

文章编号:1001-9081(2007)05-1106-03

## 一种新的基于 IMS 的 SIP 重传机制

林 晖<sup>1,2</sup>, 万晓榆<sup>2</sup>, 樊自甫<sup>2</sup>

(1. 福建师范大学 数学与计算机科学学院, 福建 福州 350007;

2. 重庆邮电大学 下一代网络应用技术研究, 重庆 400065)

(hawkhui95@gmail.com)

**摘 要:** 目前, IP 多媒体子系统(IMS)中基于会话初始化协议(SIP)的会话建立时间受到无线信道约束带宽、帧错误率(FER)值的影响, 交换的消息数量、消息长度以及重传机制对于会话建立的时延有很大的影响, 延长了会话建立的时间。将基于 SIP 的会话建立与无线信道的性能结合, 提出一种新的 SIP 自适应重传机制。实验结果表明, 该重传机制缩短了会话建立的时间, 对 IMS 端到端的服务质量(QoS)性能有明显的改善。

**关键词:** IP 多媒体子系统; 服务质量; 会话初始协议; 时延

**中图分类号:** TP393 **文献标识码:** A

## A new sip retransmission mechanism based on IMS

LIN Hui<sup>1,2</sup>, WAN Xiao-yu<sup>2</sup>, FAN Zi-fu<sup>2</sup>

(1. School of Mathematics and Computer Science, Fujian Normal University, Fuzhou Fujian 350007, China;

2. Next Generation Network Application Technology Institute,

Chongqing University of Posts and Telecommunications, Chongqing 400065, China)

**Abstract:** At present, the session setup time based on the Session Initial Protocol (SIP) in IP Multimedia Subsystem (IMS) is affected by the constrained bandwidth of the wireless link, Frame Error Rate (FER). The number and size of the message exchanged affect the average session setup delay, which can result in retransmission of packets lost and can lengthen the session setup time. In this paper, both the session setup and wireless link were taken into account and a new adaptive retransmission mechanism that was adjustable to the size of signaling packets involved in the session establishment was proposed. The numerical results indicate that the proposed new mechanism can make the session setup shorter and is efficient for optimizing the time of session setup.

**Key words:** IP Multimedia Subsystem(IMS); Quality of Service(QoS); Session Initial Protocol(SIP); delay

## 0 引言

分组交换技术的快速发展,使得基于 IP 的分组交换将同时传输语音和数据流量。3GPP 和 3GPP2 都建议将基于 IP 的无线技术集成入第三代移动通信中,因此提出了适用于 UMTS 和 CDMA2000 的 IP 多媒体子系统(IMS)技术,提供诸如 VOIP 等基于 IP 的服务。目前,由于无线接入网络自身的缺点,使得 IMS 中基于会话初始化协议(SIP)的会话建立的时延太长,超出用户忍受的限度。因此有必要对呼叫建立的时延进行优化。当前,无线网络中影响 SIP 会话建立时延的因素主要包括两个:物理信道和 SIP 采用的底层协议。UAC

文献[1,2]中研究了采用排队机制建立 SIP 会话建立的数学模型。文献[3]研究了 H. 323 的平均呼叫建立时延的性能。文献[5]中通过仿真对公共 Internet 网络上基于 H. 323 和 SIP 协议的呼叫建立时延做了评估。Kist 和 Harris 在文献[6]中研究了 3GPP 环境下的 SIP 会话初始化的时延。但是都没有将 SIP 会话建立的时延与物理信道的性能结合起来考虑。

本文将 SIP 呼叫建立与无线信道的性能结合起来,建立了基于 M/M/1、M/G/1 队列<sup>[7]</sup>的 SIP 会话建

立时延模型,提出了一种新的 SIP 自适应重传机制。

## 1 IMS 中 SIP 会话的建立

IMS 中使用 SIP 实现两个用户代理间的会话的建立如图 1 所示。

在图 1 所示的会话的建立过程中,用户代理客户端(UAC)通过向用户代理服务器(UAS)发送一个 INVITE 消息开始会话的建立。INVITE 请求包含了被请求的、将要通过各个域的 P/S/I-CSCF 的会话类型的详细信息。接受到 INVITE 消息,UAS 就向 UAC 发送相应的反馈消息 183。UAC 发送一

