计算机研究与发展
Journal of Computer Research and Development
ISSN 1000-1239/CN 11-1777/TP
47(7): 1211-1218, 2010

一种基于 P2P-SIP 的可管理和可运维 VoIP 系统架构

陈 曦[1,2] 李振宇[1] 谢高岗[1]
[1](中国科学院计算技术研究所 北京 100190)
[2](中国科学院研究生院 北京 100049)
(zyli@ict.ac.cn)

An Administrable and Manageable P2P-SIP Based VoIP System

Chen Xi[1,2], Li Zhenyu[1], and Xie Gaogang[1]
[1](*Institute of Computing Technology, Chinese Academy of Sciences, Beijing* 100190)
[2](*Graduate University of Chinese Academy of Sciences, Beijing* 100049)

Abstract SIP-based VoIP systems use C/S architecture, which brings two intrinsic issues when the user number grows to a large scale, namely single point of failure and performance bottleneck problems. The systems based on P2P-SIP are scalable and fault-tolerant. However, existing systems cannot meet the administrable and maintainable requirements from the service providers. The authors propose an administrable and manageable VoIP system based on P2P-SIP, called AMAP. In an AMAP system, a number of servers, deployed by service providers, form a DHT-based service overlay network, which provides services for the user agents, such as registration, session initialization, etc. The AMAP system leverages extended SIP messages for the overlay maintenance and, thus, is compatible with traditional SIP-based VoIP systems. To further improve the fault tolerance and relieve the hot spot problem, the AMAP system replicates and caches the user information on the service overlay. A distributed meta-data aggregation and dissemination algorithm is proposed to support the management. This algorithm leverages the tree structures which are implicitly defined on the service overlay based on DHT nodes' finger table. The analysis results show that, compared with existing VoIP systems based on P2P-SIP, the AMAP system reduces the maintenance overhead by 80%. A prototype has been implemented and some functional tests have been performed.

Key words session initiation protocol (SIP); peer-to-peer (P2P); voice over IP (VoIP); distributed Hash table (DHT); Chord

摘 要 基于 SIP 的 VoIP 系统由于其客户端/服务器的服务架构,在用户规模增大时会引发其固有的单点失效和性能瓶颈问题. P2P 和 SIP 融合的 VoIP 系统具有良好的扩展性和容错能力,但是现有系统并不能满足可管理和可运维的要求. 从运营商的角度考虑,提出了一种基于 P2P-SIP 的可管理、可运维的 VoIP 系统架构 AMAP. 在 AMAP 系统中,由少量部署的服务器按照 DHT 协议组成服务覆盖网络. 服务覆盖网络使用扩展的 SIP 信令来维护,为用户代理的注册和会话建立、系统的管理和运维等服务提供支持,可有效地与现有 VoIP 系统兼容. AMAP 系统借助备份和缓存技术提高容错能力,缓解服务热点问题. 设计了分布式元数据收集和分发算法,满足了运维信息收集和管理策略下达的需求,对比分析了

收稿日期:2009-03-23;**修回日期**:2009-09-11
基金项目:国家自然科学基金项目(60873242,60903207);国家"九七三"重点基础研究发展计划基金项目(2007CB310702);中国科学院重大科研装备研制基金项目(YZ200824)

系统的维护开销，结果表明与现有基于 P2P-SIP 的 VoIP 系统相比，AMAP 系统减少约 80%的开销. 详细分析了 AMAP 系统的功能组成和所涉及的关键技术，并实现和测试了原型系统.

关键词 会话发起协议；对等网络；VoIP；分布式 Hash 表；Chord

中图法分类号 TP393

0 引 言

近年来，随着 Internet 带宽的增加，基于会话发起协议（session initiation protocol，SIP）的 VoIP（voice over IP）系统发展极为迅速. 系统为每个域提供一个注册服务器和一个代理服务器. 域内的用户代理（user agent）在注册服务器上登记地址信息完成注册，代理服务器接受用户代理的会话请求并定位被叫用户. RFC3261[1]，RFC3263[2]详细描述了 SIP 会话流程与定位服务标准.

随着用户规模越来越大，注册服务器和代理服务器将成为系统的性能瓶颈. 一旦服务器出现问题，整个系统就不能正常运行；当用户数目庞大时，单台服务器不能及时处理所有用户的请求. 为每个域部署多台服务器虽然能缓解上述问题，但这需要大量的同步和配置操作，而且实现各个域之间服务器的负载均衡也比较困难.

P2P（peer-to-peer）技术已经被广泛应用在文件共享、实时流媒体以及动态信息分发[3]等方面. 近年来，P2P 和 SIP 融合的思想被逐步提出并得到了广泛关注[4]. 基于 P2P-SIP 的 VoIP 系统可以充分利用 P2P 技术的优势，实现用户信息的分布式存储、备份和负载的均衡[5].

