doi:10.3772/j.issn.1002-0470.2014.08.001

基于流量距离因子的 P2P 缓存选址与容量分配算法①

翟海滨② 张 鸿 刘欣然 王 勇 沈时军 杜 鹏

(国家计算机网络应急技术处理协调中心(CNCERT/CC) 北京 100029)

摘 要 为缓解 P2P 应用的广泛流行给网络服务提供商(ISP)骨干网络带来的流量压力,进行了 P2P 缓存部署研究,以避免不合理的缓存部署影响缓存性能发挥和造成缓存投入浪费。首先在综合考虑骨干网络拓扑、内容热度变化和缓存状态更新等信息的基础上建立了基于骨干流量的缓存部署模型,然后定义流量距离因子辅助部署算法设计,并给出了一种基于流量距离因子的 P2P 缓存选址与容量分配(LSCA)算法,最终通过缓存位置和缓存容量的联合优化设计,充分发挥缓存性能,降低骨干网络 P2P 流量负载。仿真实验结果表明,针对典型的 H&S 型、Ladder 型骨干网络拓扑,LSCA 算法与已有部署算法相比均具有更好的性能。应用 LSCA 算法后,平均链路使用率比已有算法 Degree 低 14% ~21%,比已有算法 Centrality 低 9% ~12%;平均传输跳数比 Degree 算法低 18% ~29%,比 Centrality 算法低 11% ~20%。

关键词 P2P 缓存技术, 网络服务提供商(ISP)网络, 流量负载, 容量分配, 缓存部署

0 引 言

目前,P2P(Peer-to-Peer)流量已成为互联网流量的最主要组成部分,给网络服务提供商(ISP)带来了前所未有的运营压力。在国际上,P2P 流量占互联网流量的 60% ~75%[1],甚至高达 90%[2]。在国内,以中国电信为例,2010 年电信骨干网确知的 P2P 流量已超过 55%[3]。以 BitTorrent、迅雷等为代表的 P2P 应用已成为当今 Internet 的主流应用,P2P 应用的广泛流行使 ISP 骨干网络流量负载成倍增加,ISP 不得不增加基础设施投入以保证网络顺畅,运营压力显著提高。

P2P 缓存[4,5]是解决上述 P2P 流量问题的最有效的方法之一。P2P 缓存通过在接入网出口部署缓存设备,将 P2P 流量进行缓存并服务于后续请求,实现了 P2P 流量的本地化,从而达到减少骨干网络 P2P 流量和降低 ISP 运营压力的目标。P2P 缓存选址与容量分配用于从大量 ISP 接入网络中选择最优子集进行缓存设备部署,并为各缓存设备分配存储容量,是决定 P2P 缓存性能的关键。

本文针对 ISP 骨干网络 P2P 流量负载压力大的问题,提出了一种基于流量距离因子的 P2P 缓存选址与容量分配(location selection and capacity allocation, LSCA)算法。本研究首先在综合考虑骨干网络拓扑、内容热度变化和缓存状态更新等信息的基础上建立了与骨干流量相关的 ISP 收益函数和缓存部署模型,然后基于该模型将 P2P 缓存部署问题形式化为一个最优化问题(NP 完全问题),进而为设计具有多项式时间复杂度的缓存部署算法,定义了流量距离因子并利用该因子对原 P2P 缓存部署问题进行模型变换,最后基于变换后的模型进行部署算法设计。仿真实验结果表明,针对典型的 H&S(Hub and Spoke)型、Ladder 型骨干网络拓扑,与已有算法 Degree 和 Centrality 相比,LSCA 算法与已有部署算法相比均具有更好的性能,能降低更多骨干网络负载。

1 相关工作

部署有 P2P 缓存的 ISP 网络示意图如图 1 所示,多个接入网络连接到同一个 ISP 骨干网络,骨干

① 973 计划(2011CB302605)资助项目。

② 男,1983 年生,博士,研究方向:P2P 网络,缓存技术,分布式计算等;联系人,E-mail: zhaihaibin@ict.ac.cn

(收稿日期:2014-02-12)

网络进而连接到 Internet,具有不同存储容量的缓存设备被部署于接入网络出口,截获并服务 P2P 请求。在缓存替换算法一定的前提下,缓存容量越大则可服务请求越多,服务效果自然越好。但是 ISP 不可能无限投入,因此本文关注 ISP 投入缓存总容量有限时的缓存部署问题,重点研究如何将总容量进行分配并部署才能充分发挥缓存性能。

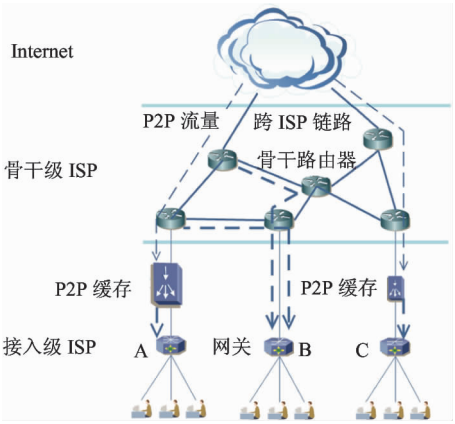

图 1 部署有 P2P 缓存的 ISP 网络示意图

缓存选址与容量分配问题是一个受多条件约束的复杂问题,原因如下:如果假设所有缓存设备的容量相同,P2P 缓存设备部署问题则退化为传统的无容量约束选址(uncapacitated facility location, UFL)问题[6],可见 P2P 缓存选址与容量分配问题难于 UFL 问题。UFL 问题已被公认是 NP 难问题,所以 P2P 缓存选址与容量分配问题也是 NP 难问题。

关于 P2P 缓存部署问题的研究在近几年刚刚起步,目前仅有的少量 P2P 缓存选址与容量分配算法[7-10]尚存在不少不足:已有方法缺乏对 P2P 内容热度分布和缓存替换算法等因素的分析和考虑,仅仅从接入网络用户数目和出口流量规模等层面设计缓存部署方案,同时部署算法主要采用按比例分配等简单低效的实现方式,缺乏高效的部署实现算法。

1.1 传统服务节点部署算法

文献[11-13]先后提出了基于线形拓扑、树形拓扑和随机图拓扑的服务节点部署模型,这些模型将网络流量分布和网络拓扑作为输入,将服务节点与用户之间的流量与两者之间的网络距离的乘积作为最小化目标。基于上述模型,设计了高效的搜索算法如贪婪算法、随机算法、热点选择算法、次优化算法等求解最优节点部署集合。文献[14]提出了适用于大规模图形拓扑的 Web 缓存部署模型,并设计了一种分布式求解算法,其特点是:以最小化服务节点与用户之间的流量和距离的乘积为目标,将该问题抽象为无容量约束的 K 中位和无容量约束节点选址两种模型;设计分布式算法进行求解,根据备选节点周围网络拓扑环境信息搜索出较优解;求解所需输入信息少,适用于大规模场景。文献[15]从 AS 拓扑层面研究了服务节点部署问题,以减小请求响应时间和服务器负载为目标,将节点部署问题描述为 *K-Median* 问题,并提出了启发式部署算法求解部署策略。该文实验结果表明,只需少量增加内容镜像节点数即可明显提升系统性能并降低服务节点负载。文献[16]研究了单服务器选址问题,以最小化网络流量和平均响应时间为目标,提出了一种面向线性和环形网络拓扑的部署算法。文献[17]则从路由拓扑层面研究了服务节点部署问题,以用户体验和网络容量为目标设计缓存部署策略。文献[18]从 ISP 角度出发,以最小化路由跳数为目标设计服务节点部署策略,通过在 ISP 部署服务节点增加 ISP 网络的连通性。