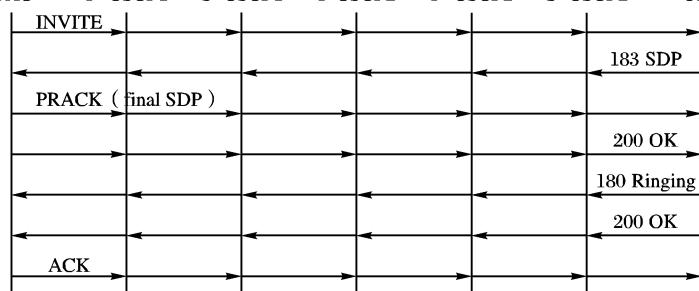


图 1 基于 UDP 的 SIP 会话建立

收稿日期:2006-11-14;修订日期:2007-01-29 基金项目:重庆市教委科学技术研究资助项目(KJ050512;KJ050513);重庆市科委自然科学基金资助项目(CSTC2005BB2054);重庆邮电学院青年教师基金资助项目(A2005-26)

作者简介:林晖(1977-),男,讲师,硕士研究生,主要研究方向:下一代网络、网络 QoS 技术; 万晓榆(1950-),男,教授,博士,主要研究方向:下一代网络; 樊自甫(1977-),男,讲师,硕士,主要研究方向:下一代网络。

个2000K用于保留经过的路径上所需要预留的资源,并且接受到一个响铃音和180响应。当UAS决定接受呼叫,就发送一个2000K响应。最后,确认UAC发送的ACK请求建立媒体会话。

## 2 会话建立时延模型

本文研究的SIP会话建立是从UAC发出INVITE请求到UAS被通知客户端已经接收到服务器关于本次会话的响应(即UAS接收到ACK请求)。会话的建立包括了客户端和服务端的消息处理,因此SIP会话建立的时延可以认为是客户端、服务器端完成所有消息处理的时延的总和。会话建立的时延与许多的因素有关,其中最主要的是在网络上的传输时延,包括分组丢失造成的重传时延和排队时延。这种传输时延将受到使用的传输协议和错误恢复策略的影响,是传输、排队和Internet的累积时延。

假设会话建立过程中有 $N$ 个必须消息,定义会话建立时延的数学模型:

$$T_{\text{会话}} = N \times T_{\text{源端}} + T_{\text{无线传输}} + N \times T_{\text{qp-CSCF}} + N \times T_{\text{ql-CSCF}} + N \times T_{\text{qs-CSCF}} + N \times T_{\text{有线传输}} + N \times T_{\text{目的端}} \quad (1)$$

模型中的 $T_{\text{无线传输}}$ 是无线传输中SIP协议采用的底层传输协议产生的时延(如使用RLP协议的 $T_{\text{TCP}}$ 、 $T_{\text{UDP}}$ ); $T_{\text{有线传输}}$ 是SIP消息在有线传输中产生的时延,主要和数据报传输路径上的链路类型和路由器的数量有关。由于很难标准化不同种类的传输路径和计算传输的时延,本文将 $T_{\text{有线传输}}$ 设置为一个常数; $T_{\text{源端}}$ 是源端的平均排队时延; $T_{\text{目的端}}$ 是目的端的时延; $T_{\text{qp-CSCF}}$ 、 $T_{\text{ql-CSCF}}$ 、 $T_{\text{qs-CSCF}}$ 分别代表P/I/S-CSCF服务器上的平均排队时延。

### 2.1 传输协议

SIP可以采用TCP或UDP作为底层的传输协议。如果SIP采用的是UDP,所有会话建立时延是UAC和UAS成功的接收到各种消息处理中所涉及的消息所需要的时间。如果SIP采用的是TCP,所有会话建立时延是TCP会话建立时间加上建立SIP会话需要的所有消息成功传输的时间。

### 2.2 排队时延

这一节中我们将考虑SIP消息在用户终端、中间的CSCF服务器和目的地上的排队时延。对于源终端(ST)和中间的CSCF服务器我们采用M/M/1的排队模型,目的终端(DT)采用M/G/1的排队模型。假设CSCF为多个ST服务,因此在ST上的SIP消息的到达速率 $\lambda_M$ 是CSCF上SIP消息到达速率 $\lambda$ 的一部分: $\lambda_M \leq \lambda$ 。

