

Tor 网桥分发中融合物理-社交属性的女巫节点检测机制

史鑫, 郭云飞, 王亚文, 孙小丽, 梁浩

(信息工程大学, 河南 郑州 450001)

摘要: 作为目前应用范围最广的网络审查规避系统之一, Tor 在网桥分发过程中面临着严重的女巫攻击威胁。具有丰富网络和人力资源的审查者往往会部署大量女巫节点, 它们通过伪装成正常节点来获取网桥信息并将其封锁或屏蔽。在此过程中, 由于女巫节点和正常节点身份、目的和意图的不同, 在网络活动中会产生个体或群体行为差异, 称为节点行为特征。针对上述女巫攻击威胁, 在分析节点行为特征的基础上提出了融合物理-社交属性的女巫节点检测机制。设计了节点物理域和社交域属性评估方法。采用客观反映节点上网桥正常运行状态的节点积分值和体现网桥屏蔽情况的节点风险指数来评估节点的物理域属性; 用描述节点静态属性标签的社交相似度和刻画节点动态交互行为特征的社交信任度来评估节点的社交域属性。进而, 融合节点的物理域和社交域属性定义可信度指标, 表征当前节点为女巫节点的可能性, 并以此为指导推测节点的真实身份, 实现对女巫节点的精准检测。基于构建的 Tor 网络运行情况模拟器和 MicroblogPCU 数据集对所提机制的检测性能进行仿真实验。结果表明, 所提机制能够有效提高女巫节点识别率, 降低误检率, 对于审查者的迷惑行为具有更强抵御能力, 并且在节点社交属性缺失情况下仍具有良好检测性能。

关键词: Tor 网桥分发; 女巫节点检测; 行为特征; 物理-社交属性

中图分类号: TP393

文献标志码: A

DOI: 10.11959/j.issn.2096-109x.2023014

Physical-social attributes integrated Sybil detection for Tor bridge distribution

SHI Xin, GUO Yunfei, WANG Yawen, SUN Xiaoli, LIANG Hao

Information Engineering University, Zhengzhou 450001, China

Abstract: As one of the most widely utilized censorship circumvention systems, Tor faces serious Sybil attacks in bridge distribution. Censors with rich network and human resources usually deploy a large number of Sybils, which disguise themselves as normal nodes to obtain bridges information and block them. In the process, due to the different identities, purposes and intentions of Sybils and normal nodes, individual or group behavior differences

收稿日期: 2022-01-17; 修回日期: 2022-07-06

通信作者: 史鑫, shi1995@pku.edu.cn

基金项目: 国家重点研发计划 (2021YFB1006200, 2021YFB1006201); 国家自然科学基金 (62072467, 62002383)

Foundation Items: The National Key R&D Program of China (2021YFB1006200, 2021YFB1006201), The National Natural Science Foundation of China (62072467, 62002383)

引用格式: 史鑫, 郭云飞, 王亚文, 等. Tor 网桥分发中融合物理-社交属性的女巫节点检测机制[J]. 网络与信息安全学报, 2023, 9(1): 103-114.

Citation Format: SHI X, GUO Y F, WANG Y W, et al. Physical-social attributes integrated Sybil detection for Tor bridge distribution[J]. Chinese Journal of Network and Information Security, 2023, 9(1): 103-114.

occur in network activities, called as node behavior characteristics. To handle the Sybil attacks threat, a Sybil detection mechanism integrating physical-social attributes was proposed based on the analysis of node behavior characteristics. The physical-social attributes evaluation methods were designed. The credit value of nodes objectively reflecting the operation status of bridges on the nodes and the suspicion index of nodes reflecting the blocking status of bridges, were utilized to evaluate the physical attributes of nodes. The social attributes of nodes were evaluated by the social similarity, which described the static attribute labels of nodes and their social trust characterizing the dynamic interaction behaviors of nodes. Furthermore, integrating the physical-social attributes, the credibility of nodes were defined as the possibility of the current node being a Sybil, which was exploited as a guidance on inferring the true identifies of nodes, so as to achieve accurate detection on Sybils. The detection performance of the proposed mechanism based on the constructed Tor network operation status simulator and the Microblog PCU dataset were simulated. The results show that the proposed mechanism can effectively improve the true positive rate on Sybils, and decrease the false positive rate. It also has stronger resistance on the deceptive behavior of censors, and still performs well in the absence of node social attributes.

Keywords: Tor bridge distribution, Sybil detection, behavior characteristics, physical-social attributes

0 引言

随着信息技术的发展进步, 互联网正逐渐成为人们自由地分享观点和获取信息的主要途径。为了控制网络舆情信息的传播和扩散, 许多国家和地区对节点接入互联网的行为进行了监控和限制, 称作网络审查^[1]。常用的网络审查技术包括 IP 地址黑名单、域名劫持和污染、深度包检测等^[2-4]。为了抵御上述网络层面上的审查和封锁, 研究人员开发了代理接入^[5]、域前置^[6]、诱饵路由^[7]等多种审查规避技术, 其中应用最为广泛的代理接入技术是指通过在审查区域之外部署代理, 为处于审查区域内的节点提供到被屏蔽站点的间接访问服务^[8]。

Tor 是目前世界上使用人数最多、范围最广的基于代理接入的网络审查规避系统, 用位于审查区域外的志愿者运行的“网桥”作为代理为受到审查的节点提供网络接入服务, 同时基于洋葱路由技术为节点的网络活动提供匿名性保障^[9]。截止到 2021 年 6 月, Tor 的全球用户数量峰值接近 600 万, 常用用户数量超过 200 万, 其中通过网桥接入的用户数量峰值超过 20 万^[10]。

Tor 网络的基本组成与通信流程如图 1 所示^[11]。Tor 中运行着 10 台由官方管控的权威目录服务器和 6 000 余台分布在全球各地的备用目录服务器, 目录服务器的信息通过硬编码方式写入 Tor 源码中并定期更新。通常情况下, 节点接入

Tor 网络的流程如图中黑色实线所示。Tor 客户端启动后, 首先尝试与目录服务器建立连接, 成功建立连接后开始下载中继节点快照和描述符。当获取到的中继信息足够多后, 客户端根据路径选择算法选择中继节点并尝试构建用于正常节点和通信目标间匿名通信活动的专用通信链路, 即“电路”^[9]。在成功构建足够数量的电路后, 即可通过 Tor 匿名网络活动^[12]。然而, 在网络审查严格的国家和地区, 由于目录服务器被屏蔽, 节点难以直接与目录服务器建立连接并获取中继信息。鉴于此, Tor 项目为审查区域内的节点提供了网桥接入服务(如图中虚线所示), 正常节点可以通过电子邮件等带外方式从网桥分发服务处获取网桥信息, 而后使用网桥作为代理访问目录服务器, 进而获取中继信息并构建电路, 实现通过 Tor 匿名上网^[13]。

与此同时, 具有丰富网络及人力资源的审查者往往会通过部署女巫节点的方式来获取网桥信息, 进而将其封锁或屏蔽, 使正常节点无网桥可用, 这种行为被称作女巫攻击^[14]。作为 Tor 网络中的宝贵资源, 网桥信息的安全保密直接关系到审查区域内的节点能否成功接入 Tor 网络, 因此具有重要的保护价值。为了确保网桥分发服务能够有针对性地将网桥信息分发给具有审查规避需求的节点, 而非由审查者部署的女巫节点, 现有文献通过挖掘和分析节点的行为特征, 在针对女巫攻击的防御策略方面取得了系列研究成果^[8,15-20]。

