

公用IPv6网络承载服务SLA监测研究

成卫青^{1,2}, 龚俭¹

¹(东南大学 计算机科学与工程学院, 江苏 南京 210096)

²(南京邮电大学 计算机学院, 江苏 南京 210003)

摘要: 分析了 NGI 中网络承载服务 SLA(NSLA)的应用需求, 针对 NSLA 可能的应用场合及 IPv6 特点定义了适用于 NSLA 的 QoS 参数: IP 分组传送延迟(IPTD)、IP 分组延迟变化(IPDV)、IP 分组传送失败率(IPFR)和 IP 分组吞吐能力(IPTC), 其中 IPTC 能有效刻画公用 IP 网段集维持业务流量模式的能力。分组识别和时钟同步是 IPv6 网络 NSLA 监测的两个关键问题。文中给出了分组识别的解决思路; 为降低对同步的要求、支持较低成本的 NSLA 监测, 除 IPTD 外其余参数均可容忍出入测量点之间存在有限的时钟不同步。此外, 综合应用特点、用户感知、IPv6 协议等因素, 讨论了面向应用 NSLA 的 QoS 参数设定。最后设计了一个包括采集器、前端机、分析机、原始数据库和结果库等部件的 NSLA 监测系统, 采用 ON-OFF 系统抽样方式被动监听自然流量, 在各部件的协作下可实现各 QoS 参数的测量。

关键词: 服务质量;网络承载服务 SLA; IP 分组吞吐能力; IP 分组延迟变化; IP 分组传送失败率

1 引言

SLA(Service Level Agreement)^[1-3]是用户和服务供应者就服务特性、责任和优先权商定的一份正式合同, 可包括性能、资费、服务交付以及经济补偿等方面的陈述。下一代互联网(NGI)的主要目标之一是为端用户提供 SLA 服务, 互联网上常用的客户-服务器(C/S)模式的应用服务需要应用服务器端、核心网络和用户端三个方面的支持, 端用户与应用服务供应商(ASP)签订的 SLA 称为 EASLA^[4]。核心网络由分别属于不同网络运营商管辖的多个公用 IP 网络(段)组成, EASLA 需要有关网络运营商的支持, 也即需要公用 IP 网络提供有具体参数和指标要求的承载服务, 称为 NSLA 服务。根据服务目的, NSLA 可分为两种: 一种用于约定公网对指定两网络段间流量的**通用**承载服务质量, 一种用于约定公网对指定两网络段间指定应用流量的**面向应用**承载服务质量。

为支持 NSLA, 网络运营商必须使网络提供能根据网络地址或应用类别等因素区别服务的 QoS 保证机制, 还必须能够定量测量 SLA 中的 QoS 参数以了解服务合同的执行情况(即实际服务的 SLA 遵从程度), 测量结果既可作为收费依据提供给用户, 也可促进运营商自己总结经验不断提高区分服务的能力与效率, 增强竞争力。ITU-T 和 IETF 等机构对 IPv4 网络性能测量做了较多研究^[5-13], 但有些性能参数的定义仍有改进空间, 比如 IP 分组延迟变化, 再如吞吐量, 已有吞吐量相关参数都不能够反映“IP 网络维持给定 IP 分组传送速率的能力”^[5]。此外, 对 NSLA 的应用场合尚无详细的建议; 对 NGI 必须解决的问题之一——IPv6 网络性能监测缺乏系统研究。因此本文以 IPv6 网络的 NSLA 监测为主题, 具体研究内容为: 分析 NSLA 应用需求, 说明 NSLA 监测的重要性和必要性; 基于 NSLA 可能的应用场合与 IPv6 特点定义若干适用于 NSLA 的 QoS 参数, 讨论应用分类并给出面向应用 NSLA 的 QoS 参数, 解决测什么的问题; 设计一个 NSLA 监测系统, 讨论如何测的问题。

本课题得到国家“九七三”计划课题(2003CB314804), 国家 863 计划课题(2005AA103001-1)资助。成卫青, 副教授, 主要研究方向为网络测量与网络行为学; 龚俭, 教授, 博士生导师, 主要研究领域为网络行为学、网络安全。

2 NSLA 应用需求

下一代互联网中, 用户为获得有 QoS 保证的应用服务可能会与 ASP 签订 EASLA, 见图 1。为保证合同能够履行, ASP 需要两方面的支持: 一方面 ASP 将使应用服务器具备区分服务能力, 另一方面它还可能与其网络服务供应商(NSP, 网络运营商担当)签订 NSLA(如图 1 中的 NSLA₁)以使应用流获得有保证的网络传输质量。为完成 NSLA₁ 合同, NSP3 可能需要向其他相关 NSPs 外购有质量担保的网络服务, 可采用迭代法与其分别签订 NSLA(如 NSLA₂, NSLA₃)约定单个网络段的传送质量, 也可以采用递归法先与 NSP2 签订 NSLA 令 NSP2 担保 NSP2 至用户网络 1 外围的网段集(如图 1 中 P1 与 P2 两点之间)的传送质量, NSP2 再与 NSP1 签订 NSLA 等。此外, 对于联系较紧密的两个用户网络, 如果双边用户希望他们之间通信质量有保证, 则可以结盟与某个 NSP(如 NSP2)签订 NSLA(如 NSLA₄), NSP2 再采用迭代法或递归法与其他相关 NSPs 外购服务。