已有工作已开始对云计算环境下服务节点的部署展开研究,文献[19]研究了云服务传递网络服务器部署问题,首先提出了云服务传递网络多维设备选址模型,进而给出了启发式的求解算法。然而,该研究是以设备选址模型为基础进行模型建立的。

关于分布式缓存系统中缓存部署算法的研究目前已有较多成熟方案[20,21],但是这些工作关注的是如何优化缓存副本部署的问题,与本文研究内容不属于同一问题。

综上所述,尽管目前已经有大量关于传统服务节点部署算法的研究,但是这些方案均将服务节点部署问题建模为设备选址问题,仅能为设备选取部署位置,并不能得出各个设备的最优存储容量,并不适用于直接解决 P2P 缓存部署问题。

1.2 P2P 缓存部署算法

文献[7,8]提出了一种最大化 ISP 收益的缓存设备选址算法 BBA,ISP 收益为所有骨干链路收益的总和,骨干链路收益定义为缓存部署后该链路流量降低比例的函数,链路流量降低越多,ISP 收益越大。该文所提算法可以减低 ISP 骨干网络负载,然而,该算法并未解决缓存容量分配问题,并做出所有缓存设备命中率为相同定值的假设。

文献[9]提出了 Centrality 算法,在综合考虑接入网络用户数目和出口流量规模的基础上给出节点中心系数的定义,并将接入网络 i 的中心系数 $H(i)$ 定义为

$$H(i) = \frac{1}{\sum_{j=1}^{N} t_j h_{ij}} \tag{1}$$

式中 N 为骨干网络的接入网络数目,t_j 为接入网络 j 产生的流量在网络总流量中所占比例,h_{ij} 为接入网络节点 i 与节点 j 之间的路由路径跳数。该文用中心系数 $H(i)$ 来描述接入网络 i 对整个骨干网络流量负载产生的影响,认为中心系数值越大,为骨干网络带来的负载越高。同时设计了基于中心系数的缓存部署算法,实验结果表明该算法可以明显降低骨干网络流量负载。

在实际应用中,ISP 会采用 Degree 算法[10]等进行缓存部署。Degree 算法根据各接入网络的节点度数设计缓存部署位置及其容量,这种算法的优势是实现简单快速,然而很难充分发挥缓存性能,容易造成 ISP 缓存投入的浪费。

综上所述,目前关于 P2P 缓存部署问题的研究较少,仅有的少量 P2P 缓存部署算法尚存在不足,例如,P2P 内容热度分布和缓存替换算法等因素决定缓存存储状态与服务效率,已有方法缺乏对这些关键因素的深入分析和考虑,仅仅从接入网络用户数目和出口流量规模等层面设计缓存部署方案,而且部署算法主要采用按比例分配等简单低效的实现方式,缺乏高效的部署实现算法。

2 基于流量距离因子的缓存部署问题建模

2.1 问题建模

面向骨干网络 P2P 流量优化的缓存部署问题(以下简称 P2P 缓存部署问题)定义为在 ISP 投入缓存总容量 T(字节)已知的前提下,求解分配给每个接入网络 i 的最优缓存容量 S_i(字节)且满足

$$\sum_{i=1}^{N} S_i = T \tag{2}$$

最终达到最小化骨干网络流量负载的目的。在部署算法确定的情况下,缓存部署容量越大,缓存所服务 P2P 请求越多,网络负载降低效果越明显。但是 ISP 不可能无限投入,因此本文研究 ISP 投入缓存总容量有限时的 P2P 缓存最优部署问题。假设 $S = \{S_1, S_2, \cdots, S_N\}$ 为缓存容量分配集合,如果 $S_i = 0$,则说明在接入网络 i 的出口没有部署缓存。假设根据 S 进行缓存容量分配和部署后 ISP 的代价为 $F(S)$,描述网络负载的典型代价函数形式如下:

$$F(S) = \sum_{i=1}^{L} (Y_i - \Delta Y_i) \tag{3}$$

其中 Y_i 为缓存部署前链路 i 的 P2P 流量,ΔY_i 为缓存部署后链路 i 减少的 P2P 流量,L 为 ISP 骨干网络所有链路数目总和。式(3)定义的代价函数表示缓存部署后所有链路流量负载总和。由于经过路由节点处理的流量最终都会经骨干链路传输,因此该代价函数不仅能够反映骨干链路负载,也能够反映路由节点处理负载,这是 ISP 骨干网络负载情况的充分体现。ISP 也可以根据需求灵活定义其它不同形式的代价函数,比如考虑链路平均使用率、添加链路权重因子、添加与位置相关的部署成本、设计非线性代价函数等。P2P 缓存部署问题可以描述为如下最优化问题:

$$\begin{aligned} &\min_S \quad \sum_{i=1}^{L} (Y_i - \Delta Y_i) \\ &s.t. \quad \sum_{i=1}^{N} S_i = T \\ &\quad S_i = \{0, T_{\min}, \cdots, T\} \ \forall i \end{aligned} \tag{4}$$

缓存的安装和部署存在与容量无关的固定开销,当 S_i 很小时,这部分部署开销不变,而此时缓存在降低网络负载方面的效果很小,反而得不偿失,因此 S_i 应存在一个部署容量下限 $T_{\min}$。上述模型形式简单且能满足 P2P 缓存部署问题求解的需求:首先,在给定缓存总容量 T 的条件下,通过求解(4)式所示问题不仅可以得出缓存部署位置,而且可以得出使 ISP 总代价最小的各个位置对应缓存容量;其次,采用关于骨干流量的目标函数,从而只需考虑缓存部署对骨干流量的影响,无需确知请求服务节点以及请求服务距离矩阵等信息,与传统设备选址模型存在较大差别;最后,骨干流量是网络拓扑、骨干路由和 P2P 流量分布等信息的集中反映,无需再考虑 P2P 请求服务节点随机分布等因素,既满足 P2P 内容请求访问模式需求,又降低了模型复杂度。

2.2 问题理论分析

定理 1. 基于流量距离因子的 P2P 缓存部署问题(如式(4)所示)为 NP 完全问题。

证明:首先,证明 P2P 缓存部署问题是 NP 问题。我们用集合 S 作为证书,一旦给定 S,ISP 的代价值即可通过式(3)求出。这个过程可以在多项式时间完成,因此 P2P 缓存部署问题是 NP 问题。其次,我们用无容量约束选址(UFL)问题证明 P2P 缓存部署问题是 NP 难题,即 UFL 问题 $\leqslant_p$P2P 缓存部署问题。规约过程如下。

