

一种基于语义相似的 P2P 搜索机制

曹迎春 王义麟 陈建忠 陆桑璐 陈道蓄

(南京大学计算机软件新技术国家重点实验室 南京大学计算机科学与技术系 南京 210093)

摘 要 现有 P2P 网络规模大、动态性高、异构性强,有效的搜索技术一直是 P2P 系统研究中的核心问题。本文针对无结构 P2P 网络泛洪搜索机制的盲目性所导致的查询开销大、效率低的问题,提出了一种基于语义相似的 P2P 搜索机制 SRVN,通过积累历史搜索经验获得路由指引信息,帮助搜索结点快速发现与搜索内容最相关的结点,从而提高搜索效率和目标命中率。实验结果显示,使用 SRVN 搜索机制,有效地提高了 Gnutella 查询性能。

关键词 P2P, 语义路由, 虚结点, Gnutella

Similarity-based Semantics Searching in Peer-to-Peer Networks

CAO Ying-Chun WANG Yi-Lin CHEN Jian-Zhong LU Sang-Lu CHEN Dao-Xu

(State Key Laboratory for Novel Software Technology, Department of Computer Science and Technology, Nanjing 210093)

Abstract Due to the increasing size of P2P systems and their characteristic of being dynamic and heterogeneous, effective searching techniques have become an important research issue and received much attention recently. On studying the key facts in unstructured P2P searching technologies and analyzing their limitations, this thesis proposes a similarity-based semantics searching technology. It gets routing information according to searching history to help recover the target which is most related to the searching content, thus improving the efficiency of searching and the hit rate. Simulation results prove the effectiveness and efficiency of the proposed searching mechanism.

Keywords Peer-to-Peer, Semantic routing, Virtual node, Gnutella

1 引言

P2P 网络中存在着丰富的信息资源,正确高效地搜索信息是 P2P 系统实现有效的资源共享和协同的关键。当前,应用较广的 P2P 系统大多为无结构 P2P 网络,这是由于其结构较为简单,容易实现,能提供较为灵活的搜索方式。然而,在无结构 P2P 网络中,由于缺乏可被利用的拓扑结构信息,因此信息搜索的算法往往带有一定的盲目性。这类网络大多采用泛洪搜索或相应的改进搜索技术。为提高搜索的成功率,泛洪方式通常需要涉及较大的搜索空间,消息量会随着搜索范围的扩大而呈指数增长,造成网络流量急剧增加,从而影响网络性能,导致 P2P 网络的可扩展性降低^[1]。此外,对于稀疏资源的搜索不能避免产生大量的查询信息,而且搜索结果不可知。

目前,针对无结构 P2P 系统搜索机制的研究主要是通过控制和减少由于泛洪搜索而带来的大量不必要流量。一个研究趋势是考虑结点和数据之间的相关性,利用结点间的语义相似性,建立结点间的快捷连接,以提高搜索性能。本文针对 P2P 网络的搜索机制进行了研究,分析了现有 P2P 搜索机制的不足,并提出了一种基于语义相似的 P2P 搜索机制 SRVN,根据网络资源访问的局部性原理,应用相似度计算理论分析相似搜索,以此指引路由,并根据一定的策略存放文件副本信息、更新搜索经验,从而进一步提高全局搜索性能。实验结果显示,使用 SRVN 搜索机制,可以有效提高查询效率和查询命中率。

本文第 2 部分讨论相关工作,分析现有 P2P 搜索机制存在的不足。第 3 部分提出基于语义相似的 P2P 搜索机制 SRVN,并着重介绍 3 个关键技术:虚结点的划分、路由指引信息的获取和维护,以及副本放置策略。第 4 部分给出实验

