

基于小世界的物联网中副本定位算法

高 辉, 张冬悦, 赵俊涛, 吴吉红

(辽宁大学信息学院, 沈阳 110036)

摘 要:物联网具有数据量大,节点异构且分布广泛等特点。采用副本存储可以提高数据访问性能。针对物联网中对副本定位快速高效要求的问题,提出一种基于小世界的类-域层次的物联网中副本定位方法。从物联网中数据访问的局部聚类性考虑,将节点划分为类层次和域层次,分别采用根据类-域数据访问特点的双向环形消息扩散机制,实现副本的局部和全局定位,从而提高副本定位的效率。

关键词:物联网;副本定位;小世界

中图分类号:TP393

文献标识码:A

引 言

物联网^[1-2]把物体通过网络互联,实现更广范围上的通信。在这种分布式的网络环境下,考虑到物联网^[3-4]很大,数据分布广泛,节点数量大、异构性强,智能程度和存储能力差别很大,通信过程对实时性和安全性的要求很高,因此对数据资源进行适当的复制是必不可少的。不仅可以提高环境的容错能力,还能平衡节点负载,减少不必要的数据传输。因此副本存储是提升物联网性能的重要技术。副本存储势在必行。在物联网中,如何快速高效的定位副本是分布式副本存储模式中面临的重要问题。

目前,提出了很多方法解决副本定位问题。文献^[5]采用层次式索引的方法,由顶到底查找,从全局定位到区域定位,逐层接近目标副本节点。但此方法需要遍历所有的索引节点,时间复杂度高,在物联网中,节点众多,单一的层次式索引方法消耗时间太大。文献^[6]提出的基于分布式哈希表的 P2P 副本定位服务 PRLS,但没有考虑到数据访问的局部聚类性,物联网中节点有很强的类别差异性,构造定位方法时应该考虑这一特点。

根据上述文献,本文引入小世界理论^[7-8],提出一种基于小世界的类-域层次的副本定位方法。将节点划分层次,在类和域的范围上,分别连成双向环,沿环的两个邻居方向查找并采用不同策略选取远方的节点扩散

查找。

1 节点的类聚集特性分析

为了将小世界理论^[9]应用到物联网的副本定位算法中,提高效率,首先来讨论物联网中节点类聚集特性。在文献^[10]中,叙述了分布式环境下数据访问具有局部聚类特征。根据集合论^[11]和数据挖掘相关知识,本文结合物联网的特点,设计采用模糊聚类法,证明物联网中数据访问也具有聚类特征。选取物联网中的节点对象,并通过属性进行标识。假设有 n 个节点,记为 O_1, O_2, \dots, O_n , 每个节点选取 m 个属性,记为 $1, 2, \dots, m$ 。这样得到一个 $n * m$ 的矩阵 $Y = [y_{ij}]$, 其中 y_{ij} 就是第 i 个节点对象对应于第 j 个属性的标度。行向量 Y_i 描述了 O_i 对象。

$$Y = \begin{bmatrix} y_{11} & y_{12} & \cdots & y_{1m} \\ y_{21} & y_{22} & \cdots & y_{2m} \\ \cdot & \cdot & \cdot & \cdot \\ y_{n1} & y_{n2} & \cdots & y_{nm} \end{bmatrix} \quad (1)$$

节点种类很多,各项属性的数据参差不齐,对各指标的数据(矩阵各列)进行标准化,按比例缩放,即把属性数据转换到 $[0, 1]$ 区间内。选择最小-最大规范化方法:令 A_{ij} 为 y_{ij} 相应列的最小值,即 $A_{ij} = \min \{y_{ij}, (i = 1, 2, \dots, n)\}$, B 为 y_{ij} 中相应列的最大值,即 $B_{ij} = \max \{y_{ij}, (i = 1, 2, \dots, n)\}$ 。得到标准化的矩阵 X , 其每一元素为

收稿日期:2012-02-24

基金项目:国家自然科学基金资助项目(60703068)

作者简介:高 辉(1987-),男,辽宁朝阳人,硕士,主要从事计算机网络、物联网方向的研究,(E-mail)ghui0318@163.com

$$x_{ij} = \frac{y_{ij} - A_{ij}}{B_{ij} - A_{ij}} \quad (2)$$

在标准化对象属性后,计算节点对象的相似度(相异度),采用欧几里得距离度量 $d(ij)$ 。进而得到相似关系矩阵 D 。

$$D = \begin{bmatrix} 0 & & & & & \\ d(21) & 0 & & & & \\ d(31) & d(32) & 0 & & & \\ \dots & \dots & \dots & \dots & \dots & \\ d(n1) & d(n2) & d(n3) & \dots & 0 & \end{bmatrix} \quad (3)$$

