

基于 SIP 无线 IP 网络移动管理的研究

叶德谦, 李锁雷

(燕山大学中德信息技术研究所, 秦皇岛 066004)

摘要: IETF 组织提出的 SIP 是建立、修改和终止多媒体会话的端到端应用层协议, 能够满足移动性的要求。但是在会话切换的过程中会丢失数据包, 产生不能忍受的延迟。该文分析了延迟产生的原因, 并在此基础上提出了充分利用网络资源、由用户终端发起的 BS 服务器来完成软切换的解决方法。并证明这种方法能够解决切换延迟问题, 有效地提高了通话质量。

关键词: 会话初始协议; 软切换; 切换延迟

Research on Mobile Management of Wireless IP Network Based on SIP

YE Deqian, LI Suolei

(ICDZ, Yanshan University, Qinhuangdao 066004)

【Abstract】 Session initiation protocol is an application-layer protocol standardized by IETF for establishing, modifying, and terminating peer-to-peer multimedia communication, and it provides mobility in such network. However, the handoff procedure with SIP suffers from undesirable delay and packet loss. In this paper, the cause of handoff delay is analyzed and it is proposed that the soft handoff processing is initiated by MA and carried out by BS. It is proved that this is an effective solution for avoiding the hand off delay and improving the session QoS.

【Key words】 Session initiation protocol (SIP); Soft handoff; Handoff delay

在用户漫游时进行的会话, 可能因为接入点改变等原因而中断。对移动通信协议来说, 很重要的一部分是在用户终端跨越不同的网络时如何获得不间断的会话。这是一个改变当前连接参数、适应新的通信需要、保持不间断连接的过程。对于那些实时性的会话, 如电话, 提供连续不间断的服务显得尤为重要。通过 SIP 建立起来的实时会话, 媒体流和信令流相互独立, 媒体流传输协议通常应用 RTP, 传输层协议是 UDP, 从而实现端到端的多媒体会话。对于这种基于 UDP 的连接, 主要参数是发起方和接受方的 IP 地址, 接入点改变, 终端用户代理的 IP 地址改变, 切换的过程主要是 UDP 数据包中源端、目的端 IP 地址的改变。SIP^[1] 协议通过发送 re-INVITE 的方法来更新当前会话阐述, 这种方法简单有效, 但是在当前网络环境下, 会产生延迟, 在采用 FR 连接的 WLAN 无线 IP 网络中, 大约是 6s, 这是不能忍受的。

1 SIP 简介以及对移动性的支持

SIP 是一种应用控制协议, 用于创建、修改或终止多媒体会话。这里“会话”可以是 Internet 多媒体会议、Internet 电话及远程教育等类似应用。会话对象包括自然人和机器, 如各种媒体存储设备。

在 SIP 中, 不仅支持终端移动性, 个人移动性还支持服务移动性, 服务移动性包括会话前的移动性和会话过程中的移动性。会话前的移动性可以通过二次注册来实现, SIP 协议已经解决得很好。在一个会话进行的过程中, 移动终端也有可能从一个无线 IP 网络进入另一个无线 IP 网络, 从而改变了它的 IP 地址。在这种情况下, 用户发送一个 re-INVITE 消息就能够更新 Contact URI 并且在携带的 SDP 消息体中更新媒体信息, 如图 1 所示。MA1 检测到一个新的无线 IP 网

络, 通过 DHCP 协议获得一个新的 IP 地址, 发送一个 re-INVITE 消息, 媒体流转向新的 IP 地址。

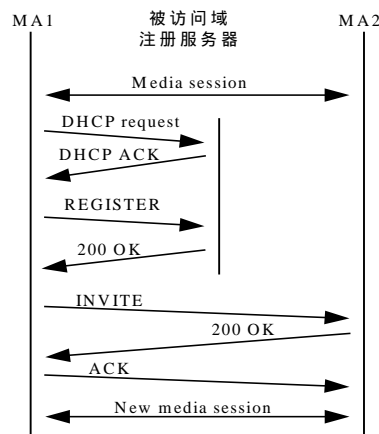


图 1 SIP 协议中会话切换过程

2 软切换模式实现无中断连接

2.1 问题描述

在 SIP 会话进行的切换过程中, 涉及到两个问题: 如何尽快改变 UDP 数据报中的 IP 地址数据参数和尽量减少数据报丢失。这种情况下基本的 SIP 协议是这样解决的: 发送一个 re-INVITE 消息, 更新 Contact URI 并且在携带的 SDP 消息体中更新媒体信息。但是, 发送 re-INVITE 之前, 要完成

基金项目: 河北省自然科学基金资助项目(F2004000252)