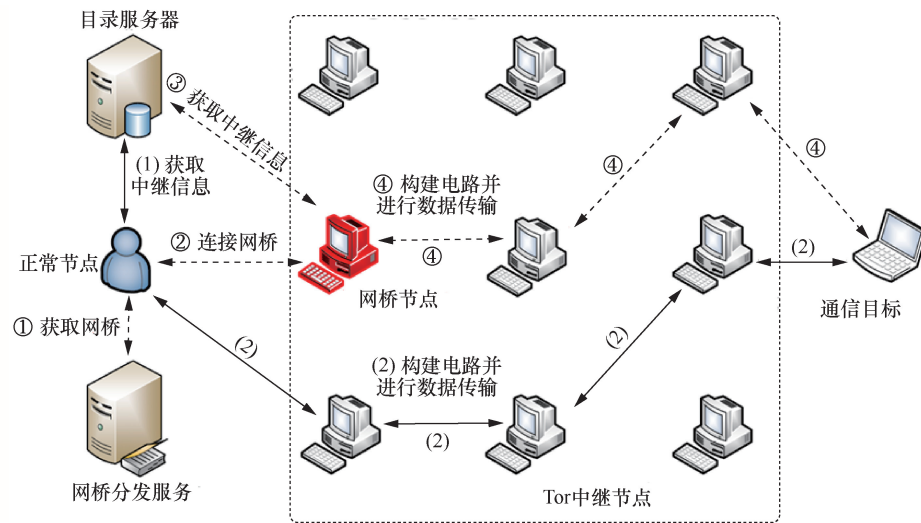


图1 Tor网络的基本组成与通信流程

Figure 1 Basic components and communication flow for Tor

文献[15]首次将审查规避系统中的网桥分发建模为算法问题,对于女巫节点和正常节点数量确定的网络,推导并证明了为全体正常节点提供审查规避服务所需网桥数量的理论界限值。后续研究大多基于此,将针对女巫攻击的防御策略集成于Tor网桥分发技术,探究如何设计安全高效的网桥分发方案。例如,文献[16]为Tor设计了网桥分发协议TorBricks,能够实现近似文献[15]中得出的理论界限值。文献[17]将女巫节点获取网桥信息的概率模型建模为泊松过程,提出Proximax网桥分发机制,实现了网桥利用率和网桥信息泄露风险的折中优化。然而上述研究普遍存在对审查者和女巫节点的行为假设过于理想化这一问题,缺乏在实际应用中的指导意义。文献[8]设计了基于节点声誉积分值的网桥分发方案rBridge,通过节点上网桥的正常运行时间这一物理域属性来刻画节点行为特征;同时提出了审查者激进型和保守型两种行为模型,具有一定的实际应用价值。文献[18]根据节点上运行网桥的屏蔽情况,提出了节点怀疑程度等物理域指标,并基于此设计了网桥分发机制Salmon。文献[19]进一步考虑了如网桥使用率、网桥使用时间和网桥连接数等多重物理域属性,将网桥分发过程建模为网桥资源与节点间的匹配博弈问题,推导了审查者的最优行动策略,设计算法求解博弈均衡,并基于此提出HypoTor网桥分发机制。

由于上述针对女巫攻击的防御策略高度集成

于Tor网桥分发技术,普遍存在难以与Tor适配和实际应用困难等问题。近年来,研究者开始尝试将针对女巫攻击的防御策略与Tor网桥分发技术解耦,探究技术上独立的女巫节点检测机制。例如,文献[20, 21]根据节点与不可用资源间的历史连接关系或访问信息构建二分图,进而将Tor网络中女巫节点的检测转化为二分图的最小顶点覆盖问题,并基于贪婪算法和启发式算法来求解该NP-Complete问题。

上述针对Tor网桥分发过程中女巫攻击的防御策略研究,大多集中于基于数学理论的建模分析,以及基于节点上网桥运行状态等物理域属性的女巫节点检测及网桥分发机制设计。由于其对节点行为特征挖掘不充分,普遍存在难以兼顾检测识别率和误检率两方面性能,以及对审查者采取的迷惑行为抵御能力较差等问题。在社交网络和无线传感器网络等背景下,研究者针对女巫攻击问题,提出了基于节点间相似度和信任值等社交域属性的女巫节点检测机制^[22-25],并通过仿真实验证明了所提机制能够有效提高对女巫节点的检测性能。

本文针对Tor网桥分发中的女巫攻击问题,提出了一种融合物理-社交属性的女巫节点检测机制。首先,分析节点上网桥的正常运行状态和屏蔽情况,用节点积分值和节点风险指数来评估物理域属性;其次,通过分析节点的静态属性标签和动态行为特征,用社交相似度和社交信任度来评估社交域属性。在此基础上,综合考虑节点

的物理域和社交域属性,定义可信度指标,来表征当前节点为女巫节点的可能性,并基于可信度实现女巫节点检测。仿真实验表明,相较现有工作,本文所提机制能够更加高效和准确地检测女巫节点,并且对审查者采取的迷惑行为具备更强的抵御能力。

1 Tor 网桥分发系统模型

Tor 网桥分发系统模型如图 2 所示。网络审查区域内存在着具有审查规避需求的正常节点和由审查者部署的女巫节点,全体节点共同构成集合 $\mathcal{U} = \{u_1, \dots, u_i, \dots, u_{N_u}\}$, $N_u = |\mathcal{U}|$ 。Tor 网桥分发服务存储着当前网络中全部可用网桥的信息,表示为 $\mathcal{B} = \{b_1, \dots, b_j, \dots, b_{N_b}\}$, $N_b = |\mathcal{B}|$ 。为规避网络审查,正常节点会通过电子邮件等带外方式向 Tor 网桥分发服务发送请求。与此同时,审查者会部署一定数量的女巫节点,通过伪装成正常节点的方式向 Tor 网桥分发服务请求网桥,一旦女巫节点获取到网桥信息后,则会上报给审查者,由审查者采取屏蔽策略。本文工作主要考虑了审查者可能采取的两类策略,即激进型和保守型策略^[8,19]。激进型策略是指审查者在收到女巫节点上报的网桥信息后立刻将其屏蔽。而保守型策略是指审查者为了保护女巫节点、混淆女巫节点和正常节点间的行为特征采取的随机延迟屏蔽策略。即对于某个已获取到信息的网桥 b_j ,审查者将在等待一段随机时间 $T_{\text{wait}}^{b_j}$ 后再将其屏蔽。Tor 网桥分发服务会在固定时刻处理收到的网桥请求并进行响应,Tor 网桥分发服务连续两次处理网桥请求的时间间隔被称作一个时间周期。

