

基于 ICN 网络架构的社区感知型 MSN 路由机制*

石峻岭¹, 王兴伟¹, 黄敏²

¹(东北大学 计算机科学与工程学院, 辽宁 沈阳 110169)

²(东北大学 信息科学与工程学院, 辽宁 沈阳 110819)

通讯作者: 王兴伟, E-mail: wangxw@mail.neu.edu.cn



摘要: 移动社交网络(mobile social network, 简称 MSN)利用移动用户之间的社交关系,通过节点间的协作式转发实现消息交付.然而,随着大数据时代的到来,MSN 需要满足移动用户日益增长的对内容(如视频)的需求.由于信息中心网络(information-centric networking, 简称 ICN)对移动性的支持,基于 ICN 架构,提出了一种 MSN 中基于社区划分的路由机制.在兴趣决策中,利用节点请求中的内容名字获取用户的兴趣偏好,进而计算用户间的兴趣差异度量;根据兴趣差异将节点划分为兴趣社区,依据这些兴趣社区进行兴趣包路由.在数据决策中,根据节点历史相遇信息计算用户间的相遇规律度量,根据相遇规律将节点划分为社交社区,依据这些社交社区进行数据包路由.同时,根据兴趣社区和社交社区信息优化节点的内容缓存,以快速满足未来的内容请求.进行了仿真实验,通过与现有机制在包交付率、平均延迟、平均跳数和网络开销方面的性能对比,表明所提出的机制是可行且有效的.

关键词: 移动社交网络;信息中心网络;路由;社区发现

中图法分类号: TP393

中文引用格式: 石峻岭,王兴伟,黄敏.基于 ICN 网络架构的社区感知型 MSN 路由机制.软件学报,2020,31(6):1786-1801.
<http://www.jos.org.cn/1000-9825/5708.htm>

英文引用格式: Shi JL, Wang XW, Huang M. Community aware MSN routing scheme based on ICN architecture. Ruan Jian Xue Bao/Journal of Software, 2020,31(6):1786-1801 (in Chinese). <http://www.jos.org.cn/1000-9825/5708.htm>

Community Aware MSN Routing Scheme Based on ICN Architecture

SHI Jun-Ling¹, WANG Xing-Wei¹, HUANG Min²

¹(School of Computer Science and Engineering, Northeastern University, Shenyang 110169, China)

²(School of Information Science and Engineering, Northeastern University, Shenyang 110819, China)

Abstract: MSN (mobile social network) realizes message delivery by leveraging social relationships of mobile users via cooperation forwarding of nodes. However, with the coming of the big data era, MSN should satisfy the daily increasing content (e.g., video) requests of the mobile users. Considering that ICN (information-centric networking) supports mobility natively, in this study, a community aware routing scheme in MSN is proposed, which is based on ICN architecture. In interest decision, the proposed interest distance metrics among users are calculated based on the interest preferences of users, which are obtained from the content name of the requests of nodes. Then, nodes are detected into interest communities based on the interest distances, and interest packets are routed based on these detected interest communities. In data decision, the proposed encounter regularity metrics are calculated according to the history encounter information of nodes. Then, based on the encounter regularities, nodes are detected into social communities, and data packets are routed based on these detected social communities. Meanwhile, the proposed routing scheme optimizes content caching of nodes based on the detected interest communities and social communities, in order to satisfy the future content requests rapidly. By comparing with the

* 基金项目: 国家杰出青年科学基金(71325002); 国家自然科学基金(61572123)

Foundation item: National Science Foundation for Distinguished Young Scholars of China (71325002); National Natural Science Foundation of China (61572123)

收稿时间: 2017-10-25; 修改时间: 2018-09-13; 采用时间: 2018-11-15

existed schemes on packet delivery, average hops, average delay and network overhead, simulation experiments show that the proposed scheme is feasible and effective.

Key words: mobile social network; information-centric networking; routing; community detection

移动社交网络(mobile social network,简称 MSN)是一种利用移动用户之间存在的社交关系进行无线通信的系统^[1]。随着大数据时代的到来,移动用户对多媒体内容(如视频)的需求日益增加,为满足移动用户的内容需求,配备高效的路由机制尤为重要。现有的 MSN 路由机制^[2-10]均基于已知消息的目的节点进行交付,即假设已知内容提供者地址。然而这种假设在实际中是不可行的,因为一旦内容提供者地址发生变化,请求者将无法通过变化前的地址获取内容。因此,需要寻求一种能够克服这种移动性的路由模式实现内容查找。此外,尽管在 MSN 路由中,通过利用移动用户的社交关系(如兴趣相似程度)能够使包的转发更加准确,并实现改善路由效率的目的,然而由于用户的兴趣等有效的社会信息难以准确获取^[11],因此需要有效的方法实现用户的社交信息挖掘。进一步,在 MSN 中,“存储-携带-转发”式的消息交付模式要求节点对消息进行本地缓存。对于包含内容数据的消息,现有的 MSN 路由仅仅根据消息的跳数^[12]、到达时间等指标对消息进行缓存,没有提取消息中的内容,也没有对内容单独进行缓存。然而,对内容进行有效的缓存,能够更加快速地满足用户后续相同或者相似的内容需求,从而不需要重新从内容提供者处获取内容。因此,需要对现有的 MSN 架构进行改善,以满足用户对内容的需求。

信息中心网络(information-centric networking,简称 ICN)将面向主机的通信方式转变为面向内容的通信方式,关注内容“是什么”,而不关心内容“在哪里”^[13]。将 ICN 的这种以内容为中心的通信方式应用于 MSN 中,用户仅需知道自己所需内容的名字,而不需要知道内容的地址,从而有助于解决 MSN 中由于内容提供者地址变化而导致的无法有效路由的问题。此外,在 ICN 中,内容以名字进行命名,用户的兴趣请求使用的是内容的名字,这些内容名字体现了用户对内容的偏好。通过有效挖掘用户对内容的偏好,能够计算节点的兴趣,获得节点之间兴趣的相似程度,因而有助于解决 MSN 中用户间社交关系难以准确获取的问题。最后,ICN 关注节点的内容缓存,支持节点对缓存内容的优化处理^[14]。当缓存空间不足时,尽可能缓存后续可能被其他用户以较高概率访问的内容,后续被访问概率低的内容不被缓存或被替换。因此,通过在 MSN 的节点中引入内容缓存机制,对内容进行有效管理,可以提高对用户后续内容访问需求的满足程度。基于上述考虑,本文基于 ICN 架构,提出一种 MSN 中的新型路由机制。

社区发现通常是一种能有效解决 MSN 路由的社会学方法,它能够使路由更高效、稳定和具有可扩展性^[15]。在基于社区的路由^[5,8,16-18]中,包通过先发送到目标社区、再发送给目标节点的方式实现交付。与 ICN 相同,本文使用两种包来满足用户对内容的访问需求,即用于查找内容的兴趣包和用于返回内容的数据包。为了实现两种包不同的目标,本文根据不同的社交度量划分了两种社区:一种是基于用户兴趣偏好划分的兴趣社区,在路由兴趣包时,将兴趣包发送给内容提供者的兴趣社区,再发送给内容提供者;另一种是基于节点的相遇规律划分的社交社区,在路由数据包时,将数据包发送给内容请求者的社交社区,再交付至内容请求者。

在 MSN 中,基于社交关系的路由存在多副本和单副本两种方式:多副本路由在将消息转发给下一跳节点后,保存消息的副本并进一步转发,因此交付率相对较高但转发开销巨大;单副本路由在将消息转发给下一跳节点后,节点不存储消息的副本,因此网络开销小,但是需要准确的社交关系才能保证交付率。单副本路由容易陷入局部最优,即容易陷入“死胡同”(dead end)^[16],因为节点总是将消息转发给与目的节点社交关系更紧密(即社会地位更高)的节点,但有时联系更紧密的节点由于突发事件而无法与目的节点及时相遇,因此造成消息不能继续转发而无法交付。本文提出一种消息转发的回溯策略,如果节点持有消息超过预定时间,则不再将消息发送给社会地位更高的节点,而是把消息发送给社交地位较低的节点,以防止路由陷入“死胡同”。

