(19)中华人民共和国国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 104660594 A
(43)申请公布日 2015.05.27

(21)申请号 201510066831.3

(22)申请日 2015.02.09

(71)申请人 中国科学院信息工程研究所
地址 100093 北京市海淀区闵庄路甲 89 号

(72)发明人 李书豪 云晓春 张永铮

(74)专利代理机构 北京君尚知识产权代理事务所(普通合伙) 11200
代理人 司立彬

(51)Int.Cl.
H04L 29/06(2006.01)

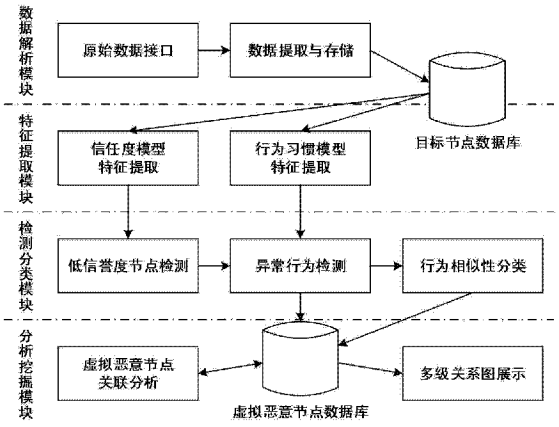
权利要求书2页 说明书8页 附图3页

(54)发明名称

一种面向社交网络的虚拟恶意节点及其网络识别方法

(57)摘要

本发明公开了一种面向社交网络的虚拟恶意节点及其网络识别方法。本方法为:1)从目标社交网络中获取未识别账户的属性数据、行为数据和通信数据;2)对于每一未识别账户,根据提取数据计算得到的信任度模型特征向量计算信誉度;3)对于信誉度低于设定阈值的每一未标识账户,根据其与正常用户的行为习惯统计数据进行对比,判断是否为虚拟恶意节点;4)对虚拟恶意节点集合进行分类并对每个分类结果中的虚拟恶意节点进行关联,形成虚拟恶意节点网络;再利用贝叶斯网络算法进行评估,确定出最终的虚拟恶意节点网络。本发明能够有效识别高伪装的恶意节点,且能够高效的识别出协同类虚拟恶意节点网络。



CN 104660594 A

1. 一种面向社交网络的虚拟恶意节点识别方法，其步骤为：

1) 从目标社交网络中获取未识别账户行为数据和通信数据；

2) 对于每一未识别账户，根据其行为数据建立该账户的好友关系网络，根据其通信数据建立该账户的通信关系网络；然后根据好友关系网络和通信关系网络计算得到的该账户好友出度变化值、好友入度变化值、通信出度向量、通信入度向量，建立该账户的信任度模型特征向量；然后根据该信任度模型特征向量计算该账户的信誉度；

3) 对于信誉度低于设定阈值的每一未标识账户，根据该账户的通信时间序列和消息内容生成该用户的行为习惯统计数据，然后将其与正常用户的行为习惯统计数据进行对比，如果差异大于设定阈值，则将该账户标记为虚拟恶意节点；否则将其标记为正常用户节点。

2. 一种面向社交网络的虚拟恶意节点网络识别方法，其步骤为：

1) 从目标社交网络中获取未识别账户的属性数据、行为数据和通信数据；

2) 对于每一未识别账户，根据其行为数据建立该账户的好友关系网络，根据其通信数据建立该账户的通信关系网络；然后根据好友关系网络和通信关系网络计算得到的该账户好友出度变化值、好友入度变化值、通信出度向量、通信入度向量，建立该账户的信任度模型特征向量；然后根据该信任度模型特征向量计算该账户的信誉度；

3) 对于信誉度低于设定阈值的每一未标识账户，根据该账户的通信时间序列和消息内容生成该用户的行为习惯统计数据，然后将其与正常用户的行为习惯统计数据进行对比，如果差异大于设定阈值，则将该账户标记为虚拟恶意节点；否则将其标记为正常用户节点；

4) 将根据虚拟恶意节点的行为习惯统计数据及其属性数据生成的特征向量作为分类器的输入特征向量，对虚拟恶意节点集合或包含虚拟恶意节点的节点集合进行分类；然后根据社交网络的网络结构和通信关系对每个分类结果中的虚拟恶意节点进行关联，形成虚拟恶意节点网络；再利用贝叶斯网络算法对每一虚拟恶意节点网络进行评估，确定出最终的虚拟恶意节点网络。

3. 如权利要求 1 或 2 所述的方法，其特征在于，所述行为习惯统计数据包括根据通信时间序列生成的通信间隔分布曲线和根据消息内容哈希值列表、通信时间序列生成的账户活跃度曲线。

4. 如权利要求 1 或 2 所述的方法，其特征在于，所述信任度模型特征向量为 $\langle D_i, D_o, D_i(T), D_o(T), M_i(T), M_o(T)\rangle$；其中，$D_i$ 表示好友入度，D_o 表示好友出度，T 表示时间窗口长度，$D_i(T)$ 表示在时间 T 内的好友入度，$D_o(T)$ 表示在时间 T 内的好友出度，$M_i(T)$ 表示在时间 T 内的消息入度，$M_o(T)$ 表示在时间 T 内的消息入度。

