受控账户检测技术研究①

宋 晨, 王 远, 王利明

(中国科学院信息工程研究所 信息安全国家重点实验室, 北京 100093)

摘 要: 账户是社交网络、即时通信、电子商务等 WEB 应用领域中最常使用的用户标识方法. 已有工作主要针对社交网络的攻击检测展开, 检测对象多为垃圾消息和伪造账户. 由此可见, 现有研究存在检测领域覆盖不全以及检测对象缺少统一描述的问题. 为了更好地进行该领域研究, 首先提出以账户作为研究对象, 依据恶意账户具备由攻击者控制并实施控制的特点, 将该类账户统一定义为受控账户. 其次, 根据受控程度对研究对象进行分类, 并将现有检测方法进行重新划分. 再次, 提出了使用统计学方法进行账户分类的思想, 并在实验部分进行了受控账户的存在性验证. 最后给出该领域问题的相关讨论, 为受控账户的检测提供了新思路.

关键词: 基于账户的应用; 受控账户分类; 异常检测; 幂律分布; 行为分析

Research of Manipulated Account Detection

SONG Chen, WANG Yuan, WANG Li-Ming

(State Key Laboratory of Information Security, Institute of Information Engineering, Chinese Academy of Sciences, Beijing 100093, China)

Abstract: Account has been widely used as the user identifier in WEB applications suck as Social Networks (SN), Instant Massager (IM) and E-Commerce. At present, most researchers focus on the social network attack detection which includes spam and fake accounts detection among SN envrionment. Obviously, current works are unable to cover all related fields and put forward unified description to the study objects. In order to solve these problems better, in this paper, firstly, we propose that the account is the only study object, and define malicious accounts controlled by attackers executing abnormal behaviors as manipulated account. Then, we classify the manipulated account according to the degree of control they suffered so as to categorize the detection methods by different kinds of manipulated account. Moreover, a statistical methods to classify the account have been promoted to verify the existence of manipulated account. Finally, we discussed some challenges and show some prospects about this area.

Key words: account-based application; manipulated account classification; anomaly detection; power-law distribution; behavior analysis

账户是由真实用户申请并由 WEB 应用系统进行分配和保存的使真实用户能够使用该 WEB 应用系统服务的凭证. 账户是 WEB 应用系统进行身份鉴别和访问控制的依据, 不同账户由于其所处角色的不同, 其拥有的权限、实施的行为、造成的影响同样存在差异. 目前基于账户的 WEB 应用系统包括:邮件系统、社交网络、即时通信、电子商务等, 这些应用极大的改变了人们的工作生活方式, 拓宽了沟通渠道、加速了协作进程. 但问题也随之而来, 账户只作为 WEB 应用系统中虚拟的身份标识, 仅在注册过程中依据用户提供的信息完成账户与现实个体之间对应关系的建立, 这种关系依赖于用户所提供信息的真实性, 当真实性不能得到保证时, 这种对应关系也不可信, 即使能够通过额外的身份信息实现账户与用户真实身份的对应,

① 基金项目:广东省省级科技计划(2013B091300019);所级前瞻项目(Y5Z0031105)
收稿时间:2016-03-08;收到修改稿时间:2016-05-12 [doi:10.15888/j.cnki.csa.005493]

也无法避免单账户被多用户同时使用或账户被非授权使用的问题. 因此, 当账户被滥用, 或者应用中充斥大量伪造账户时, 不但用户本身的隐私和利益无法得到保证, WEB 应用本身以及社会舆论也可能面临威胁.

目前, 随着社交网络的兴起, 大多数研究集中在该领域, 但是由于研究多着眼于已知的攻击, 对研究对象缺少统一的分类, 对基于账户的应用是否存在通用的研究方法也缺乏深入探讨, 因此本文旨在基于已有研究成果, 重新定义基于账户应用的研究对象, 并基于研究对象对已有研究方法进行划分, 为基于账户应用的攻击检测问题提供统一的分类标准.

账户本身并无优劣之分, 当正常用户通过账户完成其在应用中的操作时, 不会对应用产生影响, 当恶意用户使用账户完成恶意活动时, 将会对应用带来不可预知的威胁. 因此, 对于基于账户的应用, 由于账户的描述信息、操作过程等信息体现了用户的操作意图, 能够通过对账户相关属性的研究推断出使用该账户的用户的类别, 所以该领域的研究对象应当为账户本身. 本文的研究面向由恶意用户使用的账户, 这类账户有如下性质:

1) 该账户为受控账户, 在其被控制之后, 按照恶意用户的意图来完成恶意行为, 其本身不具有自主发生行为的能力和定义属性的能力;

2) 该账户发出的行为难以确认是否由真实用户所产生.

基于上述描述可知, 该账户的性质与僵尸网络中的被感染主机(bot or zombie)非常类似, 均受控于攻击者(botmaster). 另外, 大量账户通常被少量用户所控制而共同完成一个攻击任务, 其特点与僵尸网络类似, 即大量受控的被感染主机所形成的以发起各种网络攻击为目的的群体[1], 因此将由恶意用户控制使用的账户称为受控账户, 而其组成的群体称为受控账户网络. 该定义方式有以下两方面的优势:

首先, 受控账户是恶意行为的发起者, 基于行为主体的定义方式能够涵盖所有恶意用户通过操作账户所发起的攻击, 较以攻击类型进行划分的方式更能切中问题的本质.

其次, 受控账户从产生到发展是一个变化的过程, 通过对该主体的研究, 能够分析清楚问题出现的原因以及背景, 从而得到问题解决的思路以及手段, 同时能够提炼出该领域的发展轨迹和趋势.

从检测方法角度, 由于受控账户类别的不同, 其通常采用机器学习、数据挖掘等方法对数据进行分析. 然而, 除了以上两种方法之外, 在统计学领域有学者使用统计指标对数据进行分析[2], 该方法希望能够通过统计分布找到数据分类阈值, 之后再分别对不同阈值下的样本进行细致分类和分析, 这种基于统计学的分析方法虽然较少被使用在异常检测领域, 但是在异常账户占比较小以及行为模式变化较大的场景下有较好的效果.

综上, 本文的主要贡献有以下三个方面:

重新描述受控账户检测问题, 从行为主体的角度出发, 对目前检测方法只针对具体已知攻击进行分析的视角进行了补充, 同时按照受控账户受控程度的差异将其进一步划分为 Sybil 账户和 Compromised 账户, 并依此对检测方法进行了描述;

提出采用统计学方法对账户特征分布进行描述, 进而对受控账户进行分析的思路, 一方面能够基于幂律分布的特点得到正常账户与异常账户的阈值, 另一方面通过对异常账户的细致分析能够证明受控账户的存在性;

归纳总结现有技术的基础上讨论了受控账户检测领域已有方法存在的问题, 并提出了一些可能的改进方向, 为该领域的进一步发展做出努力.

本文主要包括以下几个部分:

第二章针对受控账户问题进行分析, 进一步详细阐述受控账户的分类方法以及分类的原因; 第三章和第四章依据不同的分类方法将目前针对受控账户检测的研究成果进行总结; 第五章针对电子银行个人账户的数据采用统计学方法进行了分析, 验证了受控账户的存在性; 第六章总结目前受控账户研究的研究现状, 并提出存在的问题和面临的挑战; 第七章对整个文章进行了总结.

