

基于拓扑关键点保护的 P2P 网络拓扑优化

李 榴¹ 唐九阳² 张 章¹ 肖卫东² 汤大权²

(北方计算中心 北京 100091)¹ (国防科技大学信息系统与管理学院 长沙 410073)²

摘要 网络连通性是对 P2P 网络拓扑结构进行优化的前提。为了确保 P2P 网络中每一个节点的相互连通,增强网络拓扑的抗毁性,应以分布式的方法有效地检测到拓扑关键点并合理地将其消除,以从本质上增强网络应对分割的抵抗力,同时显著地提高系统的容错性。对非结构化 P2P 网络的 CAM 拓扑关键点发现算法进行研究分析,提出 ECAM 拓扑关键点发现算法,以提高探测效率,降低网络消耗。模拟实验结果表明,ECAM 算法能够有效地发现拓扑关键点,提高网络可靠性。

关键词 P2P 覆盖网,拓扑优化,CAM

中图法分类号 TP391 文献标识码 A

Optimization of Peer-to-Peer Overlay Network Topology Based on Topologically-critical Nodes' Protection

LI Liu¹ TANG Jiu-yang² ZHANG Zhang¹ XIAO Wei-dong² TANG Da-quan²

(North Computing Center, Beijing 100091, China)¹

(School of Information System and Management, National University of Defense Technology, Changsha 410073, China)²

Abstract Connectivity of network is the premise to optimize the topology of P2P network. In order to ensure that each node has connected to each other in the P2P network and enhance the survivability of the network topology, an effective distributed method was carried out in the unstructured P2P network which detects the topologically-critical nodes and eliminates them appropriately, it will strengthen the overlay network to defeat the partition and improve the system's fault tolerance significantly. This paper proposed an ECAM topologically-critical nodes discovery algorithm, which improves the CAM. The theoretical analysis and simulation results show that, the ECAM reduces the network consumption and improves the discovery efficiency significantly by maintaining the high accuracy at the same level.

Keywords Peer-to-Peer overlay network, Topology optimization, CAM

1 引言

在 P2P 网络中,节点均离散地分布在物理网络的不同地方,P2P 网络通过在应用层建立的覆盖网将这些节点连接起来。每个节点各自存储与覆盖网中部分节点之间的路由信息,通过节点间的合作转发实现覆盖网中的消息路由,并以此为基础,为丰富多样的网络应用提供支持。为保证应用系统的高可用性,使构建的覆盖网拓扑具有良好的弹性,避免因节点离开造成系统崩溃和服务质量明显下降,通过发现 P2P 覆盖网拓扑的薄弱环节,有针对性地加强和优化,是维护系统稳定可靠的重要手段。

在非结构化 P2P 网络中,虽然在功能上节点互相平等,但实际上,其中的某些节点对于整个网络却有着特殊意义。一方面,这些节点作为两个或多个独立子网之间的唯一通道,对网络拓扑结构有重要影响,其失效将导致整个网络被分割成两个或多个互不连通的网络;另一方面,这些节点的能力影响到系统在动态环境下工作的多方面性能,如查询成功率、节

点定位跳数、通信拥塞和时延等,成为系统性能的瓶颈。学者将这种节点称为拓扑关键点(Topologically-Critical Node, TCN),并对其展开了深入的研究。其研究的核心在于如何发现网络中的拓扑关键点。

发现拓扑关键点的传统方法都是以掌握网络拓扑的全局信息为前提。例如在图论中,为了探测一幅图中的割点,以任意节点为根,建立一个包含所有节点的深度优先搜索(DFS)树,在树中检查每个节点。如果某节点后继的邻居中不包含其祖先,则该节点为一个割点;或是把图中的节点划分成若干个块,则所有连接多个块的节点是割点,但块的划分也需要遍历整个网络,以区分所有点不相交的通路。而对于规模庞大且动态性强的非结构化 P2P 网络,要了解网络拓扑的全局信息是完全不可能的。

非结构化 P2P 网络作为典型的自组织网络,任何节点均无法获得网络全局信息或控制网络拓扑,因此需要采用分布式算法来解决此问题。这就要求不需要任何集中式的控制,每个节点能够独立地运行自己的发现算法,且节点执行该算

到稿日期:2010-09-16 返修日期:2010-12-19 本文受国家自然科学基金(60903225,60903206)资助。

李 榴(1985-),男,硕士,助理工程师,主要研究方向为对等计算、无线传感器网络等,E-mail:2815010@sina.com;唐九阳(1978-),男,博士,副教授,主要研究方向为对等计算、信息集成、知识管理等;张 章(1986-),女,助理工程师,主要研究方向为信息集成、知识管理等;肖卫东(1968-),男,博士,教授,主要研究方向为信息管理、信息系统集成、对等计算;汤大权(1971-),男,博士,副教授,主要研究方向为信息资源管理、信息检索。

法的通信开销低。在去掉网络中拓扑关键点这一关键而脆弱的元素后,即使面临节点不断失效的高动态性环境,系统查询成功率仍然能够保持基本稳定,不会出现急剧下降的情况。而且算法应该充分考虑到节点之间的异构性,优化节点连接,实现负载均衡。

