

基于 SIP 协议的应用层个人移动性的分析

张超, 陈鸣, 李兵

(中国人民解放军理工大学 指挥自动化学院, 江苏 南京 210007)

摘要:移动 IP 技术对于移动性的支持在网络层, 需要对网络协议进行修改, 而 SIP 协议在应用层提供了对移动性的支持, 无需底层网络协议的改变。文中深入探讨了 SIP 协议在应用层支持个人移动性的工作原理, 设计了两个典型的个人移动性应用场景, 进行了试验验证并对试验结果进行了分析讨论。试验充分证明了 SIP 协议在应用层对于个人移动性的良好支持, 为解决底层移动协议对移动性支持所带来的种种问题以及移动 IP 缺失的环境提供了有益的补充。

关键词:SIP; 应用层移动性; 代理服务器; 重定向服务器; 定位服务器

中图分类号: TN929.5

文献标识码: A

文章编号: 1673-629X(2010)07-0109-05

Analysis of SIP - Based Application Layer Personal Mobility

ZHANG Chao, CHEN Ming, LI Bing

(Institute of Command Automation, PLA University of Science and Technology, Nanjing 210007, China)

Abstract: Mobile IP technology for support for mobility in the network layer, network protocols need to be modified, while the SIP protocol at the application layer provides support for mobility without changing the underlying network protocol. This in-depth study of the SIP protocol at the application layer mobility support individuals working principle, design of two typical individual mobility scenarios were tested and the test results are analyzed and discussed. Test fully proved SIP protocol at the application layer mobility for individuals with good support, to address the underlying mobility protocols for mobility support for the various problems arising from missing as well as mobile IP environment provides a useful supplement.

Key words: SIP; application layer mobility; proxy server; redirect server; location server

0 引言

随着移动通信技术的飞速发展, 移动通信的应用及概念也不断地丰富和扩展。特别是在近年来引入协议会话发起协议(SIP)以后, 促进了应用层移动性概念的发展, 因此, 有必要对与移动性相关的概念进行分析梳理, 以便能够进行深入的研究, 设计和实现更好的支持移动性的机制和系统。现有的移动通信技术对移动性的支持主要体现在网络层和链路层技术上, 包括位置管理、切换管理和漫游等。移动 IP (Mobile IP) 以及衍生协议、Mobile IPv6 和 GPRS 等都是从网络层解决移动性问题的主要方案。而应用层移动性就是要在不改变现有 IP 协议栈的基础上在应用层来处理移动性问题, 并可以解决底层移动协议所固有的一些技术难题如三角路由、需要两个地址以及封装报文开销太大

等, 形成对网络层和链路层移动性解决方案的有力补充, 特别在移动 IP 设施缺失的环境中^[1]。由于因特网是一个投资巨大的基础设施, 不可能独立于现有支撑网络建立新的支撑环境; 不仅仅提供基站和子网之间的切换, 还必须从新型应用需求出发来更好地分析解决移动问题。其中 IETF 研制的信令协议 SIP 起着核心作用, 文中从支持应用层移动性的角度, 探讨其工作原理。

1 SIP 协议概述

SIP 是一种应用层的控制协议, 用于管理各种通信媒体的会话。它能控制有两个或多个参与者创建包括由多个媒体流组成的会话, 并能在会话中动态调整会话属性, 如会话带宽、传输的媒体类型和编码格式等^[2]。一个用户的媒体流能分布在多个设备上, 即除了计算机还有专用视频、音频和任何基于因特网通信机制的设备。SIP 在建立和终止多媒体通信方面支持如下 5 种特性: 用户定位、用户有效性、用户能力、建立会话和会话管理。SIP 通常要与其他协议一起协同工作, 例如它与实时协议(RTP)、实时控制协议(RTCP)、

收稿日期: 2009-11-06; 修回日期: 2010-02-12

基金项目: 国家 863 计划项目(2007AA01Z418); 江苏省自然科学基金(BK2009058)