5. 如权利要求 4 所述的方法，其特征在于，所述信誉度的计算公式为：

$$R(T)=\frac{D_i}{D_i(T)}\cdot\sum_{N'\in F_{(i,T,N)}}\left(M_i^{N'}(T)-M_o^{N'}(T)\right)$$

R(T) 为信誉度，$F_{(i,T,N)}$ 表示时间区间 T 内给节点 N，即未识别账户，发送通信消息的好友节点集合，$M_i^{N'}(T)$ 表示节点 N' 在时间 T 内的消息入度；$M_o^{N'}(T)$ 表示节点 N' 在时间 T 内的消息出度。

6. 如权利要求 1 或 2 所述的方法，其特征在于，所述属性数据包括：用户标识、用户昵

称、用户关联账户、是否实名认证、账户创建时间、真实身份信息；所述通信数据包括：通信出度、通信入度、通信消息标识、通信消息类型、通信时间、通信内容；所述行为数据包括：用户登陆的 IP 地址、用户所在地、用户上网方式、好友出度、好友入度。

7. 如权利要求 2 所述的方法，其特征在于，所述虚拟恶意节点网络的可疑攻击源识别方法为：首先基于未识别账户的好友关系网络和通信关系网络，以及节点社交活动在时间和空间上的相似性、相关性，分层地从该虚拟恶意节点网络中定位出重要的节点和源头的节点，以及它们之间发动恶意攻击的关系和角色；然后将这些定位出的节点的逻辑地址与地理信息相对应，标识出可疑攻击源。

8. 如权利要求 2 所述的方法，其特征在于，根据确定的所述虚拟恶意节点网络，生成虚拟恶意节点多级关系图，所述多级关系图为基于所述虚拟恶意节点网络中选取的节点或者边进行扩展，围绕其生成新的下一层关系图。

9. 如权利要求 2 所述的方法，其特征在于，所述分类器为决策树算法。

一种面向社交网络的虚拟恶意节点及其网络识别方法

技术领域

[0001] 本发明属于网络信息安全技术领域，涉及网络安全态势感知与处理技术，特别涉及一种面向社交网络的虚拟恶意节点及其网络识别方法。

背景技术

[0002] 随着互联网的发展，社交网络已经成为人们在日常生活工作中进行交流的重要途径和平台。广义上的“社交网络”是指由人类社会活动构成的关系网络，而计算机科学与技术领域所提的“社交网络”是指基于互联网构建的虚拟人类关系网以及相关的网络服务支撑平台，英文全称为“Social Network Site”，本发明所涉及的“社交网络”属于后者。社交网络用户能够通过发布信息，分享资源，实时通信等形式与好友交流；著名的社交网络如Facebook（脸谱）、Twitter（推特）、新浪微博（Weibo）、腾讯微信（WeChat）等。社交网络的发展与流行使人们的社交活动趋于多元化、虚拟化和信息化，社交网络相关信息系统记录着海量用户的身份信息和活动数据，这些资料蕴含着难以估量的价值。很多组织机构已经开始分析利用这些资料，辅助决策，如中国互联网络信息中心（CNNIC）、知微（社交信息情报机构）、中科院网络安全团队（NSTeam）等。

[0003] 人们在享受社交网络服务便利的同时，也面临着严重的安全威胁，而威胁制造者是社交网络中的虚拟恶意节点。此类节点是指由攻击者所控制的社交网络账号，从事多种恶意行为，主要有传播恶意代码，散布垃圾信息，进行网络欺诈、操控话题导向，干扰用户活动等。例如2011年爆发的新浪微博攻击事件，攻击者通过跨站脚本攻击手段，利用虚拟人际关系网快速传播蠕虫病毒，不到1小时至少波及三万用户。攻击者往往出于黑色利益或险恶目的，非法控制大量的虚拟恶意节点，在社交网络中建立庞大的虚假信息源，形成巨大的噪音数据场，破坏网络秩序，危害用户安全，亟待治理和防范。虚拟恶意节点识别技术是有效解决上述问题的防御手段之一，得到了业界和学术界的广泛关注，并取得了一定的研究成果。例如新浪微博的“智能反垃圾系统”能够基于账户信息，自动清除垃圾广告帐号或帐号状态异常的微博帐号；美国德州农工大学的杨超等人针对推特中的垃圾信息发送账号，提出了一种基于关联性节点评估的虚拟恶意节点检测方法，具有较高的准确率。

[0004] 现有的虚拟恶意节点识别技术主要有如下几类：1）基于社交网络账户属性特征的分类识别技术，如根据微博账户的信息完整程度、关注数与粉丝数比值、发帖频率进行分析过滤；2）基于社交网络消息内容的语义分析识别技术，如根据由垃圾信息关键字组成的语义特征向量进行分析检测；3）基于社交网络用户行为模型的异常检测技术，如根据已知正常用户和虚拟恶意节点的账户属性变化和社交行为活动，构建用户行为模型，进而对未知节点进行分类。然而，社交网络中虚拟恶意节点不断演变进化，呈现伪装度高、潜伏期长、隐蔽性好、控制力强、攻击形式多样、角色分工细化等趋势，这导致现有的虚拟恶意节点识别方法适用性和准确率大幅下降，某些种类的恶意节点甚至能够完全绕过现有防御措施。现有的虚拟恶意节点识别方法有以下不足：1）难以识别高伪装的虚拟恶意节点；2）识别精度较好的算法计算复杂度高，检测结果时效性差，例如基于语义分析的检测方法难以在大