作者简介: 叶德谦(1954 -), 男, 教授、博士, 主研方向: 数据仓库, 人工智能和 SIP 应用; 李锁雷, 硕士生

收稿日期: 2006-08-14 E-mail: icdz@sina.com

几项操作，首先，移动终端要检测到一个新的无线 IP 网络。其次，通过 DHCP 协议获得一个新的 IP 地址。这个可能造成无法忍受的延迟。会话切换中的延迟产生如图 2 所示。

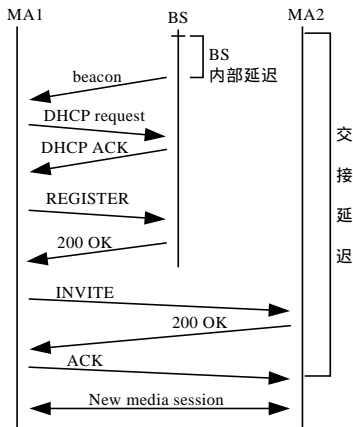


图 2 会话切换中的延迟产生

2.2 软切换过程

针对以上提到的这些情况，SIP 没有提供解决方法。其实在移动会话的切换过程中，除了上述提到的硬切换（在当前会话结束后再发起的切换）外，还可以利用软切换，在当前会话结束之前就建立新连接，这样在切换的过程中会话就不会中断，但是假如在某一时刻，用户终端收到来自两个网络的媒体信息，这种干扰通常可以忽略，如果不能忽略，一些 RTP 数据包会丢失，会话可能暂时中断。其过程见图 3。

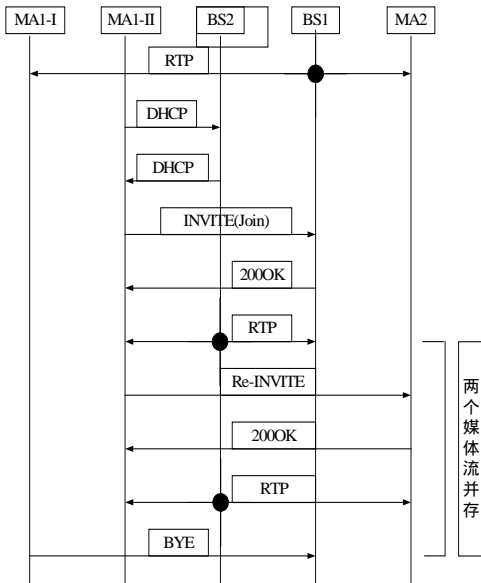


图 3 软切换过程及延迟的产生

软切换的过程由 MA1 发起，可以由 MA1 完成也可以由 BS 来完成，考虑到 MA1 的处理能力，由 BS 来完成比较合理。这时 BS 不仅具有 SIP Proxy、Redirect Server 的功能，还应装备 B2BUA。MA1 发现 BS2 的存在，就意味着要发生会话切换，如果原会话还没有中断，MA1 立即发起软切换过程，通过 DHCP 获得新的 IP 地址后，向 BS1 发出带有 join 标题头的 INVITE 消息，BS1 接到请求后，复制当前会话的所有数据包，通过 BS2 建立另一个会话，这时有两个会话同时存在。MA1 随即向 MA2 发出 re-INVITE 消息，来更新会话，成功更新后，MA1 和 MA2 通过 BS2 进行通话，MA1 向 BS1 发送 BYE 消息强制终止通过 BS1 的会话。

3 延迟分析

3.1 所用网络环境

目前存在各种各样的无线 IP 网络，本文进行延迟分析的 WLAN 网络结构如图 4 所示，MA1 和 BS 之间 FR 连接，SIP 信令传递采用 TCP 方式。媒体流采用 RTP 协议，UDP 连接。

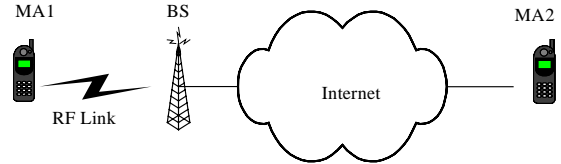


图 4 无线 IP 网络构成模型

3.2 延迟分析

3.2.1 正常切换分析

MA1 产生 SIP 消息来更新当前会话，SIP 消息发送到最近的 BS，再转发到接收机 MA2，MA2 可以是移动终端，也可以是固定终端或服务器。当会话建立以后，MA1 移动，改变接入点，MA2 就能够接收到更新消息。因此切换延迟主要是 re-INVITE 消息从 MA1 到达 MA2，主要的延迟来自于：(1) MA1 对 SIP 消息的处理延迟；(2) MA1 和 BS 之间的无线 IP 接入延迟；(3) IP 网络固有延迟；(4) MA2 的处理延迟。Internet 的传输延迟与消息排队方式以及路由器的数量，路径的选择有关，很难得到精确的时间，由于篇幅限制，本文采用固有延迟，进一步研究请参考文献[2]。并且，假定 MA1 和 BS 的排队方式为 M/M/1，MA2 的排队方式为 M/G/1。排队延迟如图 5 所示。其中 μ 是 MA1 的 SIP 消息到达率， λ 是 BS 的 SIP 消息到达率， μ_s 是 MA2 的 SIP 消息到达率， μ_1 是 MA2 的其他消息到达率， Δ 是 Internet 的固有延迟。