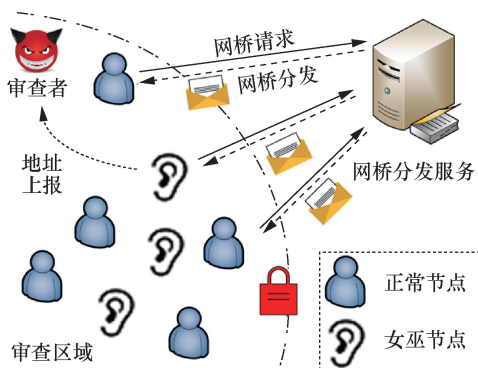


图2 Tor 网桥分发系统模型
Figure 2 Tor bridge distribution system model

由于网络审查者和女巫节点的存在,在收到节点发送的网桥请求后,若 Tor 网桥分发服务不加以检测地为全体节点分发网桥,将会导致大量网桥被审查者屏蔽,具有审查规避需求的正常节点难以接入 Tor 网络。因此,设计精准高效的检测机制来区分网络中的正常节点和女巫节点,有选择性地为正常节点分发网桥至关重要。

具体而言,检测机制的性能指标主要包括识别率和误检率两方面^[23]。识别率也称作真正率 (TPR, true positive rate),表示全体女巫节点中被正确检出的比例。TPR 越接近于 1,说明该检测机制对女巫节点的成功检出率越高,检测性能越高效。误检率又称作假正率 (FPR, false positive rate),表示全体正常节点中被错误检测为女巫节点的比例。FPR 越接近于 0,表示检测机制对于正常节点的误检概率越低,检测性能越精准。完美的检测机制应能够实现女巫节点和正常节点的准确区分,即在识别率趋近于 1 的同时误检率趋近于 0。本文的目标在于通过充分挖掘和分析节点的行为特征,设计精准高效的检测机制,能够在成功检出女巫节点的同时尽可能少地误检正常节点。

需要注意的是,本文工作建立在审查区域内的全体节点仅能通过如电子邮件等带外方式向 Tor 网桥分发服务请求网桥信息这一基础假设之上。事实上,除女巫攻击之外,审查者还可能会采取如中间人攻击、流量指纹攻击和探测攻击等其他网络层面的攻击手段来获取网桥信息,达成封锁和屏蔽的目的^[14, 26],这与本文工作并不冲突。现有研究已提出多种流量混淆技术^[6, 27]和客户端认证机制^[28]来抵御网络层面的网桥识别攻击,并在实际应用中发挥了显著作用。

2 融合物理-社交属性的女巫节点检测

2.1 设计思路

节点的行为特征是指节点在网络活动中由于个体或群体差异而体现出的行为要素组合,能够在一定程度上表现出节点的身份、目的和意图,通常可分为物理域属性和社交域属性两类。物理域属性是指节点上的程序指令在真实物理网络环境下的运行情况。而社交域属性是指节点作为社

交个体,在与社交网络上的其他个体交互过程中通过长期的行为累积得到的评价。相较于物理域属性,社交域属性往往具有短期波动小、篡改难度高、伪装代价大等特点。

鉴于上述分析,设计精准高效的女巫节点检测机制,重点在于如何评估节点的物理-社交属性,进而分析节点的行为特征,并以此为指导实现女巫节点的精准检测。

不失一般性地,本文采用了能够客观反映节点上网桥正常运行状态的节点积分值和体现网桥屏蔽情况的节点风险指数来评估节点的物理域属性,用描述节点静态社交属性标签的社交相似度和刻画节点动态交互行为特征的社交信任度来评估节点的社交域属性。本节详细介绍了节点的物理-社交属性评估方法,并在此基础上定义可信度指标,用以表征当前节点为女巫节点的可能性;最后介绍了利用可信度指标实现女巫节点检测的具体流程。

2.2 节点物理-社交属性评估方法

2.2.1 节点积分值

定义 $T_{u_i, b_j}^{t_k}$ 表示时间周期 t_k 内,网桥 b_j 在节点 u_i 上的正常运行时间,则节点 u_i 在时间周期 t_k 内获取的积分 $\gamma_{u_i}^{t_k}$ 可表示为

$$\gamma_{u_i}^{t_k} = \rho \cdot (\min\{\sum_{b_j \in \mathcal{B}} T_{u_i, b_j}^{t_k}, T_{\max}\} - T_{\min})^+ \quad (1)$$

其中, $(\cdot)^+ = \max(\cdot, 0)$, $T_{\max}$ 和 $T_{\min}$ 分别表示单个时间周期内积分时长的上下限,即节点在一个时间周期内的最大积分时长为 $T_{\max} - T_{\min}$ 。设置积分时长上限是为了避免正常节点和女巫节点为获取大量积分值而长时间占用 Tor 网桥,造成 Tor 网络滥用和网桥资源浪费。设置积分时长下限一方面是为了激励保守型审查者将已获取到信息的网桥保留更长时间,另一方面能够增加正常节点和激进型审查者部署的女巫节点间物理行为特征的区分度。 $\rho = 1/(T_{\max} - T_{\min})$ 表示归一化积分速率。

综合节点在历史时间周期的积分信息,定义到时间周期 t_k 为止,节点 u_i 的总积分值 $\phi_{u_i}^{t_k}$ 如下。

$$\phi_{u_i}^{t_k} = \sum_{k=n_i}^k \frac{e^{-\omega(k-\hat{k})}}{\sum_{\hat{k}=n_i:k} e^{-\omega(k-\hat{k})}} \cdot \gamma_{u_i}^{t_k}, k \geq n_i \quad (2)$$

其中, $e^{-\omega(k-\hat{k})}$ 是周期衰减项, $\omega \in [0, 1]$ 为衰减因子, n_i 表示节点 u_i 在时间周期 t_{n_i} 内首次加入网络。从式(2)中可以看出,节点总积分值 $\phi_{u_i}^{t_k}$ 不仅考虑了节点在当前时间周期 t_k 内获取的积分值 $\gamma_{u_i}^{t_k}$,还将自节点加入网络以来各个时间周期获取的历史积分值纳入考虑范畴,同时利用加权系数项 $e^{-\omega(k-\hat{k})} / \sum_{\hat{k}=n_i:k} e^{-\omega(k-\hat{k})}$ 进行调控。这意味着节点在越接近当前时间周期内所获得的积分值占节点的总积分值 $\phi_{u_i}^{t_k}$ 的比重越大。

2.2.2 节点风险指数

定义 $\mathcal{S}^{t_k}$ 表示到时间周期 t_k 为止已经被审查者屏蔽的网桥构成的集合, $\mathcal{U}_{b_j}^{t_k}$ 表示当前所有已知网桥 b_j 信息的节点集合,则到时间周期 t_k 为止节点 u_i 关于网桥 b_j 的风险指数 $\varphi_{u_i}^{t_k}$ 可由下式计算得到^[18]。

$$\varphi_{u_i, b_j}^{t_k} = \frac{1}{|\mathcal{U}_{b_j}^{t_k}|} \cdot 1_{b_j \in \mathcal{S}^{t_k}} \cdot 1_{u_i \in \mathcal{U}_{b_j}^{t_k}} \quad (3)$$

根据初等概率知识,此时节点 u_i 的风险指数 $\varphi_{u_i}^{t_k}$ 可表示为

$$\varphi_{u_i}^{t_k} = 1 - \prod_{b_j \in \mathcal{S}^{t_k}} (1 - \varphi_{u_i, b_j}^{t_k}) \quad (4)$$

