**基于信任的网络群体异常行为发现**

李 乔 何 慧 方滨兴 张宏莉 王雅山

（哈尔滨工业大学网络与信息安全研究中心　哈尔滨　150001）

**摘　要**　现今大规模网络群体异常事件往往由多个复杂安全事件融合，且这些安全事件之间隐藏着社会化利益与联系，表现出典型的群体性与控制性．对恶意网络群体事件的感知与响应是网络安全管理的重要任务之一．传统的异常检测机制与基于偶图模型的群体异常行为发现方法均未深入分析这些恶意网络行为潜在的社会化关系，且没有考虑交互过程对节点关系的影响．基于此，文中提出一种基于信任的网络群体异常行为发现模型．该模型首先使用网络交互拓扑信息建立网络节点间的信任矩阵；进而结合直接信任度和相关信任度计算网络节点间的相似度，并通过松弛谱聚类算法中的约束条件，增强类数目的自动识别能力，提高节点聚类准确性．实验表明该模型交互能够有效感知网络中的分布式拒绝服务攻击、蠕虫与僵尸网络的异常行为，并对潜伏期内的安全事件行为有较高识别度，同时比基于偶图的行为分类模型具有更高的准确性．

关键词　网络行为；行为聚类；网络信任；信任计算

**中图法分类号** TP394　 **DOI号** 10.3724/SP.J.1016.2014.00001

**Awareness of the Network Group Anomalous Behaviors Based on Network Trust**

LI Qiao HE Hui FANG Bin Xing ZHANG Hong Li WANG Ya Shan

(Research Center of Network and Information Security, Harbin Institute of Technology ,Harbin 150001)

**Abstract** The current large-scale network abnormal event consists of several complex security events which imply inherent social relations, such as groupment and controllability．Perception and response of malicious network incidents is the important task of network security management．The previous network monitoring and detection mechanisms neglect the potential of social relations in these malicious network behaviors．However，the network behavior is the mapping of the social behavior in essence．This paper proposes a novel network group behavior model based on trust through exploring the behavior similarity．This model establishes the trust relationship between the nodes of the network using the network communication topology．Then we use relevant trust concept which is used to increase the trust value between two weak correlation nodes．Based on the network trust matrix，we modify the spectral clustering algorithm to analyze the evolution of the network group anomalous behaviors．Finally，the experiment results show that，our model could detect anomalous behaviors such as denial of service attacks，worm propagations and botnet．Moreover，the classification accuracy of our model is better than the bipartite graph．

**１**　引　言

随着互联网技术的不断进步以及网络用户数量的日益增长，恶意网络行为呈现出复杂化、多样化及智能化趋势．当前已有诸多研究人员对网络行为监测、建模、预警及评估进行深入研究.由于互联网自身的复杂性、异构性与动态性导致在互联网上的各种网络恶意行为日趋隐蔽，同时群体化网络恶意行为可能在某一时间段内在不同空间中进行交叉融合，形成危害较大的网络攻击，对互联网的基础设施或中心服务节点产生巨大威胁．尤其随着僵尸网络技术的提升，更强的受控性、隐蔽性与融合性使得互联网安全所面临的挑战愈加严峻．传统的网络恶意行为监测与发现方法主要通过对网络数据本身进行分析获得，如利用深度包检测技术对网络恶意代码进行检测；分析网络四元组信息（源IP、源端口、目的IP、目的端口）对已知协议漏洞进行分析；利用流数据的特征（如数据包长度、数据包首部大小、流到达时间间隔等）采用机器学习方法对匿名数据流进行分类，挖掘网络主机群体的潜在相似性．然而随着社会工程学在网络攻击中的应用，攻击手段呈现出多元化、智能化趋势．在互联网开放性的分布式异构环境下，对不同的时间段、不同的地理位置所产生的群体网络恶意行为进行监测、追踪、溯源，尤其对于群体攻击行为的潜伏期与发展期的潜在威胁发现与预警是当前网络安全研究热点．以往的网络监测检测机制并没有深入分析这些恶意网络行为潜在的社会化关系，网络信任的提出在现有的网络安全技术基础上增加对网络历史行为过程与交互过程的量化与参考，强化了网络用户之间的交互行为分析.通常而言，常态化网络中用户交互行为存在较大随机性，而对于在某个时空内网络交互呈现突发聚集性或大范围相似性行为，则可标记为异常点，但其所引发的网络流量压力与连接压力对网络基础设施带来一定的风险，对这些具有潜在恶意的非常态化网络事件感知与响应是网络安全管理的重要任务之一．当前的异常检测机制对隐蔽化、智能化及控制性强的网络群体攻击具有滞后性与开销大的局限性，且基于特征及流量异常的方法并没有分析网络行为背后的社会化关系，导致很难在安全事件潜伏期内发现潜在的网络险．

考虑到网络行为本质上是社会化行为在网络中的映射，可采用信任关系对网络行为进行建模，发掘恶意群体行为与常态化行为的区别，及时发现网的异常行为群体，为网络应急提供预警．本文提出一种基于信任的群体行为模型，该模型不仅考虑到历史行为与时间衰减因素，还在此基础上建立信任行为拓扑视图，提出相关信任度描述弱相关网络节点间的潜在联系，从而构造基于信任度的相似度矩阵，并基于该矩阵对网络群体进行谱聚类，以此发现网络中的潜在恶意群体，为网络安全宏观预警提供参考．

本文第２节介绍国内外相关工作；第３节对网络行为进行形式化描述，并给出信任模型与定义;第4节描述相应的行为聚类算法；第５节进行实验分析；第６节总结全文．

**2**　相关工作

网络信任作为社会关系中较为复杂的一点，其不确定性、不对称性、局部传递性、主观性等因素是难以准确对信任度进行量化的根本原因．然而随着P2P(Peer to Peer)与电子商务的兴起，为解决海量用户数据分析推荐，对网络信任的研究逐渐成为当前学术界与业界的关注焦点．信任概念首先应用于分布式系统管理，而后逐步被应用于以互联网为基础的各种大规模分布式应用与服务.早期的信任模型主要针对系统子模块间的运行关系，如建立可信计算模块、依据可信第三方修正系统决策，整个系统主要处于自封闭与特定用户群体的环境．伴随互联网及无线传感网络的兴起，基于网络的开放式大规模分布式应用的使用者大多为未知群体，大量的不确定性因素引发了研究人员对信任概念与信任模型的关注．文献[6]对动态信任进行了定义，并分析了典型的动态信任模型与信任计算方法．

