基于行为特征分析的社交网络女巫节点检测机制

吴大鹏 司书山* 闫俊杰 王汝言
(重庆邮电大学通信与信息工程学院 重庆 400065)
(重庆高校市级光通信与网络重点实验室 重庆 400065)

摘 要：通过制造大量非法虚假身份，女巫攻击者可以提高自身在社交网络中的影响力，影响网络中社交个体中继选择意愿，窃取社交个体隐私，对其利益造成严重威胁。在对女巫节点行为特征分析的基础上，该文提出一种适用于社交网络的女巫节点检测机制，通过节点间静态相似度和动态相似度评估节点影响力，并筛选可疑节点，进而观察可疑节点的异常行为，利用隐形马尔科夫模型推测女巫节点通过伪装所隐藏的真实身份，更加精确地检测女巫节点。分析结果表明，所提机制能有效提高女巫节点的识别率，降低误检率，更好地保护社交个体的隐私和利益。

关键词：社交网络；女巫节点检测；行为特征；隐形马尔科夫模型
中图分类号：TP393 **文献标识码**：A **文章编号**：1009-5896(2017)09-2089-08
DOI: 10.11999/JEIT170246

Behaviors Analysis Based Sybil Detection in Social Networks

WU Dapeng SI Shushan YAN Junjie WANG Ruyan
(*School of Telecommunication and Information Engineering, Chongqing University of Posts and Telecommunications, Chongqing* 400065, *China*)
(*Optical Communication and Network Key Laboratory of Chongqing, Chongqing* 400065, *China*)

Abstract: Sybil attackers can improve their own influence in social networks by creating a large number of illegal illusive identities then affect the social individuals' choice of relays and steal individuals' privacy, which seriously threatens the interests of social individuals. Based on the analysis of the Sybil's behaviors, a Sybil detection mechanism applied to social networks is proposed in this paper. The influence of nodes is calculated according to static similarity and dynamic similarity and then selecting the suspicious nodes based on the influence. Next, using the Hidden Markov Model (HMM) to infer the true identity of suspicious nodes by observing their abnormal behaviors, thus detecting the Sybil more precisely. Analysis results show that the proposed mechanism can effectively improve the recognition rate and reduce the false detection rate of the Sybil and thereby protecting the privacy and interests of social individuals better.

Key words: Social networks; Sybil detection; Behavior characteristics; Hidden Markov Model (HMM)

1 引言

社交网络，即社交网络服务(Social Network Service, SNS)的出现为信息的传播、人们彼此间的交流提供了便利[1,2]。然而，社交网络社交环境及信息的开放性，社交方式的多样化以及信息存储集中等特点[3]也给恶意节点发起攻击，获取非法利益提供了契机，对社交个体的个人利益与隐私造成严重威胁[4]。

在诸多攻击手段中，女巫攻击对系统的影响较为严重[5]，女巫攻击者制造大量非法虚假身份即女巫节点，女巫节点通过与其他节点频繁交互提升自身在整个网络中的影响力，从而更方便地在社交网络中发起欺骗社交个体，干扰社交个体中继选择，拦截和篡改数据等攻击行为[6]。可见，在数据转发过程中有效地识别女巫节点对维护社交个体利益及系统安全至关重要。

社交网络中女巫节点的检测方法可分为两类，一类是基于网络拓扑结构的女巫节点检测机制，文献[7]中提出的面向群体融合的女巫节点检测方案

收稿日期：2017-03-29；改回日期：2017-07-20；网络出版：2017-08-11
*通信作者：司书山 1063603919@qq.com
基金项目：国家自然科学基金(61371097)，重庆高校创新团队建设计划(CXTDX201601020)
Foundation Items: The National Natural Science Foundation of China (61371097), The Program for Innovation Team Building at Institutions of Higher Education in Chongqing (CXTDX201601020)

SybilGuard，女巫节点与正常节点交互异常，不具备群体融合的特点，整个网络可以很明显地划分为正常节点与女巫节点两个群体。在 SybilGuard 的基础上，文献[8]对利用群体融合假设划分出的女巫群体进一步处理，只剔除其中影响力较高的节点。另外一类是基于用户行为分析的女巫节点检测方法，Krishnamurthy 等人[9]以粉丝数与关注数的比值为指标对 Twitter 恶意用户进行检测，但检测指标过于单一；为此，Chu 等人[10]以多个属性为依据借助朴素贝叶斯模型对 Twitter 恶意用户进行检测；Tan 等人[11]收集用户多个属性信息，运用机器学习分析多个属性之间的相关性，并以此为依据利用 KNN 分类算法检测恶意用户群体。这些方法在检测女巫节点时，只考虑女巫节点作为恶意节点发起攻击时破坏数据传输的共性，缺乏针对女巫节点行为特征的分析，容易导致误判。另外，上述机制的准确性受女巫节点攻击次数的影响较大。

针对上述问题，本文提出了一种基于节点行为特征分析的社交网络女巫节点检测机制(Behaviors Analysis-based Sybil Detection, BASD)，对女巫节点行为特征进行分析，从女巫节点高相似度伪装正常节点、频繁切换假名和异常协作的特征出发，首先，评估节点影响力，筛选可疑节点；然后，观察可疑节点的异常交互行为，利用隐形马尔科夫模型推测可疑节点隐藏的真实身份，更加精确地检测女巫节点。

2 网络结构与女巫节点行为模型

2.1 网络结构

根据社交个体行为特征，社交网络中社交个体主要分为两类：正常社交个体和女巫攻击者。

正常社交个体：智能终端、便携式通信设备所有者，可以与其他社交个体进行双向通信，社交个体需向管理中心注册身份标识，每个正常的社交个体可申请多个合法身份标识，本文将这些身份标识看作节点。

女巫攻击者：其每个设备有多个非法虚假身份来掩饰自己的真实身份，这些虚假身份即为女巫节点。

此外，社交网络需要部署管理中心并设置跟踪节点。

管理中心：与节点直接通信并收集节点所携带的信息。其功能如下：及时收集节点反馈的信息，配置节点社会属性文件，更新节点影响力，并分析可疑节点是女巫节点的概率。

跟踪节点：由管理中心指定的特殊节点用于监测可疑节点异常交互行为。

每个节点随身携带自己的社会属性文件，管理中心定期更新其控制范围内节点的社会属性文件。社会属性文件主要包括节点的身份属性信息、节点与其好友的相似度、节点异常行为信息以及节点的状态属性标签。其中，节点状态属性标签包含 3 种节点状态：“正常节点”、“可疑节点”、“女巫节点”。