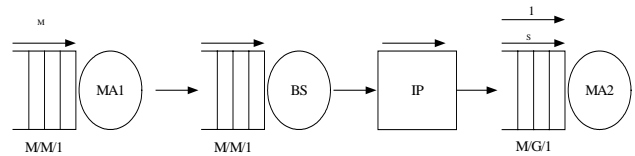


图 5 切换过程中的排队延迟模型

因为每个 BS 下有多个 MA 工作，所以 $\mu < \lambda$ ， μ 为的一部分，则 D_h 由以下几部分组成。

$$D_h = D_1 + D_2 + D_3 + D_4 + D_5$$

其中 D_1 为在 MA1 排队延迟， D_2 为接入传输延迟， D_3 为 BS 排队延迟， D_4 为 IP 网络固有延迟， D_5 为 MA2 排队延迟，由排队理论[3]可知：

$$D_1 = \frac{1}{\mu - \lambda}$$

其中， μ 是每个 SIP 消息在 MA1 和 BS 中的处理延迟。

$$D_3 = \frac{1}{\mu - \lambda} \quad D_4 = \Delta$$

$$D_5 = \frac{\frac{1}{\mu_s} (1 - \rho_1 - \rho_s) + R}{(1 - \rho_1)(1 - \rho_1 - \rho_s)}$$

其中， μ_s ：MA2 的 SIP 消息处理延迟， μ_1 ：MA2 的其他消息处理延迟， ρ_1 ：MA2 的其他消息负荷， ρ_s ：MA2 的 SIP 消息应答负荷。

$$R = \frac{\lambda_1 X_1^2 + \lambda_s X_s^2}{2} = 0.501(\rho_1^2 + \rho_s^2)$$

其中 X_1 和 X_s 分别为 $\mu_1 \mu_s$ 的期望，由于计算 SIP 消息的延迟，因此，只需考虑优先级高于 SIP 消息的其他消息处理延迟。

D_2 构成分为两种情况：

(1)当接入没有 RLP (Radio Link Protocol) 支持时,

$$D_2 = \sum_{i=1}^{N_m} [(k-1)\tau + \frac{D}{(1-q^{N_m})(1-2q)} + \frac{1-q}{1-q^{N_m}} D [\frac{q^{N_m}}{1-q} - \frac{2^{N_m+1}}{1-2q}]]$$

其中 N_m 是TCP成功传输SIP消息前传输的其他消息数目,是重传间隔时间, D 是无线信道端到端传播延迟, p 是连接发生错误的概率, q 是丢包率。 $q=1-(1-p)^k$,根据文献[4], $D=100ms$, $\tau=20ms$ 。

(2)当接入有 RLP 支持时,

$$D_2 = D + (k-1)\tau + \frac{k(P_f - (1-p))}{P_f^2} * (\sum_{j=1}^n \sum_{i=1}^j P(C_{ij})(2jD + (\frac{j(j+1)}{2} + i)\tau)) + \frac{2D_q(1-q)}{1-q^{N_m}} * [1 + \frac{4q(1-(2q)^{N_m-2})}{1-2q} - \frac{q(1-q^{N_m-2})}{1-q}]$$

其中,当 $n=3$ 是RLP重传的最大数时, C_{ij} 是第 j 次重传尝试中的第 i 帧。

$$P_f = 1-p + \sum_{j=1}^n \sum_{i=1}^j P(C_{ij}) = 1-p(p(2-p))^{\frac{n(n+1)}{2}}$$

K 是每个SIP消息需要的帧数,假定重传间隔是20ms,无线IP带宽是9.6Kbps,则设帧包含 $9.6*10^3*20*10^{-3}*1/8=24$ 字节,每个SIP消息是500个字节,这样每个SIP消息大约要传21帧。其他参数的值分别为: $\mu=4*10^{-4}s$, $s_1=0.7$, $s_2=1/\mu$, $\tau=200ms$, $N_m=10$ 。假定 $\mu=\mu_s$, $M=0.1$, $\lambda=500messages/s$,当FER (frame error rate)=0.5时,有RLP支持时, $D_h=6.42s$,没有RLP支持要高出许多,大于31s。其中 $D_2/D_h=98.7%$,可以看出,整个过程的延迟主要来自于无线接入传输,且即使有RLP支持,延迟也达到6s,这是让人无法忍受的。