近年来，对信任的研究主要集中于P2P网络与无线传感网络，Huang等人提出一种基于形式化语义的信任值计算方法，该方法采用逻辑语义建立信任关系，并在此基础上对信任值进行量化，为公钥认证与信任决策提供数值化参考；Zhou等人针对P2P网络信任中的声誉系统提出一种基于Gossip的信誉快速聚合方法，该方法就是在*n*个节点的P2P网络中通过)的消息开销有效避免节点的共谋攻击．Li等人提出一种基于反馈聚合的信任计算模型，将时间衰减因子作为自反馈参数对信任加权算法进行动态调整以此提升P2P网络中信任评估的准确性．Erman等人在信任声望管理系统中采用信念传播方式改进全局信任计算的准确性，其通过构造信念关系图模型计算出信念概率矩阵的边缘概率，从而将节点间的信念作为信任传递的参考值．Bao等人提出一种基于概率的信任树模型用于无线传感网络管理，以此提升传感网络中的服务传输能.上述这些模型主要考虑如何在无中心化网络中建立节点的互信关系以及如何量化节点间信任值，没有进一步考虑群体性行为尤其是恶意行为在网络中的演化．

Cai等人将文件流行度与节点信任度进行混合加权，并通过三维归一化方法计算节点活跃度与信誉度，以此降低P２P网络中恶意文件感染传播速度. Carol等人在协同入侵检测系统中部署信任模块，并采用Dirichlet分布作为信念分布模型计算各个检测点间信任值，以此降低内部风险.文献[14]考虑到社交群体中信任团体内部信念值与团体间信念值的差异，提出一种基于偏见主义的随机行走算法构造社交网络拓扑，并以此作为认证的随机算法防范共谋攻击．

对于基于特征与流量异常的网络监测检测技术已较为成熟，现今主要使用图模型对流量进行析，试图发掘主机间的共享关联，文献［17］采用偶图模型对目标网络异常行为进行分类．但尚未发现使用信任模型的相关文献.

３　信任模型构建

**3. 1**　网络行为形式化

随着互联网用户数量与应用的飞速增长，对海量网络数据的处理是影响网络安全监测系统性能的主要因素，但对于数据的检测依赖于恶意攻击的特征，因此通过流量建模、行为建模的方法被应用于大规模网络检测．在宏观层面对网络行为进行监测分析，主要通过关注网络节点的交互过程及交互历史对当前的网络环境进行评估．网络异构性、复杂性、动态性及交互行为的主观性或被动性导致人们无法用准确的度量对网络节点间的交互关系进行刻画，从而采用动态信任关系对当前上下文环境下的网络节点关系进行建模．动态信任关系建模的关键在于

如何更加合理量化影响信任的各个因素．在量化信任度之前，首先需要对网络行为进行形式化描述．

定义 **1**．　交互行为，表示在网络*N*中，节点与节点在*t*时间内的通信动作，其在网络中的表现可以为数据包的发送与接收或网络流的传输．

假设，…，为网络*N*中的*n*个节点，则节点与节点在*T*时间内的交互过程可以表示为时间序列集合 ={ ， ,... }.在某个时刻或或时间段内，通过端到端的节点交互行为可构造出基于交互行为的网络拓扑图*G=(V,E)*（可能*G*不连通），其中边集E=×．如图１所示，交互拓扑分成两个不连通的子图，节点间的边表示两个节点间有直接的交互行为．

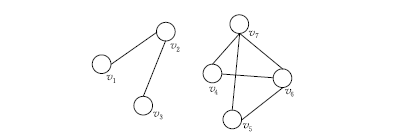


图１　交互拓扑图示例

为了在交互拓扑图中进一步发现隐藏的节点群体行为，文献［17-18］用偶图对网络交互进行刻画，通过对偶图中一侧节点的投影挖掘出另一侧节点的行为相似性．如图２（a）中的偶图，节点与，和另一侧节点均有交互，则在投影图中将，和互连，并根据每个左侧节点所连接的右侧节点交集与并集比例作为投影图中的边权值，如式（１）所示，其中表示节点与的单侧投影图边权值，*NS*表示节点的目标节点集合.该方法对于目标集合重叠的节点聚类具有较好的挖掘效果．然而在计算行为相似度时，单纯依赖节点的节点集合重叠度定义边权值具有很大的局限性．首先该权值并未考虑交互的时长或频率，例如在*t*时间内节点与节点交互１次，与交互10次，而与交互10次，与交互１次，很明显节点和的行为并不相似，然而它们的投影图边权值是１，见图２（b）．其次并没有考虑在目标集合重叠度较低的情况下对构造相似度矩阵的影响，例如蠕虫随机扫描行为，不同节点所扫描的目标地址集合重叠度可能很低，但该时间段内的扫描行为具有极大的相似性．由于交互行为是网络信任的基本属性，利用信任度对权值进行计算，从而构造出基于信任关系的相似度矩阵，改进对相似度的计算．

![C:\Users\aliali\Documents\Tencent Files\3028286534\Image\C2C\EVA54W@7A}S@6GQ32GO]({V.png](data:image/png;base64,)

(a)偶图 (b)单侧投影图

图２　偶图与单侧投影图

= （1）



**3.2** 信任关系模型建立

信任关系在本质上是社会关系的一种，网络信任则是社会行为在网络空间中的映射．与社会信任相似，网络信任具有不确定性、不对称性、时空衰减性与部分传递性等特征．当前多数动态信任模型都采用信任关系图进行描述，区别主要在于对信任值的计算方式，为便于描述信任算法，首先给出一些相关定义．

定义 **2**．　网络信任，是网络空间中的节点与节点的一种交互行为上所表现出的主观性判断．

定义 **3.**　 网络信任度，网络信任的量化值，与信任属性直接相关，是描述网络节点行为间交互程度的定量表示．

虽然信任是一种主观的模糊概念，但通过对特定的上下文环境下的信任属性进行获取与计算，能够对信任进行判断与推测.

定义 **4**． 总体信任度，节点与节点间的信任度可以表示为*F*（,,,,*t*），其中表示与直接交互行为产生的直接信任值，表示与的相关信任值．直接信任与直接交互的次数、间隔、总字节大小相关，节点对的直接信任度计算公式如下：

＝*nor*()+*nor*(*it*)+*nor(siz*) (2)

其中表示交互次数，*it*表示交互平均间隔，*siz*表示交互的总字节大小，*nor*（）表示归一化函数,如*nor*()将节点对的交互次数除以当前的总交互次数，因此一般而言≠,σ是参数权值，Σ＝１，＞０．而相关信任值为相关信任度的量化形式．

3. 2. 1　相关信任度定义与量化

在网络信任模型中，反馈信任与激励函数是增强网络鲁棒性的常用参数，其中反馈信任表示当前节点通过收集其它节点的信息对信任值进行修正，而激励函数则是为降低网络节点恶意行为而进行奖励机制。由于本文所关注的对象是网络节点交互行为，目的在于更准确地挖掘出具有相似行为的群体，因此这两类信任属性并不具有实际的参考价值．与文献［５］的总体信任函数相比，本文增加了相关信任度，定义如下.