由式(2)和式(4)可知,节点积分值越高,表示该节点上网桥的正常运行时间越长,在一定程度上可以认为该节点更加安全。同时,节点的风险指数越高,表示该节点与越多的网桥屏蔽事件相关,即该节点越有可能为女巫节点。但需注意的是,不能排除有一些正常节点由于多次被分发到了与女巫节点相同的网桥,导致该节点上网桥的正常运行时间较短,获取积分值较低,同时具有较高的风险指数。因此,仅使用当前时间周期内的物理域属性来区分正常节点和女巫节点是不精准的,可能会导致部分正常节点被误检为女巫节点,影响正常用户接入 Tor。

为了消除网桥分发的随机性对于正常节点的影响,在上述物理域属性的评估过程中,充分考虑了历史周期信息对于节点积分值和风险指数的影响。例如,在计算节点总积分值时,综合了各时间周期的历史积分值并采用指数加权系数项进行调控,节点风险指数的计算式则包含了到当前

时间周期为止的所有网桥屏蔽情况。除了将历史信息纳入考量之外,本文进一步引入了节点的社交域属性来解决上述问题。

2.2.3 社交相似度

节点在社交网络中通常拥有各自的属性标签,以新浪微博平台为例,节点的属性标签通常包括年龄、性别、住址、注册时间、职业、粉丝数量、关注用户数量等固有属性。这类属性标签往往是基于长期累积得到的,通常不会在短时间内发生改变,或需要付出较大代价进行修改。

定义向量 $\mathbf{a}_{u_i}^{t_k} = \{a_{u_i,1}^{t_k}, a_{u_i,2}^{t_k}, \dots, a_{u_i,N_a}^{t_k}\}$ 表示到时间周期 t_k 为止,节点 u_i 的属性标签。 N_a 是标签中属性种类数目, $a_{u_i,j}^{t_k} \in [0,1], \forall j \in \{1,2,\dots,N_a\}$ 表示归一化后节点属性值,则此刻节点 u_i 和 u_j 间的社交相似度 $\alpha_{u_i,u_j}^{t_k}$ 可用余弦相似度表示如下^[29]。

$$\alpha_{u_i,u_j}^{t_k} = \cos(\angle \mathbf{a}_{u_i}^{t_k}, \mathbf{a}_{u_j}^{t_k}) = \frac{\mathbf{a}_{u_i}^{t_k} \cdot \mathbf{a}_{u_j}^{t_k}}{\|\mathbf{a}_{u_i}^{t_k}\| \|\mathbf{a}_{u_j}^{t_k}\|} \quad (5)$$

其中, $\|\mathbf{a}\|$ 表示向量 $\mathbf{a}$ 的范数。特别地,节点与自身的社交相似度为 1,即 $\alpha_{u_i,u_i}^{t_k} = 1$ 。进而,到时间周期 t_k 为止,节点 u_i 的社交相似度可以表示为其与网络中所有节点间社交相似度的平均值,即

$$\alpha_{u_i}^{t_k} = \frac{1}{N_u} \sum_{u_j \in \mathcal{U}} \alpha_{u_i,u_j}^{t_k} \quad (6)$$

2.2.4 社交信任度

社交信任度指的是节点在动态行为中表现出作为正常节点的可能性,往往由节点在交互过程中的行为特征决定。以新浪微博平台为例,节点的动态行为主要由转发、点赞、评论、关注等构成。定义 $\mathbf{I}_{u_i,u_j}^{t_k} = \{I_{u_i,u_j,1}^{t_k}, I_{u_i,u_j,2}^{t_k}, \dots, I_{u_i,u_j,N_{u_i,u_j}^{t_k}}^{t_k}\}$ 表示在时间周期 t_k 内,两节点 u_i 与 u_j 之间的交互行为向量。其中 $N_{u_i,u_j}^{t_k}$ 是本周期内两节点间的交互行为次数, $I_{u_i,u_j,j}^{t_k} \in [0,1], \forall j \in \{1,2,\dots,N_{u_i,u_j}^{t_k}\}$ 表示在第 j 次交互行为中,节点 u_i 从 u_j 处获取的社交信任度。

根据节点间的交互行为向量,时间周期 t_k 内,节点 u_i 和 u_j 间的社交信任度 $\delta_{u_i,u_j}^{t_k}$ 表示如下。

$$\delta_{u_i,u_j}^{t_k} = \frac{1}{N_{u_i,u_j}^{t_k}} \sum_{j=1}^{N_{u_i,u_j}^{t_k}} I_{u_i,u_j,j}^{t_k} \quad (7)$$

特别地,节点与自身的社交信任度定义为 1,即 $\delta_{u_i,u_i}^{t_k} = 1$ 。进而,得到在时间周期 t_k 内,节点 u_i 的社交信任度 $\delta_{u_i}^{t_k}$ 如下,即节点 u_i 与网络中所有节点间社交信任度的平均值。

$$\delta_{u_i}^{t_k} = \frac{1}{N_u} \sum_{u_j \in \mathcal{U}} \delta_{u_i,u_j}^{t_k} \quad (8)$$

与节点积分值类似,计算社交信任度时应考虑到历史时间周期上节点的社交信任度对于当前时刻节点社交信任度的影响。因此,到当前时间周期 t_k 为止,节点 u_i 的社交信任度 $\beta_{u_i}^{t_k}$ 如下。

$$\beta_{u_i}^{t_k} = \sum_{k=n_i}^k \frac{e^{-\omega(k-\hat{k})}}{\sum_{\hat{k}=n_i:k} e^{-\omega(k-\hat{k})}} \cdot \delta_{u_i}^{\hat{k}}, k \geq n_i \quad (9)$$

2.3 检测流程

基于上述对节点物理-社交属性评估方法的分析介绍,定义可信度 $\eta \in [0,1]$ 指标表示当前节点为女巫节点的可能性。具体而言,到时间周期 t_k 为止,节点 u_i 的可信度 $\eta_{u_i}^{t_k}$ 可表示为其物理-社交属性的加权平均值,即

$$\eta_{u_i}^{t_k} = \omega_\phi \cdot \phi_{u_i}^{t_k} + \omega_\psi \cdot (1 - \phi_{u_i}^{t_k}) + \omega_\alpha \cdot \alpha_{u_i}^{t_k} + \omega_\beta \cdot \beta_{u_i}^{t_k} \quad (10)$$

其中, $\omega_\phi, \omega_\psi, \omega_\alpha, \omega_\beta \in [0,1]$ 分别表示对应项的权重系数,且 $\omega_\phi + \omega_\psi + \omega_\alpha + \omega_\beta = 1$ 。由定义式可以看出,节点的可信度越低,说明当前节点是女巫节点的可能性越高。

在定义可信度指标的基础上,本文提出了融合社交-物理属性的女巫节点检测机制,所提机制的总体流程如图 3 所示。在时间周期 t_{k+1} 的开始时刻,网桥分发服务将会获取前一时间周期 t_k 内网络中全体节点上的网桥运行情况和社交行为属性,并根据式(1)~式(9)分别计算得到时间周期 t_k 为止,网络中全体节点的积分值、风险指数、社交相似度和社交信任度。进而通过式(10)计算得到节点的可信度。而后,网桥分发服务开始处理并响应时间周期 t_k 内收到的网桥请求。网桥分发服务将会根据历史信息来预测当前网络中女巫节点所占比例 π ,并依据节点的可信度筛选出可能为女巫节点的部分节点,进而按照预设策略为剩余正常节点分发网桥。