本文的主要贡献如下:基于 ICN 架构提出了一种 MSN 中的路由机制,以满足移动用户对内容的需求;针对兴趣包和数据包,分别提出了社区发现方法,以高效地实现包交付;提出了一种路由转发回溯策略,以解决路由陷入“死胡同”的问题;节点根据划分的兴趣社区和社交社区对内容进行有效缓存,以满足后续用户内容请求。

本文第 1 节介绍相关工作。第 2 节是本文机制的系统框架。第 3 节描述社区发现方法。第 4 节介绍路由决策。第 5 节是仿真与性能评价。第 6 节给出结论。

1 相关工作

关于 MSN 路由已经有一些研究工作.文献[3]提出了一种基于社交簇的路由机制,每个节点都选择与之关系紧密的其他节点形成本地簇,将消息首先转发给目的节点的簇成员节点,再由簇成员节点发送给目的节点.文献[4]在 MSN 中提出了一种截止期限敏感(dead line sensitive)的基于效用的路由模型,如果一个消息在截止期限前成功得到交付,则它的源节点将得到一个积极的收益;否则,源节点不会得到任何收益.文献[5]提出了一种 MSN 中基于社交关系的路由方法,引入社会能量来衡量节点的消息转发能力,将消息转发给社会能量强的节点.文献[6]在 MSN 中提出了一种基于参数最优化的路由协议,在选择中继节点时,综合考虑 3 个社交度量,即 LinkRank、相似度和接触强度.3 个度量的权重由配对学习算法(pair-wise learning algorithm)推导得到.文献[7]面向移动用户提出了一种两段式动态路由转发算法,设计了基于节点社交活跃度的多副本传播策略和基于节点物理接触因子的单副本转发策略.文献[8]提出了一种社区感知的网络模型,将 MSN 转变为仅包含社区的网络,进而给出了一种分布式最优社区感知机会路由算法.文献[9]指出,MSN 呈现出一种嵌套式的分层结构,少数活跃的节点构成网络的核心,而多数不活跃的节点构成网络的边缘.提出了一种上传-下载式的路由协议,消息通过迭代地转发给更活跃的中继节点被上传至网络核心,基于布鲁姆过滤器,从网络核心下载消息到目的节点.文献[10]提出一种基于蚁群算法的 MSN 路由机制,根据节点的传输路径信息得到节点对之间的信息列表.设计信息素的更新方法,在转发数据给其他节点时,提供有效的信息以选择合适的中继节点.与上述工作不同,本文基于 ICN 架构提出了一种 MSN 中的路由机制,根据用户请求中的内容名字挖掘用户之间关于兴趣的社交度量进行路由.

已经有一些研究工作基于 ICN 架构设计了 MSN 路由.文献[17]根据移动节点的移动性,在每个社区中确定一个代理节点,负责接收和转发用户的兴趣包.由于代理节点可能需要同时转发多个兴趣包,根据用户兴趣和请求数量设计需求度,以确定每个兴趣包的优先级.通过比较不同节点彼此相遇的可能,选择转发节点传输数据包.由于具有更高社会地位的节点在网络中更受欢迎,文献[18]令用户将兴趣包发送给具有更高社会地位的节点.如果不能在请求者的社区实现兴趣包的内容查找,它将被转发给其他社区.基于节点的相遇历史,数据包转发给更有可能与请求者相遇的节点.但是,文献[17,18]的社区中存储的为相遇频繁的节点,在划分社区时没有考虑用户对内容的兴趣.然而,有效挖掘用户对内容的偏好,能够有助于为兴趣包查找内容.本文的社区发现方法不仅考虑了节点的相遇规律,同时考虑了用户兴趣的相似程度,能够有效实现内容查找,实现兴趣包和数据包的交付.此外,文献[17,18]没有在路由数据包的过程中对节点缓存的内容有效管理,因此对于后续节点相同或相似内容请求,仍需要从原始的内容提供者那里获取.而本文设计了节点的内容缓存机制,满足后续到达的内容请求.文献[19]通过使用命名数据网络(named data networking,简称 NDN)规则,提出了一种社交感知的 NDN 架构,将具有高物理临近度和内容相似度的亲密用户定义为朋友圈,根据朋友圈相遇频率构建路由表,在朋友圈之间路由兴趣包和数据包.虽然在形成朋友圈的过程中,文献[19]同时考虑了相遇规律与用户的兴趣相似程度,但没有考虑在节点中引入内容缓存机制.此外,作为基于单副本转发的路由,文献[6-9]均存在消息转发陷入“死胡同”的问题.本文路由机制为了避免消息转发陷入“死胡同”,提出了一种回溯策略,可以将陷入“死胡同”的消息发送给具有较低社会地位的节点.

2 系统框架

2.1 系统模块

如图 1 所示,本文提出的基于 ICN 网络架构的社区感知型路由机制(ICN based community aware routing scheme,简称 ICRS)主要包含两个组件,分别是社区发现和路由决策.前者为基于用户的兴趣偏好划分兴趣社区,同时基于节点的相遇规律划分社交社区;后者为负责转发兴趣包的兴趣决策和数据包的数据决策.具体而言,在兴趣决策中包含社区内转发和社区间转发;在数据决策中包含网络内缓存、社区内转发和社区间转发.

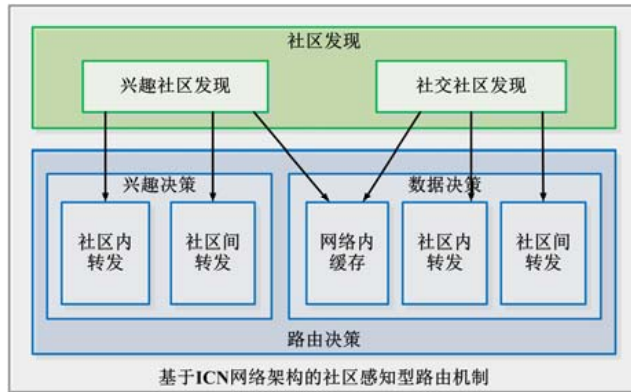


Fig.1 System framework

图 1 系统架构

2.2 工作流程

MSN 分为集中式、分布式和混合式这 3 种网络结构.在集中式 MSN 中,节点通过基础设施与 Internet 相连;在分布式 MSN 中,节点通过设备对设备(device to device)的通信模式进行数据传输;而混合式是这两种网络结构的结合.ICRS 采用混合式网络结构,如图 2 所示.