S. Saroiu^[1]等人通过细致的测量来研究非结构化 P2P 网络 Gnutella 的容错性。其实验结果表明,在 Gnutella 中,绝大多数节点连接度很低,只有少数节点连接度非常高。更具体地说,Gnutella 符合 $\alpha=2.3$ 的幂律分布,对随机节点失效有高容错性,但是对恶意攻击的抵抗力则非常弱。他们观察到,在具有 1771 个节点的 Gnutella 这样的 P2P 拓扑环境中,如果随机删除 30% 的节点,剩余的 1300 个节点中有 1106 个节点还是连通的,它们构成的 P2P 网络依旧可以正常工作。但是,如果有选择地删除其中 4% 的节点,P2P 网络被割裂为数百个互不相连的部分。这说明在非结构化 P2P 拓扑中,不同的节点对 P2P 网络的稳定性的影响是有较大区别的。虽然其认识到高连接度的节点对系统容错性的重要意义,但却没有能够进一步回答到底哪部分高连接度节点属于网络的拓扑关键点、哪些节点失效会导致网络分割。

Liu^[2]等人最先从拓扑关键点的角度研究非结构化 P2P 网络的分割问题。他们设计了一种称为 CAM 的分布式算法来检测拓扑关键点。CAM 算法有效地检测到节点间的可达关系,避免了重复的探测工作,将通信开销降得很低。然而,CAM 算法却无法准确检测割点,除非消息路由的跳数 TTL 没有限制。

任浩^[4]从图论的角度去考虑拓扑关键点的定义,将 P2P 网络考虑成一个无向图,拓扑关键点即为无向图的割点。在连通的无向图中,若将割点从图中删除,那么图的连通性将被破坏,导致图中的一些节点不再可达。并提出一种自适应分布式割点发现算法,使节点能够自主判断自己是否为割点,并进一步提出一种主动式割点发现算法来验证与弥补自适应发现算法的缺漏。该算法发现准确率高,但实现较为复杂。

李振华^[5]等人鉴于非结构化 P2P 网络的特点,认为割点不能充分、合理地描述非结构化 P2P 网络的拓扑关键点。因为割点是全局的、静止的,而非结构化 P2P 网络则是在动态环境下分布式工作、局部性路由的,既不可能静止地获取网络全局状态,也不可能做全局性的定位。所以从跳数 TTL 的有限范围考虑,提出分点的概念,即认为分点既是其邻居子集间互相连接的唯一通道,也是邻居子集间互相通信的性能瓶颈。并由此对 CAM 算法进行了改进,将其应用于分点的检测与避免工作中,使系统对网络分割的抵抗力得到本质的增强。分点避免后,系统查询成功率有所上升,在高动态性网络环境下系统的容错性得到明显提高。

2 基于扩展式 CAM 的拓扑关键点发现算法

2.1 CAM 算法

Liu 等人提出的 CAM 算法是最为经典的拓扑关键点发现算法。该算法基于泛洪的消息传递方式,通过 TTL 的控制,能够以被探测节点(候选节点)为中心,局部范围内获取网络拓扑,有效地判断候选节点是否为拓扑关键点,如图 1 所示。CAM 算法的核心思想为:在候选节点 C 不参与消息路由的前提下,探测其邻居节点间的可达关系,再依据可达关系

将邻居节点分成互不可达的子集。若子集只有一个,那么 C 不是拓扑关键点,否则 C 是拓扑关键点。

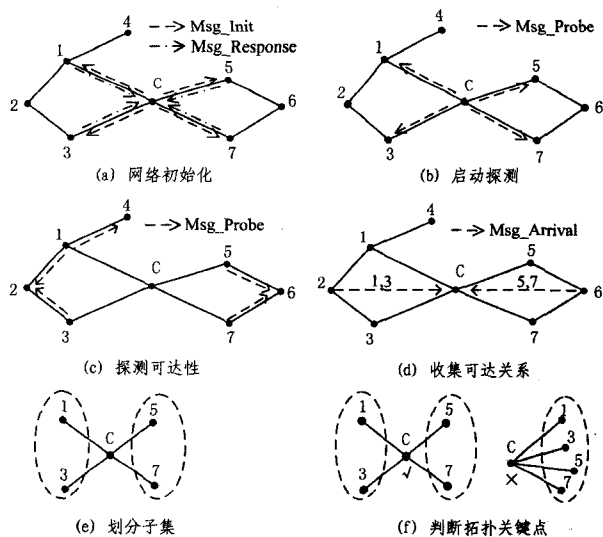


图 1 CAM 算法发现拓扑关键点的过程

CAM 算法是一种简单、有效、分布式的拓扑关键点发现算法,发现准确率非常高。但其基于泛洪的消息传递方式,每次发送的消息数量呈级数增长,网络消耗非常大。若某一时刻所有节点同时发起探测行为,则网络中将出现大量的探测消息,引起网络拥塞,严重时甚至会导致网络瘫痪。而且各候选节点的探测行为相互完全独立,之间并不进行信息的交互,由此带来的局限性是:只有当一个候选节点在整个探测行为结束之后,才能够获得完整的探测结果。实际上,如果能够在各个候选节点的探测行为之间建立起一种信息的交互机制,就不需要等到一个完整的探测行为结束,而只在中间的某个步骤就可以获得完整的结果,从而做出是否为拓扑关键点的判断。这样能够大大减少数据的泛洪投递次数,提高发现效率,降低网络消耗。