UFL 问题可以描述为如下形式：

输入：服务站点集合 F，用户集合 C，用户到最近服务站点的距离矩阵。

输出：使 $\sum_{i\in C} dist(i,Q)$ 值最小的集合 Q，满足 $Q \subseteq F$。$dist(i,Q)$ 为 i 到 Q 中距离 i 最近服务站点的距离。

假设所有缓存设备的容量相同，ISP 网络拓扑为一个完全图，当前网络只有 1 个文件，因而单个缓存最大容量是 1。定义一个与 S 一一对应的新集合 S^*，S^* 是一个整数集合，包含了满足 $S_i = 1$ 且 $1 \leqslant i \leqslant N$ 的每一个整数元素 i。根据 S^* 的定义不难发现，S 与 S^* 可以进行相互唯一转换。此时，ISP 拓扑图中任意两个节点之间的最短路由路径就是两点之间的链路。将骨干网络所有接入网络集合 $\{1,2,\cdots,N\}$ 抽象为 UFL 问题中的服务站点集合，将链路集合 E 抽象为用户集合。对于任意链路 i，假设其两个端点分别为 i_1、i_2，两节点之间的 P2P 流量满足 $t_{i_1i_2} = t_{i_2i_1}$，其中为 $t_{i_1i_2}$ 为从 i_1 流向 i_2 的流量。将缓存部署于 S 前后链路 i 上的流量差值抽象为用户到最近服务站点的距离，该距离表达式的定义需要借助集合 S^*，具体表示如下：

$$dist(i,S) = \begin{cases} 2t_{i_1i_2} & |\{i_1,i_2\} \cap S^*| = 0 \\ t_{i_1i_2} & |\{i_1,i_2\} \cap S^*| = 1 \\ 0 & |\{i_1,i_2\} \cap S^*| = 2 \end{cases} \tag{5}$$

这样 P2P 缓存部署问题即可抽象为如下所示的 UFL 问题。

输入：服务站点集合 $\{1,2,\cdots,N\}$，用户集合 E，由式(5)表示的用户到最近服务站点的距离矩阵。

输出：使 $\sum_{i\in E} dist(i,S)$ 值最小的集合 S。

因此，UFL 问题 $\leqslant_p$P2P 缓存部署问题，问题得证。

3 基于流量距离因子的缓存部署算法设计

如 2.2 节所述，P2P 缓存部署问题的求解复杂度较高。因此，本文首先对如式(4)所示的 P2P 缓存部署模型进行变换，建立新的变换求解模型。然后，基于该模型设计具有多项式时间复杂度的部署算法。

3.1 模型变换

(1) 流量距离因子定义

如式(4)所示的 P2P 缓存部署问题模型存在的问题是当缓存根据容量分配集合 S 进行部署后，需要对网络所有链路进行遍历计算才能知道网络负载变化情况，才能计算 ISP 代价值，计算复杂度较高。本文中我们引入流量距离因子 d 解决上述问题。

流量距离因子：接入网络 i 的流量距离因子 d_i 定义为

$$d_i = t_i \sum_{j=1}^{N} t_j h_{ij} \tag{6}$$

式中 N 为骨干网络的接入网数目，t_i、t_j 分别为接入网络 i 和 j 产生的流量在网络总流量中所占比例，h_{ij} 为接入网络 i 与 j 之间的路由路径跳数。流量距离因子 d_i 的物理意义为在没有缓存部署的情况下，来自接入网络 i 的 P2P 流量所产生的骨干网络负载。

(2) 基于流量距离因子的骨干网络收益函数定义

在定义骨干网络收益函数之前，需要首先确定缓存存储的内容集合，该集合是由缓存替换算法决定的。常用的 P2P 缓存替换算法包括最小频率优先(LFU)，最近最少优先(LRU)以及最少发送字节数优先(LSB)等。其中 LSB 算法不但考虑了内容大小而且考虑了内容热度，是目前可以获得最高字节命中率(byte hit ratio, BHR)的算法。因而，本文假设 ISP 采用 LSB 算法作为其缓存替换算法。需要说明的是，本文所提部署策略并不是仅适用于 LSB 算法。当采用其它缓存替换算法时，只需利用该算法对应缓存命中率计算公式更新收益函数中的热度计算公式，就可以得出适用于该替换算法的缓存部署策略。

基于流量距离因子，我们定义 $D(i,j)$ 为接入网络 i 的收益，表示接入网络 i 的缓存容量为 j 时为骨干网络产生的收益，形式如下：

$$D(i,j) = Bd_i \sum_{m=1}^{j} p_{r(m)} \tag{7}$$

式中 B 为内容平均大小(字节)，$\sum_{m=1}^{j} p_{r(m)}$ 表示缓存 i 所存储所有内容的热度之和，p 为内容热度分布函数，$r(m)$ 为发送字节数第 m 多的内容的热度排名，注意缓存容量 j 的单位不是字节数，而是内容数。

收益函数(式(7))的值表示容量为 j 的缓存部署于接入网络 i 后，该接入网络 P2P 流量占用骨干

网络的总传输跳数的降低值。该降低值越大,说明缓存效果越好,骨干网络的流量负载越低。

(3) 基于流量距离因子的变换求解模型

基于式(7)所示的收益函数,我们利用如下所示的模型代替如式(4)所示原 P2P 缓存部署模型:

$$\max_K \sum_{i=1}^{N} D(i,k_i) = \sum_{i=1}^{N} Bd_i \sum_{j=1}^{k_i} p_{r(j)}$$

$$s.t. \quad \sum_{i=1}^{N} k_i = \lfloor T/B \rfloor \tag{8}$$

$$k_i = \{0, k_{\min}, \cdots, k_{\max}\} \ \forall i$$

式中 $K = \{k_1, k_2, \cdots, k_N\}$ 为缓存容量分配集合,但是单位不再是字节数,而是内容数,其中 k_i 为在接入网络 i 出口部署缓存能存储的平均内容个数,$k_i = \lfloor S_i/B \rfloor$。问题(式(8))的目标函数值为所有接入网络的收益总和,即骨干网络总收益,表示最大化缓存部署于 K 后所有接入网络 P2P 流量占用骨干网络的总传输跳数的减少值。第一个约束条件表示所有缓存容量之和为 ISP 投入总容量;第二个约束条件对单个缓存容量进行限制:接入网络 i 可以不部署缓存(即 $k_i = 0$),一旦部署缓存就必须遵守容量限制。如 2.1 节所述,缓存容量下限为 $T_{\min}$,因此得出容量下限 $k_{\min} = \lfloor T_{\min}/B \rfloor$;缓存容量不应超过总容量 T 或网络中内容总数 M,因此容量上限 $k_{\max} = \min\{\lfloor T/B \rfloor, M\}$。通过求解问题(式(8))得到最优缓存容量分配集合 K 后,将其乘以 B 即可得到最终的缓存容量分配策略。通过求解问题(式(8))可以实现最小化骨干网络链路负载的目标,原因如下:尽可能降低各个接入网络 P2P 流量流经骨干网络的跳数总和,不仅可以减少接入网络流入骨干网络的 P2P 流量,而且可以减少流量流经的骨干链路数目,从而最小化骨干链路负载。