作者简介: 张超(1983-), 男, 硕士研究生, 研究方向为分布式计算; 陈鸣, 教授, 研究方向为计算机网络、网络测量、分布式计算。

媒体网关控制协议(MEGACO)等,才能提供完整的通信服务,然而 SIP 并不依赖于这些协议。SIP 本身只是一种网络信令协议而不直接提供通信服务,但却提供了一些可用来实现各种服务的原语。例如它可以用于定位用户,也可以传输一个不透明的对象到对方当前位置。此外,SIP 本身还提供了一套安全服务,包括防止拒绝服务、认证服务、完整性保证、加密和隐私服务等^[3]。

为了支撑 SIP 提供的服务,SIP 系统定义了若干种服务器和用户代理(UA)。通过 UA 与服务器之间的请求报文和响应报文的交互以完成呼叫和传送层的控制。UA 是指为了向服务器发送请求而与服务器建立连接的应用程序,又可进一步分为用户代理服务器 UAS 和用户代理客户机 UAC。代理服务器 UAS 是用于向 UAC 发出的请求提供服务并回送应答的应用程序^[4]。在 SIP 系统中共有 3 类基本服务器:代理服务器(proxy server)、重定向服务器(redirect server)、注册服务器(registrar)。代理服务器本身既作为服务器向提出的请求提供转发服务,同时又可作为客户机向其他服务器发出请求。代理服务器按照对请求的处理方式不同又可以分为有状态模式和无状态模式。重定向服务器接收请求并将包括目的地的数据发送回主叫方,而不是直接发往被叫方。注册服务器是接收 REGISTER 请求并将用户注册信息保存至位置服务器的服务器^[5]。另外,在 SIP 系统中,还有位置服务器(location server)。位置服务器存储向请求的用户返回可能的位置信息,这些位置信息是由注册服务器接收到时立即上传到位置服务器上的。位置服务器作为全局性的数据库,其作用就像 DNS,对它的访问和接入可以是非 SIP 方式的。

SIP 报文分为两类:从客户机到服务器的请求 Request 报文和从服务器返回的响应 Response 报文。SIP 请求报文定义了几种重要的方法:INVITE、ACK、OPTIONS、BYE、CANCEL 和 REGISTER,其中 INVITE、ACK 和 CANCEL 用于建立会话,BYE 用于终止会话,REGISTER 用于注册联系信息,OPTIONS 用于查询服务器能力。响应报文则主要定义了 6 类应答:1xx 为临时应答,2xx 为成功处理,3xx 为重定向,4xx 为客户机错误,5xx 为服务器错误,6xx 为全局错误^[6]。

SIP 是基于文本的协议,使用 UTF-8 字符集。SIP 报文都由一个起始行、一个或者多个头域和一个可选的报文体组成,其中起始行又分为请求行和状态行。起始行、每个头域行和空行都必须由回车换行符组成。即使没有报文体,也必须有一个空行。除了字符集上的区别外,很多 SIP 的报文和头域的格式都同

HTTP/1.1 的一样。

相对其他协议而言,SIP 设计采用了基本功能与扩展功能相分离的思想,故有很强的可扩展性。基本功能相对简单稳定以保证 SIP 设备间的互通,任何支持 SIP 的设备都必须支持 SIP 的基本功能。随着 SIP 影响力日益增加,其基本功能已不能满足需求了。在基本功能的基础上,针对特殊应用如即时讯息和呈现(Presence)服务、第三方呼叫控制和资源管理等,IETF 又推出了很多 SIP 扩展标准,定义了 SIP 的扩展功能,每个扩展功能解决一个具体问题。例如, SIMPLE (RFC3265)就是针对即时消息和呈现服务提出的^[7]。