2.2 基于扩展式 CAM 的拓扑关键点发现算法

如图 2(a) 所示,在节点 C 完成拓扑关键点的探测后,节点 1 作为候选节点,也开始发起一轮新的探测行为。按照 CAM 算法,它所发出的 Msg_Probe 一直要传递到节点 3,才能由节点 3 返回一个“节点 2 和 C 相互可达”的 Msg_Arrival 消息。然而,在之前节点 C 的探测过程中,由于节点 2 接收到了节点 C 的消息,因此,节点 2 早已知道“节点 1、2 和 C 是相互可达的”,同理节点 C 也掌握了这些信息。因此,若节点 1 能够利用这些有价值的信息,在网络初始化阶段,通过 Msg_Init 和 Msg_Response 消息就可以收集到“节点 2 与节点 C 相互可达”这条信息(如图 2(b) 所示),这就大大加快了探测过程,减轻了通信开销。

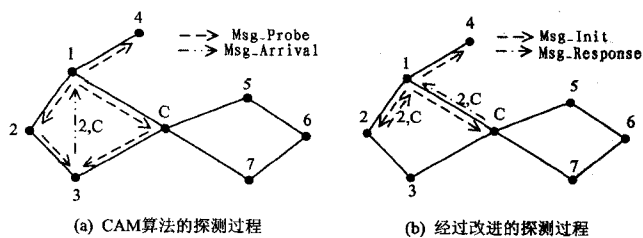


图 2 节点 1 可以从节点 2 与节点 C 直接得到连通信息

那么,该如何保存这些有价值的信息呢?考虑对 CAM 算法中所设计的 CAM 表进行修改。对于一个候选节点的探测过程,CAM 表只存在于除它之外的其他邻接节点中,候选节点本身并不保存 CAM 表。因此提出让候选节点本身也维护一张 CAM 表,即对于一个候选节点发起的探测过程,它及其所有涉及的节点的缓存中都会建立一张 CAM 表,称为 ECAM 表,它的结构设计如图 3 所示。

Target Node	Connected Nodes List
C	N_1, N_3

图 3 ECAM 表

ECAM 表中每一条数据表示 Connected Nodes List 中的节点与 Target Node 是相互可达的。ECAM 表定义为描述节点之间的相互可达关系的表。例如图 3 中的 ECAM 表描述了节点 C、节点 N_1 和节点 N_3 相互可达。对于邻居节点,ECAM 表的维护方式与 CAM 表相同,其中 Target Node 记录的是候选节点;Connected Nodes List 记录的是收到的 Msg_Probe 消息的新的邻居节点,候选节点及其邻居节点是必然相互可达的。对于候选节点,ECAM 表记录的是其收到的 Msg_Arrival 消息的情况,其中 Target Nodes 记录的是 Msg_Arrival 消息的发送节点;Connected Nodes List 记录的是 Msg_Arrival 消息中的邻居节点。根据 CAM 机制,发送节点与邻居节点是必然相互可达的。

ECAM 表中的数据并不是保持不变的,而是随着网络拓扑结构的变化而不断维护与更新。由 ECAM 表的定义可知,保持该示例表的节点必定与 N_1, N_3 连通。为验证 Target Node 为 C 的记录是否依然正确,即节点 C 是否仍然与节点 N_1 和 N_3 连通,设计针对 ECAM 表的维护机制:假设该节点为候选拓扑关键点,发起一次探测行为,依据探测结果便可验证该记录是否正确。若不正确,则将该记录从 ECAM 表中删除。

基于 ECAM 表的定义,ECAM 算法步骤如图 4 所示。

从图 4 可以看出,ECAM 算法对 CAM 算法的改进在于步骤(b)与步骤(d)对历史信息利用。

在步骤(b)中,通过节点所保存的已有的 ECAM 信息,可以迅速获取到候选节点的邻居之间的可达关系,而不必进行泛洪探测。各节点所保存的 ECAM 信息越丰富,则越有可能直接在步骤(b)获得全部可达关系的信息,从而立刻终止探测行为,直接转入步骤(f)划分子集,从而极大地节省了发现时间,提高了发现效率。步骤(b)改进的效果直接取决于节点 ECAM 信息的丰富程度。节点已有的 ECAM 信息是该节点在参与其他节点探测行为过程中保存下来的。考虑 P2P 网络的动态性与随机性,则认定网络中的候选节点不会同时发起探测行为,因此当稍迟的候选节点发起探测行为时,在探测范围中的部分节点上必定已经保存有有价值的历史 ECAM 数据,由此让越迟开始探测的候选节点越受益。故步骤(b)的改进必定能够带来发现效率的提高。