系统并分析实验结果。最后总结全文。

2 相关工作

针对无结构 P2P 搜索机制中存在的问题,已经有多种改进的算法,主要可以分为两类:路由机制的改进、基于缓存的改进。路由机制的改进方法试图通过改进泛洪搜索机制,不是将查询发送给所有的邻居,而是有选择地发送到一些邻居,尽量减少查询消息、减少被访问的结点,同时又能够得到满意的查询结果。如改进的 BFS(Modified BFS)^[2]、有向 BFS(Directed BFS)^[3]、迭代加深技术(Iterative Deepening)^[3]、随机漫步法(Random Walker)^[4]、范围扩张(Expanding Ring)^[4]等,它们在一定程度上减少了通信量,提高了搜索效率。基于缓存的改进主要有内容缓存^[5]和索引缓存^[3]。内容缓存将文件的内容复制到结点上,以减少某些热门结点的负载,减少响应时间。索引缓存通常在结点上保存一定半径以内的结点资源元数据索引,当该结点接收到查询消息时,可以代表该半径内的结点处理相应搜索请求,这样可以在跳数不增加的情况下扩大搜索范围,既能返回搜索结果,又大大降低了搜索代价。这些改进算法,在一定程度上控制了消息量,降低了网络负载。但是,这些算法大多只是利用局部信息进行一定程度的改进以降低搜索的盲目性,在搜索性能上的提高较为有限。

实际上,结点和数据之间存在着一定的相关性,结点上的共享文件在一定程度上代表了结点的兴趣偏好。此外,搜索请求往往也与其历史搜索有较大的相关性。如果某结点曾经对相似的历史搜索作出响应,那么该结点也很可能对当前搜索作出响应。文[11]对 P2P 网络中文件聚集程度的测量结果给出了同样的结论:在自己以往下载或得到搜索结果的结点上更容易找到自己需要的信息(局部性现象)。因此,我们考虑可以通过挖掘结点与数据间的相关性,并通过考察一段

时期内历史搜索记录来获得更为有效的路由指引信息,使得路由尽可能指向最有可能提供相关信息的结点。

NeuroGrid^[6]、Alpine^[7]、APPN^[8]、GS^[9]以及Buddy Web^[10]等对算法进行了改进,体现了这种相关性。它们的特点是基于结点的语义相似性,建立结点间的快捷连接。文件的查询请求被路由到最可能的物理位置,其性能与基于分布式散列函数的定位算法性能基本相当。但这些算法分别存在如下问题:过于依赖用户对于规则的定义,而用户对规则定义的不一致性使得搜索算法的有效性受到影响;采用了复杂的语义相似度计算方法而导致在分布式环境中计算开销过大;采用被动的学习方式,使得收敛速度过慢,等等。

针对这些问题,本文提出的SRVN(Semantic Routing Based on Virtual Node,简称SRVN)利用了P2P网络的拓扑特性,以及网络资源访问的局部性原理,各结点在查找过程中积累“有用的”搜索经验以更有效地指引路由,使得路由总是指向能提供更多信息的邻近结点。同时,在应答消息回路经过的各跳结点上,可以根据一定的策略存放文件副本信息,更新搜索经验,从而进一步提高全局搜索性能。此外,SRVN提供了一个附功能,当用户提交查询请求后,通过该机制可以得到一组相关的信息(文件),从而扩大了搜索的范围。

3 基于语义相似的P2P搜索机制

由于结点和数据之间存在着一定的相关性,结点上的共享文件在一定程度上代表了结点的兴趣偏好,SRVN搜索机制将每个P2P结点按照其上共享数据的不同语义,逻辑划分成若干个具有不同语义的虚结点。并且在P2P系统的运行过程中,不断演化,使得具有相似语义的虚结点逐渐“聚集”在一起,从而形成语义路由网络(Semantic Routing Peer-to-Peer Network,简称SRPNet),如图1所示。

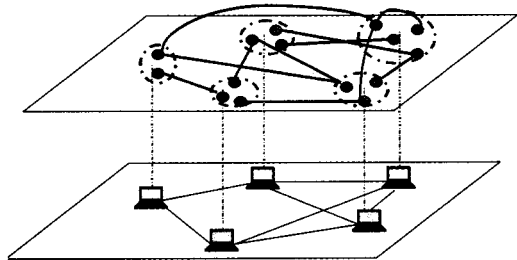


图1 SRPNet网络结构示意图

图1中,逻辑层上的各个虚线框对应着物理层上的每个结点,表示该结点对应的虚结点集。颜色相同的虚结点表示存放着同类信息(文件)。虚结点间经过演化,相互建立联系,形成覆盖网。

