Internet 中支配延迟的特征行为研究

李 超, 赵 海, 张 昕, 袁韶谦
(东北大学信息科学与工程学院, 辽宁沈阳 110004)

摘 要: 通过对 CAIDA 机构授权的原始海量样本数据处理得到单向链路延迟, 在此基础上计算了路径上最大的链路延迟对端到端延迟的比例以及路径上链路个数分布, 基于此定义了支配延迟. 针对链路延迟对端到端延迟的影响进行分析, 表明支配延迟之间的差异是导致端到端延迟呈现多峰分布的主要原因. 支配延迟更多地出现在路径的中间部分和 AS 自治域内部, 说明延迟瓶颈从网络的接入部分向传输部分转移, 得出传输延迟在总延迟中的影响逐渐减小, 路径中的支配延迟将主要由传播延迟决定的结论.

关键词: Internet 测量; 链路延迟; 支配延迟; 跳数
中图分类号: TP393.6 文献标识码: A 文章编号: 0372-2112 (2008) 06-1063-05

Research on the Characteristic Behavior of Internet Dominant Delay

LI Chao, ZHAO Hai, ZHANG Xin, YUAN Shao-Qian
(*School of Information Science and Engineering, Northeastern University, Shenyang, Liaoning* 110004, *China*)

Abstract: By dealing with the raw giant data samples authorized by CAIDA, we obtain one-way link delay. Together with the ratio of the biggest link delay to the end-to-end delay and the distribution of the Internet hops, the dominant delay is defined. Analysis on the impact of link delay on the end-to-end delay reviews that discrepancy of dominant delay primarily accounts for the multi-modal distribution of end-to-end delay. Besides, dominant delay tends to position more preferentially on the middle part of paths and inside the AS, which indicates that the delay bottleneck was shifted from access part to transmission part. At last, conclusion is draw that propagation delay is the major composition of the dominant delay as the effect of transmission delay declined.

Key words: Internet measurement; link delay; dominant delay; hops

1 引言

一般认为, 由于网络接入端的带宽受限导致延迟瓶颈出现在网络的边界. 随着网络接入技术的发展, 该问题已经得到很好地解决, 因此网络上延迟瓶颈的特征也在逐渐发生改变.

目前, 国内外已经有相关工作针对 Internet 的路径链路延迟进行研究. 文献[1]从 AS(autonomous systems)层面分析了 Internet 的路径链路延迟. 在文献[2]中, 采用 traceroute 服务器以对测的方式保证路由对称以获取路径上的单向链路延迟, 但由于服务器个数受限, 使其探测到的网络样本数据受到了一定程度的局限.

本文基于CAIDA(The Cooperative Association for Internet Data Analysis)授权的海量 Internet IP 层数据, 对 Internet 的链路延迟行为特征做初步分析. 首先在第二节详细讨论了 Internet 测量方法及其样本数据的获取, 在此基础上, 针对链路延迟样本数据进行分析, 并给出支配延迟的概念. 第四节分析了支配延迟对端到端延迟的影响, 第五节给出支配延迟在网络中容易出现的位置, 最后是对全文进行总结.

2 Internet 测量方法及样本数据获取

2.1 理想 Internet

网络测量是分析网络性能的必要手段. 一般来说, 测量方式有多种, 根据分类标准有主动测量和被动测量, 根据测量点的多少, 分为单点测量和多点测量[3], 本文以 CAIDA skitter[4]项目对 Internet 进行主动测量.

分布在全球范围内的 skitter 监测节点采用类似 traceroute 的方式对网络目的端发送探测包. 本文首先对探测过程中 IP 数据包经过的路径做如下形式化描述.

定义 1 路径

IP 数据包从测量节点 SRC 到被测目的节点 DST 依次经过了网络节点 $a_1, a_2, \cdots, a_{m-1}$, 令 $R=(SRC, a_1, a_2, \cdots, a_{m-1}, DST)$, 则称 R 为 IP 数据包从测量节点 SRC 到被测目的节点 DST 经过的路径.