Skype[6]是基于 P2P 技术的 VoIP 系统，它将节点分为普通节点与超级节点 2 种，并采用混合式 P2P 结构[7]. 上层结构是由能力高的超级节点（用户代理）组成的覆盖网络，而普通节点和超级节点连接构成下层结构. Skype 使用私有的会话协议，不能和基于 SIP 的 VoIP 系统兼容.

Bryan 等人提出了基于 P2P-SIP 的 SOSIMPLE 框架[8]. 用户终端按照基于 DHT（distributed Hash table）的 Chord 协议[9-10]组成平面结构. 虽然结构化 P2P 网络结构支持资源的快速定位，但是基于 Chord 协议的平面结构维护开销较大[11]，不能有效适用于用户终端动态性较大的应用场景.

Singh 等人[12]提出了基于 P2P-SIP 的层次化 VoIP 系统 SIPPeer. 该系统借助能力较高的节点组成基于 DHT 的上层结构. 上层节点的动态性将会增加 DHT 的维护信令开销. 更重要的是，上层用户节点的不可信性不利于管理和运维操作. 不可信节点恶意收集整个系统的信息或发送错误的管理消息都可能破坏整个上层结构的稳定性.

上述系统存在的共同问题是不能有效解决用户代理节点的不可信问题和动态性问题. 不可信的节点可能不合作，甚至提供错误信息，而节点的动态性（随意加入和离开）会带来大量的维护开销. 这些都限制了现有系统对可管理和可运维的支持.

本文从运营商的角度考虑，提出了一种基于 P2P-SIP 的可管理、可运维 VoIP 系统 AMAP（administrable and manageable VoIP system based on P2P-SIP）. 在 AMAP 系统中，由少量部署的服务器按照 DHT 协议组成支持管理和控制功能的服务覆盖网络. 服务覆盖网络使用扩展后的 SIP 信令构建和维护，支持传统 SIP 用户代理的注册和会话建立方案. AMAP 系统借助备份和缓存技术增加容错能力、缓解服务热点问题. 进一步，为了支持运维信息收集和管理策略的下达等服务，提出了构建于服务覆盖网络上的分布式元数据收集和分发算法. 区别于现有系统，AMAP 系统具有以下优势：

1. 服务覆盖网络是由运营商部署的、可信且稳定的服务器构成的，不仅减少了维护开销，而且为用户认证、计费、流量统计等功能的实现奠定了基础. 另一方面，可信的服务覆盖网络为负载均衡、资源定位提供有效支持.

2. 构建于服务覆盖网络上的分布式元数据收集和分发算法利用了隐含在 DHT 路由表中的树状结构，以较低的开销实现了运维等数据的收集和管理策略的下发，满足了运维需求.

3. 系统支持服务覆盖网络中服务器的灵活增加和删除. 运营商可以根据系统的运维状况删除或者增加服务器. 借助 DHT 协议，服务器的增加和删除仅需要很少的配置.

4. 分析得出，由于部署服务节点的稳定性，与现有基于 P2P-SIP 的 VoIP 系统相比，AMAP 系统可节约 80%的控制开销.

1 AMAP 系统设计

1.1 AMAP 系统概述

AMAP 系统是面向服务运营商的可管理、可运

维的 VoIP 系统. 运营商部署的服务器按照基于 DHT 的 Chord[8-9] 协议组成支持管理和运维功能的覆盖网络. 该覆盖网络是系统的核心组成部分，实现了用户管理、服务器管理、用户接入和定位、NAT 穿越以及元数据收集等功能，称为服务覆盖网络. 为了叙述方便，我们称服务覆盖网络上的服务器为服务节点(service node，SN).

图 1 给出了 AMAP 系统框架和主要功能. 服务节点的主要功能包括：服务节点管理、用户代理管理、运维支持和 NAT 穿越等.

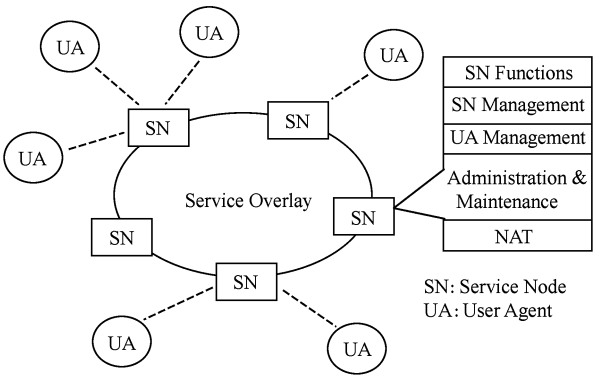

Fig. 1　Structure and functions of AMAP system.

图 1　AMAP 系统的框架和服务的主要功能

1. 服务节点管理模块负责服务覆盖网络的建立和维护，实现了基于 Chord 协议的服务覆盖网络的维护功能，处理服务节点的加入、离开以及失效等动态行为. 该模块使用了扩展的 SIP 消息来实现内部通信. 2.1 节详细描述了 SIP 消息的扩展.