这个过程主要涉及到两个关键问题:①虚结点的划分;②路由指引信息的获取和维护(包括副本放置策略)。

3.1 虚结点的划分

将一个物理结点划分成若干个逻辑结点,每个逻辑结点包含不同的语义信息,称这样的逻辑结点为虚结点(Virtual Node,简称VN)。因为结点本身存放的文件一定程度上代表了结点的兴趣偏好,所以需要首先确定结点上共享文件的语义类别。

每个结点上有用户界面、虚结点集、元数据库、语义分析模块和管理模块,如图2所示。

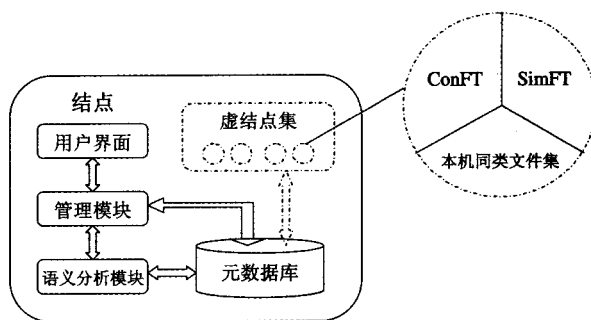


图2 结点示意图

结点上存放的共享文件存在一定的语义差异,将它们进行分类,同一类文件可理解为存放在同一个虚结点上的共享文件。虚结点的语义属性即为该类文件的中心向量。规定结点上共享文件之间的关系是:在同一个虚结点上的文件之间语义相似度高于不同虚结点上的文件之间的语义相似度。我们采用简单的空间向量模型(VSM)来表示文本,设置了两个阈值: ϵ 和 ϵ' , $\epsilon < \epsilon'$,可以计算出两个文件之间的语义距离,用 d 来表示。定义两个文件之间的关系如下:

(1)当 $d \leq \epsilon$ 时,两个文件属于同一类,称它们为同类文件;

(2)当 $\epsilon < d \leq \epsilon'$ 时,表示这两个文件语义相似,但不属于同一类别,称它们为相关文件;

(3)当 $d > \epsilon'$ 时,认为两个文件没有相关性。

在虚结点的共享存储空间中保存了两张表,一张表用于缓存历史搜索中的同类文件数据以及原始宿主ID,称为ConFT,属性是 $\langle \text{OID}, \text{NID}, \text{URL}, \text{VNID}, \text{VECT}, \text{data size}, \text{data} \rangle$,其中OID表示文件ID,NID表示结点ID,URL表示结点的位置,VNID表示文件所在虚结点ID,VECT指文件的向量描述。另一张表用于缓存历史搜索中相关文件的三元组和宿主ID,称为SimFT,属性是 $\langle \text{OID}, \text{NID}, \text{URL}, \text{VNID}, \text{VECT} \rangle$ 。这两张表的数据都来自历史搜索。通过这两组信息来指导搜索请求的路由指向,更有可能存放目标文件的结点。

通过用户界面,用户可以管理和维护本地文件、设置文件共享范围及其共享权限、产生和提交查询。虚结点集是一个虚拟的集合,由本机所有虚结点构成,结点上所有的共享文件都划分到各个虚结点。元数据库中存放本地所有虚结点的语义属性、虚结点上文件的索引、文件语义向量、文件的共享权限、ConFT和SimFT表、分类规则等信息。语义分析模块主要功能是计算语义相似度,分析两个文件的关系。它从管理模块那里获得从消息中获得的文件语义向量,通过计算得出它和元数据库中虚结点或相关文件的关系,然后将结果交给管理模块进一步处理。管理模块主要用于管理邻居结点、相关结点的NID、URL,以及元数据库等。通过管理模块迅速判断本机是否有目标文件,或者在虚结点上是否保存有目标文件的记录信息;对虚结点上ConFT、SimFT的构造和维护也由管理模块来完成。一旦收到查询,管理模块进行预处理,获得消息中关于文件语义的向量描述,交给语义分析模块进行分析,然后根据分析结果找到目标文件或者路由指引信息,决定下一步消息转发。