式中:

$$d(ij) = \sqrt{\sum_{t=1}^m q_t (x_{it} - x_{jt})^2} \quad (i, j = 1, 2, \dots, n) \quad (4)$$

网络中的节点通过属性来标识,当两个节点通信时,相互依靠属性判断识别。选取三个对象 O_p, O_q, O_r , 当对象 O_p 和 O_q 各自的相对应的属性维度上 x_{pj}, x_{qj} 值相同或相近, $|x_{pj} - x_{qj}| \approx 0$, 相似度高,两个对象容易相互识别,容易建立关系,通信的概率大, $P(O_p \leftrightarrow O_q) \rightarrow 1$ (表示 O_p 和 O_q 相互通信的概率)。相反,当对象 O_p 和 O_r 各自的相对应的属性维度上 x_{pj}, x_{rj} 值相差很大, $|x_{pj} - x_{rj}| \approx 1$, 相似度低,表现出很强的跨属性倾向,这样的关系很难建立,通信的概率小, $P(O_p \leftrightarrow O_r) \rightarrow 0$ 。因此, $P(O_p \leftrightarrow O_q) \geq P(O_p \leftrightarrow O_r)$ 。例如在一起玩的小朋友会形成一个圈子。若集合 G 中的两个元素的距离 $d(ij)$ 都满足

$$d(ij) \leq T \quad (i, j \in G) \quad (5)$$

则 G 对阈值 T 组成一个类。

根据公式分析,当两个对象属性越相似或越“近”, $d(ij)$ 值越接近 0, 两个对象越相同。数据访问有局部性,随着相近对象增加,表现出聚类性质。由此,数据访问呈现出类意义上的局部聚集性。

已知在两个地理域 $\text{Domain} - S$ 和 $\text{Domain} - T$ 中, $S = \{O_1, O_2, O_m, O_n, O_r, \dots\}$, $T = \{O_1, O_2, O_p, O_q, O_r, \dots\}$, 当能满足相同的通信需求时,可以认为通信概率 $P(O_p \in T \leftrightarrow O_r \in T) \geq P(O_p \in T \leftrightarrow O_r \in S)$ 。因此,数据访问优先选择相同地域或近地域的对象。

综上,在物联网中访问对象时,相近域、相似对象通信概率大,呈现局部聚类的小世界特性。

2 体系结构

如图 1 为本文体系结构,分为三层:底层将属性相同或相似度高的普通节点组成一个类,每个类选出一个类中心节点。中间层把地域近的各个类中心节点组成一个域,域表示一定地域范围内的节点,每个域选取一

个域中心节点。顶层选取部分域中心节点做为全局定位节点。

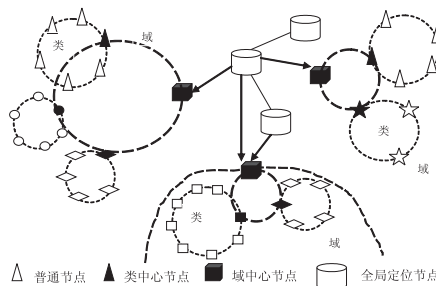


图 1 体系结构

根据全局唯一的逻辑资源名找到一个或多个物理副本即为副本定位过程。

普通类节点:分布在各地,存放实际的物理资源,副本信息表,邻居表。

类中心节点:存放本类中的类别号,各个节点的访问表,本类中逻辑资源名。

域中心节点:存放各类副本信息汇总后的域索引,域中节点类别号表,本域中的逻辑资源名。

全局定位节点:存放各域副本信息汇总后的全局索引,全局路由表,全局逻辑资源名。

3 基于小世界的环形定位方法

由前文得知,一个域可以看作一个小世界。根据小世界理论,在类和域层次上采用基于小世界的环形消息扩散方法来定位副本。在该方法中,寻找远方节点是关键。

3.1 类层次上消息扩散

在类层次上,基于小世界模型的环形消息扩散方法过程为:按普通类节点进入类的先后顺序,赋予一个单调递增的编号,从小到大按升序的方式将普通节点连成一个环。当某节点要定位数据资源的副本时,将定位消息分别沿两个邻居节点方向扩散,同时也将副本定位消息发向一个或多个远方节点。以此迭代,直到该副本信息被找到或者各个节点都被访问过,消息扩散过程结束,如图 2 所示。