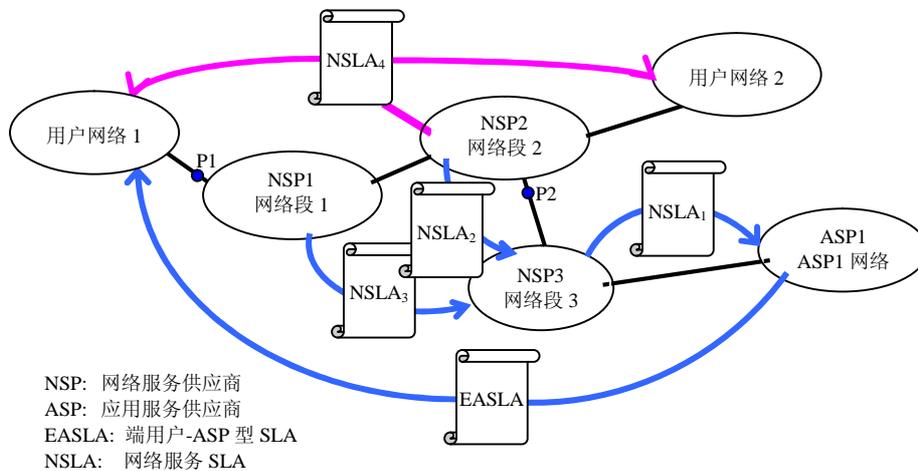


图 1 NSLA 举例

综上所述 NSLA 既可能是互联网上用户(组织)与网络运营商之间的服务协议, 也可能是两个网络运营商之间签署的服务协议, NSLA 有着广泛的应用需求。

3 QoS 参数定义

要提供 NSLA 服务, 首先必须确定 QoS 参数, QoS 参数应能反映公用 IP 网络的承载服务的服务质量, 可从速度、准确度、可靠性三个方面考虑^[5]。常见的反映网络性能可作为 NSLA 参数的 QoS 参数有: 时延、时延抖动和丢包率。ITU-T Y.1540 建议^[5]定义了如下若干 IP 分组传送性能参数: IP 分组传送延迟、IP 分组延迟变化、IP 分组差错率、IP 分组丢失率、每秒出现的乱 IP 分组数、IP 分组严重丢失块百分比、IP 分组吞吐量和基于 8 位组的 IP 分组吞吐量, 其中传送延迟和吞吐量反映速度, 差错率反映准确度, 延迟变化和丢失率反映可靠性。Y.1540 建议指出其定义的参数可用于刻画 IPv4 网络所提供的 IP 服务, 而 IPv6 网络的 IP 服务性能参数要做进一步研究, 这正是本文的主要研究内容。

在定义 NSLA 的 QoS 参数之前先定义几个适于 IPv6 网络的术语。

(1) 网络段和网段集

在同一(或协作)管辖权下提供源宿主机之间一部分 IP 服务的一组主机及其所有互连交换链路构成一个网络段(network section, NS)^[5]。一个网段集(NSE)指 1 个或多个连通的网路段及其互连交换链路。

(2) 进入事件(ingress event) / 外出事件(egress event)

IP 分组进入(离开)一个 NSE 事件的简称, 当一个 IP 分组穿过一个入(出)测量点(入/出 MP, 见图 2)进入(离开)一个 NSE, 并且分组头内源和目的地址为期待的源和目的主机的 IP 地址时, 发生 IP 分组传送进入(外出)事件。

(3) 成功的 IP 分组传送输出

IPv6 规定 IPv6 分组只可以在源主机处分段^[14], 在接收主机处进行重组, 分组传递路径中的路由器不能对分组进行分段或重组, 因此在网络中 IPv6 分片可看成独立的分组。如果在某许可的入测量点 MP₀ 处的一个 IP 分组传送进入事件, 在随后的 D_{max} 内导致在某许可的出测量点 MP_i 处发生一个相应外出事件, 并且递交的 IP 分组中所有不随传送而改变的部分(不变部分)与原分组的完全一致, 即发生了一个成功的 IP 分组传送输出。分组从入 MP 穿越 NSE 到达出 MP 所经历的延迟大于 D_{max} 时当作丢失处理, D_{max} 一般设为 3s^[5]。

(4) 失败的 IP 分组传送输出

IPv6 头比 IPv4 头新增了流标签字段, 少了标识和头校验和字段, 只在分段扩展头中有标识字段。对于 IPv4 网络, 根据 IPv4 头能够分辨在出测量点采样到的 IP 分组与在入测量点采样到的某分组(原分组)是否为同一分组或为其分片, 若能收齐所有分片则可拼接信息字段再与原分组的信息字段比较, 即能辨别分组经过 NSE 传送后是正确、出错还是丢失。对于 IPv6 网络, 除非都是分片分组否则无法仅根据 IPv6 头和扩展头区分一个流内的各个分组, 只有检查 IP 分组所有(或部分)不变部分及采样时间才能较好区分各分组。而一旦某分组的不变部分在传送中发生差错, 在出测量点无法辨别(或代价太大)它究竟该对应于在入测量点采集到的哪个分组。因此对于 IPv6 网络性能测量应统一处理分组传送出错和丢失, 如果在某许可的入 MP 处的一个 IP 分组传送进入事件, 在随后的 D_{max} 内没有导致在许可的出 MP 处发生一个相应的正确的传送外出事件, 即发生了一个失败的 IP 分组传送输出。