规模实时数据系统中应用。

发明内容

[0005] 针对现有社交网络中虚拟恶意节点识别方法检测种类有限，难以发现高伪装、协同类虚拟恶意节点的问题，本发明基于节点信任度模型和行为习惯模型，公开了一种面向社交网络的虚拟恶意节点及其网络识别方法及系统。本发明主要包括以下几个方面：

[0006] (1) 能够识别社交网络中高伪装的虚拟恶意节点。本发明的方法提出了节点信任度模型，并且把社交网络账号难以伪造的属性作为识别特征，解决了针对克隆伪装、感染受控等社交网络节点的检测问题；

[0007] (2) 能够识别社交网络中协同类的虚拟恶意节点网络。本发明的方法提出了社交网络账户的行为习惯模型，基于单位时间内节点行为的异常程度和异常相似性，检测协同类的虚拟恶意节点，并挖掘相似恶意节点间的关系网；

[0008] (3) 能够达到针对大数据检测的准实时性要求。本发明的方法遴选了区分度好的分类特征，并且融合了两种高效的模式识别算法，取长补短，能够满足社交网络虚拟恶意节点检测系统的识别率和实时性需求。

[0009] 本发明的技术方案为：

[0010] 一种面向社交网络的虚拟恶意节点识别方法，其步骤为：

[0011] 1) 从目标社交网络中获取未识别账户行为数据和通信数据；

[0012] 2) 对于每一未识别账户，根据其行为数据建立该账户的好友关系网络，根据其通信数据建立该账户的通信关系网络；然后根据好友关系网络和通信关系网络计算得到的该账户好友出度变化值、好友入度变化值、通信出度向量、通信入度向量，建立该账户的信任度模型特征向量；然后根据该信任度模型特征向量计算该账户的信誉度；

[0013] 3) 对于信誉度低于设定阈值的每一未标识账户，根据该账户的通信时间序列和消息内容生成该用户的行为习惯统计数据，然后将其与正常用户的行为习惯统计数据进行对比，如果差异大于设定阈值，则将该账户标记为虚拟恶意节点；否则将其标记为正常用户节点。

[0014] 一种面向社交网络的虚拟恶意节点网络识别方法，其步骤为：

[0015] 1) 从目标社交网络中获取未识别账户的属性数据、行为数据和通信数据；

[0016] 2) 对于每一未识别账户，根据其行为数据建立该账户的好友关系网络，根据其通信数据建立该账户的通信关系网络；然后根据好友关系网络和通信关系网络计算得到的该账户好友出度变化值、好友入度变化值、通信出度向量、通信入度向量，建立该账户的信任度模型特征向量；然后根据该信任度模型特征向量计算该账户的信誉度；

[0017] 3) 对于信誉度低于设定阈值的每一未标识账户，根据该账户的通信时间序列和消息内容生成该用户的行为习惯统计数据，然后将其与正常用户的行为习惯统计数据进行对比，如果差异大于设定阈值，则将该账户标记为虚拟恶意节点；否则将其标记为正常用户节点；

[0018] 4) 将根据虚拟恶意节点的行为习惯统计数据及其属性数据生成的特征向量作为分类器的输入特征向量，对虚拟恶意节点集合或包含虚拟恶意节点的节点集合进行分类；然后根据社交网络的网络结构和通信关系对每个分类结果中的虚拟恶意节点进行关联，形

成虚拟恶意节点网络；再利用贝叶斯网络算法对每一虚拟恶意节点网络进行评估，确定出最终的虚拟恶意节点网络。

[0019] 进一步的，所述行为习惯统计数据包括根据通信时间序列生成的通信间隔分布曲线和根据消息内容哈希值列表、通信时间序列生成的账户活跃度曲线。

[0020] 进一步的，所述信任度模型特征向量为 $<D_i, D_o, D_i(T), D_o(T), M_i(T), M_o(T)>$；其中，$D_i$表示好友入度，$D_o$表示好友出度，T 表示时间窗口长度，$D_i(T)$ 表示在时间 T 内的好友入度，$D_o(T)$ 表示在时间 T 内的好友出度，$M_i(T)$ 表示在时间 T 内的消息入度，$M_o(T)$ 表示在时间 T 内的消息入度。

[0021] 进一步的，所述信誉度的计算公式为：$R(T)=\frac{D_i}{D_i(T)}\cdot\sum_{N'\in F_{(i,T,N)}}(M_i^{N'}(T)-M_o^{N'}(T))$ R(T) 为信誉度，$F_{(i,T,N)}$表示时间区间 T 内给节点 N，即未识别账户，发送通信消息的好友节点集合，$M_i^{N'}(T)$表示节点 N'在时间 T 内的消息入度；$M_o^{N'}(T)$表示节点 N' 在时间 T 内的消息出度。