网络运行过程如下所述：在当前时刻，节点 B, C, D 均为节点 A 的好友，当节点 A 试图与节点 B 交互时，首先向管理中心发送查询指令，管理中心反馈节点 B 最新的状态属性标签，若 B 当前时刻为可疑节点或女巫节点，则 A 不与其交互，管理中心指派跟踪节点继续对节点 B 跟踪监测。若节点 B 当前时刻为正常节点，则 A 可以与其交互并交换社会属性文件，交互后更新节点 B 社会属性文件中的动态相似度信息，并计算当前时刻节点 B 对 A 的影响力，根据影响力更新节点 B 的状态属性标签，若节点 B 的状态属性标签更新为可疑节点，则由管理中心指派跟踪节点进一步观察其社交行为，判断其是否存在异常假名变换行为和异常协作行为，管理中心根据跟踪节点反馈的信息，利用隐形马尔科夫模型推测其是否为女巫节点。

2.2 女巫节点行为模型

女巫攻击者制造大量女巫节点，女巫节点通过相互协作伪装成正常节点发起攻击，向目标节点重复发送其他节点的诽谤信息以影响社交个体的思维偏好和中继选择意愿[12]。假如节点 A 想要共享信息给节点 E，但与节点 E 不是好友，则节点 A 将其共享信息的需求广播给其好友，好友在收到广播后会根据实际情况反馈，而女巫攻击者则与其所控制的女巫节点合谋反馈虚假信息以获得节点 A 的信任，从而影响控制节点 A 的社交个体的中继选择意愿，最终使节点 A 选择女巫节点作为下一跳中继节点。针对上述问题，本文在网络中部署了管理中心，节点 A 可以向管理中心查询其好友节点的状态信息，判断其好友是否为女巫节点。

3 可疑节点筛选

跟踪节点负责检测可疑节点异常行为，并将结果报告给管理中心。但在实际应用中，社交网络资源很有限[13]，虽然令跟踪节点对每个节点都进行跟踪和监测能够显著提升检测准确度，但会增加大量不必要的资源消耗。通常，对节点影响力较高的好友，其发表的内容更容易影响控制该节点的社交个体的选择意愿和偏好，但这类好友既可能是与节点关系密切的好友，也可能是女巫节点，一旦这些女

巫节点对目标节点发起攻击，所产生的影响也是巨大的。因此，本文在对节点好友进行女巫节点检测之前，首先根据节点好友对自身的影响力筛选好友节点组成可疑节点集合，然后对可疑节点进一步监测，确定其是否为女巫节点。

3.1 节点影响力评估

节点相似度反映两个节点身份信息、兴趣爱好以及行为特征之间的相似程度[14]。本文通过计算节点与其好友之间的相似度来评估好友对节点的影响力。节点之间的相似度包含静态相似度和动态相似度。静态相似度描述节点固有属性之间的相似程度，不受节点行为影响，短时间内不变；动态相似度描述节点行为属性上的相关性，通过分析节点之间交互行为特征评估节点间动态相似度。

3.1.1 静态相似度评估　节点在社交网络中交互时普遍采用年龄、住址、职业等身份属性来实现区分[15]，女巫节点在试图对节点发起攻击时，通常会为自己贴上与目标节点相同或相关的标签以达到伪装自己，降低社交个体警惕性的目的，本文在计算节点静态相似度时考虑节点身份属性的相似程度。通过计算节点给定属性对应标签的方差评估节点身份属性差异度，进而评估节点身份属性相似度。节点身份属性差异越小，相似度越高。

假设每个节点都拥有 n 个属性，节点 A 与其好友 N_k 第 n 个属性可分别表示为 $\mathrm{a}_n^{\mathrm{A}}$ 和 $\mathrm{a}_n^{\mathrm{N}_k}$，则节点 A 和 N_k 的身份属性标签列表分别为 $\mathrm{user(A)}=\left\{\mathrm{a}_1^{\mathrm{A}},\mathrm{a}_2^{\mathrm{A}},\cdots,\mathrm{a}_n^{\mathrm{A}}\right\}$ 和 $\mathrm{user}\left(\mathrm{N}_k\right)=\left\{\mathrm{a}_1^{\mathrm{N}_k},\mathrm{a}_2^{\mathrm{N}_k},\cdots,\mathrm{a}_n^{\mathrm{N}_k}\right\}$。$\mathrm{a}_n^{\mathrm{A}}$ 为目标节点身份属性标签，其取值设置为 1，则 $\mathrm{user(A)}=\{1,1,\cdots,1\}$，对于 $\mathrm{a}_n^{\mathrm{N}_k}$，若 N_k 的第 n 个属性与 A 相同，则 $\mathrm{a}_n^{\mathrm{N}_k}=1$，若不同，则 $\mathrm{a}_n^{\mathrm{N}_k}$ 为节点 N_k 好友中第 n 个属性与 A 相同的节点数量与 N_k 好友数量的比值。

节点 A 和好友 N_k 之间的身份属性差异度如式(1)所示。

$$\mathrm{IDSim}_{\mathrm{A,N}_k}=\left[\sum_{i=1}^{n}\left(\mathrm{a}_i^{\mathrm{A}}-\mathrm{a}_i^{\mathrm{N}_k}\right)^2\right]\Bigg/\left(\sum_{i=1}^{n}\left|\mathrm{a}_i^{\mathrm{A}}-\mathrm{a}_i^{\mathrm{N}_k}\right|\right)^2 \tag{1}$$

其中，$\left(\sum_{i=1}^{n}\left|\mathrm{a}_i^{\mathrm{A}}-\mathrm{a}_i^{\mathrm{N}_k}\right|\right)^2\geq\sum_{i=1}^{n}\left(\mathrm{a}_i^{\mathrm{A}}-\mathrm{a}_i^{\mathrm{N}_k}\right)^2$，所以 $0\leq\mathrm{IDSim}_{\mathrm{A,N}_k}\leq 1$。

