

节点相似度感知的社会化间断连接无线网络结构检测机制

吴大鹏* 靳继伟 吕翊 王汝言
(重庆邮电大学宽带泛在接入技术研究所 重庆 400065)

摘要: 间断连接无线网络中的节点采用“存储、携带、转发”的方式传输消息,网络中的节点具有较强的社会属性。该文提出一种基于相似度和标签传播的网络结构检测机制,节点以运动过程所经历的社区状态为依据,描述自身的中心度,进而,采用带有约束条件的标签传播方法,更新网络结构检测状态,最终完成社区结构划分。仿真结果表明,所提出的检测算法具有较高的准确性,适用于多种复杂网络结构。

关键词: 间断连接无线网络; 社区检测; 中心节点; 邻居节点

中图分类号: TP393.04

文献标识码: A

文章编号: 1009-5896(2013)01-0141-06

DOI: 10.3724/SP.J.1146.2011.01178

Similarity Aware Community Detecting Method for Social Intermittently Connected Mobile Network

Wu Da-peng Jin Ji-wei Lü Yi Wang Ru-yan

(Broadband Ubiquitous Network Research Laboratory,

Chongqing University of Posts and Telecommunications, Chongqing 400065, China)

Abstract: With the “store-carry-forward” transmitting manner in social intermittently connected mobile network, the nodes always behave in community movement model. A community structure detecting method is proposed based on similarity and community label propagation. According to the statistics of roaming epochs in community movement, the centrality of nodes can be evaluated accurately; moreover, the network update correlative situation after controlling the label propagation with certain constraint condition. Finally the whole network is divided into several communities logically, and centrality nodes are also probed. Simulation results show that the method is effective.

Key words: Intermittently connected mobile network; Community detecting; Central node; Neighbor node

1 引言

区别于传统的移动自组织网络(Mobile Ad hoc NETworks, MANET),社会化间断连接无线网络^[1]采用“存储-携带-转发”的方式,充分利用节点间的相遇机会中继数据,直到数据送达目的节点。现实生活中,移动通信设备的持有者的运动过程具有一定的规律性,多个终端组成的网络呈现出社会网络学中的“小世界,大世界”现象^[2,3]。此种模式下,节点以自组织的方式形成社区,网络从逻辑上表现为由多个社区所构成。

传统的 MANET 中,节点大多采用基于 IEEE802.11 标准的模块作为射频传输单元,其通信范围较大,节点覆盖范围内的邻居数量相对较多,网络结构检测过程中的相关控制信息可以被多个节

点接收,检测过程相对简单。该领域内具有代表性的社区检测方法包括 K-CLIQUE^[4], Weighted Network Analysis (WNA)^[5], 受限最大跳数方法^[6], 社会活跃度检测方法^[7]。K-CLIQUE 方法需要预先设定社会图中边的权重,与其类似, WNA 算法需要获取全部节点的关系向量,两者具有一定的局限性。限制最大跳数的方法主要面向风琴网络模型^[8],并不适用于社会化间断连接无线网络。

区别于 MANET,在社会化间断连接无线网络中,节点大多采用基于蓝牙标准的模块作为射频传输单元,发射功率较小,通信范围内的邻居节点相对较少,在节点比较稀疏的情况下,存在连接的节点数量较少,大多数情况下是依靠节点相遇的机会来转发数据。社会化间断连接无线网络具有较强的特殊性,相关网络结构检测的研究成果极为有限。针对社会网络,文献[9]利用相遇矩阵检测相遇概率较大的源节点和目的节点。文献[10]提出了标签泛洪算法导致边缘节点随机性地划分到各个社区。文献