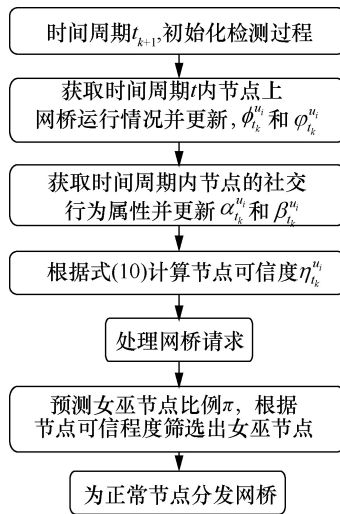


图 3 所提机制总体流程

Figure 3 Overall flow of the proposed mechanism

为实现本文所提检测机制, Tor 网桥分发服务可以在获得正常节点授权后, 调用平台的开放社交数据接口来获取节点社交属性。在实际应用中, 可以通过支持多种主流社交平台账户, 如 Google、Twitter、Facebook 和微博等来保证所提机制的实用性^[18, 30]。同时, 为实现所提女巫节点检测机制, 网络中正常节点的部分隐私数据会被泄露给 Tor 官方服务, 但其泄露范围和泄露对象均有严格边界。泄露范围包括所需的节点静态属性标签和动态行为数据, 泄露对象仅限于 Tor 网桥分发服务。因此, 本文所提机制具备一定的隐私数据保护能力。

3 数值结果分析

3.1 仿真实验设置及对比方法

本节基于 Matlab 平台构建了用于模拟真实 Tor 网络运行情况的离散网络事件模拟器, 来获取 Tor 网络中节点的物理域属性。其参数设置主要参考了 Tor 网络的真实配置^[10, 31]和相关文献中的参数设置^[8, 19]。具体而言, 每次实验中独立生成 1 000 个正常节点和 50 个网桥, 其中女巫节点比例默认设置为 0.05。女巫节点所占比例预测值 π 默认设置为 0.2。为每个节点随机分发 3 个网桥^[31], 每个网桥最多可同时为 40 个节点提供代理服务^[19]。积分时长上下限 $T_{\max}$ 和 $T_{\min}$ 分别为 0.8 和 0.2^[8]。在审查者保守型策略下的时延屏蔽等待时间 T_{wait}^b 服从 $[T_{\min}, 1]$ 上的独立均匀分布^[19]。权重

系数的默认设置为 $\omega_\phi = \omega_\alpha = \omega_\beta = 0.25$ 。将 100 次独立蒙特卡洛仿真实验数据的平均值作为节点物理域属性的输入测试集。

同时, 国家或地区级别的审查者一般具有较为丰富的网络及人力资源, 有能力在国内外主流社交平台上为部署的女巫节点申请社交账户。由于这些社交账户是统一管理和运维的, 它们在不同社交平台上的社交行为特征往往呈现出一定的相似性和关联性。不失一般性地, 本文使用微博实测数据集 MicroblogPCU 来模拟节点的社交域属性。MicroblogPCU 数据集主要由用户信息文件、用户关系文件和用户发表文件构成, 是被广泛用于国内外社会信息学领域相关研究的测试数据集^[23, 25, 32]。

本节通过仿真实验评估了所提机制的识别率和误检率性能, 并与基于二分图和基于最大熵原理的检测机制进行对比。基于二分图的检测机制是 Tor 网络中女巫节点检测的代表性方法^[20-21], 而基于最大熵的检测机制是社交网络中女巫节点检测研究的常见方法^[23, 25], 其相对优势均已在上述相关研究工作中得到验证。此外, 对比了本文所提机制两种退化方法的检测性能, 即仅基于物理属性和仅基于社交属性的检测机制, 同时分析了权重系数设置、女巫节点比例预测值和节点数据混合情况对于所提机制检测性能的影响。

3.2 检测性能对比

审查者激进型和保守型策略下, 网络中女巫节点比例对检测性能的影响分别如图 4 和图 5 所示。识别率性能方面, 基于最大熵原理的检测机制能够实现对所有女巫节点的成功检测, 且对网络中女巫节点比例的变化不敏感。本文所提机制在女巫节点比例较低时, 能够达到 100% 的识别率, 但随着女巫节点比例的增加, 识别率性能缓慢下降。当网络中女巫节点比例增加到 20%, 本文所提机制在两种审查者策略下的识别率分别为 83.6% 和 80.8%。其他 3 种机制在女巫节点比例较低时具有较高的识别率性能, 但随着女巫节点比例的增加, 识别率性能会迅速下降。

误检率性能方面, 基于二分图的检测机制对正常节点的误检概率最低, 不超过 3.71%, 且对网络中女巫节点比例影响不敏感。在其他几类检

测机制中, 本文所提机制具有较低的误检率, 并且会随着网络中女巫节点比例的增加而逐渐降低, 最终接近甚至低于基于二分图的检测机制的误检率性能。在女巫节点比例为 20% 时, 本文所提机制在两种审查策略下的误检率性能分别为 3.28% 和 3.83%。而基于最大熵原理的检测机制由于检测门限设置过于严格, 导致其误检率性能较差, 牺牲了网络中正常节点的服务需求。

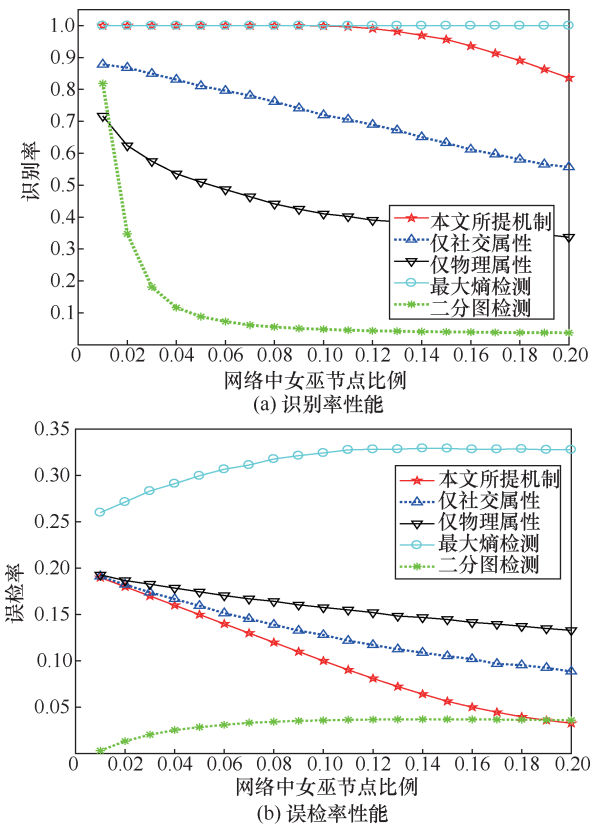


图 4 激进型策略下, 女巫节点比例对检测性能的影响
Figure 4 Impact of Sybil ratio on detection performance under aggressive strategy