定义 5．　相关信任度，与节点的行为相似度相关，描述交互行为相似的节点间信任属性．

相关信任度是描述在弱相关节点间的信任强化因素．在现实社会中，两个人都具有购物的行为，虽然购物的地点大多不

C:\Users\aliali\Documents\Tencent Files\3028286534\Image\C2C\)YNY)%``T_FP9SFI5TX$(SY.png

其中表示节点的邻接向量．余弦定理是数据相似性分析的基本定理，当向量的夹角越小，其相似度越高．根据余弦

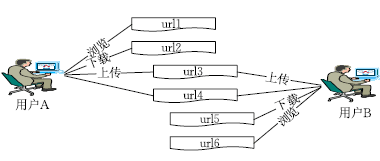


图3　网络访问行为示意

定理可知当为０时，两个向量垂直，即相似度为０，当为１时两向量平行，相似度为１．

由于与维度可能不同，这将引起相关信任度计算出现误差，即维度的缺失降低了一些节点对相关信任度计算的影响．考虑到信任关系中存在部分传递性，可通过信任传递方法对向量进行填充．假设节点与４个节点有交互行为，节点与３个节点有交互行为，如图４所示．需要计算到图中的节点０的传递信任值，从而保证两个节点维度相同，计算公式如下：

= ，0<<1 (4)式（４）中的表示节点i到节点j通过节点k的传递信任值，其值等于节点i对k的直接信任值与节点k对i的直接信任值乘积，α是间接信任调节参数．

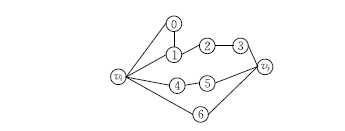


图４　不同维度的节点交互图

由3. 1节可知通过原始交互数据得到的交互矩阵存在不连通现象．因此本文引入虚拟边对交互拓扑图进行连通性恢复．在社会网络中，社团领袖的影响力越高则可能吸引更多的新成员．对于社团领袖的发现，当前已有诸多算法，然而在本文中并不考虑社区中语义、情感等因素，且对数据的提取并不深入到语义层次，因此度是最显著的节点重要性特征．同时在一定时间内，实际网络中度越大的节点表明其活跃度或知名度越高，则越可能吸引新的节点.

定义 **6. 虚拟边，对于交互拓扑图*G*，若图*G*由*m*个不连通子图，…，构成，则表示将子图与（１*i*，*k**m*）中度最大的点相连形成的虚拟边，如度最大的点有多个，则任选其一相连．**

**在修复过程中，对新加入的交互边权值的赋值是影响传递信任的关键性因素．在网络交互拓扑中，不连通子图的结构在一定程度上表现了该子图内节点的交互特征，而节点在子图内的影响力在一定程度上也表现出该节点在图中的地位．为了表示节点在子图内的活跃度，本文定义节点i在子图内的影响力如下：**

C:\Users\aliali\Documents\Tencent Files\3028286534\Image\C2C\JTKZIO(E2O3D{X1KT409RSU.png

其中**表示第*n*个子图内的节点个数，*dgr(i)*表示节点*i*的度**

**显然０＜*IM*１. *IM*越高表明该节点在子图内的活跃度越高，同时也直观表明在子图内的群体行为越集中．因此对于不同子图中领袖节点的*IM*越相近，表明集中性的群体行为越相似，可分为３类情形：（１）两个领袖节点*IM*都高，说明两个子图领袖节点的影响力很高，此时两个子图在未来可能产生交集的概率较大；（２）两个领袖节点*IM*都低，说明两个子图结构较为松散，子图内的交互行为尚未稳定，其可能的产生新的交互概率较高；（３）两个*IM*差异较大，说明这两个子图的交互行为并不相似，则可能产生新的交互概率较低．计算两个子图领袖节点虚拟边权值可表示为**

C:\Users\aliali\Documents\Tencent Files\3028286534\Image\C2C\KELIEXNA1SXLKOZ1SA8]7P2.png

**其中表示领袖节点*i*与领袖节点*j*之间的虚拟边权值，是调节参数，显然当它们差异越小时，越大，在一定程度上表示了上述的３类情形．**

**整个填充算法如算法１所示，首先通过虚拟边对交互拓扑图进行连通性恢复，其次根据先广搜索算法分别计算邻接节点集合中所有与****及****无直连的节点的传递信任值．**

**算法１．　节点交互向量填充算法**BVRCA(Behavior Vector Recovery and Completion Algorithm).

**输入：行为交互图邻接矩*M*，节点****，节点**

**输出：新的邻接向量，**

**１．对交互邻接矩阵*M*进行连通性恢复，得到新邻接矩**

**阵*M*′；**

**２．通过*M*′得到节点****与****的节点邻居集合和**

**；**

**３．计算和的差集；**

**４．对于，采用先广搜索**BFS (Breadth Frist Search) **算法得出****到***n***对其的路径；**

**５．对于，采用先广搜索**BFS (Breadth Frist Search)**算法得出****到***n***对其的路径；**

**６．根据式（４）分别计算出节点****和****对中每个节点的传递信任值，并更新对应的邻接向量；**

**７．**END

3. 2. 2**信任矩阵构建**

**通过交互矩阵的连通恢复与节点交互向量填充，我们可通过式（３）计算出节点****与****的相关信任值，此时总信任值***F*（,,,,*t*）**的计算公式可表示为**

C:\Users\aliali\Documents\Tencent Files\3028286534\Image\C2C\_4`@}%RNH$0OY1~{T8@9IYU.png

（7）**其中是权重参数，且＞０，Σ＝１；***L***是获取交互数据的总次数，可视为距离当前***t***时间段内的总快照次数;表示距离当前时间越近的交互行为价值越高．**

**由于获取的交互拓扑数据只是整个网络交互演变过程中的序列快照，若事先固定的值将导致计算出的总信任值存在较大误差．为了降低误差对行为相似度矩阵的影响，本文采用动态的参数赋值策略对和进行指定．信任的不确定性是影响信任量化准确度的重要因素，而信任值的大小则表示未来交互概率的高低，考虑到不确定性特征在信任中的影响，我们使用信息熵方法对权重进行自动修正，进一步降低信任误差．信息熵是表示事件不确定性的一个重要概念，熵值越低表示不确定性越小，计算公式如下：**

C:\Users\aliali\Documents\Tencent Files\3028286534\Image\C2C\3JFTTB1K41QOFZW6UJS@HY1.png

**则对于直接信任值与相关信任值，分别计算其对应的熵值，例如从交互矩阵中计算得出节点****与****的**=0.7，=0.5，对应的*H(w)*计算如下：

*H=-*0.7log(0.7)-0.3log(0.3)=0.2652,

*H=-*0.5******log(0.5)***-***0.5log(0.5)=0.3010.

表示本次数据中所获得直接信任值的不确定性低于相关信任值，根据式（９）计算得出相应的****，显然Σ****＝１.

