

基于衰落 Bloom Filter 的 P2P 网络弱状态路由算法*

朱桂明^{1,2+}, 郭得科³, 金士尧¹

¹(国防科学技术大学 计算机学院 并行与分布处理国防科技重点实验室, 湖南 长沙 410073)

²(江南计算技术研究所, 江苏 无锡 214083)

³(国防科学技术大学 信息系统与管理学院 信息系统工程国防科技重点实验室, 湖南 长沙 410073)

P2P Weak State Routing Scheme Based on Decaying Bloom Filters

ZHU Gui-Ming^{1,2+}, GUO De-Ke³, JIN Shi-Yao¹

¹(Key Laboratory of Science and Technology for National Defense of Parallel and Distributed Processing, College of Computer, National University of Defense Technology, Changsha 410073, China)

²(Jiangnan Institute of Computing Technology, Wuxi 214083, China)

³(Key Laboratory for Information System Engineering, College of Information System and Management, National University of Defense Technology, Changsha 410073, China)

+ Corresponding author: E-mail: guiming.zhu@yahoo.com.cn

Zhu GM, Guo DK, Jin SY. P2P weak state routing scheme based on decaying Bloom Filters. Journal of Software, 2011, 22(11): 2810-2819. <http://www.jos.org.cn/1000-9825/3863.htm>

Abstract: The current weak state routing schemes cannot facilitate in-network queries effectively. When a query is given for an item at an arbitrary node, disturbance in unrelated routing entries are likely stronger than the useful information in the right routing entries. Consequently, the majority of queries are moved towards wrong nodes. To solve this problem, this paper presents DWalker, an efficient weak state routing scheme-based design based on decaying bloom filters. DWalker uses a bloom filter to represent the summary information of resources at each node and then propagated a bloom filter within a propagation range. The amount of information in each bloom filter decreases exponentially with the distance from the source. DWalker makes the max propagation range less than the expected distance between any two nodes; hence, effectively tackling the problem of multi-path propagation of a decaying bloom filter. DWalker replaces a single bloom filter with multiple bloom filters as a routing entry and holds more routing information. DWalker can ensure that any query can be forwarded in the right direction with high a probability that is based on the proportion of bits set to 1 in a bloom filter and the least number of bits set to 1 in an entry bloom filter for a query. The analysis and simulation of results show that DWalker can make many queries that can be forwarded in the right direction to achieve a high query hit rate with low cost and low delay.

Key words: peer-to-peer computing; directed random network; weak state routing; decaying Bloom Filter; disturbance

* 基金项目: 国家自然科学基金(60903206, 61170284, 60970094, 61003304); 中国博士后科学基金(201104439); 湖南省自然科学基金(09ZZ4034); 国防科学技术大学预先研究项目(JC10-05-01)

收稿时间: 2009-07-07; 修改时间: 2010-03-01; 定稿时间: 2010-04-27

摘要: 在 P2P 网络中,基于衰落 Bloom Filter 的弱状态路由算法试图将每条查询消息沿着成员资格信息量最强的方向传递,并最终以较低的传输代价和传输时延确保较高的查准率.衰落 Bloom Filter 在传递过程中存在严重的多径叠加和噪音问题,这直接导致查询消息会以很高的概率沿着错误的方向传播,甚至会退化为泛洪路由算法.为了解决这一挑战性难题,提出了 DWalker 这种基于衰落 Bloom Filter 的高效弱状态路由算法.DWalker 基于有向随机网络,采用指数衰落 Bloom Filter 来发布和传播每个节点共享资源的信息,且其最大传播距离小于网络中任意两点之间距离的期望值,从而有效抑制了衰落 Bloom Filter 在传播过程中的多径叠加问题.DWalker 采用多个 Bloom Filter 而不是单个 Bloom Filter 来表达一项路由条目,在单个 Bloom Filter 的错误发生概率达到设计上限时,可按需动态增加新的 Bloom Filter,以将更多资源对象信息纳入到当前路由条目中.DWalker 仅根据当前节点的各项路由条目中值为 1 的比特位所占的最大比例,以及查询消息在正确转发方向对应的路由条目中对应比特位中值为 1 的个数的临界值,就能使进入目标对象传播范围内的查询消息以较高的概率辨认出正确的路由方向.理论分析和实验结果表明,DWalker 能够以较低的查询消息代价、较小的路由条目存储开销以及较短的查询时延,使绝大多数查询消息沿正确方向转发,从而获得较高的查准率.