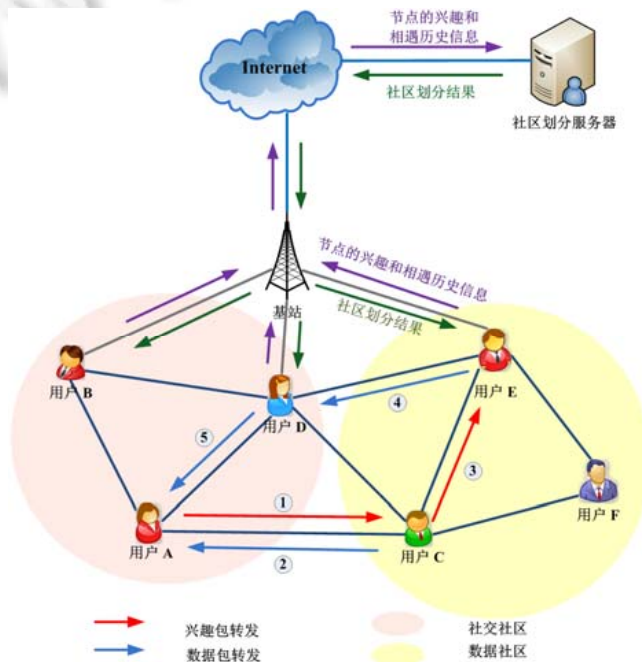


Fig.2 Workflow and scenario

图 2 工作场景与流程

当节点在基站的信号覆盖范围内时(图中节点 B 和 D),它们将自己的兴趣信息(具体为用户对于不同内容分类的历史请求次数)和社交信息(具体为节点与其他节点的历史相遇情况)通过基站和 Internet 发送给负责社区划分的服务器.服务器根据节点上传的兴趣信息和社交信息将节点划分为兴趣社区和社交社区,通过基站下发给各个节点,因此,社区发现组件采用集中式结构.基于得到的社区划分结果,节点相互协作进行路由决策以完

成兴趣包和数据包的交付,因此,路由决策组件为分布式结构。

例如,在图 2 中,假设节点 A, B 和 D 属于同一个社交社区,节点 C, E 和 F 属于同一个兴趣社区.当请求者 A 生成一个兴趣包后, A 根据兴趣决策中的社区间转发(因为 A 不属于目标兴趣社区)将请求发送给 C .如果 C 缓存有请求内容,则它会产生相应的数据包返回给 A ;否则, C 会根据兴趣决策中的社区内转发(因为 C 属于目标兴趣社区)将内容请求发送给 E .假设 E 为内容提供者,则它产生数据包返还给 A .之后, E 根据数据决策中的社区间转发(因为 E 和 A 不属于同一社交社区)将数据包发送给 D , D 根据网络内缓存规则对数据包中的内容进行本地缓存.最后, D 根据数据决策中的社区内转发(因为 D 和 A 属于同一社交社区)将数据包交付给 A .

3 社区发现

3.1 兴趣社区发现

3.1.1 兴趣距离

在 ICN 架构中,请求者生成的请求由分层的内容名字构成,如 $/c/.../c/fname/version/s$.用户的历史请求能够体现用户对于不同类型内容的偏好.当用户对某一类型的内容感兴趣时,其请求会多次包含这一内容对应的字段;否则,用户的请求会较少包含对应字段.因此,本文根据用户的历史请求所包含的内容字段计算用户对于不同内容分类的兴趣偏好.假设存在 n 个不同的内容字段, $field_k$ 为任意内容字段, $1 \leq k \leq n$.设网络中有 m 个节点, R_i 为其中任意节点, $1 \leq i \leq m$.

定义 1(兴趣权值). 兴趣权值为节点对一个字段的历史请求次数.令 $iweight_i^k$ 代表 R_i 对 $field_k$ 的兴趣权值,则当 R_i 生成一个包含 $field_k$ 的请求时:

$$iweight_i^k = iweight_i^k + 1 \quad (1)$$

接下来,根据最大兴趣权值 $iweight_i^{\max}$ 对 $iweight_i^k$ 进行归一化:

$$iweight_i^k = \frac{iweight_i^k}{iweight_i^{\max}} \quad (2)$$

将 n 个兴趣权值降序排列,和它们对应的内容字段一起组成 R_i 的兴趣集合.令 In_i 代表 R_i 的兴趣集合, In_i 中存储的元素为二元组(字段,兴趣权值).假设 $n=2$, R_i 的兴趣集合中包含 $field_2, field_6$ 以及它们对应的兴趣权值,则:

$$In_i = \{ \langle field_2, iweight_i^2 \rangle, \langle field_6, iweight_i^6 \rangle \} .$$

定义 2(兴趣距离). 兴趣距离为两个节点的兴趣集合中,相同字段对应兴趣权值的绝对值之和,如果两个节点的兴趣集合中没有相同的字段,则兴趣距离为 $+\infty$.

假设 R_j 为网络中任意节点, $1 \leq j \leq m$ 且 $i \neq j$.用 $ides_{ij}$ 表示 R_i 和 R_j 之间的兴趣距离,则:

$$ides_{ij} = \sum_{k=1}^n |iweight_i^k - iweight_j^k| \quad (3)$$

兴趣距离衡量和节点之间的兴趣相似度.根据公式(3)可得:两个节点之间兴趣距离的值越小,节点间的兴趣相似度越高.

3.1.2 基于节点兴趣的社区发现

基于文献[20]中的社区划分方法,将节点划分为不同的兴趣社区.首先,根据收集到的节点的历史请求内容的信息,能够计算节点之间的兴趣距离.本文将网络抽象为无向带权图 $G(V, E, W)$,其中: V 是网络中所有的节点集合; E 是边集合(每对节点存在一条边); W 是边的权值集合,每条边的权值为其两端节点兴趣距离的倒数.兴趣社区划分方法步骤如下.

- 1) 将所有节点作为一个初始社区;
- 2) 计算对应无权图中所有边的边介数,再将边介数除以其权重得到边权比;

/*这里的边介数采用最短路径边介数方法,即一条边的边介数是指从某个源节点出发通过该边的最短路径的数目,对所有可能的源节点,重复做同样的计算,并将得到的相对于各个不同的源节点的边介数相加,所得

的累加和为该边的边介数*

3) 删除边权比最高的边,并计算网络的模块度 Q (见公式(4))^[20]来衡量网络中社区结构的显著性.在计算时,若边权比最高的边有多条,则同时移除这些边,并将移除的边和模块度进行存储:

$$Q = \frac{1}{M} \sum_{i,j} \left[\left(a_{ij} - \frac{k_i k_j}{M} \right) \delta(\sigma_i, \sigma_j) \right] \quad (4)$$

其中, a_{ij} 为网络邻接矩阵的元素,若节点 R_i 和 R_j 相连,则 a_{ij} 为边的权重,否则等于 0; δ 为隶属函数,若 R_i 和 R_j 属于同一个社区,则其值为 1,否则等于 0; $M = \sum a_{ij}$, 即网络中边的权重之和.在网络划分结构固定且两节点的边随机连接时,节点间存在边的可能性为 $k_i k_j / (2M)$, 其中, k_i 为节点 i 的点权,计算方法为对连通矩阵的第 i 行求和.

4) 重复步骤 2)、步骤 3),直到所有的边被删除.

当所有边均被删除时,对应于模块度最大的社区划分(每个连通图为一个社区)即为最终的兴趣社区划分.

3.2 社交社区发现

3.2.1 相遇规律

节点间的历史相遇体现了它们之间的相遇规律,包括相遇的频率、持续时间和周期性等.节点间规律强的历史相遇往往意味着它们未来相遇的频率可能高、持续时间可能长、周期性可能比较强.