根据排队理论<sup>[7]</sup>,UAC上的排队时延 $T_{\text{源端}}$ 为:

$$T_{\text{源端}} = \frac{1}{\mu - \lambda_M} \quad (2)$$

这里的 $\mu$ 是SIP消息在ST的服务速率。P/I/S-CSCF上的排队时延为:

$$T_{\text{qp-CSCF}} = T_{\text{ql-CSCF}} = T_{\text{qs-CSCF}} = \frac{\rho_s}{\lambda(1 - \rho_s)} \quad (3)$$

这里 $\rho_s$ 是DT和CSCF的负载。DT上的时延 $T_{\text{目的端}}$ 将通过非抢先的基于优先级的M/G/1排队模型来获得<sup>[7]</sup>:

$$T_{\text{目的端}} = \frac{\frac{1}{\mu_s}(1 - \rho_0 - \rho_s) + R}{(1 - \rho_0) + (1 - \rho_0 - \rho_s)} \quad (4)$$

$$R = \lambda_0 \bar{X}_1^2 + \lambda_s \frac{\bar{X}_s^2}{2} \quad (5)$$

这里的 $\rho_0$ 是DT上非SIP消息的负载, $\mu_s$ 是SIP消息在DT上的服务速率, $\mu_0$ 是非SIP消息在DT上的服务速率, $\bar{X}_s^2$ 和 $\bar{X}_1^2$ 分别是 $\mu_s$ 和 $\mu_0$ 的力矩。

## 3 基于UDP的SIP自适应重传机制

从上述的数学模型可以看出 $T_{\text{无线传输}}$ 是整个SIP会话建立时延的重要组成部分,本文的研究主要考虑UDP协议。由于SIP会话发起使用的是带宽约束和错误率高的无线信道,交换的消息的数量和长度将影响会话建立时延。因此本文主要考虑消息重传机制的改进以减少会话建立时延。定义SIP的第 $i$ 次重传时间:

$$Tr(i) = 2^{i-1} \times Tr \quad (6)$$

初始化的 $Tr$ 直接影响到会话建立的延时。如果太短,那么在反馈的应答消息还在半路中没有被接收到数据报就开始重传了;如果太长,那么当丢失出现的时候,就会增加不必要的会话建立时间。目前SIP中的 $Tr$ 采用的是固定的时间间隔,无法根据当前的消息的传输时间做出适当的调节,因此本文提出一种自适应的重传机制。定义客户端的 $Tr$ 如下:

$$Tr = D + (k_1 - 1) \times \tau + D + (k_2 - 1) \times \tau + T_{\text{排队}} \quad (7)$$

对于服务器端来说 $Tr$ 的值将影响2000K( $K_2$ )和ACK( $K_3$ )中的帧的数目,采用和客户端相同的方法定义服务器端的 $Tr$ 如下:

$$Tr = D + (k_3 - 1) \times \tau + D + (k_4 - 1) \times \tau + T_{\text{排队}} \quad (8)$$

其中 $D$ 是端到端的帧的传输时延; $\tau$ 是帧间时间; $K_1$ 代表INVITE请求中包含 $K_1$ 个帧; $k_2$ 代表183反馈中包含 $k_2$ 个帧; $K_3$ 代表服务器端2000K消息中包含 $K_3$ 个帧; $k_4$ 代表ACK中包含 $k_4$ 个帧; $T_{\text{排队}}$ 是2.2中所提到的排队时延。

假设 $p$ 为一个帧在空中链路中出现错误的概率一个UDP包中包含 $k$ 个帧,则 $(1 - p)^k$ 是UDP数据报不出现错误的概率,数据报的丢失率为 $1 - (1 - p)^k$ 。假设 $q$ 是重传的概率:发送的第一个包(包含 $K_1$ 个帧的INVITE请求)丢失或第一个包被受到但是反馈(包含 $K_2$ 个帧的183消息)丢失。因此,客户端出现INVITE重传的概率为:

$$q = [1 - (1 - p)^{k_1}] + [1 - (1 - p)^{k_2}] \times [(1 - p)^{k_1}] \quad (9)$$