关键词: 对等计算;有向随机网络;弱状态路由;衰落 Bloom Filter;噪音

中图法分类号: TP393 **文献标识码:** A

非结构化 P2P 网络由于其资源定位方法的简单性和易用性,在 Internet 上得到了大量应用.然而,由于拓扑结构的不确定性、资源对象放置的随意性等特点,在没有辅助机制的条件下,非结构化 P2P 资源定位方法很难同时获得较低的查询延迟、少量的定位成本和较高的查询命中率^[1].近年来,出现了大量非结构化 P2P 网络的提示性资源定位技术的研究,其主要思想是,由各资源共享节点发布共享信息,并在网络中传播和维护这些共享信息,用于指导资源定位消息的转发^[2,3].对于不采用压缩方法的提示性路由方法,维护和更新共享资源对象元信息需要消耗大量的存储空间和网络带宽,资源对象元信息的传播和维护通常只能在一个相对较小的范围内进行,因而对路由指导作用十分有限.针对该问题,提示性路由的研究重点是基于 Bloom Filter 技术对资源对象元信息进行压缩存储和传播,并基于 Bloom Filter 进行查询路由^[1-3].Bloom Filter 被广泛用于非结构化对等网络^[3-6]、无线传感器网络^[7]、ad hoc 网络^[8-11]以及 mesh 网络^[12]中实现概率路由.其共同的思想^[1]是,每个节点将其所共享的全部资源对象用一个 Bloom Filter 表示,并将其传播到给定逻辑跳步(例如 h 跳)内的节点;节点把从相同邻居链路传递进来的 Bloom Filter 集合作为该邻居链路的路由条目.如果一个节点需要将查询消息路由到距离其 h 跳步内的潜在目的节点,则通过检查每个邻居链路的路由条目就能挑选出正确的路由方向,并转发给相应的邻居节点.这些方法的缺点是,虽然单个 Bloom Filter 的存储开销不大,但由于每个节点需要存储从每个链路传递进来的所有 Bloom Filter,且同一节点的 Bloom Filter 往往通过不同链路多次到达同一节点,从而对存储资源提出了极高的要求.Kumar 等人设计了一种依指数衰减的 Bloom Filter,并在此基础上提出了非结构化对等网络中的弱状态路由机制 SQR^[13].SQR 中,每个节点产生本地 Bloom Filter,以距离为衰减指数向周围节点传递;节点把从相同邻居链路传递进来的 Bloom Filter 进行并运算,并将产生的联合 Bloom Filter 作为该邻居链路的路由条目.该弱状态路由机制要求在查询路由过程中,查询消息沿着成员资格信息量^{**}最强的方向进行.

基于 Bloom Filter 的弱状态路由机制的核心思想是:每个节点为其所属对象构造从自身向周边邻居依规则衰减的信息梯度;如果节点转发以其他节点上对象为目标的查询消息,并且处于该消息的潜在目标节点为查询对象构造的信息梯度范围内,则要求该节点能够为该查询消息选择正确的方向进行转发.文献[1]指出,SQR 将从相同链路传递进来的 Bloom Filter 的联合作为该链路的路由条目,虽然大大缩小了对存储资源的要求,但是由于噪音的问题,且噪音很可能接近甚至比正确路由条目中的成员资格信息还要强,导致大部分查询沿着背离目的节点的方向传递,使得弱状态路由机制十分低效.图 1 给出了弱状态路由示例,在查询消息进入目标节点 A 的衰减范围之前,即路径 $K \rightarrow H \rightarrow G \rightarrow E$,采取随机路由;之后,根据非噪音与噪音强度对比关系,或沿着错误方向进

^{**} 给定 Bloom Filter bf ,元素 x 的成员信息量定义为 bf 中位置为 $h_1(x), h_2(x), \dots, h_t(x)$ 的比特位值为 1 的个数.

行(如图 1(a)所示),或沿着正确方向进行(如图 1(b)所示).文献[1]深入分析了基于 Bloom Filter 的弱状态路由机制 SQR 存在的问题,并给出了该机制正常运行所需满足的约束条件.然而,该模型建立在局部网络拓扑度数为常数的树型结构基础上,每个节点共享相同个数的资源,约束条件太苛刻;且相关参数的设置和获取也并非易事.

针对弱状态路由机制 SQR 低效率的问题,本文提出了一种基于衰落 Bloom Filter 的高效 P2P 网络弱状态路由算法 DWalker. DWalker 仅根据本地知识,以严格的数学推导为基础,使得进入目标对象传播范围内的查询消息既能以较高的概率辨认出正确的路由方向,又能以较小的概率产生冗余的、沿着错误方向的路由.

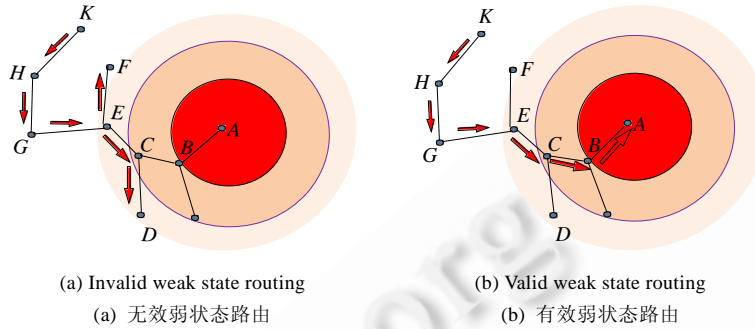


Fig.1 An example for weak state routing

图 1 弱状态路由示例

1 Bloom Filters

Bloom Filter 是一种结构简单而空间复杂度很小的随机数据结构^[14].其主要用一个长度为 m 的位向量 V 对大小为 n 的集合 X 进行精确压缩表示,并代替被表示的集合回答类似“元素 x 是否在集合 A 中”的判定. Bloom Filter 使用 k 个相互独立且随机分布的哈希函数 h_1, h_2, \dots, h_k 将每个元素散列到一个长度为 m 的数组中.图 2 给出了将元素 x, y 添加到 Bloom Filter 中的结果示例.对于被 Bloom Filter 表示的集合 X 中元素 e ,数组中位置为 $h_1(e), h_2(e), \dots, h_k(e)$ 的比特位均被设置为 1.对于给定的元素 x ,若位置为 $h_1(x), h_2(x), \dots, h_k(x)$ 的比特位均为 1,则认为元素 $x \in X$,否则, $x \notin X$. Bloom Filter 对该问题的回答不存在假阴性错误(false negative),但是存在着假阳性错误(false positive),即 Bloom Filter 技术不可能把属于集合 X 的元素误判为不属于 X ,但可能把不属于集合 X 的元素误判为 X 的元素.出现假阳性错误的概率为^[1]