对给定路径 $R=(SRC, a_1, a_2, \cdots, a_{m-1}, DST)$, 设相应的往返延迟时间为 $(RTT_1, \cdots, RTT_i, RTT_{i+1}, \cdots, RTT_n)$, 其中 RTT_i 为 SRC 到 a_i 的往返延迟.

由于网络的复杂性(例如目的端主机设置对探测包

收稿日期: 2007-02-09; 修回日期: 2007-09-17
基金项目: 国家自然科学基金(No. 69873007)

的 ICMP 请求不作应答或者网络繁忙等原因) 有可能造成 IP 探测包在网络目的端不可达, 并且在探测过程中, 有可能出现路径上某一跳或几跳路由器对 IP 探测包不作响应, 但是监测节点在未收到中间路由器响应的情况下, 仍可通过增加探测包的 TTL 值继续向前探测直至到达目的端为止[3]. 将这种情况下探测到的往返延迟时间称为路径不完整样本数据. 由于样本数据集存在较大程度的冗余, 针对探测结果的有效性本文对 Internet 作如下定义:

定义 **2** 理想 Internet

忽略 Internet 中目的端不可达样本, 只考虑可达样本, 在此基础上只考虑路径完整的样本数据, 称为有效样本, 以这种方式抽象出来的 Internet 称为理想 Internet.

本文仅针对理想 Internet 的有效样本数据做统计分析.

2.2 单向链路延迟

Internet 在高负载的网络情况下, 网络流量复杂多变, 加上负载平衡的应用, 沿着同一路径连续发送的分组延迟经常会有比较大的变化. 另外, 由于新的路由策略的应用, 一个分组从监测节点到达目的节点的路径与从目的节点返回到监测节点的路径并不一定相同. 因此文献[5] 认为不能通过简单的将往返延迟时间除以 2 来准确估计单向延迟时间, 其原因主要由以下两点所造成: (1) 路由不对称, (2) 双向队列不对称.

源端和目的端以 traceroute 服务器进行对测的方式可以保证路由对称, 但是这需要源端和目的端进行协作, 因此只能对以局限于 traceroute 服务器的目的端进行测量, 其探测到的路径会受到一定程度的限制.

为最小程度降低路由不对称造成的影响, 考虑到路由仍具有一定的稳定性, 本文以连续多次测量 Internet 的方式, 针对同一目的端 IP, 选取跳数和路由同时相同的路径, 通过取多次测量往返延迟时间的最小值, 消除排队延迟的影响, 尽可能减小由于双向队列不对称导致的排队延迟不对称. 在保证路由对称以及多次测量取最小值的前提下, 可以近似地认为单向链路延迟为往返延迟的一半.

2.3 有效 Internet 样本数据

目前 CAIDA 在全球范围有 20 多个监测节点对 Internet 进行多点测量, 其探测到的样本可靠性高, 权威性强. 由于目的端地址列表数量大, 本文只从时间维度上选取具有代表性的 riesling* 监测节点对网络进行连续探测, 探测时间范围从 2006 年 9 月 1 日到 9 月 30 日**.

针对测量的连续样本, 本文对其有效性做统计. 从总体上来看, 在连续探测过程中, 样本数据集比较稳定, 样本总数在 9.6×10^5 条附近波动, 有效样本总数在 3.5×10^5 条附近波动, 有效样本占总体样本的比例在 36% 左右, 说明由于各种原因导致目的端无法响应监测节点的样本占总体样本的比例接近 2/3.

在连续测量得到的有效样本中, 对同一路径多次测量的延迟时间取最小值, 在此基础上, 除去路径不完整的样本, 共得到有效路径 268 214 条, 相对于样本总数的 9.6×10^5 条数据, 有效路径大约占总体样本的 27.9%, 虽然所占比例较小, 但是绝对条数依然很大, 可以进行分析.