本文所提模型大大降低了缓存部署策略的计算复杂度:首先,在原部署模型中,任意接入网络 i 部署缓存设备都会对整个网络产生影响,通过引入流量距离因子 d 来刻画接入网络对骨干网络负载的影响,可将缓存影响限制到每个接入网络,从而使得问题(式(8))具有最优子结构性质,可以在多项式时间内求解,大大降低求解复杂度。其次,本模型引入平均内容大小 B 作为分配单位,与以字节为分配单位相比计算速度必然有显著提升,同时,用 k_iB 近似求解 S_i 也是合理的。利用本文所提模型可达到降低部署策略计算复杂度的目标,但是以牺牲部署策略的准确性为代价。本文所提模型为变换模型,利用流量距离因子刻画的网络负载降低值与实际网络负载降低值必然存在差距,因而利用本文所提模型求得的缓存部署策略为近似最优部署策略,将在第 4.4 节对近似最优部署策略与实际最优部署策略进行对比分析。

3.2 部署算法设计

基于经过变换后的 P2P 缓存部署模型(如式(8)所示),本文提出一种面向骨干网络 P2P 流量优化的缓存部署(LSCA)算法。

LSCA 算法采用迭代增量式思想计算最优部署策略,假设 N 个接入网络已经被依次编号。首先计算在接入网络 1 部署缓存时的最优部署策略,只需遍历求解所有可能缓存容量产生的收益,最优策略显然是为接入网络 1 分配 $k_{\max}$;然后计算在接入网络 1 和 2 部署缓存时的最优部署策略,此次循环中将前一次循环结果(只在接入网络 1 部署)作为已知条件,只需在总容量和单个缓存容量限制下,遍历求解在接入网络 2 部署所有可能缓存容量产生的收益(保存中间计算结果),并从中选择收益最大的容量作为接入网络 2 的缓存容量,剩余缓存容量即为接入网络 1 的容量。其余循环计算过程相同,直至求解出在接入网络 1 到 N 部署缓存的最优策略。

LSCA 算法的描述如表 1 所示。在循环 i 中,计算在接入网络 1 到 i 部署缓存时的最优容量分配策略。循环 i 中能够分配的总容量最小为 $k_{\min}$,最大为 $\min\{ik_{\max}, \lfloor T/B \rfloor\}$(第 2 行)。当接入网络 i 分配容量为 0 时,最大收益即为前一次循环(在接入网络 1 到 $i-1$ 部署)中产生的最优策略(第 3 至第 5 行)。当接入网络 i 分配容量不为 0 时,计算当前收益并与最大收益值比较,从中选择收益更大的容量作为接入网络 i 的当前最优缓存容量(第 6 至第 12 行)。通过算法执行过程中保存的中间结果,可以依次得出接入网络 1 到接入网络 $i-1$ 的最优缓存容量(第 16 至第 21 行),从而确定最优缓存容量集合 S(第 22 行)。

3.3 算法复杂度分析

设 $V = \lfloor T/B \rfloor$ 表示单个缓存可部署的最大容量,并设骨干网络中骨干节点数目为 N。LSCA 算法中存在三层循环,第一层循环(第 1 行)的执行次数为 N,第二层(第 2 行)和第三层循环(第 6 行)的复杂度均为 $O(V)$,所以 LSCA 算法具有多项式时间复杂度 $O(NV^2)$。

表 1 缓存部署算法 LSCA 的描述

输入:缓存总容量 T, 平均内容大小 B, 骨干网络初始流量分布集合 t, 接入网络间路由路径跳数矩阵 h, 热度分布函数 p。

输出:最优缓存容量分配集合 S。

变量定义:

$D(i, j)$: 如(7)所示的接入网络 i 的收益函数;

$value(i, j)$: 在接入网络 1 到 i 部署总容量为 j 的缓存时,骨干网络最大收益值;

$current(i, j)$: 在接入网络 1 到 i 部署总容量为 j 的缓存时,为使收益最大,接入网络 i 所部属缓存容量大小;

$previous(i, j)$: 在接入网络 1 到 i 部署总容量为 j 的缓存时,为使收益最大,接入网络 1 到 $i-1$ 所部署缓存总容量大小, $previous(i, j)$ 等于 j-$current(i, j)$。

算法描述:

```
LSCA (T, B, t, h, p)
/* 循环 i 中计算在接入网络 1 到 i 部署缓存时骨干网络最大收益值及当前最优缓存部署策略 */
1.  For i = 1 to N do {
2.  For total = k_min to min{ik_max, ⌊T/B⌋} do {
/* 考虑接入网络 i 当前容量为 0 的情况 */
3.   value(i, total) ← value (i − 1,total);
4.   current(i, total) ← 0;
5.   previous (i, total) ← total;
/* 遍历求解在接入网络 i 部署所有可能缓存容量产生的收益并保存中间结果 */
6.  For current = k_min to k_max do{
7.   tmp←D(i,current) + value(i − 1, total-current);
/* 更新当前最优策略 */
8.    If tmp > value(i, total)
9.     value(i, total). value ← tmp;
10.     current (i , total) ← current;
11.     previous (i , total) ← total-current;
12.    End If
13.   }End For
14.  }End For
15. }End For
/* 根据中间计算结果依次得出接入网络 1 到接入网络 N 的最优缓存容量 */
16. pre ←⌊T/B⌋;
17. For i = N to 1 do {
18.   K(i) ← current(i, pre);
19.   pre←previous(i, pre);
20.   S(i) ← K(i)B; //利用 K 乘以 B 求出 S
21. }End For
22. return S
```

4 实验结果与评价

以参考文献[9]的 Centrality 算法和参考文献[10]的 Degree 算法作为性能评价的对比方案,主要从以下两个方面对 LSCA 算法进行性能评价分析:(1) LSCA 算法与对比方案的比较分析,主要指标包括骨干网络的平均链路使用率、平均传输跳数等;(2) LSCA 算法的影响性分析,反映 LSCA 在降低网络负载方面的效果,主要包括骨干链路使用率的变化情况、骨干节点和骨干链路流量负载变化情况等。

4.1 实验参数和环境

文献[9, 22]根据节点度数将 ISP 骨干网络拓扑划分为 H&S 型网络和 Ladder 型网络:当节点相对最大度数(节点最大度数除以节点总数)大于等于 0.4,或者节点平均度数大于 3 时,该网络为 H&S 型网络,否则,该网络为 Ladder 型网络。CAIDA 项目在其主页上公开了部分著名的商用和学术用 ISP 网络拓扑[23],我们从中挑选两个 H&S 型网络:Cable and Wireless (CW)和 Qwest;两个 Ladder 型网络:Netrail Incorporated (NI)和 CAIS Internet (CAIS)作为本文实验的网络拓扑,从而验证本文所提算法是否适用于不同类型的网络,所选网络拓扑结构如图 2所示,每个拓扑中的二元组表示(节点数目,链路数目),节点为接入网络对应骨干路由器,而链路为路由器之间的连接链路。选择节点间跳数最小的路径作为路由路径,该路径可通过 Floyd 算法[24]得出。对于缓存部署问题而言,关键参数不是描述节点动态性的节点加入速率和节点退出概率等,而是网络中 P2P 流量分布、网络内容总数以及内容热度分布函数等,下面介绍这些关键参数的设置。