本文根据节点之间的历史相遇信息计算节点的相遇规律,将其作为划分社交社区的度量.首先,将 R_i 和 R_j 在 τ 时间内的相遇历史划分为 σ 个相等的时间段,其中每个时间段的时长为 $\frac{\tau}{\sigma}$. 令 H_{ij} 代表 R_i 和 R_j 的这 σ 个时间段的集合,则:

$$H_{ij} = \{h_{ij}^p \mid p \in N^*, p \leq \sigma\} \quad (5)$$

定义 3(相遇项). 相遇项代表两节点在一个时间段内是否相遇.令 $eitem_{ij}^p$ 代表 R_i 和 R_j 在 h_{ij}^p 内的相遇项,则:

$$eitem_{ij}^p = \begin{cases} 1, & R_i \text{ 在 } h_{ij}^p \text{ 内与 } R_j \text{ 相遇} \\ 0, & \text{其他} \end{cases} \quad (6)$$

定义 4(单位相遇强度). 一个时间段内的单位相遇强度为相遇项与时间段个数的比值.令 $edsty_{ij}^p$ 代表 h_{ij}^p 内的单位相遇强度,则:

$$edsty_{ij}^p = \frac{eitem_{ij}^p}{\sigma} \quad (7)$$

定义 5(周期比对集). 设 Γ 为比对周期,则一个周期内的时间段个数 noh 为

$$noh = \frac{\sigma \cdot \Gamma}{\tau} \quad (8)$$

将所有 σ 个时间段中上标对 noh 取余后得到余数相等的相遇项定义为一个周期比对集,则可得到 noh 个周期比对集,每个周期比对集中有 $\frac{\tau}{\Gamma}$ 个时间段和 $\frac{\tau}{\Gamma}$ 个相遇项.令 $Compare_{ij}$ 代表 R_i 和 R_j 所有的周期比对集合,则:

$$Compare_{ij} = \{C_{ij}^q \mid q \in N^*, q \leq noh\} \quad (9)$$

其中,

$$C_{ij}^q = \left\{ eitem_{ij}^r \mid r = q \% noh + l \cdot noh + 1, l \in N, l < \frac{\tau}{\Gamma} \right\} \quad (10)$$

定义 6(周期度量). 周期度量是一个周期比对集中所有非零的相遇项之和与这个周期比对集中的相遇项总数量的比值,代表两个节点相遇周期性的强度.令 $period_{ij}^q$ 代表 C_{ij}^q 的周期度量,则:

$$period_{ij}^q = \frac{\sum_{eitem_{ij}^r \in C_{ij}^q \wedge eitem_{ij}^r = 1} eitem_{ij}^r}{\tau / \Gamma} \quad (11)$$

定义 7(单位相遇规律). 两个节点在一个时间段的单位相遇规律为其单位相遇强度和对应周期度量的乘

积.令 reg_{ij}^p 代表 R_i 和 R_j 在 h_{ij}^p 内的单位相遇规律,则:

$$reg_{ij}^p = edsty_{ij}^p \cdot period_{ij}^q \tag{12}$$

其中, $q=p\%noh+1$.

R_i 和 R_j 的相遇规律为它们所有的单位相遇规律的总和.令 reg_{ij} 代表 R_i 和 R_j 的相遇规律,其计算如下:

$$reg_{ij} = \sum_{i=1}^{\sigma} reg_{ij}^p \tag{13}$$

举例计算相遇规律如下:设 $\sigma=10, \tau=50$ 小时,则每个时间段的时长为 $50/10=5h$.假设图 3 表示 R_i 和 R_j 在 50h 的时间内的相遇历史.从图 3 可见:共有 10 个时间段,每个时间段有对应的相遇项和单位相遇强度.

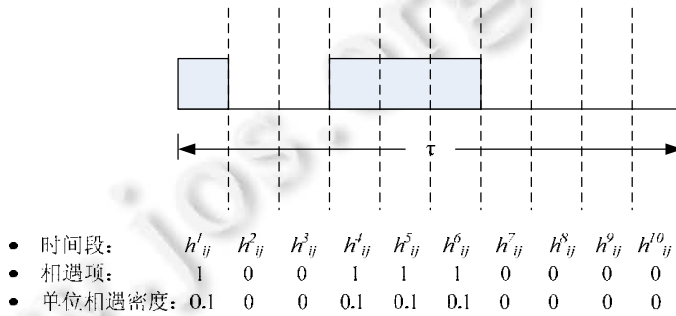


Fig.3 Example of encounter regularity calculation

图 3 相遇规律计算举例

设 $T=10h$,则一个周期内的时间段个数为 $noh=(10 \times 10)/50=2$,两个周期比对集 $C_{ij}^1 = \{eitem_{ij}^2, eitem_{ij}^4, eitem_{ij}^6, eitem_{ij}^8, eitem_{ij}^{10}\} = \{0,1,1,0,0\}$ (时间段的上标对 2 取余等于 0) 和 $C_{ij}^2 = \{eitem_{ij}^1, eitem_{ij}^3, eitem_{ij}^5, eitem_{ij}^7, eitem_{ij}^9\} = \{1,0,1,0,0\}$ (时间段的上标对 2 取余等于 1).进而得到两个周期比对集合的周期度量 $period_{ij}^1 = 2/5, period_{ij}^2 = 2/5$.据此可得 R_i 和 R_j 的单位相遇规律 $reg_{ij}^1 = 0.1 \times (2/5) = 0.04$;同理可得 $reg_{ij}^4 = 0.04, reg_{ij}^5 = 0.04, reg_{ij}^6 = 0.04$,而其余项均为 0.最后,将各单位相遇规律求和,可得 R_i 和 R_j 的相遇规律 $reg_{ij}=0.16$.

3.2.2 基于节点社交关系的社区发现

基于节点社交关系的社区发现与第 3.1.2 节基于节点兴趣的社区发现相类似.不同地,边的权值为两个端节点的相遇规律.

4 路由决策

4.1 兴趣决策

4.1.1 社区内转发

本文中,每个节点存储兴趣社区成员持有的内容情况.令 $Cont_i$ 代表 R_i 存储的兴趣成员持有内容的集合, $Cont_i$ 中存储二元组(内容名字,节点).特别地,如果兴趣社区中有多个成员持有内容,则 $Cont_i$ 中存储对应于这个内容的节点为这些节点中与 R_i 社交规律值最大的节点.令 cn_{ip} 代表兴趣包 ip 请求的内容字段,如果:

$$cn_{ip} \in Cont_i \tag{14}$$

则 R_i 对 ip 执行社区内转发.

根据 $Cont_i$ 能够获知持有请求内容的节点,兴趣包路由即从未知目的节点的情形转化为已知目的节点的情形.由于本文的数据包转发过程已知目的节点(详见第 4.2.2 节),因此兴趣包的社区内转发根据数据包转发策略进行(详见后文第 4.2.2 节和第 4.2.3 节).

4.1.2 社区间转发

如果接收到兴趣包的节点的兴趣社区成员均未持有请求内容,则节点进行社区间兴趣包转发.令 $Field_{ip}$ 代表 ip 请求的内容名字中包含的所有字段的集合,即:

$$Field_{ip} = \{field_k | field_k \in cn_{ip}\} \quad (15)$$

定义 8(兴趣度量). 节点的兴趣度量为请求内容名字包含所有字段的兴趣权值总和.令 $ite_i^{cn_{ip}}$ 代表节点 R_i 对 cn_{ip} 的兴趣度量,则:

$$ite_i^{cn_{ip}} = \sum_{field_k \in Field_{ip}} iweight_i^k \quad (16)$$

为了防止单副本转发方式易造成路由陷入“死胡同”,本文提出如下回溯策略进行兴趣包社区间转发.当节点持有兴趣包的时间超过时间阈值 ψ ,则认为兴趣包陷入“死胡同”.令 $holdtim_{ip}^i$ 代表 R_i 持有 ip 的时间,即,如果:

$$holdtim_{ip}^i > \psi \quad (17)$$

则认为 ip 陷入“死胡同”.

如果兴趣包没有陷入“死胡同”,则将兴趣包转发给具有更高兴趣度量的相遇节点.令 R_x 代表 R_i 的任意相遇节点, $1 \leq x \leq m$ 且 $x \neq i$. $ite_x^{cn_{ip}}$ 代表相遇节点 R_x 对 cn_{ip} 的兴趣度量,则,如果:

$$ite_x^{cn_{ip}} - ite_i^{cn_{ip}} > 0 \quad (18)$$

则 R_i 将 ip 转发给 R_x .

若不满足公式(17),即 ip 陷入“死胡同”,则降低 ip 的转发条件,即不再只将 ip 发送给对兴趣包的内容名字的兴趣度量更高的相遇节点,而是对于具有低于范围 $[0, \xi]$ 的兴趣度量的相遇节点,同样将 ip 转发给它,即,如果:

$$ite_x^{cn_{ip}} - ite_i^{cn_{ip}} < \xi \quad (19)$$

则 R_i 将 ip 转发给 R_x . 兴趣决策见算法 1.