3.2 路由指引信息的获取和维护

无结构P2P的泛洪搜索机制的路由带有盲目性,所以带来大量不必要的查询消息,导致网络负载大,可扩展性弱。为

了使得查询消息向更有可能存放目标文件的结点路由,需要在路由过程中寻找“有用的”路由指引信息,因为结点上的共享文件在一定程度上代表了结点的兴趣偏好,而且搜索请求往往与其历史相似搜索有很大关系。如果某结点曾经对相似的历史搜索作出响应,那么该结点也很可能对当前搜索作出响应,因此可以在虚结点上保存相似的历史搜索信息用于指引路由。在 SRVN 机制中,语义路由的指引信息来自虚结点上的两张表 ConFT 和 SimFT,在结点加入、查询的过程中构造和维护这两张表。

算法 1: 虚结点上 ConFT 和 SimFT 的构造和维护算法

Step1. 计算消息中获得的文件(或虚结点)和自己的语义距离 d ;

Step2. 若 $d \leq \epsilon$, 则向宿主结点发出请求,复制该文本。如果 ConFT 中容量足够,则存放该文本的 $\langle \text{OID}, \text{NID}, \text{URL}, \text{VECT}, \text{data size}, \text{data} \rangle$ 到 ConFT 中,如容量不足,则执行 ConFT 的替换算法。过程结束;

Step3. 若 $\epsilon < d \leq \epsilon'$, 如果 SimFT 中有空表项,则将该文本的 $\langle \text{OID}, \text{NID}, \text{URL}, \text{VECT} \rangle$ 放到空表项中,如果没有空表项,则执行 SimFT 的替换算法。过程结束。

当 $d > \epsilon'$, 就表示该文本和虚结点没有相关性,对 ConFT 和 SimFT 没有影响。

两张表空间满的时候,需要执行各自的替换算法。SimFT 的替换较简单,计算各表项中文本与虚结点的语义距离,和 d 相比,把语义距离最大的替换出去。下面介绍 ConFT 的替换算法。

算法 2: ConFT 的替换算法

Step1. 计算各文本与虚结点的语义距离 d_i , 分别比较 d_i 和 d 的大小;

Step2. 若对于任意 d_i , 都有 $d_i > d$, 则将该文本的 $\langle \text{OID}, \text{NID}, \text{URL}, \text{VECT} \rangle$ 存放到 SimFT 中,过程结束;

Step3. 若存在某个 d_i , 有 $d_i < d$, 则考虑语义距离最大的 d_j , 其文本的 data size 和 ConFT 中剩余容量是否足够存放新文本,如可以,则保存新文本的 $\langle \text{OID}, \text{NID}, \text{URL}, \text{VECT}, \text{data size}, \text{data} \rangle$ 到 ConFT 中,如不够,则考察语义距离次大的文本替换出去是否可行,依次找到大小合适的文本,将新文本记录替换进来,对于替换出去的文本,将它的 $\langle \text{OID}, \text{NID}, \text{URL}, \text{VECT} \rangle$ 存放到 SimFT 中。过程结束。

通过上述 ConFT 和 SimFT 的构造和维护过程,随着时间的推移,语义相近的虚结点越来越聚拢,在覆盖网上聚集成为若干语义相关的簇集。

3.3 基于虚结点的语义路由

基于虚结点的路由是一种语义路由算法。所谓语义路由就是利用搜索关键字和结点的映射关系,将查询请求转发给一部分认为最有可能满足请求的结点。该算法通过将在无结构的 P2P 网络中将虚结点按照语义分类,搜索限制在相关类别内传播,这样可以大幅度提高搜索的成功率和降低搜索开销。SRVN 使得系统的信息查找包括两个阶段:一是采用路由指引信息的查找,二是根据邻居结点表进行查找。结点接收到请求后,首先查找本机共享文件中是否有请求的目标文件。如果没有,则计算目标文件和本机各虚结点的语义距离 d 。如 $d \leq \epsilon'$, 则采用路由指引信息进行查找。否则按照原来的路由方式,将请求转发给邻居结点。具体的查找过程如下:

算法 3: SRVN 的信息查找算法

Step1. 在虚结点的本机同类文件和 ConFT 中查找,是否存

放目标文件 OID 的副本,如果有,就返回回答,查找过程结束;

Step2. 在虚结点的 SimFT 中查找,是否存放目标文件 OID 的三元组信息,如果有,就将该查询请求转发给目标结点;

Step3. 计算目标文件和 ConFT、SimFT 中各文件的语义距离,如 $d > \epsilon'$, 选取最相似的前 k (k 值由系统设定,限制消息传播的宽度)个结点转发请求,查找结束;