2.4 链路延迟样本

一般来说, 往返延迟时间 RTT_i 满足 $RTT_1<RTT_2,\ldots,<RTT_n$, 尽管通过取延迟时间的最小值将动态网络负载的影响降到最低, 但实际上仍然可能出现在测量较远路由器的时候, 网络处于较轻负载, 而较近的路由器则恰好相反, 因此会出现 $RTT_i>RTT_{i+1}$, 针对这种情况, 本文对样本数据作修正, 具体算法如下:

(1) 如果, $0<d_i-d_{i+1}<1$s, $d_i=d_{i+1}-\varepsilon(0<\varepsilon<1$, ε 为随机变量, 概率密度服从区间 $[0,1]$ 上的均匀分布).

(2) 如果 $d_i-d_{i+1}>1$s, 异常情况在某条路径中出现在 3 次以内, 且在 $d_{i-1}<d_{i+1}$ 的前提下, $d_i=(d_{i-1}+d_{i+1})$. 如果异常情况出现超过 3 次以上, 认为该条路径的延迟时间失效, 不处理该条路径.

对目的端可达且路径完整的 268 214 条数据以修正算法处理后, 共得到有效路径 257 382 条, 未处理路径 10 832 条, 有效路径占 95.96%, 有效路径所占比例较大, 认为修正后的数据能较真实的反映链路延迟情况.

3 链路延迟分析

3.1 链路延迟特征

设路径 j 的链路延迟时间为 $(\Delta t_1^j,\Delta t_2^j,\ldots,\Delta t_n^j)$, 按照从大到小的顺序对其排序, 排序后的链路延迟为 $(\Delta t_1^{j^*},\Delta t_2^{j^*},\ldots,\Delta t_n^{j^*})$, 对最大的前 m 条链路延迟求和, 令 $T_m=\sum_{i=1}^{m}\Delta t_i^{j^*}\ (1\leqslant m\leqslant n)$, 设 $P_m=T_m/T$, 则 P_m 为最大的前 $m(1\leqslant m\leqslant n)$ 条链路延迟之和占该路径的端到端延迟的比例.

对于单向链路延迟, 分别计算所有路径的 p_1, p_2, p_3, 按从小到大的顺序依次排序, 其分布如图 1 所示.

从图 1 可以看出, Internet 中超过 90% 以上的路径其最大的链路延迟占端到端延迟的 1/4 以上, 超过 50% 以上的路径其最大的链路延迟占端到端延迟的一半以

* 选取 riesling 监测节点的原因是因为 riesling 监测节点属于 CAIDA, 成立时间较早, 其数据集比较稳定.

** 由于监测节点当天未对网络进行探测, 因此存在某一天数据不存的情况, 一般不对缺失数据做补充.

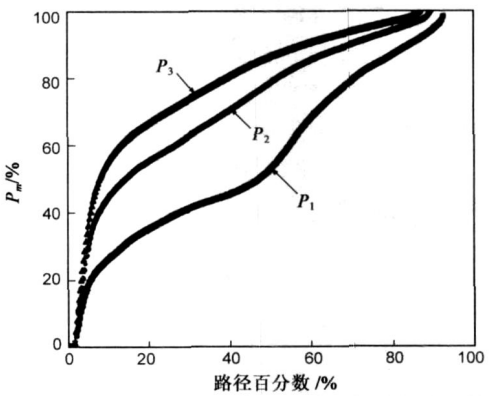


图 1　最大前 3 条链路延迟对端到端延迟比例分布情况

上, 说明总体样本中有一半的路径其某一跳消耗的延迟大于等于所有其他跳消耗的链路延迟的总和.

此外, 80% 以上的路径其最大前 2 条链路延迟之和与端到端延迟的比例超过 50%, 最大前 3 条链路延迟之和占端到端延迟的 60% 以上. 对所有路径的 p_1, p_2, p_3 求期望, 分别为 56.24%, 71.93%, 78.47%, 计算所有有效路径的最大链路延迟的均值为 41.24ms.

3.2　Internet 跳数分布

跳数反映了 Internet 上从源节点到目的节点 IP 包经过的路由器个数, 可以用跳数表示一条路径上链路延迟的个数. 对于网络拓扑特性, 跳数在一定程度上体现了某一路径的路由效率和传输质量. 图 2 给出了跳数的累计分布, 可以看出超过 95% 以上的路径其跳数都在 10 跳以上, 说明大多数 IP 数据包一般需要经过较多的中间路由节点转发才能到达目的端.