在步骤(d)中同样利用了节点已保存的 ECAM 数据来实现改进。该节点若能在收到 Msg_Probe 消息后,通过 ECAM 得到有价值的信息,就可以在该节点终止 Msg_Probe 消息向下的泛洪传递。而按照 CAM 算法的设计,若 TTL 不为零,则必须将 Msg_Probe 消息继续以泛洪的方式继续传递给它的邻居。通过步骤(d)的改进,忽略了 TTL 的控制,以提前结

束探测来实现泛洪次数的降低,在提高发现效率的同时,也大大降低了网络消耗。同样地,步骤(d)改进的效果也取决于节点 ECAM 信息的丰富程度。由上分析,可知在实际 P2P 网络中,步骤(d)的改进也必定能够带来发现效率的提高。

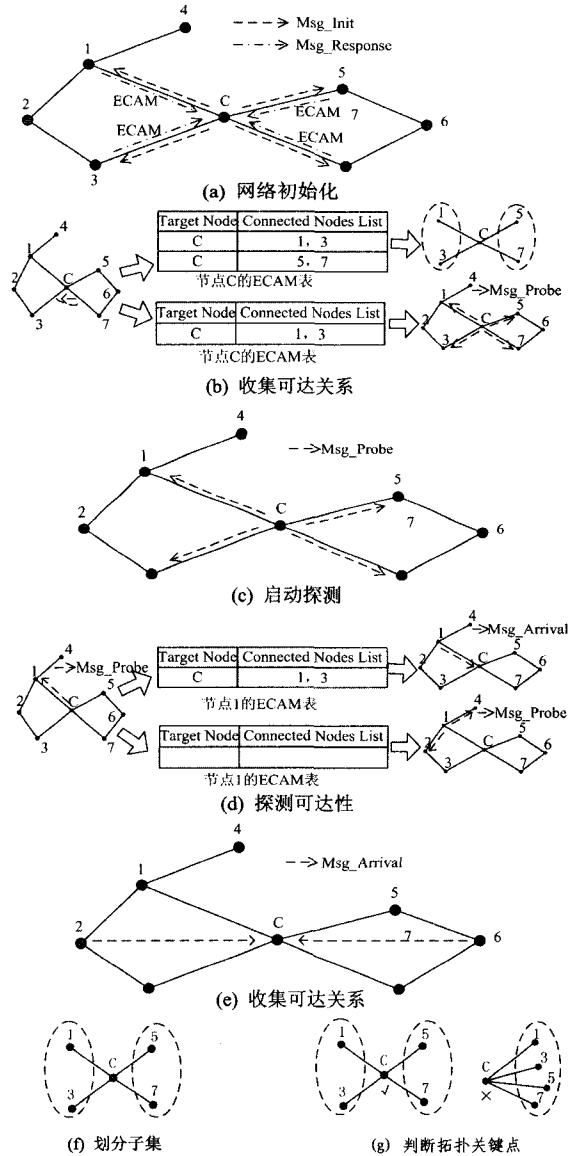


图 4 ECAM 算法发现拓扑关键点的过程

2.3 算法比较

2.3.1 发现准确率

拓扑关键点发现算法的准确率是指在作出判断的结果集合中真实的拓扑关键点所占的比率。对于 CAM 算法与 ECAM 算法,因为拓扑关键点的发现受到探测消息跳数 TTL 的限制,会有部分非拓扑关键点被误判为拓扑关键点,或者拓扑关键点被误判为非拓扑关键点。

假设网络拓扑是静态不变的,且泛洪消息的 TTL 初始值为无穷大,那么只要在两个节点之间有通路,这条通路必然能够被泛洪搜索所发现。因此,只要提供充足的时间,网络中所有可能的路径都会被泛洪搜索消息所遍历,这保证了算法判断的基础,即获得节点邻居间的所有路径信息是能够实现的,因此算法的正确性可以保证。我们知道,如果能够对网络拓扑进行有效控制,让网络中的节点均匀分布,则搜索从一个

节点到任意另外一个节点的平均距离为 $O(\ln N)$, 这里 N 表示网络的规模。因此, 任意两个节点之间的路径长度通常是在网络规模的对数级内, 即采用较小 TTL 值已经能够获得足够的准确率。

从另外一个方面看, 由于缺少全局知识, 可能有很少的一部分节点会被错误地判定成拓扑关键点, 主要原因是它们位于某些非常长的圈上。如果采用有限的 TTL, 这样的圈就不会被发现。然而, 在 P2P 网络中这样的圈非常少, 而即使这些相关的节点都被误判为拓扑关键点, 则产生的唯一问题仅仅是拓扑关键点消除算法会因此在网络中多建立几条连接, 这对网络不会造成任何重大损害。

对于 CAM 算法, 判断的依据来自于实时收集的邻居节点互通的 CAM 数据。而 ECAM 算法判断的依据来自于两部分: 实时收集的 CAM 数据和节点保存的 ECAM 数据。因此, ECAM 算法的准确性取决于节点保存的 ECAM 数据的正确性。由 ECAM 算法可知, 其准确率最低与 CAM 算法保持相同。随着节点所保存的 ECAM 数据的所占比率越大, 其算法准确率将会不断提高。当所占比率达到 100% 时, 算法的发现效率达到最高, 此时 ECAM 数据的正确性直接决定着算法的准确性。根据 ECAM 算法的数据维护机制, ECAM 数据的正确性由维护机制启动频率所决定。