算法 1. 兴趣决策.

输入:兴趣包 ip ;

输出:转发节点或 $Null$.

```

01: IF  $cn_{ip} \in Cont_i$  //存在兴趣社区成员缓存有请求内容
02:   DO 算法 2; //转化为已知目的节点情况
03: ELSE IF  $ne > 0$  //社区间转发,  $ne$  为相遇节点个数
04:   FOR  $x=1$  to  $ne$ , DO
05:     IF  $holdtim_{ip}^i > \psi$  //陷入 dead end
06:       IF  $ite_x^{cn_{ip}} - ite_i^{cn_{ip}} < \xi$  //符合转发条件
07:         RETURN  $R_x$ ;
08:       ELSE
09:         IF  $ite_x^{cn_{ip}} - ite_i^{cn_{ip}} > 0$  //未陷入 dead end 且符合转发条件
10:           RETURN  $R_x$ ;
11:       END FOR
12:     END IF
13:   RETURN  $Null$ ; //不转发继续携带  $ip$ 

```

4.2 数据决策

4.2.1 网络内缓存

由于在 MSN 中的节点存储空间有限,因此,节点无法实现在本地存储所有内容.本文设定节点相遇时交换各自的兴趣集合.当节点接收到一个数据包时,根据数据包中的内容名字,为每一个内容计算一个存储权值.存

储权值的计算考虑所有社交社区和兴趣社区共同成员对这个内容的偏好程度.如果社交社区成员对其偏好程度较大,则存储时间较长.存储权值为所有的社交社区和兴趣社区的共同成员对这个内容名字的兴趣度量的总和.当缓存空间不足时,将权值最低的内容首先丢弃.令 $SCoM_i$ 代表 R_i 的上述共同成员集合.假设 $SCoM_i$ 共包含 d 个节点,令 R_c 代表其中任意一个节点, $1 \leq c \leq d$.令 $cachwei_i^{cn}$ 代表 R_i 对名字为 cn 的内容的缓存权值,则:

$$cachwei_i^{cn} = \sum_{R_c \in SCoM_i} ite_c^{cn} \quad (20)$$

4.2.2 社区内转发

在 MSN 中,节点的移动会造成网络拓扑动态变化.因此在 ICRS 中,为了将数据包发送给请求节点而不受变化的拓扑影响,在兴趣包中标记请求节点.当内容提供者收到兴趣包后产生数据包时,同样为数据包标记请求节点.这样,数据包路由过程即转化为已知目的节点的路由过程.令 R_r 代表一个数据包 dp 的内容请求者,当 R_i 接收到 dp 时, R_i 判断 R_r 是否为 R_i 的社交社区成员.即,如果:

$$R_r \in SCoM_i \quad (21)$$

则 R_i 执行社区内转发规则;否则执行社区间转发规则(后文第 4.2.3 节).

根据第 3.1.2 节社交社区的划分方法,一个社交社区的成员具有相互之间比较紧密的相遇规律,即相互均具有较高的相遇概率.首次相遇路由^[21]将包发送给第 1 次相遇的节点,其转发机制简单且计算复杂度低.本文基于首次相遇路由,提出社区内首次相遇路由(intra-community first contact routing,简称 ICFC)机制,将数据包发送给首次相遇的、从未接收过该数据包的社交社区成员.

令 Pk_x 代表任意相遇节点 R_x 携带的包集合,如果:

$$R_x \in SCoM_i \wedge dp \notin Pk_x \quad (22)$$

则 R_i 将 dp 发送给 R_x .

4.2.3 社区间转发

与第 4.1.2 节兴趣决策中的社区间转发相类似,数据包的社区间转发包括两种转发:陷入“死胡同”时基于回溯策略的转发,没有陷入“死胡同”时的转发.当节点持有 dp 的时间超过阈值 ε ,则认为 dp 陷入“死胡同”.令 $holdtim_{dp}$ 代表节点持有 dp 的时间,即,如果:

$$holdtim_{dp} > \varepsilon \quad (23)$$

则认为 dp 陷入“死胡同”.

如果数据包没有陷入“死胡同”,则将数据包转发给与 R_r 相遇规律更高的相遇节点.令 reg_{xr} 代表任意相遇节点 R_x 与 R_r 的相遇规律,如果:

$$reg_{xr} - reg_{ir} > 0 \quad (24)$$

则 R_i 将 dp 转发给 R_x .

若未陷入“死胡同”,则将 dp 发送给这样的节点 R_x ,其中, R_i 和 R_r 的相遇规律与 R_x 和 R_r 的相遇规律的差值范围在 $[0, \varpi]$, ϖ 是决定回溯节点范围的阈值.即,如果:

$$reg_{ir} - reg_{xr} < \varpi \quad (25)$$

则 R_i 将 dp 转发给 R_x .数据决策见算法 2.

算法 2. 数据决策.

输入:数据包 dp ;

输出:转发节点或 Null.

01: **IF** $ne > 0$

02: **FOR** $x=1$ to ne , **DO**

03: **IF** $R_r \in SCoM_i$ //社区内转发

04: **IF** $R_x \in SCoM_i \wedge dp \notin Pk_x$ //执行 ICFC

05: **RETURN** R_x ;

```

06:     ELSE IF  $holdtim_{dp} > \varepsilon$            //陷入 dead end
07:         IF  $reg_{ir} - reg_{xr} < \varpi$        //符合转发条件
08:             RETURN  $R_x$ ;
09:     ELSE
10:         IF  $reg_{xr} - reg_{ir} > 0$          //未陷入 dead end 且符合转发条件
11:             RETURN  $R_x$ ;
12:         END IF
13:     END FOR
14: RETURN Null;           //不转发继续携带  $dp$ 

```

5 仿真实现与性能评价

5.1 仿真设置

采用两种拓扑:剑桥轨迹(Cambridge trace)^[22]和在文献[18]中使用的合成轨迹(synthetic trace)在机会网络环境(opportunistic network environment,简称 ONE)^[23]上进行仿真实验.在 Cambridge 轨迹中,Haggle Project 将 iMotes 分发给 Cambridge 计算机实验室的 36 个大学生,并收集他们历时 11 天的实验性数据.Synthetic 轨迹包含 120 个节点,它们被分为两个社区:一个社区包含 50 个节点,另一个包含 70 个节点且其中有 20 个节点频繁地在两个社区之间移动.将本文提出的 ICRS 与文献[18]提出的社区间基于社交联系的内容查找(social-tie based content retrieval among communities,简称 STCRC)策略进行对比,因为 STCRC 同样是基于 ICN 和社区结构的路由机制,与 ICRS 较为相似.STCRC 令社区中具有高社交地位的节点记录社区成员所持有内容的情况,节点将内容请求发送给具有较高社会地位的节点.如果社区成员没有匹配的内容,则进行社区间内容查找.一旦获知内容提供者,则根据与内容提供者的相遇可能进行路由.

采用如下 4 个指标进行性能评价:包交付率(packet delivery ratio,简称 PDR)、平均跳数(average HoP,简称 AHP)、平均延迟(average DeLay,简称 ADL)和网络开销(network OverHead,简称 NOH).PDR 是交付的包数量(包括兴趣包和数据包)与生成包的总量的比值,AHP 是所有交付的包所经历的跳数与交付的包的数量的比值,ADL 是所有交付的包经历的时长与交付的包的数量的比值,NOH 是所有生成的包被转发的次数与所有交付的包的数量的比值.