$$P_{err} = (1 - (1 - 1/m)^{kn})^k \approx (1 - e^{-kn/m})^k.$$

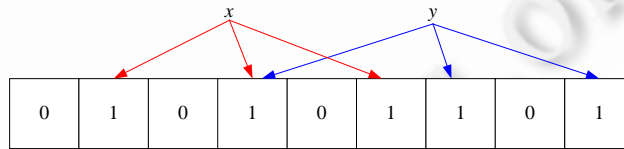


Fig.2 An example for adding x and y into a Bloom Filter

图 2 将元素 x, y 添加到 Bloom Filter 中的结果示例

2 DWalker

DWalker 的基本思想是,基于有向随机网络,采用依指数衰落 Bloom Filter 来发布和传播每个节点共享资源的信息,且其最大传播距离小于网络中任意两点之间距离的期望值,从而有效抑制了衰落 Bloom Filter 在传播过程中的多径叠加问题;采用多个 Bloom Filter 而不是单个 Bloom Filter 来表达一项路由条目,在单个 Bloom Filter 的错误发生概率达到设计上限时,可按需动态增加新的 Bloom Filter,以将更多资源对象信息纳入到当前路由条目中,具有较好的扩展性;仅根据本地知识,即 Bloom Filter 的比特位个数 m 、哈希函数的个数 k 、衰减指数 d 、

最大传输距离 h 及 Bloom Filter 中值为 1 的比特位个数,以严格的数学分析为基础,计算出各个 Bloom Filter 值为 1 的比特位个数的最大比例,以及下一跳转发节点的路由条目至少被查询匹配的值为 1 的比特位个数的临界值,使得进入目标对象传播范围内的查询消息既能以较高的概率辨认出正确的路由方向,又能以较小的概率产生冗余的沿着错误方向的路由。

2.1 基于有向随机网络及指数衰落 Bloom Filter 的资源对象发布与传递

基于衰落 Bloom Filter 的概率路由能否沿着正确的方向传递,取决于节点路由条目中关于目标对象成员资格信息量的增大趋势是否如实反映了各节点与目标对象所在节点的距离远近关系.双向随机网络中衰落 Bloom Filter 的传递,存在着回流、多径叠加和噪音问题,往往使得查询消息沿着错误的方向传递.对于有向随机网络而言,衰落 Bloom Filter 沿着单个方向衰减传递,资源对象成员资格信息量在最短路径上以最小的衰减到达各节点.通过设定衰落 Bloom Filter 最大传递距离小于随机网络任意两个节点之间距离的期望值,节点从非最短路径上收到关于目标对象的成员资格信息量的期望值接近为 0,从而能够有效抑制衰落 Bloom Filter 回流和多径叠加问题。

定理 1. 假设有向随机网络 G ,其网络规模为 n ,节点出度为 c ,令 G 中任意两个节点之间距离的期望值为 AVG ,则资源对象的 Bloom Filter 传递的最大距离小于 AVG 时,节点从非最短路径上接收到的资源对象的成员资格信息量的期望值为 0。

证明:由于 G 中任意两个节点之间距离的期望值为 AVG ,因此,任意节点 u 从非最短路径上接收到节点 v 共享资源对象的 Bloom Filter,经 AVG 次的衰减传递;由于共享资源对象的 Bloom Filter 传递的最大距离小于 AVG ,因此,节点 u 从非最短路径上接收到节点 v 共享资源对象成员资格信息量的期望值为 0.定理 1 得证. \square

基于有向随机网络的节点接收和转发 Bloom Filter 的算法如算法 1 所示.节点将从上行(UP)邻居节点 src 接收到的 Bloom Filter $bloomfilter$ 加入到上行邻居节点 src 原路由条目数组 $srcBfArray$ 中.原路由条目数组 $srcBfArray$ 最初仅有 1 个 Bloom Filter,当该 Bloom Filter 与待加入的 $bloomfilter$ 进行并运算后值为 1 的比特位个数达到设定的上限值时,额外创建一个新的 Bloom Filter,以容纳待加入的 $bloomfilter$.将 $bloomfilter$ 加入到 $srcBfArray$ 的具体过程为:依次遍历数组 $srcBfArray$ 中的各个 Bloom Filter,若存在一个 Bloom Filter bf 与 $bloomfilter$ 进行并运算后值为 1 的比特位个数未达到上限值,则保存 bf 与 $bloomfilter$ 的并运算结果,否则将 $bloomfilter$ 插入到数组的末尾.将衰减后的 $bloomfilter$ 传递给所有下行(DOWN)邻居节点,由其作进一步处理。

算法 1. $rcv_snd(Node\ src, BloomFilter\ bloomfilter)$.

```

1  get src node BloomFilter arraylist srcBfArray from UP neighbors list;
2  //将 bloomfilter 加入到对应 src 节点的路由条目中
3  if ( $\exists bf \in srcBfArray! full(bf \cup bloomfilter)$ ){
4       $bf \leftarrow bf \cup bloomfilter$ ;
5  } else {
6       $srcBfArray.add(bloomfilter)$ ;
7  }
8   $bloomfilter \leftarrow bloomfilter/d$ ;           //衰减 bloomfilter
9  for all neighbor  $i$  from DOWN neighbors list //更新所有下行邻居节点的路由条目
10      $rcv\_snd(i, bloomfilter)$ ;
```

- $bf \cup bloomfilter$: 定义为 bf 与 $bloomfilter$ 的按位“或”;
- $full(bf \cup bloomfilter)$: 定义为若 bf 与 $bloomfilter$ 进行按位“或”运算后,值为 1 的比特位个数达到设定的上限时其值为 $true$, 否则为 $false$.

2.2 参数计算

定义 1. 给定集合 X 及其 Bloom Filter bf , 令 $\theta(x, bf)$ 表示在对应 bf 中任意元素 $x \in X$ 的成员资格信息量, 即

Bloom Filter 地址 $Bfaddress(x)$ 中设置为 1 的比特位的个数, ($Bfaddress(x)$ 由 $h_i(x)$ 构成, 其中, $1 \leq i \leq k$); 令 $\theta(X, bf)$ 表示集合 X 在对应 bf 中的信息量, 即 bf 中被设置为 1 的比特位的个数.