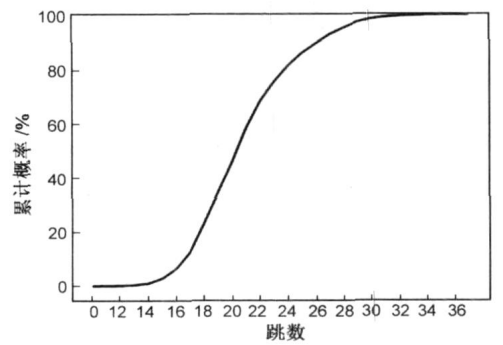


图 2　跳数的累计分布

3.3　支配延迟

从图 1 看到超过 90% 以上的路径其最大的链路延迟占端到端延迟的 1/4 以上, 根据图 2, 超过 95% 以上的路径其跳数都在 10 跳以上, 相对于最大链路延迟对端到端延迟的比例以及路径上链路延迟的个数分布来说, 可以认为路径上存在支配延迟, 对支配延迟作如下定义:

定义 **3**　支配延迟

设 Internet 上的一条路径有 n 跳, 每一跳的延迟分别为$(\Delta t_1^j, \Delta t_2^j, \cdots, \Delta t_n^j,)$, 端到端延迟为 T, $\Delta t_{\max}^j = \max\{\Delta t_i^j \mid i = 1, 2, \cdots, n\},$, 如果 $\Delta t_{\max}^j > T/4,$, 且 $n \gg 4$, 则 $D_{\max}^j$处于路径延迟的支配地位, 称 $\Delta t_{\max}^j$, 为支配延迟.

综合以上分析, 可以认为支配延迟对端到端延迟有较大影响, 有必要对链路上的延迟时间与端到端延迟的关系做进一步研究.

4　支配延迟与端到端延迟的关系

4.1　理想 Internet 端到端延迟分布

端到端延迟在一定程度上能反映网络访问速度, 表明网络整体传输质量. 对有效路径的端到端延迟作统计, 考虑到实际中端到端延迟大于 300ms 时意义不大, 且为更加突出其分布特征, 本文只对小于 300ms 的延迟时间作图, 分布如图 3.

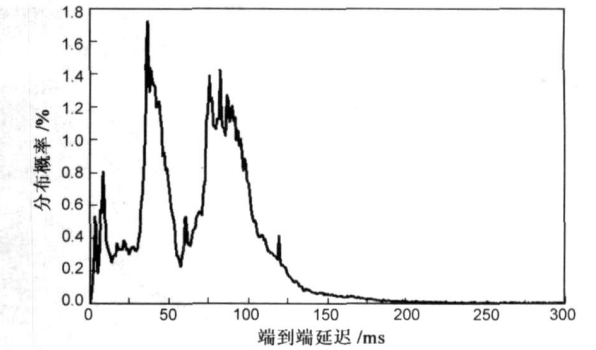


图 3　理想 Internet 的端到端延迟分布

从图 3 看到, 理想 Internet 端到端延迟服从多峰重尾分布[6]. 延迟时间主要集中在两个峰值附近, 大于 150ms 的端到端延迟只占很小比例, 说明从总体上来看, 当前 Internet 延迟性能较好.

4.2　支配延迟对端到端延迟的影响

针对端到端延迟的波峰, 分别观察端到端延迟在 [25, 50] 和 [75, 100] 路径上的支配延迟, 对其按照从小到大的顺序排序之后, 两区间上支配延迟的数值分布如图 4 所示:

图 4 表明区间 [75, 100] 有效路径上的支配延迟在总体上远大于区间 [25, 50] 的支配延迟. 更进一步, 对两延迟区间上的支配延迟取平均值, 分别为 23.88ms 和