1 受控账户的问题分析

受控账户问题的根源在于目前基于账户的 WEB 应用系统对单用户注册账户的数量无限制, 同时对于通过 WEB 使用账户的用户是否为注册用户本身无法进行控制, 同时更改 WEB 应用系统增加认证因子的做法难以实现, 因此只能够通过其他的技术手段帮助基于账户的 WEB 应用系统找到并剔除或标识出该类账户.

该类账户由恶意用户所控制, 其能够执行的恶意行为包括: 发送恶意链接、传播垃圾(spam)消息、影响评价系统、窃取用户隐私等. 目前, 受控账户依据其受控程度主要分为两种类型: Sybil 账户和 Compromised 账户.

Sybil 账户, 即恶意用户在应用系统中创建并控制的多个账户, Sybil 账户从一开始就被伪造, 是完全受控账户, 因此所有从 Sybil 账户产生的行为均定义为恶意行为, 即使操纵 Sybil 账户的恶意用户尽力去模仿真实用户的操作轨迹, 例如: 与真实用户成为好友、自动执行操作、修改个人信息使之看起来更可信等, 但是由于其意图与真实用户有差异, 因此行为与真实用户不同. 恶意用户使用该类账户实现垃圾消息推送、钓鱼等操作, 这种情况多出现于在线社交网络(Online Social Network 简称 OSN)、论坛、即时通信等应用场景中, 会导致用户体验下降甚至爆发大规模蠕虫病毒等; 在电子商务领域, 恶意用户使用该类账户进行自动转账、查询等操作, 以提高操作效率, 或者通过该类账户来影响商品的推荐排名, 从而误导真实用户的购买行为. 目前针对 Sybil 账户攻击的研究主要集中在 OSN 领域.

Compromised 账户与 Sybil 账户不同, 即恶意用户通过攻击应用服务的数据库服务器或者使用社会工程学手段获取的正常账户, 是部分受控账户. Compromised 账户在起始阶段的行为由正常用户自主发出, 只有当该账户信息被盗取或者账户被攻陷之后, 该账户才会沦为受控账户. 由于 Compromised 账户保存了真实用户大量的信息, 因此操纵这类账户能够为攻击者直接带来更大的利益, 危害也相应更大. 恶意用户使用该类账户进行非用户本人授权的操作, 实现其利益需求, 这种情况除了社交网络、论坛、即时通信等应用场景之外还存在于电子商务领域, 针对于社交网络、即时通信领域, Compromised 账户与其他账户之间存在信任关系, 因此使用 Compromised 账户发动攻击较 Sybil 账户更容易; 针对于电子商务领域, Compromised 账户的危害显然不仅仅是获取个人信息, Compromised 账户能够被恶意用户使用直接获取经济利益.

对于恶意攻击者来说, 无论是 Sybil 账户还是 Compromised 账户, 受控的受控账户的数量越多, 其能够发动攻击的强度越大, 类型越复杂, 检测也越困难. 后面几章内容主要围绕这两类受控账户展开讨论, 将现有的检测方法进行重新梳理, 对受控账户的现状进行全面的描述.

2 Sybil账户检测

2.1 概述

Sybil 账户按照第 2 章的描述定义为由恶意用户创建的伪造账户, 该账户的特点是从账户创建之时起就以从事恶意操作为目的. 由 Sybil 账户发出的攻击称为 Sybil 攻击.

Sybil 这个词最早出现在一本叫做“Sybil”的小说中, 小说描述了一个病人拥有 16 个不同的身份. 2002 年, Sybil 攻击的概念首次由 Docueur 提出[3], 文中指出除非建立一个可信的机构来进行身份认证, 否则 Sybil 攻击将无法阻止. 随后, 首次针对 Sybil 攻击所讨论的应用场景是端对端(Peer to Peer, 简称 P2P)系统[4], 可以说 2008 年之前, 针对 Sybil 账户检测方法的研究一直集中在 P2P 系统中, 从 2008 年开始, 随着 OSN 的兴起, 其中 Sybil 账户数量增加所带来的一系列问题才逐渐开始受到研究者的重视. 据 CNN 以及 NYT 的统计, 截止到 2012 年 8 月 Facebook 存在 8.3 亿的 Sybil 账户[5], 截止到 2013 年 4 月 Twitter 存在 2000 万的 Sybil 账户, 这些庞大的 Sybil 账户群体所带来的潜在威胁不容小觑. 在 OSN 领域, 研究主要集中在如何检测 Sybil 账户, 但是也有部分专门针对垃圾(spam)消息或者垃圾消息推送者(spammer)展开的研究, 考虑到 spam 消息能够作为 Sybil 账户的一个行为属性, spammer 实际就是 Sybil 账户的一种类型, 因此这个问题也划分到本章节进行论述, 而且由于检测方法存在相似性, 因此本文在 Sybil 账户检测部分采用统一的标准对已有研究成果进行归类:

基于社交图谱特征的检测方法, 该检测方法使用系统中账户之间所形成的关系作为检测的依据, 该特征又被称为结构化特征.

基于群体样本行为特征的检测方法, 该检测方法依赖于多数正常账户的统计指标和行为序列特征, 采用机器学习的方法对特征进行训练, 最后使用训练得到的模型对账户进行检测.

2.2 基于社交图谱特征的检测方法

在进行基于社交图谱特征的检测方法归纳之前, 首先对该方法使用的模型进行详细的描述, 由于基于

社交图谱的Sybil检测是最早被使用的方法, 主要应用在分布式系统中, 采用的模型较为简单, 如图 1 所示.

模型使用图论的方法对系统进行抽象:

节点: 表示系统中的一个账户;

边: 表示账户与账户之间建立信任关系;

攻击边: 表示一个Sybil账户与一个正常账户建立的信任关系.

基于上述模型, 常用的基于社交图谱的特征包括以下几种:

1) 节点的度[6], 单个账户与其他账户建立的边的数量.

2) 社交图谱的直径[7], 任意两个账户之间的最长距离.

3) 聚合系数[7], 度量社交网络连接的紧密程度, 设定每个节点的聚合系数为该节点邻居之间实际的连接数量与该节点所有邻居之间最大(即任意两个互联)的连接数量的比值, 整体的聚合系数为所有节点聚合系数的均值.

4) 传导率[8], 该属性被认为是与汇聚时间(mixing time)[9]相关的一个属性, 通俗来讲, 如果两个节点同时为正常节点, 则这两个节点的随机行走路径会快速的汇聚, 但是如果一个节点为 Sybil 账户创建, 由于Sybil 账户创建的节点与正常节点的距离较远, 则将不会在随机行走过程中快速汇聚. 传导率越高, 说明两个节点之间的连通性越好, 汇聚时间越小; 而传导率越低, 说明两个节点之间的连通性较差, 汇聚时间越长.

5) 节点双向连接性, 单个账户与其他账户建立双向连接的数量.

6) 节点邻居特性, 度量测试账户所建立连接的邻居节点的属性, 描述了账户周围社交图谱的局部特性, 通过这些邻居的属性能够反映该测试账户的类别.

7) 居间向心性, 该属性主要度量一个节点作为其所在局部较短路径区域内中心的程度, 即该节点是否总处于较短路径上.

以上七个属性中, 文献[10]的作者详细讨论了前四个属性, 并且证明前三个属性对于正常账户与 Sybil 账户的区分性不好, 而传导率在攻击概率小于 0.01 的条件下表现较好, 而当攻击概率增大的时候效果不明显.

