

具有传输成功率保证的 概率路由算法的 TTL 预测模型

马学彬,王应彪,高瑞超,张晓娟

(内蒙古大学计算机学院,内蒙古自治区呼和浩特 010021)

摘要: 概率路由算法是机会网络中一种常用的路由算法,其 TTL 的设置将直接影响路由的性能. 本文首先利用马尔可夫链对概率路由的转发过程进行建模,并从消息的传输成功率和传输延时方面进行性能评估,不仅可以计算出一个消息从源节点到目的节点所需的传输延时,还可以预测出在给定传输成功率下所需要的最短消息生存周期 TTL,这对于 TTL 的设置具有理论上的指导作用. 其次,本文建立的预测模型还可以使一些无法到达目的节点的消息及时被删除,从而减少不必要的转发,达到节约网络资源的目的. 最后,在两个真实数据集中的实验结果表明,本文提出的模型能够为概率路由中 TTL 值的准确设置提供有效参考.

关键词: 机会网络; 概率路由算法; 预测模型; 传输成功率; 传输延时

中图分类号: TN92 **文献标识码:** A **文章编号:** 0372-2112 (2018)11-2679-09

电子学报 URL: <http://www.ejournal.org.cn> **DOI:** 10.3969/j.issn.0372-2112.2018.11.015

The TTL Prediction Model of Probability Routing Algorithm with the Transmission Rate Guaranteed

MA Xue-bin, WANG Ying-biao, GAO Rui-chao, ZHANG Xiao-juan

(School of Computer, Inner Mongolia University, Hohhot, Inner Mongolia 010021, China)

Abstract: Probabilistic routing algorithm is a common routing algorithm in opportunistic networks, and its TTL will directly affect the routing performance. Firstly, in the paper we uses Markov chain model to evaluate the transmission rate and the transmission delay of probabilistic routing. It can not only calculate the transmission delay for a message from a source node to a destination node, but also can predict the shortest TTL in a given transmission rate, which can provide theoretical guidance for the setting of TTL. Secondly, this prediction model can delete the messages which cannot be relayed to destination nodes to reduce unnecessary forwarding. Therefore, The model can be used to save the network resources maximally. Finally, we experimentally evaluate our model on two real data set, and demonstrate that it can provide a powerful reference for the setting of TTL values accurately in probabilistic routing.

Key words: opportunistic networks; probability routing algorithm; prediction model; transmission rate; transmission delay

1 引言

机会网络 (OppNets)^[1] 不同于传统的网络,其采用的路由模式也不是传统网络所使用的“存储—转发”模式,而是一种适用于高移动性节点的“存储—携带—转发”模式,所以传统的路由协议不再适用于机会网络,需要设计和开发符合机会网络应用环境的路由协议.

路由协议一直是机会网络的一个研究热点,文献

[2]中,针对喷雾路由自适应能力不强的问题,提出了 ASSR 路由协议. 文献[3]中针对 AD Hoc 网络中按需距离矢量路由协议开销大的问题,提出了 LAODV 协议. 本文将路由协议分为三类. 一类是通过增加消息副本数来提高消息传输成功率并降低消息传输延时的多副本路由协议. 这类路由比较典型的有 Epidemic^[4]、Spray&Wait^[5]、Spray&Focus^[6]等. 第二类是通过选择合适的中继来提高网络性能的单副本概率路由协议,这

类路由中典型的协议有 Two-hop^[7]、MV^[8]等. 但为了保持网络中的负载均衡, 机会网络中使用最多的是第三类路由协议, 这类典型的路由协议包括 ProphetRouter^[9]、Bubble Rap^[10]、SimBet^[11]等. 由于基于社团结构的路由已成为机会网络最有效的路由解决方案之一^[12,13], 因此, 本文针对存在社团结构的第二和第三类的概率路由进行研究.

2 相关工作研究

机会网络中整个网络的性能与网络拓扑具有很大的关系, 因为机会网络中消息的传递是依赖于人的接触的, 人的移动带来了转发机会. 在对路由算法进行评估时, 对人类的移动模型模拟的准确性会很大程度上影响路由算法评估模型的准确性, 所以对人类移动模型的模拟是影响路由算法评估的重要因素.

在对人类移动模型建模时, 需要知道人们之间的间隔接触时间, Hui Pan 在文献[14]中提出现实中人们的间隔接触时间服从幂律分布, 并且研究者发现接触时间间隔具有“胖尾性”, 这正是幂律分布的特点. 所以, 在本文中假设间隔接触时间服从幂律分布.

在文献[1]提出之后, 研究者在使用人类的接触规律时, 开始使用幂律分布或是帕累托分布的假设. C Boldrini在文献[15~18]中都假设人类的间隔接触时间服从帕累托分布. 文献[15]对社团结构类路由算法(非概率路由算法)进行建模评估, 作者发现最好的路由算法是两跳路由, 但是作者并没有给出计算消息传输延时的具体方法, 只是从消息传输延时收敛的情况下讨论路由算法的性能. 在文献[18]中, C Boldrini 给出了各类社团结构类路由算法消息传输延时收敛的条件, 并证明了对于不同的网络结构应该采用不同的路由算法. 但是, 作者同样并没有给出消息传输延时的具体计算方法, 只给出了消息传输延时的收敛条件.

在文献[16,17]中不再将建立的路由算法评估模型拘泥于一种特定的假设, 作者将模型应用到了更广的范围. 虽然文献[16,17]虽然对概率路由算法的评估模型给出了评估方法, 但主要是在单指标下对单副本概率路由算法进行评估, 对于多副本概率路由算法只能计算出延迟收敛的条件, 而不能计算出多副本概率路由算法的具体传输延时. 文献[19,20]都是采用马尔可夫链模型对传染病路由算法进行建模评估, 并给出了当消息覆盖整个网络时的传输延时表达式, 同时还给出了该表达式的上下界.

目前还没有人研究在保证传输成功率下概率路由算法 TTL 的预测, 之前的研究只是评估概率路由算法的延时, 这也是本文的研究重点.