文献[25]指出,接入网络 i 产生的流量与其人口数成正比,因此本文令网络 i 产生的流量为其相对人口比例 P_i, 即网络 i 的人口数与所有接入网络总人口数的比值。节点 V_i 与 V_j 之间的流量 t_{ij} 即为 $P_i * P_j$。网络中的所有内容集合为 M 且 $|M| = 1000$,此设置与文献[22]的设置相同。内容热度分布服从 Mandelbrot-Zipf 分布,其中内容 j 的热度 $p(j) = a/(j+q)^b$, a 是归一化常系数,比如 $a = 1/\sum_{j=1}^{|M|} p(j)$, b 为偏度系数, q 为高原因子。根据初始链路负载设置各链路的带宽,满足各条链路的初始链路使用率在 50% ~80% 之内。文献[22]已经给

出了 BitTorrent 系统内容大小分布情况:大小为 10Mbyte 的内容占 8%,大小为 100Mbyte 的内容占 28%,大小为 300Mbyte 的内容占 40%,大小为 700Mbyte 的内容占 11%,剩余 13% 的内容大小为 1.4Gbyte,所以平均内容大小 B 为 400Mbyte。我们根据此分布随机为内容设置大小,同时将缓存总容量设为 1 Tbyte。实验中将挑选如图 2 所示 4 种算法进行对比分析。

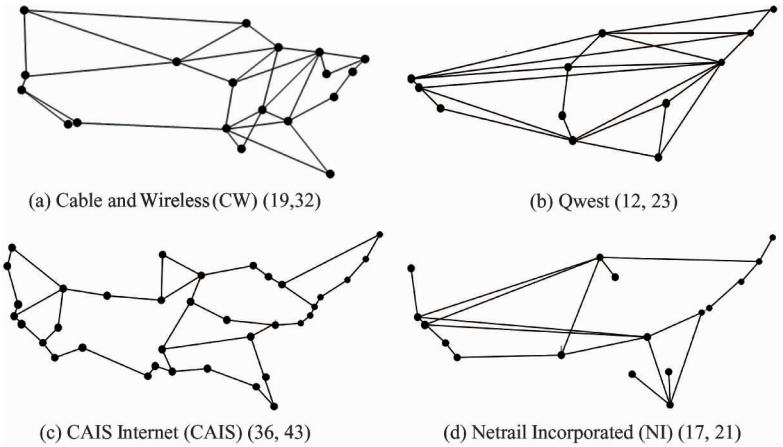

图 2　实验采用的网络拓扑结构图

4.2　与其他算法的性能比较分析

将 LSCA 算法应用于不同类型网络,并与已有算法 Degree、Centrality 以及最优解 OPT 进行比较。Centrality 算法:文献[9]提出的基于中心系数的部署算法;Degree 算法:根据各接入网络的节点度数所占比例设计缓存部署位置及其容量[10];OPT 算法:采用穷举法遍历所有可行部署策略,因而得出的是最优策略,由于本实验骨干网络拓扑节点数目通常小于 100 个[23],因此穷举法的运行时间可以接受;LSCA 算法:本文所提算法。

本实验中算法的评价指标包括骨干节点平均流量负载和平均传输跳数。骨干节点平均流量负载表示所有骨干节点 P2P 流量负载的平均值,可以反映骨干网络 P2P 流量负载的多少;平均传输跳数表示流量在骨干网络传输所需跳数的平均值,不仅可以反映骨干网络负载的多少,而且可以反映用户访问体验的高低。

4.2.1　平均链路使用率比较分析

图 3 至图 6 显示了不同算法应用于不同类型网络拓扑后骨干网络平均链路使用率的变化情况,“Original Value”表示缓存部署前的平均链路使用率初始值。首先,在不同网络拓扑下 LSCA 算法所得结果均与 OPT 算法几乎相同,原因如下:LSCA 算法采用流量距离因子进行缓存部署策略制定,该因子将流量和传输距离进行综合考虑,可以合理地反映出骨干网络负载;同时,LSCA 算法虽然用内容平均大小作为分配单位可能导致误差的产生,但是站在计算缓存总容量的角度,这种误差可以忽略。其次,LSCA 算法在降低骨干节点负载方面的性能明显优于 Centrality 算法和 Degree 算法,而 Centrality 算法的性能优于 Degree 算法。Degree 算法根据接入网络节点度数进行缓存容量分配,是一种简单快速的缓存部署方案。相比 Degree 算法而言,Centrality 算法不仅考虑了接入网络拓扑信息(路由跳数),而且结合接入网络对应用户数目进行缓存容量分配,所以性能明显优于 Degree 算法。实验结果表明,随着缓存总容量的增加,骨干网络平均链路使用率均明显降低。同时,不同算法间平均链路使用率差值随缓存总容量的增加变化不大,应用LSCA算法后,骨