[0022] 进一步的，所述属性数据包括：用户标识、用户昵称、用户关联账户、是否实名认证、账户创建时间、真实身份信息；所述通信数据包括：通信出度、通信入度、通信消息标识、通信消息类型、通信时间、通信内容；所述行为数据包括：用户登陆的 IP 地址、用户所在地、用户上网方式、好友出度、好友入度。

[0023] 进一步的，所述虚拟恶意节点网络的可疑攻击源识别方法为：首先基于未识别账户的好友关系网络和通信关系网络，以及节点社交活动在时间和空间上的相似性、相关性，分层地从该虚拟恶意节点网络中定位出重要的节点和源头的节点，以及它们之间发动恶意攻击的关系和角色；然后将这些定位出的节点的逻辑地址与地理信息相对应，标识出可疑攻击源。

[0024] 进一步的，根据确定的所述虚拟恶意节点网络，生成虚拟恶意节点多级关系图，所述多级关系图为基于所述虚拟恶意节点网络中选取的节点或者边进行扩展，围绕其生成新的下一层关系图。

[0025] 进一步的，所述分类器为决策树算法。

[0026] 本发明公开的面向社交网络的虚拟恶意节点识别方法具体内容如下：

[0027] (1) 基于社交网络数据接口或爬虫工具，获取未识别账户的属性数据，行为数据和通信数据，并设定时间窗口，提取该时间区间内的属性变化情况、行为活动情况，以及通信情况，为后续恶意节点网络分类提供数据输入，恶意节点是否呈现相似的属性变化、行为活动和通信行为，将作为判断是否为同一个恶意节点网络的因素。

[0028] 所述的“属性数据”是指社交网络账户的身份标识码、用户性别、创建时间等注册信息和状态信息；所述的“行为数据”是指社交网络账户在使用过程中产生的活动记录，如建立新的好友关系；所述的“通信数据”是指社交网络账户间的通信记录。

[0029] (2) 基于 (1) 的数据，形成未识别账户好友关系网络和通信关系网络，通过解析单位时间内目标账户（即未标识账户）的好友出度变化值、好友入度变化值、通信出度向量、通信入度向量，建立信任度模型特征向量，计算未识别账户的信誉度。其中，根据活动记录建立好友关系网络（即社交网络），根据通信记录建立该账户的通信关系网络；从网络结构

上来讲，通信关系网络是好友关系网络的一部分。

[0030] 所述的“好友出度”是指把目标账户标识为好友的账户数量；所述的“好友入度”是指被目标账户标识为好友的账户数量；所述的“通信出度”是指目标账户发送的通信消息量；所述的“通信入度”是指目标账户接收的通信消息量。

[0031] (3) 基于 (2) 的账户信誉度，根据阈值设定，提取信誉低的可疑账户集合，生成每个可疑账户在时间窗口内的通信时间序列和消息内容哈希值列表，形成账户活跃度曲线和通信间隔分布曲线。基于上述行为习惯统计数据，与正常用户相应数据进行对照，若其中一条曲线存在明显差异，则把可疑账户标记为虚拟恶意节点；若无明显差异，则把可疑账户标记为正常用户节点；从而在每一时间窗口得到一批标注为正常的节点和一批标注为恶意的节点。

[0032] 所述的“行为习惯”是指社交网络账户在行为数据和通信数据上表现出的统计规律。

[0033] (4) 基于 (3) 中虚拟恶意节点的行为习惯统计数据，以及其他属性信息，形成特征向量，利用决策树 (Decision Tree) 算法对所有已经标定的虚拟恶意节点进行分类，形成多个虚拟恶意节点网络；再利用贝叶斯网络 (Bayes Net) 算法对已形成的虚拟恶意节点网络进行评估和调整，进一步提高分类精度，进而挖掘出行为习惯相似的具有协同特征的虚拟恶意节点网络。

[0034] 结合上述虚拟恶意节点识别方法，本发明同时公开了一种面向社交网络的虚拟恶意节点识别系统。本系统主要由数据解析模块、特征提取模块、节点检测模块和分类挖掘模块四部分构成。

[0035] (1) 数据解析模块。本模块基于目标社交网络系统数据接口，能够利用爬虫工具或数据分流方式，实时获取目标账户集合的原始数据，并解析为属性数据、行为数据和通信数据，存入数据库。

[0036] (2) 特征提取模块。本模块基于数据库中的节点属性、行为和通信数据，能够提取并生成节点信任度模型特征向量，目标账户的信誉度，以及其行为习惯模型特征向量。

[0037] (3) 检测分类模块。本模块基于特征提取模块所生成的数据库信息，能够按照阈值设置和正常行为模型对照，对虚拟恶意节点进行标定，并根据其异常行为相似性进行划分。