3 模型设计

3.1 网络设计

假设网络中共有 N 个移动的节点, 不考虑节点的动态加入及退出行为, 每对节点只能在每次接触的时候交换消息, 且在每次接触过程中总能完成消息的交付. 另外, 假设所有的消息都是原子信息, 即不可以再被分割, 且每对节点接触时, 消息的传输时间相对于整个消息传输的延迟时间可以忽略不计. 因为本节是从理论上建立模型计算概率路由的消息传输延时, 所以假设每个节点的缓存空间是无限的, 以排除缓存空间对消息传输延时的影响.

消息的延时主要受两个因素的影响: 间隔相遇时间 $I_{m,n}$ 和剩余间隔相遇时间 $R_{m,n}$. 其中, m, n 为任意两节点, t_{\min} 为全网中的最小间隔相遇时间. 假设两节点对之间的间隔相遇时间服从幂律分布, $\alpha_{m,n}$ 为 $I_{m,n}$ 的幂指数, 则 $R_{m,n}$ 也服从幂律分布, 且其幂指数 $\alpha_{m,n}^* = \alpha_{m,n} - 1$ 且 $\alpha_{m,n} > 1$, 对于所有的 $\alpha_{m,n}$ 符合某一分布 $F(\alpha)$, 其均值为 μ_α . 当某一个携带消息的节点遇到其它节点并将该消息或是该消息的副本交付给与其相遇的节点的概率为 $p_{m,n}^\varphi$, 记 $p_{m,n}^\varphi$ 的均值为 p^φ , 其中 φ 为概率路由的详细转发策略.

由于该模型模拟的是两个节点之间的相遇过程, 所以需要知道任意两个节点之间的平均间隔相遇时间. 另外, 因为在机会网络中, 一个消息到达网络中的某一节点的时间是随机的, 所以也需要知道任意两节点之间的剩余间隔相遇时间. 由于节点间的间隔相遇时间服从幂律分布, 所以可求出平均间隔相遇时间和剩余间隔相遇时间的平均值, 分别为:

$$E(I_{m,n}) = \frac{t_{\min}}{\alpha_{m,n} - 1} \quad (1)$$

$$E(R_{m,n}) = \frac{t_{\min}}{\alpha_{m,n}^* - 1} = \frac{t_{\min}}{\alpha_{m,n} - 2} \quad (2)$$

由公式(1)(2)可知, 若要使间隔相遇时间的期望是收敛的, 必须满足 $\alpha_{m,n} - 1 > 0$, 此时表示 m, n 两节点总会相遇; 若使剩余间隔相遇时间的期望收敛, 则应满足 $\alpha_{m,n} - 2 > 0$ 表示从任何时刻开始, m, n 两节点必能在有限的时间内相遇. 然后可得到节点之间的连接频率为:

$$u_{m,n} = \frac{\alpha_{m,n} - 1}{t_{\min}} \quad (3)$$

由于是给定的常数, 所以用 $\alpha_{m,n}$ 代替 $u_{m,n}$ 作为两节点之间的连接频率, 在每一个节点上使用一个连接矩阵存储其与各节点之间的连接频率 $\alpha_{m,n}$, 这些连接频率可以用真实数据集通过最大似然估计得到.

3.2 马尔可夫链模型

设任意一个消息从源节点被转发到目的节点共经

历 $S(1 \leq S \leq N-1)$ 个节点. 其中, 一个消息所在的节点用来表示消息的一个状态, 同时, 消息的转发过程就表示状态迁移的过程. 因此, 若一个消息经历 S 个节点, 则这个消息共有 S 个状态. T_M 表示消息在前 M 个状态处所花费的时间; $T_{k,k+1}$ 表示消息从状态 k 转移到状态 $k+1$ 所花费的时间.

模拟一个马尔可夫链的转发模型, 除了上面的状态空间, 还需要各个状态之间的转移概率. 在每个节点上, 用一个概率矩阵存储消息各状态之间的转移概率. 由于在转发过程中一对节点之间定义的转发概率是不对称的, 所以每个节点上的矩阵是一个完全矩阵, 且每个状态到自身的转移概率为零, 所以该矩阵的主对角线上的元素除最后一个都为零. 该矩阵的最后一行表示的是消息已经达到目的节点, 所以此时马尔可夫链处于吸收态, 即退出该状态的概率为零. 上述状态转移概率矩阵如下:

$$P = \begin{bmatrix} 0 & P_{1,2} & \cdots & P_{1,S-1} & P_{1,S} \\ P_{2,1} & 0 & \cdots & P_{2,S-1} & P_{2,S} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

对于任意一个消息, 源节点为 I , 目的节点为 d , 则

$$\sum_I^d P_{i,j} = 1.$$

由马尔可夫链的性质可知:

命题 1 马尔可夫链中消息退出 M 个状态共消耗的时间为:

$$T_M = \sum_{k=1}^M T_{k,k+1} \quad (4)$$

当 $M=S$ 时表示消息已经交付到目的节点, 并且 T_M 为递增的.

定理 1 (消息传输延时) 一个消息从源节点 I 到目的节点 d 的总的传输延时为:

$$E(D) = \begin{cases} 0, & I = d \\ E(T_1) + \sum_{k=2}^S P_{k-1,k} * E(T_{k,k+1}), & I \neq d \end{cases} \quad (5)$$

证明 由上面描述的马尔可夫链可知, 当消息已经到达目的节点 d 时, 马尔可夫链处于吸收态, 状态不再转移, 所以此时无延时发生; 当源节点和目的节点不是同一节点时, 则消息总延时为马尔可夫链中消息经过的各个状态的延时之和.

3.3 成功率保证的状态转移

在机会网络中, 节点主要以手持设备为主, 而手持设备的存储空间一般是有限的, 因此需要对消息的生存期 TTL 严格限制, 以使网络中的消息副本尽快被删

除, 节约网络资源. 这里的 TTL 与传统意义上以跳数作为其度量指标的定义不同, 根据机会网络转发延时较长的特点, 本文是以时间作为 TTL 的度量指标.