2011-11-14 收到, 2012-10-29 改回

国家自然科学基金(60772109, 61001105)和重庆市教委科学技术项目(KJ100521)资助课题

*通信作者: 吴大鹏 wudapengphd@gmail.com

[11]统计节点为其他节点在最短路径上中继消息的次数,并以此参数来衡量节点的中心程度。此外,文献[12]采用集合的相关性来衡量重要性。文献[13]选择此类节点作为社区中心节点。较为常用的最大连接度算法(Highest Connectivity Degree cluster Algorithm, HCDA)操作相对简单,但准确程度有限。以上文献中虽能够以一定的成功率完成社区中心节点的检测,但是总体来说精度较低,动态性较差,对于“存储-携带-转发”的社会化间断连接无线网络中,网络结构检测方法并不完善。

为了提高社会化间断连接无线网络中的资源利用率,改善网络性能,需要节点在未知网络拓扑信息的情况下,在移动过程中充分感知网络逻辑结构,包括社区中心节点和邻居节点。根据网络中节点的社会属性,以低开销为目标,本文提出一种适用于社会化间断连接无线网络的快速网络结构检测策略,克服了社区邻居节点检测的难点。根据运动过程中节点状态变换情况,节点对自身的中心度进行估计,同时,利用节点运动过程的历史信息,确定归属社区的邻居节点,进而,在未知社区规模和个数的情况下采用标签传播方法检测网络结构。

2 网络结构检测方法

根据社会化间断连接无线网络中的节点运动特性,本文提出的网络结构检测方法主要包含社区中心节点检测方法、邻居节点检测方法以及标签扩散过程3个部分。

2.1 社区节点中心度

根据社会网络中的“小世界,大世界”理论,社会化间断连接无线网络的拓扑可以从逻辑上划分为若干个区域,每个区域称为社区,社区内部节点采用短距离无线通信协议进行数据传输,且网络中的节点具有唯一的归属社区,因此,根据节点运动过程中所处的社区情况,可以将节点划分为两种状态,分别为本地状态(L)和漫游状态(R)如图1所示。本地状态对应节点的归属社区,当节点处于本地社区的时候,通常在下一个运动周期内以较大的概率处于本地状态,较小的概率漫游到其他社区;若当前节点处于漫游状态,则以较小的概率继续漫游在其他社区,而以较大的概率回到本地社区。

根据社区模型原理^[4]可知,节点运动过程包括两个状态,即本地状态和漫游状态。各个状态之间的转换情况如图1所示。其中 P_l 表示当前节点状态为本地,下一个运动周期内节点状态依然为本地的概率; P_r 表示当前为漫游状态,下一个运动周期内节点状态依然为漫游状态的概率。假设给定的某个

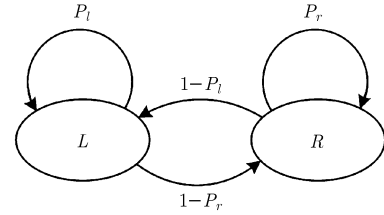


图1 节点运动状态转换原理

运动周期内节点在本地或者漫游状态的概率分别是 $\pi_l^{(i)}$ 和 $\pi_r^{(i)}$,由Markov定理可以得到稳态下运动周期内节点处于本地和漫游状态的概率分别如式(1)和式(2)所示:

$$\pi_l^{(i)} = \frac{1 - p_r^{(i)}}{2 - p_l^{(i)} - p_r^{(i)}} \quad (1)$$

$$\pi_r^{(i)} = \frac{1 - p_l^{(i)}}{2 - p_l^{(i)} - p_r^{(i)}} \quad (2)$$

在给定时间段内,节点漫游在归属社区之外的情况可以用其在任一运动周期内处于漫游状态的概率来衡量。假设存在节点 i 和 j ,对应的处于漫游状态的运动周期数量平均值为 \hat{e}_i , \hat{e}_j ,满足式(3)的关系。