2 SIP 协议对应用层移动性支持分析

SIP 协议设计简单灵活、扩展性好,随着它的不断完善,其支持移动通信的得天独厚能力越来越被认可,用于提供应用层移动性。

下面分析为满足通信系统移动性的要求,SIP 所采用的机制。

1) 名字映射机制。

SIP 设计采用了逻辑地址和联系地址相分离的思想。逻辑地址用于标识用户,而联系地址用于标识用户当前的实际网络位置,一个逻辑地址可以对应多个联系地址。SIP 中逻辑地址采用了电子邮件格式,形如“user@host”,称为 SIP URI 或 SIPS URI。其中 user 部分是用户名或电话号码,指的是主机特定资源的地址;host 部分包含了一个完整的 URL 或者 IP 地址,指明了提供了资源的主机^[2]。一个 SIP URI 可以绑定一个或者多个实际的联系地址,这表明一个用户可以同时使用多个终端进行注册,如在使用移动电话的同时还可以使用笔记本多媒体通信系统。

2) 注册、注销机制。

UA 通过注册/注销机制向系统告知自己在移动过程中的当前位置,SIP 通过该机制为特定域的位置服务创建绑定关系。SIP 定义了注册服务器和 REGISTER 报文,UA 通过向注册服务器发送 REGISTER 请求报文来完成注册和注销过程,该报文中的 Contact 域包含注册的联系地址信息。当 UA 向注册服务器发送一个 REGISTER 请求后,注册服务器通过检查请求报文中的 Contact 头域来判断用户意图:是否是首次注册,是否是更新绑定信息,是否是添加地址映射记录,是否是删除绑定等。UA 在注册服务器上注册成功后,就可以接收呼叫了。主叫方根据对该用户地址映射信息的查询结果,将呼叫请求转发到该 UA 对应的联系地址。当 UA 要注销时,可以在 Contact 域中填写要删除的联系地址信息,并将 Expires 参数置 0,注册

服务器收到请求后就会删除该映射记录。若将 Contact 域设为“*”，并将 Expires 头域设为 0，则会删除该用户所有联系地址映射记录^[8]。

3) 搜索机制。

如果某用户希望建立与其他用户的会话，SIP 必须查找到对方用户正在使用的当前主机，这就需要用到 SIP 提供的搜索机制。SIP 通过位置服务来实现。位置服务则通过名字映射和注册/注销机制所建立的地址绑定关系来实现。地址绑定功能先将输入的 SIP 或 SIPS URI 映射为一个或者多个更加“接近”目标用户的 URI，代理服务器再通过查询该位置服务，并将输入的 URI 映射为最终用户的联系地址即可完成。

4) 重定向机制。

SIP 对移动性支持的重要方面就是采用了重定向机制。当被叫方移动位置时，主叫方发送请求报文给重定向服务器，重定向服务器给请求方返回一个 3xx 应答，这时它会在 Contact 头域填写一个或多个其他位置信息，告知请求方可以尝试与 Contact 头域中填写的位置信息联系。主叫方向新的联系地址发起 INVITE 呼叫请求，该请求会被路由到联系地址所在的用户终端。

5) 目标更新机制。

目标更新机制通过 SIP 中定义的 re-INVITE 报文来实现，使用户告知通信对方自己的联系地址发生了变化。当用户在通信过程中的位置发生改变时，UA 可以使用正在进行中对话的相关信息构造一个 re-INVITE 请求报文，将新的联系地址放入 Contact 头域发送给对方。对方收到请求报文后，发现 Contact 头域值的改变，就使用新的联系地址更新远端目标，从而保证了通信的持续性。用户还要及时向注册服务器更新注册信息。此外，目标更新机制还可用于修改当前会话中的参数。

事实上，在应用层的移动性有多层含义，其中包括终端移动性、个人移动性、会话移动性和服务移动性^[8]。我们常讲的移动性指的是终端移动性，而从移动性的发展趋势看则要达到个人移动性的要求。应用层的这些移动性所包含的具体内容如下所述。