通常情况下，当两个节点拥有较大数量的共同好友时，控制这两个节点的社交个体往往关系密切。但节点间共同好友数量受节点自身好友数量影响，与其他节点共同好友数量占自身好友数量的比重更能反映其他节点对自身的影响力大小，比重越大，影响力越大。女巫攻击者主动发起攻击时，为了加大攻击强度，会利用其控制的多个女巫节点组成攻击协作团体对同一目标节点发起攻击，女巫节点会同时主动关注目标节点，成为其粉丝，即好友。同时，同一协作团体内的女巫节点之间会互加好友，以便协作和伪装，因而女巫节点与目标节点共同好友数量占目标节点自身好友数量比重通常也高于普通节点。本文用节点间共同好友占目标节点自身好友的比例计算好友与目标节点的好友相似度用以评估好友对目标节点的影响力。节点 A 的第 k 个好友表示为 N_k，N_k 的第 i 个好友为 M_i^k，则节点 A 与其好友 N_k 的好友相似度 $\mathrm{NeiSim}_{\mathrm{A,N}_k}$ 为

$$\mathrm{NeiSim}_{\mathrm{A,N}_k}=\frac{\left|\left\{\mathrm{N}_1,\mathrm{N}_2,\cdots,\mathrm{N}_k\right\}\bigcap\left\{\mathrm{M}_1^k,\mathrm{M}_2^k,\cdots,\mathrm{M}_i^k\right\}\right|}{\left|\left\{\mathrm{N}_1,\mathrm{N}_2,\cdots,\mathrm{N}_k\right\}\right|} \tag{2}$$

其中，$|\{*\}|$ 表示集合 $\{*\}$ 中元素数量。

当节点间的好友相似度增加，节点身份属性差异度减小时，其静态相似度相应增加，且增加到一定程度后渐趋稳定，采用式(3)描述两节点的静态相似度与身份属性差异度以及好友相似度的关系：

$$\mathrm{StaSim}_{\mathrm{A,N}_k}=\frac{1}{2}+\frac{2}{\pi}\arctan\left(\mathrm{NeiSim}_{\mathrm{A,N}_k}-\mathrm{IDSim}_{\mathrm{A,N}_k}\right) \tag{3}$$

3.1.2 动态相似度评估　动态相似度由节点间的交互行为特征决定，两节点在一段时间内的交互次数和交互持续时间可以反映两节点之间的交互亲密度。在一段时间内交互亲密的两个节点对应的社交个体往往关系密切[16]，因此在计算节点动态相似度时需要考虑节点之间的交互亲密度；同时，还应考虑节点之间交互发生的时间，时间相隔越久的历史信息准确度越低，其参考价值也越小。

对于时间窗口 T，节点的交互状态仅包括交互持续时间和交互间隔。可用节点交互平均间隔时间占 T 的比例表示两节点的交互亲密度[17]，在时间段 T 内，节点 A 与其好友 N_k 第 i 次交互过程的起始时刻和终止时刻分别用 ST_i 和 ET_i 表示，则两节点时间段 T 第 i 与 $i+1$ 次交互过程间隔 Δt_i 可表示为 $\mathrm{ST}_{i+1}-\mathrm{ET}_i$，进而，节点 A 与其好友 N_k 的当前 T 时间段内的交互亲密度参数 $\mathrm{CF}_{\mathrm{A,N}_k}^{T}$ 如(4)式所示。

$$\mathrm{CF}_{\mathrm{A,N}_k}^{T}=\begin{cases}T_{\mathrm{inter\_avg}}/T=\left[\sum_{i=1}^{n}\left(\mathrm{ST}_{i+1}-\mathrm{ET}_i\right)\right]\Big/n\cdot T, & n\neq 0\\ \mathrm{CF}_{\mathrm{A,N}_k}^{T^*}, & n=0\end{cases} \tag{4}$$

其中，n 表示两节点在 T 时间段内的交互次数，$T_{\mathrm{inter\_avg}}$ 为 T 时间段内两节点的平均交互间隔，$\mathrm{CF}_{\mathrm{A,N}_k}^{T^*}$ 为上一个时间段 T 内节点 A 和 N_k 的交互亲密度参数。$\mathrm{CF}_{\mathrm{A,N}_k}^{T}$ 越小，节点交互越亲密。

对于节点交互发生的时间，间隔越长的历史信

息准确度越低，参考价值越小，本文采用指数衰减机制处理节点间历史交互过程，综合评估各交互过程对反映当前时刻控制两节点的社交个体间关系的重要程度：

$$\mathrm{CI}_{\mathrm{A},\mathrm{N}_k}^{T}=\begin{cases}\sum_{i=1}^{n}\exp\left(-\dfrac{\left(T-\mathrm{ST}_i\right)}{T}\right)\cdot\dfrac{T}{\mathrm{ST}_{i+1}-\mathrm{ET}_i}, & n\neq 0\\ \mathrm{CI}_{\mathrm{A},\mathrm{N}_k}^{T^*}, & n=0\end{cases}\tag{5}$$

$\mathrm{CI}_{\mathrm{A},\mathrm{N}_k}^{T^*}$ 为上一个时间段 T 内的 $\mathrm{CI}_{\mathrm{A},\mathrm{N}_k}^{T}$ 值。$\mathrm{CI}_{\mathrm{A},\mathrm{N}_k}^{T}$ 越大，说明该好友与节点在这个时间段 T 内联系越密切。

如上所述，节点动态相似度由节点交互亲密度和交互过程 $\mathrm{CI}_{\mathrm{A},\mathrm{N}_k}^{T}$ 值共同决定，节点间的动态相似度随节点交互过程 $\mathrm{CI}_{\mathrm{A},\mathrm{N}_k}^{T}$ 值的增加以及交互亲密度参数 $\mathrm{CF}_{\mathrm{A},\mathrm{N}_k}^{T}$ 的减小而增加，并逐渐趋于稳定。根据上述变化趋势，节点动态相似度的评估方式如式(6)所示。

$$\mathrm{DynSim}_{\mathrm{A},\mathrm{N}_k}=\frac{2}{1+\exp\left(-\mathrm{CI}_{\mathrm{A},\mathrm{N}_k}^{T}/\mathrm{CF}_{\mathrm{A},\mathrm{N}_k}^{T}\right)}-1\tag{6}$$

