(19)中华人民共和国国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 110750897 A
(43)申请公布日 2020.02.04

(21)申请号 201910986341.3

(22)申请日 2019.10.17

(71)申请人 中国民航大学
地址 300300 天津市东丽区津北公路2898号

(72)发明人 樊智勇 腾达 刘哲旭

(74)专利代理机构 天津中环专利商标代理有限公司 12105
代理人 王凤英

(51)Int.Cl.
G06F 30/20(2020.01)
H04L 29/08(2006.01)

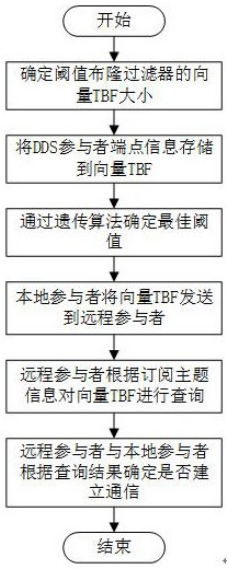
权利要求书3页 说明书7页 附图10页

(54)发明名称

一种基于阈值布隆过滤器的DDS自动发现方法

(57)摘要

本发明公开了一种基于阈值布隆过滤器的DDS自动发现方法。其包括按顺序进行的下列步骤:阈值布隆过滤器的设计;阈值布隆过滤器与DDS自动发现机制的结合;最佳阈值的确定过程。本发明优点:①该方法通过阈值布隆过滤器存储DDS发现阶段的端点描述信息,能够减小内存消耗和网络数据传输,提高DDS在分布式仿真的实时性。②该方法提出阈值优化方式,实现较小的误报率和较大的精确度,特别是当DDS端点数较多、TBF较小时,效果更为显著。③该方法为DDS自动发现过程的改进提供了一种新的思路。



CN 110750897 A

1.一种基于阈值布隆过滤器的DDS自动发现方法,其特征在于,包括按顺序进行下列步骤:1、阈值布隆过滤器的设计;2、阈值布隆过滤器与DDS自动发现机制的结合;3、最佳阈值的确定过程;

所述的阈值布隆过滤器的设计步骤是:阈值布隆过滤器使用一个m位的一位向量来保存信息,向量初始值为0;当储存一个参与者端点信息元素时,该端点信息元素通过k个不同的哈希函数映射到阈值布隆过滤器向量中,该向量记为TBF(1);每个哈希函数作用的结果分布范围是[1,m];根据k个映射结果,向量TBF(1)中对应的k个位置的值由原来的0变为1;当存储多个端点信息元素时,映射结果会叠加;因此,当阈值布隆过滤器存储n个不同的端点信息元素x_i时,通过每个端点信息元素x_i映射结果的和得到向量TBF(1),即

$$\mathbf{TBF}(\mathbf{1})=\sum_{i=1}^{n}x_i \quad \text{----------} \quad (1)$$

对于一个集合S(x_1～x_n),每个端点信息元素x_i通过k个不同的哈希函数映射到向量TBF(1),当查询一个端点信息元素x_i是否属于集合S时,通过设置不同的二值化阈值θ和判定阈值T来判断端点信息元素x_i是否属于集合S,其中二值化阈值θ满足0≤θ≤k,判定阈值T满足0≤T≤k;首先,如果向量TBF(1)中每个位置的值小于或等于θ,那么这个位置置0;这时,向量TBF(1)变成一个新的向量TBF(2);然后,通过判定阈值T判断,即当一个端点信息元素x_i的映射结果与向量TBF(2)之间的点积值大于或等于判定阈值T时,则判定这个端点信息元素x_i属于集合S。

2.根据权利要求1所述的一种基于阈值布隆过滤器的DDS自动发现方法,其特征在于,所述的阈值布隆过滤器与DDS自动发现机制的结合步骤是:

设数据从节点A发送到节点B,两节点A、B分别定义为本地参与者和远程参与者;在参与者发现阶段,将本地参与者端点即数据写入者、数据读取者的描述信息存储在向量TBF(2)中,并通过本地参与者数据包一起发给其他远程参与者,这些端点描述信息是每个本地参与者唯一的关键字,通常是主题名称;在端点发现阶段,当远程参与者的端点订阅的一个或多个主题时,远程参与者首先查询向量TBF(2)中是否存在其订阅的主题;如果存在,则远程参与者向本地参与者发送存在主题订阅信息;本地参与者发送与远程参与者相关的主题数据包和服务质量进行进一步匹配,如果匹配成功,本地参与者与远程参与者建立通信。

3.根据权利要求1所述的一种基于阈值布隆过滤器的DDS自动发现方法,其特征在于,所述的最佳阈值的确定过程步骤是:

假设一个端点信息元素x_i经过哈希函数映射的结果互不相同,即不同的端点信息元素映射到不同的位置,将端点信息元素x_i的映射看作单个实现的超几何分布,其中,分布的容量是w即w的值等于向量TBF(1)的位数m的值,置1的位数是r即r的值等于哈希函数的数量k的值,所以,在端点信息元素x_i的映射中,将向量TBF(1)特定位置置1的概率p_1是:

$$p_1=\frac{k}{m} \quad \text{----------} \quad (2)$$

将向量TBF(1)的第i个位置记为I的值看做是一个离散的随机变量,n个端点信息元素在向量TBF(1)中第i位置映射的结果服从二项分布B(n,p_1);向量TBF(1)中第i个位置被映射f次,则第i个位置值为v,第i个位置值v的值等于映射次数f的值;所以,向量TBF(1)中第i个位置的值I＝v的概率P(I＝v)为:

$$P(I=v)=\binom{n}{v}p_1^v(1-p_1)^{n-v} \quad \text{----------（3）}$$

然后得到向量TBF(1)中第i个位置值为v的位置数量的期望值l(v)：

$$l(v)=mP(I=v)=m\binom{n}{v}p_1^v(1-p_1)^{n-v} \quad \text{----------（4）}$$

通过设置二值化阈值θ，由向量TBF(1)即得到向量TBF(2)，根据公式(3)可知向量TBF(2)中一个位置值为0的概率P_0为：

$$P_0=\sum_{v=0}^{\theta}P(I=v)=\sum_{v=0}^{\theta}\binom{n}{v}p_1^v(1-p_1)^{n-v} \quad \text{----------（5）}$$

在向量TBF(2)中，一个位置值为1的概率P_1为：

$$P_1=1-P_0=1-\sum_{v=0}^{\theta}\binom{n}{v}p_1^v(1-p_1)^{n-v} \quad \text{----------（6）}$$

集合S中的端点信息元素x_i在向量TBF(2)中映射结果与向量TBF(2)的点积值为d_x，点积期望值$\bar{d}_x$由公式(4)得出；当θ＝0时，$\bar{d}_x=k$，即对于集合S中任意一个端点信息元素的点积值都是k；因此，当v≤θ时，所有位置的点积期望值为$\bar{d}_x$：

$$\bar{d}_x=k-\frac{m}{n}\sum_{v=0}^{\theta}vP(I=v) \quad \text{----------（7）}$$

同样，不属于集合S的端点信息元素y在TBF(2)中映射结果与向量TBF(2)的点积值为d_y，点积期望值$\bar{d}_y$通过向量TBF(2)中非零位置数计算得出：

$$\bar{d}_y=kP_1 \quad \text{----------（8）}$$

由此，即可计算点积值d_x和点积值d_y的概率特征，点积值d_x和点积值d_y都具有离散随机变量的特征，它们遵循二项分布：$d_x\sim B(k,p_x)$，$d_y\sim B(k,p_y)$，计算点积过程中对应于某一位置的乘积为1的概率，即可从公式(7)和公式(8)中得到概率P_x和概率P_y：

$$p_x=\frac{\bar{d}_x}{k}=1-\frac{m}{nk}\sum_{v=0}^{\theta}P(I=v) \quad \text{----------（9）}$$

$$p_y=\frac{\bar{d}_y}{k}=P_1 \quad \text{----------（10）}$$

通过引入两个参数来描述优化目标，即召回率TPR和误报率FPR，召回率TPR表示主题名称查询成功的概率，即正确的判断端点信息元素x_i属于集合S，召回率TPR的范围是0≤TPR≤1；相反，误报率FPR表示主题名称查询失败的概率，即错误的判断端点信息元素y属于集合S，误报率FPR的范围是0≤FPR≤1，考虑到判定阈值T，召回率TPR可由点积值d_x的概率质量函数求得，具体公式如下：

$$\mathrm{TPR}=\sum_{d=T}^{k}P(d_x=d)=\sum_{d=T}^{k}\binom{k}{d}p_x^d(1-p_x)^{k-d} \quad \text{----------（11）}$$