定义 2. 令 bf_i 表示一个 bf 在第 i 轮衰减后的结果, 其中 $i \geq 1$. 在 bf_i 中保留有 $\theta(bf_i)$ 个为 1 的比特位. 对 bf 依指数衰减, 则有

$$\theta(bf_i) = \begin{cases} \theta(bf), & i = 1 \\ \lfloor \theta(bf_{i-1}) / d \rfloor, & i > 1 \end{cases}$$

定义 3. 对于任意资源对象 x 及路由条目的 Bloom Filter $entry$, $\theta(x, entry)$ 中不是由当前节点接收到的任何 $bf_i(x)$ ($1 \leq i \leq h$) 设置的值为 1 的比特位个数, 定义为资源对象 x 相对于路由条目 $entry$ 的噪音.

传统的基于依指数衰减 Bloom Filter 的弱状态路由, 由于噪音的原因, 且噪音很可能接近甚至比正确路由条目中的成员资格信息还要强, 导致大部分查询沿着背离目的节点的方向传递. 为此, DWalker 依据本地知识, 即 Bloom Filter 的比特位个数 m 、哈希函数的个数 k 、衰减指数 d 及最大传输距离 h , 考虑设置单个 Bloom Filter 中值为 1 的比特位的占总比特位个数的最大比例 (以下称为 Bloom Filter 的饱和率, 记为 α), 以限定噪音的影响. 并结合具体路由条目的 Bloom Filter 中值为 1 的比特位个数, 计算出当前路由条目的噪音概率分布及非噪音与噪音叠加后的分布概率, 并进一步计算该路由条目的 Bloom Filter 与查询对象 x 的 Bloom Filter 进行并运算后值为 1 的比特位个数的临界值, 使得大于等于该临界值时, 查询以较高的概率认为该路径为正确路由; 小于该临界值时, 查询以较大的概率认为该路径为错误路由. 从而, DWalker 使得进入目标对象传播范围内的查询消息既能以较高的概率辨认出正确路由方向, 又能以较小的概率产生冗余的沿着错误方向的路由.

定理 2^[1]. 如果 x 被 bf 表示, 当 $i=1$ 时, $\theta(x, bf_i)$ 等于 k ; 否则, $\theta(x, bf_i)$ 是一个离散随机变量, 记为 U_i , 其概率分布函数为

$$P(U_i = a) = \frac{\binom{k}{k-a} \binom{\theta(bf) - k}{\theta(bf) - \theta(bf_i) - k + a}}{\binom{\theta(bf)}{\theta(bf) - \theta(bf_i)}} \quad (1)$$

定理 3. 给定 bf , 设 bf 中值为 0 和 1 的比特位的比例分别为 p_0 和 p_1 , 并分别用作 bf 中任一比特位取 0 和 1 的概率, 于是, 从 bf 中任选 k 个比特位中值为 1 的比特位个数是一个离散随机变量, 记为 V , 其概率分布函数为

$$P(V = b) = \binom{k}{b} p_1^b p_0^{k-b} \quad (2)$$

证明: 由于 bf 中任一比特位取 0 和 1 的概率分别为 p_0 和 p_1 , 则对于任取的 k 个比特位中, 任取 b 个比特位的值为 1, 从而其余比特位的值为 0 的概率为 $p_1^b p_0^{k-b}$; 又由于从 k 个比特位中取 b 个比特位的取法共有 $\binom{k}{b}$ 种, 从而从 bf 中任选 k 个比特位中值为 1 的比特位个数为 b 的概率为 $\binom{k}{b} p_1^b p_0^{k-b}$. 定理 3 得证. \square

定理 4. 设 x 被 bf 表示, 路由条目的 Bloom Filter $entry$ 由 bf_i 与其他 Bloom Filter 的并运算得到. $entry$ 中值为 0 和 1 的比特位的比例分别为 p_0 和 p_1 , 并分别用作 $entry$ 中任一比特位取 0 和 1 的概率. 于是, $\theta(x, entry)$ 是一个离散随机变量, 记为 Z_i , 其概率分布函数为

$$P(Z_i = c) = \sum_{u=0}^c P(U_i = u) \binom{k-u}{c-u} p_1^{c-u} p_0^{k-c} \quad (3)$$

证明: $entry$ 是由 bf_i 与其他 Bloom Filter 的并运算得到, 从而对于 x 而言, $\theta(x, entry)$ 是由噪音与非噪音叠加后的结果. 因此, Z_i 取 c 的概率可由非噪音 U_i 取 u 的概率与噪音取 $c-u$ 的概率时各种情况的乘积求和得到, 从而

$$P(Z_i = c) = \sum_{u=0}^c P(U_i = u) \binom{k-u}{c-u} p_1^{c-u} p_0^{k-c}. \text{定理 4 得证.} \quad \square$$

给定 Bloom Filter 的比特位个数 m 、哈希函数的个数 k 、衰减指数 d 、最大传输距离 h , 若 x 被 Bloom Filter

bf 表示, bf_h 为 bf 经 h 轮衰减后的结果. 令 p_c 为路由条目的 Bloom Filter entry 接收到 bf_h 进行正确路由的概率的下界, p_b 为 entry 未接收到 $bf_i (1 \leq i \leq h)$ 进行错误路由的概率上界. 饱和率 α 的设置应满足:

$$\exists t, P(Z_h \geq t) \geq p_c \wedge P(V \geq t) \leq 1 - p_b.$$

即存在临界值 t , 当查询消息的 Bloom Filter 与 entry 进行“与”运算后值为 1 的比特位数大于等于 t 时, 进行正确路由的概率大于等于 p_c , 进行错误路由的概率小于等于 p_b . 值得说明的是, 对于接收到 bf_h 的路由条目, 若 α 的设置使得 $\exists t, P(Z_h \geq t) \geq p_c \wedge P(V \geq t) \leq 1 - p_b$ 成立, 则对于接收到 $bf_i (1 \leq i \leq h)$ 的路由条目, 该 α 的设置同样使得 $\exists t, P(Z_i \geq t) \geq p_c \wedge P(V \geq t) \leq 1 - p_b$ 成立.