$$q = 1 - [(1 - p)^{k_1 + k_2}] \quad (10)$$

服务器端的INVITE重传的概率为:

$$q = 1 - [(1 - p)^{k_3 + k_4}] \quad (11)$$

假设最大重传的次数为 $Nm$ (对于SIP, $Nm = 7$ )。成功消息处理的时延为成功的传输一个包含SIP消息UDP数据报和成功的接收到相应的响应的时延。因此,成功传输第 $i$ 个UDP数据报的时延 $Tt(i)_{\text{UDP}}$ 为:

$$\begin{aligned} Tt(i)_{\text{UDP}} &= \frac{1}{1 - q^{Nm}} \times \left\{ (1 - q) [D + (k - 1)\tau] + \right. \\ &\quad (1 - q)q [Tr + D + (k - 1)\tau] + \\ &\quad (1 - q)q^2 [3Tr + D + (k - 1)\tau] + \dots + \\ &\quad \left. (1 - q)q^{Nm-1} [(2^{Nm-1} - 1)Tr + D + (k - 1)\tau] \right\} \\ &= D + (k - 1)\tau - Tr + \frac{(1 - q)[1 - (2q)^{Nm}]}{(1 - q^{Nm})(1 - 2q)} \times Tr \end{aligned} \quad (12)$$

会话的建立包含成功地完成客户端和服务端的消息处理。因此,建立会话的全部传输时延  $T_{tUDP}$  为:

$$T_{tUDP} = \sum_{i=1}^N T_t(i)_{UDP} \quad (13)$$

## 4 性能分析

### 4.1 实验参数设置

实验相应的参数设置如下:每一个无线帧的持续时间为 20ms;根据文献[3]中的设置:时延 D 为 100ms,帧间时间  $\tau$  为 20ms,最大的重传数  $N_m$  为 7;对于排队时延,假设  $\mu_s = \mu$ ,  $\lambda_M = 0.1\lambda$ ;在目的地的排队时延计算中假设服务速率的标准偏差  $\sigma$  为平均值的 5%。 $\bar{X}_1^2 = E[X_1]^2$ ,  $\bar{X}_s^2 = E[X_s]^2$ ,  $E[X_1]^2 = \sigma_1^2 + (E[X_1])^2$ ,  $E[X_s]^2 = \sigma_s^2 + (E[X_s])^2$ 。将  $\mu_s$  代入得出  $R = 0.501[\rho_0^2 + \rho_s^2]$ 。SIP 在 ST 的到达速率  $\lambda_M = 50$  请求/s;  $\mu = 70$  请求/s;  $\rho_s = \lambda \times \mu; \rho_0 = 0.7; T_{有线传输} = 100ms$ 。

### 4.2 性能分析结果

表 1 消息长度以及帧的数量

消息	有效载荷/B	消息长度/B	帧的数量(19.2kbps)
SI PINVITE	700	728	19
SI P183	835	863	23
SI PPRACK	558	586	16
SI P200OK	545	573	15
SI P180	349	377	10
SI PACK	300	328	9

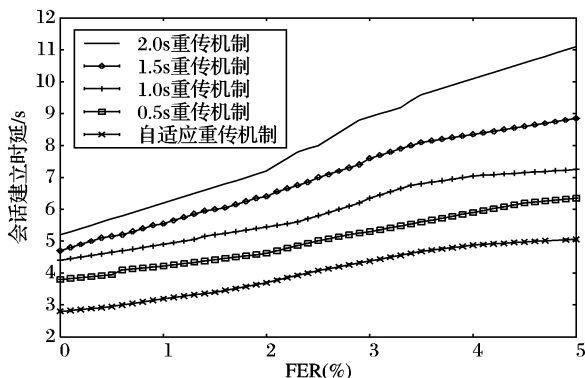


图 2 两种重传机制的平均会话建立时延性能比较

通过前面数学模型的分析可以看出 SIP 会话建立时延随着 FER(模型中的  $p$ ) 增长以指数规律增长。交换的消息的数量、消息长度以及重传机制对于会话建立的时延有很大的影响,这些方面的优化将可以缩短会话建立时延。假定每一