同样，误报率FPR可由点积值d_y的概率质量函数求得，具体公式如下：

$$\mathrm{FPR}=\sum_{d=T}^{k}P\left(d_y=d\right)=\sum_{d=T}^{k}\binom{k}{d}p_y^{d}\left(1-p_y\right)^{k-d} \quad \text{----------（12）}$$

最后将传输精度作为优化目标函数ACC：

$$ACC=\frac{\left(\mathrm{TPR}+\left(1-\mathrm{FPR}\right)\right)}{2}=\max \quad \text{----------（13）}$$

其中公式(13)的约束条件分别为公式(2)、公式(3)、公式(9)、公式(10)、公式(11)和公式(12)，公式(13)的最大值是一个非线性问题，通过遗传算法来求解。

一种基于阈值布隆过滤器的DDS自动发现方法

技术领域

[0001] 本发明涉及分布式仿真技术,特别是涉及一种基于阈值布隆过滤器的DDS(Data distribution service,数据分发服务)自动发现方法。

背景技术

[0002] 分布式仿真是将庞大的仿真计算任务分解为若干细小的任务,由多台计算机分担完成,广泛应用于军事、交通、电力系统、医疗等领域。DDS是对象管理组织制定了以数据为中心的发布/订阅通信模型规范,由于DDS的高效性和实时性,越来越多的分布式仿真系统采用DDS进行数据传输。现有的DDS标准的自动发现方法基于简单发现协议,在中小型分布式仿真系统中取得了良好的效果。但仿真系统增大时,大量的数据需要频繁、实时交换,现有的DDS标准的自动发现方法会产生高的内存消耗和网路数据传输,所以不再适用于规模较大的分布式仿真系统。

发明内容

[0003] 为了解决上述问题,本发明的目的在于提供一种在大型分布式仿真系统中,基于阈值布隆过滤器的能够实现低内存消耗、低网络传输量、低误报率的DDS自动发现方法。

[0004] 为了达到上述目的,本发明采取的技术方案是:一种基于阈值布隆过滤器的DDS自动发现方法,其特征在于,包括按顺序进行下列步骤:(1)、阈值布隆过滤器的设计;(2)、阈值布隆过滤器与DDS自动发现机制的结合;(3)、最佳阈值的确定过程。

[0005] 所述的阈值布隆过滤器的设计步骤是:阈值布隆过滤器使用一个m位的一位向量来保存信息,向量初始值为0;当储存一个端点信息元素时,该端点信息元素通过k个不同的哈希函数映射到阈值布隆过滤器向量中,该向量记为TBF(1);每个哈希函数作用的结果分布范围是[1,m];根据k个映射结果,向量TBF(1)中对应的k个位置的值由原来的0变为1;当存储多个端点信息元素时,映射结果会叠加;因此,当阈值布隆过滤器存储n个不同的端点信息元素x_i时,通过每个端点信息元素x_i映射结果的和得到向量TBF(1),即

[0006] (3) $$\mathbf{TBF(1)}=\sum_{i=1}^{n}x_i$$ ---------- (1)

[0007] 对于一个集合S(x_1~x_n),每个端点信息元素x_i通过k个不同的哈希函数映射到向量TBF(1),当查询一个端点信息元素x_i是否属于集合S时,通过设置不同的二值化阈值θ和判定阈值T来判断端点信息元素x_i是否属于集合S,其中二值化阈值θ满足0≤θ≤k,判定阈值T满足0≤T≤k;首先,如果向量TBF(1)中每个位置的值小于或等于θ,那么这个位置置0;这时,向量TBF(1)变成一个新的向量TBF(2);然后,通过判定阈值T判断,即当一个端点信息元素x_i的映射结果与向量TBF(2)之间的点积值d大于或等于判定阈值T时,则判定这个端点信息元素x_i属于集合S。

[0008] 本发明所述的阈值布隆过滤器与DDS自动发现机制的结合步骤是:

[0009] 设数据从节点A发送到节点B,两节点A、B分别定义为本地参与者和远程参与者;在

参与者发现阶段，将本地参与者端点即数据写入者、数据读取者的描述信息存储在向量TBF(2)中，并通过本地参与者数据包一起发给其他远程参与者，这些端点描述信息是每个本地参与者唯一的关键字，通常是主题名称；在端点发现阶段，当远程参与者的端点订阅的一个或多个主题时，远程参与者首先查询向量TBF(2)中是否存在其订阅的主题；如果存在，则远程参与者向本地参与者发送存在主题订阅信息；本地参与者发送与远程参与者相关的主题数据包和服务质量进行进一步匹配，如果匹配成功，本地参与者与远程参与者建立通信。

[0010] 本发明所述的最佳阈值的确定过程步骤是：

[0011] 假设一个端点信息元素x_i经过哈希函数映射的结果互不相同，即不同的端点信息元素映射到不同的位置，将端点信息元素x_i的映射看作单个实现的超几何分布，其中，分布的容量是w即w的值等于向量TBF(1)的位数m的值，置1的位数是r即r的值等于哈希函数的数量k的值，所以，在端点信息元素x_i的映射中，将向量TBF(1)特定位置置1的概率p_1是：

[0012] (4) $p_1=\frac{k}{m}$ ----------（2）

[0013] 将向量TBF(1)的第i个位置记为I的值看做是一个离散的随机变量，n个端点信息元素在向量TBF(1)中第i位置映射的结果服从二项分布B(n,p_1)；向量TBF(1)中第i个位置被映射f次，则第i个位置值为v，第i个位置值v的值等于映射次数f的值；所以，向量TBF(1)中第i个位置的值I＝v的概率P(I＝v)为：

[0014] (5) $P(I=v)=\binom{n}{v}p_1^v(1-p_1)^{n-v}$ ----------（3）

[0015] 然后得到向量TBF(1)中第i个位置值为v的位置数量的期望值l(v)：

[0016] (6) $l(v)=mP(I=v)=m\binom{n}{v}p_1^v(1-p_1)^{n-v}$ ----------（4）

[0017] 通过设置二值化阈值θ，由向量TBF(1)即得到向量TBF(2)，根据公式(3)可知向量TBF(2)中一个位置值为0的概率P_0为：

[0018] (7) $P_0=\sum_{v=0}^{\theta}P(I=v)=\sum_{v=0}^{\theta}\binom{n}{v}p_1^v(1-p_1)^{n-v}$ ----------（5）

[0019] 在向量TBF(2)中，一个位置值为1的概率P_1为：

[0020] (8) $P_1=1-P_0=1-\sum_{v=0}^{\theta}\binom{n}{v}p_1^v(1-p_1)^{n-v}$ ----------（6）

[0021] 集合S中的端点信息元素x_i在向量TBF(2)中映射结果与向量TBF(2)的点积值为d_x，点积期望值$\bar{d}_x$由公式(4)得出；当θ＝0时，$\bar{d}_x=k$，即对于集合S中任意一个端点信息元素的点积值都是k；因此，当v≤θ时，所有位置的点积期望值为$\bar{d}_x$：

[0022] (9) $\bar{d}_x=k-\frac{m}{n}\sum_{v=0}^{\theta}vP(I=v)$ ----------（7）

[0023] 同样，不属于集合S的端点信息元素y在TBF(2)中映射结果与向量TBF(2)的点积值为d_y，点积期望值$\bar{d}_y$通过向量TBF(2)中非零位置数计算得出：

[0024] (10) $\bar{d}_y=kP_1$ ----------（8）

[0025] 由此，即可计算点积值d_x和点积值d_y的概率特征，点积值d_x和点积值d_y都具有离散随机变量的特征，它们遵循二项分布：$d_x \sim B(k,p_x)$，$d_y \sim B(k,p_y)$，计算点积过程中对应于某一位置的乘积为1的概率，即可从公式(7)和公式(8)中得到概率P_x和概率P_y：

[0026] (11) $$p_x = \frac{\overline{d}_x}{k} = 1 - \frac{m}{nk}\sum_{v=0}^{\theta} P(I=v)$$ ---------- (9)