定义 1 设消息的生存周期为 TTL, 则在该 TTL 下消息的传输成功率为:

$$P(E(T_{M_{\max}}) \leq \text{TTL})$$

其中, M_{\max} 表示在给定的网络中在所有传输成功的消息中的最大跳数, 所以只要是不超过该跳数即表示该消息都有可能成功传递到目的节点. 而如果消息跳数超过 M_{\max} , 该消息很有可能已经超时, 将会被丢弃. $T_{M_{\max}}$ 表示与之对应的时间, 而 $E(T_{M_{\max}})$ 表示其平均时间. 这样就可以将该传输成功率问题转化成寻找网络中消息的平均路径长度不超过 M_{\max} 的消息所占的比例.

对于任意给定的网络, 给出 TTL 就可以通过上面的定义 1 求出其对应的最大消息跳数 M_{\max} .

定义 2 设消息的生存周期为 TTL, 则该网络中允许的最大消息转发跳数为:

$$M_{\max} = \max \{M: E(T_M) \leq \text{TTL}\} \quad (6)$$

由李昊在文献[21]的研究可知, 当网络中节点的度服从幂律分布时, 则网络中的路径长度具有如下关系:

命题 2 设网络中的节点的度分布服从幂指数为 λ 的幂律分布, C 为该网络的聚类系数, 则该网络的路径长度 l 的分布为^[20]:

$$P(l) = \left(1 - \frac{1}{\lambda}\right)^{\frac{C}{3}} * l^{-\frac{\log(C-\lambda)}{\log \lambda} - 1} \quad (7)$$

通过上面的分析, 可以得到如下定理:

定理 2 在机会网络中, 已知节点度分布服从幂指数为 λ 的幂律分布, 聚类系数为 C , 若要求该网络中消息传输的平均成功率不小于 β , 则需满足:

$$\left(1 - \frac{1}{\lambda}\right)^{\frac{C}{3}} \sum_{M=1}^{M_{\max}} M^{-\left(\frac{\log(C-\lambda)}{\log \lambda} + 1\right)} \geq \beta \quad (8)$$

证明 由定义 1 可知, 若使网络中消息传输的成功率不小于 β , 即使得 $P(E(T_{M_{\max}}) \leq \text{TTL}) \geq \beta$ 在本文中路径长度 l 对应于马尔可夫链的状态空间的大小, 即消息被转发的次数, 所以令 $l=M$. 这样就可以将该传输成功率问题转化为寻找网络中平均路径长度小于等于 M_{\max} 的消息所占的比例, 即 $P(M \leq M_{\max}) \geq \beta$.

由公式(7)可知 M 的分布函数为:

$$P(M) = \left(1 - \frac{1}{\lambda}\right)^{\frac{C}{3}} * M^{-\frac{\log(C-\lambda)}{\log \lambda} - 1}$$

所以:

$$\begin{aligned} P(M \leq M_{\max}) &= \sum_{M=1}^{M_{\max}} P(M) \\ &= \sum_{M=1}^{M_{\max}} \left(1 - \frac{1}{\lambda}\right)^{\frac{C}{3}} * M^{-\frac{\log(C-\lambda)}{\log \lambda} - 1} \end{aligned}$$

$$= \left(1 - \frac{1}{\lambda}\right)^{\frac{C}{3} M = M_{\max}} M^{-\frac{\log(C-\lambda)}{\log \lambda}} - 1$$

由此可知定理 2 成立.

根据定理 2, 在给定网络中, 对于任意给定一个要求的最低消息传输成功率 β , 都可以获得一个网络允许的最大传输路径 M_{\max} , 然后根据公式 (6) 可以获得满足该要求的最小消息生存周期 TTL.

定理 3 在机会网络中, 已知节点度分布服从幂指数为 λ 的幂律分布, 聚类系数为 C , 若要求该网络中消息传输的平均成功率不小于 β , 则需要设定的 TTL 为:

$$\text{TTL} = E[T_{M_{\max}}] \quad (9)$$

其中, M_{\max} 是和 β 有关的, 可以表示为:

$$\sum_{M=1}^{M=M_{\max}} M^{-\frac{\log(C-\lambda)}{\log \lambda}} - 1 \geq \beta \left(1 - \frac{1}{\lambda}\right)^{-\frac{C}{3}}$$

证明 对于固定网络, 给定任意一个 β , 由定理 2 可以得到一个对应的 M_{\max} , 且 M_{\max} 与 β 的关系通过公式 (8) 简单的移项就可得到. 又因为每一个 M_{\max} 都对应网络中实际允许的消息的最大传播路径, 所以也就得到唯一的 $E[T_{M_{\max}}]$, 然后由公式 (6) 可以得到定理 3 的结果.

4 概率路由模型的状态转移过程

4.1 单副本概率路由的状态转移

对于不同类型的路由, 变化的是 $T_{k,k+1}$ 和 $P_{k,k+1}$, 所以本节接下来求状态转移时间和状态转移概率.

设 $p_{m,n}^{\varphi}$ 表示在某种转发策略 φ 定义下两节点相遇后进行消息转发的概率, 两节点的相遇频率为 $u_{m,n}$, 则节点 m 遇到节点 n 并转发给 n 的频率为 $p_{m,n}^{\varphi} * u_{m,n}$, 而在整个网络中, 节点 m 退出状态 k , 实际上相当于找节点 m 遇到除本身之外的所有节点的剩余间隔相遇时间的最小值, 即 $\min\{R_{m,n}\}$, 其中 n 可以是除 m 外的所有可能节点, C_m 为 n 节点的集合, 概率路由的转发过程就是寻找比自己更频繁遇见目的节点的中继节点, 也就是 m 节点寻找最先与之相遇的且比自己更优的节点做中继, 该过程实际上是计算 $\min\{R_{m,n}\}$, 其中 $n \in C_m$. 由幂律分布的知识可知: 当随机变量 $R_{m,n}$ 服从幂指数为 $\alpha_{m,n}^* = \alpha_{m,n} - 1$ 的幂律分布时, 则 $\min\{R_{m,n}\}$ 也服从幂律分布, 幂指数为:

$$\alpha_m = \sum_{n \in C_m} \alpha_{m,n}^* - 1 = \sum_{n \in C_m} (\alpha_{m,n} - 1) - 1 \quad (10)$$

则节点 m 在状态 k 与其它节点相遇的平均频率为:

$$u_m = \sum_{n \in C_m} u_{m,n} - 1 \quad (11)$$

而在实际中, 退出状态 k 不仅需要考虑到两节点的相遇情况, 对于特定的路由 φ , 还应该考虑到两节点相遇后

做出转发决定的概率, 所以当两节点相遇时, 消息携带者将消息转发给该相遇节点的概率为 $p_{m,n}^{\varphi}$ 时, 则消息从状态 k 转发到其下一个真正的中继的真实概率应该是:

$$p_m^{\varphi} = \sum_{n \in C_m} p_{m,n}^{\varphi} * \frac{u_{m,n}}{u_m} \quad (12)$$

通过上面的分析就可以获得马尔可夫链在状态 k 与 $k+1$ 之间的转移时间.