个 UDP 数据报由一个 IP 包来传送。表 1 中给出了 UDP 数据报的长度以及信道上每个数据报所包含的帧的数量。SIP 重传计时器  $Tr$  用于实现 SIP 消息的重传,缺省值为 0.5s,但是由于 SIP 协议是设计用于带宽丰富的有线链路的协议,当应用于无线网络时,由于无线信道的带宽约束和高错误率的特点以及采用的底层传输协议(UDP),其重传的时间将变长(大于 0.5s)<sup>[8]</sup>。因此,本文将  $Tr$  从 0.5s 到 2s 的会话建立时延与自适应重传机制的会话建立时延作了比较。图 2 中显示了比较结果。从图中的比较结果可以看出采用自适应重传机制的性能比固定值重传机制的性能好,对 IMS 中基于 SIP 的会话建立时延有明显的改善。

## 5 结语

本文采用排队机制建立了 IMS 中的端到端时延的数学模型,分析了影响时延的各个因素。根据目前 IMS 中的 SIP 会话采用的固定值重传机制的缺点,本文提出了一种新的可以根据会话建立中的信令包的长度做出调节的自适应重传机制。结果表明自适应重传机制能有效地改善 SIP 会话建立时延。而且,性能的分析还表明 SIP 自适应重传机制的性能可以通过压缩算法来减少 SIP 消息的长度,通过纠错机制或混合 ARQ 校正 SIP 消息避免无线链路上的重传来改进会话建立的时延。

### 参考文献:

- [1] GURBANI VK, JAGADEESAN LJ, MENDIRATTA VB. Characterizing Session Initiation Protocol (SIP) Network Performance and Reliability[A]. ISAS 2005[C]. 2005. 196-211.
- [2] RAJAGOPAL N, DEVETSIKIOTIS M. Modeling and Optimization for the Design of IMS Networks[A]. Proceedings of the 39th Annual Simulation Symposium (ANSS'06)[C]. 2006.
- [3] DAS S, LEE E, BASU K, et al. Performance Optimization of VoIP Calls over Wireless Links Using H.323 Protocol[J]. IEEE Transactions Computers, 2003, 52(6): 742-752.
- [4] ITU-T Recommendation E. 721—Network Grade of Service Parameters and Target Values for Circuit-Switched Services in the Evolving ISDN[S].
- [5] EYERS T, SCHULZRINNE H. Predicting Internet Telephony Call Setup Delay[A]. Tony Eysers and Henning Schulzrinne IPTel 2000 (First IP Telephony Workshop)[C]. 2000.
- [6] KIST A, HARRIS RJ. SIP Signaling Delay in 3GPP[A]. Sixth International Symposium on Communications Interworking of IFIP[C]. 2002. 211-222.
- [7] KLEINROCK L. Queueing systems, vol 1: theory[M]. New York: John Wiley and Sons Ltd, 1975.
- [8] RFC 3261. SIP: Session Initiation Protocol[S].

(上接第 1088 页)

## 4 结语

为被动式红外传感器成功建立了节点模型,从模型中可以发现很多对此后系统设计具有指导意义的原则。依据该模型设计实现了运动人体的监测定位系统。系统试验验证了模型的可靠性同时证实了系统的性能满足准确性应用要求。但是其中也存在着一些问题有待进一步完善。

### 参考文献:

- [1] PAVLIDIS I, MORELLAS V, TSIMMYRTZIS P, et al. Urban Surveillance Systems: From the Laboratory to the Commercial World[J]. Proceedings of the IEEE, 2001, 89(10): 1478-1497.

- [2] WERNER-ALLEN G, JOHNSON J, RUIZ M, et al. Monitoring Volcanic Eruptions with a Wireless Sensor Network[A]. Proceedings Second European Workshop on Wireless Sensor Networks (EWSN'05)[C]. 2005. 108-120.
- [3] NOVOTNY PM, FERRIER NJ. Using Infrared Sensors and the Phong Illumination Model to Measure Distances[A]. Proceedings of the 1999 IEEE International Conference on Robotics & Automation[C]. 1999, 2: 1644-1649.
- [4] 梁振,于峰崎,程勇. CMOS 红外传感器的模型研究[J]. 传感器技术. 2005, 24(8): 36-38.
- [5] 张智丰,李向新,彭群生. 基于大气传输模型的动态目标红外成像仿真[J]. 系统仿真学报, 2000, 12(5): 524-527.