3.3 权重系数设置分析

权重系数的设置体现了检测机制运行者 (Tor 网桥分发服务) 对于获取到的节点各方面属性的信任程度。一方面, 由审查者部署的女巫节点能够伪装自身的物理域属性。例如, 采取保守型策略的审查者, 会在等待随机时间后再屏蔽已获取到信息的网桥, 以此使女巫节点的物理行为特征更具有迷惑性, 此时节点的物理域属性可信度较低。另一方面, 节点的社交域属性往往来源于被认为是半可信的社交平台^[30]。因此, 检测机制的运行者需要根据实际情况来判断节点物理域属性

和社交域属性的可信度, 并根据此设置权重系数。若权重系数设置得当, 即与数据的实际可信度一致时, 能够提高检测性能。若权重系数的设置不得当, 即与数据的实际可信度不一致时, 则会损害检测性能。因此, 在缺乏节点物理域和社交域属性数据来源可信度先验知识的情况下, 本文将权重系数默认设置为相等值 0.25。同时, 本文考虑了节点的社交域属性或物理域属性完全不可信两种特殊情况, 以及其他权重系数设置对于所提机制检测性能的影响, 具体的权重系数设置策略及相关说明如表 1 所示。

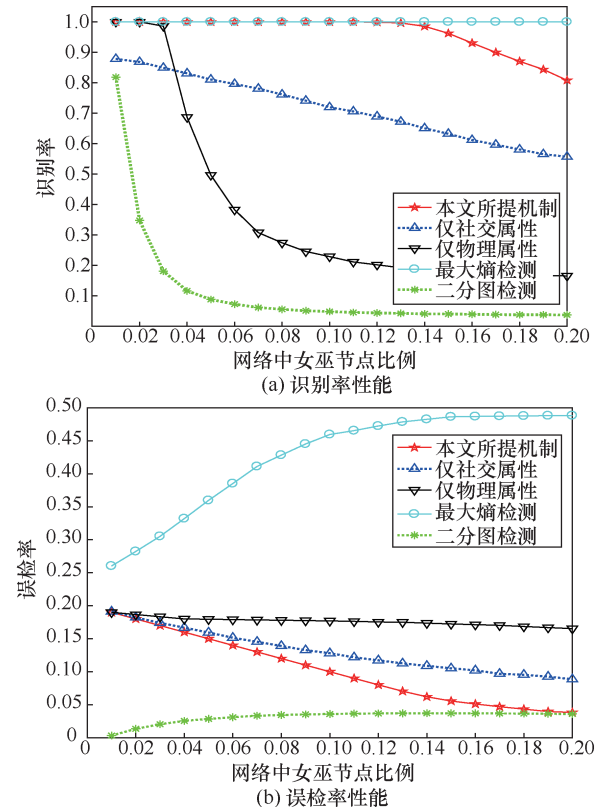


图 5 保守型策略下女巫节点比例对检测性能的影响
Figure 5 Impact of Sybil ratio on detection performance under conservative strategy

表 1 权重系数设置策略及相关说明
Table 1 Weight coefficient setting strategies and descriptions

权重系数设置 ($\omega_\theta, \omega_\phi, \omega_\alpha, \omega_\beta$)	策略代号	说明
(0.25, 0.25, 0.25, 0.25)	A	均衡设置
(0.5, 0.5, 0, 0)	B	仅物理属性
(0, 0, 0.5, 0.5)	C	仅社交属性
(0.4, 0.4, 0.1, 0.1)	D	物理属性增强
(0.1, 0.1, 0.4, 0.4)	E	社交属性增强

通过遍历网络中女巫节点所占比例的预测值 π , 得到审查者激进型和保守型策略下不同权重系数设置策略的接收者操作特征 (ROC, receiver operating characteristic) 曲线, 如图 6 所示。从中可以看出, 融合物理-社交属性的检测机制 (图中 A、D、E) 检测表现始终优于仅物理或仅社交属性检测。说明本文所提机制通过充分挖掘和分析节点行为特征, 能够兼顾识别率和误检率性能, 在实现高效检出女巫节点的同时, 尽可能地降低被误检的正常节点数量。在激进型策略下, 由于审查者在收到女巫节点上报的网桥信息后立即将该网桥屏蔽, 女巫节点的物理行为特征较为明显。因此, 物理属性增强的权重系数设置策略此时检测表现更好, 甚至略微优于均衡设置策略。在保守型策略下, 由于审查者通过延长网桥使用时间和随机化网桥屏蔽时刻使女巫节点的物理行为特征更具有迷惑性, 仅物理属性检测和本文所提融合物理-社交属性的检测机制在保守型策略下的检测性能均有一定损失。但相较于检测表现明显恶化的仅物理属性检测机制, 本文所提机制对审查者的迷惑行为表现出了更好的抵御能力。而仅社交属性的检测机制不依赖于网桥使用时间和屏蔽情况等节点的物理行为特征, 因此, 其检测性能在保守型策略下不会降低。此外, 由于物理属性增强的权重系数设置策略更多地考虑了节点物理行为特征, 相较均衡和社交属性增强两种策略, 其检测性能下降更明显。

根据图 6 计算得到不同权重系数设置策略下所提机制的曲线下面积 (AUC, area under curve) 性能如表 2 所示。从表中可以看出, 在审查者激进型和保守型两类策略下, 融合物理-社交属性的检测机制 AUC 性能均优于仅物理或仅社交属性检测。同时, 相较于仅物理属性检测, 融合物理-社交属性的检测机制对审查者的迷惑行为具有更好的抵御能力。在均衡、物理属性增强和社交属性增强 3 种具体的权重系数设置策略中, 由于更加信任节点的社交行为特征, 社交属性增强的权重系数设置策略对审查者的迷惑行为抵御能力最强, AUC 性能损失仅为 0.57%。

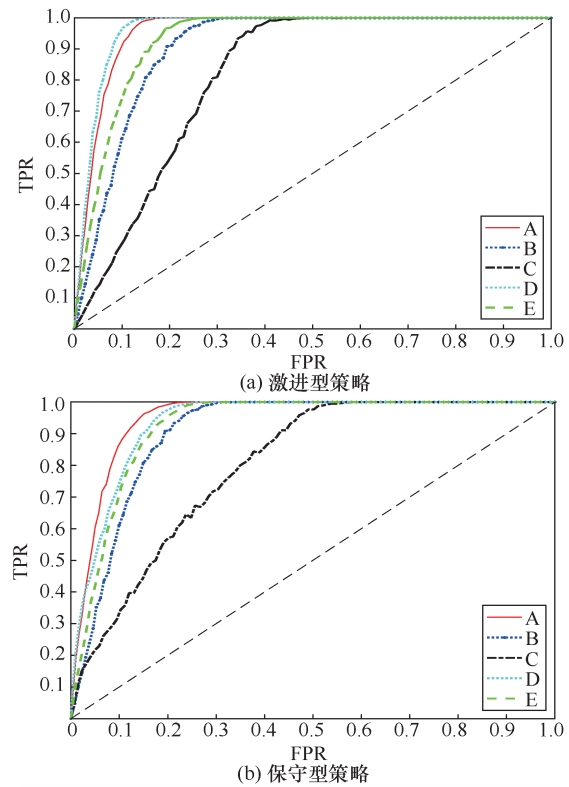


图 6 不同权重系数设置策略下所提机制的 ROC 曲线
Figure 6 ROC curves of the proposed mechanism under different weight coefficient setting strategies