根据节点 A 与其好友 N_k 的动态相似度和静态相似度的评估结果 $\xi_{\mathrm{A},\mathrm{N}_k}$，可获知节点好友 N_k 对节点的影响力。$\xi_{\mathrm{A},\mathrm{N}_k}$ 的计算方式如(7)所示。

$$\xi_{\mathrm{A},\mathrm{N}_k}=(1-\eta)\cdot\mathrm{StaSim}_{\mathrm{A},\mathrm{N}_k}+\eta\cdot\mathrm{DynSim}_{\mathrm{A},\mathrm{N}_k}\tag{7}$$

其中，η 为动态权值，可用节点交互亲密度参数 $\mathrm{CF}_{\mathrm{A},\mathrm{N}_k}^{T}$ 确定 η，$\mathrm{CF}_{\mathrm{A},\mathrm{N}_k}^{T}$ 越小，节点交互越亲密，η 越大，$\xi_{\mathrm{A},\mathrm{N}_k}$ 的评估更依赖于节点间动态相似度，η 为

$$\eta=\frac{\left(1-\mathrm{CF}_{\mathrm{A},\mathrm{N}_k}^{T}\right)}{\mathrm{StaSim}_{\mathrm{A},\mathrm{N}_k}+\left(1-\mathrm{CF}_{\mathrm{A},\mathrm{N}_k}^{T}\right)}\tag{8}$$

节点的历史影响力也在一定程度上反映两节点对应的社交个体在当前时刻的关系。所以，综合历史影响力以及 $\xi_{\mathrm{A},\mathrm{N}_k}$ 得到好友在当前时间段对该节点的影响力 $\delta_{\mathrm{A},\mathrm{N}_k}$，如(9)式所示。$f^{-1}\left(\delta_{\mathrm{A},\mathrm{N}_k}^{*}\right)$ 包含之前所有 T 时间段内的 $\xi_{\mathrm{A},\mathrm{N}_k}$ 信息，$\delta_{\mathrm{A},\mathrm{N}_k}^{*}$ 为上一时间段 T 好友 N_k 对节点 A 的影响力。

$$\delta_{\mathrm{A},\mathrm{N}_k}=f\left(f^{-1}\left(\delta_{\mathrm{A},\mathrm{N}_k}^{*}\right)+\xi_{\mathrm{A},\mathrm{N}_k}\right)\tag{9}$$

其中，

$$f(x)=1-\frac{1}{2}\exp(-x)\tag{10}$$

将式(10)代入式(9)中，得到

$$\delta_{\mathrm{A},\mathrm{N}_k}=1-\left(\left(1-\delta_{\mathrm{A},\mathrm{N}_k}^{*}\right)\cdot\exp\left(-\xi_{\mathrm{A},\mathrm{N}_k}\right)\right)\tag{11}$$

随着历史 T 时间段内的节点影响力以及当前时间段 T 内节点 $\xi_{\mathrm{A},\mathrm{N}_k}$ 值的增加，节点在当前时间段 T 内的影响力随之增加。

3.2 可疑节点集合

影响力高的好友可能是与社交个体关系亲密的社交个体控制的可信节点，也可能是由女巫节点伪装成的可信节点。若影响力较高的好友为女巫节点，会对社交个体甚至整个网络产生极大的危害。本文根据节点好友的影响力，筛选影响力较高的好友组成可疑节点集合进一步检测，其余好友标记为正常节点。

跟踪节点负责对可疑节点行为进行监测，若跟踪节点只对可疑节点而不是所有的好友节点进行监测的话，可以减少大量不必要的网络资源消耗。合理地设定影响力阈值 δ_{th} 能够有效控制可疑节点集合大小，显著降低可疑节点筛选过程所消耗的系统资源。本文根据不同时间段节点好友的影响力，利用最大熵原理[18]设定动态阈值 δ_{th}，提高可疑节点筛选的准确度，节约系统资源。具体过程如下：

设节点 A 所有好友的影响力集合为 Ω_{A}，设定阈值 δ 将该集合分为两部分：$\Omega_1=\left\{\delta_{\mathrm{A},\mathrm{N}_k}\left|\delta_{\mathrm{A},\mathrm{N}_k}\geq\delta,\right.\right.$ $\left.\left.\delta_{\mathrm{A},\mathrm{N}_k}\in\Omega_{\mathrm{A}}\right\}$ 和 $\Omega_2=\left\{\delta_{\mathrm{A},\mathrm{N}_k}\left|\delta_{\mathrm{A},\mathrm{N}_k}<\delta,\ \delta_{\mathrm{A},\mathrm{N}_k}\in\Omega_{\mathrm{A}}\right.\right\}$。其中，$\Omega_1$ 表示筛选出的可疑节点集合，Ω_2 表示正常节点集合，分别计算两个集合的信息熵 H_1 和 H_2：

$$H_1=-\sum_i\frac{\delta_{\mathrm{A},\mathrm{N}_k}}{\delta}\cdot\lg\left(\frac{\delta_{\mathrm{A},\mathrm{N}_k}}{\delta}\right),\ \delta_{\mathrm{A},\mathrm{N}_k}\in\Omega_1\tag{12}$$

$$H_2=-\sum_i\frac{\delta_{\mathrm{A},\mathrm{N}_k}}{1-\delta}\cdot\lg\left(\frac{\delta_{\mathrm{A},\mathrm{N}_k}}{1-\delta}\right),\ \delta_{\mathrm{A},\mathrm{N}_k}\in\Omega_2\tag{13}$$

按照最大熵原理，合理地确定节点影响力阈值 δ_{th} 如式(14)：

$$\delta_{\mathrm{th}}=\arg\max_{\delta}\left(H_1+H_2\right)\tag{14}$$

进而，由阈值 δ_{th} 可构建可疑节点集合 Ω_1，准确筛选可疑节点，并降低网络开销。

4 女巫节点检测

针对女巫节点假名频繁切换以及女巫节点之间相互协作的特点，本文通过假名变换检测算法(Pseudonym Changing Algorithm, PCA)检测可疑节点的异常假名变换行为；利用可疑好友与节点其他好友的关系，检测可疑节点的异常协作行为，最后根据节点异常行为利用隐形马尔科夫模型(Hidden Markov Model, HMM)计算可疑节点在下一时间段 T 为女巫节点的概率，判定该可疑节点是否为女巫节点。