Haifeng Yu 在文献[11]中基于这个模型将这类方法进行了总结. 从文章中可知, 目前的基于社交图谱检测 Sybil 的方法假设正常账户与正常账户之间的联系要紧密于 Sybil 账户与 Sybil 账户之间的联系, 并且系统要求账户通过非在线的形式建立会话密钥, 这就使得 Sybil 账号更加难以与正常账户之间建立信任关系. 因此得到图中 Attack edges 的数量不会随着 Sybil 账户的增加而无限制的增加, 同时攻击边的数量与 Sybil 账户数量的商应该较小.

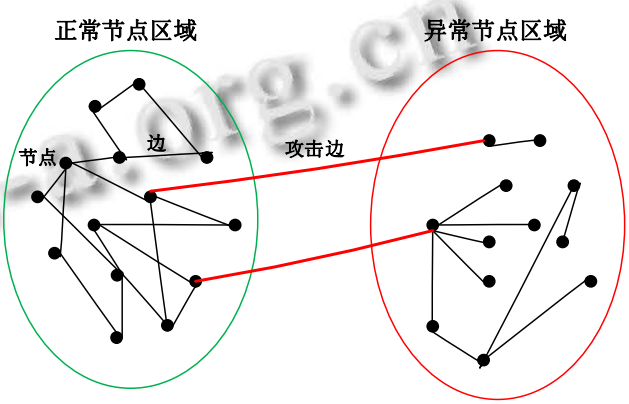

图 1 社交图谱模型示意图

SybilGuard 最早使用这个思想进行 P2P 系统中Sybil 账户的检测[12], 基于上述推论如果找到图的一个割(cut), 该割(cut)能够使得其中的边的数量与分割开的图中的节点的比值较小, 那么就认为这些边为攻击边, 作者使用随机行走的算法寻找商较小的割(cut), 从而得到攻击边. 随机行走算法对于Sybil账户检测的支持在文献[13,14]中进行了详细的说明.

在这个方法提出之后, Haifeng Yu[11]提出了一种改进的系统叫做 SybilLimit, 该方法基于随机行走产生的 mixing time 进行 Sybil 检测, 文中作者认为如果可信节点以汇聚时间为限制进行随机行走所产生的路径与检测节点以相同方式产生的路径之间有交集, 则认为测试节点为正常节点, 否则为异常节点. 这两种方法作为检测 P2P 系统中的 Sybil 账户的基础方法, 在这之后又有一些方法被提出, 这些方法均在这两种方法的基础上进行了改进, 主要包括文献[15-18].

上述工作均基于“正常账户与正常账户之间的联系紧密于异常账户与异常账户之间的联系”, 并且这些工作都需要非常严密的假设进行支撑, 当假设条件不满足的时候, 这些方法将不能很好地发挥作用, 例如, 当选取的种子节点的位置与攻击边的距离非常短,

这种方法会产生较高的误报率[19], 又如, 对于模型中正常边的建立需要使用带外方式, 以期来构建强壮的信任关系, 但是这样的要求本身就不适用于目前的实际环境. 因此该模型对于复杂环境而言过于简单. 而且由于应用背景为 P2P 网络, 对于社交网络中 Sybil 账户的检测没有达到较好的效果, 原因在于社交网络中通信双方不需要会话密钥的交换, 因此该应用场景中 Sybil 账户与正常账户之间不能够满足基于社交图谱方法所提出的假设, 但是基于社交图谱特征的方法仍然为针对社交网络中 Sybil 账户检测的研究提供了帮助.

以上方法使用随机行走算法, 但是该方法的计算效率与账户数量成反比, 为了提高计算效率, Qiang Cao 提出了一种在 OSN 中使用基于社交图谱进行受控账户检测方法[20], 该方法虽采用本类别中的基本假设, 但是能够比之前的系统实现更低的计算资源消耗, 通过使用 Hadoop 中的 MapReduce 分布式计算框架能够支持并行计算. 作者在文中认为, 基于特征的方法由于行为的不确定性导致系统将会产生较多的误报与漏报, 使用人工干预的方法由于自动化程度低并且扩展性差不能够被普遍使用, 因此提出了一种 SybilRank 系统, 该系统作为 OSN 中受控账户检测系统的一部分, 能够弥补上述两种方法所带来的问题. 该系统的基本思想为, 当随机行走路径足够长的时候, 即使从非 Sybil 节点开始进行随机行走过程, 其在 Sybil 节点和在非 Sybil 节点结束的概率相当. 但是当路径较短的时候, 非 Sybil 节点的随机行走过程更倾向于停留在非 Sybil 区域中, 即随机行走不容易遍历到极少的攻击边, 因此作者依据此进行节点等级标注, 当可信节点经过短路径的随机行走路径最终能够到达测试节点的时候, 测试节点的等级较高, 反之, 等级较低, 该方法与以上方法相比较为明显的优势在于, 其在印度的社交网站 Tuenti 中进行了实际的应用, 并且获得了较好的效果.

由于社交网络中使用基于社交图谱特征进行 Sybil 账户检测, 本质上就是使用协同的思想来剥离出异常账户, 但是由于受控账户本身能够通过串通、攻陷等方式来影响协同处理过程, 因此会对检测结果造成一些影响, 针对该问题, Cai[21]提出了一种潜在团体模型, 该模型认为攻击者通过创建一系列受控账户以组成受控账户群从而影响协同推荐结果, 为了掩饰受控账户群, 将在群内节点之间建立任意的网络拓扑结构, 同时攻陷一些正常账户, 使受控账户与正常账户之间建立联系, 但是无论使用何种掩饰方法, 受控账户所组成的团体与其他正常账户所组成的团体之间无连接或者连接较少是受控账户群天然的特征, 该特征能够作为检测受控账户群的一个标识, 因此作者在文中表明只要找到与主要团体通过相对稀少的连接进行关联的团体就能够定位出这些受控账户群. 该方法不同于之前的基于随机行走所采用单个节点进行检测的方法, 而是首先依据账户之间的关联紧密程度将整个社交图谱划分为不同的团体, 接着通过团体中所包含的账户的数量信息、团体内账户之间的连接数量、团体间账户的连接数量进行提取, 依据高斯分布计算出该团体在多维欧氏空间中的位置, 由于异常团体与正常团体之间存在账户数目、账户连接数量的差异, 因此依据该位置能够自动的找到异常的团体, 从而找到受控账户.

由于仅考虑节点的跟随关系在社交网络中不能很好地衡量该节点的类别, 对于社交网络而言, 单个节点不需要进行任何认证就能够跟随另外一个节点, 基于这样的问题有学者借用链接排名的思想进行 Sybil 账户的检测, 链接排名的思想来源于 Google 的 PageRank[22], 该方法的目的是解决搜索引擎的搜索结果排序问题, 其利用网络自身的超链接结构给所有的页面确定一个重要性的等级数, 例如, 当存在页面 P1 和 P2 时, 如果 P1 中包含到 P2 的链接, 那么就认为 P1 对 P2 投出一票, 即增加了 P2 的重要性. 该方法[23]在 Sybil 账户检测中建立双向链接特性, 使用朋友邀请图代替原有的朋友关系图. 其主要的思想概括为:

1) 对于 Sybil 账户, 满足收到的朋友邀请数量少, 发出的朋友邀请数量多, 被拒绝的朋友邀请数量多的条件.

2) 对于正常账户, 细分为两种情况:

i. 对于较为活跃的正常账户, 满足收到的朋友邀请数量多, 发出的朋友邀请数量多, 被拒绝的朋友邀请数量少的条件.

ii. 对于不太活跃的正常账户, 满足收到的朋友邀请数量少, 发出的朋友邀请数量少, 被拒绝的朋友邀请数量少的条件.