2. 用户代理管理模块负责用户代理的注册、信息存储以及定位等操作. 服务节点接收用户代理的注册信息(REGISTER)，借助 Chord 协议分布式地保存用户代理的信息. 服务节点同样接收用户代理的会话建立请求(INVITE)消息，协助用户代理的初始化会话. 1.2 节讨论了注册和会话建立的过程.

3. 运维支持模块实现对运维相关操作的有效支持，包括运维信息收集和管理策略的下达等. 2.2 节详细介绍构建于服务覆盖网络上的分布式信息收集和策略下达算法. 需要说明的是，若用户代理支持通话状态的周期性汇报，该模块可根据通话的时间等信息实现 VoIP 增值业务.

4. NAT 穿越模块应用 STUN[13] 协议实现用户代理的 NAT(network address translator)[14] 穿越. 服务节点作为 STUN 服务器协助用户代理确定是否处在 NAT 后、NAT 的类型以及所对应的公网地址信息.

1.2　AMAP 系统服务

用户代理随机选择某个服务节点连接，并缓存若干个其他服务节点的信息作为备份. 若当前服务节点失效，用户代理可主动和其他服务节点连接.

注册时，用户 U 向与其连接的服务节点 A 发送注册消息(REGISTER)提交自己的信息. 服务节点使用 Chord 提供的相容 Hash(consisting Hash)来获取用户 ID($ID(U)=Hash(U)$)，并借助 Chord 路由协议把用户 U 的信息存储在服务覆盖网络中 $ID(U)$ 的后继(successor)节点上. 这样，所有用户代理的信息将分布式地存储在服务覆盖网络上的节点中.

当用户 V 呼叫 U 时，向与其连接的服务节点 B 发送会话建立请求(INVITE)消息. 服务节点 B 提取被叫的用户名 U，获得 $ID(U)$，并利用 Chord 路由协议获得 U 的地址信息. 最后，服务节点 B 作为

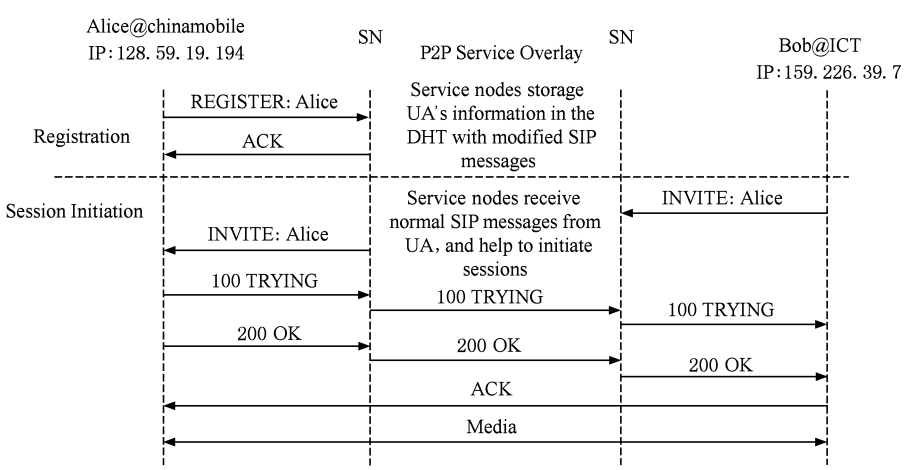

Fig. 2　Message flow of user registration and session initiation.

图 2　用户注册和会话建立的消息流程

代理节点把 INVITE 消息转发给用户 U. 接下来的会话建立过程和标准 SIP 协议类似. 用户注册和会话建立的消息流程如图 2 所示. 由于 Chord 路由算法的复杂度为 $O(\log n)$(n 为服务节点的数目)，所以注册和会话建立过程中所涉及的定位操作所需时间是很短的.

为了提高数据的可用性，用户代理 U 的信息被存储在从 $ID(U)$ 开始的、连续的若干个节点上. 存储节点的多少与可用性要求和服务节点的失效概率相关. 当负责用户代理 U 的服务节点失效时，U 的信息在失效服务节点的后继上依然是可用的，避免了单点失效问题. 用户代理信息更新后的同步由 $ID(U)$ 的后继节点控制，以保证备份信息的正确性.

用户的活跃程度是不同的：某些用户可能会被多个人不断地请求，这会使得保存活跃用户信息的服务节点成为服务的热点. 采用负载均衡的方法可以缓解热点的负担，进一步提高系统的服务性能[15]，在 AMAP 中，为了缓解服务热点问题和加快用户信息的查询，系统根据服务节点收到查询请求的速率来决定是否沿查询路径缓存用户信息. 当服务节点收到针对某个用户的查询请求的速率高于某个阈值时，该节点在回复的消息上增加标识，显式地要求返回路径上的节点缓存查询结果，从而缓解了热点问题.