当 $\theta(bf) \gg k$, 对于不同的 $\theta(bf)$, U_i 具有极为相似的概率分布. 因此, 当 $\theta(bf) \gg k$ 时, 假设 U_i 的概率分布服从 $\theta(bf)$ 取 $\theta(bf) = m \cdot \alpha$ 时的概率分布, 其中, $m \cdot \alpha \gg k$. 从而对于给定的 p_c 和 p_b , 各个节点仅依据本地知识可计算出路由条目的 Bloom Filter 的饱和率; 依据各个路由条目的 Bloom Filter 中值为 1 的比特位的实际比例及目标资源对象的 Bloom Filter 到达该路由条目的衰落次数, 可计算出各个路由条目的 Bloom Filter 的临界值 t . 而目标资源对象的 Bloom Filter 到达该路由条目的衰落次数可以采用保守的方法, 通过最大传输距离 h 及查询消息已经遍历的被认为接收到目标资源对象的 Bloom Filter 衰落结果的节点个数进行保守估计, 并以估计值代替目标资源对象的 Bloom Filter 到达该路由条目的实际衰落次数计算临界值 t .

2.3 路由算法

传统的基于衰落 Bloom Filter 的弱状态路由算法进行路由时, 选择路由条目的 Bloom Filter 与查询的 Bloom Filter 进行“与”运算后值为 1 的比特位数最大(即成员资格信息最强)的路由条目转发查询消息. 由于噪音的影响, 且噪音很可能接近甚至比正确路由条目中的成员资格信息还要强, 导致大部分查询会沿背离目的节点的方向传递. 与传统弱状态路由不同的是, 当查询消息进入目标对象的传播范围内, 对于每一个路由条目的 Bloom Filter, DWalker 计算出该路由条目的 Bloom Filter 与查询的 Bloom Filter 进行“与”运算后值为 1 的比特位个数的临界值 t , 使得该路由条目的 Bloom Filter 与查询的 Bloom Filter 进行“与”运算后值为 1 的比特位数大于等于 t 时, 以较大的概率 p_c 认为该路由条目为正确路由条目, 以较小的概率 p_b 认为该路由条目为错误路由条目. 查询消息进入目标对象的传播范围内的 DWalker 路由算法如算法 2 所示. 算法 2 中, 若当前节点具有满足查询 $query$ 的资源, 则直接向查询请求节点返回相应的资源; 否则, 若尚未达到查询消息传递的最大距离, 则继续遍历所有上行邻居节点. 对于每个上行邻居节点, 针对其路由条目数组中的每个 Bloom Filter bf 计算临界值 t , 若该 bf 与查询 $query$ 进行“与”运算后值为 1 的比特位数大于或等于 t , 则将查询进一步转发给该邻居节点.

算法 2. DWalker(Node curNode, Node srcNode, BloomFilter query, int distance). {

```

1   if contain_resource(query) {
2       send resource to srcNode;
3       return;
4   }
5   if distance  $\geq$  h return;           //h 为最大传播距离
6   for all neighbor from UP Neighbors
7       for all BloomFilter bf of neighbor's routing entry BloomFilter ArrayList
8           t=critical(bf,h-distance,m,k, $\alpha$ ,pc); //计算临界值
9           if  $\theta(query \wedge bf) \geq t$  {
10              DWalker(neighbor,srcNode,query,distance+1);
11              break;
12          }
13   return;
14 }
```

2.4 算法分析

定理 4 给出了资源对象 x 在由 $bf_i(x)$ 与噪音叠加形成的路由条目 $entry$ 中的成员资格信息量 $\theta(x, entry)$ 对应随机变量 Z_i 的概率分布. 由定理 4, 对于给定的 p_c 即可计算出 $\theta(x, entry)$ 的临界值 t , 使得 $P(Z_h \geq t) \geq p_c$, 从而保证了路由算法沿正确方向转发查询消息的概率下界 p_c . 对于未接收到 $bf_i(x) (1 \leq i \leq h)$ 的路由条目 $entry$, 定理 3 给出了资源对象 x 在该路由条目中成员资格信息量 $\theta(x, entry)$ 对应随机变量 V 的概率分布.

对于给定的 p_b , 要求 $P(Z_h \geq t) \geq p_c \wedge P(V \geq t) \leq 1 - p_b$, 从而使得路由算法错误选择该路由条目的概率上界为 p_b . 而对于路由条目中的任意 Bloom Filter, 其饱和率 α 的设置应满足 $\exists t, P(Z_h \geq t) \geq p_c \wedge P(V \geq t) \leq 1 - p_b$.

对于尚未进入目标对象的传播范围内的查询消息, DWalker 以随机路由的方式在节点之间进行转发; 对于进入目标对象传播范围内的查询消息, 各节点将查询消息沿正确方向转发的概率的下界为 p_c . 对于未接收到目标对象的成员资格信息的节点路由条目, 节点错误选择该路由条目进行转发的概率上界为 $1 - (1 - p_b)^s$, 其中, s 为该路由条目的 Bloom Filter 个数. 若节点 u 从节点 v 接收到目标对象的第 i 次衰落结果, 其中, $1 \leq i \leq h$, 则 DWalker 将查询消息从节点 u 沿着正确方向以路由延迟 i 成功路由到节点 v 的概率的下界为

$$P(u \rightarrow v) = p_c^i \quad (4)$$

若各节点的路由条目的平均个数为 c 、各个路由条目的 Bloom Filter 的个数均为 s , 令 $p_e = 1 - (1 - p_b)^s$, 则 DWalker 算法访问节点个数的期望值的上界为

$$\begin{aligned} Cost(u \rightarrow v) &= ((c-1)p_e + (c-1)p_e c p_e + \dots + (c-1)p_e (c p_e)^{i-1}) + \\ &\quad p_c ((c-1)p_e + (c-1)p_e c p_e + \dots + (c-1)p_e (c p_e)^{i-2}) + \dots + \\ &\quad p_c^{i-1} ((c-1)p_e) + (1 + p_c + p_c^2 + \dots + p_c^i) \\ &= (c-1)p_e \sum_{j=1}^i p_c^{i-j} \frac{1 - (c p_e)^j}{1 - c p_e} + \frac{1 - p_c^{i+1}}{1 - p_c} \end{aligned} \quad (5)$$

公式(4)、公式(5)给出的分析结果将在模拟实验中进一步验证.