基于该思想, 文章完成了 VoteTrust 系统, 该系统能够阻止 Sybil 用户收集其他用户信息, 同时发送大量

邀请请求. 文章在进行检测的同时明确提出“正常账户与正常账户之间的联系紧密于异常账户与异常账户之间的联系”的假设不能够完全支撑社交网络中 Sybil 账户的识别问题, 体现了社交网络结构复杂难于描述的特点. 该方法存在的问题也比较明显, 对于以上特征受控账户都可以不同程度的进行模拟, 从而逃避检测.

之前提到spammer作为Sybil账户的一种类型, 主要的行为是发送spam消息, 目前也有学者使用基于社交图谱的方式来进行spam账户的检测. Song[24]提出了一种通过判断发送者和接受者之间关系的方法来进行spammer 的检测, 作者在文中提出已有的检测方法主要针对 spam 账户特征以及 spam 消息本身, 但是存在两方面问题, 账户特征容易被攻击者所改变, 从而无法被特征检测环节所检出; spam 消息检测只有当 spammer 发生恶意行为之后才能够进行, 当 spammer 不进行任何行为的时候该方法将失效. 基于这两点, 文章使用社交图谱信息作为特征提取的依据, 作者认为社交图谱如前所述较为客观地反映了账户之间的关系, 不容易被攻击者所操纵, 因此文章主要采用的特征包括: 接受者与发送者之间的距离, 使用用户之间的连通需要的边的数量来进行度量, 经过作者的统计, 接收方收到与其距离大于 4 的发送方发来的消息均为spam, 同时研究表明 70.5%的互发消息用户对之间的距离都小于等于 4[25], 因此文中将检测的范围定位为研究距离等于4的用户对之间的社交图谱关系;接受者与发送者之间的连通性, 使用接受者与发送者之间通路的数量作为衡量其之间 连通强弱的标准. 用这两种特征进行 spammer 检测的同时, 该解决方案能够支持实时检测, 解决了检测延迟的问题.

使用社交图谱特征进行 spam 账户检测的文章还包括文献[26,27], 其中文献[26]较为全面的总结了目前spam账户检测用到的基于社交图谱的特征, 包括局部聚合系数、居间向心性、双向连接率, 其中局部聚合系数体现的思想是正常账户通常跟随的对象是其朋友、同事或者家庭成员, 这些账户之间会存在一些联系, 而 spam 账户通常盲目的跟随一些账户, 这些账户之间的联系较为稀疏, 该特征使用数学的方式描述了这种稀疏性; 居间向心性描述了账户是否总是处在最短路径中, 正常账户有选择的跟随一些账户, 而 spam 账户无特定目的, 因此其会创建更多的最短路径, 因此其居间向心性更高; 双向连接率通过衡量账户之间是否互相跟随来判断账户是否为spam账户, 因为正常账户通常不会对来路不明的账户进行跟随, 因此 spam 账户的双向连接率较低. 文献[27]中也使用了局部聚合系数和双向连接率来进行spam账户的检测, 但是文献[27]与文献[26]的不同在于除了提出一些检测方法, 该文章解决的问题是分析恶意账户是如何融入并且在 Twitter 空间中生存的, 即给出了恶意账户如何融入到 twiiter 空间并且在此空间中生存的分析. 从这一小节的分析可以看出, 无论是 P2P 领域还是 OSN 领域, 采用社交图谱特征的最主要的原因是该特征反映了整个网络的结构特性, 并且结构特性理论上较难为被受控账户所更改, 到目前为止, 只有文献[10]对以上提到的部分结构化特征的强度进行了部分论述, 其通过受控账户对改变结构化特征的难易程度来衡量该结构化特征的健壮程度.

除了使用以上提到的特征之外, 文献[26]还提出了使用节点邻居特征的检测方法, 包括平均邻居跟随者、平均邻居推文、节点邻居与邻居跟随者中位数的比值. 其中值得一提的是文中首次提出了使用邻居特征来进行spam账户检测的方法, 这种方法的思想在于利用账户所跟随账户的性质来判断该账户是否为正常账户, 这种方法利用了一些全局性的特征, 因此更难于被 spam 账户所操纵.

2.3 基于群体样本行为特征的检测方法

基于行为特征统计的检测方法主要采用机器学习的思路, 一般步骤分为特征训练和特征匹配两部分. 该方法的关键在于如何选择能够准确表现异常的行为, 并将该行为抽象为特征, 当特征确定后通过机器学习算法对特征进行训练生成模型, 当检测新样本时直接进行匹配得到该样本的匹配结果. 主要采用的算法集中在分类和聚类两种, 这两种方法分别属于监督和非监督类型, 主要区别在于是否进行样本的标记. 分类算法需要对样本进行标记后再进行模型计算, 模型能够给出明确的阈值和样本类型信息, 但是前期预处理工作繁琐. 聚类算法不需要对样本进行标记, 可以直接进行计算, 因此省去了前期与 处理的工作, 但是聚类只是将相似的样本聚合在一起, 并不知道聚合样本的类别信息, 所以需要增加后续处理步骤对聚合样本进行标记. 基于行为特征的检测方法不同于基于社交图谱特征的方法, 更偏重对于主体行为研究而非主体

的结构化关系研究, 因此该方法更有针对性, 但是由于主体行为的变化复杂并且不可预知, 该方法的误判率较高, 目前该方法主要的应用背景为 OSN 系统.

来自 Santa Barbara 大学的 Gang Wang 等学者提出了应用点击流模型在 OSN 中检测受控账户[22]的方法, 该方法将账户行为分为八大类别, 同时对已有 Sybil 账户数据和正常账户数据进行分析, 发现 Sybil 账户的行为集中在朋友邀请等类别的活动, 基于该观察作者将用户行为抽象为点击序列特征和时间间隔序列特征, 通过三种相似度度量的方法计算出序列之间的相似程度并进行聚类计算, 最后通过种子用户标记的方式对聚类结果进行标记, 能够检测出其中的受控账户, 该成果已经被应用于人人网的实际网络环境中.

同样基于人人网的实际数据, Yang[23]等学者不但提出了一种基于特征的受控账户检测方法, 并且通过对实际数据的观察得到了“OSN 中的 Sybil 账户没有形成紧密连接的团体, 相反, 它们恰恰类似正常用户参与到社交图谱中”的结论. 文章采用人人网提供的 1000 个正常账户样本与 1000 个受控账户样本信息用于行为特征的统计, 基于两类样本的统计结果研究者归纳出四种行为特征, 包括: 邀请频度, 一个账户在单位时间内对其他账户发出邀请的次数; 发出邀请接受率, 账户发送邀请被接受的数量与账户发送邀请数量之间的比值; 进入请求接受率, 账户接受的邀请数量与账户收到的邀请数量之间的比值; 聚类系数, 账户朋友之间的相互连通性度量. 基于这四种特征采用支持向量机的分类算法进行训练, 得到每种特征的分类阈值. 实验过程中, 系统对人人网的账户进行实时监控, 如果发现其对应特征的阈值高于分类阈值, 则认为该账户为受控账户, 这种方法能够在保证准确率的同时实现准实时检测.