每个缓存信息都被赋予一个生存期，超过生存期后缓存信息将被标识为过期. 服务节点判断是否沿路径缓存的阈值是由该节点能够处理的查询请求的能力决定的. 更好的方法是根据服务网络中所有节点的平均负载水平作为阈值. 当收到查询请求的速率大于系统平均负载水平时，沿路径缓存查询结果. 平均负载水平可以借助 2.2 节描述的运维数据收集和分发算法来获取和分发.

另一方面，考虑到某个特定用户呼叫的对象往往比较集中，所以用户代理或者与用户代理连接的服务节点缓存最近呼叫过的用户信息. 同样，缓存数据信息被赋予一个生存期. 在 AMAP 系统中，最近呼叫过的用户信息缓存在与呼叫发起用户连接的服务节点上.

最后需要说明的是，虽然用户代理可以任意选择服务节点连接，但服务节点可根据自己的负载水平来决定是否允许连接. 当服务节点的负载水平远远高于服务网络中所有节点的平均负载水平时，它将拒绝接收新的用户代理的连接，或者返回重定向消息.

2 SIP 协议扩展和元数据管理

2.1 SIP 协议的 P2P 扩展

为了兼容传统 VoIP 用户代理，服务节点实现了对标准 SIP 协议的支持，接收并解析用户代理发来的注册和会话建立消息. 此外，考虑到与现有 VoIP 系统兼容并保持 SIP 协议的简单性、可扩展性等优点，AMAP 系统扩展了 SIP 消息，使用扩展后的消息来构建和维护服务覆盖网络.

SIP URI(uniform resource indicators)是通过 SIP 呼叫他人的 SIP 地址方案. 在系统中，服务节点的 SIP URI 定义为：sip：*NodeID*@*hostIPport*；user＝node. 其中，*hostIPport* 为服务节点的 IP 地址和端口号，服务节点标识(*NodeID*)是 Hash(Chord 提供的相容 Hash)*hostIPport* 得到的. user＝node 表示该消息是服务节点发出的，以区别用户代理发送的消息. 例如，地址为 10.21.20：5274 的服务节点，若 Hash 该地址得到的节点标识为 7c，则该节点的 URI 为：sip：7c@10.21.2.20：5274；user＝node.

为了在服务覆盖网络的维护消息中显式地标明消息的源和相应的指针表结构，在用来维护操作的 SIP 消息中增加了 DHT-NodeID 和 DHT-Link 两个扩展头字段. DHT-NodeID 字段表示该消息的源，而 DHT-Link 字段表示前驱、后继或者指针表的一项.

以 DHT-NodeID 扩展头字段开始的消息内容的基本格式为 DHT-NodeID：〈*NodeURI*〉；algorithm，其中 algorithm 字段说明了计算节点 ID 用到的散列算法.

以 DHT-Link 扩展头字段开始的消息内容的基本格式为 DHT-Link：〈*NodeURI*〉；link＝type，其中 type 字段的取值为 P0，S0，F0，F1，…，Fm(m 为服务节点的标识长度). P0 和 S0 分别表示这一项为前驱和后继，Fi 表示这一项为指针表(finger table)的第 i 项.

图 3 给出了一个完整的维护操作的 SIP 消息格式. 该消息是服务覆盖网络上的服务节点 10.21.2.20：5150 对加入节点 10.21.2.22：5274 的请求响应. 在节点加入前，覆盖网络中只有节点 10.21.2.20：5150，所以此消息中 DHT-Link 字段包含的前驱、后继以及指针表中的所有表项都指向自己. 待加入节点收到该消息后根据 Chord 协议构建自己的前驱、

后继以及指针表，并利用扩展后的 SIP 消息发送自己的信息.

```
200 OK Response:
Via: SIP/2.0/UDP 10.21.2.22:5274;branch=z9hG4bKc15vv6Dc6QKKQ
From: <sip:7c@10.21.2.22:5274;user=node>;tag=caX9NDZgS38vc
To: <sip:7c@10.21.2.22:5274;user=node>;tag=cvepX5pryXNNm
Call-ID: 0c376847-219d-122c-0bbc-0030488e81aa
CSeq: 106593136 REGISTER
Contact: <sip:7c@10.21.2.22:5274;user=node>
Content-Length: 0
DHT-Link: <sip:00@10.21.2.20:5150;user=node>;link=P0
DHT-Link: <sip:00@10.21.2.20:5150;user=node>;link=S0
DHT-Link: <sip:00@10.21.2.20:5150;user=node>;link=F0
DHT-Link: <sip:00@10.21.2.20:5150;user=node>;link=F1
.....
DHT-NodeID: <sip:00@10.21.2.20:5150;user=node>; algorithm=sum-8
```

Fig. 3　Extended SIP message to maintain P2P overlay.