设 ECAM 数据所占比率为 p , ECAM 数据维护频率为 f , 则 ECAM 算法的发现准确率 E_{ECAM} 计算公式如下:

$$E_{ECAM} = (1-p)E_{CAM} + pQf$$

式中, Q 为 ECAM 数据的正确因子。若要 E_{ECAM} 与 E_{CAM} 保持相同的发现水准, 则 E_{ECAM} 的下浮应该保持在 -5% 之内, 即要保证 $Qf \geq 0.95E_{CAM}$ 。在第 3 节的模拟实验中将证明该公式成立, 即证明 ECAM 算法能够保持 CAM 算法的发现准确率不变。当然, 在 P2P 网络动态性环境下, 在泛洪探测过程中, 由于探测消息步长 TTL 的限制, 或者某些节点在拓扑关键点发现之前离开网络, 使得节点不能获得足够的消息, 从而影响最终的准确率。

2.3.2 通信开销

网络通信开销是指节点通讯所消耗的网络带宽和其它相关参数的一个函数值。这是网络管理员最为关心的一个参数, 因为它将对网络的服务性能产生重大的影响。极大的网络通信开销将严重限制 P2P 系统的可扩展性, 同时也是网络管理员禁止 P2P 系统在网络中运行的原因之一。

基于当前计算机能力, 节点在本地的计算开销要远远低于算法产生的通信开销, 因此只对通信开销进行评估。假设非结构化 P2P 网络中共有 n 个节点, 平均节点度为 c , 跳数限制 TTL 为 T , 由于采用泛洪探测模式, 探测产生的消息数量要远远多于应答消息和请求建立连接消息的数量, 因此整个算法产生的通信开销中探测消息占绝大多数。

对于 CAM 算法, 在候选节点探测过程中每个节点只会向外发送或转发一次探测消息, 在距候选节点 TTL 长度范围内的所有连接只会传送一次探测消息。因此, 对于一个候选节点来说, 它产生的探测消息数约为 cT , 再加上网络初始化的 Msg_Init 消息和 $Msg_Response$ 消息, 则全部节点进行探测的通信开销为 $O(nc(T+2))$ 。从另外一个角度出发, 图中的每条边都对应了两个点, 则图中所有的边的数量是 $\frac{nc}{2}$, 每

条边最多传送一次消息, 因此即使每次探测都传遍图中的所有边(如果网络规模比较小, 而 TTL 的初始值比较高, 就会产生这种情况), 全部节点都进行一次探测, 总的通信开销的上限为 $\frac{nc}{2}$ 。综合这两个计算结果, CAM 算法产生的通信开销

为 $\min(O(nc(T+2)), O(\frac{nc}{2}))$ 。对于真正的大规模 P2P 网络来说, c 的值往往远远小于网络的规模 n 。如果为了控制探测的通信开销而把 TTL 值限定在一定范围内, 普通情况下是 $cT < n$, 因此我们认为 CAM 算法的通信开销 S_{CAM} 为:

$$S_{CAM} = O(nc(T+2))$$

对于 ECAM 算法, 最好的情况是候选节点仅通过一次网络初始化操作便能够获得所有信息而结束探测, 此时全部节点进行探测的通信开销为 $O(2nc)$, 设其概率为 $p_1(t)$, 其中 t 为时间参数。在第 k 次探测时获得所有信息而结束探测, 通信开销为 $O(nc(k+2))$, 设其概率为 $p_2(t)$ 。因此最坏的情况则是一个完整的 CAM 算法过程, 此时通信开销为 $O(nc(T+2))$, 其概率为 $1-p_1(t)-p_2(t)$, 可以得出 ECAM 算法的通信开销公式 S_{ECAM} 为:

$$S_{ECAM} = O(nc[2p_1 + (k+2)p_2 + (T+2)(1-p_1-p_2)])$$

化简后为:

$$S_{ECAM} = O(nc(T+2-p_1T+p_2k-p_2T))$$

因为 $k \leq T$, 则

$$S_{ECAM} = O(nc(T+2-p_1T+p_2k-p_2T))$$

$$\leq O(nc(T+2-p_1T))$$

$$\leq O(nc(T+2)) = S_{CAM}$$

即

$$S_{ECAM} \leq S_{CAM}$$

由此可证 ECAM 算法的通信开销要小于 CAM 的通信开销, 且 $p_1(t)$ 与 $p_2(t)$ 越大, 通信开销就越小。

3 模拟实验

仿真实验在 PC 机上完成。PC 机的 CPU 为 Intel Core2 Quad 2.5GHz, 内存为 2G, 硬盘为 320G, 操作系统为 Windows XP SP3。仿真平台选用 Peersim, 仿真程序由 JAVA 编写。使用 WireScaleFreeDM 算法来实现 4 个大规模非结构化 P2P 实验网络的生成, 其参数如表 1 所列。