ICRS 有网络内缓存机制,而 STCRC 没有缓存机制.由于本文的网络内缓存机制基于第 3.1.1 节提出的兴趣集,因此无法直接用于 STCRC.为了更好地对比这两种算法的转发机制,为 ICRS 和 STCRC 配置先进先出(first in first out,简称 FIFO)缓存机制,从而产生了两种新的算法,即 ICRS-F 和 STCRC-F.其中,FIFO 缓存机制的规则是:当节点缓存空间不足时,删除最早进入缓存区的内容.仿真参数设置如下: $\sigma=6$, $\xi=\varpi=0.2$.在 Cambridge trace 中, $\tau=24h$, $T=8h$, $\psi=\varepsilon=0.5h$;在 Synthetic trace 中, $\tau=3600s$, $T=1200s$, $\psi=\varepsilon=60s$.

5.2 性能评价

5.2.1 社区结构

每个场景(Cambridge 和 synthetic)的社区划分情况如图 4 和图 5 所示,图中虚线对应的社区划分状态为采用的社区划分结构,此时对应于图 4(a)、图 4(b)、图 5(a)和图 5(b)的模块度分别为 0.354 5、0.240 9、0.493 8 和 0.499 5.此外,通过在 Cambridge 中为 τ 配置不同取值(3h 和 24h),能够得到不同的社交社区结构(不同社区结构的模块度对应于 3h 和 24h 分别为 $Q=0.2338$ 和 0.3545).

为了分析社区划分的结果对路由性能的影响,测试这两种模块度的社交社区结构下的包交付率,如图 6 所示.结果表明,较高的社区结构的模块度能够得到较高的包交付率.

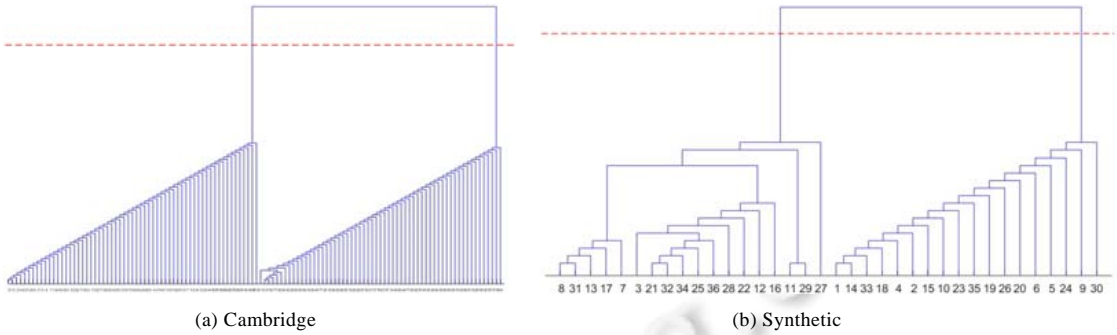


Fig.4 Social community detection

图 4 社交社区划分

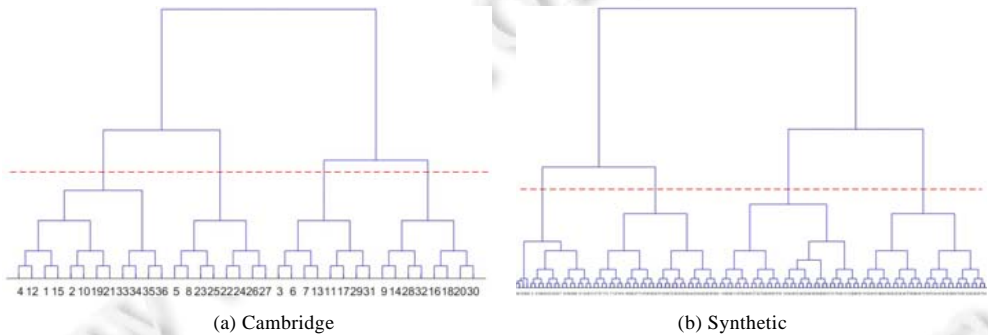


Fig.5 Interest community detection

图 5 兴趣社区划分

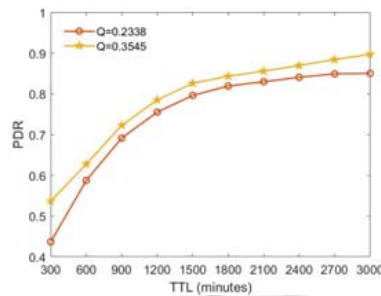


Fig.6 Performance influence by communities

图 6 社区对性能的影响

5.2.2 包交付率

ICRS,STCRC,ICRS-F 和 STCRC-F 在两种拓扑、不同的包生存时间(time to live,简称 TTL)下的包交付率如图 7(a)、图 9(a)所示;在两种拓扑、不同运行时间下的包交付率如图 8(a)、图 10(a)所示.对比 ICRS 和 STCRC,可以发现 ICRS 具有更高的包交付率.原因如下:ICRS 进行有效的网络内缓存,基于其他节点在未来访问的可能性对缓存的内容进行有效的管理.因此,ICRS 能够在有限的包的 TTL 内响应更多的兴趣包,而不是仅仅由源内容提供者获得内容.进一步,在路由兴趣包时,ICRS 基于根据节点兴趣划分的兴趣社区,而在路由数据包时,基于根据节点相遇规律划分的社交社区.这使对兴趣包和数据包的转发更加精确,从而提高了包交付率.然而,STCRC 根据基于节点间的相遇规律得到的社区转发兴趣包和数据包,这会使兴趣包的路由不准确,因为相遇规律强的节点不一定持有请求的内容.

对比 ICRS-F 和 STCRC-F 可知,ICRS-F 具有较高的包交付率.这说明除去网络内缓存为 ICRS 带来的优势,

ICRS 的转发机制相比于 STCRC 仍然具有更好的性能.进一步,对比 ICRS 和 ICRS-F 可知,ICRS 具有更高的包交付率.这是因为 ICRS 使用本文提出的网络内缓存机制,允许节点缓存后续以较大可能被其他节点请求的内容,进而提高了网络的交付率.相比之下,FIFO 缓存机制在缓存空间不足时,仅仅将先到达缓存空间的内容进行删除,没有考虑内容是否可能被其他节点在未来请求.同样,由于 STCRC-F 进行了网络内缓存,它比 STCRC 具有更高的包交付率.

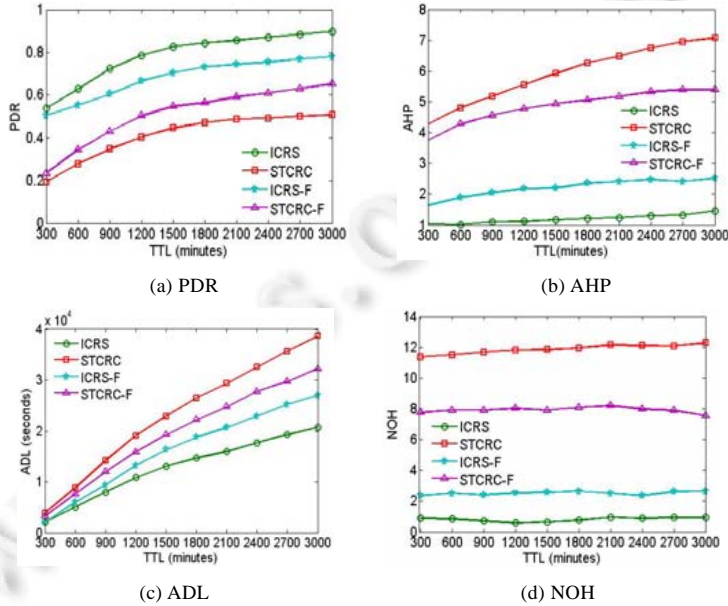


Fig.7 Cambridge (varying TTL)

图 7 Cambridge(变化 TTL)