图 3　针对 P2P 维护操作扩展的 SIP 消息

2.2　分布式元数据收集和分发算法

为了支持管理和运维功能，AMAP 系统不仅需要向运营商提供服务覆盖网络的运维情况（如服务节点的负载情况等），而且需要提供策略下达服务（如阻止某些用户接入等）. 运维状况信息和管理策略通称为元数据. 由于服务覆盖网络规模较大，借助某个服务器集中式地收集和分发元数据是不合理的. AMAP 系统借助隐含在 Chord 服务覆盖网络中的树状结构来分布式地收集和分发元数据.

在 Chord[9] 协议中，每个节点拥有一个 m 位的标识，标识空间大小为 2^m. 标识空间映射到 Chord 环状结构上，分散在环状结构上的每个节点负责一部分标识区域. 服务节点 A 的指针表包含 $O(\log n)$ 个其他节点的信息（n 为服务节点的数目），其中指针表第 i 项是沿顺时针方向与 A 的距离为 $1/2^{\log n-i}$ 个环的节点（$0\leqslant i<\log n$）.

在 Chord 环中，任意节点 A 都可以借助路由协议定位到其他所有的节点. 如果把从 A 出发、到其他节点的定位消息路径连接起来，则会形成以节点 A 为根的树状结构. 我们可以借助这个过程来隐式地构建树状结构收集和分发元数据信息. 起始节点 A 拥有整个标识空间，它利用自己的路由表把标识空间分为大小不等的若干块，指针表中的每个节点负责一块，这些节点将作为节点 A 的孩子节点. 节点 A 的孩子节点继续利用自己的路由表来划分所负责的区域. 以此类推，直到划分所得区域内不再有节点为止. 当服务节点 x 需要收集或者分发元数据信息时，它调用算法 1 所描述的函数 $region_partition(region_x)$ 来构建树状结构，其中 $region_x=[x,(x-1+2^m)\bmod 2^m]$，$x.finger[j].node$ 指节点 x 的指针表的第 j 项“succ.”域对应的节点.

算法 1. 服务节点 x 构建树状结构的算法.

$x.region_partition(region_x)$

① if $x.id+1>region_x.\mathrm{end}$ then /* 节点 x 是这个区域内唯一的节点，无需继续划分区域 */

②　return

③ end if

④ for j from m to 1 do

⑤ if $x.finger[j].node\in[region_x.\mathrm{start},region_x.\mathrm{end}]$ then

⑥　$x.children=x.children\cup x.finger[j].node$;

⑦　$sub_region=[x.finger[j].node,region_x.\mathrm{end}]$;

⑧　$x.finger[j].node.region_partition(sub_region)$;

⑨　$region_x.\mathrm{end}=x.finger[j].node-1$;

⑩　end if

⑪ end for

上述建树过程类似于 FPCast[3] 中的树状结构建立过程，不同的是 AMAP 系统不限制每个节点最多拥有的孩子节点数目. 这是因为树状结构仅用来传输元数据信息，给服务节点带来的负载不会很大. 可以分析得到树状结构的高度为 $O(\log n)$[3].

元数据信息可沿着树状结构自底向上发送，中间节点可对元数据作处理后发送给父节点，根节点会收集到整个系统的元数据信息. 元数据也可沿着树状结构自顶向下传输. 需要说明的是，树状结构中的节点 A 到孩子节点的链路集是节点 A 到路由表中的节点的链路集的子集. 这些链路本来就存在于服务覆盖网络中，所以并不需要额外维护树状结构.

3　系统性能分析

通过建立数学模型对服务覆盖网络的性能进行分析. 可以对 AMAP 系统大规模部署的情况有所估计[16]. 用 M 表示服务节点的数目，用 N 表示用户代理的数目，令 $\beta=N/M$ 为服务节点的平均服务率.

3.1　服务性能要求

每个服务节点存储系统中一部分用户代理的信息，针对这些用户的请求将会发送到该服务节点. 假设每个用户代理发起会话的平均速率是 q，在不使用任何缓存技术的情况下，服务节点的平均处理查找请求的速率为：

$$R=q\cdot N/M=\beta\cdot q. \tag{1}$$

假设每个用户代理每分钟平均发起 1 个会话操作，即 $q=1/60$，设用户代理的数目 $N=1000000$，服务节点的数目 $M=1000$，那么服务节点的平均处理

查找请求的速率 $R=16$ 个/s. 由于用户信息近似平均分配到服务节点上，在用户活跃程度相同的情况下，每个服务节点处理查找请求的速率相近. 所以，对服务器的性能要求是完全可以接受的.