3.1 IP 分组传送延迟(IPTD)

对穿越指定 NSE 的所有感兴趣 IP 分组的成功的分组传送输出计算 IP 分组传送延迟(IPTD)。IPTD 是两个相应 IP 分组传送参考事件发生的时间差, 设在 t₁ 时刻发生进入事件, 在 t₂ 时刻发生外出事件, 则分组 i 的传送延迟

$$IPTD_i = \begin{cases} t_2 - t_1 & \text{if } t_2 - t_1 \leq D_{\max} \\ \text{无} & \text{if } t_2 - t_1 > D_{\max} \text{ (失败)} \end{cases}$$

测量该参数需要解决两个关键问题: (1) 维持各处高速采集器时钟同步, 已有经验表明通过 NTP 等纯软件同步方法效果不佳^[15], 综合考虑成本和效果, 实现时钟同步尤其是大规模分布式测量所需要的时

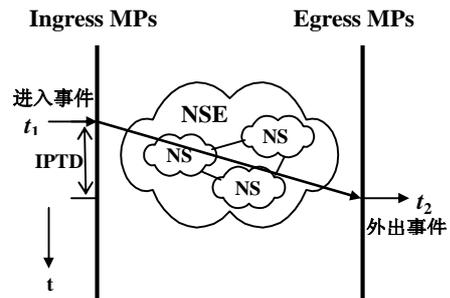


图 2 IP 分组传送延迟测量

钟同步，可行方法为借助外部高稳定性和高准确度时钟源，比如 GPS 时钟，其误差小于 1 μ s，但由于它并不总靠得住，还可利用 GPS / GLONASS 双模接收机、GPS / 北斗双模终端进行时钟同步。测量 IPTD 必需测量点时钟精确同步。(2) 计算两点测度必需的分组识别，即辨别在出入测量点采样到的 IP 分组是否为同一分组。

关于**分组识别**，IPv4 分组在 IPv4 网络中传送时可能会被分段，因此可用源 IP 地址、标识、标志、段偏移和目的 IP 地址(可选)字段来识别一段时间内的各个分组。IPv6 关于分组的分段与重组的规定有利于分组识别，不利的是 IPv6 头仅有流标签，流内分组没有其他标识，只在分段扩展头中有标识字段，封装安全载荷(ESP)扩展头有序号字段。为简单快速辨别流经两测量点的分组是否为同一分组，可以在采集时对监听到的每个 IP 分组不变部分的全部或部分进行特定运算，如 32 位 CRC 运算，得到该**分组的摘要**。如果核心 IPv6 网络中不设安全网关^[16]，则 IPv6 分组在核心网络传送途中不变的部分包括：IPv6 头中的版本号、流标签、负载长度、下一个头、源地址，分段扩展头，认证扩展头，ESP 扩展头和**高层协议数据**。**分组的摘要**连同源地址、目的地址(指最终目的地址，不包含路由扩展头时即为 IPv6 头中的目的地址)、流标签一起可作为**分组的识别标记**。

IPTD 统计参数有：(1) 平均 IP 分组传送延迟，定义为在一段时间内一组感兴趣分组的 IPTD 的算术平均；(2) IPTD 的下侧 a 分位数，例如下侧 95%分位数即为在一段时间内一组感兴趣分组的 IPTD 中的 95%最大值。

3.2 IP 分组延迟变化(IPDV)

如果测量设备间支持时钟同步，则一段时间内单个感兴趣分组的延迟变化可定义为分组的实际传送延迟与额定延迟的差值，可使用这段时间内所有成功传送分组的传送延迟的算术平均、最小延迟或任意分组的传送延迟作为额定延迟。

如果出入测量点设备之间不支持精确的时钟同步，但测量设备的时钟具有极小的相对脉冲相位差(relative skew)和极小至可忽略的时钟漂移(drift)，即时钟频率基本一致且高稳定，则经过同一对测量点穿越 NSE 的单个分组的 IPDV 的测量方法如下所述。在一段时间内在两测量点分别按各自的时钟记录感兴趣分组的到达时间，设共有 n 个成功传送的分组，分组 k 到达测量点 1(MP1)的时间为 $a_{1,k}$ ，到达测量点 2(MP2)的时间为 $a_{2,k}$ ，以分组 r ： $a_{2,r} - a_{1,r} = \min\{a_{2,i} - a_{1,i}, i = 1, \dots, n\}$ 作为参考分组。考虑到两点不精确同步，在 MP2 应在更长的时间范围内（详见 3.3 小节）监听从 MP1 来的分组，避免监测造成分组传送失败误判。另外，因时钟不同步， $a_{2,r} - a_{1,r}$ 不一定是正数，并不能作为分组 r 的 IPTD。单个分组 k 的延迟变化 $IPDV_k$ 定义为：

$$IPDV_k = (a_{2,k} - a_{2,r}) - (a_{1,k} - a_{1,r}) \quad (1)$$

下面分析该测度定义的有效性和适用性。

为描述方便，称 MP1 的时钟为时钟 1，MP2 的时钟为时钟 2。设时钟 1 的时钟频率与理想时钟频率的比值为 k_1 ，例如每 24 小时慢 5.184 秒(即每秒慢 60 微秒)的时钟， $k_1=0.99994$ ；