(1) 个人移动性。

个人移动性^[8]是指用户在任何地方、任何时间、使用任何设备访问特定网络服务的能力。SIP 通过 SIP 注册服务器和定位服务器提供了名字映射机制和重定向服务，这些组件的应用使得 SIP 系统能够定位用户所使用的任何终端，从而保证了 SIP 与生俱来的对个人移动性支持。用户可以在任何情况下使用 SIP URI 或 SIPS URI 来联系其他用户而不需要考虑被呼叫用

户使用何种设备。可以使用一对多(一个地址、多个可能的终端)和多对一(多个地址绑定到一个终端)的映射。在一个用户使用多个终端的情况下，SIP 通过注册机制、定位机制和分支(forking)机制来提供个人移动性，一个呼叫请求或者通信报文最终能够到达被呼叫用户正在使用或希望使用的终端，从而将用户与终端和网络位置相分离，提供个人移动性。

(2) 终端移动性。

终端移动性^[8]是指用户在移动中使用某一终端，用户访问服务的接入点不是固定的，网络具有标识和定位终端的能力。SIP 通过注册和重定向机制提供终端移动性。当终端设备在不同网络之间移动时，终端被分配新的 IP 地址，但是终端则拥有唯一的 SIP URI 或 SIPS URI。通过注册机制，建立起终端新的 IP 地址和 SIP URI 或 SIPS URI 的绑定关系。因此无论终端如何移动，所对应的位置信息都是最新的，因此任何指向它的呼叫和通信报文都能够成功到达并维持已有会话的持续。

(3) 会话移动性。

会话移动性^[8]是指用户在不同终端间进行切换时保持会话的能力。SIP 通过目标更新机制来支持会话移动性。例如，用户 A 使用手机与用户 B 通话，当用户 A 进入办公室后希望使用个人计算机继续通话，并保持本次通话不中断。用户 A 的终端向用户 B 的终端发送 re-INVITE 请求报文，告知用户 B 的终端用户 A 希望使用新的终端代替现有设备，并将新设备相关参数传递给对方，用户 B 的终端就在通话没有中断的情况下与用户 A 新的终端建立了连接。

(4) 服务移动性。

服务移动性^[8]是指网络以合适的 QoS 持续地为用户提供服务而不考虑用户位置的能力。它保证了用户在移动或在终端和网络服务提供商之间进行切换时服务的可访问性。针对服务移动性，必须考虑两方面因素：网络切换时保持合适的 QoS，保证用户访问服务而不考虑其接入点。SIP 可以通过资源预留机制来提供 QoS 保证，通过 AAA 和注册机制来保证用户访问服务的能力。

3 SIP 支持应用层个人移动性试验及分析

根据用户具有多个终端和用户漫游到其他域中的两种不同情况，设计了下列试验来分析 SIP 对应用层个人移动性的支持过程。在试验网络中包括了 3 个局域网，设置了 SIP UA 若干、两个代理服务器/注册服务器以及 1 个定位服务器。尽管在逻辑上注册服务器与代理服务器是分离的，具体实现时它们通常是运

行在同一设备上的。SIP 本身对实现位置服务器没有特别要求,只要求某些域的注册服务器必须能够访问位置服务器上的数据,并且这些域的代理或者重定向服务器必须能够兼容读取相同的数据。采用 DNS 域名服务器作为位置服务器,即将代理服务器的地址映射到对应域的域名存储在该 DNS 服务器上。通过网络协议分析仪 Wireshark 获取 SIP 报文以及通信报文,从而绘制出通信时序图。