定理 3 (单副本状态转移时间) 对于单副本概率路由, 马尔可夫链从状态 k 转移到状态 $k+1$ 的时间为:

$$E(T_{k,k+1}) = \frac{t_{\min}}{\sum_{n \in C_m} p_{m,n}^{\varphi} * u_{m,n}} \quad (13)$$

证明 由前面分析知, 消息从状态 k 退出到状态 $k+1$ 的时间是由节点 m 遇到节点 n 的平均相遇频率和该真实转发概率的乘积决定, 所以:

$$\begin{aligned} E(T_{k,k+1}) &= \frac{t_{\min}}{u_m * p_m^{\varphi}} = \frac{t_{\min}}{\left(\sum_{n \in C_m} u_{m,n} - 1\right) * p_m^{\varphi}} \\ &= \frac{t_{\min}}{\left(\sum_{n \in C_m} u_{m,n} - 1\right) * \sum_{n \in C_m} p_{m,n}^{\varphi} * \frac{u_{m,n}}{u_m}} \\ &= \frac{t_{\min}}{\sum_{n \in C_m} p_{m,n}^{\varphi} * u_{m,n}} \end{aligned}$$

由于单副本概率路由可以记录已经转发过该消息的节点, 所以在状态 k 时, 表明整个网络中已经有 k 个节点携带过该消息, C_m 中最多有 $N-k$ 个节点. 为了方便计算, 本文在此处引入 $p_{m,n}^{\varphi}$ 的均值 p_m^{φ} , 节点之间的接触率均值 μ_{α} 来计算, 则有:

定理 4 (简化的单副本状态转移时间) 对于单副本概率路由, 在转发均值为 p^{φ} , 节点之间的接触率均值为 μ_{α} 的网络, 马尔可夫链从状态 k 转移到状态 $k+1$ 的时间为:

$$E(T_{k,k+1}) \approx \frac{t_{\min}}{p^{\varphi} (N-k) (\mu_{\alpha} - 1)} \quad (14)$$

证明 当用 p^{φ} 代替 $p_{m,n}^{\varphi}$ 时, 对于给定路由策略 φ , p^{φ} 为常数, 则

$$\begin{aligned} \sum_{n \in C_m} p_{m,n}^{\varphi} * (\alpha_{m,n} - 1) &= p^{\varphi} \sum_{n \in C_m} (\alpha_{m,n} - 1) \\ &= p^{\varphi} \sum_{n \in C_m} \alpha_{m,n} - (N-k) \end{aligned}$$

又因 μ_{α} 为 $\alpha_{m,n}$ 的均值, 所以若节点数较大时, 可以简单的简化为

$$\sum_{n \in C_m} \alpha_{m,n} \approx (N-k) \mu_{\alpha}, \text{ 所以有公式 (14) 成立.}$$

4.2 多副本概率路由的状态转移

对于多副本概率路由, 基本的假设和单副本情况下相同, 但在单副本概率路由的情况下, 马尔可夫链的每个状态中, 全局网络里只有一个节点持有消息, 而在

多副本概率路由的情况下则不同,在不同的状态下网络中持有消息的节点数是不同的. 这里假设 C_k 为在第 k 个状态持有消息的 k 个节点的集合, B_k 为剩余的无消息的 $N-k$ 个节点的集合.

在多副本概率路由下,仍需求解的是 $T_{k,k+1}$ 和 $P_{k,k+1}$,而求 $T_{k,k+1}$ 的关键仍是要找到这时的最小剩余间隔相遇时间. 在多副本情况下,最小剩余间隔相遇时间是不同于单副本的情况的. 因为在多副本路由下,持有消息的 m 节点在下一个时间段首先遇到的节点可能是已经持有消息的节点,所以本文将每一个阶段的节点按是否持有消息分作两个集合, C_k 和 B_k ,在任意时刻,仍利用单副本概率路由中计算最小剩余间隔相遇时间的思想,只分析一个持有消息的节点和 B_k . 这样就变成全局共有 $N-k+1$ 个节点的单副本概率路由,则由前面的公式(10)可以推出,此时只考虑一个节点 m 的最小剩余间隔相遇时间 $\min\{R_{m,n}\}$ 也服从幂律分布,幂指数为:

$$\overline{\alpha}_m = \sum_{n \in B_k} \alpha_{m,n}^* - 1 = \sum_{n \in B_k} (\alpha_{m,n} - 1) - 1 \quad (15)$$

此时节点 m 的平均相遇频率为:

$$\overline{u}_m = \sum_{n \in B_k} u_{m,n} - 1 \quad (16)$$

节点 m 从状态 k 将消息转发到其下一个真正的中继的真实概率是:

$$p_m^\varphi = \sum_{n \in B_k} p_{m,n}^\varphi * \frac{u_{m,n}}{u_m} \quad (17)$$

定理 5 (多副本状态转移时间)多副本概率路由,消息从状态 k 转移到状态 $k+1$ 的所需的时间为:

$$E(T_{k,k+1}) = \frac{t_{\min}}{\sum_{m \in C_k} \sum_{n \in C_{k+1}} p_{m,n}^\varphi * (\alpha_{m,n} - 1)} \quad (18)$$

证明 节点 m 将消息转发出去的频率为 $p_m^\varphi * u_m$,而多副本过程中同时有 k 个这样的过程, k 个节点的集合为 B_m ,所以多副本路由中,状态发生转移的频率为:

$\sum_{m \in C_k} \overline{p}_m^\varphi * \overline{u}_m$, 所以:

$$\begin{aligned} E(T_{k,k+1}) &= \frac{t_{\min}}{\sum_{m \in C_k} u_m * \overline{p}_m^\varphi} \\ &= \frac{t_{\min}}{\sum_{m \in C_k} (\sum_{n \in B_k} u_{m,n} - 1) * \overline{p}_m^\varphi} \\ &= \frac{t_{\min}}{\sum_{m \in C_k} (\sum_{n \in B_k} u_{m,n} - 1) * \sum_{n \in B_k} p_{m,n}^\varphi * \frac{u_{m,n}}{u_m}} \\ &= \frac{t_{\min}}{\sum_{m \in C_k} \sum_{n \in B_k} p_{m,n}^\varphi * u_{m,n}} \end{aligned}$$

为了计算方便,本文仍用转发频率和接触率的均值 \overline{p}^φ 和 $\overline{\mu}_\alpha$ 来代替每一个 $p_{m,n}^\varphi$ 和 $\alpha_{m,n}$,因此,

定理 6 (简化的多副本状态转移时间)对于多副本概率路由,在转发均值为 \overline{p}^φ 、节点之间的接触率均值为 $\overline{\mu}_\alpha$ 的网络,消息从状态 k 转移到状态 $k+1$ 的时间为:

$$E(T_{k,k+1}) \approx \frac{t_{\min}}{\overline{p}^\varphi k (N-k) (\overline{\mu}_\alpha - 1)} \quad (19)$$

证明 同定理(4)的证明过程,因为在公式(18)中 B_m 中共有 m 个节点,所以 B_m 中共有 m 个节点,于是,

$$\begin{aligned} &\sum_{m \in B_m} \sum_{n \in C_{k+1}} p_{m,n}^\varphi * (\alpha_{m,n} - 1) \\ &\approx \sum_{m \in B_m} \overline{p}^\varphi (N-k) (\overline{\mu}_\alpha - 1) \\ &\approx k * \overline{p}^\varphi (N-k) (\overline{\mu}_\alpha - 1) \end{aligned}$$

5 仿真证明

实验选取了文献[17]中满足幂律分布的两个真实数据集对该模型进行了仿真. 一个是 Cambridge^[21] 数据集,另一个是 Sassy^[21] 数据集.

该实验是从两个方面进行. 首先,本节选取了 ONE 中的典型的概率路由 ProphetRouter 在两个真实数据集上测试了多副本路由协议下传输成功的消息的平均传输延时. 对于单副本路由的情况下,本文改动了原先的 ProphetRouter,使之符合本文提到的单副本概率路由,上述在 ONE 中所得的结果作为真实情况下一个消息在相应策略下的平均传输延时. 本文应用最大似然估计法分别对上述数据集在幂律分布的假设下估计其任意一对节点之间的幂指数 α ,并使用 MATLAB 进行编程,模拟了本文所提出的 TTL 预测模型,此时需要知道给定数据集中节点的度分布的幂指数 λ 和该网络的聚类系数 C . 然而,对于给定数据集,寻找其连接图时,目前仍然没有一个固定的标准判断两节点之间是否存在一条固定的链路,这就导致计算节点的度没有固定的标准.

在文献[18]中,作者提出只有当两个节点之间的间隔连接时间服从幂律分布,且幂指数 $\alpha \geq 2$ 时认为两节点之间传递消息时可以直接交付,无需借助中继节点. 所以,本文认为只有两节点之间的间隔连接时间幂指数 $\alpha \geq 2$,则这两个节点之间存在一定固定的链路,这两个节点的度值分别加 1.

在文献[21]中,作者通过实验证明,现实中大部分网络的聚类系数 C 接近于其节点度分布所成的幂指数 α 的 2 倍,所以,本文取 $C \approx 2\alpha$.

由于在 Sassy 数据集中有三个节点在参加数据收集的过程中离开了,Sassy 数据集中实际收集到的数据只有 25 个节点上传的数据. 因此,在用 ONE 仿真器做实验时,节点数也设置为 25 个. 表 1 是在 Cambridge 数

据集和 Saasy 数据集上做实验时的一些重要参数的设置.

表 1 ONE 中实验参数设置

	Cambridge	Saasy
节点数	35	25
传输速率	25K	25K
缓存大小	20M	20M
消息大小	20k	20k
仿真时间	7200s	7200s
传输范围	20m	20m

为了验证本文提出的模型的合理性,以及找到对应成功率要求下网络中需要设置的最低消息生存期,本文从如下几个方面进行了实验:

5.1 成功率对消息的平均传输延时(TTL)的影响

本部分使用 MATLAB 和 ONE 在 Cambridge 和 Saasy 两个数据集上验证了本文所提模型的准确性.如图 1,是利用该模型计算出的不同的成功率要求下所对应的平均消息传输延时.两图分别是单副本和多副本情况下,两数据集上的预测结果.