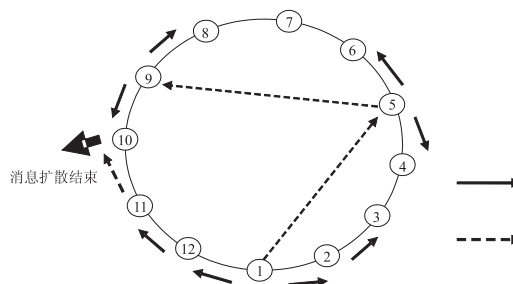


图 2 类层次环形消息扩散定位方法

远方节点的选择条件如下:根据节点访问表,选取

类中最近访问量大的节点,这些最近访问的远方节点与当前节点的编号形成新的环,环上各节点编号间的距离均大于等于 $3 + m$, (m 表示扩散轮数,即初始扩散源寻找远方节点时 m 为 0,每迭代一次, m 加 1)。远方节点选取依据下列条件:

$$|N_{i+1} - N_i| \geq 3 + m \text{ 且 } [|n - N_{i+1} + N_i| \geq 3 + m (\text{当 } N_{i+1} > N_i) \text{ 或 } |n - N_i + N_{i+1}| \geq 3 + m (\text{当 } N_{i+1} < N_i)]$$

其中 n 为类中节点个数, N_i 为当前节点编号, N_{i+1} 为新选节点编号, m 表示扩散轮数。

算法 1:消息扩散

- ①当前节点从邻居表中取得两个邻居的编号和最近访问节点的编号;
- ②节点将最近访问节点的编号、 n 、 m 运用于表达式 1,从而得到符合条件的远方节点的编号;
- ③节点根据邻居节点的编号和远方节点的编号,从邻居表中找到其 IP 地址,将定位消息发送到这些节点;
- ④收到信息的节点是否已经遍历,如果已遍历转向⑧;如果未遍历转向⑤;
- ⑤节点根据消息来自于何处,如果来自于邻居,则转向⑥;如果来自于远方节点,转向⑦;
- ⑥邻居节点从邻居表中取出各自的两个邻居,并将消息发送它们,转向④;
- ⑦远方节点产生一个新的当前节点,并将 m 加 1 后,转向①;
- ⑧节点结束扩散,并返回 NULL;

图 2 为类层次消息扩散过程,节点 1 为消息源节点,向邻居节点 2、邻居节点 12 和远方节点 5 扩散消息,依次迭代,节点 5 向其邻居节点和远方节点扩散消息。最终在节点 10 消息扩散结束。

3.2 域层次上查找

如图 3 所示,在域层次上,基于小世界的环形消息扩散方法过程与类层次上的相似,但在远方节点的选择策略上有所不同。

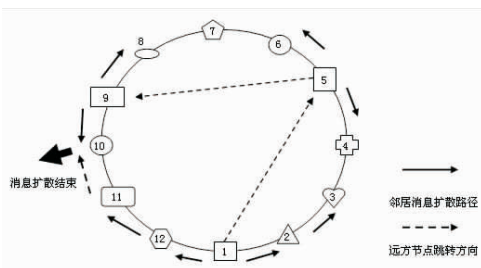


图 3 域层次环形消息扩散定位方法

一个域中有很多个类,根据上文中模糊聚类法,同时结合物联网特点,依照节点的属性,计算任意两类节点的相似度(相异度),见表 1。

远方节点的选择条件如下:根据节点属性,建立类相似度表,根据相似度表,选择与当前节点相似度高的节点。然后运用表达式(1),从而得到远方节点。

表 1 节点 1 的相似度表

节点编号	相似度	
节点 5	95%	□
节点 9	90%	□
节点 11	80%	□
节点 4	70%	⊕
节点 7	60%	◇
节点 2	50%	△
节点 12	50%	◇
节点 6	20%	○
节点 10	20%	○
节点 8	10%	◌
节点 3	10%	♡

3.3 副本定位过程

在本文的体系结构中,基于小世界的环形消息扩散方法的副本定位过程如下:

本地节点收到副本定位信息后,首先将副本名与自己的副本信息表比较,如果表中有想要的信息,则将结果返回,如果没有找到所需信息,则调用算法 1 将副本定位消息扩散,收到消息的节点与自己的信息表比较,如果找到该副本信息,则将结果返回,并结束消息扩散。如果没有找到,则在域层次调用算法 1 将该副本定位消息扩散,收到消息节点与自己的信息表比较,如果找到将结果返回,直到本域所有节点都收到该消息,如果未找到,将消息发散到顶层全局定位节点,转向全局副本定位,重新定位新的域节点。

算法 2:副本定位

- ①本地节点与自身副本定位信息表比较,看本地节点是否有副本资源,如果有,取出副本资源,转向⑨;如果没有找到,转向②;
- ②本地节点调用算法 1 进行消息扩散;
- ③类中节点收到定位消息后,与自己的副本定位信息表进行比较,如果找到,则将自己的 IP 地址和编号返回给源节点,消息扩散结束,转向④;如果没有找到,转向②;
- ④源节点判断结果是否为 NULL,如果不是 NULL,则取出结果中的 IP 和编号,转向⑨;如果是 NULL,则转向域层次副本定位⑤;
- ⑤域中节点调用算法 1 进行消息扩散;如果消息扩散结束,转向⑧;
- ⑥域中节点收到定位消息后,与自己的副本定位信息表比较,如果找到,则将自己的 IP 地址和编号返回给源节点,消息扩散结束,转向⑦;如果未找到,转向⑤;