[0038] (4) 分析挖掘模块。本模块基于节点检测模块标定的虚拟恶意节点集合，能够根据社交网络关系网络和通信网络，对虚拟恶意节点进行关联分析，以事件序列的形式展示攻击脉络，并形成虚拟恶意节点多级关系图，标识可疑攻击源。具体方法为：基于挖掘出的虚拟恶意节点集合构建虚拟恶意节点网络，然后基于未识别账户的好友关系网络和通信关系网络，以及节点社交活动在时间和空间上的相似性、相关性，分层地从一个虚拟恶意节点网络中，定位出重要的节点和源头的节点，以及它们之间发动恶意攻击的关系和角色。然后把这些定位出的节点的逻辑地址与地理信息相对应，比如 IP 地理位置数据库，发送社交网络消息所标注的 GPS 地标信息，可以把虚拟网络中的恶意节点与背后的攻击者关联起来，进而标识可疑攻击源。例如：某个 IP 就是网络水军的一个重要接入点，这个 IP 从事的社交活动就是可疑的概率就增加了。

[0039] 所述的“多级关系图”是指可以基于已有关系图中的某点，或者某条边进行扩展，围绕其生成新的下一层关系图。多级关系图主要是用于可视化分析和展示的，其不是将所

有攻击脉络相关的节点和边都多级展示，而是根据分析者的操作，对指定的节点或边进行指定层次的扩展展示。扩展是根据通信关系网络为基础的，例如：在单位时间内，某个节点和其它 3 个节点进行了通信，那么分析者对于该节点的 2 层关系图展示，就是一个 4 个节点构成的星形网络。

[0040] 本发明公开的方法及系统可以为社交网络系统管理者和安全防御者提供技术支持，帮助他们更好地完成恶意代码入侵阻断、垃圾消息传播监测、网络欺诈识别溯源、舆论环境净化等工作。

[0041] 与现有技术相比，本发明的积极效果：

[0042] 本发明公开了一种面向社交网络的虚拟恶意节点识别方法与系统，有效解决了针对社交网络中虚拟恶意节点的发现与处理问题。与已公开的相关技术相比，本方法及系统预期产生如下积极效果：

[0043] (1) 对于社交网络中具有高伪装特点的恶意账户，能够利用本方法及系统进行有效识别，进而检测出其他方法不易发现的虚拟恶意节点；

[0044] (2) 对于社交网络中具有协同性攻击特征的恶意账户，能够利用本方法及系统挖掘出虚拟恶意节点网络，并且具有较高的准确率。

[0045] (3) 对于社交网络账户海量原始数据处理，能够利用本方法及系统快速提取有效数据特征，并生成高效分类器进行识别，具有良好的实时性。

附图说明

[0046] 图 1：社交网络账户数据获取与初步解析。

[0047] 图 2：节点通信关系示意图。

[0048] 图 3：虚拟恶意节点检测流程图。

[0049] 图 4：系统基本原理框图。

具体实施方式

[0050] 本发明公开的方法根据获取到的社交网络账户数据进行虚拟恶意节点的识别，通过对账户属性特征、行为特征和通信特征的深度分析，判断该账户是否为恶意节点。该方法的主要思想是：在对社交网络账户数据预处理的前提下，通过对账户属性数据和社交网络消息的解析，计算节点信任度模型特征向量和行为习惯模型特征向量，进而利用融合两种具有代表性的模式分类算法进行分析和评估，对可疑节点进行标识，并进一步检测其关联节点，最终得到虚拟恶意节点集合和多种虚拟恶意网络集合。

[0051] 如发明内容相关部分所述，本发明将所公开的虚拟恶意节点识别方法分为四个步骤，接下来将具体阐述。

[0052] (1) 未识别账户相关数据的自动化获取与初步解析。

[0053] 图 1 给出了步骤 (1) 的流程图，具体实施方式如下：

[0054] a) 根据目标社交网络系统的账户数据结构与形式，实现对应的数据接口和爬虫工具，形成原始数据集。本方法适用于以用户交互活动为主的主流社交网络，如 Facebook、Twitter、新浪微博、腾讯微信等；而不适用于以信息发布为主的社交网络，如论坛、贴吧等。

[0055] b) 基于可访问的目标社交网络原始数据集，获取目标账户的属性数据，包括：用

户标识、用户昵称、用户关联账户、是否实名认证、账户创建时间、真实身份信息。

[0056] c) 根据设定的时间窗口，从原始数据集中提取目标账户的行为数据和通信数据。行为数据带有时间点，包括：用户登陆的 IP（互联网协议）地址、用户所在地、用户上网方式、好友出度、好友入度；通信数据包括：通信出度、通信入度、通信消息标识、通信消息类型、通信时间、通信内容、通信时长（可选）、消息评论数（可选）、消息转发数（可选）、超级链接（可选）、链接文件（可选）。需要说明的是，上述通信数据中的可选信息字段是指部分特定社交网络可获取的数据，如新浪微博，具有这些可选信息时对于步骤 (4) 的分类效果会有进一步提升。而非可选信息字段是指所有类型的社交网络中均可获取的数据。