设时钟 2 的时钟频率与理想时钟频率的比值为 k_{f2} 。

设 MP1 监听到分组 r (k_{f1} 和 k_{f2} 相等时, 分组 r 为最小传送延迟分组) 的时刻, 真实时间(理想时钟计时)为 T_r , 时钟 1 的时间为 $a_{1,r}$, 时钟 2 的时间为 $a_{1,r} + \Delta C_r$, ΔC_r 为时钟 2 在 T_r 时刻相对于时钟 1 的偏移量(offset)^[7], 即两时钟各自报告的时间之差, 也等于两时钟各自相对于理想时钟的偏移量之差。经过 t_r 时间(理想时钟计时) MP2 监听到分组 r , 此时真实时间为 $T_r + t_r$, 时钟 1 的时标为 $a_{1,r} + t_r * k_{f1}$, 时钟 2 的时标为 $a_{1,r} + \Delta C_r + t_r * k_{f2}$, 也即 $a_{2,r}$ 。对于分组 k 同理, 如图 3 所示, 则

$$(a_{2,k} - a_{2,r}) - (a_{1,k} - a_{1,r}) = \Delta C_k + t_k \times k_{f2} - \Delta C_r - t_r \times k_{f2} \quad (2)$$

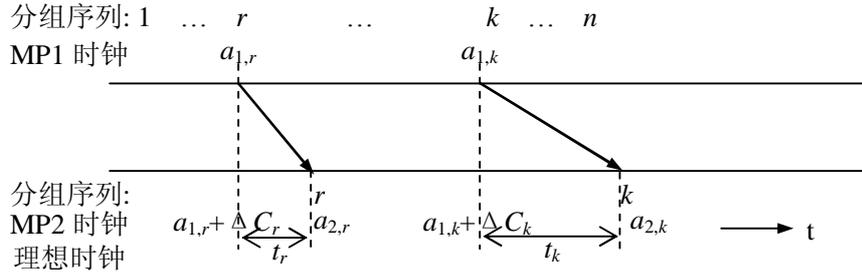


图 3 时钟不同步情况下 IP 分组延迟变化计算

对于 MP1, 时钟 1 记录分组 k 和分组 r 到达时间相隔为 $a_{1,k} - a_{1,r}$, 而真实时间间隔为 $(a_{1,k} - a_{1,r}) / k_{f1}$; 分组 k 到达 MP1 的时刻, 时钟 2 的时间为 $a_{1,k} + \Delta C_k = a_{1,r} + \Delta C_r + ((a_{1,k} - a_{1,r}) / k_{f1}) k_{f2}$, 可得 $\Delta C_k - \Delta C_r = (a_{1,k} - a_{1,r}) (k_{f2} - k_{f1}) / k_{f1}$, 由此式(2)化为

$$(a_{2,k} - a_{2,r}) - (a_{1,k} - a_{1,r}) = (t_k - t_r) \times k_{f2} + (a_{1,k} - a_{1,r}) \frac{k_{f2} - k_{f1}}{k_{f1}} \quad (3)$$

由式(3)推导得到式(1)的 $IPDV_k$ 与真实值 $(t_k - t_r)$ 之间的误差:

$$IP\tilde{D}V_k = (t_k - t_r) \times (k_{f2} - 1) + (a_{1,k} - a_{1,r}) \frac{k_{f2} - k_{f1}}{k_{f1}} \quad (4)$$

式(4)表明误差与两时钟的相对偏移量无关, 这一点很有意义, 因为除非借助外部的精准时钟, 两个现实时钟的相对偏移量实际上是很难准确估计的^[15]。不过为降低存储与分组识别的代价, 应保证两时钟在任意时刻相对偏移量小于数秒(如 2s), 每隔数小时利用 NTP 做一次调整即可做到。设 $k_{f1} = 0.999994$, $k_{f2} = 1.000006$, $a_{1,k} - a_{1,r} = 300s$, 当真实值为 50ms 时, 由式(4)得到误差约为 36ms, 且式中后一分量占 99.991668%。故式(1)的适用条件是: 两时钟频率稳定, 相对脉冲相位差极小, 并且时钟频率的准确度较高; 当 $|k_{f1} - k_{f2}| \leq 0.000005$, 两时钟每 24 小时比真实时钟快慢不超过 5s, $a_{1,k} - a_{1,r} = 300s$ 时, 时延抖动真值为 10ms、20ms、50ms 时误差均小于 1.503ms。误差最主要因素是相对脉冲相位差, 其次是时钟频率的准确度, 实际测量值中还包括测量主机的中断响应机制(并不一定是每个到达的分组产生一个中断, 而很可能是多个接连到达的分组产生一个中断, 因而这几个分组的到达时间相同)引入的误差。

当时钟频率稳定但其他条件不满足时，可根据式(3)将 $IPDV_k$ 定义为：

$$IPDV_k = \frac{(a_{2,k} - a_{2,r})}{k_{f2}} - \frac{(a_{1,k} - a_{1,r})}{k_{f1}} \quad (5)$$

式(5)理论值就是真值，但 k_f 的估计、测量主机的中断响应机制、无法分辨在入 MP 接连出现的多个相同分组在出 MP 各自的到达时间等因素使实际测量值可能存在误差。

IPDV 统计参数有：(1) 平均 IP 分组延迟变化，即在一段时间内一组感兴趣分组的各自分组延迟变化的算术平均；(2) 基于延迟变化间隔的统计参数，计算一段时间内一组感兴趣分组各个 IPDV 落入指定延迟变化间隔的百分比^[5]；(3) 基于分位数限制的统计参数，定义为在一段时间内一组感兴趣分组的延迟变化分布中上侧 a 分位数与下侧 a 分位数之差^[5]，例如 a 为 1%；(4) IPDV 的下侧 a 分位数，例如 a 为 98%。