$$\frac{\hat{e}_i}{\hat{e}_j} = \frac{\pi_r^{(i)}}{\pi_r^{(j)}} \quad (3)$$

根据节点运动模型的特性,给定时间段内,社区切换较为频繁的节点,处于漫游状态的运动周期平均数 \hat{e} 较多,该类节点更适合协助其他节点转发数据包,通常被其他节点选为备用辅助节点。因此节点处于漫游状态的运动周期平均数 \hat{e} 可用于衡量其中心度,处于漫游状态的运动周期数量越多,节点也就越活跃,更适合作为社区备用辅助节点。由于节点的每个运动周期只有本地和漫游两种状态,通过统计这两种状态各自的运动周期数量,根据式(4)可以得到节点的中心度:

$$C(i) = \frac{\hat{e}(i)_r}{\hat{e}(i)_r + \hat{e}(i)_l} \quad (4)$$

其中 $C(i)$ 是用来衡量节点中心度的参数; $\hat{e}(i)_l$ 是节点 i 在所有的运动周期中处于本地状态的平均个数; $\hat{e}(i)_r$ 是节点 i 在所有的运动周期中处于漫游状态的平均个数。社会化间断连接无线网络的无向图 $G_s=(V_s, E_s)$ 中, V_s 为节点集合, E_s 为链路集合。对于任意节点 i ,存在

$$C_i \leq C_j, \quad i, j \in V_s \quad (5)$$

则节点 j 定义为社区中心节点,节点 j 的中心度为 C_j ;节点 i 定义为普通节点,中心度为 C_i 。

2.2 邻居节点发现过程

定义 社会化间断连接无线网络可用社会无向

图 $G_s=(V_s,E_s)$ 表示, 其中 V_s 为节点集合, E_s 为链路集合。若 $\text{Edge}(u,v)$ 属于 E_s , 则认为节点 u 和 v 间存在社会关系。

与 MANET 中节点较为密集的情况有所不同, 在社区网络模型中节点比较稀疏的情况下, 节点通信范围内的邻居数量较少, 同时, 与本地社区的节点相遇机会较多, 与其他社区的节点接触机会较少, 且节点间相遇接触的时间相当短暂。基于社会化间断连接无线网络的此种特点, 本文使用节点间的接触频繁程度来确定社会图中边的权重, 进而, 根据节点的历史信息中可以确定邻居节点。历史信息保存结构如图 2 所示。

节点 u_1	节点 u_2	...	节点 u_n	平均值
(u_1, c_{u_1})	(u_2, c_{u_2})	...	(u_n, c_{u_n})	C_{ave}

图 2 节点相遇次数映射集合

假设网络中存在节点 u , 与其相遇的节点分别是 $u_1, u_2, \dots, u_i, v_1, v_2, \dots, v_j$ 。其中 u_1, u_2, \dots, u_i 代表与 u 处于同一社区的节点; v_1, v_2, \dots, v_j 代表与 u 处于不同社区的节点; u 与各个节点相应的相遇次数为 $c_{u_1}, c_{u_2}, \dots, c_{u_i}, c_{v_1}, c_{v_2}, \dots, c_{v_j}$, 根据式(6), 式(7)可以得到节点 u 与其他所有相遇节点的平均相遇次数 C_{ave} 。

$$C_{sum} = \sum_{n=1}^i c_{u_n} + \sum_{m=1}^j c_{v_m} \quad (6)$$

$$C_{ave} = C_{sum} / (m + n) \quad (7)$$

对于 u 的相遇节点 u_i 来说, 若 u 与 u_i 相遇的次数 c_{u_i} 满足式(7)条件

$$c_{u_i} \geq C_{ave} \quad (8)$$

则 u 认为 u_i 是邻居节点, 否则视为不同社区的节点。

2.3 标签传播和社区的形

基于标签传播的思想^[15], 本文提出了一种低复杂度的网络结构检测机制, 相似度感知的检测机制 (Similarity Aware Community Detecting, SACD)。