C:\Users\aliali\Documents\Tencent Files\3028286534\Image\C2C\2B1`4L~EMN(]Q}O`]C8V239.png

由于传递信任的计算是指数性下降，因此可能出现*H(w) *1的情况，此时说明该数据的不确定性极大，数据值所包括的信息量极低，对于这类数据直接判定对应的****为０，降低不确定性对信任计算的影响．

通过上述的交互矩阵的连通恢复，节点交互向量填充及确定信任值权重参数，我们使用算法２计算节点信任矩阵．

**算法2．**信任矩阵构建算法CTMA(Constructing Trust Matrix Algorithm).

输入：交互邻接矩阵***M***

输出：信任矩阵***T***

1. 根据***M***及式（２）得到直接信任矩阵，由得到的图记为（*V,E*）;

２．根据算法１对进行连通性恢复以及相应的传递信任计算，得到相关信任矩阵***M***′；

３．对*i,j*，根据式（９）计算和对对应的ω，其中和分别是与***M***′对应位置的元素；

４．根据式（７）得到总信任值，所有的组成信任矩阵***T***；

５．END

３．２．３　算法时间复杂度分析

**整个算法的时间复杂度分为两个部分：（１）相关信任矩阵的计算，该部分中的连通性恢复需要对图进行遍历，复杂度为***O(|V|+|E|)***，其中|***V|***表示节点个数，***|E*|**表示边条数；对于计算传递信任值最坏的情况下是***O()***，即对任意的两个节点**

**都计算它们间的最短路径，但在实际上由于网络交互矩阵通常都是稀疏矩阵，因此一般情况下传递信任计算的时间复杂度为O((|***V*|+|*E***|)log|***V*|**)；对于最后相关信任矩阵的建立时间复杂度为***O***，因为对于任意两个节点都需要计算它们的相关信任度；（２）总信任矩阵的建立，该部分的时间复杂度为***O***，因为对于直接信任矩阵与相关信任矩阵中每个位置都要进行加权计算．因此整个算法的时间复杂度为***O***．**

**4　行为聚类算法**

**4. 1　相似度矩阵的构建**

**通过前文的信任建模，我们将原始的网络行为交互图转化为信任关系图，目的就是通过信任的方式对网络中的节点进行分类，在连续的时间序列中发掘出节点分类的演变趋势，从而判定网络中是否存在异常．**

**对于我们得到的总信任矩阵*T*，该矩阵是非对称矩阵，保持了信任有向性特征．但这种非对称性对分类造成了极大的困难，为了能够对网络节点进行分类，我们需要从*T*中提取出节点的相似性，构建对称的相似度矩阵，相似度矩阵的构建算法如下．**

**算法3．　相似度矩阵构建算法CSMA(Constructing Similarity Matrix Algorithm).**

**输入：交互邻接矩阵*M*，信任矩阵*T***

**输出：相似度矩阵*SIM***

**１．对于交互图中的任意两个节点,，根据信任矩阵*T*获得对应的信任向量,；**

**２．根据余弦定理，计算,间的夹角，得到节点**

**与的相似度；**

**３．对所有的节点,，根据构成相似度矩阵*SIM*；**

**４．END**

**整个算法的时间复杂度是***O***，即对任意两个节点都要计算它们间的相似度．通过以上算法可知*SIM*是对称阵，此时可使用聚类算法对网络节点进行分类，进而对网络的群体进行分析．**

**4. 2**　**基于信任的行为聚类算法**

聚类算法当前已经被广泛应用于对网络终端主机的分析，例如通过IP数据包头的协议特征、字节流大小、平均TTL(Time to Live)等等．但本文主要关注终端的网络行为，并以此建立信任矩阵，通过信任向量的相似性对网络节点进行分类．谱聚类算法作为数据聚类的一种重要方法，具有严密的数学理论，被广泛应用在图分割领域．原始的谱聚类需要给定犽值作为期望的簇个数，具有一定的局限性．文献［１７］对谱聚类算法进行改进，提出使用拉普拉斯矩阵***L***中的特征值前*k*个最大的特征值，其中特征值满足公式且，其中是***L***的特征值．由于该约束条件的形式太强，导致类的数目太小从而使得网络簇结构不明显，而且参数化公式不具备自调节能力．文献［２０］指出第*k*个和第*k*＋１个特征值之间的差越大，选取的*k*个特征向量所构成的子空间就越稳定．假设拉普拉斯矩阵***L***为*n*阶，特征值个数为，是***L***的特征值，且．我们定义，显然＞０．考虑到二阶导

数的几何意义为函数斜率变化的快慢，类似的，我们只需找出犵犻的变化趋势即可．因此，我们采用式（１０）对约束条件进行松弛，增强类数目的自动识别能力，其中*k*值是满足<0或中最小的*i*．

 (10)

我们通过对矩阵***L***的特征值进行约束，确定目标数据集的簇个数，进而通过算法４对相似度矩阵进行谱聚类．

**算法4．**　相似度矩阵自动谱聚类算法SMASCA(Similarity

Matrix Automated Spectral Clustering Algorithm).

**输入：相似度矩阵*SIM*，*SIM*维度为***n*

**输出：类集合**

**１．计算对角阵*D*，其中且=0；**

**２．计算拉普拉斯矩阵*L=D-SIM*；**

**３．计算*L*的特征值，并按降序排列生成特征值序列**

**；**

**４．计算，生成序列，，，；**

**５．根据式（１０）计算出簇数目***k***；**

**６．根据前***k***个最大的特征值获得对应的特征向量;**

**７．采用***k***均值算法对特征矩阵*V*进行聚类，得出聚类结果；**

**８．**END

**算法的前３步是典型的谱聚类算法计算拉普拉斯矩阵特征值方法，接着为了增加自动识别簇的能力，使用式（１０）作为特征值选取的约束条件，决定簇的数量．第７步是使用***k***均值算法对这***k***个最大特征值所对应的特征向量组成的实矩阵进行***k***分类，最后得到聚类结果．对于计算矩阵的特征值的时间复杂度一般为，***n***为矩阵维度；对于降序排列特征值时间为，其中***m***为特征值个数；而约束条件匹配的时间复杂度通常都较小，可视为常数；对于***k***均值算法事件复杂度一般为，其中***t***为迭代次数，***k* **为簇数目，***n***为节点个数，因此总的时间复杂度上界为.**

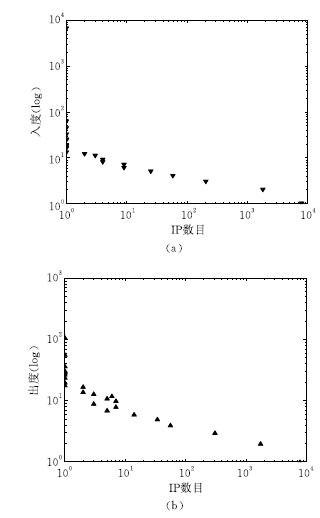
**5　实验分析**

**为了验证本文基于信任的行为分类算法的有效性，本节通过在OMNeT＋＋**C:\Users\aliali\Documents\Tencent Files\3028286534\Image\C2C\I1WC35N0AH)6_FA15V){OMO.jpg**平台中进行DDoS(Distributed Denial of Service)、蠕虫与僵尸网络3种典型互联网攻击模拟，分析攻击每个阶段的群体交互行为及分类结果，并与文献［17］提出的基于偶图的网络行为聚类方法进行对比．**