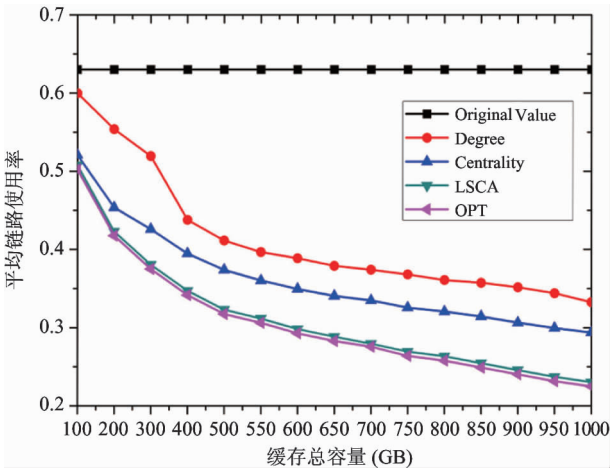

图 3　CW 网络平均链路使用率变化情况

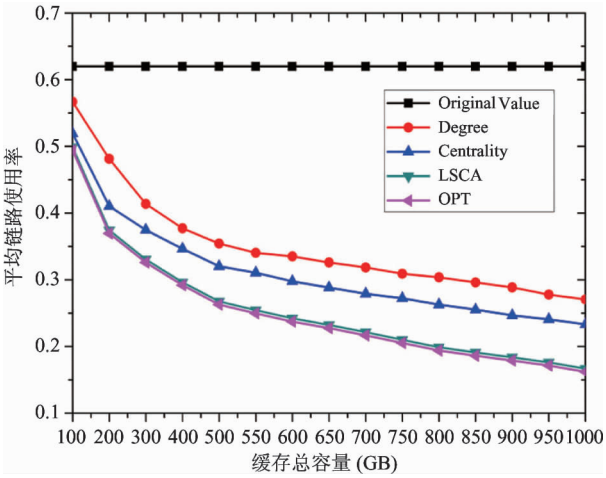

图 4 Qwest 网络平均链路使用率变化情况

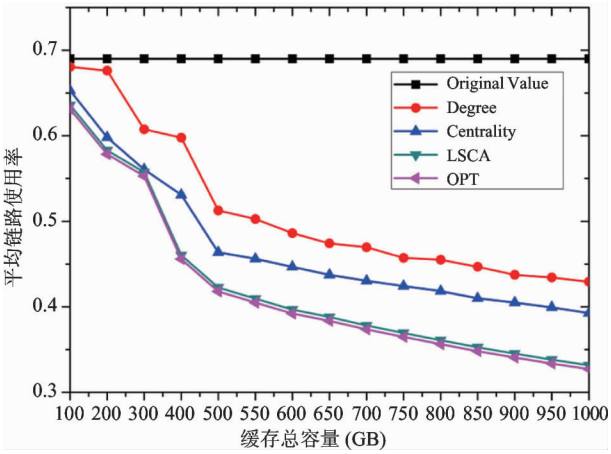

图 5 CAIS 网络平均链路使用率变化情况

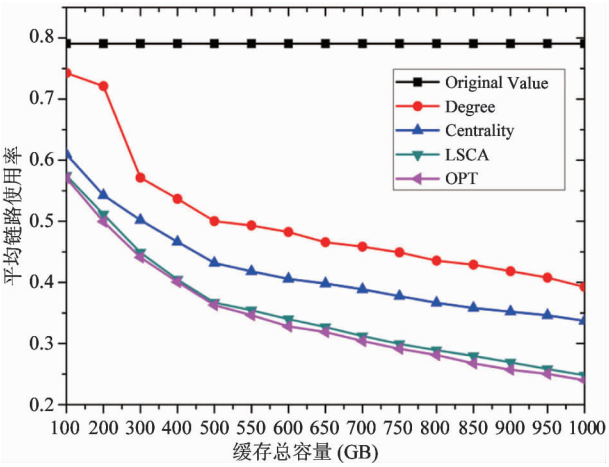

图 6 NI 网络平均链路使用率变化情况

干网络的平均链路使用率比 Degree 算法低 14% ~ 21%,比 Centrality 算法低 9% ~12%。

4.2.2 平均传输跳数比较分析

平均传输跳数表示流量在骨干网络传输所需跳数的平均值,可以反映骨干网络负载的多少,也可以反映用户访问体验的高低。平均传输跳数越小,用户获取 P2P 内容时延越小,用户体验越好。图 7 至图10所示为不同算法应用于不同类型网络拓扑后

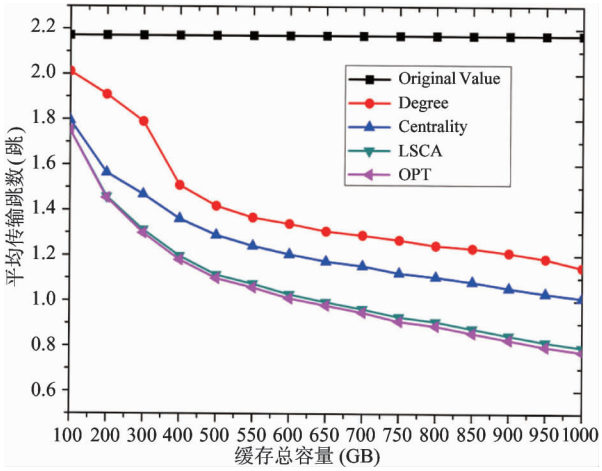

图 7 CW 网络下平均传输跳数对比

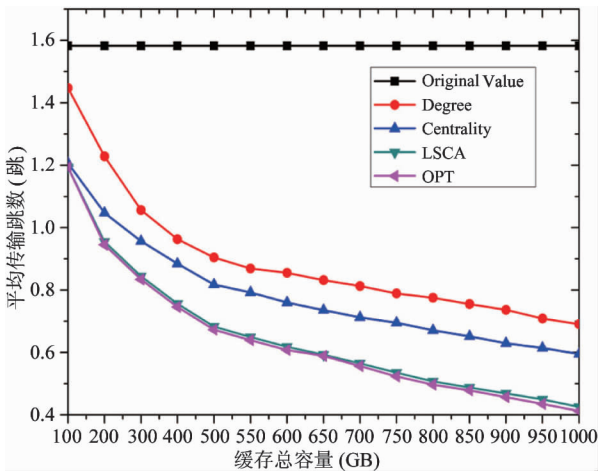

图 8 Qwest 网络下平均传输跳数对比

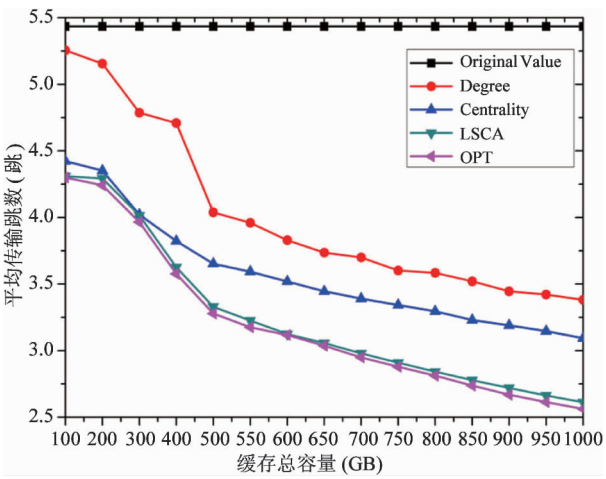

图 9 CAIS 网络下平均传输跳数对比

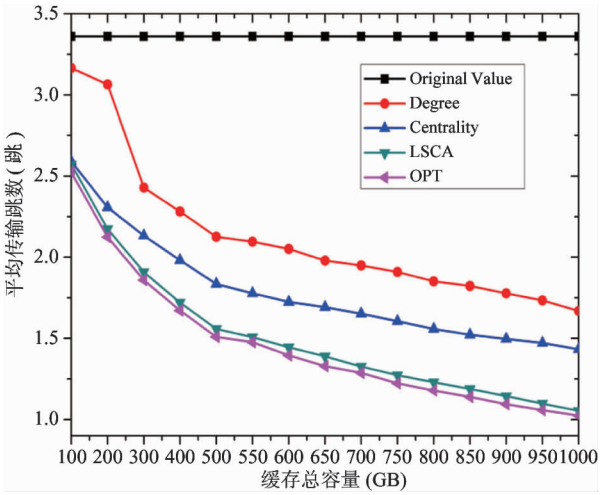

图 10　NI 网络下平均传输跳数对比

平均传输跳数的变化情况,“Original Value”表示缓存部署前的初始平均传输跳数值。实验结果表明,LSCA 算法的性能与 OPT 算法几乎相同,应用 LSCA 算法后,骨干网络平均传输跳数比 Degree 算法低 18% ~29%,比 Centrality 算法低 11% ~20%。我们还发现 Ladder 型网络 CAIS 和 NI 的平均传输跳数明显高于 H&S 型网络 CW 和 Qwest,这是因为 Ladder 型网络链路数目少于 H&S 型网络,所以流量传输所需跳数较多。