表 2 不同权重系数设置策略下的 AUC 性能对比
Table 2 Comparison of AUC performance under different weight coefficient setting strategies

权重系数设置	激进型策略 AUC 值	保守型策略 AUC 值	性能损失
A (均衡设置)	0.955 5	0.949 2	0.66%
B (仅物理属性)	0.815 5	0.801 5	1.72%
C (仅社交属性)	0.906 1	0.906 1	0
D (物理属性增强)	0.960 1	0.944 9	1.58%
E (社交属性增强)	0.934 0	0.928 7	0.57%

3.4 女巫节点比例预测值分析

本文所提机制的检测性能随女巫节点所占比例预测值 π 的变化曲线如图 7 所示。从图中可以看出, 随着预测值 π 增大, 所提机制的识别率逐渐提高, 最终能够检出全部女巫节点。与此同时, 由于检测门限趋于严格, 所提机制的误检率也会缓慢增加, 但最终不超过女巫节点比例预测值和实际值的差值。例如, 当女巫节点所占比例预测值 $\pi = 20\%$ 时, 误检率不超过 15%。此外, 所提机制在审查者激进型和保守

型策略下的检测表现相近,说明通过合理引入社交域属性能够实现审查者采取的诸如延长网桥使用时间和随机化网桥屏蔽时刻等迷惑行为的有效辨别。

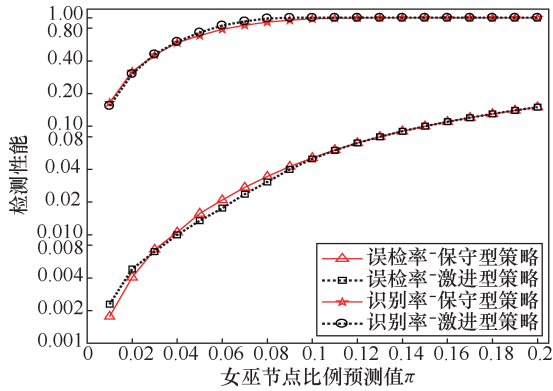


图 7 女巫节点比例预测值对检测性能的影响
Figure 7 Impact of predicted Sybil ratio on detection performance

3.5 节点数据混合情况分析

实际中,可能有部分正常节点没有注册相关社交账户或其社交数据无法获取,这时会存在被检测节点数据混合情况。针对上述现象,可以通过为正常节点提供多种社交平台账户和社交数据选项的方法来尽可能避免网络中过多节点出现数据混合问题。本文将没有注册相关社交账户或社交属性无法获取的节点统称为社交属性缺失节点,将其占全体正常节点的比例表示为 γ ; 进而通过仿真实验验证了网络中社交数据缺失节点比例较低时,所提机制的检测性能仍然具有相对优势,仿真实验结果如图 8 所示。

从图 8 中可以看出,随着网络中社交属性缺失节点比例的增加,本文和基于最大熵的检测机制的检测性能在不断恶化,而由于基于二分图的检测机制不依赖于节点的社交属性,其检测性能不会变化。具体而言,在审查者采取激进型策略且社交属性缺失节点比例 γ 不超过 25% 时,或审查者采取保守型策略且 γ 不超过 15% 时,本文所提机制相较于基于最大熵和基于二分图的检测机制仍然具有优势。而当网络中社交属性缺失节点比例 γ 超过一定值时,本文所提机制的识别率性能会迅速下降,同时其误检率略微增加,但不会超过女巫节点比例预测值 π 。

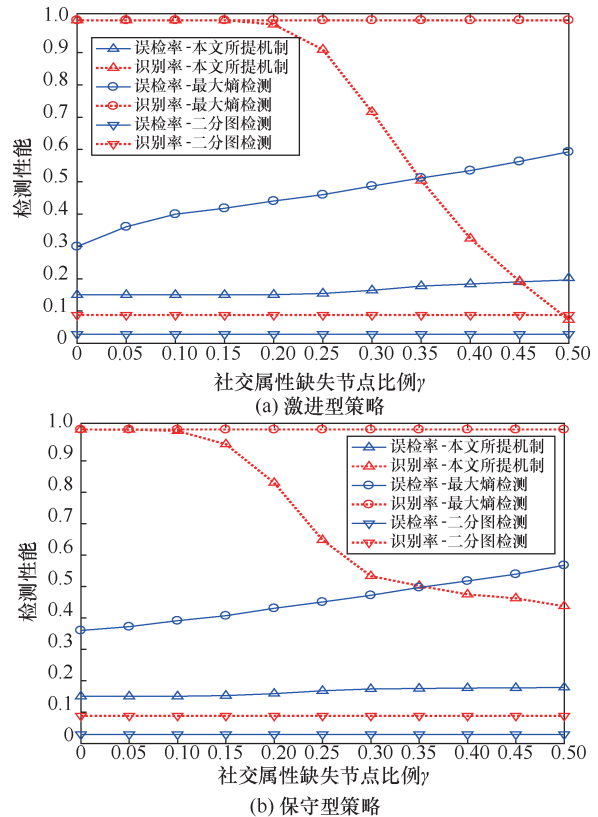


图 8 社交属性缺失节点比例对检测性能的影响
Figure 8 Impact of nodes ratio with missing social attributes on detection performance

4 结束语

作为当前应用范围最广的审查规避系统之一, Tor 在网桥分发过程中面临着严重的女巫攻击问题。为解决这一问题,本文通过充分挖掘节点的行为特征,用节点积分值、风险指数、社交相似度和社交信任度来刻画节点的物理-社交属性,进而定义了可信度指标来表征当前节点为女巫节点的可能性。基于上述工作,设计了融合物理-社交属性的女巫节点检测机制,并在激进型和保守型两类审查者策略下进行了仿真实验。结果表明,所提机制能够有效提高女巫节点检测过程中的识别率性能,降低误检率,同时对于审查者的迷惑行为具有较强抵御能力。本文工作为基于 Tor 协议栈构建的匿名通信系统所面临的女巫攻击威胁提供了可行防御策略。同时,为相关网络审查机构嗅探 Tor 网络中的可用网桥资源、探究更有效的女巫节点伪装方法提供了借鉴思路。在后续研究中,

将进一步考虑女巫节点检测和网桥分发机制相结合的问题,旨在保护网桥信息的同时,为正常节点提供更优质的服务。

参考文献:

- [1] DOUGLAS F E. Circumvention of censorship of internet access and publication[D]. University of Illinois at Urbana-Champaign, 2017.
- [2] FIFIELD D. Threat modeling and circumvention of internet censorship[D]. University of California Berkeley, 2017.
- [3] LEBERKNIGHT C S, CHIANG M, WONG F M F. A taxonomy of censors and anti-censors part II: Anti-censorship technologies[J]. International Journal of E-Politics (IJEP), 2012, 3(4): 20-35.
- [4] TSCHANTZ M C, AFROZ S, PAXSON V. Sok: towards grounding censorship circumvention in empiricism[C]//IEEE Symposium on Security and Privacy. 2016: 914-933.
- [5] NOBORI D, SHINJO Y. VPN gate: a volunteer-organized public vpn relay system with blocking resistance for bypassing government censorship firewalls[C]//11th USENIX Symposium on Networked Systems Design and Implementation. 2014: 229-241.
- [6] ZOLFAGHARI H, HOUMANSADR A. Practical censorship evasion leveraging content delivery networks[C]//ACM SIGSAC Conference on Computer and Communications Security. 2016: 1715-1726.
- [7] NASR M, ZOLFAGHARI H, HOUMANSADR A. The waterfall of liberty: Decoy routing circumvention that resists routing attacks[C]//ACM SIGSAC Conference on Computer and Communications Security. 2017: 2037-2052.
- [8] WANG Q, LIN Z, BORISOV N, et al. rBridge: user reputation based tor bridge distribution with privacy preservation[C]//Network and Distributed Systems Security Symposium. 2013.
- [9] 杨云, 李凌燕, 魏庆征. 匿名网络 Tor 与 I2P 的比较研究[J]. 网络与信息安全学报, 2019, 5(1): 66-77.
YANG Y, LI L Y, WEI Q Z. Comparative study of anonymous network Tor and I2P[J]. Chinese Journal of Network and Information Security, 2019, 5(1): 66-77.
- [10] PORTAL T M. The Tor project[EB].
- [11] 王啸, 方滨兴, 刘培朋, 等. Tor 匿名通信网络节点家族的测量与分析[J]. 通信学报, 2015, 36(2): 84-91.
WANG X, FANG B X, LIU P M, et al. Measuring and analyzing node families in the Tor anonymous communication network[J]. Journal on Communications, 2015, 36(2): 84-91.
- [12] 罗军舟, 杨明, 凌振, 等. 匿名通信与暗网研究综述[J]. 计算机研究与发展, 2019, 56(1): 103-130.
LUO J Z, YANG M, LING Z, et al. Anonymous communication and darknet: a survey[J]. Journal of Computer Research and Development, 2019 56(1): 103-130.
- [13] 杜捷, 何永忠, 杜晔. 基于改进 IPD 质心的 Tor 网络流水印检测方法[J]. 网络与信息安全学报, 2019, 5(4): 91-98.
- DU J, HE Y Z, DU Y. Improved method of Tor network flow watermarks based on IPD interval[J]. Chinese Journal of Network and Information Security, 2019, 5(4): 91-98.
- [14] WINTER P, LINDSKOG S. How the great firewall of China is blocking Tor[C]//USENIX Workshop on Free and Open Communications on the Internet (FOCI). Bellevue, USA, 2012.
- [15] MAHDIAN M. Fighting censorship with algorithms[C]//Internet Conference on Fun with Algorithms. 2010: 296-306.
- [16] ZAMANI M, SAIA J, CRANDALL J. TorBricks: blocking-resistant Tor bridge distribution[C]//International Symposium on Stabilization, Safety, and Security of Distributed Systems. 2017: 426-440.
- [17] MCCOY D, MORALES J A, LEVCHENKO K. Proximax: A measurement based system for proxies dissemination[J]. Financial Cryptography and Data Security, 2011, 5(9): 1-10.
- [18] DOUGLAS F, PAN W, CAESAR M. Salmon: robust proxy distribution for censorship circumvention[J]. Proceedings on Privacy Enhancing Technologies, 2016, (4): 4-20.
- [19] NASR M, FARHANG S, HOUMANSADR A, et al. Enemy at the gateways: Censorship-resilient proxy distribution using game theory[C]//Network and Distributed Systems Security Symposium. 2019.
- [20] GE K, HE Y. Detection of sybil attack on Tor resource distribution[C]//IEEE International Conference on Power, Intelligent Computing and Systems. 2020: 328-332.
- [21] SANATINIA A, NOUBIR G. Honey onions: a framework for characterizing and identifying misbehaving tor HSDirs[C]//2016 IEEE Conference on Communications and Network Security (CNS). 2016: 127-135.
- [22] 康恺, 张颖君, 连一峰, 等. 一种社交网络 Sybil 用户检测方法[J]. 计算机科学, 2016, 43(1): 172-177.
KANG K, YANG Y J, LIAN Y F, et al. Compound approach for sybil users detection in social networks[J]. Computer Science, 2016, 43(1): 328-332.
- [23] 吴大鹏, 司书山, 闫俊杰, 等. 基于行为特征分析的社交网络女巫节点检测机制[J]. 电子与信息学报, 2017, 39(9): 2089-2096.
WU D P, SI S S, YAN J J, et al. Behaviors analysis based sybil detection in social networks[J]. Journal of Electronics and Information Technology, 2017, 39(9): 2089-2096.
- [24] 方晓汾, 方凯, 汪小东, 等. 基于能耗信任值的无线传感器网络 Sybil 攻击检测方法研究[J]. 传感技术学报, 2020, 33(6): 907-915.
FANG F X, FANG K, WANG X D, et al. Sybil attack detection method based on energy consumption trust value in WSN[J]. Chinese Journal of Sensors and Actuators, 2020, 33(6): 907-915.
- [25] WU D, SI S, WANG H, et al. Social influence aware sybil detection in social networks[C]//2017 IEEE/CIC International Conference on Communications in China (ICCC). 2017: 1-4.

- [26] GEDDES J, SCHUCHARD M, HOPPER N. Cover your acks: Pitfalls of covert channel censorship circumvention[C]//ACM SIGSAC Conference on Computer & Communications Security. 2013: 361-372.
- [27] MOHAJERI MOGHADDAM H, LI B, DERAKHSHANI M, et al. Skypemorph: protocol obfuscation for Tor bridges[C]//ACM Conference on Computer and Communications Security. 2012: 97-108.
- [28] SMITS R, JAIN D, PIDCOCK S, et al. BridgeSPA: improving tor bridges with single packet authorization[C]//Proceedings of the 10th Annual ACM Workshop on Privacy in the Electronic Society. 2011: 93-102.
- [29] WU D, ZHOU L, CAI Y. Social-aware rate based content sharing mode selection for D2D content sharing scenarios[J]. IEEE Transactions on Multimedia, 2017, 19(11): 2571-2582.
- [30] LIU Z, LIU Y, WINTER P, et al. TorPolice: Towards enforcing service-defined access policies for anonymous communication in the Tor network[C]//2017 IEEE 25th International Conference on Network Protocols (ICNP). 2017: 1-10.
- [31] SYVERSON P, DINGLELINE R, MATHEWSON N. Tor: The second generation onion router[C]//Usenix Security. 2004: 303-320.
- [32] XU G, ZHOU D, LIU J. Social network spam detection based on ALBERT and combination of Bi-LSTM with self-attention[J]. Security and Communication Networks, 2021.



郭云飞（1963- ），男，河南郑州人，信息工程大学教授、博士生导师，主要研究方向为网络空间安全、云安全和电信网安全。



王亚文（1990- ），男，河南郑州人，信息工程大学助理研究员，主要研究方向为拟态防御和云计算。



孙小丽（1992- ），女，河南南阳人，信息工程大学助理研究员，主要研究方向为信息超材料、隐蔽通信和毫米波。

[作者简介]



史鑫（1995- ），男，吉林长春人，信息工程大学博士生，主要研究方向为匿名网络、隐蔽通信和博弈论。



梁浩（1987- ），男，河南郑州人，信息工程大学副研究员，主要研究方向为网络空间安全和主动防御技术。