[0027] (12) $$p_y = \frac{\overline{d}_y}{k} = P_1$$ ---------- (10)

[0028] 通过引入两个参数来描述优化目标，即召回率TPR和误报率FPR，召回率TPR表示主题名称查询成功的概率，即正确的判断端点信息元素x_i属于集合S，召回率TPR的范围是$0 \leq TPR \leq 1$；相反，误报率FPR表示主题名称查询失败的概率，即错误的判断端点信息元素y属于集合S，误报率FPR的范围是$0 \leq FPR \leq 1$，考虑到判定阈值T，召回率TPR可由点积值d_x的概率质量函数求得，具体公式如下：

[0029] (13) $$TPR = \sum_{d=T}^{k} P(d_x = d) = \sum_{d=T}^{k} \binom{k}{d} p_x^d (1-p_x)^{k-d}$$ ---------- (11)

[0030] 同样，误报率FPR可由点积值d_y的概率质量函数求得，具体公式如下：

[0031] (14) $$FPR = \sum_{d=T}^{k} P(d_y = d) = \sum_{d=T}^{k} \binom{k}{d} p_y^d (1-p_y)^{k-d}$$ ---------- (12)。

[0032] 最后将传输精度作为优化目标函数ACC：

[0033] (15) $$ACC = \frac{(TPR + (1-FPR))}{2} = \max$$ ---------- (13)

[0034] 其中公式(13)的约束条件分别为公式(2)、公式(3)、公式(9)、公式(10)、公式(11)和公式(12)，公式(13)的最大值是一个非线性问题，通过遗传算法来求解。

[0035] 本发明与现有方法相比具有以下优点：

[0036] ①该方法通过阈值布隆过滤器存储DDS发现阶段的端点描述信息，能够减小内存消耗和网络数据传输，提高DDS在分布式仿真的实时性。

[0037] ②该方法提出阈值优化方式，实现较小的误报率和较大的精确度，特别是当DDS端点数较多、TBF较小时，效果更为显著。

[0038] ③该方法为DDS自动发现过程的改进提供了一种新的思路。

附图说明

[0039] 图1为本发明的基于阈值布隆过滤器的DDS自动发现方法流程图。

[0040] 图2为阈值布隆过滤器的构造的实例分析图；

[0041] 图3为阈值布隆过滤器自动发现方法执行过程图；

[0042] 图4为本地参与者与远程参与者建立通信的流程图；

[0043] 图5为基于阈值布隆过滤器的DDS自动发现方法阈值选择的实例示意图；

[0044] 图6为二值化阈值θ和判定阈值T的取值不同，误报率TPR的分布图；

[0045] 图7为二值化阈值θ和判定阈值T的取值不同，召回率FPR的分布图；

[0046] 图8为二值化阈值θ和判定阈值T的取值不同，传输精度ACC的分布图；

[0047] 图9为TBFAD与SDPBloom的召回率TPR分析曲线图；

[0048] 图10为TBFAD与SDPBloom的误报率FPR分析曲线图；

[0049] 图11为TBFAD与SDPBloom的传输精度ACC分析曲线图；

[0050] 图12为TBFAD与SDPBloom性能召回率TPR分析曲线图；

[0051] 图13为TBFAD与SDPBloom性能误报率FPR分析曲线图；

[0052] 图14为TBFAD与SDPBloom性能传输精度ACC分析曲线图；

[0053] 图15为TBFAD与SDPBloom性能召回率TPR分析曲线图；

[0054] 图16为TBFAD与SDPBloom性能误报率FPR分析曲线图；

[0055] 图17为TBFAD与SDPBloom性能传输精度ACC分析曲线图。

具体实施方式

[0056] 下面结合附图和具体实施方式对本发明进行详细说明。

[0057] 如图1所示，本发明提供的基于阈值布隆过滤器的DDS自动发现方法包括按顺序进行的下列步骤：

[0058] (1)阈值布隆过滤器的设计

[0059] 阈值布隆过滤器存储n个不同的端点信息元素x_i时，通过每个端点信息元素x_i映射结果的和，可以得到向量TBF(1)，即

[0060] (16) $$\mathrm{TBF}(1)=\sum_{i=1}^{n}x_i \quad \text{----------} \quad (1)$$

[0061] 如图2所示，阈值布隆过滤器的构造的实例：即向量TBF位数m＝20，哈希函数数量k＝3。对于集合S(x_1～x_4)，(x_1～x_4)中的每个端点信息元素都通过3个不同的哈希函数映射到3个不同的位置。

[0062] 当查询一个端点信息元素x_i是否属于集合S时，通过设置不同的二值化阈值θ和判定阈值T来判断端点信息元素x_i是否属于集合S，其中二值化阈值θ满足0≤θ≤k，判定阈值T满足0≤T≤k；首先，如果向量TBF(1)中每个位置的值小于或等于θ，那么这个位置置0；这时，向量TBF(1)变成一个新的向量TBF(2)；然后，通过判定阈值T判断即当一个端点信息元素x_i的映射结果与向量TBF(2)之间的点积值d大于或等于判定阈值T时，则判定这个端点信息元素x_i属于集合S.

[0063] (2)阈值布隆过滤器与DDS自动发现机制的结合

[0064] 如图3所示，使用两个节点之间通信的简单示例说明阈值布隆过滤器自动发现方法执行过程。数据从节点A发送到节点B，两节点A、B分别定义为本地参与者和远程参与者。

[0065] 如图4所示，在参与者发现阶段，将本地参与者端点即数据写入者、数据读取者的描述信息存储在向量TBF(2)中，并通过本地参与者数据包一起发给其他远程参与者，这些端点描述信息是每个本地参与者唯一的关键字，通常是主题名称；在端点发现阶段，当远程参与者的端点订阅的一个或多个主题时，远程参与者首先查询向量TBF(2)中是否存在其订阅的主题；如果存在，则远程参与者向本地参与者发送存在主题订阅信息；本地参与者发送与远程参与者相关的主题数据包和服务质量进行进一步匹配，如果匹配成功，本地参与者与远程参与者建立通信。

[0066] (17)最佳阈值的确定过程

[0067] 如图5所示，图5举例说明二值化阈值θ和判定阈值T的选择是阈值布隆过滤器DDS

自动发现方法的关键部分，具体自动发现过程描述如下：

[0068] 在如图2得到向量TBF(1)的基础上，通过阈值布过滤器DDS自动发现方法查询一个端点信息元素y是否属于集合S(x_1～x_4)。在本例中，使用向量y_1表示端点信息元素y在TBF中映射的结果。

[0069] 图5(b)和图5(c)表示公式(5)的二值化阈值分别取θ＝0和θ＝1。当θ＝0时，根据公式(8)端点信息元素y和向量TBF(2)的点积值d是3。因为k＝3，判定阈值T范围是[0,k]，则d≥T总是成立，故无论判定阈值T取何值，结论总是端点信息元素y属于集合S。然而，实际上端点信息元素y不是集合S中的成员。所以，此查询过程错误，将会导致自动发现过程的失败。

[0070] 当二值化阈值θ＝1时，端点信息元素y和向量TBF(2)的点积值d是1。如果判定阈值范围是T≤1，那么d≥T。在这种情况下，端点信息元素y仍被判定属于集合S，产生了错误判断。如果判定阈值范围T≥2时，将会产生正确的判断，即端点信息元素y不属于集合S，最终自动发现过程成功。

[0071] 上述讨论表明，阈值的选择对自动发现过程非常重要。本发明中通过以下方法求取最佳阈值。将传输精度(ACC)作为优化目标函数：

[0072] (18) $$ACC=\frac{(TPR+(1-FPR))}{2}=\max \quad \text{----------（13）}$$

[0073] 其中约束条件分别为公式(2)、公式(3)、公式(9)、公式(10)、公式(11)和公式(12)。通过遗传算法迭代出传输精度ACC最大值，并求出传输精度ACC为最大值时的二值化阈值θ和判定阈值T。