4.1 异常假名变换检测

女巫节点在与其他节点交互时会使用多个假名

伪装自己，使正常节点或跟踪节点很难检测到女巫节点的真实身份，但同时，这也要求女巫节点在发起攻击时需要频繁切换假名。利用女巫节点的这个行为特征。本文利用跟踪节点观察可疑节点假名变换行为并将观察结果报告给管理中心。

给定时间段 T 内，女巫节点在发起攻击时其假名变换次数会远远高于正常节点[19]，所以其对应的数据会偏离样本平均值很远，可以看作异常值，因此，可利用异常值检测的方法辨识假名变换异常的可疑节点。本文使用格拉布斯(Grubbs)准则[20]对可疑节点假名变换次数逐一识别。

统计 T 时间段内可疑节点的假名变换次数，$\mathrm{CL}_{i,k}$ 表示跟踪节点 U_i 与可疑节点 SN_k 在 T 时间段内的交互列表，$\Phi_{i,k,t_n}=(\mathrm{pse}_i,\mathrm{pse}_k^n)$ 表示跟踪节点 U_i 与可疑节点 SN_k 在 t_n 时刻分别使用假名 pse_i 和 pse_k^n 发起的一次交互，如果 $\Phi_{i,k,t_n}\notin \mathrm{CL}_{i,k}$，则将这个交互过程添加到跟踪节点 U_i 与可疑好友 SN_k 的交互列表 $\mathrm{CL}_{i,k}$。T 时间段内可疑节点 SN_k 的假名变换次数 AbNum_k^T 即为交互列表 $\mathrm{CL}_{i,k}$ 的大小 $\left|\mathrm{CL}_{i,k}\right|$，对 AbNum_k^T 从小到大排序，得到 T 时间段内可疑节点假名变换次数样本 $\mathrm{AbNum}^{T^*}=\{\mathrm{AbNum}_{i1}^T<\mathrm{AbNum}_{i2}^T<\cdots<\mathrm{AbNum}_{ik}^T\}$，进而，可获知可疑节点假名变换次数的平均值 E_{AbNum} 和标准差 S_{AbNum}。

异常值应为样本中较大的值，首先选择假名变换次数样本中的最大值作为可疑值，计算其与平均值的差，得到可疑值的偏离值，然后计算偏离值与标准差的比值得到该可疑节点假名变换次数的 G_i 值。

$$G_i=\left[\max\left(\mathrm{AbNum}^{T^*}\right)-E_{\mathrm{AbNum}}\right]\Big/S_{\mathrm{AbNum}} \tag{15}$$

将该可疑节点假名变换次数的 G_i 值与格布拉斯表给出的临界值 $G_p(n)$ 比较。如果计算出的 G_i 值大于临界值 $G_p(n)$，则跟踪节点判定该可疑节点的假名变换次数为异常值，从样本中剔除，并判定该可疑节点在当前时间段 T 内发生了异常假名变换行为，将该可疑节点所使用的假名以及使用该假名所建立的交互过程列表发送给管理中心，管理中心更新该节点的异常行为信息。然后继续选择剔除异常值后样本中的最大值重复上述步骤，直到所计算的节点假名变换次数的 G_i 值小于临界值 $G_p(n)$。其中，n 为样本中数据数量，p 为置信概率，与检出水平 α 有关，$p=1-\alpha$，α 为人为设定，取值越小检验要求越严格。

4.2 可疑节点恶意攻击估计

隐形马尔科夫模型(Hidden Markov Model, HMM)[21]可以通过观察节点外在表现行为进行自主学习，并将观察的结果与其所隐藏的状态相互关联，评估节点当前的状态和潜在的攻击性。本文利用隐形马尔科夫模型观察可疑节点的假名变换行为和异常协作行为，推测可疑节点的真实身份。

可疑节点的身份 Q 有两种：正常节点 $\mathrm{Q_N}$ 和女巫节点 $\mathrm{Q_S}$，可观察的节点表现为节点历史异常行为，包括异常假名变换行为和异常协作行为。建立每一个可疑节点身份状态切换的隐形马尔科夫模型 $\mathrm{HMM}(\lambda)=(\boldsymbol{\pi},\boldsymbol{A},\boldsymbol{B})$，其中各符号含义如下：

$\boldsymbol{\pi}$ 为可疑节点的初始状态概率矩阵，即上一时间段 T 内可疑节点为女巫节点和正常节点的概率。若针对该可疑节点执行首次检测，则处于两种状态的概率分别为 1/2。

$\boldsymbol{A}=\left[a_{\mathrm{Q}_i,\mathrm{Q}_j}\right]$ 为可疑节点的身份状态转移矩阵，其中 $a_{\mathrm{Q}_i,\mathrm{Q}_j}$ 表示可疑节点身份由 Q_i 转换为 Q_j 的概率。设某可疑节点在前 $n-1$ 个 T 时间段内，有 m 个时间段被检测为女巫节点，则其第 n 个时间段 T 内在对应的状态转移矩阵为 $\left[a_{\mathrm{Q}_i,\mathrm{Q}_s}\right]=m/(n-1)$，初始值设为 1/2。

$\boldsymbol{B}=\left[b_{N_{\mathrm{Ab}},\mathrm{Q}_i}\right]$ 表示节点可观察的行为表现概率矩阵，其中 $b_{N_{\mathrm{Ab}},\mathrm{Q}_i}$ 表示可疑节点状态为 Q_i 时，历史异常行为次数为 N_{Ab} 的概率。异常假名变换行为如 4.1 节所述。对于异常协作行为，通常，一个节点有超过 1/3 的好友为女巫节点时，则称该节点被女巫节点完全感染[8]。计算时间段 T 内可疑节点对目标节点其他好友的影响力，若可疑节点对超过 1/3 的目标节点好友影响力大于影响力阈值 δ_{th}，则称该可疑节点在当前时间段 T 内有异常协作行为。随着可疑好友节点异常行为的增加，其为女巫节点的概率逐渐增加，增加到一定程度后趋于稳定，具有单调递增趋势，用式(16)计算节点为女巫节点时历史异常行为次数为 N_{Ab} 的概率 $b_{N_{\mathrm{Ab}},\mathrm{Q_S}}$：

$$b_{N_{\mathrm{Ab}},\mathrm{Q_S}}=1-\exp\left(-N_{\mathrm{Ab}}\right) \tag{16}$$