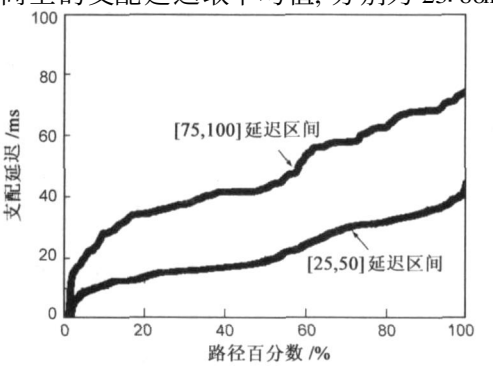


图 4　两峰值区间上有效路径的支配延迟分布情况

55.71ms, 两者相差 31.83ms. 对[25, 50] 和[75, 100] 延迟区间上的端到端延迟取均值, 分别为 41.73ms 和 85.71ms, 两者相差 43.98ms. 以上分析结果表明, 正是由于支配延迟相差较大, 且端到端延迟主要由支配延迟所组成, 导致了端到端延迟呈现出较为明显的多峰分布特征.

很显然, 尽可能地减小支配延迟是提高网络访问速度的有效手段之一. 支配延迟在网络上出现的位置对于发现网络延迟瓶颈, 指导网络规划具有重要意义. 本文针对支配延迟在网络中出现的位置做初步分析.

5 支配延迟在网络上出现的位置

5.1 相对于跳数在路径上的位置

假设支配延迟出现在路径 j 路由节点 (d_i^j, d_{i+1}^j) 之间, 并且路由节点 d_i^j 出现在路径的第 n 跳, 如果该条路径的总跳数为 N, 那么支配延迟相对于跳数在路径上出现的位置为 $ratio$, $ratio = n/N$.

针对所有的有效路径计算 $ratio$, 为清楚看到 $ratio$ 分布, 对 $ratio$ 按照小到大排序, $ratio$ 的分布如图 5.

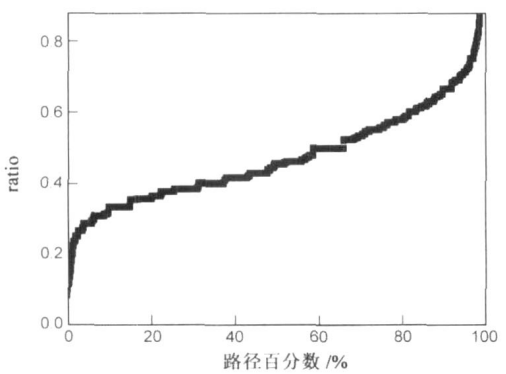


图 5 有效路径上支配延迟相对整条路径出现的位置

从图 5 看到, $ratio$ 整条曲线上升较为缓慢, 并且大多数路径的 $ratio$ 集中在 0.3~ 0.6 之间, 说明支配延迟更多的出现在整条路径的中间部分. 对所有路径的 $ratio$ 求均值后, 得到均值为 0.453, 表明支配延迟较少地出现在网络的接入端.

5.2 支配延迟与 AS 域的关系

由于探测方式本身的原因, 监测节点能得到沿途路由器接口的 IP 地址, 仅从路由层考察链路上的支配延迟, 这对于网络性能评价意义较小.

由于网络本身是由大量的 AS 自治域组成, 自治域系统是 IP 层上更高层次的抽象. 针对 BGP 支持路由策略, AS 之间存在 provider2customer, peer2peer, customer2provider 等关系[8], 将产生支配延迟的路由节点 IP 地址映射到 AS 域, 可以从较高层次上分析网络的支配延迟特征.

一般来说, 有两种常用的方法将 IP 地址映射到 AS 域, 一是通过 BGP 路由表找到对应 IP 地址的 AS 号, 二是通过 IRR (Internet Routing Registries)[9] 路由注册信息数据库进行查询. 由于 Internet 的动态复杂性可能会导致 IP 地址不断变化, 因此通过 IRR 数据库查询可能会出现 IP 地址过时等问题. 考虑到 BGP 路由表实时性较好, 本文选择 Routeviews 机构提供的 2006 年 9 月 BGP 路由表作为映射数据源.