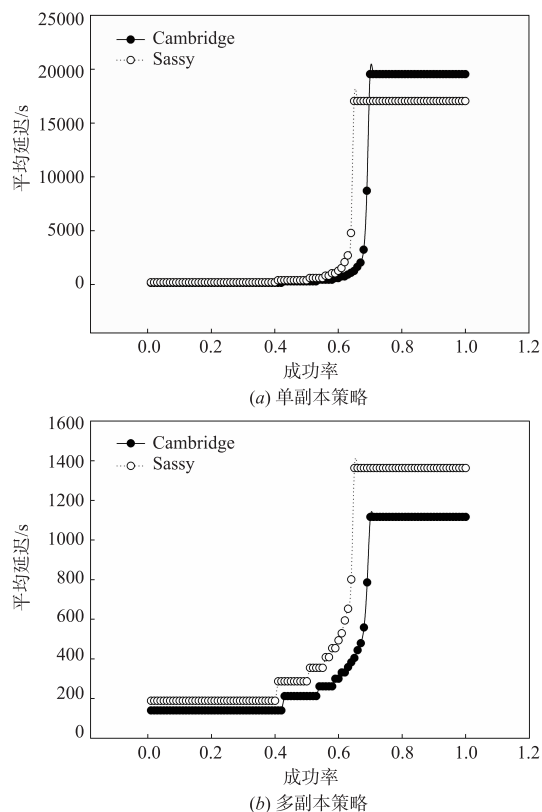


图1 预测所得的平均延迟(TTL)随传输成功率的变化

由图 1 可知,当需要保证的传输成功率较小时,预测出的 TTL 也相对较小,并且变化幅度很小,但此时传输延迟也并非局限于相对固定的值,而是需要根据对

传输成功率的具体要求来设定该模型预测得出的 TTL 值,从而取得传输延迟与传输成功率的一个均衡;当成功率达到一定程度时(Cambridge 数据集对应 0.7, Saasy 数据集对应 0.6)所需要设置的 TTL 急剧增加,引起的开销也会急剧增加,而成功率提高的幅度很小.所以,实际应用中,要求将成功率尽量控制在该临界点之下.

5.2 预测所得的消息平均传输延时(TTL)的准确性

图 2 和图 3 是本文提出的具有成功率保证的 TTL 预测模型预测出的结果与在 ONE 实际环境中得到的结果进行的对比,图 2 是单副本路由协议的对比结果,图 3 是多副本路由协议的对比结果.由于在 ONE 中得到的消息的转发成功率是实验得出的结果,具有不可控性.所以 ONE 中的数据点与 MATLAB 中得到的数据点有一定差异;另外,该模型将预测所得的平均传输延时作为将来的 TTL,所以在图 2 和图 3 统一将纵坐标称作平均传输延时.

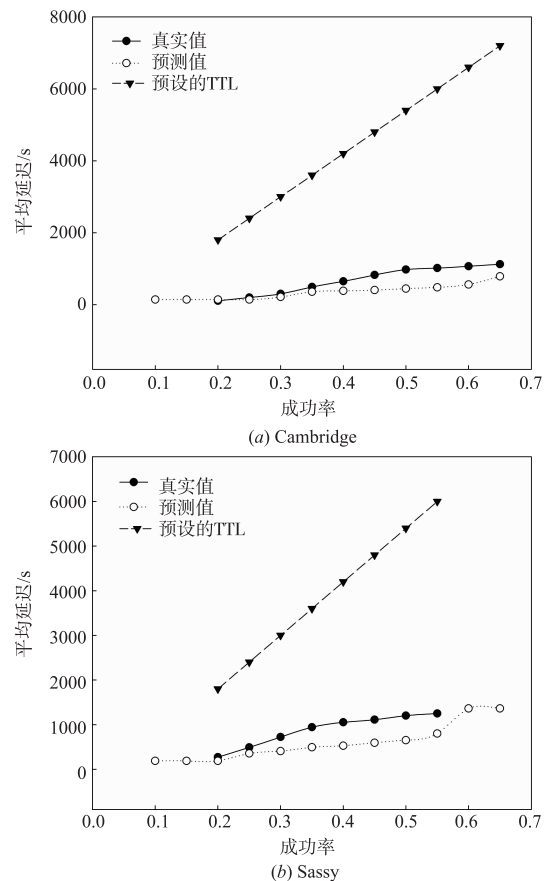


图2 单副本下预测所得TTL和ONE中所得结果的对比

由图 2 和图 3 可见,在 ONE 中测出的实际结果总是比预测结果大一些,这是由于 ONE 中得出的结果包含消息的传输时间,以及消息在缓冲中的调度时间,而在 MATLAB 中预测的是理想情况下的消息传输延时,

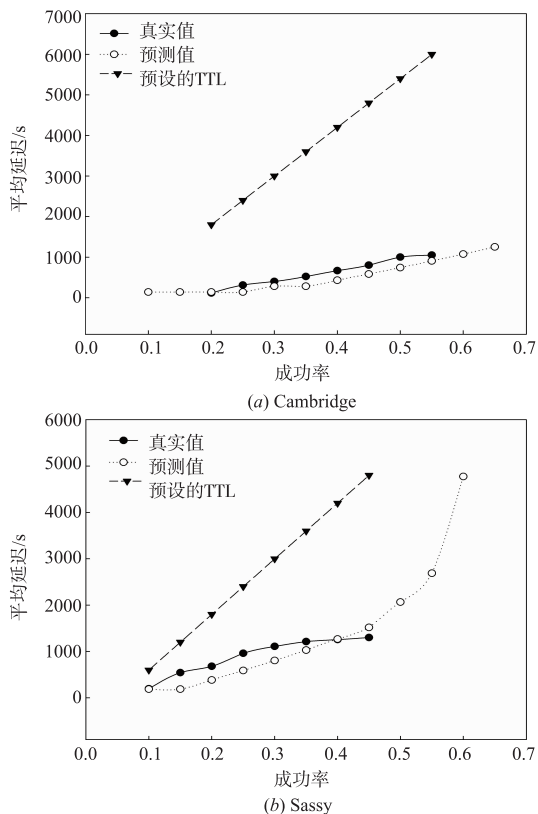


图3 多副本下预测所得TTL和ONE中所得结果的对比

所以总会略小于真实的延迟. 由实验可知,使用预测出的延迟作为 TTL, 不仅可以及时删除无法发送到目的节点的消息,还可以及时删除已经到达目的节点的消息的副本,从而减少转发次数,减少消息的存储时间,最终为网络节约资源.

5.3 消息被成功转发时成功率对需要经历的最大跳数的影响

图 4 是预测的消息成功率对网络中消息被成功转发时需要经历的最大跳数的影响. 从图 1 和图 4 可以看到,消息传输成功率对 TTL 的要求和对最大跳数的要求曲线基本相同;并且,对于一个给定的路由协议,给定的数据集上,当要求消息的传输成功率达到一定的程度时,就能保证整个网络是全连通的,否则,在给定条件下是无法通过其它方式达到该要求.

图 5 是本文统计的两个数据集中每个节点的平均 α 值,横坐标代表数据集中节点的编号,纵坐标表示节点的 α 值,该 α 值越大表示节点与其它节点的接触频率越大,节点越活跃.