[0074] 为了验证本发明提供的基于阈值布隆过滤器的DDS自动发现方法的有效性，本发明对其进行了实验，过程如下：

[0075] 为了进一步验证本发明提供的基于阈值布隆过滤器的DDS自动发现方法的性能，与目前比较认可的改进自动发现方法(SDPBloom)进行了4组对比实验。

[0076] 实验中，SDPBloom的召回率TPR总是1，误报率FPR_{SB}为：

[0077] (19) $$FPR_{SB}=\left[1-(1-\frac{1}{m})^{kn}\right]^k\approx\left(1-e^{-\frac{kn}{m}}\right)^k \quad \text{----------（14）}$$

[0078] SDPBloom的传输精度ACC也由公式(13)得出。

[0079] 当阈值布隆过滤器位数m、存储端点信息数量n、哈希函数数量k为定值时，本方法与SDPBloom对比分析如下：

[0080] 实验选取Windows 10系统作为实验平台的操作系统，其硬件的主要参数如下：CPU型号为Intel CPU Core i5-3210，CPU主频为2.5GHz，内存12.0GB。假设m＝500，n＝1000，k＝30，通过阈值布隆过滤器DDS自动发现方法(TBFAD)求得最优二值化阈值θ＝5和判定阈值T＝20。与SDPBloom对比如表1。

[0081] 表1 SDPBloom与TBFAD方法性能对比

[0082]

方法类型	TPR	FPR	ACC
SDPBloom	1	0.93	0.54
TBFAD	0.85	0.24	0.81

[0083] 通过TBFAD方法，误报率FPR从0.93减小到0.24，传输精度ACC从0.54增加到0.81，这个结果表明数据传输错误率大幅度下降。尽管召回率TPR略微减小，但是整个过程通信性能明显提高。

[0084] 如图6所示，二值化阈值θ和判定阈值T的取值不同，召回率TPR的分布情况验证了这个方法对最优阈值的有效性。

[0085] 如图7所示，二值化阈值θ和判定阈值T的取值不同，误报率FPR的分布情况验证了这个方法对最优阈值的有效性。

[0086] 如图8所示，二值化阈值θ和判定阈值T的取值不同，传输精度ACC的分布情况验证了这个方法对最优阈值的有效性。

[0087] 此外，还对不同的阈值布隆过滤器位数m、存储端点信息数量n、哈希函数数量k进行了相同的实验，也得到了类似的结论。

[0088] 当存储端点信息数量n变化时，TBFAD与SDPBloom对比分析如下：

[0089] 如图9、图10、图11所示，在上述实验的基础上，假设阈值布隆过滤器位数m、哈希函数数量k不变，存储端点信息数量n从10到150变化，对TBFAD与SDPBloom的召回率TPR、误报率FPR、传输精度ACC进行分析，可以得到以下结论：

[0090] (1)在存储端点信息数量n增加到30之前，两种方法的召回率TPR都是最大值1。此后，TBFAD的误报率TPR下降，但在存储端点信息数量n增加到150之前，TBFAD的召回率TPR减小量在0.2以内。

[0091] (2)在存储端点信息数量n增加到30之前，这两种方法的误报率FPR都是最小值0。此后，两者的误报率FPR均增加，但TBFAD的变化(n＝150时为0.28)明显小于SDPBloom(n＝150时为1)。

[0092] (3)传输精度ACC由召回率TPR和误报率FPR计算得出。对于SDPBloom，在存储端点信息数量n大于30时开始迅速下降，当n＝150时下降到0.5。但对于TBFAD，在n大于70时开始缓慢下降，当n＝150时，传输精度ACC仅降至0.77。

[0093] 存储端点信息数量n由本地参与者端点数量决定，在大型分布式仿真系统中，存储端点信息数量n的值一般会很大。根据上述结论，当本地参与者端点数目较大时，与SDPBloom相比，TBFAD方法具有明显的优势，保证了大规模分布式仿真中数据传输的正确性。

[0094] 当阈值布隆过滤器位数m变化时，TBFAD与SDPBloom对比分析如下：

[0095] 如图12、图13、图14所示，假设存储端点信息数量n、哈希函数数量k不变，阈值布隆过滤器位数m从100到1000变化，对TBFAD与SDPBloom的召回率TPR、误报率FPR、传输精度ACC性能进行分析，可以得到以下结论：

[0096] (1)在100≤m≤1000，TBFAD的召回率TPR始终略低于SDPBloom，但随阈值布隆过滤器位数m的增加TBFAD的召回率TPR增大。当m＝1000时，TBFAD的召回率TPR可达0.9。

[0097] (2)两种方法的误报率FPR均随阈值布隆过滤器位数m的增加而降低，但TBFAD的误

报率FPR远小于SDPBloom。

[0098] (3)两种方法的传输精度ACC均随阈值布隆过滤器位数m的增加而增加,但TBFAD的传输精度ACC始终高于SDPBloom。当阈值布隆过滤器位数m较小时,两种方法的传输精度ACC都较低,但TBFAD明显优于SDPBloom。

[0099] 在DDS的参与者发现阶段,较小长度的向量可以减少的带宽使用。在本实验中,当布隆过滤器位数m的位数较少时,本发明提出的TBFAD方法取得了良好的效果。因此,在大规模分布式仿真中,可以保证本地参与者和远程参与者之间的具有较小的数据传输.

[0100] 当哈希函数数量k变化时,TBFAD与SDPBloom对比分析如下:

[0101] 如图15、图16、图17所示,假设阈值布隆过滤器位数m、存储端点信息数量n不变,哈希函数数量k从5到50变化,对TBFAD与SDPBloom的召回率TPR、误报率FPR、传输精度ACC进行分析,可以得到以下结论:

[0102] (1)当哈希函数数量k≤10时,TBFAD与SDPBloom的召回率TPR都为1。此后,TBFAD的召回率TPR开始下降,但在哈希函数数量k增加到50之前,变化范围一直在0.25之内。

[0103] (2)当哈希函数数量k≤10时,TBFAD的误报率FPR与SDPBloom具有相同的趋势。随着哈希函数数量k的增加,TBFAD的误报率FPR保持在[0.1,0.25]的较小范围内,而SDPBloom的误报率FPR迅速上升。

[0104] (3)对于SDPBloom,随着哈希函数数量k的增加,传输精度ACC开始迅速下降,当哈希函数数量k=50时,传输精度ACC=0.5。但TBFAD的传输精度ACC缓慢下降,当哈希函数数量k=50时,仅下降到0.8。

[0105] 在构造向量TBF时,哈希函数数量决定了计算的复杂性。结果表明,在分布式仿真系统中,当哈希函数数量k选择较小时,可以减少计算量,TBFADF优于或等于SDPBloom的性能。

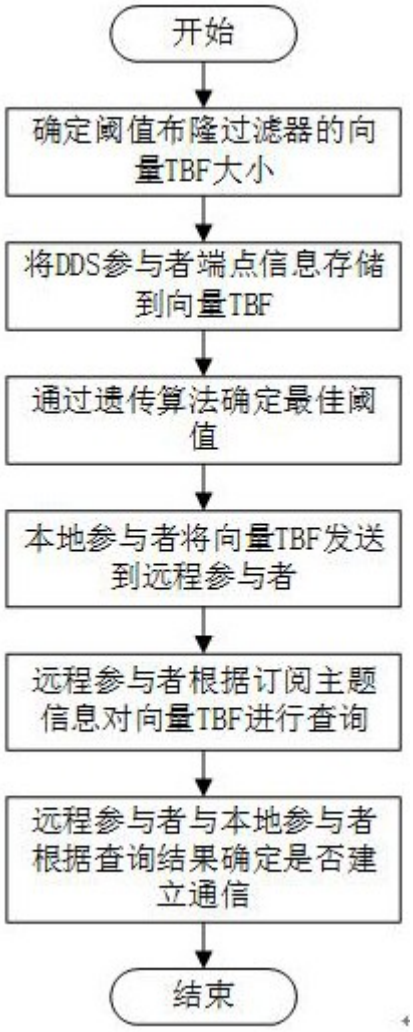

开始
确定阈值布隆过滤器的向量TBF大小
将DDS参与者端点信息存储到向量TBF
通过遗传算法确定最佳阈值
本地参与者将向量TBF发送到远程参与者
远程参与者根据订阅主题信息对向量TBF进行查询
远程参与者与本地参与者根据查询结果确定是否建立通信
结束