4.2.3　开销比较分析

与 3.3 节一样,设 $V=\lfloor T/B \rfloor$ 表示单个缓存可部署的最大容量,N 为骨干网络中的骨干节点数目。采用遍历方法寻找最优缓存部署策略所需计算时间为 $O(V^N)$,显然 LSCA 算法的时间复杂度远低于遍历法。

与 Degree 算法和 Centrality 算法相比,LSCA 算法的性能更好,但是也具有更高的时间复杂度。LSCA 算法具有多项式时间复杂度 $O(NV^2)$,而 Degree 算法和 Centrality 算法的时间复杂度均为 $O(N)$。尽管如此,由于在实际应用中缓存部署算法多采用离线方式运行,因此具有多项式时间复杂度 $O(NV^2)$ 的 LSCA 算法是完全可以接受的。

4.3　LSCA 算法性能分析

以 CW 网络为例,展示应用 LSCA 算法后骨干链路使用率、骨干链路流量负载和骨干节点流量负载随缓存部署容量的变化情况,从而反映本文算法在降低网络负载方面的效果。LSCA 算法应用于其它网络拓扑的流量负载变化情况与 CW 网络类似。

4.3.1　骨干链路使用率的变化情况

图 11 显示了应用 LSCA 算法后,CW 网络各条骨干链路使用率的变化情况,“Original Value”表示缓存部署前骨干链路的初始使用率值。随着缓存部署容量的增加,骨干链路使用率明显减小。实验结果表明,当缓存容量为 1Tbyte 时,LSCA 算法可使 CW 网络的各条骨干链路使用率降低 35% ~75%。

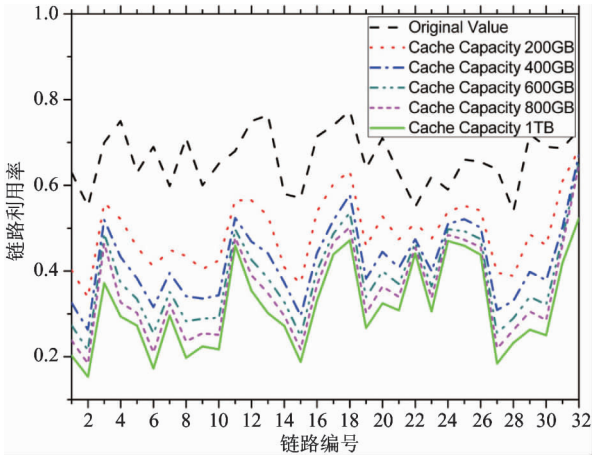

图 11　CW 网络骨干链路使用率变化情况

4.3.2　骨干链路流量负载的变化情况

图 12 显示了应用 LSCA 算法后,CW 网络各条骨干链路流量负载的变化情况,“Original Value”表示缓存部署前的初始骨干链路流量负载值。骨干链路流量负载变化情况与骨干节点类似,随着缓存部署容量的增加,骨干链路负载减小。但是,流量负载减小速度越来越慢,这是因为当缓存容量到达一定规模后,热门内容多被缓存,在此基础上增加缓存容量只能缓存非热门内容,由此带来的负载降低值变小。实验结果表明,当缓存容量为 1Tbyte 时,LSCA 算法可使骨干链路 P2P 流量负载降低 40% ~72%。

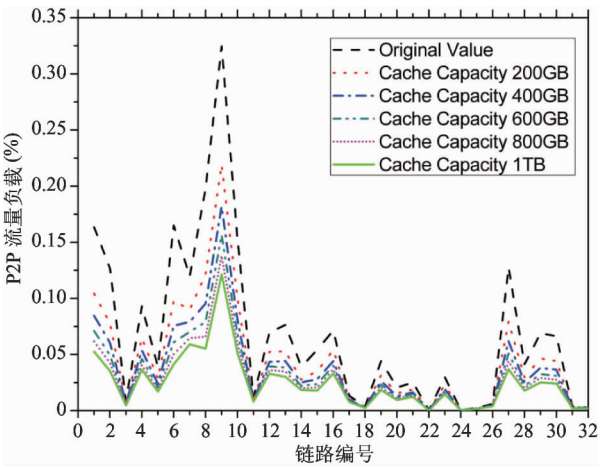

图 12　CW 网络骨干链路流量负载变化情况

4.3.3 骨干节点流量负载的变化情况

图 13 所示为 LSCA 算法应用于 CW 网络后，骨干节点流量负载随缓存总容量的变化情况，“Original Value”表示缓存部署前的初始骨干节点流量负载值。骨干节点 i 的流量负载包括两部分：一是由来自接入网络 i 的请求产生的流量，我们称之为始发流量；二是由其他接入网络产生但是流经接入网络 i 对应骨干路由器的流量，我们称之为过路流量。缓存部署前，始发流量和过路流量都需要骨干路由节点的处理；缓存部署后，始发流量中缓存命中部分由缓存处理，未命中部分以及过路流量仍然需要骨干路由节点进行处理。缓存设备部署后，始发流量不受影响，过路流量明显减少，因此，骨干节点流量负载降低本质上是过路流量减少所产生的效果。实验结果表明，应用 LSCA 算法进行缓存部署时，骨干节点流量负载随着缓存容量的增加而减小。但是，流量负载减小速度越来越慢，同样是因为当缓存容量到达一定规模后，热门内容多被缓存，在此基础上增加缓存容量带来的负载降低效果变小。在本文实验环境下，当缓存容量为 1Tbyte 时，LSCA 算法可使骨干节点 P2P 流量负载降低 50% ~70%。

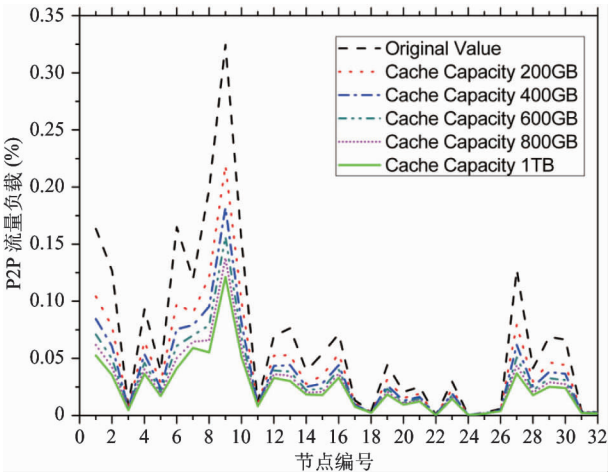

图 13 CW 网络骨干节点流量负载变化情况

5 结 论

本文提出了一种基于流量距离因子的缓存选址与容量分配算法 LSCA。该算法不仅能为骨干级 ISP 选取缓存设备部署位置，而且能够得出各个设备对应的近似最优存储容量。

首先建立了基于骨干流量的 P2P 缓存部署问题数学模型，并证明该问题为 NP 完全问题。由于 P2P 缓存部署问题求解复杂度较高，本文对 P2P 缓存部署模型进行了变换：引入流量距离因子来描述 P2P 流量为骨干网络带来的传输负载，进而基于该因子构建 ISP 收益函数，并建立新的变换求解模型。基于该模型，设计具有多项式时间复杂度的迭代增量式部署算法 LSCA。