**5. 1　数据描述**

**本文的网络交互数据来自于北京国家计算机网络应急技术与处理协调中心某出口．整个数据集包括1392387个数据包，共计11852个**IP**，图５直观表示出整个数据集中节点的出度分布、入度分布，在双对数坐标下呈现出明显的幂律分布，与文献［20］互联网终端交互应用拓扑分析结果一致．为了减少**

**OMNeT＋＋的模拟负载，与文献［17］的数据处理类似，我们对数据集进行基于**IP**地址24位前缀的节点聚合，得到2527个**IP**地址，463912次交互．我们采用这2527个**IP**作为整个模拟实验的基础拓扑，同时将交互次数作为最原始的交互矩阵权值，并根据式（２）计算每个节点间的直接信任值，生成直接信任矩阵，记为*DTM*，为模拟提供基础输入．**



**图５　数据集节点交互入度与出度分布**

**5. 2**　模拟实验

OMNET＋＋是一个免费的、开源的多协议网络仿真软件，是学术界研究网络协议与交互的重要仿真工具之一．由于实际数据只是网络节点行为在某个时间段的快照，为了验证模型的

**\_\_\_\_\_\_\_\_\_\_\_\_**

1. OMNET++. [**http://www.omnetpp.org/**](http://www.omnetpp.org/)

有效性，我们将实际的网络交互拓扑数据作为模拟的启动参数，如表１所示．

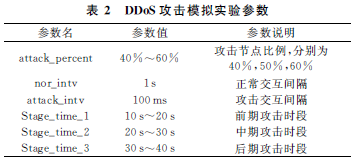


为使模拟实验更加真实，在模拟初始阶段根据直接信任矩阵***DTM***提取基本网络交互列表，并考虑到用户访问行为具有一定的倾向性，且不对称概率的设置使实验更加贴近实际，一般而言，我们令节点以３０％的概率与自身的中节点进行交互以维持原本具有交互行为节点在模拟过程中的持续性交互，同时以７０％概率与网络中其它节点交互．

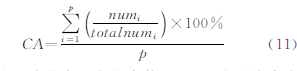
５．２．１　DDoS模拟实验

　　DDoS攻击是典型的网络恶意行为之一，通常的检测手段是通过链路流量的饱和度以及服务端连接数进行判断．为验证本文模型对整个DDoS攻击前期、中期和后期３个阶段的识别能力，我们设计以下实验进行分析．

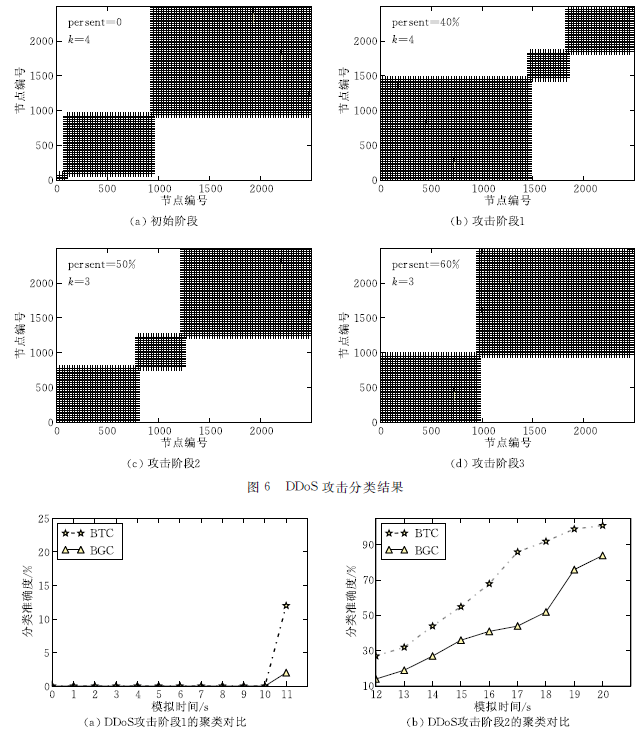
整个模拟过程的参数设置如表２所示，节点数及算法参数同表１．整个攻击模拟时长为４０s，被攻击节点序号为第２５２７号节点，模拟共分为４个阶段：（１）初始阶段（initialstage）．持续时间为０～１０s，该时期内所有节点根据前文提到的初始阶段进行网络交互；（２）攻击阶段１（attackstage１）．持续时间为１０s～２０s，４０％的节点（节点序号１５１０～２５１６）参与DDoS攻击；（３）攻击阶段２（attackstage２）．持续时间为２０s～３０s，５０％的节点（节点序号１２６０～２５１６）参与DDoS攻击；（３）攻击阶段３（attackstage３）．持续时间为３０s～４０s，６０％的节点（节点序号１００８～２５１６）参与DDoS攻击．每秒进行１次网络交互拓扑导出，共计产生４０个交互矩阵，则４个阶段的交互矩阵均为１０个，根据式（５）对每个阶段的总信任值进行迭代计算，图６显示了４个阶段的分类结果．图６（a）显示了初始阶段的分类结果，根据特征值选择约束条件得出类数目＝４，可看出大约有６０％的节点被划分到同一类中，根据图５的交互度分布可知大量的节点的出度小于２，虽然它们的目标节点大多不同，然而在行为上具有很高的相似度．图６（b）显示了攻击节点占总节点数量４０％的分类结果，两个偏小的阴影区域是攻击节点聚类集合，同时被攻击节点被单独分出，最大的阴影区域是正常节点集合．由于传递信任的影响导致攻击节点被分成２类，但所包括非攻击节点数极少，说明分类准确度较高．图６（c）显示攻击节点占５０％的情况，此时分类数目被自动识别为３，攻击节点与被攻击节点均被划分至同一类，说明随着攻击节点比例的增加，传递信任的影响逐渐增多．而图６（d）展示了攻击节点为６０％的分类结果，此时所有的攻击节点在同一类中，被攻击节点再次被单独分类，普通节点间的聚合度也进一步提升．我们可看出随着攻击节点数目的增加，整个网络的分类趋势逐渐变小，即值从４变至３，同时产生新的独立类别节点，即被攻击节点，同时还可看出当攻击节点比例较小时，分类的效果并不理想，没有将所有攻击节点划分至同一类别中，说明传递信任在网络行为聚集度不太高的情况下对分类结果产生一定影响，造成节点的相似度距离变大，当攻击节点比例上升至６０％，整个网络分类结果非常明显，攻击、被攻击及普通节点被分至不同的３类．整个过程同时也表明分类识别度随着攻击节点覆盖面的增加而增加．



为了对聚类算法中的特征值约束条件进行分析，我们进一步将分类算法与文献［１７］进行对比，首先定义分类准确度*CA*作为评价指标，由于我们更加关心攻击节点的分类准确性，因此只计算攻击节点居多的类别，公式如下：