Step4. $d \leq \epsilon'$, 无路由指引信息,按结点上维护的邻居结点表,随机选取 k 个邻居结点转发请求。

4 SRVN 在 Gnutella 中的应用

本部分以无结构 P2P 网络的典型系统 Gnutella 为例,将 SRVN 应用到 Gnutella 中,并通过模拟实验进行相应的性能分析和评估。

4.1 基于 Gnutella 的 SRPNet 模型

首先,我们基于 Gnutella 构造 SRPNet。其实现较为简单,主要改进在于:

结点在加入系统时,增加的工作是在本机发起虚结点的划分,构造各虚结点的 ConFT 和 SimFT。

在查找-应答过程中,当收到查询请求时,使用虚结点上语义路由信息指导路由;在收到应答返回的消息时,对语义路由信息进行更新和维护。具体的步骤参考算法 1 至 3。

在 P2P 系统中,结点退出包括正常退出和异常退出两种情况。正常退出情况下,结点会主动向邻居结点发送退出的消息通知,而异常退出不可预知。所以两种情况下,处理办法不同。对于 ConFT 表来说,因为已经有了该结点上的文件,所以不受影响;而对于 SimFT 表,保存的仅仅是文件记录,包括文件名、向量描述、文件大小、宿主结点位置信息等等,所以会受到结点退出的影响。在算法中,因为有这样的前提:结点上的共享文件在一定程度上代表了结点的兴趣偏好。那么各 ConFT 和 SimFT 上所记录的结点,它们和退出结点的兴趣相似。因此,除了给邻居结点表中的结点发送通知之外,同时向 ConFT 和 SimFT 表中的结点发送退出通知。而对于异常退出的情况,则采用基于时间驱动的路由更新机制处理异常失败,即只有在系统中发现不可达链路时才启动路由更新机制。

4.2 模拟实验和性能分析

我们基于 PeerSim 模拟器建立了实验模拟环境,评估和分析了与 SRVN 算法性能相关的参数。主要评价指标包括搜索命中率、平均搜索长度、平均通信开销等。此外,我们也在实验中比较了两种搜索算法:SRVN 搜索以及有向 BFS。

图 3 显示了 TTL 取值分别为 2、4、6 时,采用了 SRVN 后的搜索命中率与之前的搜索命中率的比较。从图上可以看出,采用优化算法后,搜索命中率有了大幅度提高,并可以看到明显的演化和收敛过程。利用 SRVN 算法,其搜索效率明显优于原来的搜索方法,这主要是因为 SRVN 通过语义路由将请求转发到与搜索内容最相关的结点,逼近了目标结点,有效地缩小了查找范围。从图上还可以看出,优化的算法不能保证所有的请求都能被正确处理。造成该现象的原因是:语义路由依赖于语义分类方法的精确度,精确度越高,搜索命中率越高。

当 TTL=2 时,搜索命中率普遍不高,比较这种情况下的平均搜索跳数没有意义。图 4 是在 TTL=4、6 情况下,采用 SRVN 改进路由机制之前、之后,对于平均搜索跳数的比较。