通过实验对 LSCA 算法进行了性能评价，包括：(1) LSCA 算法与对比算法的比较分析，主要指标包括平均链路使用率、平均传输跳数等；(2) LSCA 算法的性能分析，反映 LSCA 算法在降低网络负载方面的效果，主要包括骨干链路使用率变化情况、骨干节点和骨干链路的流量负载变化情况等。实验结果表明，LSCA 算法可降低更多骨干网络负载，有效减轻 ISP 骨干网络流量压力。针对典型的 H&S 型、Ladder 型骨干网络拓扑，应用 LSCA 算法后，平均链路使用率比已有算法 Degree 低 14% ~21%，比已有算法 Centrality 低 9% ~12%；平均传输跳数比 Degree 低 18% ~29%，比 Centrality 低 11% ~20%。

参考文献

[1] Slyck News. Cachelogic study - P2P is changing. http://www.slyck.com/story914_CacheLogic_Study_P2P_is_Changing：Slyck，2005

[2] Ipoque. Ipoque Company Internet study 2008/2009. http://www.ipoque.com/sites/default/files/mediafiles/documents/internet-study-2008-2009.pdf:Ipoque，2009

[3] Wei Leping. Development trends and challenges of telecommunications industry and telecommunications technology. http://wenku.baidu.com/view/be139cc78bd63186bcebbcdc.html：Baidu，2010

[4] Wierzbicki A，Leibowitz N，Ripeanu M，et al. Cache replacement policies revisited：the case of P2P traffic. In：Proceedings of the 2004 IEEE International Symposium on Cluster Computing and the Grid，Chicago，Illinois，USA，2004.182-189

[5] Hefeeda M，Saleh O. Traffic modeling and proportional partial caching for Peer-to-Peer systems. *IEEE Transactions on Networking*，2008，16(6)：1447-1460

[6] Laoutaris N，Smaragdakis G，Oikonomou K，et al. Distributed Placement of Service Facilities in Large-Scale Networks. In：Proceedings of the 26th Annual IEEE Conference on Computer Communications，Anchorage，Alaska，USA，2007. 2144-2152

[7] Ye M，Wu J. Caching the P2P traffic in ISP network. In：Proceedings of the 2008 IEEE International Conference on Communications (ICC2008)，Beijing，China，2008.5876-5880

[8] Ye M，Wu J. Modeling of peer-to-peer traffic caching deployment. *Journal of Tsinghua University* (*Science and Technology*)，2009，49(1)：110-113

[9] Kamiyama N, Mori T. ISP-Operated CDN. In: Proceedings of the 28th Conference on Computer Communications (INFOCOM2009), Rio de Janeiro, Brazil, 2009. 1-6
[10] Kamiyama N, Mori T. Analyzing influence of network topology on designing ISP-operated CDN. Telecommunication Systems, DOI: 10.1007/s11235-011-9605-2, 2011
[11] Li B, Deng X. On the optimal placement of Web proxies in the Inter-net: Linear Topology. In: Proceedings of High Performance Networking (HPN1998), Vienna, Austria, 1998. 485-495
[12] Li B, Golin M J. On the Optimal Placement of Web Proxies in the Internet. In: Proceedings of IEEE Conference on Computer Communications (INFOCOM1999), New York, USA, 1999. 339-344
[13] Qiu L, Padmanabhan V N. On the Placement of Web Server Replicas. In: Proceedings of IEEE Conference on Computer Communications (INFOCOM2001), Anchorage, USA, 2001. 1587-1596
[14] Laoutaris N, Smaragdakis G. Distributed Placement of Service Facilities in Large-Scale Networks. In: Proceedings of the 26th Annual IEEE Conference on Computer Communications (INFOCOM2007), Anchorage, USA, 2007. 2144-2152
[15] Jamin S, Jin C. Constrained mirror placement on the Internet. In: Proceedings of IEEE Conference on Computer Communications (INFOCOM2001), Anchorage, USA, 2001. 31-40
[16] Krishnan P, Raz K D. The cache location problem. *IEEE/ACM Transaction on Networking*, 2000, 8(5): 568-582
[17] Radoslavov P, Govindan R. Topology-informed Internet replica placement. In: Proceedings of Web Caching and Content Distribution Workshop (WCW2001), Boston, USA, 2001. 384-392
[18] Cha M, Moon S. Placing relay nodes for intra-domain path diversity. In: Proceedings of IEEE Conference on Computer Communications (INFOCOM2006), Barcelona, Spain, 2006. 1-12
[19] 史佩昌, 王怀民, 尹刚等. 云服务传递网络资源动态分配模型. 计算机学报, 2011, 34(12): 2305-2318
[20] Ip ATS, Liu JC, Lui JCS. COPACC: An architecture of cooperative proxy-client caching system for on-demand media streaming. *IEEE Trans on Parallel and Distributed Systems*, 2007, 18(1):70-83
[21] Li W Z, Chen D X, Lu S L. Graph-Based optimal cache deployment algorithm for distributed caching systems. *Journal of Software*, 2010, 21(7): 1524-1535
[22] Kamiyama N, Kawahara R. Optimally designing caches to reduce P2P traffic. *Computer Communications*, 2011, 34(7): 883-897
[23] Kamiyama N, Mori T. ISP-Operated CDN. In: Proceedings of the 28th Conference on Computer Communications (INFOCOM2009), Rio de Janeiro, Brazil, 2009. 1-6
[24] Floyd R. Algorithm 97: Shortest path. *Communications of the ACM*, 1962, 15(6): 345-345
[25] Cho K, Fukuda K. The impact and implications of the growth in residential user-to-user traffic. In: Proceedings of the ACM SIGCOMM Conference on Data Communication (SIGCOMM2006), Pisa, Italy, 2006. 207-218

A novel P2P cache location selection and capacity allocation algorithm based on traffic-distance factor

Zhai Haibin, Zhang Hong, Liu Xinran, Wang Yong, Shen Shijun, Du Peng
(National Computer Network Emergency Response Technical Team/Coordination Center of China, Beijing 100029)

Abstract

To reduce the heavy traffic burden on the backbone network of internet services providers (ISPs) caused by the widespread peer-to-peer (P2P) application, a study on P2P cache deployment was conducted with the aim to overcome an irrational P2P cache deployment's negtive impact on the effectiveness of P2P caching. Firstly, a backbone traffic-based cache deployment model was established based on the synthetical consideration of the information of backbone network topology, renewal of cache states, etc., and then, a traffic-distance factor was defined and used to develop an approximate optimal cache deployment algorithm, the P2P cache location selection and capacity allocation (LSCA) algorithm based on the traffic-distance factor. The experimental results showed that the LSCA algorithm outperformed previous algorithms for the two typical networks of H&S and Ladder. The average backbone link utilization of the LSCA algorithm was 14% ~21% lower than that of the Degree algorithm, and 9% ~12% lower than that of the Centrality algorithm. The average transmission hop number of LSCA was 18% ~29% lower than that of Degree, and 11% ~20% lower than that of Centrality.

Key words: P2P caching technology, internet services providers (ISPs) network, traffic load, capacity allocation, cache deployment