3 模拟实验

我们在 Peersim^[15]上实现了 DWalker, 并将其与 SQR、传统方法(节点通过邻居链路接收给定逻辑跳步 h 内所有节点共享资源的 Bloom Filter, 并将这些 Bloom Filter 的并集作为该邻居链路的路由条目)进行了比较.

实验考察各算法在给定参数 Bloom Filter 的比特位个数 m 、哈希函数的个数 k 、衰减指数 d 、最大传输距离 h 及 p_c 和 p_b , 饱和率 α 的取值与临界值 t 的取值范围的关系, 从接收目标对象的第 h 次衰落 Bloom Filter 的节点发起的查询的命中率、查询代价(即查询消息访问节点的个数)及各个路由条目中 Bloom Filter 的个数. 由于各算法的查询延迟均为 h , 本实验不作考虑. 实验参数见表 1.

Table 1 Experiment parameters

表 1 实验参数

Parameter	Value	Parameter	Value
Number of nodes n	2 000	Decaying coefficient d	1.2
Initial out-degree c	4	Max routing hops h	4 (AVG=5.53)
Initial topology	random	p_c	0.95, 0.98
Number of bits of a Bloom Filter m	6 000	p_b	0.05
Number of hash functions k	16	α	0.25

实验过程: 各节点发布相同个数资源对象的 Bloom Filter; 任选 100 个节点作为查询的目标节点, 对于任意目标节点, 从距离该目标节点距离为 h 的所有节点发起针对目标节点任意共享资源的查询, 并统计查询命中率的均值、查询代价的均值以及各节点路由条目中 Bloom Filter 个数的均值.

图 3 给出了临界值 t 的取值范围与饱和率 α 的关系. 从图 3(a)、图 3(b)中可以看出: Bloom Filter 的衰落次数 i 越小, t 取值的上限越大; 饱和率 α 的值越大, 临界值 t 的取值范围越小; 若 α 的设置使得 Bloom Filter 衰落 i 次时临界值 t 存在, 则该 α 的设置同样使得 Bloom Filter 衰落 j 次时存在临界值 t , 其中, $i \geq j$. 在以下的模拟实验中, 取

$\alpha=0.25$ 对于 $(p_c=0.95, p_b=0.05)$ 和 $(p_c=0.98, p_b=0.05)$, t 均存在, 且实验中 t 尽可能取较大的值.