该实验是在幂律分布的假设下采用最大似然估计法对两数据集进行估计,得出的每对节点之间的接触率 $\alpha_{m,n}$,然后对所有具有 m 的节点对 $\alpha_{m,n}$ 取算术平均数得到的结果如图 5 所示. 可以看出, Cambridge 数据集 35 个节点中, α 值在 1.5 到 2.0 之间的节点数为 21, α 值

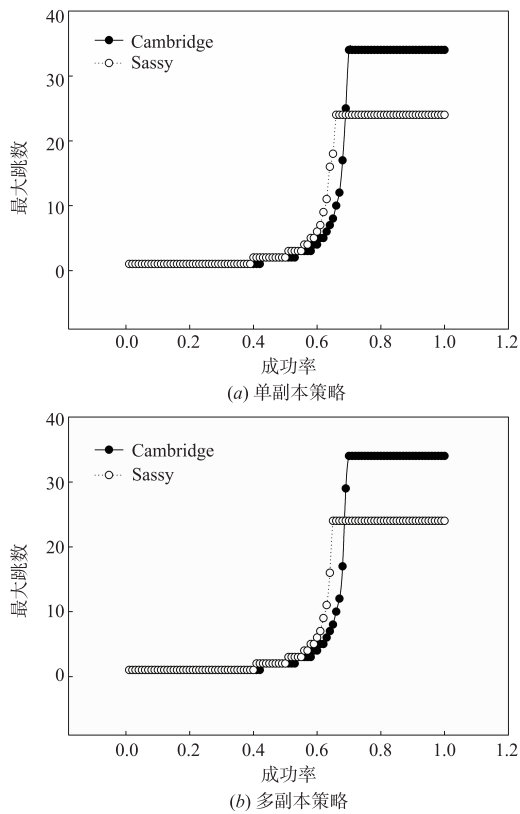


图4 预测所得的最大跳数随成功率的变化

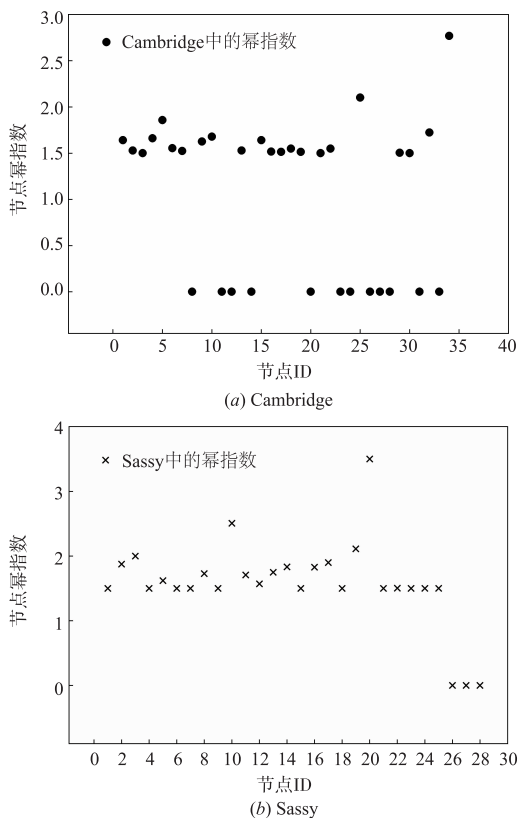


图5 两个数据集中不同节点的 α 值

大于 2 的节点数为 2, α 值为 0 的节点数为 12; 在 Sassy 数据集中, 共有 28 个节点, 其中有 3 个节点的 α 为 0, 3 个节点的 α 大于 2.

表 2 是对两数据集的不同特点的一个统计, 主要是参与人数、孤立节点数、特别活跃节点数以及全网的 α 和全网节点 α 的方差.

表 2 两个数据集的详细对比

数据集名称	Cambridge	Sassy
节点数	35	28
孤立节点数	12	3
高活跃节点数	2	33
全网的平均 α	1.042	1.569
α 的方差	0.6988	0.4795

结合图 5 和表 2 可以看出, 本文提出的模型在分级明显的社会结构上实验效果更好, 主要原因有两个: (1) 本文采用的两个数据集 Cambridge^[23] 和 Sassy^[23] 是一种通用的数据集, 它们在孤立点数量、节点之间接触频率、 α 的方差、节点活跃度分布等方面有如下特征: Cambridge 数据集中孤立节点较多, 整个数据集的平均 α 较小, 节点之间的接触频率较小, α 的方差较大, 即节点的活跃度分布不均匀, 使得网络中有明显的分层现象, 而 Sassy 数据集中孤立节点数少, α 的方差较小, 网络的分层现象要比 Cambridge 数据集弱. (2) 正如文献 [22] 中提到的, 节点间的均匀接触率在实际情况下通常是难以实现的, 尤其是在像机会网络这种具有社会属性的网络中. 因此, 本文中提出的 TTL 预测模型对实际应用也有一定的参考价值. 同时, 该模型所获得的预测结果在 Cambridge 数据集上的准确性比在 Sassy 数据集上高, 所以说本文的模型在分级明显的社会结构上使用效果会更好.

6 结果与展望

本文针对机会网络中具有成功率保证的概率路由转发模型, 在保证一定成功率的条件下, 可以找到满足该要求的最小 TTL, 改变了 TTL 只能依赖经验设置的现状. 另外, 由于机会网络的反馈机制很难实现, 使得消息在到达目的节点后不能把网络中剩余的副本及时删除, 给有限的存储空间带来较大的压力, 在不影响消息投递时, 设置最小的 TTL, 可保证所有消息及其副本被目的节点接收或被及时删除, 释放缓存空间, 从而提升网络的性能.

本文所讨论的模型还具有普适性, 该模型将单副本概率路由 TTL 预测模型和多副本概率路由 TTL 预测模型进行了结合, 对于不同的社团结构, 只要该社团结构中人类间的间隔接触时间和节点的度分布服从幂律

分布, 具有无标度网络的特点, 该模型都适用.