[0057] d) 对上述属性数据、行为数据和通信数据进行格式化处理，保证不同数据源记录格式的一致性，并结构化存储，写入数据库。

[0058] (2) 建立信任度模型特征向量，计算未识别账户的信誉度。

[0059] 步骤 (2) 的具体实施方式如下：

[0060] a) 基于目标社交网络的网络结构获取通信数据，然后根据通信数据构建社交网络节点信任度模型。该信任度模型的建模思想是：节点之间的信任关系是不对称的，通信消息的方向、数量和频率可以表征节点之间的信任程度，而节点的好友出度、好友入度，以及这些参数单位时间内的变化量可以表征该节点的信誉度。对于节点 N，D_i表示好友入度，D_o表示好友出度，T 表示时间窗口长度，$D_i(T)$ 表示在时间 T 内的好友入度，$D_o(T)$ 表示在时间 T 内的好友出度，$M_i(T)$ 表示在时间 T 内的消息入度，$M_o(T)$ 表示在时间 T 内的消息入度，那么，节点 N 的信任度模型特征向量可表示为 $<D_i, D_o, D_i(T), D_o(T), M_i(T), M_o(T)>$。图 2 给出了任意一个用户（标记为 N）的好友关系和时间区间 T 内的通信情况，设 T = 1，那么其信任度模型特征向量可表示为 <4, 3, 4, 2, 8, 7>。

[0061] b) 计算未识别账户的信誉度，基于节点信任度模型建模思想，标记节点 N 在时间区间 T 内的信誉度为 R(T)，其计算公式如下：

[0062] $$R(T) = \frac{D_i}{D_i(T)} \cdot \sum_{N' \in F_{(i,T,N)}} (M_i^{N'}(T) - M_o^{N'}(T))$$

[0063] 其中，$F_{(i,T,N)}$表示时间区间 T 内给节点 N 发送通信消息的好友节点集合。$M_i^{N'}(T)$ 表示节点 N' 在时间 T 内的消息入度；$M_o^{N'}(T)$表示节点 N' 在时间 T 内的消息出度。

[0064] (3) 生成可疑账户通信内容哈希值，并计算通信活跃度和时间间隔分布，与正常用户对照，确定虚拟恶意节点。

[0065] 图 3 给出了步骤 (3) 的流程图，其具体实施方式如下：

[0066] a) 从数据库中提取目标账户的信誉度，根据已标记的样本数据得出信誉度阈值，比较目标账户的信誉度与阈值关系，若低于阈值，则标记为可疑账户。根据计算公式，节点信誉度值越低，说明该点存在异常的可能性越大。

[0067] b) 提取可疑账户的通信内容，利用哈希算法（如 MD5 算法）将每条通信内容映射为较短的固定长度的二进制值，这个二进制值即为“哈希值”，形成对应通信内容的带有时间点（通信内容发生的时间点）标识的哈希值列表。该哈希值列表主要用于虚拟恶意节点网络挖掘中快速比较节点间通信内容的相似程度。

[0068] c) 提取可疑账户的通信时间序列，计算其每日、每周和节假日的活跃度曲线，

以＜时间点，通信量，时间粒度＞为向量格式存入数据库；计算其通信时间间隔，形成以秒为单位的时间间隔集合，生成通信间隔分布曲线。正常用户的通信间隔曲线为重尾分布(Heavy-Tailed Distribution)，可以用 Zeta 分布来表示，其概率密度函数如下：

[0069]

$$f_{Zeta}(x;\tau)=\frac{x^{-\tau}}{\varsigma(\tau)},\tau>1$$

[0070] 其中，$\varsigma(\tau)$是 Riemann Zeta 函数。

[0071] d) 把可疑账户的通信活跃度曲线、通信间隔分布曲线与正常用户的相应曲线进行对照，如果存在明显差异，说明可疑账户的行为习惯存在异常，我们能够将其标识为虚拟恶意节点。

[0072] (4) 融合模式识别分类算法，挖掘协同类虚拟可疑节点以及网络。

[0073] 步骤 (4) 的具体实施方式如下：

[0074] a) 基于数据库中标识为虚拟恶意节点和正常节点的账户属性数据、行为数据和通信数据构造高难伪造的特征向量，作为模式识别分类算法输入的特征向量。

[0075] b) 根据具体需求，确定目标账户集合，可以是所有账户、可疑账户或标记为虚拟恶意节点的账户。选择并融合两种模式识别分类算法，对目标账户集合中的节点进行分类。本方法采用决策树和贝叶斯网络这两种算法对节点进行分析评估，上述两种算法简介和优缺点如下：

[0076] 决策树算法是由一个决策图和可能的结果组成，使得决策树易于理解和实现。相对于其他算法往往要求处理的数据属性的单一性，决策树能同时处理数据型和常规型的属性，而且易于通过静态测试来对模型进行评测，其擅长处理非数值型数据的特点，使得在处理大型数据源时可以做出良好的结果。

[0077] 贝叶斯网络算法是一种概率图型模型，借助于有向无环图得知一组随机变量及其条件概率分配的性质。贝叶斯网络可以根据变量的一些特征就计算出该变量的异常概率，它对于评估复杂变量的不确定性和节点的关联性引起的差异有很大的优势。

[0078] 所谓融合上述两种算法，具体是指：首先基于人工标注的样本集合，利用决策树算法，对所有已经标定的虚拟恶意节点进行分类，形成多个虚拟恶意节点集合，然后根据社交网络的网络结构和通信关系，对每个集合中的恶意节点进行关联，形成虚拟恶意节点网络。接下来，利用贝叶斯网络算法对已形成的虚拟恶意节点网络进行评估和调整，进一步提高分类精度，进而挖掘出行为习惯相似的具有协同特征的虚拟恶意节点网络。