TTL=4 时,虚结点的缓存表使得平均搜索长度降低了

15%,而在 TTL=6 时效果更明显,使用 SRVN 算法后,平均搜索跳数得到有效的收敛。

图 5 是 TTL 取值为 4 和 6 时,网络的平均开销(平均通过结点的报文数量)。

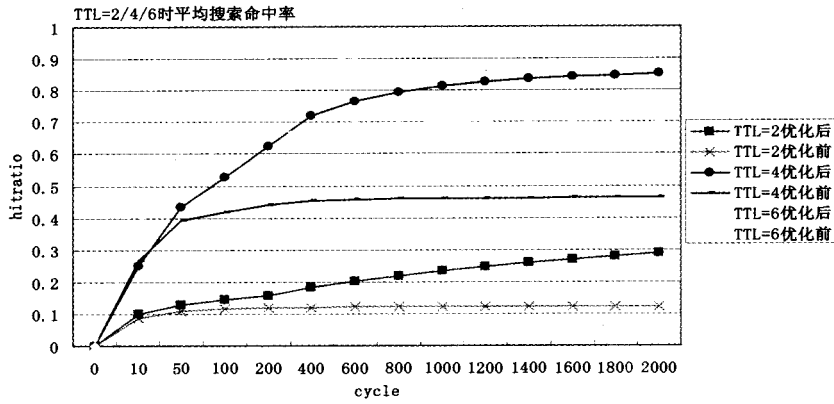


图 3 TTL=2、4、6 时平均搜索命中率的比较

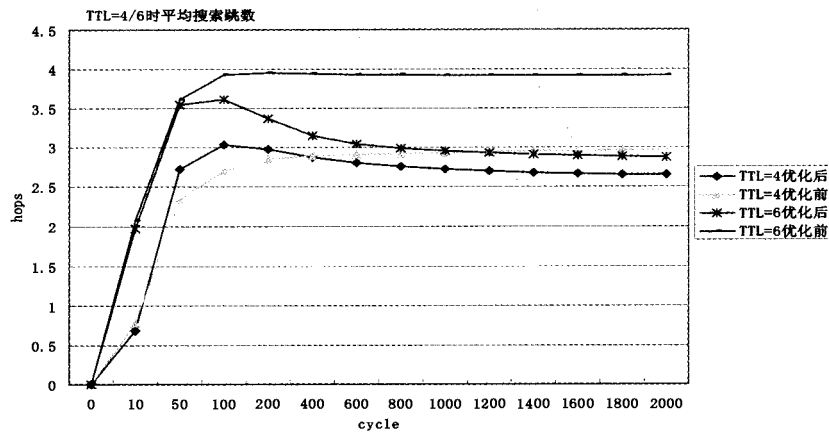


图 4 TTL=4、6 时平均搜索跳数的比较

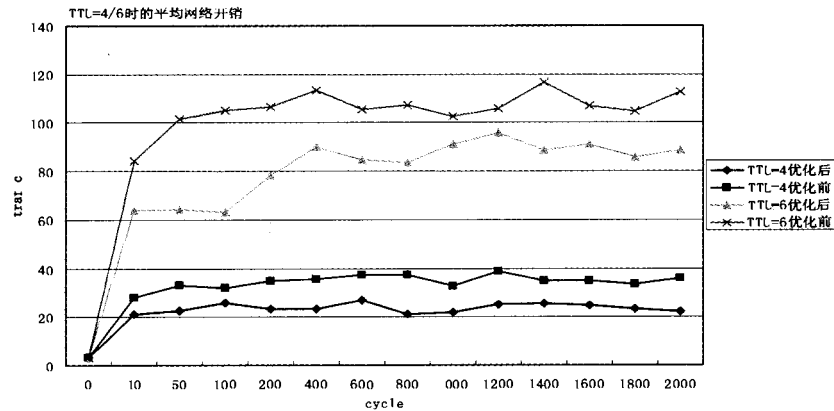


图 5 TTL=4、6 时网络平均开销

从图 5 可以看出,由于使用了历史搜索经验作为路由的指引信息,因此很大程度上减少了泛洪搜索的消息量,而且设置了 ConFT 缓存,它的高命中率降低了泛洪搜索的频率,从这两个方面,极大地降低了网络开销。

从模拟实验的结果可以得出结论:本文提出的 SRVN 搜索机制和有向 BFS 相比,可用性和高效性有明显的提高。而且,随着搜索过程的进行,ConFT 和 SimFT 使得虚结点的聚集度越来越高,搜索效果的改进越来越明显。

但是,我们也发现,目前 SRVN 搜索算法的效率对系统发布的文档集合变化敏感,所以该方法的使用还有一定的局限

性,需要进一步研究更好的指导路由的方法。另外,现在对历史记录所做的筛选是使用相似度计算的方法。因为向量特征值的提取、语义距离的计算开销较大,所以对于网络的演化有一定影响,需要进一步研究更好的处理历史搜索经验的方法。

结束语 本文提出了基于虚结点的语义路由策略 SRVN,该策略是对 P2P 泛洪搜索方式的一种优化。SRVN 根据 P2P 网络的拓扑特征和网络资源访问的局部性现象,使用语义相似度计算的方法,过滤历史搜索信息,积累搜索经验,获取路由指引信息,帮助搜索结点快速发现与搜索内容最相关的结点。通过模拟实验和性能分析,说明对于无结构