利用维特比算法和上述 $\mathrm{HMM}(\lambda)$ 模型，根据可疑节点历史异常行为预测下一时间段 T 可疑节点为女巫节点的概率。可观测序列 $N=\left\{N_{\mathrm{Ab}}^{T_1},N_{\mathrm{Ab}}^{T_2},\cdots,N_{\mathrm{Ab}}^{T_n}\right\}$ 为节点在每个 T 时间段内的累积异常行为次数。其中 T_n 为当前 T 时间段。定义维特比变量：

$$\delta_{T_n}\left(\mathrm{Q_S}\right)=\max_{\mathrm{Q}_i^{T_1},\mathrm{Q}_i^{T_2},\cdots,\mathrm{Q}_i^{T_{n-1}}} P\left\{\mathrm{Q}_i^{T_n}=\mathrm{Q_S},\mathrm{Q}_i^{T_1},\mathrm{Q}_i^{T_2},\cdots,\mathrm{Q}_i^{T_{n-1}},N_{\mathrm{Ab}}^{T_1},N_{\mathrm{Ab}}^{T_2},\cdots,N_{\mathrm{Ab}}^{T_n}\right\} \tag{17}$$

初始化维特比变量：

$$\delta_{T_1}\left(\mathrm{Q_S}\right)=\boldsymbol{\pi}\cdot b_{N_{\mathrm{Ab}}^{T_1},\mathrm{Q_S}} \tag{18}$$

由定义得到变量 δ 的递推公式：

$$\delta_{T_{n+1}}(\mathrm{Q_S}) = \max_{j\in\{\mathrm{Q_S},\mathrm{Q_N}\}}\left(\delta_{T_n}(\mathrm{Q}_j)\cdot a_{\mathrm{Q}_j,\mathrm{Q_S}}\right)\cdot b_{N_{\mathrm{Ab}}^{T_1},\mathrm{Q_S}} \tag{19}$$

最终根据动态规划方法得到下一时间段 T 内可疑节点是女巫节点的概率：

$$P_{T_{n+1}}(\mathrm{Q_S}) = \max_{j\in\{\mathrm{Q_S},\mathrm{Q_N}\}}\delta_{T_n}(\mathrm{Q_S}) \tag{20}$$

若可疑节点在下一时间段 T 为女巫节点的概率大于节点恶意概率阈值 P_{Sybil}，则该可疑节点为女巫节点。恶意概率阈值 P_{Sybil} 受网络环境影响，网络环境恶意程度越高，恶意概率阈值越小，本文用网络中可疑节点数量评估网络环境的恶意程度，确定节点恶意概率阈值，如式(21)：

$$P_{\mathrm{Sybil}} = 1 - N_{\mathrm{sus}}/N \tag{21}$$

其中，$N_{\mathrm{sus}} = |\Omega_1|$ 为目标节点当前时间段 T 可疑好友数量，由第 3 节得到，N 为目标节点好友节点总数量。

最后，对所提机制复杂度进行分析，目标节点好友总数为 n，运用最大熵原理计算节点好友影响力阈值筛选可疑节点复杂度为 $O(n^2)$；假名变换行为检测过程最大复杂度为 $O(n)$；异常协作行为检测过程最大复杂度为 $O(n(n+1))$，即 $O(n^2)$；构建可疑节点隐形马尔科夫模型 $\mathrm{HMM}(\lambda)=(\boldsymbol{\pi},\boldsymbol{A},\boldsymbol{B})$ 最大复杂度为 $O(n)$；最后利用维特比算法推测可疑节点真实身份，复杂度为 $O(N^2k)$，其中，N 表示可疑节点状态数量，k 表示可观测值数量，综上所述，所提机制复杂度为 $O(n^2+N^2k)$。

5 数值结果分析

本节在 MATLAB 平台下，使用微博实测数据集 MicroblogPCU 对所提出的基于节点行为特征分析的社交网络女巫攻击检测机制 BASD 在女巫节点检测方面的准确度和可靠性进行验证，并与传统女巫攻击防御算法 SybilGuard 以及文献[8]提出的女巫防御算法(Signed Network-based Sybil Defense, SNSD)进行对比。

实验过程中，筛选出一个目标节点及其所有 438 个粉丝，粉丝认为是本文中目标节点的好友，其中包含 32 个恶意节点，统计其身份属性信息，计算目标节点与其粉丝的静态相似度，统计该目标节点与其粉丝之间 1 个月之内的互动行为，计算目标节点与其粉丝的动态相似度。一次转发或评论行为即为一次互动，联系紧密的多次互动行为构成一次交互。给定时间段 T 内，节点第 n 次互动行为发生时刻为 t_n^{inter}，则第 n 次互动与第 n+1 次互动间隔时间 $\Delta t_n^{\mathrm{inter}} = t_{n+1}^{\mathrm{inter}} - t_n^{\mathrm{inter}}$，则在时间段 T 内第 1 次交互起始时刻为时间段 T 内第 1 次互动发生时刻，即 $\mathrm{ST}_1 = t_1^{\mathrm{inter}}$，第 1 次交互终止时刻 $\mathrm{ET}_1 = t_k^{\mathrm{inter}}$，其中，$k$ 满足：

$$\begin{aligned}&\max\left\{\left(\Delta t_1^{\mathrm{inter}} - E\left(\Delta t_n^{\mathrm{inter}}\right)\right)^2,\right.\\&\left.\left(\Delta t_2^{\mathrm{inter}} - E\left(\Delta t_n^{\mathrm{inter}}\right)\right)^2,\cdots,\left(\Delta t_{k-1}^{\mathrm{inter}} - E\left(\Delta t_n^{\mathrm{inter}}\right)\right)^2\right\}\\&\le D\left(\Delta t_n^{\mathrm{inter}}\right) < \left(\Delta t_k^{\mathrm{inter}} - E\left(\Delta t_n^{\mathrm{inter}}\right)\right)^2\end{aligned} \tag{22}$$