以上两种方法最大的进步在于相比较之前的检测类别, 其真正在实际的 OSN 中进行测试, 因此实用性较好, 特别是文献[28]中使用了时间序列这个重要的维度对行为进行了描述, 是非常值得肯定的突破. 需要指出的是, 这两篇文章的数据是真实数据, 但是学习数据已经完成标记, 因此能够为后续的处理提供非常重要的依据, 如果是未标记数据, 问题将不会这么容易得到解决.

针对 spam 账户检测, 账户使用自动化工具的行为特征也被研究者进行统计, 包括使用自动化客户端推送推文的比例、使用自动化客户端推送推文中包含 URL 的比例、使用自动化客户端推送相似推文的比例, 这些由行为所转化的统计特征能够反应 spam 账户的一些属性, 但是容易被攻击者所操控从而逃过检测系统的检测[29-31], 例如这种类型的账户可以通过在推文中夹杂其他类别的内容, 或者更换自动化客户端甚至进行手动操作的方法来躲避检测. 不过通过这些特征的分析可以看出, 无论是基于社交图谱特征还是基于统计行为特征, 全局化或者局部化的特征比单个个体的特征在 Sybil 账户检测中能够更好地对恶意样本做出判断.

2.4 小结

基于社交图谱特征的检测方法主要使用用户之间的结构特征来判断用户是否为受控账户, 方法的思想容易理解, 基于图论构建模型具备扎实的理论依据, 但是这种方法依赖于全局信息, 且建立在“正常账户与正常账户之间的联系紧密于异常账户与异常账户之间的联系”的前提条件下, 对个体用户的行为考虑并不充分, 全局结构变化易导致模型失效, 容易被恶意账户所操纵; 基于群体样本行为特征的检测方法使用机器学习的思想来进行受控账户的判断, 方法以正常用户的行为特征为样本进行模型的训练, 并基于此进行受控账户检测, 由于受控账户与正常账户之间行为模式差别较大, 能够得到较好的检测结果, 但是由于样本的准确分类和标记对于结果的影响较大, 因此数据与处理至关重要, 另外目前已经出现了一些逃避检测的方法, 需要不断地改进模型以适应受控账户的检测. 由此可见, 将基于社交图谱的方法与基于群体样本行为特征的方法相结合, 能够首先将账户进行群体划分, 然后对相似群体的特征进行训练, 从而得到更准确的结果.

3 Compromised账户检测

3.1 概述

Compromised 账户按照第 2 章的描述定义为恶意用户通过攻击应用服务的数据库服务器或者使用社会工程学手段获取的正常账户, 其最主要的一个特征就是该账户被攻陷前的行为有较强的一致性(稳定性), 而被攻陷之后该账户会产生与之前行为有显著区别的行为轨迹.

针对这种账户检测的研究比针对于 Sybil 账户检

测的研究少, 但是并不表示该领域研究不重要, 例如, 社交网络账户或者即时通信账户被攻陷之后将会泄露用户信息以及用户的朋友信息, 这种信息无法通过 Sybil 账户获取, 攻击者能够利用朋友之间的信任关系发布恶意链接、传播病毒, 直接危害社交网络以及即时通网络整体的安全[32-34]; 电子商务账户被攻陷之后, 攻击者能够直接伪装成账户所有者进行金融操作, 这些操作均涉及直接经济利益, 对于用户将造成直接的经济损害, 因此针对 Compromised 账户检测的研究有其必要性.

由于 Compromised 账户的特殊性, 对其的检测方法也与以往检测 Sybil 账户的方法存在差异, 之前的工作未对 Sybil 账户以及 Compromised 账户进行区分, 这是因为检测对象主要集中在 spam 消息, 因此无论是 Sybil 账户还是 Compromised 账户均不会影响消息内容本身的检测, 这方面具体包括针对消息内容所包含的 URL 的检测[32,35,36]和针对消息相似性的检测[37], 这种检测方法虽然能够检测出一部分攻击消息从而判断出受控账户, 但是对于消息内容中不包含 URL 的情况以及类似于“Happy Birthday”类型的常用语消息则无法达到预期的效果.

由于检测对象为消息而不是账户, 这些检测方法并没有很好地解决 Compromised 的检测问题, Sybil 账户的检测方法也不能够很好地适用于 Compromised 账户检测, 原因在于 Sybil 账户的检测与 Compromised 账户的检测重点不同, 即 Sybil 只需要对全局的受控账户的行为进行统计学习就能够找到针对于受控账户的行为特征, 而 Compromised 账户的及时发现需要考虑前后行为的差异, 这里需要强调的是 Compromised 账户的及时发现, (因为如果不做这方面的要求, Compromised 账户的检测就能够使用与 Sybil 账户检测相同的方法), 及时发现 Compromised 账户的重要性在于系统能够对账户的所有者发出通知甚至通知与该账户相关的账户, 从而减少和避免严重后果的发生.

基于以上原因, Sybil 账户检测中基于社交图谱和基于统计行为特征的检测方法并不能直接适用于该场景, 因为其中没有考虑账户前后行为的变化信息. 另外, 由于服务提供者不能够轻易地删除或者限制此类的账户, 并且在 OSN 领域如果使用该账户作为跳板从而增加攻击边的数量, 也会对 Sybil 账户的检测带来一定的困难.

目前, 对于 Compromised 账户的检测主要采用基于单样本行为特征统计的检测方法, 同时检测的对象集中在 OSN 领域, 以下对这些方法进行详细的总结.

3.2 基于单样本行为特征的检测方法

与 Sybil 账户检测中基于统计行为特征的检测流程相类似, 对于 Compromised 账户的检测同样采用特征提取、模型训练以及样本匹配三个步骤进行, 但是正如 3.1 小节所述, Sybil 账户与 Compromised 账户最主要的不同在于 Sybil 账户从一开始就是受控账户, 而 Compromised 账户是从正常账户变化到受控账户, 因此 Sybil 账户的检测思路集中在如何通过社交图谱或者群体样本行为这种全局性的特征来区分正常样本与异常样本, 而 Compromised 账户的检测思路集中在如何对单个样本的历史行为进行描述从而及时找到异常发生的节点, 所以 Compromised 账户检测必须要引入时间的概念, 因为只有通过引入时间才能明确的捕获账户的变化.

Manuel Egele[38]提出一种检测 Compromised 账户的方法, 该方法不依赖于内容中的 URL 信息, 较之前基于消息内容的方法更准确, 并且是首次专门针对 Compromised 账户的问题提出解决方案. 作者使用账户发送消息的行为进行特征提取, 提出了六种特征, 分别是: 账户活动的时间, 即账户在一天之内发生典型活动的时间; 消息来源, 即提交消息的应用; 消息文本, 即消息所使用的语言类别; 消息主题, 即账户经常参与讨论的内容类型; 消息链接, 即账户是否发送与其相符合的链接;直接交互用户, 即账户是否与其历史相关的用户进行交互. 定义了这六种特征之后, 作者所采用的步骤包括: 对单个账户所发出的消息进行描述, 并采用 SMO[39](最小序列优化算法)对特征分配相应的权重, 经过训练数据得到相应的阈值;将账户当前的消息样本与阈值比较, 如果发现有特征超出阈值, 则认为该账户的行为出现问题, 但是这时并不立刻对账户进行标注, 而是通过相似消息查找的方法, 判断该消息是否存在被大量传播的情况, 从而判断该账户是否已经被攻陷. 文章基于上述思想实现了 COMPA 系统, 该系统在两大社交网络 Twitter 和 Facebook 公开的数据中进行测试, 得到较好的效果.