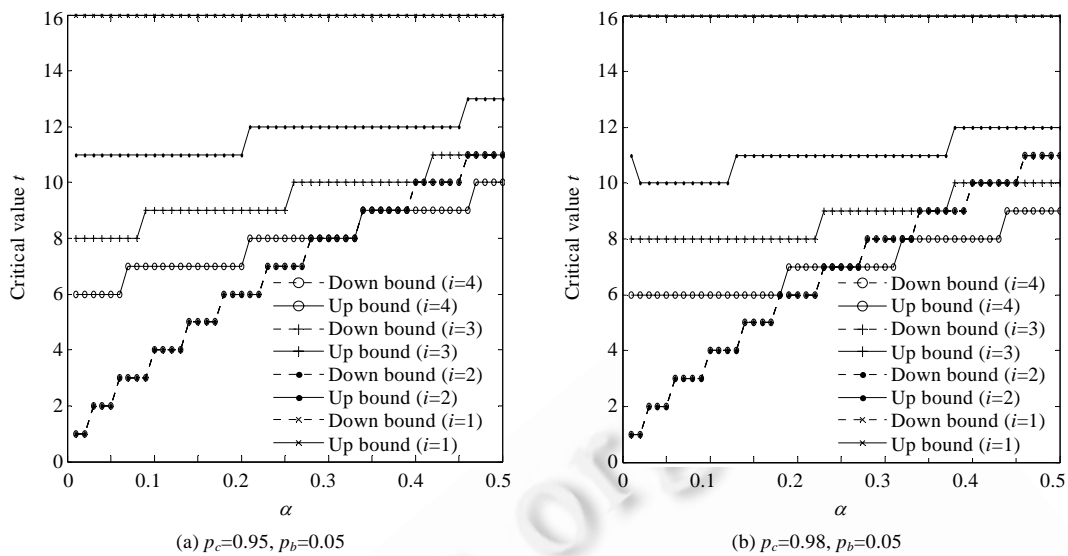


Fig.3 The bound of critical value t is a function of α

图3 临界值 t 的取值范围与饱和率 α 的关系

图4给出了查询命中率、查询代价与各节点共享资源个数的关系.在图4(a)中,当 $p_c=0.95, p_c=0.98$ 时, DWalker 对应的查询命中率,即 $DWalker(p_c=0.95, p_b=0.05)$, $DWalker(p_c=0.98, p_b=0.05)$,基本上均大于由公式(4)给出的理论查询命中率下界的期望值.传统方法(tradition)通过将同一邻居链路传递进来的所有 Bloom Filter 完整保存下来,因而命中率达到 1. SQR 采用单播策略时,即路由过程中从成员资格信息量最大的路由条目中随机选择一个节点转发,查询命中率随着各节点共享资源个数的增加而急剧下降,且最低值接近 0; SQR 采用多播策略时,即路由过程中将查询消息转发给所有成员资格信息量最大的路由条目对应的邻居节点,查询命中率随着各节点共享资源个数的增加先急剧降低、后缓慢增加,且最低达到了 0.4 左右.在图4(b)中, DWalker 的查询代价即 $DWalker(p_c=0.95, p_b=0.05)$, $DWalker(p_c=0.98, p_b=0.05)$, 其值均小于由公式(5)给出的理论查询代价上界的期望值,且与 SQR 采取单播策略和传统方法的查询代价均十分接近,远远小于 SQR 采取多播策略对应的查询代价. SQR 采取多播策略的查询代价随着各节点共享资源个数的增加而急剧增大,最大值接近 150. 结合图4(a)和图4(b)可以看出: DWalker 以较低的查询代价实现了较高的查询命中率; 采取多播策略的 SQR, 其查询命中率降低的过程中, 查询代价缓慢增加, 即对应着其他链路路由条目的噪音强度以一定的概率大于正确路由条目中信息强度, 而查询命中率增加的过程中, 查询代价急剧上升, 即查询命中率的上升是由于各节点共享资源个数达到一定数量后, 其他链路上路由条目的噪音强度以较大的概率与正确路由条目的强度等同, 产生了大量近似于“泛洪”形式的消息转发. 也正是这个原因, 使得采取单播策略的 SQR 的查询命中率逐渐降低, 最终近似于随机形式的路由.

图5给出了路由条目中 Bloom Filter 个数与各节点共享资源个数的关系. DWalker 对应路由条目的空间开销随着资源个数的增加而缓慢增加; SQR 由于仅采用 1 个 Bloom Filter 保存从同一邻居链路传递进来的所有 Bloom Filter 的联合, 因而其路由条目仅为 1 个 Bloom Filter 的空间开销; 传统方法由于将从同一邻居链路传递进来的所有 Bloom Filter 完整保存下来, 因而其空间开销十分巨大, 远远大于 SQR 和 DWalker 对应的空间开销.

从实验结果可以看出, 相对于传统方法和 SQR 算法, DWalker 以稍大于 SQR、远小于传统方法的路由条目中 Bloom Filter 的个数, 以较低的查询代价, 使绝大多数查询消息沿着正确方向行走, 达到了较高的查询命中率.

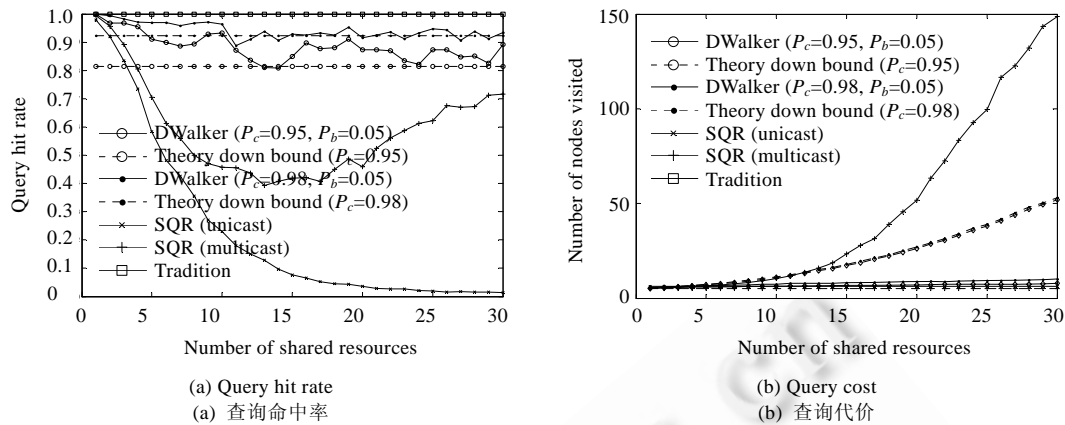


Fig.4 Query hit rate and cost are functions of shared resources of each node

图 4 查询命中率和查询代价与各节点共享资源个数关系

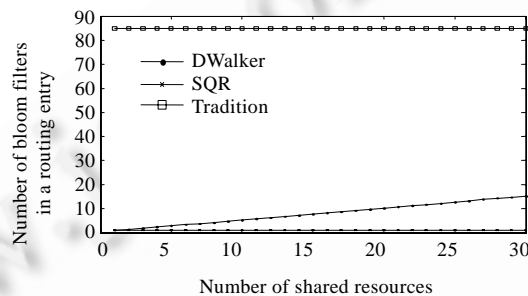


Fig.5 The number of bloom filters of each routing entry is a function of the number of resources shared by each node

图 5 路由条目中 Bloom Filter 的个数与资源个数的关系

4 结论

本文提出了一种基于衰落 Bloom Filter 的 P2P 网络弱状态路由算法 DWalker. DWalker 基于有向随机网络, 采用指数衰落 Bloom Filter 来发布和传播每个节点共享资源的信息, 且其最大传播距离小于网络中任意两点之间距离的期望值, 从而有效抑制了衰落 Bloom Filter 在传播过程中的多径叠加问题; 采用多个 Bloom Filter 而不是单个 Bloom Filter 来表达一项路由条目, 在单个 Bloom Filter 的错误发生概率达到设计上限时, 可按需动态增加新的 Bloom Filter, 以将更多资源对象信息纳入到当前路由条目中, 具有较好的扩展性; DWalker 仅通过本地知识, 即 Bloom Filter 的比特位个数 m 、哈希函数的个数 k 、衰减指数 d 、最大传输距离 h 及 Bloom Filter 中值为 1 的比特位个数, 以严格的数学分析, 计算出各个 Bloom Filter 值为 1 的比特位个数的最大比例 (即饱和率 α), 以及下一跳转发节点的路由条目至少被查询匹配的值为 1 的比特位个数的临界值 t , 使得进入目标对象传播范围内的查询消息既能以较高的概率辨认出正确路由方向, 又能以较小的概率产生冗余的沿错误方向的路由. 理论分析和实验结果表明, DWalker 能够以较低的查询代价、较小路由条目中 Bloom Filter 的个数, 使绝大多数查询消息沿着正确方向行走, 达到了较高的查询命中率.