其中，$D(\Delta t_n^{\mathrm{inter}})$ 为 T 时间段内两节点互动间隔时间方差，$E(\Delta t_n^{\mathrm{inter}})$ 为互动间隔时间期望。统计所有节点 1 个月之内互动行为所使用的假名数量，进行异常假名变换检测。置信度参数 α 设置为 0.05[22]。

性能指标为女巫节点识别率 TPR 和误检率 FPR，其具体定义如下：设网络中正常节点和女巫节点组成的节点集合分别为 N 和 M，在女巫节点检测过程中，被检测机制正确检测的女巫节点组成集合 M_{TPR}，实际为正常节点但被检测机制误判为女巫节点的节点组成集合 N_{FPR}，则女巫节点的识别率 TPR 为 $|M_{\mathrm{TPR}}|/|M|$，女巫节点的误检率 FPR 为 $|N_{\mathrm{FPR}}|/|N|$，识别率越大，误检率越小，检测机制的准确度越高。

5.1 不同女巫节点百分比下的性能分析

本节对女巫节点百分比对所提机制可疑节点筛选过程的影响进行了验证分析，并在不同女巫节点百分比下对所提机制 BASD, SybilGuard 以及 SNSD 的识别率及误检率进行了比较分析，时间窗口 T 为 3 d(天)。

由图1可以看出，随着网络中女巫节点百分比的增加，可疑节点与节点总数的比值随之增加，女巫节点数量与可疑节点数量的比值也随之增加，当网络中女巫节点百分比达到30%时，可疑节点中女巫节点的比例可达45%以上，可疑节点只占节点总数的60%，此时，只对筛选出的可疑节点进行跟踪检测可以在保证女巫节点检测准确度性能的同时，大大节约网络资源消耗。

由图2和图3可以看出，各机制的识别率和误检率均随女巫节点百分比的升高分别呈下降和上升趋势，而本文所提出的机制相比其他两种机制体现出较大的优势，在女巫节点百分比为30%时，女巫节点识别率为86%，误判率为26.5%。主要因为在检测女巫节点时本文算法既考虑了恶意节点的共性也考虑了女巫节点的个性，在女巫节点比例较小时，女巫节点的行为特征表现不明显，可以利用恶意节点的共性对其进行检测，在女巫节点比例较大时，针对女巫节点的行为特征进行检测，因此在整体准确性上会有明显提高。

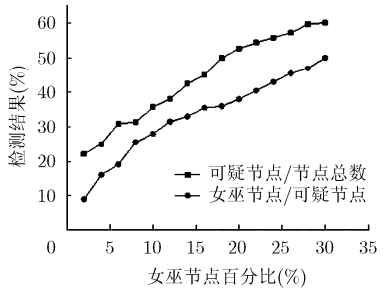

图 1 女巫节点百分比变化对可疑节点筛选的影响

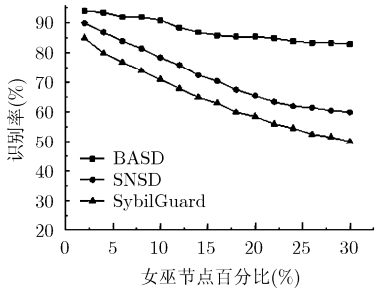

图 2 女巫节点百分比变化对识别率的影响

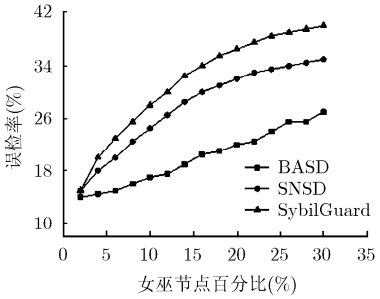

图 3 女巫节点百分比变化对误检率的影响

5.2 不同女巫节点攻击次数下的性能分析

本节在不同女巫节点攻击次数下对本文所提机制 BASD, SybilGuard 以及 SNSD 的识别率及误检率进行了比较分析，时间窗口 T 为 3 d，女巫节点比例为 30%。

图 4 及图 5 分别描述了女巫节点攻击次数对识别率和误检率的影响。当女巫节点攻击次数增加时，3 种机制的识别率和误检率均分别呈上升和下降趋势。本文所提机制其准确性受女巫节点攻击次数的影响较小。当女巫节点攻击次数较少时，即女巫节点向目标节点发起的恶意交互过程较少，女巫节点攻击力度不够，此时女巫节点与正常节点的行为区别较小，SybilGuard 以及 SNSD 无法有效地识别女巫节点。本文机制由于对女巫节点攻击的行为特性进行了具体的分析，能够根据女巫节点的特点有针对性的检测，因此，在女巫节点发起少量攻击时就已表现出良好的检测效果。

6 结束语

为了防御社交网络中的女巫攻击，更好地保护社交个体的隐私和利益，本文针对女巫节点行为特征提出一种基于行为特征分析的社交网络女巫节点检测机制，通过评估节点间的静态相似度和动态相似度获知节点的影响力，筛选可疑节点，然后根据可疑节点异常行为利用隐形马尔科夫模型推测节点真实身份，检测女巫节点，一定程度上减少网络开销。数据结果表明，所提机制在女巫节点识别率和误检率性能上都有较好的表现，能够有效地检测出社交网络中的女巫节点，更好地保护社交个体的隐私和利益。

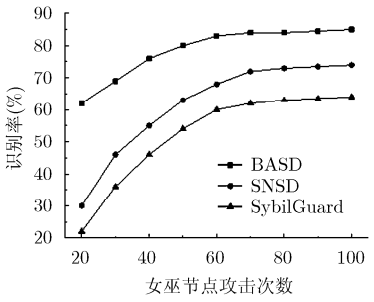

图 4 女巫节点攻击次数变化对识别率的影响

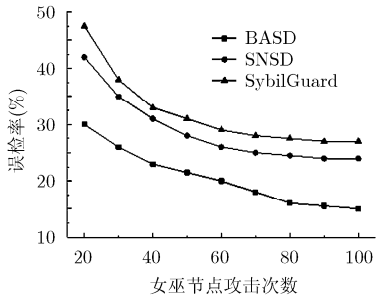

图 5 女巫节点攻击次数变化对误检率的影响

参 考 文 献

[1] CALDELLI R, BECARELLI R, and AMERINI I. Image origin classification based on social network provenance[J]. *IEEE Transactions on Information Forensics and Security*, 2017, 12(6): 1299–1308. doi: 10.1109/TIFS.2017.2656842.