其中*p*为攻击节点居多的类数目，表示该类中攻击节点的个数，表示该类中所有节点个数．为了更直观地分析分类结果，在模拟过程中每秒输出１次行为交互拓扑，在４０s过程中共计产生４０个交互行为图．图７显示了４个阶段的分类准确度*CA*，虚线是本文的聚类算法（BTC），很明显分类准确度高于文献［１７］中的基于偶图的分类模型（BGC）．从图７中可看出随着攻击节点数增加，两种分类模型的分类准确性迅速提升，但在攻击的初阶段，如图7(b)所示，当攻击持续７s时，BGC的分类准确性仅有３３％，而本文的分类准确度达到87%，这是由于BGC算法对特征值的



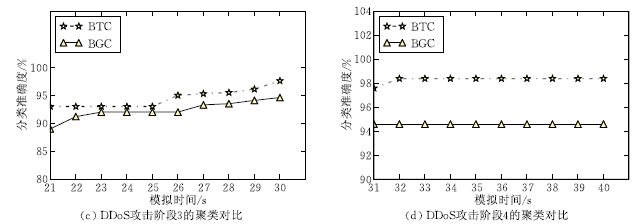
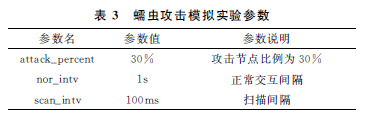


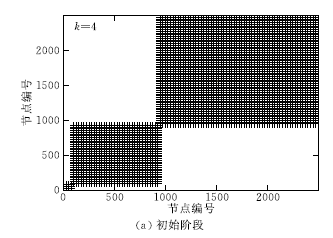
图7 DDoS攻击分类准确度对比

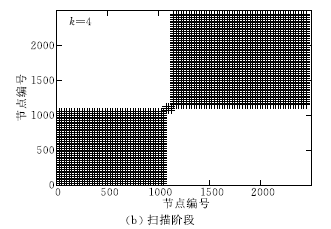
的约束条件过强，导致类数目较低，大量非攻击节点和攻击节点被划分到同一类中．随着参与攻击的节点不断增多，到第４阶段，两种算法的准确度都大于９０％，说明当攻击节点超过全网节点５０％时，偶图模型的分类方法也较为有效．

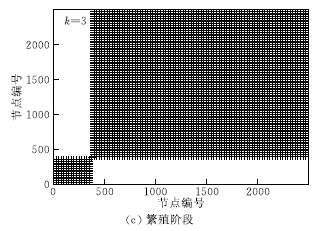
5. 2. 2　蠕虫模拟实验

蠕虫的网络行为与DDoS攻击截然不同，因为蠕虫扫描的目标集合并不完全一致，而DDoS攻击的目标几乎是唯一的．为了评估我们提出的信任模型分类方法在目标集合不同情况下的分类能力，我们在OMNET＋＋上进行蠕虫模拟实验．模拟时长及数据包交互参数与表２一致，其余参数如表３所示．整个模拟过程依然分为４个阶段：（１）正常交互过程，与DDoS模拟的阶段１一致；（２）随机扫描过程，扫描源为２个节点，目标为３０％正常节点；（３）繁殖过程，扫描源为之前３０％被扫描节点，目标为未感染的３０％节点；（４）爆发阶段，６０％被感染节点对全网进行随机扫描.









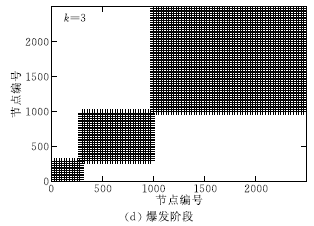
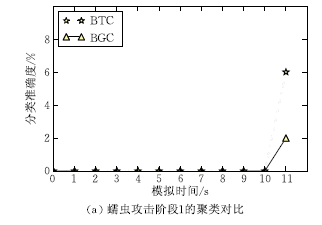
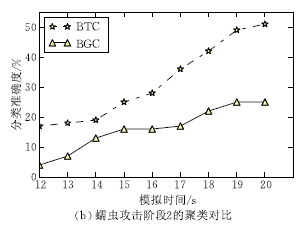


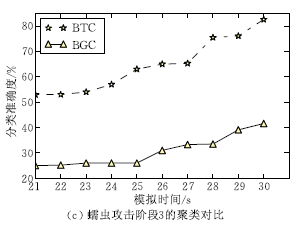
图８　蠕虫攻击分类结果

整个蠕虫传播过程的分类结果如图８所示，图８（b）是扫描阶段，可以看出类别数目*k*从４降到３，右上角阴影部分是被扫描节点所在的类，但该类中４４％的节点是正常节点,这是因为扫描行为的随机性以及传递信任导致一些普通节点与被扫描的相似度较高导致的．而图８（c）的分类效果较为明显，３０％的扫描节点与另外３０％的被扫描节点均被划分到同一类中，但也包括了将近２０％的正常节点，图８（d）是爆发阶段，比较明显的是６０％被感染的节点分到了同中．从图８的４个子图中可以看出蠕虫在初始扫描阶段的行为是较为隐蔽的，因此分类误差极大，然而随着感染节点增多，分类的准确度不断上升，并且从分类结果变化中可知在繁殖阶段，整

个网络分类的变化与其它３个阶段截然不同，存在某一类数目极多，这是由于计算相似度时，相关信任与直接信任的影响较为均衡，被感染的３０％节点与被扫描的３０％节点关联度很高，同时对于图８（d）而言，由于被感染节点比例在全网中已达到６０％，而且随着模拟时间的增加，这些扫描行为带来的影响已经完全超越了普通交互行为，从而提高了分类准确度．图９显示了４个阶段的分类准确度，当扫描节点不30%时，BTC的分类准确度最高达到50%，而BGC算法只达到20%，这是由于偶图模型依赖于目标节点集合之间的覆盖度作







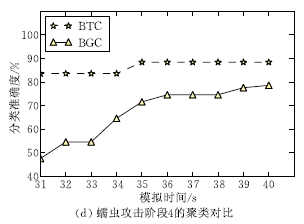


图９　蠕虫攻击分类准确度对比

为相似性判断的主要依据，而蠕虫扫描的目标节点集合覆盖度远远低于DDoS攻击．图９（c）和（d）中，本文提出的

BTC算法的分类准确度随着扫描节点不断增多大幅提升，当感染节点达到６０％时（即６０％节点均产生扫描行为），分类准确度达到８４％，同时由于扫描空间的有限性，偶图模型中的目标节点集合覆盖度也不断增加，因此BGC算法的分类准确度当感染节点数达到５０％时迅速提升至７２％．总体而

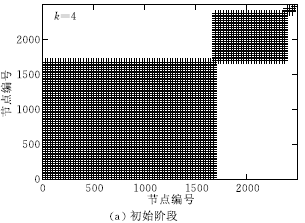
言，蠕虫行为的分类准确度低于DDoS，这是由于扫描行为分散性较强，但由于传递信任的影响增加了相关信任度数值，因此BTC方法明显优于BGC，尤其当感染节点数低于５０％时，即蠕虫传播初期与中期更明显．