在 Compromised 账户检测的基础上, Gianluca Stringhini[40]针对攻陷具有高知名度账户的攻击问题提出了一套检测方法, 针对高知名度账户的攻击主要目

的是利用其在某个领域的影响力来传播虚假或者错误的信息[41,42], 该文章同时也提到不仅仅在 OSN 领域存在这个问题, 在金融领域更需要防范这样的攻击. 文中指出, 对于高知名度账户的攻陷检测主要依赖于行为特征, 因为这些账户的前后行为均保持一致, 研究的对象同样为账户的发送行为. 作者将发送行为抽象为六种特征, 分别包括: 时间特征, 即推文在一天中发送的时间; 发送源, 即消息发送的接口, 包括标准的 WEB 接口、或者 Twitter 的客户端, 该特征描述用户发送推文所使用的应用; 链接特征, 包括推文包含链接或者网站链接的频率; 语言特征, 推文所使用的语言类别; "#"号标签, 推文中是否包含线索主题的标识; 引用标签, 推文中直接连接其他用户的标签. 通过这些特征可以看到, 该文章与文献[38]所采用的特征基本类似, 但是该文章明确提出了这种 Compromised 账户所产生的行为会对金融领域带来极大的危害, 这是之前文章所未提及的内容.

从以上方法可以看出, 基于单样本行为特征的检测方法主要提取的特征虽然与基于群体样本行为检测的特征存在重叠, 但是其异常分析主要基于单用户当前行为与历史行为的对比, 而不是基于全局特征的统计, 因此这种方法能够对 Compromised 账户的检测产生较好的效果, 然而由于样本数量的限制, 如果需要发现 Compromised 账户, 需要累积较长时间的用户数据, 因此对于数据收集时间和处理方法的要求较 Sybil 账户检测更高.

4 实验及分析

以上两类账户的分析通常采用数据挖掘、机器学习的方法提取异常账户. 如前所述, 针对账户行为分析的问题还有一类方法, 即统计学方法, 该方法通常用来对账户的类别进行划分, 一般在特征提取之后, 采用累计分布函数(2)或互补累计分布函数(3)对量化后的特征进行处理, 并对自变量和因变量进行双对数变换, 从而得到该组特征的整体分布情况, 采用的公式如下所示.

$$\Pr(a \le X \le b) = \int_a^b f_X(x)dx. \quad (1)$$

$$F_X(x) = P(X \le x). \quad (2)$$

$$\overline{F}(x) = P(X > x) = 1 - F(x). \quad (3)$$

这种方法适用于正常账户户占多数, 并且特征较为一致, 异常账户户占少数, 但是行为模式复杂的情况, 能够通过特征的统计分布情况得到区分正常与异常的阈值区间, 为异常行为模式的细致分析提供基础. 受控账户其样本数量少, 为了躲避检测通常其行为模式复杂, 与正常账户的使用模式存在较大差异, 适合采用上述方法进行阈值区间的确定, 本小节即通过该方法对受控账户的存在性进行验证.

4.1 数据描述

本文使用电子银行生产环境的真实数据进行处理, 该数据跨度为 12 天, 一共 101833505 条记录, 通过分析发现, 该批数据有一部分是通过网络爬虫访问所产生的, 无登录环境, 因此将该部分数据过滤掉, 另外, 一些带有系统错误信息的数据也被过滤掉, 最终我们得到了 23863321 条记录, 对这些记录依据记录中的会话标识信息进行进一步整理, 最后得到 4983518 个会话, 这些会话包含的用户量是 2116841.

4.2 存在性分析

考虑到会话长度是反映账户业务使用情况的重要指标, 即会话 ID 相同的记录数量, 因此本文先对会话长度按照公式(3)进行计算, 并对结果进行双对数处理, 如图 2 所示.

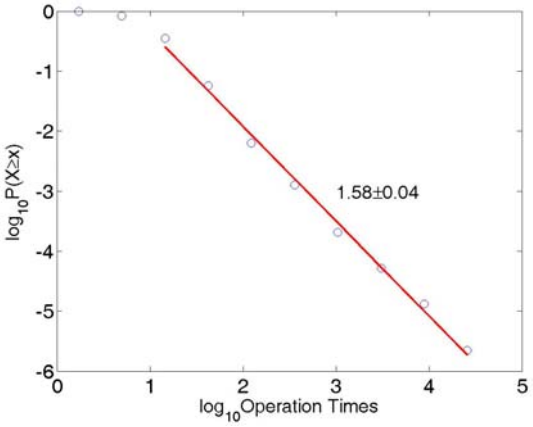

图 2 操作序列长度

该图经过双对数处理之后, 能够拟合为斜率为 1.58 的直线, 依据公式(3)可知, 该批数据符合幂律分布.

进一步的, 将会话长度大于 1000 的会话提取出来, 进行会话中操作的频率进行分析, 其中对所有会话中的一种查询操作的序列进行时间间隔的直方图分析, 分析结果如图 3 所示.

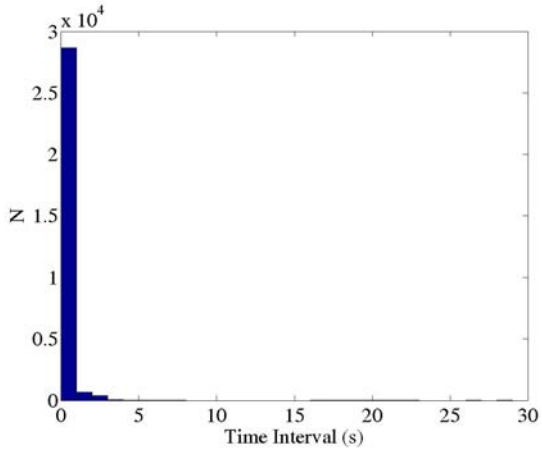

图 3　查询操作序列时间间隔直方图

由用户操作序列长度的分布(α=1.58< 2), 可以得到用户操作长度是一种重尾分布. 通常来说, 用户的操作不会超过一百次, 但是, 由分布可得到较长序列的存在不容忽视. 通过对较长序列的时间间隔的统计分析, 可以看出, 95.89%的操作间隔不大于 1 秒, 且操作次数达到将近 3 万次如图 4 所示, 显然是通过自动化客户端控制账户操作所产生, 属于受控账户的范畴.

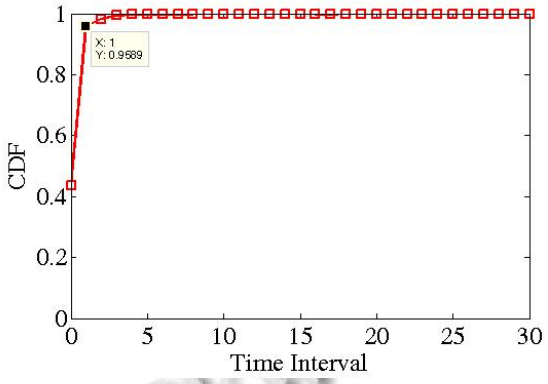

图 4　操作间隔分布示意图

5　讨论

5.1　两类受控账户对比

受控账户即由恶意用户所控制的账户, 恶意用户通过受控账户执行恶意行为, 包括发送 spam 消息、影响推荐排名、窃取用户隐私等. 目前, 受控账户主要分为两种类型:

通过第三章和第四章的描述, 将 Sybil 账户和 Compromised 账户进行对比, 如表 1 所示.