3.1 用户有多个可用设备时的通信过程

在如图 1 所示的试验环境中,假定用户 A 和用户 B 分别是 A 域、B 域的用户,Proxy A、Proxy B 分别为 A 域、B 域的代理服务器和相应的注册服务器;C 为定位服务器,提供位置服务。UA A、UA B 分别是用户 A 和用户 B 运行在主机 A 和主机 B 上的相应程序。用户 A 和用户 B 可以同时使用 IP 电话、视频和语音信箱等设备进行通信,并可与普通电话通信。当用户 A 想要打电话给用户 B,他不需要知道用户 B 的联系地址以及 B 正在使用何种设备,只需要输入 B 的逻辑地址 SIP URI: B@b.sflphone.org 即可。Proxy A 接收到的 UA A 的 INVITE 请求,分析 To 头域值为 B@b.sflphone.org,在自己的数据库中没有查找其位置,转而通过查询定位服务器查得用户 B 对应的域代理服务器 Proxy B 的地址,从而将 INVITE 请求转发给 Proxy B。当 Proxy B 收到 INVITE 请求后,通过查询本域注册服务器就得到用户 B 逻辑地址对应的联系地址。若发现用户 B 对应的联系地址不只一个时,Proxy B 则采取 SIP 分支机制来转发该 INVITE 请求。即检查每个 Contact 头域值中的 qvalue 参数值的大小,按照由大到小的顺序向目的地址转发 INVITE 请求,上一个转发完成才能够开始下一个;如果前一个成功响应即返回 2xx 响应报文,则停止后面的尝试^[9]。qvalue 值的大小标识着用户使用某种通信终端的偏爱程度,qvalue 值越大,表示用户更倾向于使用该联系地址所对应的终端。例如,当用户 B 在会议室开会,他就更希望使用语音信箱通信方式,则语音信箱所对应的 qvalue 值最大, SIP 电话对应的 qvalue 值则最小,而拥有相同 qvalue 值的目的地可以并行处理。此外,分支机制既可以采用并行转发的方式,也可以通过分组的形式,即每组之间是串行的,而组内是并行的方式。在试验中,设置了多种场景对每种终端设置了不同的 qvalue 值,例如用户离开场景中 qvalue = 1 对应于语音信箱,qvalue = 0.8 对应于 IP 电话,qvalue = 0.4 对应于普通电话等。在这些场景下,系统均较好地完成了通