表 1 4 个实验网络的参数

实验网络名称	节点数	节点平均度
Net 1	5000	5
Net 2	5000	10
Net 3	10000	5
Net 4	10000	10

3.1 发现算法的准确率

CAM 发现算法的准确率主要由消息跳数 TTL 决定, TTL 越大, 则发现算法的探测范围就越大, 所掌握的判断信息就越准确, 对拓扑关键点的发现准确率就越高。但当 TTL 增大到一定程度, 则所掌握的信息趋于一致, 发现准确率将不再有大的变化。而 ECAM 算法的准确率则由 ECAM 数据的维护频率 f 决定, f 越大则维护节点的 ECAM 数据的频率就越高, ECAM 数据的正确率就越大, 那么 ECAM 算法的准确率就越高。

本小节仿真实验以维护频率 f 为核心, 设定 3 种网络状态: 一是网络无节点失效下, 维护频率 $f=0.2s$; 二是网络以

每秒失效 100 个普通节点的频率衰减下,维护频率 $f=0.2s$;三是网络以每秒失效 100 个普通节点的频率衰减下,维护频率 $f=0.1s$ 。由此来验证 CAM 算法与 ECAM 算法面对 3 种复杂网络情况在不同 TTL 下的发现准确率。

(1)无节点失效,网络维护频率 $f=0.2s$

由图 5 可以看出,ECAM 算法与 CAM 算法的准确率基本相同。在部分网络中由于节点连接度较大,ECAM 的准确率会有 1% 以内的降低。由于网络状况稳定,无节点失效,因此两种算法都能够正常稳定地执行完毕,准确率自然也基本相同。且 TTL 越大其准确率越高,这是由于 CAM 算法和 ECAM 算法在高 TTL 下获取了更多的连通信息。

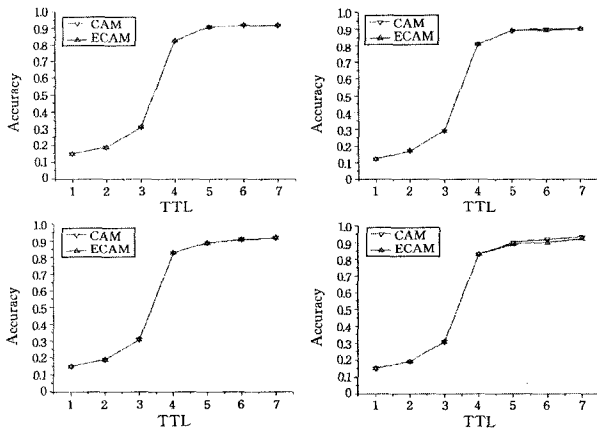


图 5 无节点失效、 $f=0.2s$ 的发现准确率

(2)每 1 秒失效 100 个普通节点,网络维护频率 $f=0.2s$

由图 6 可以看出,CAM 算法与 ECAM 算法的准确率均有下降,这是因为普通节点失效给算法造成了部分消息没有回应,以致候选节点在划分子集时判断失误。而 ECAM 算法由于维护频率低于节点失效频率,因此准确率比 CAM 算法稍低,但 ECAM 算法的准确率始终保持在 CAM 算法准确率的 -0.5% 内,验证了理论分析中 $Qf \geq 0.95E_{CAM}$ 公式成立,即 ECAM 算法的准确率基本保持与 CAM 算法相同的水准。

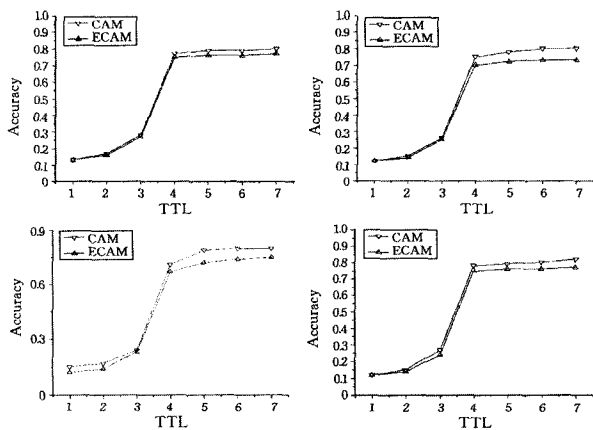


图 6 每 1 秒失效 100 个普通节点、 $f=0.2s$ 的发现准确率

(3)每 1 秒失效 100 个普通节点,网络维护频率 $f=0.1s$

由图 7 可以看出,维护频率提高 1 倍后,与节点失效频率相同,此时 ECAM 算法的准确率有了大幅度的提高,甚至高于 CAM 算法。这正是由 ECAM 算法的维护机制带来的优势,它能够及时弥补因节点失效而造成的无响应状况。而 CAM 算法没有此类维护机制,因此准确率受到了很大的影响。