P2P 系统,应用该算法能有效地减少由于采用类泛洪方式广播消息、定位对等点的方法造成的急剧增加的网络流量,提高了搜索效率和目标命中率。该算法把基于语义的搜索技术与 P2P 系统相结合,既可充分发挥 P2P 系统的优势,克服传统搜索方式存在的可扩展性瓶颈等局限性,又可克服现有 P2P 信息服务系统对语义的支持度不高的问题,实现 P2P 系统语义信息搜索与共享。

参考文献

- 1 Gnutella Website. <http://www.gnutella.com>
- 2 Kalogeraki V, Gunopulos D, Zeinalipour-Yazti D. A Local Search Mechanism for Peer-to-Peer Networks. In: CIKM, New York, 2002
- 3 Yang B, Garcia-Molina H. Efficient Search in Peer-to-Peer Networks. In: Proc. of the 22nd IEEE ICDCS'02. Vienna Austria, July 2-5, 2002
- 4 Lv Q, Cao P, Cohen E, et al. Search and Replication in Unstructured Peer-to-Peer Networks. In: Proc. of 16th ACM ICS'02. New York USA, June 2002

- 5 Saroiu S, Gummadi K P, Dunn R J, et al. An Analysis of Internet Content Delivery Systems. In: Proc. of the 5th Symposium on Operating Systems Design and Implementation (OSDI). Boston MA, December 2002
- 6 Joseph SRH. NeuroGrid: Semantically Routing Queries in Peer-to-Peer Networks. In: International Workshop on Peer-to-Peer Computing. Pisa Italy, 2002
- 7 Crowcroft J. Application Level Programmable Inter-Network Environment. <http://www.cs.ucl.ac.uk/research/alpine/alpine.html>
- 8 Cohen E, Fiat A, Kaplan H. Associative search in peer to peer networks: Harnessing Latent Semantics. In: Proc. IEEE INFOCOM. Los Alamitos, 2003
- 9 Sripanidkulchai K, Maggs B, Zhang H. Efficient content location using interest-based locality in peer-to-peer systems. In: Proc. IEEE INFOCOM. Los Alamitos, 2003
- 10 Wang X Y, Ng Wee Siong, et al. BuddyWeb: A P2P-based Collaborative Web Caching System. In: Proc. Int'l Workshop on Peer-to-Peer Computing, Berlin German, 2002
- 11 Handurukande S B, Kermarrec A M, Le Fessant F, et al. Peer Sharing Behaviour in the eDonkeyNetwork, and Implications for the Design of Server-less File Sharing Systems; [Technical Report PI-1697]. Rennes France: Institute de Recherche en Informatique et Systemes Aleatoires, February 2005

(上接第 40 页)

$BER=10^{-4}$ 的情况,当分组长度为 200 byte 时,网络的端到端通过量和相应的网络吞吐量达到最大,最佳分组长度应该是 200 byte。再看图 7,对存在信道误码的情况,分组长度为 200 byte 时,分组传输时延约为 41ms,这个数值大于分组长度为 400 byte 所对应的分组传输时延的最小值(20ms)。

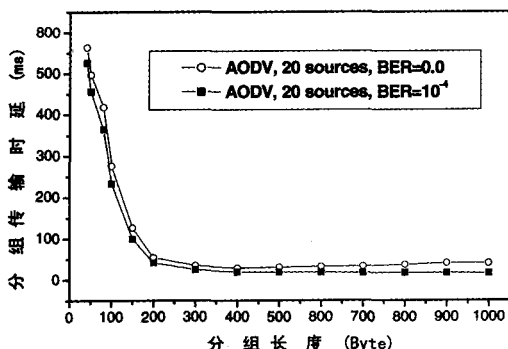


图 7 端到端分组传输时延

怎样解释上述现象呢?在存在信道误码的情况下,当分组长度很小时,信道误码造成分组遭受破坏的概率很小,几乎可以忽略。对于 CBR 业务,分组长度小意味着相同的时间内网络中各节点要处理更多的分组,MAC 层需要更多次的预约,这就会增大分组在缓冲区中的等待时延和 MAC 层的重传时延,缓冲区溢出的概率增大,从而造成端到端通过率和网络吞吐量不高;当分组长度很大时,相同的时间内网络中各节点要处理分组数较小,由于缓冲区溢出导致的分组丢失的概率迅速减小,但信道误码造成分组破坏的概率增大,此时的分组丢失主要是因为分组遭受破坏而被丢弃。在分组长度从小到大的变化过程中,会出现一个适当的长度,分组在这个长度时,由上述两种主要因素(缓冲区溢出和分组遭受破坏)导致分组丢失的概率最小,此时端到端通过率和网络吞吐量达到最高,如本文仿真条件下的 200 byte,这个长度就是存在信道误码条件下的最佳分组长度。需要指出的是,此时分组传输时延不能保证是最小,而是与分组传输时延的最小值比较接近。