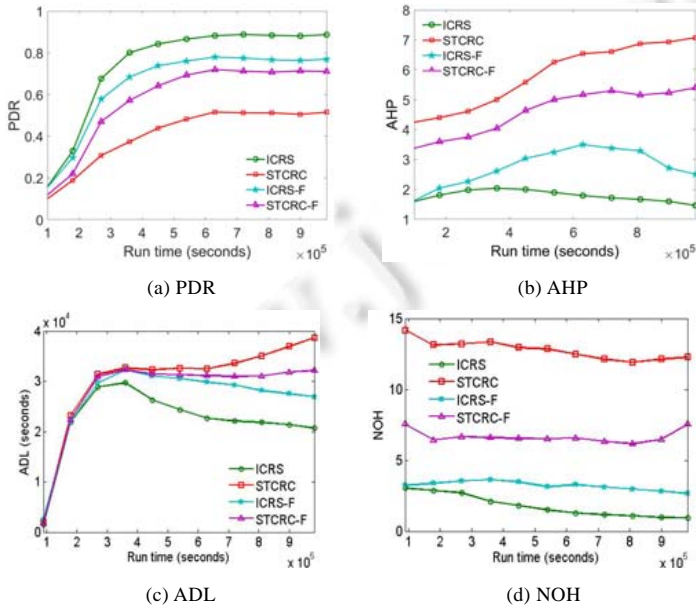


Fig.8 Cambridge (varying running time)

图 8 Cambridge(变化运行时间)

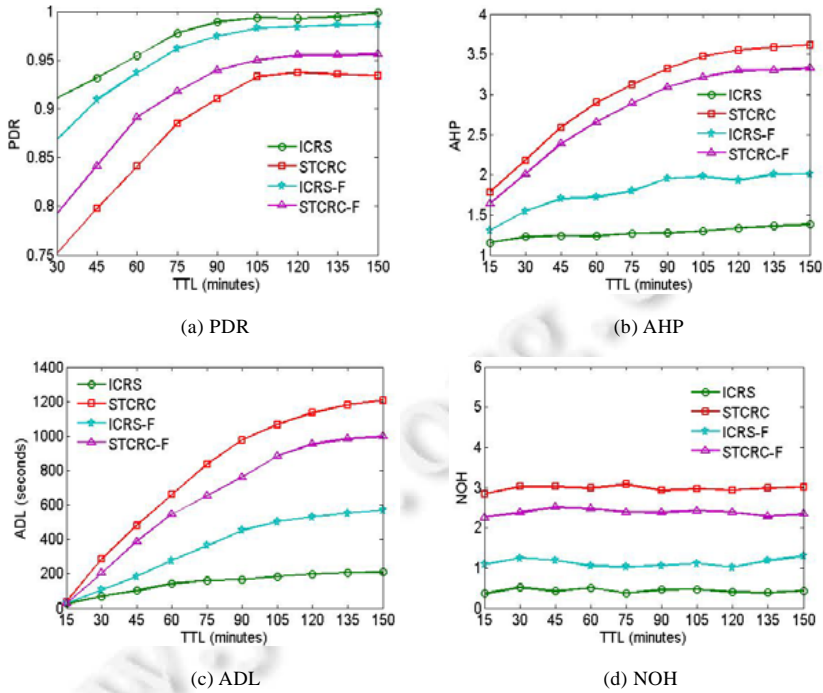


Fig.9 Synthetic (varying TTL)

图 9 Synthetic(变化 TTL)

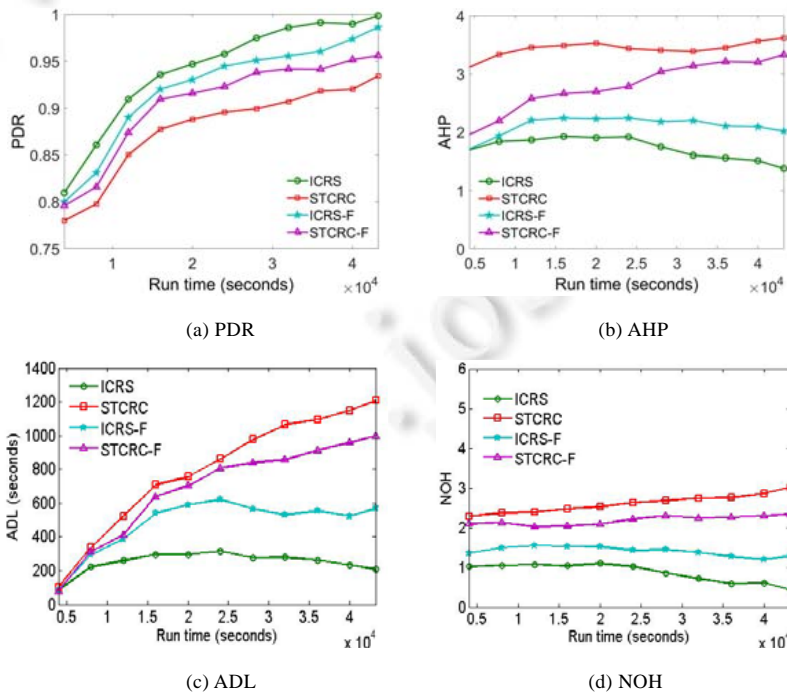


Fig.10 Synthetic (varying running time)

图 10 Synthetic(变化运行时间)

5.2.3 平均跳数

ICRS,STCRC,ICRS-F 和 STCRC-F 在两种拓扑、不同的包 TTL 下的平均跳数如图 7(b)、图 9(b)所示;在两种拓扑、不同运行时间下的平均跳数如图 8(b)、图 10(b)所示.对比 ICRS 和 STCRC 可以发现:ICRS 具有更低的平均跳数,且跳数在 1 和 2 之间.之所以 ICRS 的平均跳数只有不到 2 跳,一方面是因为 ICRS 进行了网络内缓存的缘故.根据 ICRS 的网络内缓存规则,节点基于其社交社区和兴趣社区的共同成员对内容的兴趣偏好总和进行计算,并决定缓存时间.因此,得益于网络内缓存,节点能够以最大概率从其邻近的其他节点处获得感兴趣的内容.另一方面,是因为实验中内容数量和节点的缓存空间的设置导致的.

在 Cambridge 轨迹中,实验设置 36 个不同内容种类,节点的缓存空间设置为 10 个内容(假设内容尺寸相同).由于实验中节点对不同内容的请求符合幂律分布(实验中幂律分布的指数设为 2.5),因此经常被用户请求的内容不超过 8 个(约为 $36 \times 20\%$).所以,大多数的请求能够在不到 2 跳获得内容.

在 Synthetic 轨迹中,其原因类似,只是不同内容种类设置为 120 个,节点的缓存空间设置为 30 个.对于 STCRC,当节点为兴趣包查找内容提供者时,首先需要从社区内中心度最高的节点处获知社区内是否有成员持有该内容.如果社区内没有内容提供者,则需要进一步向社区外节点进行查找.这种内容查找方法使兴趣包的检索经历了较多转发节点,从而导致平均跳数较高.在图 7(a)、图 7(b)中可见,ICRS 和 ICRS-F 的 AHP 呈现下降趋势.这是因为随着运行时间的增加,社区结构趋于稳定,使路由过程的准确性得到了提高,包能够经历更少的跳数得到交付.

5.2.4 平均延迟

ICRS,STCRC,ICRS-F 和 STCRC-F 在两种拓扑、不同的包 TTL 下的平均延迟如图 7(c)、图 9(c)所示;在两种拓扑、不同运行时间下的平均延迟如图 8(c)、图 10(c)所示.对比 ICRS 和 STCRC 可以发现,ICRS 具有更低的平均延迟.ICRS 依据不同的社区划分结果分别对兴趣包和数据包进行路由,适当使用的社区加快了包交付的过程.虽然 STCRC 同样基于社区进行路由,然而,当路由兴趣包时,STCRC 根据基于相遇规律划分的社区无法有效找到内容提供者,从而使兴趣包交付具有较长时间的延迟.