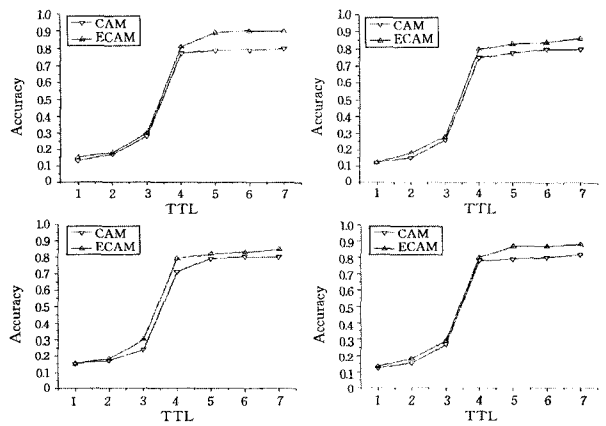


图 7 每 1 秒失效 100 个普通节点、 $f=0.1s$ 的发现准确率

通过以上 3 种情况的仿真实验,可以认定 ECAM 算法能够保证其准确率与 CAM 算法处于同一水准上。甚至当面对高度动态的网络状况时,若 ECAM 算法的维护频率能够保持一个较高的值,它的发现准确率甚至比 CAM 算法要高。也就是说,ECAM 算法能够比 CAM 更胜任复杂且高度动态的 P2P 网络环境。

3.2 发现算法的通信开销

ECAM 算法的改进依据在于节点保存的 ECAM 数据对后发节点的帮助,则节点开启发现算法的间隔时间 d 是一个非常重要的参数。当 d 值为零时,说明所有节点同时开启发现算法。 d 值越大,节点之间开启发现算法的间隔越大,节点保存的 ECAM 信息就越多,对后发节点的帮助就越大。因此在本小节的仿真实验中着重考虑 d 值对 ECAM 算法准确率的影响。由此仿真实验设计两种情况:一是所有节点同时开启发现算法;二是节点以 $0.01s$ 的间隔依次开启发现算法。在两种情况下考虑不同 TTL 对两种算法通信开销的影响。

(1)所有节点同时开启发现算法

由图 8 可以看出,ECAM 算法通信开销与 CAM 算法相比略有降低,这是因为其中部分节点所保存的 ECAM 数据对其他节点起到了帮助作用。因此,纵使所有节点同时开启发现算法,ECAM 算法也能够实现降低通信开销的效果。

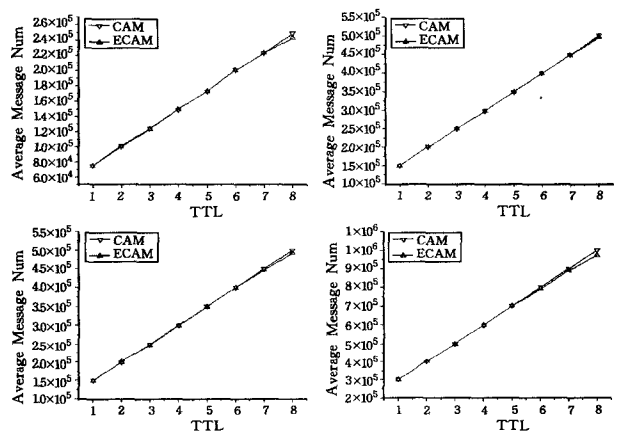


图 8 所有节点同时开启发现算法时的通信开销

(2)节点以 $0.01s$ 的间隔依次开启发现算法

由图 9 可以看出,ECAM 算法的通信开销大幅度降低,在最好情况下 ECAM 算法的通信开销甚至能够比 CAM 算

(下转第 68 页)

比较,即 Alice 用值 u 与 Bob 的 $v_4 + \varphi$ 进行比较,如果 $u \leq v_4 + \varphi$ 则目标定位比较准确,否则准确度有待提高。

结束语 不同机构或部门合作完成某项测绘作业时经常需要对各自的坐标系进行转换,为了保证合作双方的安全与利益,双方都不希望泄露各自的输入信息。本文首次提出并设计了保护私有信息的坐标系变换协议,并在该协议的基础上设计了保护私有信息不同坐标系两点距离计算协议。解决了在不向对方泄露自身私有信息的情况下计算不同坐标系两点之间的距离问题,并通过秘密比较得出两点距离与某个设定阈值 ϕ 的大小关系,从而解决了对目标定位准确性的判定问题。

参考文献

[1] 刘宗泉,贾志强,邢诚,等. GPS网 WGS-84 平差坐标向地方独立坐标转换[J]. 测绘信息与工程,2007,32(1):33-35
 [2] 雷伟伟,姜斌. 国家坐标系与城市坐标系转换方法的探讨[J]. 测绘科学,2010,35(1):22-23
 [3] Yao A C. Protocols for secure computation[C]//Proc. of the 23rd IEEE Symp. on Foundation of Computer Science. Chicago: IEEE Computer Society,1982:160-164
 [4] Golderich O, Micali S, Wigderson A. How to play any mental game[C]//Proceedings of the 19th Annual ACM Symposium on Theory of Computing. New York,1987:218-229
 [5] Goldreich O. Foundations of cryptography, basic applications