结论 本文基于场景,用仿真方法研究了不同条件下网络传输 CBR 业务时的最佳 UDP 分组长度问题。仿真中采用了 AODV 和 DSR 路由协议,考虑了不同的业务强度以及有、

无误码的情况,结果表明:(1)在无信道误码情况下,采用 AODV 时,对于 CBR 连接数等于 20、10、5,最佳分组长度分别为 400、200、100 byte,最佳分组长度与网络中业务强度有关,轻业务强度条件下的最佳分组长度比重业务强度条件下的最佳分组长度短;采用 DSR 时,对于 CBR 连接数等于 20、10、5,最佳分组长度分别为 600、300、150 byte;不同的路由协议对最佳 UDP 分组长度有直接影响,同样条件下,AODV 对应的分组长度比 DSR 的要小一些。(2)在信道误码率 $BER=10^{-4}$ 的情况下,从端到端通过量或网络吞吐量的角度来看,最佳分组长度比无误码信道条件下的最佳分组长度要短;从端到端分组传输时延的角度来看,当端到端通过量或网络吞吐量达到最大时,此时分组传输时延不能保证是最小,而是与分组传输时延的最小值比较接近。信道误码率也是影响最佳 UDP 分组长度的重要因素。

参考文献

- 1 Chen Xiang, Zhai Hongqiang, Wang Jianfeng, et al. TCP performance over mobile ad hoc networks. Canadian Journal of Electrical and Computer Engineering, 2004, 29(1): 129~134
- 2 Eshak V, Baba M D. Improving TCP performance in mobile ad hoc networks. In: the 9th Asia-Pacific Conference on Communications, Vol 2. Penang, Malaysia, Sept. 2003. 580~583
- 3 Bertocchi F, Giovanardi A, Mazzini G, et al. UDP and TCP performance investigation in energy efficient ad hoc networks. In: PIMRC 2004, Vol 3. Barcelona, Spain, Sept. 2004. 1822~1826
- 4 Cano J C, Manzoni P. Group mobility impact over TCP and CBR traffic in mobile ad hoc networks. In: 12th Euromicro Conference on Parallel, Distributed and Network-Based Processing, Madrid, Spain, Feb. 2004. 382~389
- 5 Petrovic M, Vinod M, Jacobsen H A. Content-based routing in mobile ad hoc networks. In: the Second Annual International Conference on Mobile and Ubiquitous Systems: Networking and Services, San Diego, California, Jul. 2005. 45~55
- 6 Lee Jang Yeon, et al. Optimum UDP Packet Sizes in Ad Hoc Networks. In: Proc. Workshop on High Performance Switching and Routing, Merging Optical and IP Technologies, Kobe, Hyogo, Japan, May 2002. 214~218
- 7 Perkins C E, Belding-Royer E, Das S. Ad hoc On-Demand Distance Vector (AODV) Routing. Internet Draft. Available from <http://www.ietf.cnri.reston.va.us/rfc/rfc3561.txt>, Jul. 2003
- 8 Johnson D B, Maltz D A, et al. The Dynamic Source Routing Protocol for Mobile Ad Hoc Networks (DSR). Internet Draft. Available from <http://www.ietf.cnri.reston.va.us/internet-drafts/draft-ietf-manet-dsr-10.txt>, Jul. 2004
- 9 LAN/MAN Standards Committee of the IEEE Computer Society. Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications. IEEE Std 802.11-1997. The Institute of Electrical and Electronics Engineers, New York
- 10 Wu Yueh-Ting, Ho Tsung-Yen, Liao Wanjiun, et al. Epoch, length of the random waypoint model in mobile ad hoc networks. IEEE Communications Letters, 2005, 9(11): 1003~1005