3.2 维护开销比较

在 SIPPeer[11] 系统中，用 M 表示上层节点的数目，N 表示所有参与的用户代理数目. 在 SIPPeer 中，假设上层节点的到达过程服从 Poisson 分布，而且到达速率为 λ，上层节点的逗留时间 T 服从分布 $f(T)$，那么其平均逗留时间 $D=\int Tf(T)\mathrm{d}T$. 测量结果[17]显示平均逗留时间 D 大约是 300 s. 可以证明 SIPPeer 中上层节点的数目 $M=\lambda D$.

证明. 如果节点的逗留时间 T 的分布密度是 $f(T)$，那么逗留时间在区间 $(T,T+\mathrm{d}T)$ 内的节点的到达速率是 $\lambda_T=\lambda(f(T)\mathrm{d}T)$. 从而，上层节点数目是：

$$M=\int_T \lambda_T T=\int_T \lambda(f(T)\mathrm{d}T)T=\lambda D. \qquad (2)$$

证毕.

假设 $M=1000$，$D=300$ s，则 $\lambda=3.33$ 个/s. 这表明拥有 1000 个上层节点的 SIPPeer 系统平均每秒有 3.33 个上层节点加入和离开. 如此高的动态性必将带来更多的维护开销.

DHT 覆盖网络的维护开销可以分为 2 部分：一部分是周期性维护操作带来的开销，另一部分是处理节点动态性带来的开销. 由 Chord 协议的分析可知节点更新路由表和处理节点动态性均需 $(\log M)^2$ 条消息. 设 r_s 为每个 DHT 节点确认后继节点的频率，r_f 为每个 DHT 节点更新路由表的频率，那么 DHT 覆盖网络中平均每个节点产生控制消息的速率为：

$$MS=r_s+r_f(\log M)^2+\frac{\lambda(\log M)^2}{M}. \qquad (3)$$

式(3)中前 2 项为 DHT 正常的更新维护信令开销，第 3 项则是由节点动态性带来的修复 Chord 的维护开销.

对于 SIPPeer 系统，假设上层节点更新周期为 1 min，即 $r_s=r_f=1/60$. 当 $M=1000$，$\lambda=3.33$ 个/s 时，$MS=2$. 在 AMAP 系统中，服务节点平均逗留时间较长. 设更新周期为 300 s，即 $r_s=r_f=1/300$，那么 $MS=0.35$. 所以与 SIPPeer 相比，AMAP 系统可节约 82.5%的维护开销.

上述开销的节约主要原因在于 AMAP 系统中稳定的服务覆盖网络. 对于 SIPPeer 系统，上层节点是用户节点，动态性较大，需要更多的维护消息. 而 AMAP 系统中由运营商部署的服务器更加稳定和可信，大幅度节约了维护开销.

4 系统实现与测试

我们实现了 AMAP 原型系统并在在实验室环境中使用 5 台普通 PC 模拟服务节点. 每台 PC 开启 2 个端口来模拟 2 个服务节点，共有 10 个服务节点. 另外 2 台 PC 作为用户代理通过服务覆盖网络进行注册、查询、建立会话、发送短消息等操作对系统进行功能测试. 测试的网络环境如图 4 所示：

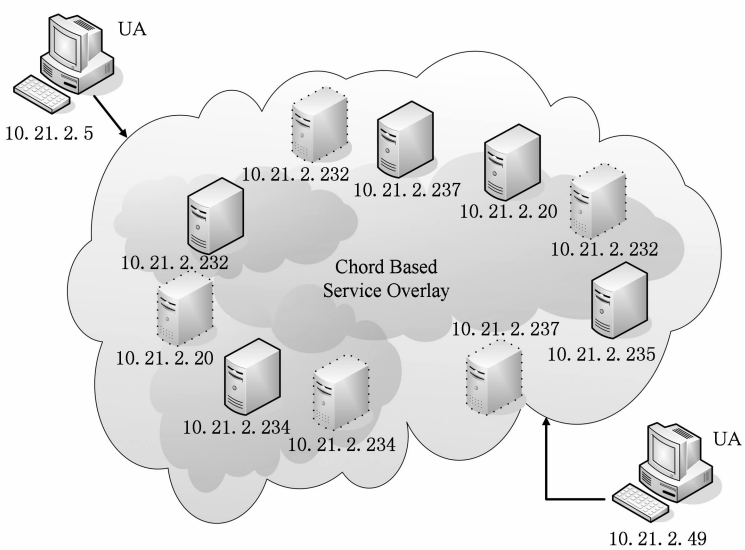

Fig. 4 Functional test environment.

图 4 功能测试的网络环境

我们使用 X-Lite[18] 传统 SIP 用户代理对 AMAP 系统进行了功能性测试，主要包括用户代理的注册和会话建立. 安装在 10.21.2.5（端口号：24464）上的用户代理使用用户名 jerry 与服务节点 10.21.2.234：5050 连接，而安装在 10.21.2.49（端口号：32846）上的用户代理使用用户名 bob 与服务节点 10.21.2.232：5000 连接，它们都已经通过各自连接的服务节点在服务覆盖网络上注册.