5.2.5 网络开销

ICRS,STCRC,ICRS-F 和 STCRC-F 在两种拓扑、不同的包 TTL 下的网络开销如图 7(d)、图 9(d)所示;在两种拓扑、不同运行时间下的网络开销如图 8(d)、图 10(d)所示.对比 ICRS 和 STCRC 可以发现,ICRS 具有更低的网络开销.得益于网络内缓存和准确的社区划分,ICRS 减少了对包无用的转发,因此具有较低的网络开销.

在图 7(d)、图 9(d)中,对于 STCRC,由于包所经历的较高平均跳数,因此当 TTL 过期时包还未实现交付,使其转发没有对全网范围内的最终交付产生贡献,导致较高的网络开销.

在图 8(d)和图 10(d)中可见,ICRS 的 NOH 呈现下降趋势.这是因为随着运行时间的增加,社区结构趋于稳定;同时,随着运行时间增加,网络内缓存趋于优化,节点存储后续被其他节点请求的重复或者相似的内容,使路由开销下降.

从图 7~图 10 可以观察到,Synthetic 比 Cambridge 总体呈现更好的性能,即更高的包交付率、更低的包交付延迟、平均跳数和网络开销.原因如下.

- Synthetic 轨迹是 120 个节点所组成的较密集拓扑,运行时间为 43 200s,节点之间的相遇频率较高且相遇间隔较小;
- 而 Cambridge 轨迹是 36 个点所组成的较稀疏拓扑,运行时间为 987 530s,节点之间的相遇频率较低且相遇间隔较大.

因此,Synthetic 能够比 Cambridge 在有限的生存时间 TTL(time to live)内用更短的时间、更少的跳数和开销交付更多的包.

6 结 论

对于移动社交网络,提出一种基于信息中心网络架构的路由机制.通过节点的历史请求内容,刻画节点的兴

趣差异度量,并划分兴趣社区;依据兴趣社区,进行兴趣包转发.通过节点的相遇历史,刻画节点间的相遇规律度量,并划分社交社区;依据社交社区,进行数据包转发.实验结果表明,本文提出的机制在包交付率、平均跳数、平均延迟和网络开销上均优于对比机制.在实际网络背景下进行实验,进一步验证和提高本文机制的实用性,是未来研究工作的重点.

References:

- [1] Hu X, Chu THS, Victor CML, Edith CHN, Philippe K, Henry CBC. A survey on mobile social networks: Applications, platforms, system architectures, and future research directions. *IEEE Communications Surveys & Tutorials*, 2015,17(3):1557–1581.
- [2] Eyuphan B, Boleslaw KS. Exploiting friendship relations for efficient routing in mobile social networks. *IEEE Trans. on Parallel and Distributed Systems*, 2012,23(12):2254–2265.
- [3] Feng Z, Nan Z, Li W. Effective social relationship measurement and cluster based routing in mobile opportunistic networks. *Sensors*, 2017,17(5):1109.
- [4] Xiao M, Wu J, Huang H, Huang L, Yang W. Deadline-Sensitive opportunistic utility-based routing in cyclic mobile social networks. In: *Proc. of the Int'l Conf. on Sensing, Communication, and Networking*. Seattle: IEEE, 2015. 301–309.
- [5] Li F, Jiang H, Li H, Wang M, Abdeldjalil T. SEBAR: Social energy based routing for mobile social delay tolerant networks. *IEEE Trans. on Vehicular Technology*, 2017,66(8):7195–7206.
- [6] Yan H, Jing Qi, Zhen L, Tao J. APPOW: An advanced routing protocol based on parameters optimization in the weighted mobile social network. *China Communications*, 2016,13(S1):107–115.
- [7] Wang X, Lin Y, Zhang S, Cai Z. A social activity and physical contact-based routing algorithm in mobile opportunistic networks for emergency response to sudden disasters. *Enterprise Information Systems*, 2015,3(2):1–30.
- [8] Xiao M, Wu J, Huang L. Community-Aware opportunistic routing in mobile social networks. *IEEE Trans. on Computers*, 2014, 63(7):1682–1695.
- [9] Zheng H, Wu J. Up-and-Down routing through nested core-periphery hierarchy in mobile opportunistic social networks. *IEEE Trans. on Vehicular Technology*, 2017,99:1–15.
- [10] Wu YF, Zhu YQ, Yang Z. Routing algorithm based on ant colony optimization for mobile social network. In: *Proc. of the IEEE/ACIS Int'l Conf. on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing*. Kanazawa: IEEE, 2017. 297–302.
- [11] Girolami M, Chessa S, Caruso A. On service discovery in mobile social networks: Survey and perspectives. *Computer Networks*, 2015,88:51–71.
- [12] Burgess J, Gallagher B, Jensen D, Levine BN. MaxProp: Routing for vehicle-based disruption-tolerant networks. In: *Proc. of the IEEE Infocom*. Barcelona: IEEE, 2007. 1–11.
- [13] Ahlgren B, Dannowitz C, Imbrenda C, Kutscher D, Ohlman B. A survey of information-centric networking. *Communications Magazine IEEE*, 2011,50(7):26–36.
- [14] Xylomenos G, Ververidis CN, Siris VA, Fotiou N, Tsilopoulos C, Vasilakos X, Katsaros KV, Polyzos GC. A survey of information-centric networking research. *IEEE Communications Surveys & Tutorials*, 2014,16(2):1024–1049.
- [15] Xiao M, Wu J, Huang L. Community-Aware opportunistic routing in mobile social networks. *IEEE Trans. on Computers*, 2014, 63(7):1682–1695.
- [16] Hui P, Crowcroft J, Yoneki E. Bubble rap: Social-based forwarding in delay tolerant networks. *IEEE Trans. on Mobile Computing*, 2010,10(11):1576–1589.
- [17] Xu ZJ, Su Z, Xu QH, Qi QF, Yang TT, Li JT, Fang DF, Han B. Delivering mobile social content with selective agent and relay nodes in content centric networks. *Peer-to-Peer Networking and Applications*, 2017,10(2):296–304.
- [18] Lu Y, Gerla M, Le T, Rabsatt V, Kalantarian H. Community aware content retrieval in disruption-tolerant networks. In: *Proc. of the Ad Hoc Networking Workshop*. Piran: IEEE, 2014. 172–179.
- [19] Pu L, Chen X, Xu J, Fu X. sNDN: A social-aware named data framework for cooperative content retrieval via D2D communications. In: *Proc. of the Int'l Workshop on Mobility in the Evolving Internet Architecture*. Paris: ACM, 2015. 14–19.

- [20] Newman ME, Girvan M. Finding and evaluating community structure in networks. *Physical Review E Statistical Nonlinear & Soft Matter Physics*, 2004,69(026113):1-16.
- [21] Jain S, Fall K, Patra R. Routing in a delay tolerant network. *ACM Sigcom Computer Communication Review*, 2004,34(4):145-158.
- [22] <http://www.haggleproject.org>
- [23] Keranen A, Ott J, Karkkainen T. The ONE simulator for DTN protocol evaluation. In: *Proc. of the 2nd Int'l Conf. on Simulation Tools and Techniques*. Rome: ACM, 2009. 1-10.



石峻岭(1989-),女,辽宁沈阳人,博士生,CCF 专业会员,主要研究领域为移动社交,网络路由.



黄敏(1968-),女,博士,教授,博士生导师,主要研究领域为智能算法设计与优化,调度理论与方法.



王兴伟(1968-),男,博士,教授,博士生导师,CCF 高级会员,主要研究领域为未来互联网,云计算,网络空间安全.

www.jos.org.cn

www.jos.org.cn