参考文献

- [1] Karamshuk D, Boldrini C, Conti M, et al. Human mobility models for opportunistic networks [J]. Communications Magazine, IEEE, 2011, 49(12): 157 - 165.
- [2] 徐佳, 孙力娟, 王汝传, 等. 机会网络中基于种子喷雾的自适应路由协议 [J]. 电子学报, 2010, 38(10): 2315 - 2321.
XU Jia, SUN Li-juan, WANG Ru-chuan, et al. Adaptive seed spray routing for opportunistic networks [J]. Acta Electronica Sinica, 2010, 38(10): 2315 - 2321. (in Chinese)
- [3] 吴克军, 于全, 田艳霞. 一种位置信息辅助的 Ad Hoc 网络按需距离矢量路由协议 [J]. 电子学报, 2010, 38(4): 983 - 988.
WU Ke-jun, YU Quan, TIAN Yan-xia. A novel location-aided Ad hoc on-demand vector routing protocol [J]. Acta Electronica Sinica, 2010, 38(4): 983 - 988. (in Chinese)
- [4] Vahdat A, Becker D. Epidemic routing for partially connected ad hoc networks [R]. CS-200006, Duke University, 2000.
- [5] Spyropoulos T, Psounis K, Raghavendra C S. Spray and wait: an efficient routing scheme for intermittently connected mobile networks [A]. Proceedings of the 2005 ACM SIGCOMM Workshop on Delay-tolerant networking [C]. ACM, 2005. 252 - 259.
- [6] Spyropoulos T, Psounis K, Raghavendra C S. Spray and focus: Efficient mobility-assisted routing for heterogeneous and correlated mobility [A]. Pervasive Computing and Communications Workshops, 2007. PerCom Workshops '07. [C]. IEEE, 2007. 79 - 85.
- [7] Hanbali A A, Nain P, Altman E. Performance of ad hoc networks with two-hop relay routing and limited packet lifetime [A]. Proceedings of the 1st International Conference on Performance Evaluation Methodologies and Tools [C]. ACM, 2006: 49.
- [8] Burns B, Brock O, Levine B N. MV routing and capacity building in disruption tolerant networks [J]. Ad Hoc Networks, 2005, 1(4): 398 - 408.
- [9] Lindgren A, Doria A, Schelén O. Probabilistic routing in intermittently connected networks [J]. ACM SIGMOBILE Mobile Computing and Communications Review, 2003, 7(3): 19 - 20.
- [10] Hui P, Crowcroft J, Yoneki E. Bubble rap: Social-based forwarding in delay-tolerant networks [J]. Mobile Computing, IEEE Transactions on, 2011, 10(11): 1576 - 1589.
- [11] Daly E M, Haahr M. Social network analysis for information flow in disconnected delay-tolerant MANETs [J]. Mobile Computing, IEEE Transactions on, 2009, 8(5):

- 606 – 621.
- [12] Kim S, Gerla M. Socio-geo: Social network routing protocol in delay tolerant networks [A]. Computing, Networking and Communications (ICNC) [C]. IEEE, 2016. 1 – 6.
- [13] 马学彬, 白婧, 郑田玉. 机会网络中基于有权社团结构图的路由协议研究 [J]. 电子学报, 2016, (10): 2449 – 2458.
MA Xue-bin, BAI Jing, ZHENG Tian-yu. A routing algorithm based on weighted community detection for Opportunistic networks [J]. Acta Electronica Sinica, 2016, 44 (10): 2449 – 2458. (in Chinese)
- [14] Chaintreau A, Hui P, Crowcroft J, et al. Impact of human mobility on opportunistic forwarding algorithms [J]. Mobile Computing, IEEE Transactions on, 2007, 6 (6): 606 – 620.
- [15] Boldrini C, Conti M, Passarella A. Less is more: long paths do not help the convergence of social-oblivious forwarding in Opportunistic Networks [A]. Proceedings of the third ACM International Workshop on Mobile Opportunistic Networks [C]. ACM, 2012. 13 – 20.
- [16] Boldrini C, Conti M, Passarella A. Performance modelling of opportunistic forwarding with imprecise knowledge [A]. Modeling and Optimization in Mobile, Ad Hoc and Wireless Networks (WiOpt) [C]. IEEE, 2012. 216 – 223.
- [17] Boldrini C, Conti M, Passarella A. Performance modelling of opportunistic forwarding under heterogenous mobility [J]. Computer Communications, 2014, 48: 56 – 70.
- [18] Boldrini C, Conti M, Passarella A. The stability region of the delay in Pareto opportunistic networks [J]. IEEE Transactions on Mobile Computing, 2015, 14 (1): 180 – 193.
- [19] Sermpezis P, Spyropoulos T. Delay analysis of epidemic schemes in sparse and dense heterogeneous contact environments [R]. Eurecom, Biot, France, 2012. 3868.
- [20] Picu A, Spyropoulos T, Hossmann T. An analysis of the information spreading delay in heterogeneous mobility dtns [A]. World of Wireless, Mobile and Multimedia Networks (WoWMoM) [C]. IEEE, 2012. 1 – 10.
- [21] Ying L, Hong-Duo C, Xiu-Ming S, et al. An estimation formula for the average path length of scale-free networks [J]. Chinese Physics B, 2008, 17 (7): 2327.
- [22] Sermpezis P, Spyropoulos T. Delay analysis of epidemic schemes in sparse and dense heterogeneous contact networks [J]. IEEE Transactions on Mobile Computing, 2017, 16 (9): 2464 – 2477.
- [23] HuggleProject [OL]: <https://crawdad.cs.dartmouth.edu/>.
- [24] 李昊, 山秀明, 任勇. 具有幂率度分布的因特网平均最短路径长度估计 [J]. 物理学报, 2004, 53 (11): 3695 – 3700.
LI Ying, Shan Xiu-Ming, Ren Yong. Average path length of Internet with power law degree distribution [J]. Acta Physica Sinica. 2004, 53 (11): 3695 – 3700. (in Chinese)

作者简介



马学彬 男, 1981 年 4 月出生, 内蒙古大学副教授, 研究方向为无线网络、移动 IP、延迟容忍网络、信息安全。
E-mail: csmxuebin@imu.edu.cn



王应彪 男, 1989 年 12 月出生, 内蒙古大学硕士研究生, 研究方向为计算机网络。