图1

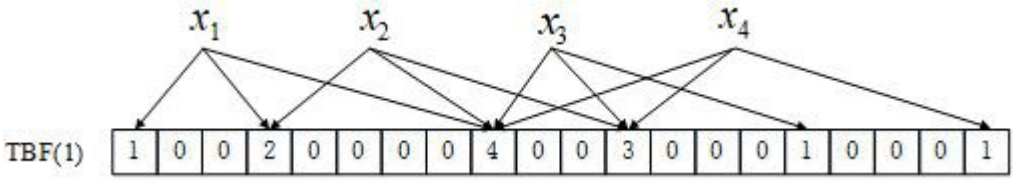

x_1
x_2
x_3
x_4
TBF(1)
1 0 0 2 0 0 0 0 4 0 0 3 0 0 0 1 0 0 0 1

图2

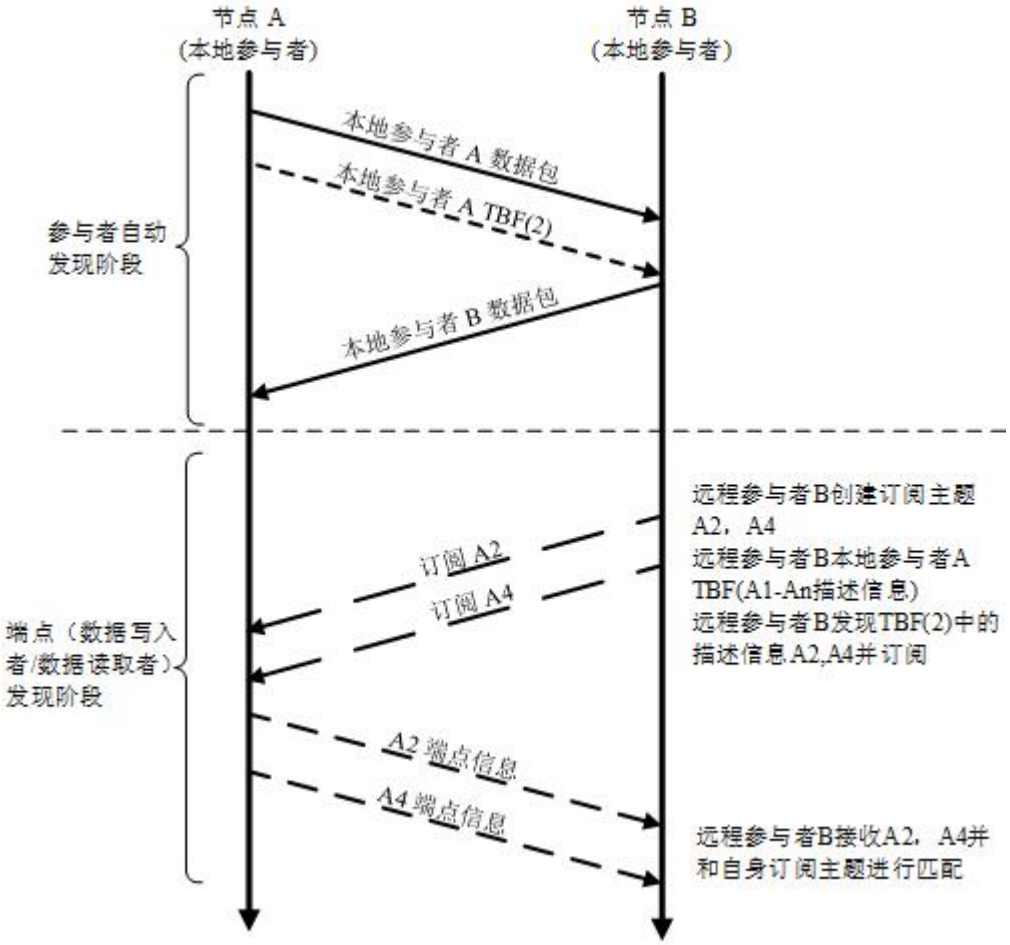

节点 A
(本地参与者)
节点 B
(本地参与者)
本地参与者 A 数据包
本地参与者 A TBF(2)
本地参与者 B 数据包
参与者自动
发现阶段
远程参与者B创建订阅主题
A2，A4
远程参与者B本地参与者A
TBF(A1-An描述信息)
远程参与者B发现TBF(2)中的
描述信息A2,A4并订阅
订阅 A2
订阅 A4
端点（数据写入
者/数据读取者）
发现阶段
A2 端点信息
A4 端点信息
远程参与者B接收A2，A4并
和自身订阅主题进行匹配

图3

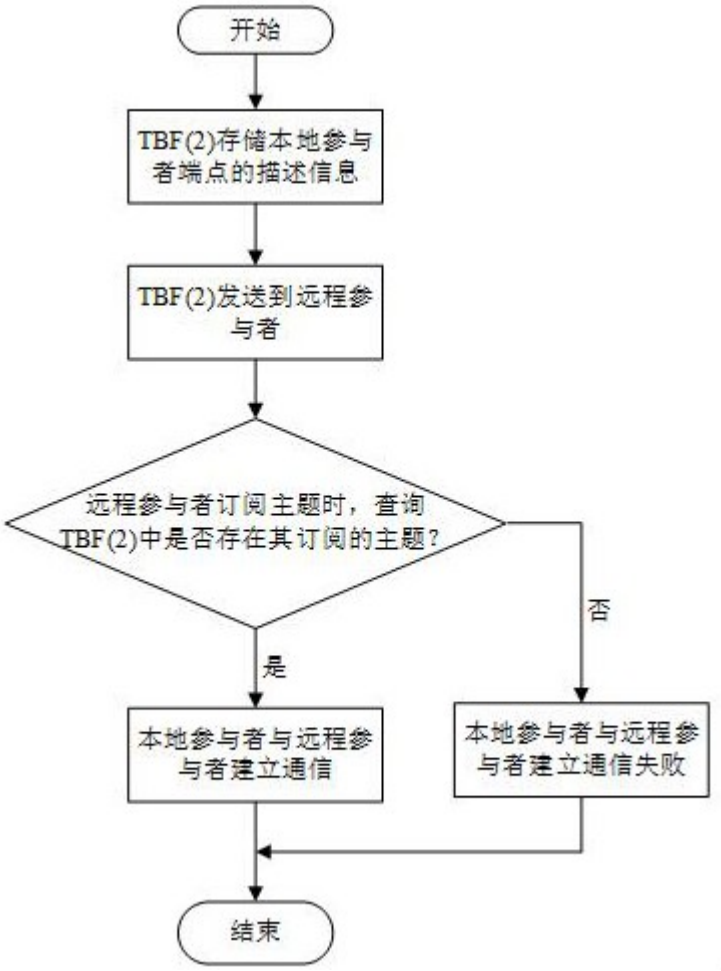

开始
TBF(2)存储本地参与者端点的描述信息
TBF(2)发送到远程参与者
远程参与者订阅主题时，查询TBF(2)中是否存在其订阅的主题？
否
是
本地参与者与远程参与者建立通信
本地参与者与远程参与者建立通信失败
结束