我们首先展开部署服务节点以构建服务覆盖网络的测试. 图 5 显示了在主机 10.21.2.20 上使用 Wireshark 工具捕获到的服务节点 3e@10.21.2.237：5154 加入服务覆盖网络的信令流程. 新节点首先向引导服务节点 00@10.21.2.20：5150 发送加入请求但被重定向到服务节点 7c@10.21.2.20：5274. 之后节点 3e 重新向节点 7c 发送加入请求，得到 200 OK 确认加入响应后成功加入服务覆盖网络. 之后捕获到的消息为节点 ba@10.21.2.237：5278 成功通过引导节点 00@10.21.2.20：5150 加入服务覆盖网络以及节点 3e 和 7c 的更新消息.

Source	Source Port	Destination	Dest Port	Protocol	Info
10.21.2.237	5154	10.21.2.20	5150	SIP	Request: REGISTER sip:10.21.2.20:5150
10.21.2.20	5150	10.21.2.237	5154	SIP	Status: 302 Moved Temporarily (1 bindings)
10.21.2.237	5154	10.21.2.20	5274	SIP	Request: REGISTER sip:7c@10.21.2.20:5274;user=node
10.21.2.20	5274	10.21.2.237	5154	SIP	Status: 200 OK (1 bindings)
10.21.2.237	5278	10.21.2.20	5150	SIP	Request: REGISTER sip:10.21.2.20:5150
10.21.2.20	5150	10.21.2.237	5278	SIP	Status: 200 OK (1 bindings)
10.21.2.237	5154	10.21.2.20	5274	SIP	Request: REGISTER sip:7c@10.21.2.20:5274;user=node
10.21.2.20	5274	10.21.2.237	5154	SIP	Status: 200 OK (1 bindings)
10.21.2.237	5154	10.21.2.20	5274	SIP	Request: REGISTER sip:7c@10.21.2.20:5274;user=node
10.21.2.20	5274	10.21.2.237	5154	SIP	Status: 200 OK (1 bindings)

Fig. 5　Trace of overlay maintaining messages on host 10.21.2.20.

图 5　主机 10.21.2.20 上捕获的服务覆盖网络维护信息

另一组实验中，主机 10.21.2.5 的用户 bob 希望与处在主机 10.21.2.49 上的 jerry 建立会话. 图 6 显示了 bob 的服务节点 10.21.2.232：5000 的处理过程以及 bob 的用户代理客户端成功与 jerry 建立通话. 服务节点（ID：0xde）利用 Hash 函数得到被叫用户 jerry 的 ID，并通过查找指针表发现此 ID 的后继就在指针表中（finger：－1）. 在获得 jerry 的地址信息后，服务节点在 bob 发起的 INVITE 消息中添加 via 字段说明消息传递的路径（指明 jerry 的用户代理回应消息的路径），最后转发修改后的 INVITE 消息到 jerry 的用户代理 10.21.2.5：24464.

Fig. 6　User bob initiates a session through service node.

图 6　bob 连接服务节点成功建立会话

5　小结和下一步工作

本文提出了基于 P2P-SIP 的可管理、可运维 VoIP 系统 AMAP. 部署的服务节点按照 Chord 协议组成服务覆盖网络，使用扩展的 SIP 消息来维护服务覆盖网络，从而有效支持传统 VoIP 用户终端的无缝接入. 系统使用了副本技术来提高数据的可用性，借助缓存技术来提高服务性能和负载均衡效果. 分布式元数据收集和分发算法有效支持了运维信息的收集和管理策略的下达，满足了运维和管理的需求. 与现有基于 P2P-SIP 的 VoIP 系统相比，AMAP 系统减少约 80％的开销. 对原型系统的测试验证了系统功能，进一步说明了其可行性.

在本文基础上，我们将从通信安全、认证和计费等方面完善系统功能. 进一步，借助流量产生设备对系统进行压力测试，给出系统的配置建议等，为系统的部署作好准备.

参　考　文　献

[1]　Rosenberg J, Schulzrinne H, Camarillo G, et al. SIP: Session Initiation Protocol (RFC 3261) [S]. Fremont, CA: Internet Engineering Task Force, 2002

[2] Rosenberg J, Schulzrinne H. Session Initiation Protocol (SIP): Locating SIP Servers (RFC 3263) [S]. Fremont, CA: Internet Engineering Task Force, 2002

[3] Li Z, Zhu Z, Xie G, et al. Fast and proximity-aware multi-source overlay multicast under heterogeneous environment [J]. Journal of Computer Communication, 2009, 32(2): 257-267