3.3 IP 分组传送失败率(IPFR)

IP 分组传送失败率定义为一段时间内感兴趣分组中失败的 IP 分组传送输出总量占传输的 IP 分组总量的比例。设出入测量点时钟相对偏移量为 ΔC ，两测量点按各自时钟在 $[t, t+T]$ 时段内采集感兴趣分组，且 $T \gg D_{max}$ ， $T \gg |\Delta C|$ ，则 IPFR 定义为：

$$IPFR = \frac{\|P\| - \|P_{succ}\|}{\|P\|} \quad (6)$$

其中， P 为在 $[\max(t - \Delta C, t), \min(t + T - D_{max} - \Delta C, t + T)]$ (入 MP 时钟计时)时段内到达入 MP 的感兴趣分组集合，内含分组数用 $\|P\|$ 表示，成功到达(正确且延迟 $\leq D_{max}$)出 MP 的 P 中分组构成 P_{succ} 。因 ΔC 是变量，在 T (如 5 分钟)时段前后的 ΔC 相差一般小于 1.5 毫秒，在实际测量中，为避免不断估计 ΔC ，可大致估计出时钟相对偏移量最大绝对值 ΔC_{max} (可取得稍大)，简化用在 $[t + \Delta C_{max}, t + T - D_{max} - \Delta C_{max}]$ 时段内到达入 MP 的分组构成集合 P ，以在 $[t, t+T]$ (出 MP 计时)时段内正确到达出 MP 的 P 中分组构成 P_{succ} 。实际测量中的 P_{succ} 与真实集合可能有误差，主要干扰是：相邻时段若存在相同分组，可能导致成功传送的误判；超时正确到达的分组也可能当作成功到达了。

3.4 IP 分组吞吐能力(IPTC)

关于吞吐量，IETF 在 RFC3148 中提出一个测量在单个拥塞感知的运输连接(如 TCP)上网络传送大量数据的能力的量度——整批运输容量(Bulk Transport Capacity, BTC)，其直观定义是单个连接上的长期平均数据率(bps)，较适于估计网络对 FTP、传递大图像或大文档时的 Web 等应用的服务质量^[11]。但是由于拥塞控制算法对 BTC 有影响，而不同的 TCP 实现可使用不同的标准拥塞控制算法，因此作为网络和端系统共同作用的结果，BTC 不适合作为网络 QoS 参数。ITU-T 在 Y.1540 中定义了 IP 分组吞吐量参数(IPPT, 定义为每秒成功传送的 IP 分组数)和基于 8 位组的 IP 分组吞吐量(IPOT, 定义为每秒成功传送的 IP 分组的 8 位组数)；还给出了使用吞吐量探针(一串 576 字节的 IP 分组)测量源宿主机之间可用网络容量的方法。网络可用容量与用户实际流量并不相关，因而它若作为 SLA 参数，其遵从性验证必须采用如上的增加网络负载且不宜大规模使用的主动测量方法，因此它也不适合作为 NSLA

参数。

吞吐量其实与两端主机的性能及其使用的拥塞控制算法、应用程序的流量模式、沿途链路带宽、网络负载等因素均有关系。为刻画公用 IP 网段集传送具有不同流量模式的感兴趣分组并维持业务流量模式的能力，本文定义 IP 分组吞吐能力(IPTC, IP packet throughput capability)参数；该参数能够不受末端主机与末端链路带宽的影响，不局限于单个连接的容量测量，支持固定比特率和可变比特率的流量模式，而且可采用被动测量方式测量。

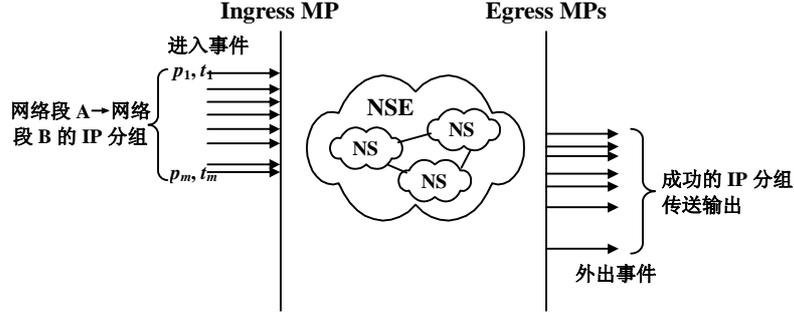


图 4 IP 分组吞吐能力测量

设感兴趣的分组为从网络段 A 发至网络段 B 的 IP 分组（见图 4），以入 MP 计时，在 $[t, t+T]$ 时段的 $[t+\Delta C_{\max}, t+T-D_{\max}-\Delta C_{\max}]$ 时间段内（参见 3.3）到达入 MP 的 IP 分组序列按到达先后排序为 $P = \{p_1, p_2, \dots, p_m\}$ ，相应到达时间为 $t_p = \{t_1, t_2, \dots, t_m\}$ 。在 P 中寻找满足下列条件的子序列

$$P_{i_s, j_s}(q_t, q_{num}) = \{p_k : 1 \leq i_s \leq k \leq j_s \leq m; t_{k+1} - t_k \leq q_t, \forall k \leq j_s - 1; \\ t_{i_s} - t_{i_s-1} > q_t, \text{if } (i_s > 1); t_{j_s+1} - t_{j_s} > q_t, \text{if } (j_s < m); j_s - i_s + 1 \geq q_{num}\}$$