图4

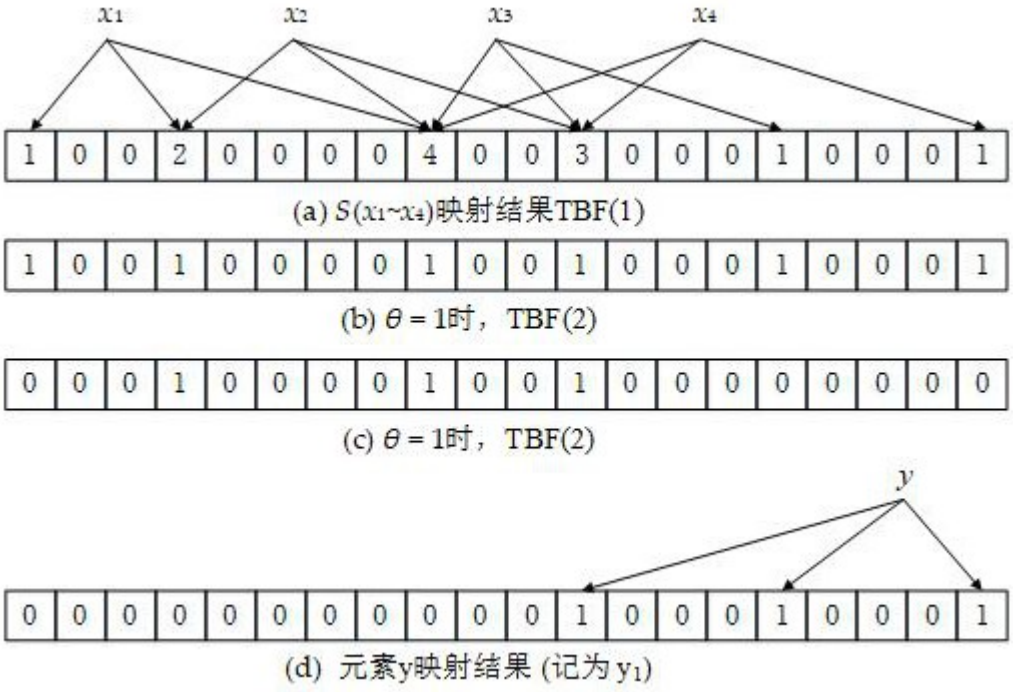

x1
x2
x3
x4
1 0 0 2 0 0 0 0 4 0 0 3 0 0 0 1 0 0 0 1
(a) S(x1~x4)映射结果TBF(1)
1 0 0 1 0 0 0 0 1 0 0 1 0 0 0 1 0 0 0 1
(b) θ = 1时， TBF(2)
0 0 0 1 0 0 0 0 1 0 0 1 0 0 0 0 0 0 0 0
(c) θ = 1时， TBF(2)
y
0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 1 0 0 0 1
(d) 元素y映射结果 (记为 y1)

图5

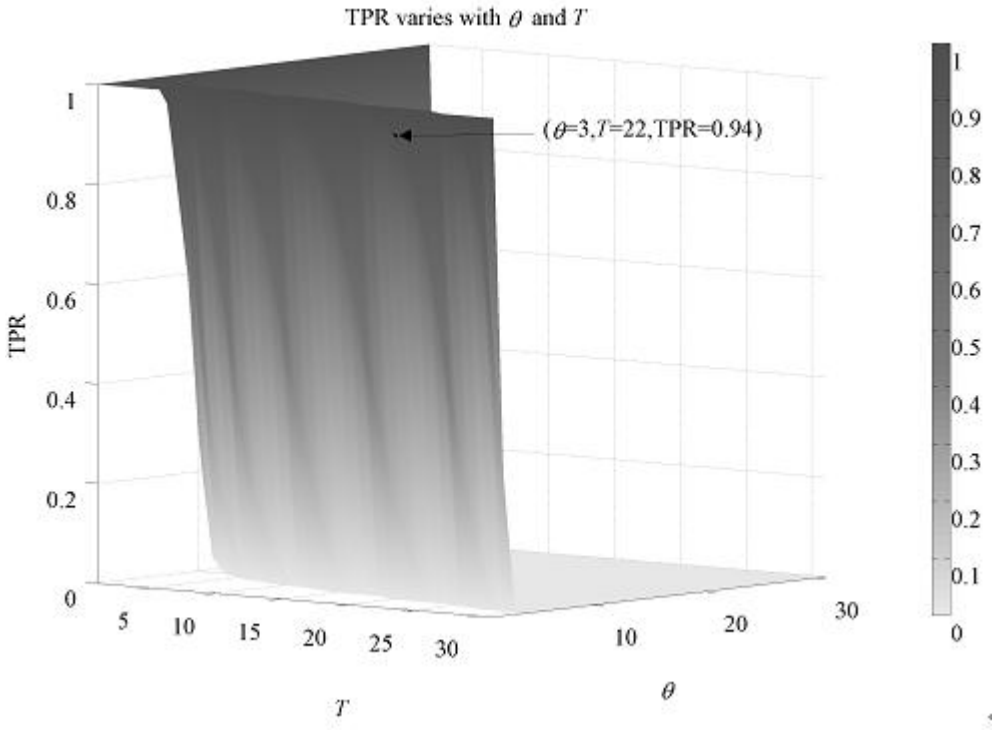

TPR varies with θ and T
(θ=3,T=22,TPR=0.94)
TPR
1
0.8
0.6
0.4
0.2
0
5
10
15
20
25
30
T
10
20
30
θ
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0

图6

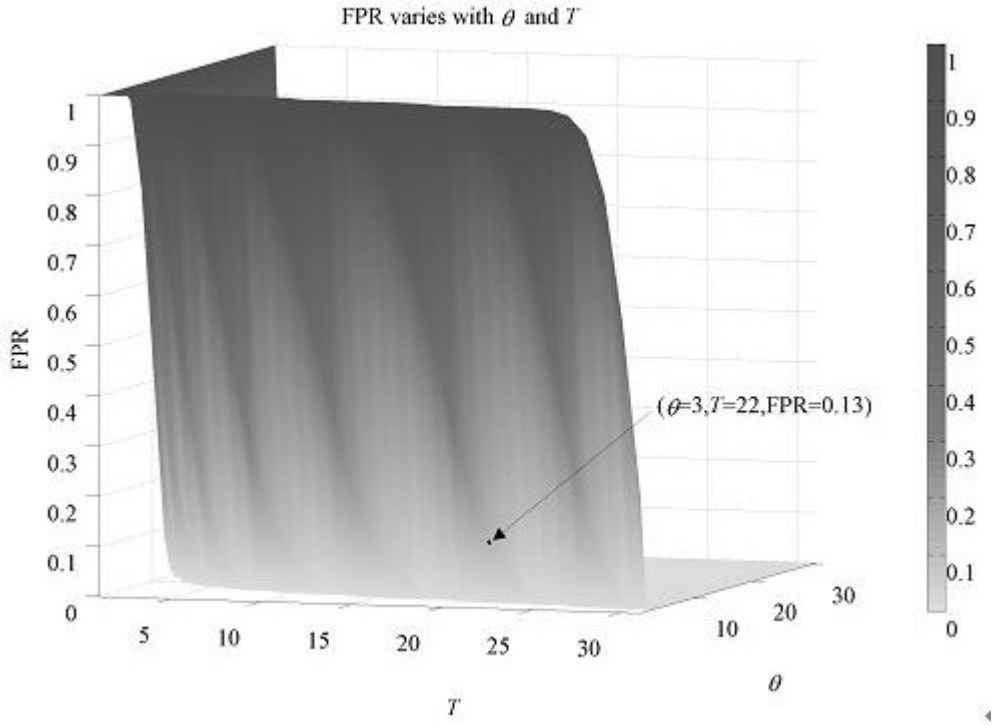

FPR varies with θ and T
FPR
T
θ
(θ=3,T=22,FPR=0.13)

图7

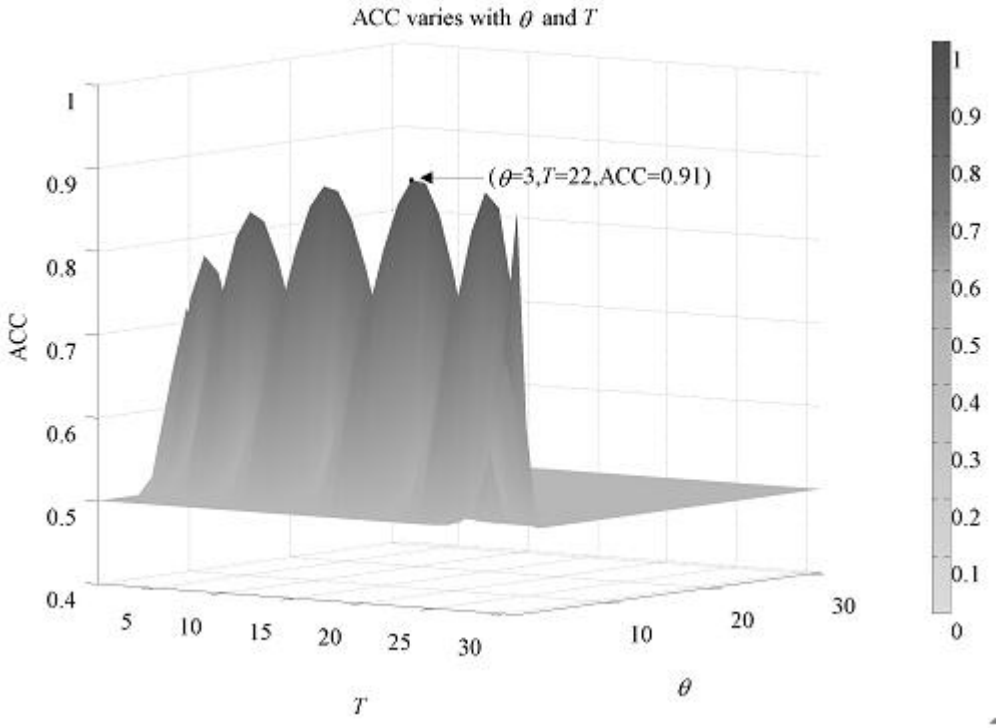

ACC varies with θ and T
ACC
T
θ
(θ=3,T=22,ACC=0.91)