[4] Bryan D, Broadbent T. P2PSIP [OL]. 2009 [2009-03-23]. http://www.p2psip.org

[5] Singh K, Schulzrinne H. Peer-to-peer Internet telephony using SIP [C] //Proc of the Int Workshop on Network and Operating System Support for Digital Audio and Video. New York: ACM, 2005: 63-68

[6] Skype Company. Skype [OL]. 2009 [2009-03-23]. http://www.skype.com

[7] Baset S, Shulzrinne H. An analysis of the skype peer-to-peer Internet telephony protocol [C] //Proc of IEEE INFOCOM. Piscataway, NJ: IEEE, 2006:1-11

[8] Bryan D, Lowekamp B, Jennings C. SOSIMPLE: A serverless, standards-based P2P SIP communication system [C] //Proc of the Int Workshop on Advanced Architectures and Algorithms for Internet Delivery and Applications. Piscataway, NJ: IEEE, 2006: 42-49

[9] Stoica I, Moms R, Karger D, et al. Chord: A scalable peer to peer lookup service for Internet applications [C] //Proc of the ACM SIGCOMM. New York: ACM, 2001: 149-160

[10] Dabek F, Brunskill E, Kaashoek M F, et al. Building peer-to-peer systems with chord, a distributed lookup service [C] //Proc of the 8th Workshop on Hot Topics in Operating Systems. Piscataway, NJ: IEEE, 2001: 81-86

[11] Castro M, Jones M, Kermarrec A M, et al. An evaluation of scalable application-level multicast built using peer-to-peer overlays [C] //Proc of the IEEE INFOCOM. Piscataway, NJ: IEEE, 2003: 1510-1520

[12] Singh K, Schulzrinne H. SIPPEER: A session initiation protocol (SIP)-based peer-to-peer Internet telephony client adaptor [R]. New York: Columbia University, 2004

[13] Rosenberg J, Weinberger J, Huitema C, et al. STUN-simple traversal of User Datagram Protocol (UDP) Through network address translators (NATs) (RFC 3489) [S]. Fremont, CA: Internet Engineering Task Force, 2003

[14] Thernelius F. SIP, NAT and Firewalls [D]. Stockholm, Sweden: Department of Teleinformatics, Kungl Tekniska Hogskolan, 2000: 1-69

[15] Li Zhenyu, Xie Gaogang. A load balancing algorithm for DHT-based P2P systems [J]. Journal of Computer Research and Development, 2006, 43(9): 1579-1585 (in Chinese)
(李振宇，谢高岗. 基于 DHT 的 P2P 系统的负载均衡算法 [J]. 计算机研究与发展，2006，43(9)：1579-1585)

[16] Xia Qizhi, Xie Gaogang, Min Yinghua, et al. IS-P2P: Index-based P2P structured networks [J]. Chinese Journal of Computers, 2006, 29(4): 602-610 (in Chinese)
(夏启志，谢高岗，闵应骅，等. IS-P2P：一种基于索引的结构化 P2P 网络模型[J]. 计算机学报，2006，29(4)：602-610)

[17] Bishop M, Rao S, Sripanidkulchai K. Considering priority in overlay multicast protocols under heterogeneous environments [C] //Proc of the IEEE INFOCOM. Piscataway, NJ: IEEE, 2006: 1-13

[18] CounterPath Corporation. X-Lite [OL]. [2009-03-21]. http://www.counterpath.net/x-lite.html

Chen Xi, born in 1983. Master. His main research interests include P2P networks, network measurement and analysis.

陈 曦，1983 年生，硕士，主要研究方向为 P2P 网络、网络测量与分析.

Li Zhenyu, born in 1980. PhD and assistant professor of the Institute of Computing Technology, Chinese Academy of Sciences. His main research interests include P2P networks, network measurement and analysis.

李振宇，1980 年生，主要研究方向为 P2P 网络、网络测量与分析.

Xie Gaogang, born in 1974. PhD, and professor of the Institute of Computing Technology, Chinese Academy of Sciences. His main research interests include peer-to-peer computing, network measurement and QoS.

谢高岗，1974 年生，博士，副研究员，主要研究方向为对等计算、网络测试与服务质量.

Research Background

This work is supported by the National Natural Science Foundation of China under grant Nos. 60873242 and 60903207, and the National 973 Basic Research Program of China under grant No. 2007CB310702.

Most of the SIP-based VoIP systems leverage C/S architecture, which brings two intrinsic issues when the user number grows to a large scale, namely single point of failure and performance bottleneck problems. Peer-to-peer (P2P) systems which are scalable and fault-tolerant can solve those problems. P2P-SIP based VoIP system has gained lots of studies. However, existing systems cannot meet the administrable and maintainable requirements from the service providers. The dynamicity of user nodes and untrustworthy user nodes are also two difficult problems. This paper designs an administrable and manageable VoIP system based on P2P-SIP. The key technologies involved in the structure are discussed in detail. A prototype has been implemented. The analysis and experiment results show the preferences of the system.