信过程。图 2 给出了在图 1 环境下对应于上述过程的通信时序图。

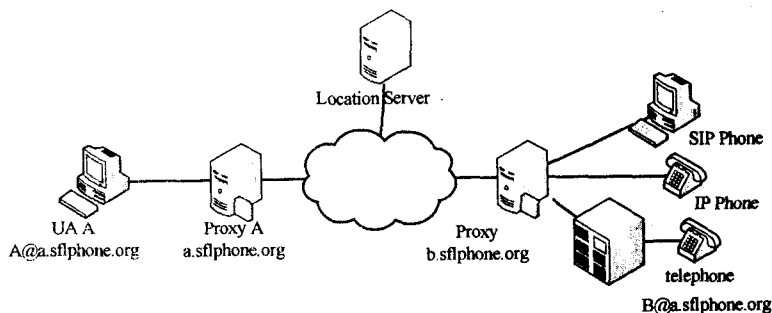


图 1 用户具有多个可用设备时的试验环境

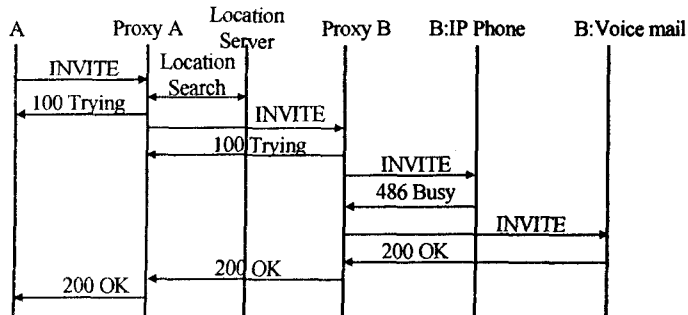


图 2 用户 B 具有多个可用设备时的通信时序图

3.2 个人移动时的通信过程

在如图 3 所示的试验环境中,Proxy A、Proxy B 分别为 A 域、B 域的代理服务器和注册服务器,C 为定位服务器,提供位置服务。C1、C2 为域 B 的用户,C3 为域 A 的用户。当用户 C1 移动到 A 域时,它无法直接通过 B 域的代理服务器/注册服务器进行工作,转而联系 Proxy A,于是 Proxy A 查询定位服务器 C 得知 B 域的代理服务器 Proxy B 的地址,并通过它向 B 域的注册服务器进行注册。当 C2 想要与 C1 通话,发送 INVITE 请求到 Proxy B,Proxy B 收到该请求后,查询本地注册服务器后发现 C1 已经漫游到 A 域,将 INVITE 请求转发给 Proxy A;Proxy A 收到请求后查询本地注册服务器中的漫游地址发现 C1 的联系地址(即实际 IP 地址)就将 INVITE 请求转发给 C1。一旦用户 C1 接听,C1 和 C2 就可以通话了。同样,当 A 域的用户 C3 想要与 C1 通话,发送 INVITE 请求到 Proxy A,Proxy A 通过查询本地数据库发现 C1 用户已经漫游到本域,就将 INVITE 请求直接转发给 C1 而不需要再转发给 C1 所属域的代理服务器 Proxy B。C1 收到后,待用户接听后连接建立,C1 和 C3 就可以通话了。上述过程对应的通信时序过程如图 4 所示。

4 结束语

SIP 协议提供了对应用层多种移动性支持,为解决底层协议对移动性支持所存在的种种问题提供了有

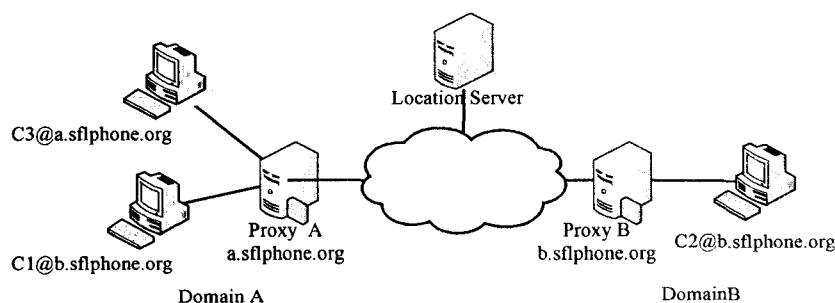


图 3 用户 C1 移动时的试验环境

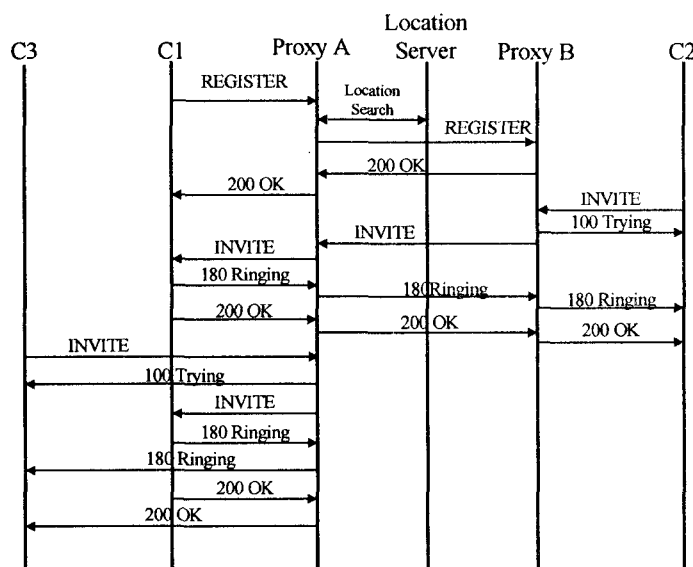


图 4 用户 C1 移动时的通信时序图

益的补充。随着人们对移动通信要求的提高,基于 SIP 在应用层解决移动性问题将具有更加广阔的应用前景。后继工作包括采用 SIP 机制来研究解决终端移

动性、会话移动性以及服务移动性所面临的技术问题。

参考文献:

- [1] Popescu L. Supporting Multimedia Session Mobility using SIP[C]// Communication Networks and Services Research Conference 2003. Moncton, New Brunswick, Canada; [s. n.], 2003.
- [2] Rosenberg J. SIP: Session Initiation Protocol[S]. IETF, RFC 3261. 2002.
- [3] 张永强, 张捍东, 赵金宝. SIP 协议栈研究[J]. 计算机技术与发展, 2007, 17(11): 55-57.
- [4] 张 荣, 武 波. SIP 协议的应用研究[J]. 计算机技术与发展, 2006, 16(6): 71-73.
- [5] Vakil F. Mobility Management in a SIP Environment[S]. IETF, Internet Draft, 2000.
- [6] Schulzrinne H, Wedlund E. Application - Layer Mobility Using SIP[J]. Mobile Computing and Communications Review, 2001, 4(3): 47-57.
- [7] 黄 斌, 李秉智. 基于 SIP 的 VoIP 的移动性研究[J]. 微计算机技术, 2008, 24(1): 145-147.
- [8] Wedlund E, Schulzrinne H. Mobility support using SIP [C]// ACM/IEEE International Conference on Wireless and Mobile Multimedia. Seattle, Washington: [s. n.], 1999.
- [9] 胡凌凌, 彭容修. SIP 协议在一个 IP 电话模型中的实现 [J]. 微机发展 (现更名: 计算机技术与发展), 2005, 15(2): 100-102.

(上接第 108 页)

④目前, 聚类融合方法科学研究和具体应用上前景广泛, 其研究有重要意义。

参考文献:

- [1] 韩家炜. 数据挖掘: 概念与技术[M]. 北京: 清华大学出版社, 2000.
- [2] 李雄飞, 李 军. 数据挖掘与知识发现[M]. 北京: 高等教育出版社, 2003.
- [3] 阳琳贤, 王文渊. 聚类融合方法综述[J]. 计算机应用研究, 2005(12): 8-10.
- [4] 蒋盛益. 基于投票机制的融合聚类算法[J]. 小型微型计算机系统, 2007(2): 306-309.
- [5] 邹远强, 李国徽, 赵梓屹. 基于遗传和蚁群算法融合的聚类新方法[J]. 科学技术与工程, 2006(23): 4700-4705.
- [6] Strehl A, Ghosh J. Cluster Ensembles: A Knowledge Reuse Framework for Combining Multiple Partitions[J]. Journal of Machine Learning Research, 2003, 3(3): 583-617.
- [7] Minaei - Bidgoli B, Topchy A, Punch W F. A Comparison of

Resampling Methods for Clustering Ensembles [C] // Int. Conf. on Machine Learning, Models, Technologies and Applications (MLMTA 2004). [s. l.]: [s. n.], 2004: 939-945.

- [8] Fern X Z, Brodley C E. Random projection for High Dimensional Data Clustering: A Cluster Ensemble Approach [C] // Proceedings of the 20th International Conference on Machine Learning. [s. l.]: [s. n.], 2003: 186-193.
- [9] Minaei - Bidgoli B, Topchy A, Punch W F. Ensembles of Partitions via Data Resampling [C] // Proceedings International Conference on Information Technology, Coding and Computing (ITCC 2004). [s. l.]: [s. n.], 2004: 188-192.
- [10] Topchy A, Minaei - Bidgoli B, Jain A K, et al. Adaptive Clustering Ensembles [C] // Proceedings of the 17th International Conference on Pattern Recognition (ICPR 2004). [s. l.]: [s. n.], 2004: 272-275.
- [11] Fred A L. Finding Consistent Clusters in Data Partitions [C] // Proceeding of the 2nd International Workshop on Multiple Classifier Systems, Volume 2096 of Lecture Notes in Computer Science. [s. l.]: Springer, 2001: 309-318.