初始化阶段, 每个节点都赋予唯一的 ID 以相互区别, 而且标签初始化为对应的节点 ID, 即每个节点假设自身为社区中心节点。随着在运动过程中与其他节点相遇, 通过比较各自的中心度数值确定社区结构以及中心节点。若与之相遇的节点中心度数值 $C(j)$ 至少为本节点 $C(i)$ 值的 N_{ji} 倍 ($N_{ji} = \pi_r^{(j)} / \pi_r^{(i)}$), 并且检测到对方是邻居节点的情况下, 则把对方的社区标签设置为自己的标签。随着标签在网络中的逐渐传播, 所有节点将以自组织的方式形成相应的社区。

$$\begin{aligned} \frac{C(j)}{C(i)} &= \frac{\hat{e}(j)_r / [\hat{e}(j)_r + \hat{e}(j)_l]}{\hat{e}(i)_r / [\hat{e}(i)_r + \hat{e}(i)_l]} \\ &= \frac{\hat{e}(j)_r}{\hat{e}(i)_r} \times \frac{\hat{e}(i)_r + \hat{e}(i)_l}{\hat{e}(j)_r + \hat{e}(j)_l} \geq \frac{\hat{e}(j)_r}{\hat{e}(i)_r} = \frac{\pi_r^{(j)}}{\pi_r^{(i)}} \quad (9) \end{aligned}$$

本文所提出的标签传播和社区中心节点选举的步骤概括如下:

步骤 1 节点初始状态下将社区标签设置为自己的 ID; 同时将中心度数值 $C(i)$ 设置为 0;

步骤 2 每个运动周期开始的时候, 更新 $C(i)$ 数值, 节点相遇之后, 彼此交换自身的 $C(i)$ 数值和社区标签。

步骤 3 节点收到其他节点相关信息之后, 更新本地邻居列表中的相遇次数表项, 若节点之间为首次相遇, 则将其添加至邻居列表。

步骤 4 根据式(6)更新平均相遇次数 C_{ave} 。

步骤 5 若相遇节点满足式(7)的条件, 则选用 $C(i)$ 最大的节点作为社区标签, 更新当前的社区标签。

该算法具体过程如下图 3 所示。

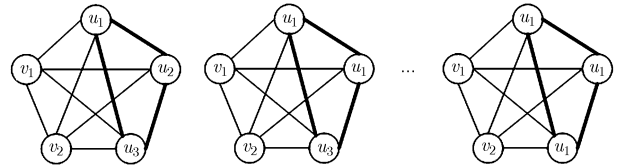


图 3 标签传播原理

网络初始化的时候每个节点都配备了独有的社区标签, 这些标签随着节点运动在网络内部进行传播。社会关系较为紧密的节点能够以较快的速度聚集为社区, 如图 3 所示。文中的连线表示节点间有相遇的历史记录, 如节点 u_1 和 u_2 之间的连线表示 u_1 的邻居列表中有节点 u_2 的相遇记录以及对应的相遇次数, 而且图中较粗的连接线表示节点间的联系较为频繁。当 u_2 与 u_1 相遇后, u_2 更新邻居列表, 若列表中 u_1 节点及其对应的相遇次数 $c_{u_2-u_1} \geq u_{2-ave}$, 则 u_2 节点的社区标签更新为 u_1 。以此类推, 最终节点密集组在标签上达成一致后, 所有相同标签的节点组成一个社区。

该方法的伪代码如表 1 所示。

3 仿真数值分析

本仿真采用 (Opportunistic Network Environment, ONE)^[16]仿真软件, 并在此基础上搭建社区运动模型, 具体仿真参数如表 2 所示。