图8

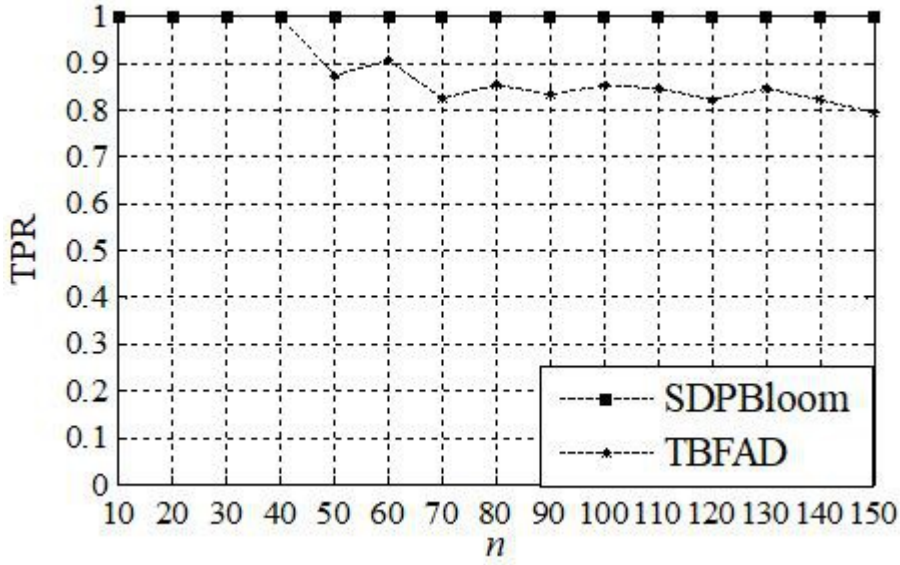


图9

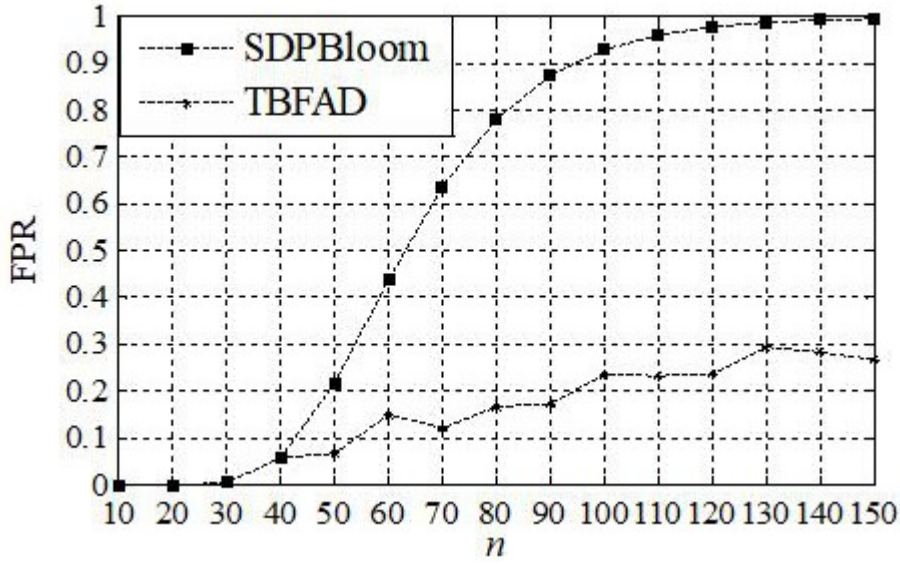


图10

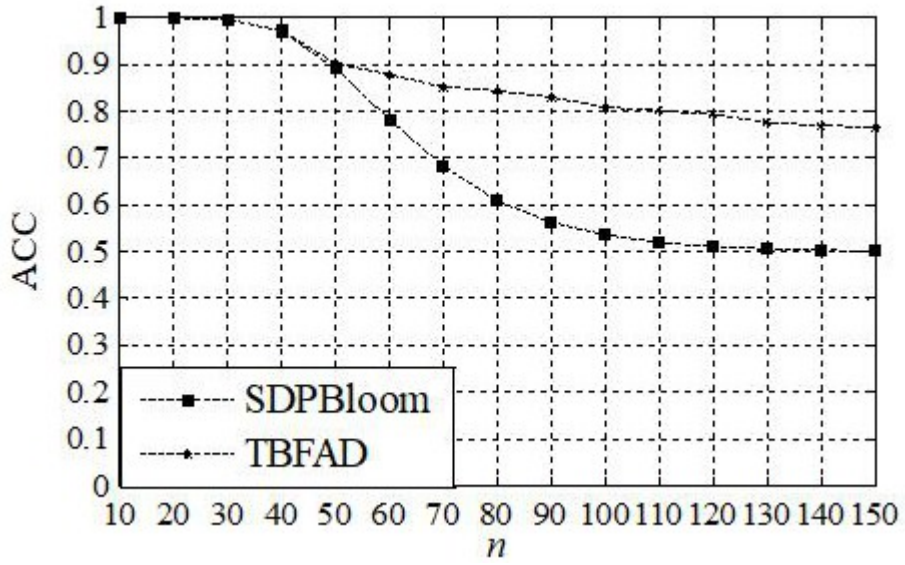

ACC
n
SDPBloom
TBFAD
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
10 20 30 40 50 60 70 80 90 100 110 120 130 140 150

图11

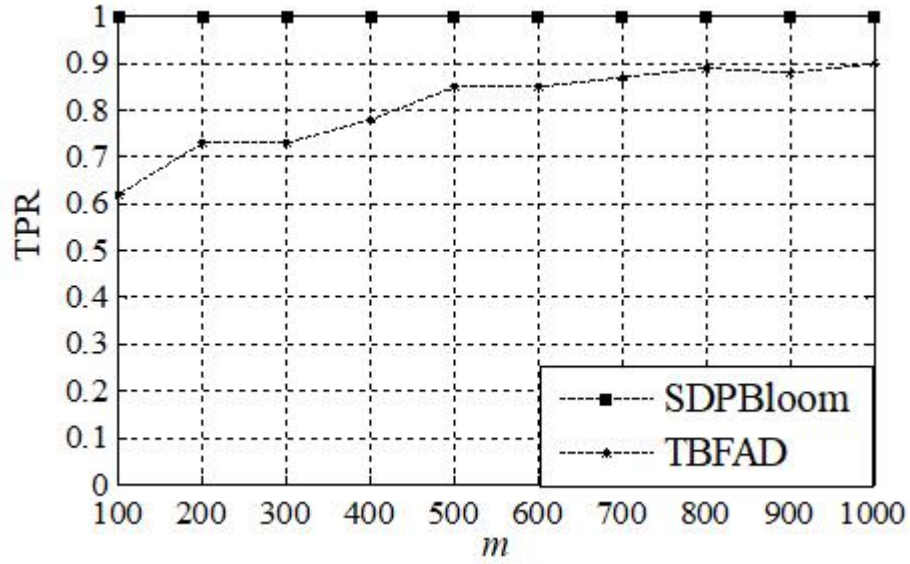

TPR
m
SDPBloom
TBFAD
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
100 200 300 400 500 600 700 800 900 1000