子序列中，相邻分组到达间隔均小于阈值 θ_t ，分组总数大于阈值 θ_{num} ； P 中可能包含若干这样的子序列，选择分组总数最多的子序列 $P_{n_1, n_2}(q_t, q_{num})$ 计算一定时间间隔内 NSE 对指定分组的 IP 分组吞吐能力：

$$IPTC = \begin{cases} EIPT/IIPT, & \text{if } IIPT \leq CMIIPT \\ EIPT/CMIIPT, & \text{if } IIPT > CMIIPT \end{cases}, \text{ 其中}$$

1) 入 IP 分组吞吐量(IIPT)定义为： $IIPT = (n_2 - n_1)/(t_{n_2} - t_{n_1})$ 。

2) 出 IP 分组吞吐量(EIPT)

P 中 IP 分组经过 NSE 传送将经由许可的 1 个或多个出 MPs 到达网络段 B。各出测量点时钟要求精确同步，出入测量点时钟要求频率一致稳定但允许存在有限的相对偏移量。以出测量点计时，设在 $[\max(t, t_{n_1} - \Delta C_{\max}), \min(t+T, t_{n_2} + \Delta C_{\max} + D_{\max})]$ 时段内 P_{n_1, n_2} 中 IP 分组共有 x 个正确到达出测量点，按到达先后排序为 $P_e = \{p_{e_1}, p_{e_2}, \dots, p_{e_x}\}$ ，其中 $\forall p_{e_i} \in P_{n_1, n_2}$ ，相应

到达时间为 $t_{P_e} = \{t_{e_1}, t_{e_2}, \dots, t_{e_x}\}$ 。出 IP 分组吞吐量定义为： $EIPT = (x-1)/(t_{e_x} - t_{e_1})$ 。EIPT 有可能大于 IPTC，尤其当有分组传送失败时，因此计算 IPTC 同时应对同一批分组再计算 IPFR = $(\|P_{n_1, n_2}\| - \|P_e\|) / \|P_{n_1, n_2}\| = (n_2 - n_1 + 1 - x) / (n_2 - n_1 + 1)$ 。实际测量中若子序列 $P_{n_1, n_2}(q_i, q_{num})$ 与其前一序列有相同分组，则该子序列的 EIPT 和 IPFR 都很可能存在误差。

3) 约定的最大入 IP 分组吞吐量(CMIIPT)

NSE 承诺当指定分组进入 NSE 的速率不超过 CMIIPT 时为其提供承诺的服务质量。

包含多个入测量点时，对每个入 MP 分别按上述方法计算 IPTC。

4 IPv6 网络承载服务 SLA 监测

4.1 通用 NSLA 监测

上节定义的参数可作为公用 IPv6 网络通用 NSLA 的 QoS 参数，并且各参数应分两个方向分别测量。

4.1.1 测量点布设与测量方法

测量点应设在网段集的边界。假设末端网络段 A 发向末端网络段 B 的 IP 分组可经交换链路 A1 或 A2 进入位于 A 和 B 网络段之间的某公用网段集 NSE1，再经三个出口链路离开 NSE1 到达网络段 B。为测量 NSE1 对始发于网络段 A 以网络段 B 为终点的所有 IP 分组的承载服务质量，将包括 2 个入测量点和 3 个出测量点，如图 5 所示，反向的同理。如果各测量点时钟能保持同步，则能够测量上述所有 QoS 参数；当不支持精确同步时，如果仅有一个出测量点，则可对各入测量点分别计算除 IPTD 之外的其余 QoS 参数。

测量方法可选用被动监听自然流量直接测量参数的方法，为降低测量成本并使计算可行可采用按时间 ON-OFF 方式系统抽样，采集器在一个测算周期(见图 6)内的测量时间间隔内以概率 1 采样感兴趣的 IP 分组。一个测量点由高速采集器和前端机^[4]组成。采集器根据采集规则(源于 NSLA 合同中限定分组属性的参数，如端网络段 IP 地址)过滤流过的 IP 分组，将符合条件的 IP 分组的识别标记(见 3.1 节)及其到达时间经专用信道发送到前端机暂存，前端机将暂存的数据依据合同需要整理，连同合同编号和测量点编号等信息，经专用或公用互联网信道存入本测量域(见下小节)的某原始数据库中。