表 1　Sybil 账户与 Compromised 账户的不同

受控账户类型	账户属性	行为特点	检测特征	检测难点
Sybil	完全受控账户	行为均为恶意	社交图谱特征 群体样本行为特征	全局结构描述 行为特征描述
Compromised	由正常账户转为受控账户	初始行为正常, 被盗后为恶意	单样本行为特征	行为特征描述 行为变化点捕获

5.2　存在问题

通过上述分析, 可以总结出目前方法存在以下几方面的问题:

其一, 基于社交图谱特征的方法是否有效. 大多数学者认为结构特征更稳定, 更不易被攻击者所改变, 因此该方法被普遍使用, 而且很多学者依据该模型进行改进以期望达到更好的效果[17,43-45], 但是这种方法存在两方面的不足: 一方面, 这种方法更加适用于 P2P 以及 OSN 的环境, 对于社交关系较弱的领域来说这些方法不能够很好适用. 另一方面, 实际效果差于实验效果, 因为实际的应用场景并不能够由简单的模型所归纳完全, 例如文献[46]在其研究成果中明确指出, 社交网络并没有显现出 Sybil 账户结合更紧密的特征, 又如文献[47]提出, 当攻击边的数量超过一定的阈值, 根据拜占庭理论将无法找到 Sybil 账户. 因此如果需要使用这类方法, 如何建立适用于实际系统的模型是需要考虑的问题.

其二, 行为特征如何与时间维度结合. 在实际应用过程中, 越来越多的学者意识到该领域研究的主体应当为账户本身, 主体所发出的各种行为能够最大限度的反应其是否为受控账户, 因此主体的行为开始被研究者们所关注, 这种特征在实际系统中被证明是有效果的, 而衡量主体行为的一个重要特征维度是时间, 但是只有文献[28]中明确采用时间序列作为特征进行检测, 因此如果能够加入时间维度则能够帮助研究者更加合理的描述账户行为, 也有助于确定 Compromised 账户的攻陷时间.

其三, 模型演进与单样本训练问题. 针对 Compromised 账户检测, 如何对用户的行为进行定时的采样和描述, 从而准确的把握行为变化的时间节点是一个非常重要的问题, 而现有的方法只是使用传统的机器学习的思想进行训练和匹配, 对于何时调整样

本基线没有做出明确的解释, 同时由于每个用户之间的行为均存在差异, 相对于 Sybil 账户检测而言 Compromised 账户检测主要是对单个样本的行为进行判断, 大量样本所获得的模型不一定适用于单个样本, 因此对于单个样本如何建立模型也是一个问题. 对于 Sybil 账户的检测, 已经产生的检测模型如何在运行的过程中动态的适应检测对象的变化, 即现有模型的有效时限也是一个值得验证的问题.

其四, 何种防御最有效果. 如文章的所述, 针对受控账户检测的研究较多, 但是只有基于社交图谱特征的方法中提到了如何利用该方法帮助系统中的节点接受/拒绝一个新生成的请求, 特别是针对 Compromised 账户, 需要研究采取何种措施来缓解或者消除账户被盗所产生的恶劣影响.

5.3 发展趋势

目前, 该领域的新趋势包括:

其一, 随着用户安全防范意识的增强, 大量 P2P 和 OSN 的用户不会再轻易接受或点击陌生人发送的好友邀请和链接, 因此Sybil账户攻击的强度和危害将有所减弱;

其二, 为了获取直接的经济利益, 攻击者开始涉足电子商务领域, 导致针对账户的欺诈攻击、资金被盗、非法洗钱套现等现象频繁发生, 因此金融领域将成为受控账户检测的另一个重要应用背景;

其三, 由少量控制节点所控制的受控账户数量呈现增长趋势, 因此受控账户以及受控账户网络本身能够学习和演进, 从而躲避已有检测系统的检测.

5.4 未来研究方向

基于存在的问题与发展趋势, 本文总结出以下几方面可能的研究方向, 希望能对后来的学者有所借鉴意义:

1) 针对账户模型的研究, 提出更加通用和符合实际的模型来描述该领域的问题, 可以考虑将统计学方法应用于该领域.

2) 针对异常行为特征的提取, 特别是如何使用时间维度对账户的行为以及行为的变化进行描述.

3) 利用海量数据信息结合团体行为对受控账户以及受控账户网络进行检测.

4) 针对金融领域存在的由受控账户(网络)所发起攻击的研究.

6 结语

本文描述了受控账户检测的问题, 对已有Sybil账户以及 Compromised 账户的检测进行总结之后可以得到如下结论:

针对 Sybil 账户检测的研究较多, 已经由起初的 P2P 领域转向 OSN 领域, 检测方法仍旧集中在基于社交图谱方法的研究, 即主要关注账户之间的结构属性, 虽然在基于社交图谱方法的基础上引入链接排名的思想强化了社交图谱特征, 但是这方面的实际成果还较少. 基于群体样本行为特征的方法由于采用机器学习的技术路线而被很多学者认为不适合用来进行 Sybil 账户的检测, 但是也有不少学者尝试使用机器学习方法进行 OSN 中 Sybil 账户的检测, 这些系统在实际应用中被证明有效. 目前虽然还没有将统计学方法广泛应用于账户分类问题的解决中, 但是通过文章的实验表明, 该方法更适用于正常用户数量多、模式固定, 异常用户数量少、模式复杂的场景.

针对Compromised账户的检测研究处于起步阶段, 虽然之前对于 spammer 的研究中已将 Compromised 账户作为其研究对象[32,35,36], 但是相对于 Sybil 账户的检测而言, 该领域研究尚未完全展开, 主要原因包括:

其一, 该领域起初的研究对象为 Sybil 账户以及 Sybil 账户所发出的 spam 消息的检测, 因此并未从 Compromised 账户的角度对带来的危害进行单独研究;

其二, 对比 Sybil 账户创建的难易程度, Compromised 账户权限的获取需要花费较高的代价, 因此研究主要针对数量较大的 Sybil 账户.

由于各种应用环境的复杂性、受控账户本身躲避策略的演进等问题, 如何有效进行受控账户以及受控账户网络的检测仍然需要深入而持续的研究.

参考文献

1 江健,诸葛建伟,段海新,吴建平.僵尸网络机理与防御技术. 软件学报,2012,23(1):82–96.

2 Ding Y, Du Y, Hu YK, Liu ZY, Wang LQ, Ross K, Ghose A. Broadcast yourself: Understanding YouTube uploaders. Proc. of the 2011 ACM SIGCOMM Conference on Internet Measurement Conference. ACM. 2011. 361–370.

3 Douceur JR. The sybil attack. Peer-to-Peer Systems. Berlin Heidelberg: Springer-Verlag, 2002: 251–260.

4 Lian Q, Zhang Z, Yang M, Zhao BY. An empirical study of

collusion behavior in the Maze P2P file-sharing system. 27th International Conference on Distributed Computing Systems, 2007. ICDCS'07. IEEE. 2007. 56–56.

5 CellanJones R. Facebook has more than 83 million illegitimate accounts. BBC News, 2012.8.2. http://www.bbc.co.uk/news/technology-19093078.

6 Barabási AL, Albert R. Emergence of scaling in random networks. Science, 1999, 286(5439): 509–512.

7 Watts DJ, Strogatz SH. Collective dynamics of 'small- world' networks. Nature, 1998, 393(6684): 440–442.

8 Leskovec J, Lang K J, Dasgupta A, Mahoney MW. Statistical properties of community structure in large social and information networks. Proc. of the 17th International Conference on World Wide Web. ACM. 2008. 695–704.