[0079] c) 根据分类结果和模糊度阈值，对节点进行标记，标记相同的节点集合表示这些节点在属性特征、行为习惯和通信特点方面具有相似性，而标记相同的虚拟恶意节点集合即为具有协同性特征的恶意节点网络。例如：传播恶意代码的受感染账户会在很短时间内发送大量含有恶意代码获取方式的相似通信消息，这就是一种典型的协同类虚拟恶意代码网络。

[0080] 本发明公开的面向社交网络的虚拟恶意节点识别系统主要根据本方法的上述四个步骤部署实施，使用 C/C++、Python 语言开发后台程序，使用 Java 语言开发前台界面，使用 MySQL 数据库管理系统搭建相关数据库，本系统涉及到两个数据库：目标节点数据库和

虚拟恶意节点数据库。目标节点数据库用于存储目标社交网络账户的属性数据、行为数据和通信数据；虚拟恶意节点数据库用于存储经过检测分类的虚拟恶意节点相关数据，包括恶意账户信息、恶意节点网络结构信息、相关恶意文件等。

[0081] 图 4 给出了本系统的原理框图，本系统由数据解析模板、特征提取模块、检测分类模块和分析挖掘模块构成，具体描述如下：

[0082] (1) 数据解析模块。本模块涉及目标节点数据库，可分为原始数据接口子模块和数据提取与存储子模块。其中，原始数据接口子模块主要负责调用目标社交网络数据接口和爬虫程序，实现对社交账户原始信息的实时访问；数据提取与存储子模块主要负责实时获取目标账户集合的原始数据，进行初步分析处理，去除噪音数据，并进行分类解析，把目标账户的属性数据、行为数据和通信数据存入数据库。

[0083] (2) 特征提取模块。本模块涉及目标节点数据库，可分为信任度模型特征提取子模块和行为习惯模型特征提取子模块。其中，信任度模型特征提取子模块主要负责基于目标节点数据库中的账户行为数据和通信数据，根据节点信任度模型，生成其特征向量；行为习惯模型特征提取子模块主要负责生成目标节点的活跃度特征和通信时间间隔分布。

[0084] (3) 检测分类模块。本模块涉及虚拟恶意节点数据库，可分为低信誉度节点检测子模块、异常行为检测子模块和行为相似性分类子模块。其中，低信誉度节点检测子模块主要基于特征提取模块输出的数据，根据节点信誉度阈值，对目标账户进行检测，标定出可疑账户；异常行为检测子模块主要基于用户正常行为模块，对照目标节点行为习惯特征数据，并标定存在异常的节点为虚拟恶意节点；行为相似性分类子模块主要负责对虚拟恶意节点以及关联节点进行通信行为相似性划分。

[0085] (4) 分析挖掘模块。本模块涉及虚拟恶意节点数据库，可分为虚拟恶意节点关联分析子模块和多级关系图展示子模块。其中，虚拟恶意节点关联分析子模块主要负责根据虚拟恶意节点关系网络和通信网络，进行关联分析挖掘，以事件序列的形式展示攻击脉络，标定可疑攻击源的账户信息和物理地址信息；多级关系图展示子模块主要根据虚拟恶意节点数据库中的多级关系结构数据，进行可视化展示，基于已显示点或边进行交互式扩展，形成新的关系图。