图 4 描述了中心度差异对网络结构检测机制的

表 1 标签传播和社区形成方法的伪代码

```

On POWER-ON for node  $u$ 
{
    Set Comm_Label = uniqueID;
    Set Centrality( $u$ ) = 0;
    Set neighborList $_u$  = null;
}
While(selectDestination)//每个运动周期开始时选择目的地
    If Node goes out of Local Community//如果节点选择
    出社区
        Centrality( $u$ )++;//出入社区次数增加 1
    End if
End while
Upon  $T_c$  timer expiration:
    Broadcast its packet containing its unique ID,
    Comm_label and Centrality( $u$ );
    Receive packets from meeting_nodes and update
    neighborList $_u$ .
    If fist time encounter these node
        Add uniqueID of those node in neighborList $_u$ ;
    End if
    Calculate  $C_{ave}$  according to equation(2)
    For all node  $i$  in meeting_nodes do
        If( $c_{ui} \geq C_{ave}$ )
            Set Set Comm_Label = Comm_Label $_{maxEpochs}$ ;
        End if
    End for
End for

```

表 2 仿真参数设置

参数设定	参数数值
网络区域(m ²)	4500×3400
网络仿真时间(h)	24
社区数量	4(默认)
社区内节点数量	30
无线传输范围(m)	10
传输速度(kbps)	250
数据包间隔时间(s)	[25,35], 服从指数分布
节点的缓存(M)	5 M
单个消息大小(kb)	[80,150], 服从指数分布
运动速度(m/s)	1-6
停留时间(s)	0-10

影响情况。随着中心节点与普通节点访问外部社区的概率比值越大, 本文所提出的 SACD 机制性能逐渐上升。当比值相近的时候, 普通节点表现出和中心节点相似的社会属性, 节点状态趋近随机运动模型的特性。在节点相遇历史信息中, 本地节点和邻

居社区节点在相遇次数上趋于一致, 用式(7)和式(8)无法判断所遇到的节点是否属于同一个社区, 本地节点极易被划分到外部社区。当节点访问外部社区的概率比值达到 10 以上, 则社区结构检测结果较为准确; 当概率比值为 1 的时候, 即中心节点和普通节点漫游状态的概率相同, 则该算法只能达到 24.21% 的准确率。HCDA 机制以节点连接度来确定网络中心节点, 当节点访问外部社区的概率比值较小时, 连接度较高的节点通信范围内所含邻居社区节点较多, 难以适应动态性较强的社会化间断连接无线网络, 检测正确率较低; HCDA 最大的缺陷在于, 其限制了连接度较大节点的社区切换频率, 只有在中心节点切换频率极低的情况下才能保证邻居节点成员的稳定。

图 5 描述了两种机制随网络运行时间的变化情况, 当运行时间为 2 h 的时候, 社区检测的准确率为 36.53%。在网络运行时间较短的情况下, 节点无法充分获得邻居节点的信息, 对式(8)来说, 节点相遇次数映射集合中 C_{ave} 和式(6)中的各个分量过于接近, 进而对邻居节点产生误判; 同时节点以某种概率进行社区切换, 在未达到稳态之前节点的中心度无法充分体现, 普通节点和中心节点访问外部社区的次数具有随机性, 从而造成对中心节点的误选。而当网络时间是 12 h 的时候, SACD 机制的准确率达到 99.08%。随着运行时间的增加, 节点获得更充分的信息来判断邻居节点, 该算法所表现的性能越来越好。HCDA 机制随着运行时间的增加, 连接度最大的节点通常选取社区中较少进行社区切换的普通节点。但是, 对于间断连接无线网络来说, 节点动态性较强, 运动轨迹难以预测, 连接度最大的节点获得的邻居节点成员具有不稳定性, 随网络运行时间的变化而变化。针对社区切换较为频繁的节点难以确定其归属社区, 仿真数据显示所检测的准确度不超过 52%。

图 6 描述了社区节点个数与检测准确率之间的关系, 结果表明两种算法性能均呈现出一定的上升趋势。节点密集程度随社区节点数目逐渐增加, 最大连接度算法 HCDA 中, 由于连接度最大的节点通信范围内节点数量增加, 其性能不断提升, 其中普通节点的邻居成员稳定性有所增加, 但是由于所选取的连接度最大的节点仍属于普通节点, 无法检测到中心节点。而本文所提出的 SACD 机制随着节点数目增加, 即导致节点相遇次数映射集合中所含元素增加, 同社区节点的通信次数明显增加, 有利于式(7)和式(8)确定邻居节点, 进而确保标签准确传播。仿真数据表明本文所提出的 SACD 算法均具有相对较高的准确度, 最高可达 99.89%。随着社区内