⑦源节点判断结果是否为 NULL,如果不是 NULL,则取出结果中的 IP 和编号,转向②;如果是 NULL,则转向全局索引定位⑧;

⑧在全局索引定位节点上查找逻辑资源,如果找到,定位出相应的域节点,转向⑤;如果未找到,转向⑨;

⑨副本定位结束。

4 实验分析

本文提出的副本定位方法,充分考虑物联网中数据访问的特点,在类层次上,利用对节点访问在时间上的聚集性,在域层次上,利用数据访问时在节点类型上的聚集性,引入小世界理论,设计了基于类-域层次的副本定位算法,提高了系统的效率。

将本文副本定位算法的平均响应时间与 PRLS 的比较:取出与 PRLS 实验中相同数量的节点和相同数量的副本定位信息(每个节点存放 10000 条副本定位信息),同时进行 1000 次副本查询,测出本文算法的平均响应时间与 PRLS 的平均响应时间的对比,其结果如图 4 所示。

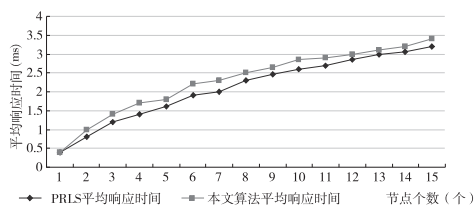


图 4 副本定位平均响应时间与 PRLS 对比

从图 4 中可以看出,相比之下本文算法的平均响应时间有所减少。

5 结束语

本文提出一种基于小世界的类-域层次的副本定位方法。该方法在论证了在物联网中数据访问具有小世

界特性的基础上提出,把小世界理论应用于副本定位中。采用环形消息扩散,建立类、域和全局的三层体系结构。实验表明,该方法具有较好的性能和可用性。

参考文献:

- [1] 李一,陈火峰.关于物联网的研究思考[J].价值工程,2009,29(8):126-127.
- [2] 姚万华.关于物联网的概念及基本内涵[J].中国信息界,2010,5:22-23.
- [3] 孙其博,刘杰,黎彝,等.物联网的概念、架构与关键技术研究综述[J].北京邮电大学学报,2010,33(3):1-9.
- [4] 蒋林涛.互联网与物联网[J].专家视点,2010,2:1-5.
- [5] 刘高嵩,张小波,龙军.数据网格中的副本定位[J].计算机系统应用,2010,19(10):71-75.
- [6] Chervenak A, Cai min, Frank M. Applying Peer-to-Peer Technology to Grid Replica Location Services[J].Journal of Grid Computing,2006,4(1):49-69.
- [7] Itchi A I, Ripeanu M, Foster I. Locating Data in (Small-World) Peer-to-Peer Scientific Collaborations [C]//Proc. of the 1st Int'l Workshop on Peer-to-Peer Systems. Cambridge, MA, USA: Springer-Verlag Press, 2002: 232-241.
- [8] Watts D J. Small Worlds-The Dynamics of Networks Between Order and Randomness [J]. ACM SIGMOD Record, 2002, 31(4): 74-75.
- [9] 傅霞玲,许力.“小世界理论”在无线传感器网络的研究现状[J].信息与电脑,2010,4:96-97,99.
- [10] Ranganathan K, Foster I. Identifying dynamic replication strategies for a high-performance data grid [C]//Proceeding of the Second International Workshop on Grid Computing. Denver, November, 2001, 75-86.
- [11] 李志敏,杨义先.哈希函数设计与分析[D].北京:北京邮电大学,2009.

Replica Location Algorithm Approach Based on the Small World in the Internet of Things

GAO Hui, ZHANG Dong-yue, ZHAO Jun-tao, WU Ji-hong

(College of Information, Liaoning University, Shenyang 110036, China)

Abstract: Internet of Things has a large amount of data, heterogeneous node, extensive distribution, etc. Replica can improve data access performance. According to the requirements of fast and efficient replica location for the Internet of things, a replica location method based on the class of small world-the class level and the domain level in the Internet of Things is presented. From the local clustering of data access in the Internet of Things, nodes are divided into the class level and domain level. Given the class-domain data access characteristics, two-way circular message diffusion mechanism to achieve replica of the local and global location is proposed to increase the replica location efficiency.

Key words: Internet of Things; replica location; small world