3.2.2 软切换的延迟分析

软切换是在会话还没有结束时,就由MA1发起切换过

程,这样有效避免了在无线接入过程中产生的延迟,但是也带来了一些影响,首先对MA1特别是BS1的处理能力提出了更高的要求,其次在切换过程中,会有两个媒体流同时并存,会产生数据丢失。由于和BS2已经建立连接,两个媒体流并存时间最长是 $D=2(D_1+D_3+D_4+D_5)=0.164s$,这种干扰通常可以忽略。在上面的计算中,假设MA2不发生会话参数的改变,第2个媒体流建立的时间和正常切换的时间一样,则整个软切换的时间为 $6.32+0.164=6.484s$,由于软切换过程中,绝大部分时间里原会话仍在进行,因此不会中断会话。

4 结束语

本文分析了在基于SIP的无线IP网络中,会话进行中切换延迟产生的原因,在此基础上,提出了软切换,这种方法能够避免产生切换延迟,同时提高了对BS的要求,这种提高很容易达到,而且由此可能产生的数据丢失在可以忽略的范围内。SIP已经被3GPP组织确定为对媒体子系统的信令协议,也是下一代网络中,软交换的核心协议之一,有效地解决其在无线IP网络领域的切换问题,对提高服务质量有极其重要的意义。

参考文献

- 1 Rosenberg J, Schulzrinne H, Camarillo G, et al. SIP: Session Initiation Protocol[S]. RFC 3261, IETF, 2002-06.
- 2 Eyers T, Schulzrinne H. Predicting Internet Tele-phony Call Setup Delay[C]//Proceedings of 1stIP-Telephony Workshop, Berlin, Germany. 2000-04.
- 3 Kleinrock L. Queuing Systems Volume I: Theory[M]. John Wiley & Sons, 1975.
- 4 Das S K, Lee E, Basu K, et al. Performance Optimization of VoIP Calls over Wireless Links Using H.323 Protocol[C]//Proceedings of the INFOCOM'02. 2002.

(上接第100页)

3 基于流行度预测的缓存替换算法及其性能研究

对每个缓存的流媒体文件,将由上述流行度预测算法得到的流行度值作为流媒体文件的综合利用价值,并以此作为缓存替换算法的排名依据,这就是基于流行度预测的缓存替换算法。算法简述如下:

当有流媒体文件需要进入缓存但空闲缓存空间不足时执行如下步骤:

(1)对每个缓存的流媒体文件 f 计算

$$\bar{B}_f = AY_f, \bar{y}_f = (1, k+1, (k+1)^2) \cdot \bar{B}_f$$

(2)循环执行,直到能容纳下要缓存的文件时,删除缓存中 \bar{y}_f 最小的流媒体文件;

(3)将要缓存的文件存入缓存。

不难看出,如果缓存中的文件数为 m ,则算法的时间复杂性与空间复杂性均为 $O(km)$ 。由于流媒体文件体积庞大,因此 m 值不可能太大,而 k 值一般不超过10,其计算复杂性并不大,完全适合在线计算。

对于算法的有效性,我们通过模拟实验进行了研究。实验输入的请求流是由山东农业大学教学实验网上视频服务器的访问日志整理筛选得到的一个实际的请求访问序列,由于篇幅所限,详细的实验过程及结果在此不再叙述,实验结果显示,该算法与采用衰减函数法的流行度估计算法相比,总

的缓存替换次数较少,缓存命中率较高。

4 总结

本文提出了一种较为精确的流媒体流行度预测算法,并以此为基础给出了一种流媒体代理缓存的替换算法,初步的实验验证该算法具有较好的性能。在不同应用环境下,可将流媒体文件的传输成本、字节有用性、文件大小等因素分别引入算法中以进一步提高系统性能。采用不同输入流的进一步实验对比,以及预测算法计算复杂性的进一步降低等是下一步要做的工作。

参考文献

- 1 覃少华,李子木,蔡青松,等.基于代理缓存的流媒体动态调度算法研究[J].计算机学报,2005,28(2):185-194.
- 2 钟玉琢,向哲,沈洪.流媒体和视频服务器[M].北京:清华大学出版社,2003.
- 3 肖明忠,李晓明,刘翰宇,等.基于流媒体文件字节有用性的代理服务器缓存替换策略[J].计算机学报,2004,27(12):1633-1641.
- 4 罗治国,孙巍,王行刚.一种基于传输成本的流媒体缓存替换算法及其性能评价[J].通信学报,2004,25(2):61-67.
- 5 Jin S, Bestavros A. Popularity-aware Greedy Dual-size Web Proxy Caching Algorithms[C]//Proceedings of ICDCS. 2000-06.