图 5 公用 IP 网络 SLA 监测系统测量点设置

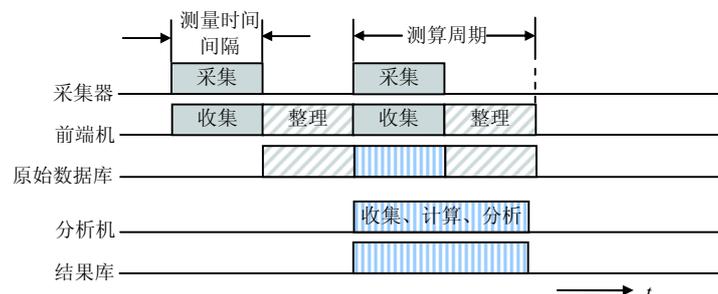


图 6 NSLA 监测系统各功能部件工作时序

4.1.2 监测系统与监测网络

如果大规模实施 NSLA 管理，可能有多份 NSLA 需要在某公用网络段的某中继链路上设置测量点，为合理利用资源，可使同一测量点服务于若干具体 NSLA 合同。根据所属管辖单位，测量点归属不同的测量域，一个测量域内包含多个出入测量点和若干数据库。除采集器、前端机、存放原始数据的数据库外，NSLA 监测系统还需要分析机和存放结果的数据库。根据 NSLA 合同、测量点与数据库分布信息，分析机可先从入测量点相关的原始数据库获得进入 NSE 的 IP 分组信息，再在出测量点相关的原始数据库中查找相应 IP 分组信息，计算 NSLA 参数值，进而可分析合同遵从程度，并将 NSLA 参数值和遵从程度存入结果库中。由于参数的计算需要协同出入测量点的原始数据，因而当 NSLA 管理大规模实施时将占用相当的带宽，若在原有互联网上传输测量数据将不时影响正常业务流量，因此若仍能使测量成本仅占网络服务供应成本的一小部分，可建立专门的 NSLA 监测网络互联前端机、原始数据库、分析机和结果库。

4.2 面向应用 NSLA 监测

4.2.1 面向应用 NSLA 的 QoS 参数

通用 NSLA 监测是测量公用 IP 网络对指定源和目的 IP 地址的 IP 分组的传送性能，并不管 IP 分组里装载的是什么数据。实际上，对不同质量需求或不同种类的应用，公用 IP 网络最好能区别服务。因而对于面向应用的 NSLA，其 QoS 参数的选择应该综合考虑下列因素：应用的特点，用户感知，引起应用交付质量下降的网络因素和 IPv6 协议特点等。

关于互联网应用，根据应用协议所实现的事务的类型，可分为 3 种应用^[17]：面向事务的，如 Web，面向吞吐量的，如 FTP，和面向流型的，如实况视频。根据交互性和实时性可分为：实时性要求最高的**会话式**应用，如会话式语音/视频业务、交互式游戏、Telnet；有一定实时要求的**响应式**应用，如 Web、Email 访问；要求及时的非会话应用，如流式音频/视频、信息速递、FTP；及非紧急后台应用，如传真、Email 服务器间邮件传送。应用传送的媒体分为音频、视频、文本、图形、静止图像和数据^[18]。音视频一般容错，但容错程度因应用不同、压缩格式不同有所不同，例如话音业务的容错程度比高质量音频(音乐)的要高；其他媒体一般不容错，因此应用还可分为容错性和非容错性应用。另外，网络应用一般都涉及两个方向的流量，但除会话式应用外其余应用的双向流量一般不对称，多以服务器到客户方向流量为主，如流式音视频和 Web，也有以客户到服务器方向流量为主的，如 Email 访问。

影响用户感知的主要参数是：延迟、延迟变化、信息丢失和吞吐量^[19]。就协议特点来说，IP 协议提供无连接的 IP 数据报服务，实现的是一种缺少规划的异步时分复用，因此 IP 分组传送延迟具有较大随机性；另外网络瞬时高负载和路由器有限的 CPU 与缓存资源的矛盾易造成分组丢失。综上，设定公用 IPv6 网络对下列应用的承载服务的 QoS 参数：

表 1 面向应用 NSLA 的 QoS 参数

应用类别	NSLA 参数	方向性	IPTD	IPDV	IPFR	IPTC
事务类(如 Web 浏览)		S→C	√		√	√

文件传输类(如FTP)	S→C	√	√	√	√
会话式面向流类(如会话式语音/视频)	双向	√	√	√	*
非会话式面向流类(如高质量流式音频 ¹⁸⁾ , 流式视频)	S→C		**	√	√

注: * 对于会话式视频应用, 包括IPTC参数

** 对于高质量流式音频应用, 包括IPDV参数

4.2.2 监测系统

面向应用 NSLA 的服务对象是指定的应用流量, 因此 NSLA 监测是测量公用 IP 网络对指定源、目的 IP 地址和应用类别的 IP 分组的传送性能。如果每个应用服务器只提供一种应用, 则区分 IP 地址也即区分了应用, 那么面向应用 NSLA 的监测系统与通用 NSLA 的基本一致, 不同之处是分析机应对不同应用计算不同的参数子集; 否则监测系统需要增加专门的应用流量识别模块。以前常通过保留端口号识别应用, 但现在存在大量使用伪装端口、未知的自由端口, 或基于约定端口号协商本次使用的端口号的应用, 因此现在也常利用动态端口解析、应用特征代码匹配等方法识别应用^[20~21]。虽然应用识别难度可能越来越大, 但只要合同中给定应用的具体特征(如端口号、应用特征代码、应用流量特征等)与识别方法, 通过修改采集器的过滤规则和原始数据的保存规则, 调整前端机的过滤整理规则, 就一定可以区分不同的应用流量。