[M]. Cambridge:Cambridge University Press,2004:233-278

[6] Du W L, Atallah M J. Privacy-preserving cooperative scientific computations [A]//Proceedings of the 14th IEEE Computer Security Workshop[C]. Nova Scotia, Canada: IEEE Computer Society Press,2001:273-82
 [7] Mikhail J, Atallah, Du W L. Secure multi-party computational geometry[C]//Lecture Notes in Computer Science 2125. Berlin: Springer,2001:165-179
 [8] Lindell Y, Pinkas B. Privacy preserving data mining [J]. Journal of Cryptology,2002,15(3):177-206
 [9] 罗永龙,黄刘生. 空间几何对象相对位置判定中的私有信息保护[J]. 计算机研究与发展,2006,43(3):410-416
 [10] Luo Yong-long, Huang Liu-sheng, Zhong Hong. Secure two-party point-circle inclusion problem [J]. Journal of Computer Science and Technology,2007,22(1):88-91
 [11] Paillier P. Public-key cryptosystems based on composite degree residuosity classes [A] // Advances in Cryptology—EUROCRYPT'99, Lecture Notes in Computer Science 1592 [C]. Springer-Verlag,1999:223-238
 [12] Cachin C. Efficient private bidding and auctions with an oblivious third party[C]//Proc. of the 6th ACM Conf. on Computer and Communications Security. Assn for Computing Machinery, 1999:120-127
 [13] 秦静,张振峰,冯登国,等. 无信息泄漏的比较协议[J]. 软件学报,2004,15(3):421-427

(上接第 57 页)

法降低 $\frac{1}{2}$, 经统计 ECAM 算法的通信开销能够比 CAM 算法节省 35%~48%。对于实际的大规模非结构化 P2P 网络,其节点开启发现算法的方式基本是间隔开启的状态,因此 ECAM 算法对通信开销的降低是非常有实际意义的,对提升 P2P 网络的可靠性与可用性有着很重要的帮助。

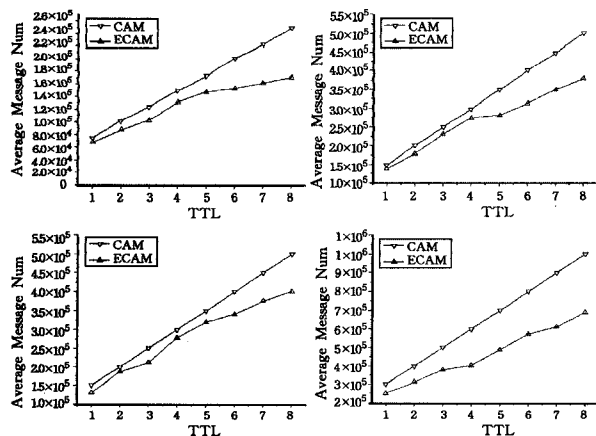


图9 节点以 0.01s 的间隔依次开启发现算法时的通信开销

结束语 本文对非结构化 P2P 网络下的拓扑优化进行了研究,从拓扑网络中最薄弱的环节——拓扑关键点出发,深入探讨了拓扑关键点的发现算法与消除算法,为保证网络的可靠连通提供了有效手段。总结了非结构化 P2P 网络拓扑优化的研究现状与发展趋势,综述了拓扑关键点的相关技术和应用,分析了众多研究中存在的不足和主要问题。对 CAM

拓扑关键点发现算法进行仿真分析,针对其存在的发现效率低、网络消耗大的问题,提出基于扩展式 CAM 的拓扑关键点发现算法 ECAM,并对其进行了模拟仿真。实验结果表明,ECAM 算法在保证发现准确率的同时,大大降低了网络消耗,提高了发现效率。且本算法可应用到实际网络中,与拓扑关键点的消除算法相结合,提高网络可靠性。

参考文献

[1] Saroiu S, Gummadi P, Gribble S. Measuring and analyzing the characteristics of napster and gnutella hosts [M]. Multimedia Systems 9, Berlin: Springer-Verlag,2003:170-184
 [2] Liu X, Xiao L, Kreling A, et al. Optimizing overlay topology by reducing cut vertices[C]//Proc. of the ACM Int'l Workshop on Network and Operating System Support for Digital Audio and Video (NOSSDAV). Newport: ACM Special Interest Group on Multimedia. <http://portal.acm.org/citation.cfm?id=1378213&jmp=references&coll=ACM&dl=ACM>,2006
 [3] Chawathe Y, Ratnasamy S, Breslau L, et al. Making gnutella-like P2P systems scalable[C]//Proc. of the Annual Conf. of the Special Interest Group on Data Communication (SIGCOMM). Karlsruhe: ACM Special Interest Group on Data Communication,2003:407-418
 [4] 任浩. P2P 覆盖网拓扑优化研究[D]. 长沙:国防科技大学,2007
 [5] 李振华,陈贵海,邱彤庆. 分点:无结构对等网络的拓扑关键点[J]. 软件学报,2008(9):2376-2388
 [6] 冯国富,张金城,姜玉泉,等. 无结构 P2P 覆盖网络的拓扑优化[J]. 软件学报,2007(11):2819-2829