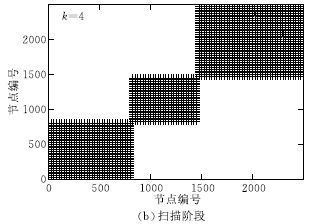
5. 2. 3　僵尸网络模拟实验

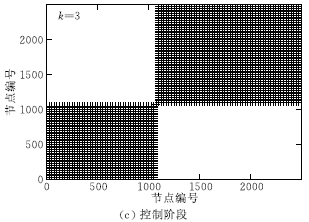
僵尸网络是当前互联网中威胁最大的恶意行为，同时由于僵尸网络的复杂度高，融合了蠕虫传播与DDoS攻击，且在控制阶段的隐蔽性强．为了进一步验证模型分类的有效性，我们在OMNET＋＋上对僵尸网络进行模拟，其中控制命令的交互与正常交互间隔一致，扫描参数与蠕虫模拟实验一致，而发动DDoS时的模拟参数与实验１一致．

整个模拟阶段依旧分为４部分：（１）正常阶段；（２）传播阶段（propagation）．该阶段为蠕虫扫描阶段，控制者向３０％的节点进行扫描，并且被感染的节点使用基于目标列表扫描方式，减少蠕虫扫描的影响范围，增强隐蔽性；（３）控制阶段（controlstage）．感染节点数占３０％，并不继续扫描，而是与控制者进行周期交互维持被控状态；（４）攻击阶段（attack stage）．控制者向被控端发送命令，进行DDoS攻击．模拟过程的结果分类如图１０所示．图１０（b）是扫描阶段，类似于图８（c），但由于僵尸网络中存在控制者的控制行为与其它行为的相似度较低，因此该控制者被单独分类，因此在僵尸网络模拟实验中传播阶段的*k*值为４．图１0（c）是控制阶段，由于该阶段并不存在扫描行为，而是单纯的控制行为，类似于僵尸网络模型中的潜伏期，此时的分类准

确度较低，很难区分控制行为与普通交互行为；图１０（d）是攻击阶段，与实验１类似，但由于僵尸网络的传播过程与控制阶段的累积影响，计算出*k*的值为６，可以看出参与攻击的３０％节点被分到同一类中．类似的，图１１显示了分类准确度的对比，在僵尸网络的传播阶段的分类准确度明显低于控制阶段，这是由于在传播初期，感染节点数目较少，整个僵尸网络处于不稳定状态．但在控制阶段中, 中，分类准确度总体呈现下降趋势，如图１１（c）所示，这表明在潜伏期的僵尸网络仅仅只存在控制行为，由于抽象度较高，单纯依赖交互行为







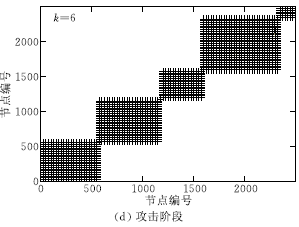
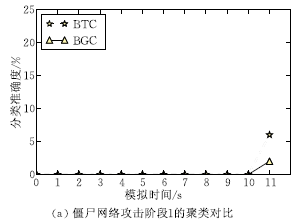
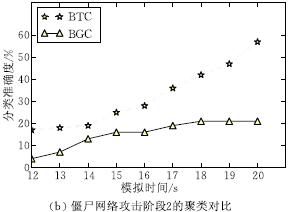
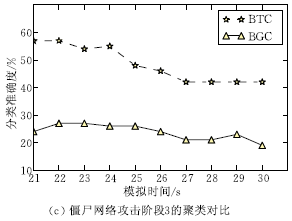


图１０　僵尸网络攻击分类结果







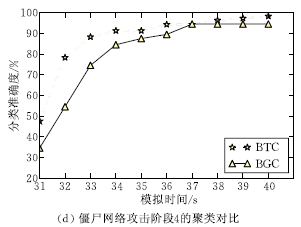


图１１　僵尸网络分类准确度对比

为是很难区分控制交互与普通交互，但BTC算法明显优于BGC，说明传递信任的优势强于偶图．图11(d)是发动DDoS攻击时的分类结果，与之前的DDoS攻击类似，分类准确度达到９０％以上

**5. 3**　小　结

总体而言，本文提出的基于信任的行为分类模型对于安全事件的初期阶段具有较好的分类准确度．在典型的３类网络攻击中，DDoS由于其集中性行为而容易被发现，而且整个网络分类的变化趋势较为显著，随着节点增加，网络内所划分的类数目*k*值呈现下降趋势，同时分类准确度也较高；蠕虫攻击则是分类准确度较低的一种网络行为，这是由于当感染节点比例较低时，传递信任的计算容易提高其它未感染节点的相关信任度值，从而导致分类的误差较高，整个蠕虫攻击实验中类的数目依然呈现下降趋势，而且随着蠕虫感染节点数的增加，整个分类结果中出现某一类中节点数急剧增加的情况，这种分类结果主要是由于我们的模型使用了相关信任度作为相似度计算的一个重要因素，而偶图模型在分类过程中过于依赖目标节点集合的重叠度，从而导致分类准确度较低．对于僵尸网络，由于其复杂的融合性与异构性，以及潜伏期的隐蔽性控制行为导致

在控制期的分类结果准确度较低，但在整个僵尸网络过程中可看出犽值变化不稳定，但仍可在其传播阶段发现典型的蠕虫扫描阶段行为及在DDoS攻击阶段发现节点聚集行为．

**6**　结论及未来工作

针对当前网络群体异常行为发现的模型采用偶图的局限性，以及这些模型并未考虑到社会化信任关系在互联网中的映射所表示出的信任关系因素，本文提出一种基于信任的群体行为模型，并在网络信任计算中增加相关信任这一新型概念，提高弱相关节点的相似度，同时在总信任度计算中采用信息熵概念降低信任计算风险，进而增加基于信任相似度聚类的准度．

下一阶段的工作重点是在加强互联网其它应用噪声环境下对信任度进行评估，同时深入分析网络信任认知过程以及不对称性在模型中的体现，进一步完善模型，提高性能，使之更适用于实际应用．

致　谢　感谢审稿老师给出的评语和修改意见！

参 考 文 献

[1] Sun Jian- Zhi, Zhang Ying-Xin,Chen Dan, Han Zhong-Ming .Effects on lifetime of opportunistic network based on

Epidemic routing under flooding attacks．Journal on Commu-

nications , 2012,33（9）：185-190（in Chinese）

（孙践知，张迎新，陈丹，韩忠明．泛洪攻击对Epidemic机制下机会网络生命期的影响．通信学报，2012,33（9）：185-190）

[2] Zang Tian-Ning，Yun Xiao-Chun，Zhang Yong-Zheng，Men Chao- Guang ，Sun Jian-Liang．A model of network device

Coordinative run．Chinese Journal of Computers,2011,34（2）：216-228（in Chinese）

（臧天宁，云晓春，张永铮，门朝光，孙建亮．网络设备协同

联动模型．计算机学报，2011,34（2）：216-228）

[3] Jiang Hong-Bo，Ge Zi-Hui，Jin Shu-Dong，Wang Jia．Network prefix-level traffic profiling：Characterizing，modeling，and evaluation．Computer Networks.2010,54(18):3327-3340