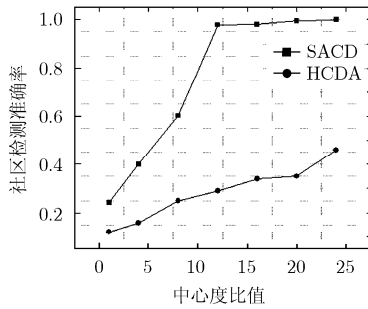


图 4 检测准确度与中心度差异关系

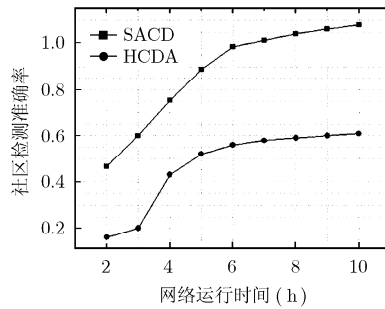


图 5 检测准确度与运行时间的关系

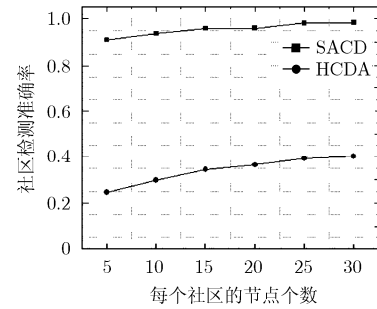


图 6 检测准确度与社区节点数量关系

节点数目的增加，SACD 机制的准确度随之增加，可见，所提出的机制具有较好的扩展性。HCDA 算法的检测准确度虽有所上升，但是效果较差，难以满足实际应用需求。

从图 7 可以看出，随着社区数量的增加，本文提出的 SACD 机制更具优势。SACD 算法扩展性较强的原因主要是社区数量的增加和中心节点数目的增加成正比例关系，对每个社区来说，该算法是以分布式方法检测中心节点和邻居节点成员；同时随着社区数目的增加，中心节点切换到每个社区的平均次数下降，从而中心节点标签传播到外部社区的机率随之降低，因此准确度随社区数目增加而增加。由仿真数据显示，当社区个数从 2 个增加到 5 个的时候，网络结构检测的难度也随之上升，SACD 机制的准确度仍然可以达到 90% 以上。对于 HCDA 来说，社区个数增加的同时，单个社区节点数目没有变化，因此，网络节点的稀疏状态比较稳定，检测出的准确度变化幅度较小。

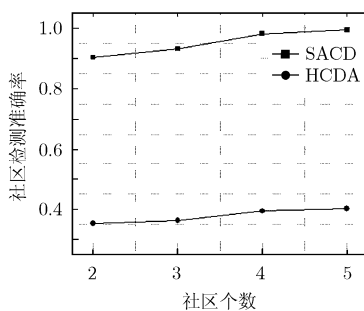


图 7 检测准确度与社区数量关系

4 结束语

根据节点运动状态改变情况，本文提出的一种适用于社会化间断连接无线网络的结构检测方法。所提出的网络结构检测机制不仅能够检测到社区的成员，还能根据网络中节点活跃程度的差异性检测出社区的中心节点，进而，有利于设计相应的路由方法以及缓存管理方法，更适用于个人泛在化网络。