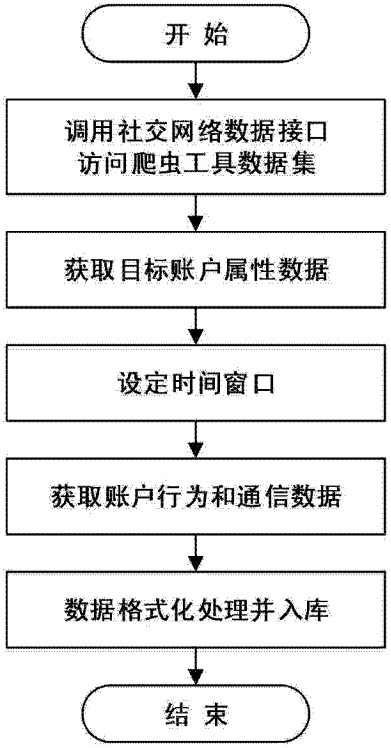

开 始
调用社交网络数据接口
访问爬虫工具数据集
获取目标账户属性数据
设定时间窗口
获取账户行为和通信数据
数据格式化处理并入库
结 束

图 1

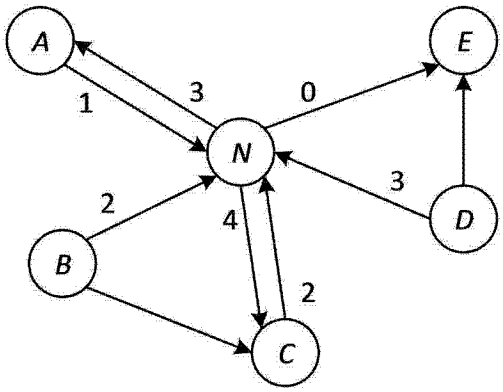

A
E
N
B
C
D
1
3
0
2
4
3
2

图 2

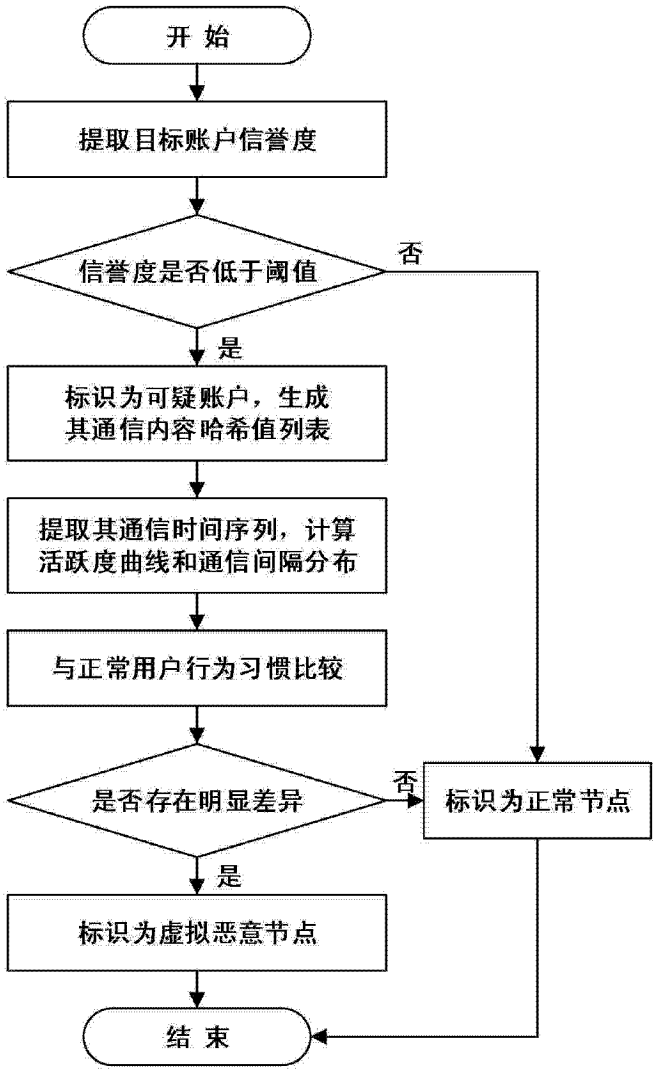

开 始
提取目标账户信誉度
信誉度是否低于阈值
否
是
标识为可疑账户，生成
其通信内容哈希值列表
提取其通信时间序列，计算
活跃度曲线和通信间隔分布
与正常用户行为习惯比较
是否存在明显差异
否
标识为正常节点
是
标识为虚拟恶意节点
结 束

图 3

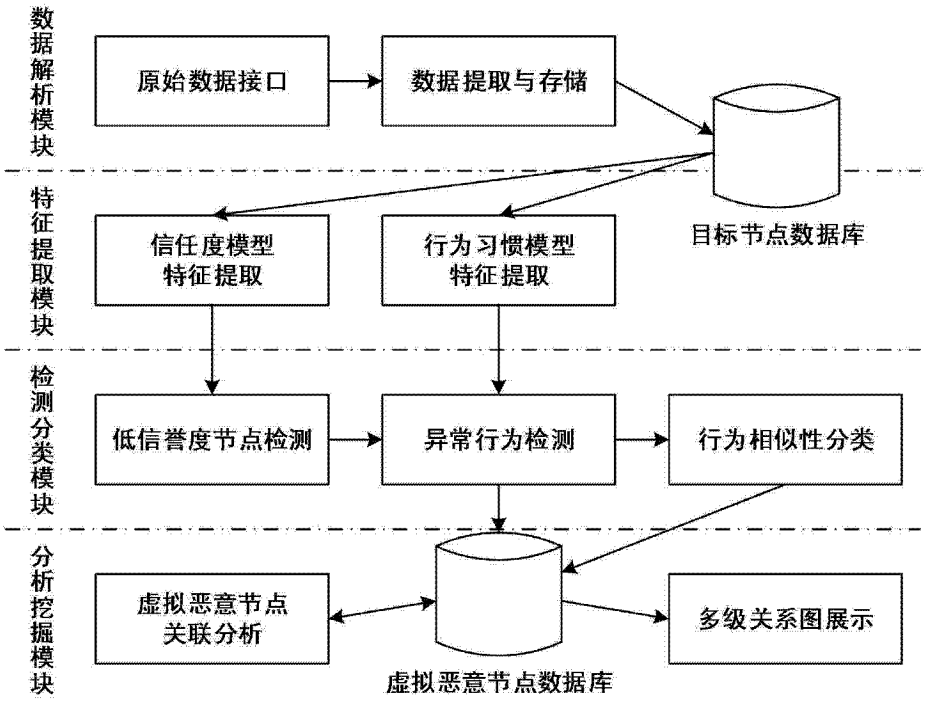

数据解析模块
原始数据接口
数据提取与存储
目标节点数据库
特征提取模块
信任度模型
特征提取
行为习惯模型
特征提取
检测分类模块
低信誉度节点检测
异常行为检测
行为相似性分类
分析挖掘模块
虚拟恶意节点
关联分析
虚拟恶意节点数据库
多级关系图展示

图 4