References:

- [1] Guo DK. Research on peer-to-peer networks based on Kautz digraph and Bloom filters [Ph.D. Thesis]. Changsha: National University of Defense Technology, 2008 (in Chinese with English abstract).

- [2] Zhang YM, Lu XC, Zheng QB, Li DS. PST: An efficient search algorithm for large-scale P2P systems. *Journal of Software*, 2008, 19(6):1473–1480 (in Chinese with English abstract). <http://www.jos.org.cn/1000-9825/19/1473.htm> [doi: 10.3724/SP.J.1001.2008.01473]
- [3] Hodes TD, Czerwinski SE, Zhao BY, Joseph AD, Katz RH. An architecture for secure wide-area service discovery. *Wireless Networks*, 2002,8(2/3):213–230. [doi: 10.1023/A:1013772027164]
- [4] Reynolds P, Vahdat A. Efficient peer-to-peer keyword searching. In: Schmidt D, Endler M, eds. *Proc. of the ACM/IFID/USENIX 2003 Int'l Conf. on Middleware*. New York: Springer-Verlag, 2003. 21–40.
- [5] Rhea SC, Kubiatowicz J. Probabilistic location and routing. In: Lee D, Orda A, eds. *Proc. of the IEEE INFOCOM 2002*. Washington: IEEE Computer Society, 2002. 1248–1257. [doi: 10.1109/INFCOM.2002.1019375]
- [6] Bauer D, Hurley P, Pletka R, Waldvogel M. Bringing efficient advanced queries to distributed hash tables. In: Jha S, Hassanein H, eds. *Proc. of the 29th Annual IEEE Int'l Conf. on Local Computer Networks*. Washington: IEEE Computer Society, 2004. 6–14. [doi: 10.1109/LCN.2004.32]
- [7] Hebden P, Pearce AR. Data-Centric routing using bloom filters in wireless sensor networks. In: Alex H, Begg R, eds. *Proc. of the 4th Int'l Conf. on Intelligent Sensing and Information Processing*. Washington: IEEE Computer Society, 2006. 72–77. [doi: 10.1109/ICISIP.2006.4286065]
- [8] Hsiao PH. Geographical region summary service for geographical routing. In: Corson MS, Das SR, eds. *Proc. of the 2nd ACM Int'l Symp. on Mobile Ad Hoc Networking & Computing*. New York: ACM, 2001. 263–266. [doi: 10.1145/501449.501455]
- [9] Yuen WH, Schulzrinne H. Improving search efficient using Bloom filters in partially connected ad hoc networks: A node-centric analysis. *Computer Communications*, 2007,30(16):3000–3011. [doi: 10.1016/j.comcom.2007.05.055]
- [10] Acer UG, Kalyanaraman S, Abouzeid AA. Weak state routing for large scale dynamic networks. In: Hou J, Ramanathan R, eds. *Proc. of the 13th ACM Int'l Conf. on Mobile Computing and Networking*. New York: ACM, 2007. 290–301. [doi: 10.1145/1287853.1287888]
- [11] Gilbert R, Johnson K, Wu SM, Zhao BY, Zheng HT. Location independent compact routing for wireless networks. In: Petrioli C, Ramjee R, eds. *Proc. of the 1st Int'l Workshop on Decentralized Resource Sharing in Mobile Computing and Networking*. New York: ACM, 2006. 57–59. [doi: 10.1145/1161252.1161267]
- [12] Chan CF. Mole: Multi-hop object location in wireless mesh networks [Ph.D. Thesis]. Hong Kong: Hong Kong University of Science and Technology, 2008.
- [13] Kumar A, Xu J, Zegura EW. Efficient and scalable query routing for unstructured peer-to-peer networks. In: Knightly E, Makki K, eds. *Proc. of the IEEE INFOCOM 2005*. Washington: IEEE Computer Society, 2005. 1162–1173. [doi: 10.1109/INFCOM.2005.1498343]
- [14] Bloom BH. Space/Time tradeoffs in hash coding with allowable errors. *Communications of the ACM*, 1970,13(7):422–426. [doi: 10.1145/362686.362692]
- [15] Jelasity M, Montresor A, Jesi GP. The PeerSim simulator. 2009. <http://peersim.sf.net>

附中文参考文献:

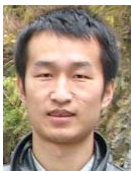
- [1] 郭得科. 基于 Kautz 图和 Bloom 滤波的对等网络研究[博士学位论文]. 长沙: 国防科学技术大学, 2008.
- [2] 张一鸣, 卢锡城, 郑倩冰, 李东升. 一种概率搜索小组: 面向大规模 P2P 系统的快速搜索算法. *软件学报*, 2008, 19(6):1473–1480. <http://www.jos.org.cn/1000-9825/19/1473.htm> [doi: 10.3724/SP.J.1001.2008.01473]



朱桂明(1981—),男,江苏兴化人,博士,主要研究领域为对等计算,数据中心网络互联.



金士尧(1937—),男,教授,博士生导师,CCF高级会员,主要研究领域为分布式计算,复杂系统仿真,实时系统,虚拟现实.



郭得科(1980—),男,博士,副研究员,CCF会员,主要研究领域为无线多跳网络,对等计算,Bloom Filters,数据中心网络互联.