5 总结

本文探讨了 NGI 中 NSLA 的应用需求; 基于 NSLA 可能的应用场合与 IPv6 特点定义了适用于 NSLA 的 QoS 参数: IP 分组传送延迟(IPTD)、IP 分组延迟变化(IPDV)、IP 分组传送失败率(IPFR)和 IP 分组吞吐能力(IPTC); 在分析应用特点的基础上, 为一些常用应用类设定了面向应用的 NSLA 的 QoS 参数; 并讨论了 NSLA 监测系统各功能部件的作用及其协作工作流程。本文主要贡献是: 给出了 IPv6 网络测量必须解决的分组识别问题的解决思路; 定义了一个能够不受末端主机等其他因素影响, 公正度量公用 IP 网段集维持业务流量模式的能力的参数——IPTC; 详细讨论了时钟不同步给 IPDV 带来的影响, 而且本文讨论参数详细定义时除 IPTD 外均兼顾了出入测量点时钟不精确同步的情形, 可供不同代价的 NSLA 监测参考。下一步工作是以计算简便快速、冲突低、节约空间和重要字段可恢复(可选)为目标, 研究应选择 IPv6 分组的哪些不变部分及使用何种算法来生成分组识别标记; 研究各种应用特征实现 Web、FTP、BT 等常用应用流量的快速识别。

参考文献:

- [1] ITU-T Study Group 2. Framework of a service level agreement (ITU-T E.860) [EB/OL]. <http://www.itu.org>. 2002-06-29
- [2] GB 917. SLA management handbook (public evaluation/version 1.5) [EB/OL]. <http://www.tmforum.org>. 2001-06-01.
- [3] TMF 701. Performance reporting concepts & definitions (public version 2.0) [EB/OL]. <http://www.tmforum.org>. 2001-11-01.
- [4] 成卫青, 龚俭, 葛立青等. 端到端SLA性能测量与分析[J]. 东南大学学报(自然科学版), 2004, 34(5): 589~593.
- [5] ITU-T Y.1540. Internet protocol data communication service –IP packet transfer and availability performance parameters[EB/OL]. <http://www.itu.org>. 2002-12-14.
- [6] ITU-T Y.1541. Network performance objectives for IP-based services[EB/OL]. <http://www.itu.org>. May 2002.

- [7] IETF RFC2330. Framework for IP performance metrics[EB/OL]. <http://www.ietf.org/>. 1998-05-05.
- [8] IETF RFC2679. A One-way Delay Metric for IPPM[S]. Sep. 1999.
- [9] IETF RFC2680. A One-way Packet Loss Metric for IPPM[S]. Sep. 1999.
- [10] IETF RFC2681. A Round-trip Delay Metric for IPPM[S]. Sep. 1999.
- [11] IETF RFC3148. A Framework for Defining Empirical Bulk Transfer Capacity Metrics[EB/OL]. <http://www.ietf.org/>. 2001-07-01.
- [12] IETF RFC3393. IP Packet Delay Variation Metric for IP Performance Metrics[S]. Nov. 2002.
- [13] 何宝宏. IP网络性能参数与指标要求[J]. 电信工程技术与标准化. 2001 (5): 26~29
- [14] IETF RFC2460. IPv6 Specification[S]. Dec. 1998.
- [15] 葛立青. 分布式SLA监测系统后处理器的研究与实现[D]:[硕士学位论文]. 南京:东南大学计算机科学与技术系, 2004.
- [16] IETF RFC2401. Security Architecture for the Internet Protocol[S]. Nov. 1998.
- [17] IETF RFC3729. Application Performance Measurement MIB[S]. March 2004.
- [18] ITU-T F.700. Framework Recommendation for multimedia services[EB/OL]. <http://www.itu.org>. Nov. 2000
- [19] ITU-T G.1010. End-user multimedia QoS categories[EB/OL]. <http://www.itu.org>. Nov. 2001
- [20] Cisco MAR. Network-Based Application Recognition and Distributed Network-Based Application Recognition[EB/OL]. http://www.cisco.com/en/US/products/ps6616/products_ios_protocol_group_home.html. 2005-8-26.
- [21] TS Choi, CH Kim, SH Yoon, *et al.* Content-aware Internet application traffic measurement and analysis[C]. In: 2004 IEEE/IFIP Network Operations and Management Symposium (NOMS 2004). 2004. 511~524.

Research On QoS monitoring for public IPv6 network bearer service SLA

CHENG Wei-qing^{1,2}, GONG Jian¹

¹(School of Computer Science and Engineering, Southeast University, Nanjing 210096)

²(College of Computer, Nanjing University of Posts & Telecommunications, Nanjing 210003)

Abstract: After the demands for network bearer service SLA (NSLA) in NGI are analyzed, taking possible application scenarios of NSLA as well as new features of IPv6 into account, this paper defines four QoS parameters for IPv6 NSLA: IP packet transfer delay (IPTD), IP packet delay variation (IPDV), IP packet fail ratio (IPFR) and IP packet throughput capability (IPTC), where IPTC is suitable for characterizing the ability of IP network section ensemble to sustain a given traffic pattern. Packet recognition and clock synchronization are two key issues for IPv6 NSLA monitoring. In this paper an approach to packet recognition is proposed, and to weaken synchronization demands and allow NSLA monitoring at lower cost, all QoS parameters but IPTD are rendered to be tolerant to limited unsynchronization between clocks of an ingress measurement point (MP) and its corresponding egress ones. Moreover we discuss the selection of QoS parameters for application-oriented NSLA based on factors of application features, user perception and IPv6 protocol. Finally, a NSLA monitoring system consisting in meters, front processors, analyzers, raw data databases and result databases is designed. It adopts ON-OFF systematic sampling method to monitor natural traffic passively, and measure QoS parameters timely under the collaboration of all components of the system.

Key words: Quality of Service; network bearer service SLA (NSLA); IP packet throughput capability (IPTC); IP packet Delay Variation (IPDV); IP packet fail ratio (IPFR)