图12

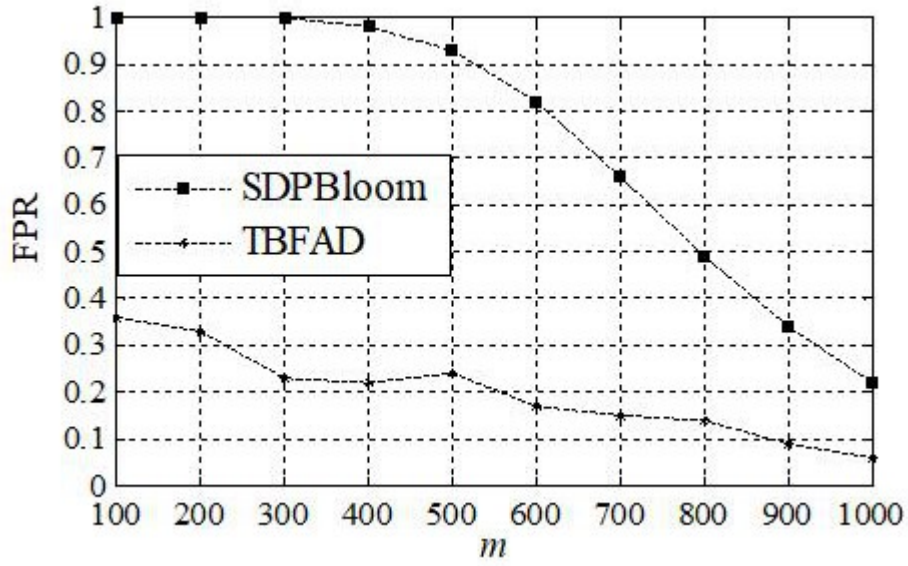

FPR
m
SDPBloom
TBFAD
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
100
200
300
400
500
600
700
800
900
1000

图13

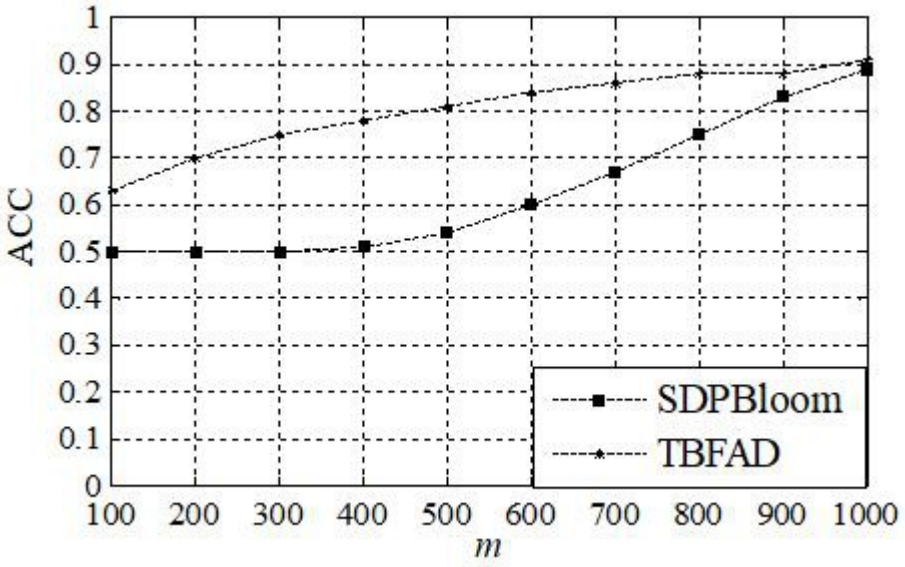

ACC
m
SDPBloom
TBFAD
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
100
200
300
400
500
600
700
800
900
1000

图14

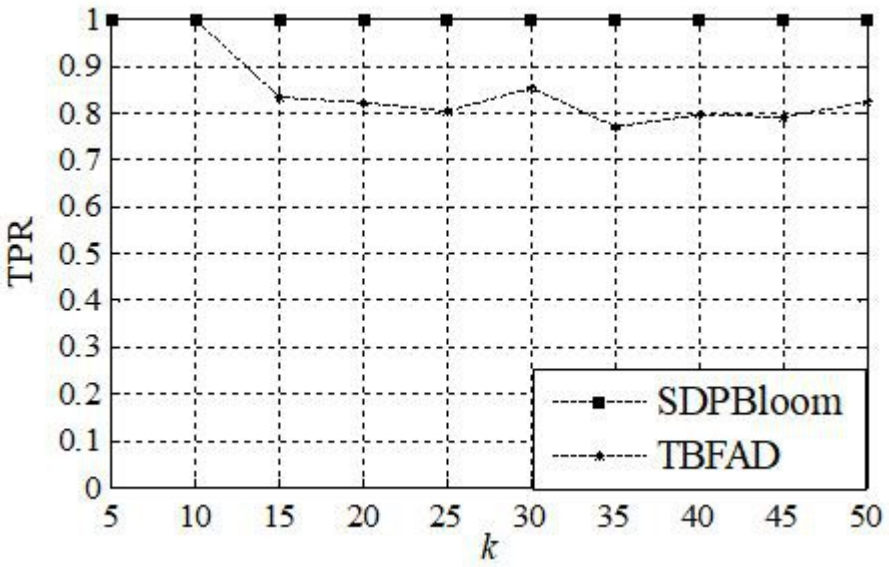


图15

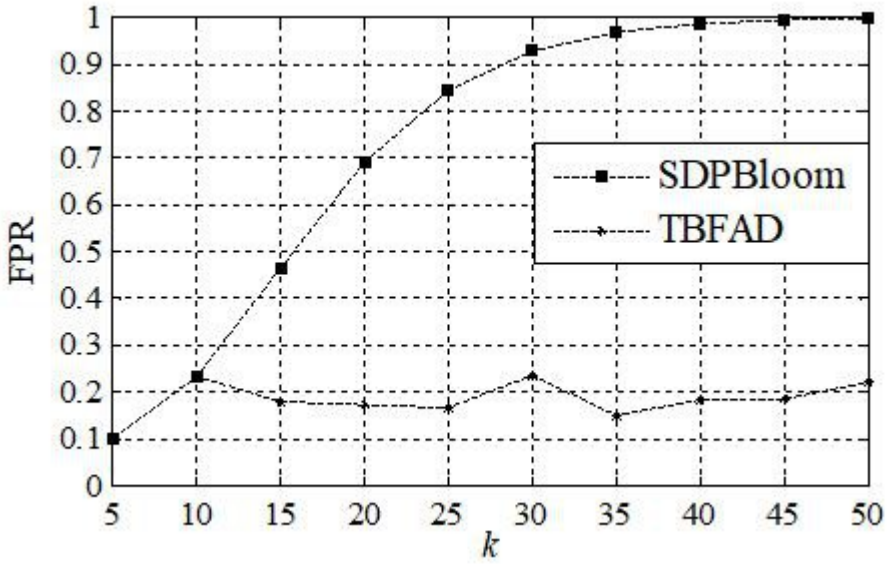


图16

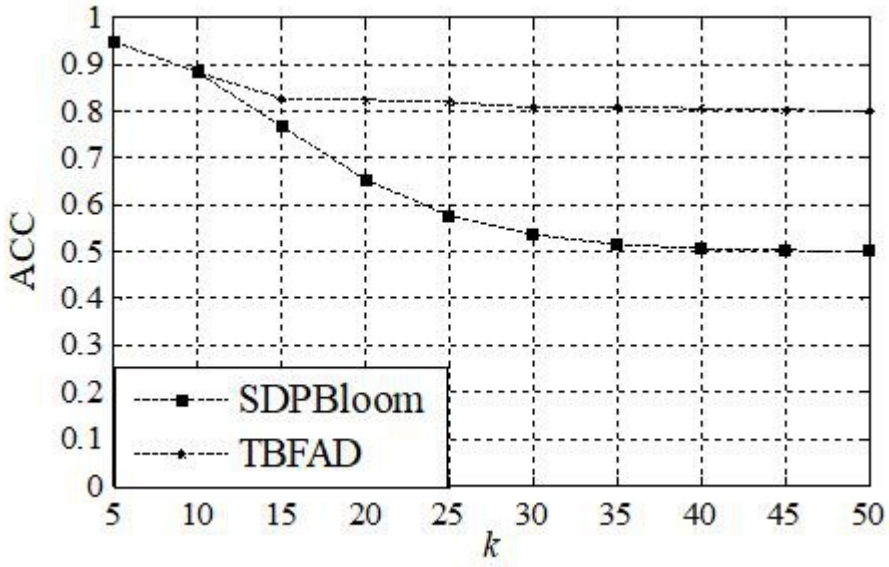

ACC
k
SDPBloom
TBFAD
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
5
10
15
20
25
30
35
40
45
50

图17