采用 CAIDA 的 ASFinder 工具, 通过对 257382 条有效路径产生支配延迟链路的两端路由器进行映射, 发现两端路由器 IP 分布在 501 个 AS 域内. 其中, 支配延迟在同一 AS 域内的链路共 186 087 条, 占 72.3%, 位于 AS 之间的链路共 65 889 条, 占 25.6%, 两端路由器节点 IP 的任意一端在 BGP 路由表中无法查询的共 5406 条, 占 2.1%.

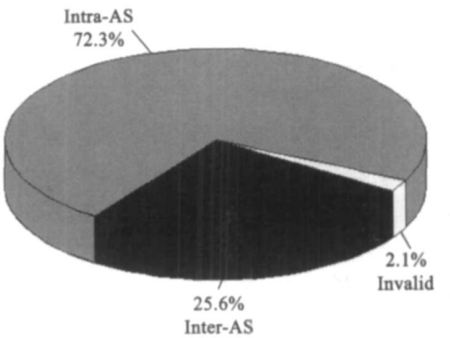


图 6 支配延迟出现在 AS 域的分布

图 6 表明支配延迟大多出现在 AS 内部, 该现象可由 Baran 提出的热土豆(Hot-Potato Routing) 路由[10] 解释. 对于提供 transit 的 AS 来说, 除非 IP 数据包到达的目的地是其 customer, 否则该数据包会尽快脱手, 因此在不同 AS 域边界支持 BGP 路由的网络节点会尽最大努力将 IP 数据包放入到输出列最短的方向上排队, 而不管数据包被传输到哪个 AS 上.

另外, 由于 IP 数据包在同一个 AS 域内会经过多个路由节点转发, 其在 AS 内部所经过的链路跳数较多. 相对于跳数相同的路径来说, 从一个 AS 域到达另一个 AS 域所经过的跳数相对总的跳数来说偏小. 因此, 尽管支配延迟出现在 AS 域之间的比例较小, 但是如果整条路径上 AS 域之间的链路个数小于 25.6%, 那么就不能说支配延迟是更容易出现在 AS 域内部还是 AS 域之间. 分析路径上 AS 域之间的链路个数相对于链路总数的比例是本文未来进一步的工作.

5.3 产生支配延迟的原因

一般的, 端到端延迟可分为三部分, 分别为传播延迟, 传输延迟, 路由器内部排队延迟. 传播延迟和传输延迟合称为固有延迟, 传播延迟与链路长度和电信号在电缆中的传播速度有关, 传输延迟与包的大小和链路带宽有关, 固有延迟主要反映了路径本身的物理特征. 排队延迟为动态延迟, 反映了延迟的可变部分, 动态延迟主

要由网络负载引起. 本文通过多次测量取延迟时间最小值, 已将动态延迟的影响降到最小程度.

为考察传播延迟的影响, 本文利用 CAIDA 的 NetGeo 工具选取一部分 4*** 产生支配延迟的路由器 a_i^j, a_{i+1}^j映射到地理位置上, 根据地理位置计算它们之间的相对距离. 结果表明, 路由器之间跨越的地理范围较大, 大部分都在 2000km~ 4000km 范围内. 对于一条长度为 4000km 的链路, 考虑到如果电信号在电缆中的传播速度为 200km/ ms, 那么传播所需时间 20ms, 相对于所有有效路径的支配延迟的均值 41. 24ms, 传播延迟大约占支配延迟的一半.

另外, 对于带宽为 10Mbps 的链路来说, 对于采用字节级的探测包, 所引起的传输延迟时间基本上可以忽略不计. 并且考虑到网络上较大的数据包一般分段传输, 并且随着网络设施的不断更新, 链路带宽的不断增长, 传输延迟在总时延中占据比例将更小.

综上所述, 链路长度将更多地成为制约网络访问速度的延迟瓶颈. 因此, 夸洲际链路和卫星通道等将是支配延迟更容易出现的地方. 对于支配延迟的解决, 应该更多地专注于提高这些链路的传输质量.