未来的主要工作是在社区划分和中心节点检测的基础上，根据网络状态确定消息副本数量，从而优化多副本路由协议在实际应用中的性能。

参考文献

- [1] Sabrina G, Elena P, and Gian P R. Strangers help friends to communicate in opportunistic networks [J]. *Computer Networks*, 2011, 55(18): 374-385.
- [2] Spyropoulos T, Psounis K, and Raghavendra C S. Performance analysis of mobility-assisted routing [C]. *Proceedings of ACM MOBIHOC*, Florence, Italy, 2006: 49-60.
- [3] 牛建伟, 周兴, 刘燕, 等. 一种基于社区机会网络的消息传输算法[J]. *计算机研究与发展*, 2009, 46(12): 2068-2075. Niu Jian-wei, Zhou Xing, Liu Yan, et al. A message transmission scheme for community-based opportunistic network [J]. *Journal of Computer Research and Development*, 2009, 46(12): 2068-2075.
- [4] Palla G, Farkas I, and Vicsek T. Uncovering the overlapping community structure of complex networks in nature and society [J]. *Nature*, 2005, 435(7043): 814-818.
- [5] Newman M E J. Analysis of weighted networks [J]. *Physical Review E*, 2004, 70(5): 056131.
- [6] Whitebeck J and Conan V. HYMAD: hybrid DTN-MANET routing for dense and highly dynamic wireless networks [J]. *Computer Communications*, 2010, 33(13): 1483-1492.
- [7] 王贵竹, 张家勇, 王炳庭. SA-DTN: 基于节点社会活跃度的 DTN 路由研究[J]. *计算机应用研究*, 2011, 28(4): 1524-1526. Wang Gui-zhu, Zhang Jia-yong, and Wang Bing-ting. SA-DTN: research of DTN routing based on node social activity [J]. *Application Research of Computers*, 2011, 28(4): 1524-1526.
- [8] Pierre-Ugo Tournoux, Jeremie Leguay, Farid Benbadis, et al. The Accordion Phenomenon: Analysis, Characterization, and Impact on DTN Routing [C]. *INFOCOM 2009*, IEEE, RIO de Janeiro, April 2009: 1116-1124.
- [9] Daly E M and Haahr M. Social network analysis for routing in disconnected delay-tolerant MANETS [C]. *Proceedings in*

- ACM MobiHoc, New York, USA, 2007: 911-922.
- [10] Raghavan U N, Albert R, and Kumara S. Near linear time algorithm to detect community structures in large-scale networks [J]. *Physical Review E*, 2007, 76(3): 036106.
- [11] Hui P, Crowcroft J, and Yoneki E. BUBBLE rap: social-based forwarding in delay tolerant networks [J]. *IEEE Transactions on Mobile Computing*, 2010, 10(11): 1536-1550.
- [12] Brin S and Page L. The anatomy of a large-scale hypertextual web search engine [J]. *Computer Networks and ISDN Systems*, 1998, 30(1-7): 107-117.
- [13] Mtibaa A, May M, Diot C, *et al.* PeopleRank: social opportunistic forwarding [C]. Proceedings of INFOCOM, San diego, CA, 2010: 1-5.
- [14] Hui P, Chaintreau A, Scott J, *et al.* Pocket switched networks and human mobility in conference environments [C]. In Proceedings of the 2005 ACM SIGCOMM Workshop on Delay-tolerant Networking, New York, USA, 2005: 244-251.
- [15] Hui P and Crowcroft J. How small labels create big improvements [C]. Pervasive Computing and Communications Workshops, 2007. PerCom WorkShops'07, Fifth Annual IEEE International Conference on, New York, USA, 2007: 65-70.
- [16] keranen A, Ott J, and Karkkainen T. The ONE simulator for DTN protocol evaluation [C]. Proceedings of the 2nd International Conference on Simulation Tools and Techniques, Rome, Italy, 2009: 1-10.
- 吴大鹏: 男, 1979 年生, 博士, 副教授, 研究方向为泛在网络。
靳继伟: 男, 1984 年生, 硕士生, 研究方向为无线机会网络。
吕 翊: 男, 1963 年生, 教授, 研究方向为互联网服务质量控制。