[2] TONG G, WU W, TANG S, *et al*. Adaptive influence maximization in dynamic social networks[J]. *IEEE Transactions on Networking*, 2017, 25(1): 112–125. doi: 10.1109/TNET.2016.2563397.

[3] KHAN M S, WAHAB A W A, HERAWAN T, *et al*. Virtual community detection through the association between prime nodes in online social networks and its application to ranking algorithms[J]. *IEEE Access*, 2016, 4: 9614–9624. doi: 10.1109/ACCESS.2016.2639563.

[4] WU D P, ZHANG P N, WANG H G, *et al*. Node service ability aware, packet forwarding mechanism in intermittently connected wireless networks[J]. *IEEE Transactions on Wireless Communications*, 2016, 15(12): 8169–8181. doi: 10.1109/TWC.2016.2613077.

[5] ZHANG K, LIANG X, SHEN X, *et al*. Exploiting multimedia services in mobile social networks from security and privacy perspectives[J]. *IEEE Communications Magazine*, 2014, 52(3): 58–65. doi: 10.1109/MCOM.2014.6766086.

[6] ZHANG J, ZHANG R, SUN J, *et al*. TrueTop: a sybil-resilient system for user influence measurement on Twitter[J]. *IEEE Transactions on Networking*, 2016, 24(5): 2834–2846. doi: 10.1109/TNET.2015.2494059.

[7] VASUDEVAN S K, SIVARAMAN R, and KARTHICK M R. Sybil guard: Defending against sybil attacks via social networks[J]. *International Journal of Computer Applications*, 2010, 5(3): 27–42. doi: 10.1145/1159913.1159945.

[8] CHANG W, WU J, TAN C C, *et al*. Sybil defenses in mobile social networks[C]. IEEE Global Communications Conference, Atlanta, GA, USA, 2013: 641–646. doi: 10.1109/GLOCOM.2013.6831144.

[9] KRISHNAMURTHY B, GILL P, and ARLITT M. A few chirps about Twitter[C]. Proceedings of the First Workshop on Online Social Networks, Seattle, USA, 2008: 19–24.

[10] CHU Z, GIANVECCHIO S, WANG H, *et al*. Who is tweeting on twitter: Human, bot or cyborg?[C]. Proceedings of 26th Annual Computer Security Applications Conference, Austin, USA, 2010: 21–30.

[11] TAN L, LIAN Y F, and CHEN K. Malicious users identification in social network based on composite classification model[J]. *Computer Applications and Softeware*, 2012, 29(12): 1–5.

[12] ZHANG K, LIANG X, LU R, *et al*. Exploiting mobile social behaviors for sybil detection[C]. IEEE Conference on Coputer Communications, HongKong, China, 2015: 271–279.

[13] FENG M, MAO S, and JIANG T. Joint duplex mode selection, channel allocation, and power control for full-duplex cognitive femtocell networks[J]. *Digital Communications and Networks*, 2015, 1(1): 30–44.

[14] IRFAN R, BICKLER G, KHAN S U, *et al*. Survey on social networking services[J]. *IET Networks*, 2013, 2(4): 224–234. doi: 10.1049/iet-net.2013.0009.

[15] YAO L, MAN Y, HUANG Z, *et al*. Secure routing based on social similarity in opportunistic networks[J]. *IEEE Transactions on Wireless Communications*, 2016, 15(1): 594–605. doi: 10.1109/TWC.2015.2476466.

[16] WU D P, YANG B R, WANG H G, *et al*. Privacy-preserving multimedia big data aggregation in large-scale wireless sensor networks[J]. *ACM Transactions on Multimedia Computing, Communications, and Applications*, 2016, 12(4): 1–19. doi: 10.1145/2978570.

[17] WANG R Y, YANG H P, WANG H G, *et al*. Social overlapping community-aware neighbor discovery for D2D communications[J]. *IEEE Wireless Communications*, 2016, 23(4): 28–34. doi: 10.1109/MWC.2016.7553023.

[18] QIN L, SUN K Q, and LI S G. Maximum fuzzy entropy image segmentation based on artificial fish school algorithm[C]. International Conference on Intelligent Human-Machine Systems and Cybernetics, Hangzhou, China, 2016: 164–168.

[19] LIANG X, LI X, ZHANG K, *et al*. Fully anonymous profile matching in mobile social networks[J]. *IEEE Journal on Selected Areas in Communications*, 2013, 31(9): 641–655. doi: 10.1109/JSAC.2013.SUP.0513056.

[20] LUO W, WU Y, YUAN J, *et al*. The calculation method with Grubbs test for real-time saturation flow tate at signalized intersection[C]. Proceedings of the Second International Conference on Intelligent Transportation, Singapore, 2017: 129–136.

[21] KITZIG A, NAROSKA E, STOCKMANNS G, *et al*. A novel approach to creating artificial training and test data for an HMM based posture recognition system[C]. International Workshop on Machine Learning for Signal Processing, Salerno, Italy, 2016: 1–6.

[22] WANG X F, LIU L, and SU J S. RLM: A general model for trust representation and aggregation[J]. *IEEE Transactions on Service Computing*, 2012, 5(1): 131–143. doi: 10.1109/TSC.2010.56.

吴大鹏： 男，1979 年生，教授，主要研究方向为泛在无线网络、无线网络服务质量管理等.

司书山： 男，1993 年生，硕士生，研究方向为社交网络攻击检测.

闫俊杰： 男，1990 年生，博士生，研究方向为 D2D 通信、雾计算等.

王汝言： 男，1969 年生，教授，主要研究方向为泛在网络、多媒体信息处理等.