[4] Wang Z W，Xia Q，Lu K ,Zhang J．An approach on detecting attack based on causality in network behavior//Proceedings of the International Conference on Computer Science and Software Engineering．Wuhan，China，2008:947-950

[5] Li Xiao-Yong，Gui Xiao-Lin．Trust quantitative model with

Multiple decision factors in trusted network．Chinese Journal

Of Computers，2009,32(3):405-416（in Chinese）

（李小勇，桂小林．可信网络中基于多维决策属性的信任量

化模型．计算机学报，2009,32（3）：405-416）

[6] Li Xiao-Yong．Gui Xiao-Lin．Research on dynamic trust

model in large-scaled is tribute environment．Journal of

Software，2007,18（6）：1510-1521（in Chinese）

（李小勇，桂小林．大规模分布式环境下动态信任模型研究．

软件学报，2007,18（6）：1510-1521）

[7] Huang J，David M N．A formal-semantics-based calculus of

trust．IEEE Internet Computing，2010,14(5):38-46

[8] Zhou R，Hwang K，Cai M．Gossip Trust for fast reputation

aggregation in peer-to-peer networks．IEEE Transactionson

Knowledge and Data Engineering，2008,20(9):1282-1295

[9] Li X，Zhou F，Yang X．Scalable feedback aggregating

（SFA）overlay for large-scale P2P trust management．IEEE

Transactions on Parallel and Distributed Systems, 2014, 23(10): 1944-1957

[10] Erman A，Faramarz F．Iterative trust and reputation man-

agement using belief propagation．IEEE Transactionson

Depend able and Secure Computing，2012,9(3):375-386

[11] Bao F，Chen I R，Chang M J，et al．Hierarchical trust man-agement for wireless sensor networks and its applications to

Trust-based routing and intrusion detection．IEEE Transtctions on Network and Service Management,2012,9(2):169-183

[12] Cai L，Roberto R C．Three-dimensional based trust management scheme for virus control in P2P networks//Proceedings of the IEEE International Conference on Communications

（ICC）．Cape Town，South Africt，2010:1-5

[13] Carol J F，Zhang J，Issam A，Raouf B．Dirichlet-based trust management for effective collaborative intrusion detection

networks．IEEE Transactions on Network and Serbice

Management，2011,8(2):79-91

[14] Abedelaziz M，Nicholas H，Yongdae K．You’re your friends close：Incorporating trust into social network-based Sybil

Defenses//Proceedings of the IEEE International Conference

On Computer Communications（INFOCOM）．Shanghai，

China，2011:1943-1951

[15] Jin Y，Sharafuddin E，Zhang Z L．Unveiling core network wide communication patterns through application traffic

activity graph decomposition//Proceedings of the ACM Special

Interest Group on Performance Evaluation（SIGMETRICS）．

Seattle，USA，2009:49-60

[16] Iliofocu M，Pappu P，Faloutsos M，et al．Network monitoring using traffic dispersion graphs//Proceedings of the 7th

ACM Special Interest Group on Data Communication

（SIGCOMM）Conference on Internet Measurement．San

Diego，CA，USA，2007:315-320

[17] Xu K，Wang F，Gu L．Network-aware behavior clustering of Internet end hosts Proceedings of the IEEE International Conference on Computer Communications（INFOCOM）．Shanghai，China,2011:2078-2086

[18] Guillaume J L，Latapy M．Bipartite graphs as models of

Complex networks．Physica A：Statistical and Theoretical

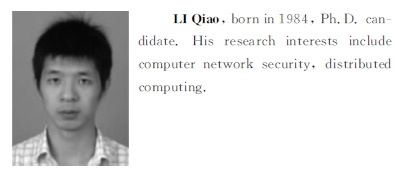
Physics，2006,371(2):795-813

[19] Sabrina D C，Sara F，Sushil J，et al．Integrating trust management and access control in data-intensive Web applications．ACM Transactions on the Web，2012,6(2):1-43

[20] Xu K，Wang F．Behavioral graph analysis of Internet applications//Proceedings of the IEEE International Conference

On Global Telecommunications（GLOBECOM）．Houston，

USA，2011:1-5



**HE Hui**，born in1974，Ph．D．，associate professor．

Her research interests include network computing，network

security．

**FANG Bin-Xing**，born in1960，Ph．D．，professor，Ph．D．supervisor，member of Chinese Academy of Engineering．His current research interests include the theory of information security，parallel computing．

**ZHANG Hong-Li**，born in1973，Ph．D．，professor，

Ph．D．supervisor．Her research interests are focused on

Internet measurement and network computing

**WANG Ya-Shan**，born in1978，M．S．candidate．His research interests include computer network security，distributed computing．

**Background**

It is important to understand network behavior for efficient network management and security monitoring，while the Internet is plagued with a wide range of security threats such as fast worm spreads and distributed denial-of-service attacks．In an open distributed environment，trust management had been proposed to decrease the network risk in the past decade．However ,researchers combine these trust systems with the decentralized reputation and recommendation systems used in P2P and social networks，they do not utilize the trust for detecting the network group behavior . Moreover，the current network security monitoring and detecting mechanisms do not consider the inherent social relationship under the network behavior．Most of the network abnormity monitoring systems have some limitations:(1) the depth packet detection technique consume so large time and space cost that it could not apply to analyze the large scale network data;(2) the current macro monitoring systems ignore the social relations behind the network communications．

The inherent essence of network behavior is the external representation of the social relationship．This paper constructs a network group behavior model based on the network trust．The mainly decision factors ,which are used to quantify the trust value in this paper include the direct interaction trust and the correlative trust which implies the potential trust between two indirect connection nodes．

To analyze the evolution of network group behavior and detect the network abnormality in a macro perspective, this paper focuses on the trust relationship among the network．The intuition to solve group abnormality is through trust evaluation , node similarity compute , and node clustering．However there are four mainly problems：(1)the network interaction topology is an unconnected graph in most cases；(2) how to compute trust value between two undirected nodes；(3) how to compute the weight of trust decision factors; （4） how to convert the trust value to the node similarity．

Firstly ，a connectedness recovering strategy is proposed in order to establish the trust between each two nodes using the node activeness as well as the social network．Secondly the definition of relevant trust is proposed to resolve the trust between the weak correlation nodes．Thirdly，the theory of Information entropy is used to quantify the weight of decision factors．At last this paper employs the vector angle method to evaluate the node similarity and then adjust the spectral clustering algorithm to classify the network nodes．The simulation results show that our model could detect anomalous behaviors such as denial of service attacks，worm propagations and botnet．

This work is partially supported by the National High Technology Research and Development Program（863 Program）of China（Grant No．2011AA010705），National Basic Research Program of China（ 973 Program）of China（Grant No．2011CB302605）；the National Natural Science Foundation of China（Grant Nos．61173145,60203021）．

**备注：在pdf转word中，各种字体、大小的切换，数学公式、图片的插入，整篇论文的排版，都依赖于word强大的功能。**