6　结论

本文对 Internet 测量过程中存在的问题做了详细讨论和分析, 并提出相应的解决方法. 通过对 CAIDA 组织授权的海量样本数据进行连续多次采样, 并做相应处理之后, 共得到有效样本路径 257 382 条, 样本数据量大, 具有较强代表性, 能较好地表现 Internet 的路径延迟特征. 在对有效路径样本分析中, 得到如下结论:

Internet 中超过 90% 以上的路径其最大的链路延迟占端到端延迟的 1/ 4 以上, 相比于路径上链路的个数分布, 可以认为大部分有效路径的端到端延迟主要由路径上最大的链路延迟引起, 并据此定义了支配延迟. 针对理想 Internet 端到端延迟呈现较明显的多峰分布, 考虑到端到端延迟主要受支配延迟的影响, 在对两峰值附近的有效路径分析之后, 认为支配延迟是导致端到端延迟呈现多峰分布的主要原因.

对于支配延迟在路径上出现的位置, 发现支配延迟较多地出现在路径的中间部分, 且以较大比例出现在 AS 域内, 并且从 AS 域的经济行为解释了支配延迟较少出现在 AS 之间的原因. 但是由于目前还无法确切知道 AS 域之间的链路个数相对于总跳数的比例, 因此还不能说支配延迟是否更容易出现在 AS 域内还是 AS 域之间.

最后, 将路由器节点 IP 映射到其所在的地理位置上, 通过计算产生支配延迟 IP 对之间的距离, 说明了链路带宽制约网络访问速度的比例较小, 链路的长度主要影响支配延迟的大小, 支配延迟主要由传播延迟所造成.

参考文献:

[1] Zeitoun A, Chuah CN, Bhattacharyya S. An AS-level study of Internet path delay characteristics[A]. Proceedings of IEEE GLOBECOM 2004[C]. Dallas, USA: IEEE, 2004. 1480- 1484.

[2] 毕经平, 吴起, 李忠诚. Internet 延迟瓶颈的测量与分析[J]. 计算机学报, 2003, 26(4): 406- 416.
BI J P, WU Q, LI Z C. Measurement and analysis of Internet delay bottlenecks[J]. Chinese Journal of Computers, 2003, 26(4): 406- 416. (in Chinese)

[3] 张宏莉, 方滨兴, 胡铭曾. Internet 测量与分析综述[J]. 软件学报, 2003, 14(1): 110- 116.
ZHANG H L, FANG B Y, HU M Z. A survey on Internet measurement and analysis[J]. Journal of Software, 2003, 14(1): 110- 116. (in Chinese)

[4] CAIDA skitter Project [EB/ OL]. http: // www. caida. org.

[5] Paxson V. End-to-end Internet packet dynamics[J]. IEEE/ ACM Transactions on Networking, 1999, 7(3): 277- 292.

[6] BRADLEY H, MARINA F, Daniel J. Distance metrics in the Internet[A]. Proceedings of the IEEE International Telecommunications Symposium 2002[C]. Natal, Brazil: IEEE, 2002.

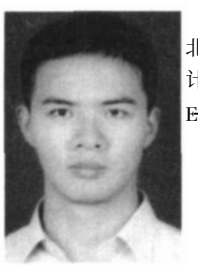
[7] HOOGHIEMSTRA G, Van P. Delay Distributions on fixed Internet Paths[R]. Delft, Netherlands: Delft University of Technology, 2001.

[8] Gao L. On inferring autonomous system relationships in the Internet[J]. IEEE/ ACM Transactions on Networking, 2001, 9(6): 733- 745.

[9] Merit IRR Services [EB/ OL]. http: // www. irr. net/

[10] Halabi S, McPherson D. Internet Routing Architectures[M]. 北京: 清华大学出版社, 2000.

作者简介:



李　超　男, 1982 年 8 月生于湖北洪湖, 东北大学信息科学与工程学院博士生. 研究方向为计算机网络、复杂网络.
E-mail: zhhai@ neuera. com

赵　海　男, 1959 年 3 月生于辽宁沈阳, 东北大学信息科学与工程学院教授、博士生导师. 主要研究方向为计算机网络、复杂网络、数据融合. E-mail: zhhai@ neuera. com

*** 4. 不能将所有 IP 都作映射的原因是因为 NetGeo 只能提供在线查询, 数据库不具有查询几十万条数据的能力.