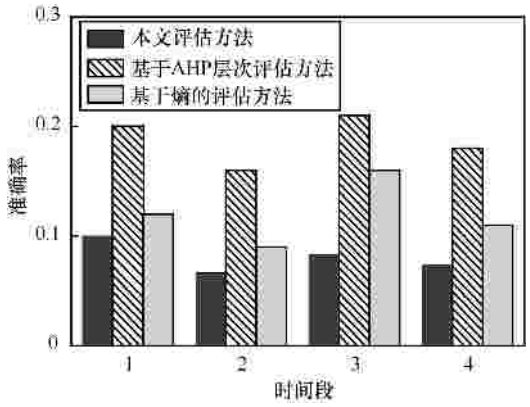

图 5　漏报率比较结果

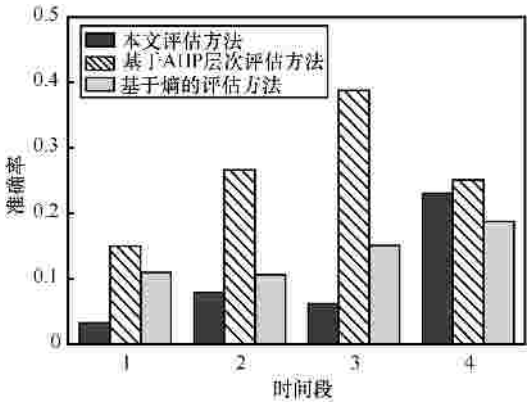

图 6　误报率比较结果

5　结束语

用户日趋扩大的规模和复杂的行为对网络正常运行造成了巨大影响，评估用户行为对网络状态变化的影响程度，能为网络状态发生非预期变化时发现相应用户及进行控制决策提供基础，进而有助

于实现用户行为的可控性。传统的用户行为评估方法存在评估指标和权重设置的主观性、静态性弊端，影响了评估用户行为网络影响力的准确性，鉴于此，本文引入粗糙集理论，利用其属性约简和属性重要度方法实现了自适应的指标选择与权重设置及调整，避免了传统方法的缺陷，因此能准确地量化评估用户行为对网络状态变化影响力，在评估用户行为基础上，下一步工作主要是结合行为识别、用户聚类等方法，进一步研究用户行为的控制机制和方法，从而实现用户行为的可预测、可控制。

参考文献：

[1] 罗军舟，韩志耕，王良民. 一种可信可控的网络体系及协议结构[J]. 计算机学报, 2009, 3(3): 391-404.
LUO J Z, HAN Z G, WANG L M. Trustworthy and controllable network architecture and protocol framework[J]. Chinese Journal of Computer, 2009, 3(3): 391-404.

[2] 林闯，雷蕾. 下一代互连网络体系结构研究[J].计算机学报, 2007, 30(5):693-711.
LIN C, LEI L. Research on next generation internet arc itecture[J]. Chinese Journal of Computers, 2007, 30(5):693-711.

[3] FUNG C, ZHANG J, AIB I, *et al*. Trust management and admission control for host-based collaborative intrusion detection[J]. Journal of Network and Systems Management, 2011, 19(2):257-277.

[4] TAJEDDINE A, KAYSSI A, CHEHAB A, *et al*. Fuzzy reputation-based trust model[J]. Applied Soft Computing, 2011, 11(1):345-355.

[5] GOVINDAN K. Trust computations and trust dynamics in mobile ad hoc networks: a survey[J]. Communications Surveys & Tu ials, 2011, 14(2):279-298.

[6] RAYA M, PAPADIMITRATOS P, GLIGOR V D, *et al*. On data-centric trust establishment in ephemeral ad hoc networks[A]. INFOCOM[C]. 2008. 13-18.

[7] XI Z Y, CHEN H, WANG X Z, *et al*. Evaluation model for computer network information security based on analytic hierarchy process[A]. Intelligent Information Technology Application[C]. 2009.

[8] CAO Y, ZHANG L, WU H. An evaluation system for network attack effect based on fuzzy[A]. E-Business and Information System Security[C]. 2009.

[9] JIN F, LIU P, ZHANG X, *et al*. The evaluation study of knowledge management performance based on grey-ahp method[A]. Software Engineering, Artificial Intelligence, Networking, and allel/Distributed Computing[C]. 2007.

[10] 张义荣，鲜明，王国玉. 一种基于网络熵的计算机网络攻击效果定量评估方法[J]. 通信学报, 2004, 25(11):158-165.
ZHANG Y R, XIAN M, WANG G Y. A quantitative evaluation technique of attack effect of computer network based on network entropy[J]. Journal on Communications, 2004, 25(11):158-165.

[11] WANG X, HE H, ZHANG H L, *et al*. Dynamic damage evaluation of network availability adjusted by index correlation[A]. Internet Computing for Science and Engineering (ICICSE)[C]. 2009.

[12] 李小勇，桂小林，毛倩等. 基于行为监控的自适应动态信任度测模型[J]. 计算机学报, 2009, 32(4):664-674.
LI X Y, GUI X L, MAO Q, *et al*. Adaptive dynamic trust measurement and prediction model based on behavior monitoring[J]. inese Journal of Computers, 2009, 32(4):664-674.

[13] 梁吉业，李德玉. 信息系统中的不确定性与知识获取[M] .北京: 科学出版社, 2005.
LIANG J Y, LI D Y. The Uncertainty and Knowledge Acqui ing in Imformation Systems[M]. Beijing: Science Press, 2005.

[14] 龚正虎,卓莹. 网络态势感知研究[J]. 软件学报, 2010, 21(7): 1605-1619.
GONG Z H, ZHUO Y. Research on cyberspace situational awareness[J]. Journal of Software, 2010, 21(7):1605-1619.

[15] 江勇，林闯，吴建平. 网络传输控制的综合性能评价标准[J]. 计算机学报, 2002, 25(8):869-877.
JIANG Y, LIN C, WU J P. Integrated performance evaluat n criteriafor network traffic control[J]. Chinese Journa of Computers, 2002, 25(8):869-877.

[16] MOORE A W, ZUEV D. Discriminators for Use in Flow-Based Classification[R]. Technical Report IRC-TR-04-028, 2004.

[17] LIN Z, FENG J. A rough set approach to feature selection based on relative decision entropy[J]. Lecture Notes in Computer Science, 2011, 6954(2011):110-119.

[18] KIM H, CLAFFY K, FOMENKOV M, *et al*. Internet traffic classification demystified: myths, caveats, and the best practices[A]. Proc of ACM CoNEXT'08[C]. New York, USA, 2008.1-12.

作者简介：

陆悠（1977-），男，江苏苏州人，苏州科技学院讲师，东南大学博士生，主要研究方向为下一代网络体系结构、用户行为控制等。

罗军舟（1960-），男，浙江宁波人，博士，东南大学教授、博士生导师，主要研究方向为下一代网络体系结构、协议工程、云计算和服务计算。

李伟（1978-），男，河南许昌人，博士，东南大学副教授，主要研究方向为下一代网络体系结构、服务计算和网络管理。

于枫（1974-），女，山东济南人，东南大学博士生、讲师，主要研究方向为下一代网络体系结构、网络管理。

夏怒（1981-），男，江苏南京人，东南大学博士生，主要研究方向为下一代网络体系结构、网络管理。