9 Mohaisen A, Yun A, Kim Y. Measuring the mixing time of social graphs. Proc. of the 10th ACM SIGCOMM Conference on Internet Measurement. ACM. 2010. 383–389.

10 Alvisi L, Clement A, Epasto A, lLattanzi S, Panconesi A. Sok: The evolution of sybil defense via social networks. 2013 IEEE Symposium on Security and Privacy (SP). IEEE. 2013. 382–396.

11 Yu H, Gibbons PB, Kaminsky M, Xiao F. Sybillimit: A near-optimal social network defense against sybil attacks. IEEE Symposium on Security and Privacy, 2008, SP 2008. IEEE. 2008. 3–17.

12 Yu H, Kaminsky M, Gibbons PB, Flaxman AD. Sybilguard: Defending against sybil attacks via social networks. ACM SIGCOMM Computer Communication Review, 2006, 36(4): 267–278.

13 Alvisi L, Clement A, Epasto A, Lattanzi S, Panconesi A. Communities, random walks and social sybil defense. [Technical Report] TR-13-04, UTCS, 2013.

14 Andersen R, Chung F, Lang K. Local graph partitioning using pagerank vectors. FOCS. 2006. 475– 486.

15 Danezis G, Mittal P. SybilInfer: Detecting sybil nodes using social networks. NDSS. 2009.

16 Tran N, Li J, Subramanian L, Chow SSM. Optimal sybil-resilient node admission control. Proc. IEEE INFOCOM, 2011. IEEE. 2011. 3218–3226.

17 Tran DN, Min B, Li J, Subramanian L. Sybil-resilient online content voting. NSDI. 2009, 9. 15–28.

18 Lesniewski-Lass C, Kaashoek MF. Whanau: A sybil-proof distributed hash table. 7th USENIX Symposium on Network Design and Implementation. 2010. 3–17.

19 Yang C, Harkreader RC, Gu G. Die free or live hard? empirical evaluation and new design for fighting evolving twitter spammers. Recent Advances in Intrusion Detection. Springer Berlin Heidelberg, 2011: 318–337.

20 Cao Q, Sirivianos M, Yang X, Pregueiro T. Aiding the detection of fake accounts in large scale social online services. Proc. of NSD, 2012: 15–15.

21 Cai Z, Jermaine C. The latent community model for detecting sybil attacks in social networks. Proc. NDSS. 2012.

22 Page L, Brin S, Motwani R, lWinograd T. The PageRank citation ranking: Bringing order to the web[Technical Report]. Stanford InfoLab. 1999.

23 Xue J, Yang Z, Yang X,Wang X, Chen L J, Dai Y F. VoteTrust: Leveraging friend invitation graph to defend against social network Sybils. Proc. IEEE INFOCOM, 2013. IEEE. 2013. 2400–2408.

24 Song J, Lee S, Kim J. Spam filtering in twitter using sender-receiver relationship. Recent Advances in Intrusion Detection. Springer Berlin Heidelberg, 2011: 301–317.

25 Kwak H, Lee C, Park H, Moon S. What is Twitter, a social network or a news media? Proc. of the 19th International Conference on World Wide Web. ACM. 2010. 591–600.

26 Yang C, Harkreader RC, Gu G. Die free or live hard? empirical evaluation and new design for fighting evolving twitter spammers. Recent Advances in Intrusion Detection. Springer Berlin Heidelberg, 2011: 318–337.

27 Yang C, Harkreader R, Zhang J, Shin S W, Gu GF. Analyzing spammers' social networks for fun and profit: A case study of cyber criminal ecosystem on twitter. Proc. of the 21st International Conference on World Wide Web. ACM. 2012. 71–80.

28 Wang G, Konolige T, Wilson C, Wang X, Zheng HT, Zhao BY. You are how you click: Clickstream analysis for sybil detection. USENIX Security Symposium (Washington, DC, 2013). 2013. 241–256.

29 Lee K, Caverlee J, Webb S. Uncovering social spammers: Social honeypots+ machine learning. Proc. of the 33rd

International ACM SIGIR Conference on Research and Development in Information Retrieval. ACM. 2010. 435–442

30 Stringhini G, Kruegel C, Vigna G. Detecting spammers on social networks. Proc. of the 26th Annual Computer Security Applications Conference. ACM. 2010. 1–9.

31 Wang AH. Don't follow me: Spam detection in twitter. Proc. of the 2010 International Conference on Security and Cryptography (SECRYPT). IEEE. 2010. 1–10.

32 Gao H, Hu J, Wilson C, Li ZC, Chen Y, Zhao BY. Detecting and characterizing social spam campaigns. Proc. of the 10th ACM SIGCOMM Conference on Internet Measurement. ACM. 2010. 35–47.

33 Grier C, Thomas K, Paxson V, Zhang M. @ spam: The underground on 140 characters or less. Proc. of the 17th ACM Conference on Computer and Communications Security. ACM. 2010. 27–37.

34 Bilge L, Strufe T, Balzarotti D, Kirda E. All your contacts are belong to us: Automated identity theft attacks on social networks. Proc. of the 18th International Conference on World Wide Web. ACM. 2009. 551–560.

35 Lee S, Kim J. Warningbird: Detecting suspicious urls in twitter stream. Symposium on Network and Distributed System Security (NDSS). 2012.

36 Thomas K, Grier C, Ma J, lPaxson V, Song D. Design and evaluation of a real-time url spam filtering service. 2011 IEEE Symposium on Security and Privacy (SP). IEEE. 2011. 447–462.

37 Gao H, Chen Y, Lee K, lPalsetia D, Choudhary A. Towards online spam filtering in social networks. Symposium on Network and Distributed System Security (NDSS). 2012.

38 Egele M, Stringhini G, Kruegel C, Vigna G. Compa: Detecting compromised accounts on social networks. Symposium on Network and Distributed System Security (NDSS). 2013.

39 Platt JC. Sequential Minimal Optimization: A Fast Algorithm for Training Support Vector Machines[Technical Report]. MSR-TR-98-14.

40 Stringhini G, Egele M. Detecting twitter compromises. The Eighth Annual Graduate Student Workshop on Computing. 2013. 25–26.

41 Manker R. Burger king twitter hack: Bk latest company. chicagotribune.com/2013-02-19/business/ct-burger-king-twitter-hack-0219-20130218_1_tweets-mcdonalds-hack.

42 Siegel E. Fox news politics twitter account hacked, disturbing tweets appear. The huffingtonpost, http://www.huffingtonpost.com/2011/07/04/fox-news-twitter-hacked n889590.html. [2011-07-04].

43 Lesniewski-Lass C, Kaashoek MF. Whanau: A sybil-proof distributed hash table. 7th USENIX Symposium on Network Design and Implementation. 2010. 3–17.

44 Mislove A, Post A, Druschel P, Gummadi PK. Ostra: Leveraging trust to thwart unwanted communication. NSDI. 2008, 8. 15–30.

45 Viswanath B, Post A, Gummadi K P, Mislove A. An analysis of social network-based sybil defenses. ACM SIGCOMM Computer Communication Review, 2011, 41(4): 363–374.

46 Yang Z, Wilson C, Wang X, lGao TT, Zhao BY, Dai YF. Uncovering social network sybils in the wild. Proc. of the 2011 ACM SIGCOMM Conference on Internet Measurement Conference. ACM. 2011. 259–268.

47 Yu H. Sybil defenses via social networks: A tutorial and survey. ACM SIGACT News, 2011, 42(3): 80–101.