

面向服务和信息的网络体系结构——SIONA

夏春梅 徐明伟

(清华大学计算机科学与技术系清华信息科学与技术国家实验室(筹) 北京 100084)

摘 要 随着网络需求的改变,信息中心网络 ICN(Information Centric Networking)被提出,目前提出的 ICN 方案基本存在可扩展和部署难的问题.文中提出了一种新的面向服务和信息的网络体系结构 SIONA.采用了 IP 地址加信息内容的 Hash 值作为信息名字,支持缓存加多源组播的路由模式,阐述了路由器反馈接收端驱动的多窗口的拥塞控制协议. IP 地址作为信息名字的前缀,不仅支持 IP 地址的路由,而且解决了 ICN 的可扩展和部署难问题,保证了 SIONA 仍具有 IP 网络的可扩展能力.而多源组播路由协议和拥塞控制协议的提出保证了 SIONA 更高效、更稳定和更公平的传输.文中模拟了多源分块路由算法和拥塞控制算法.通过分析和模拟,SIONA 不仅具有良好的可扩展性,而且更具高效性、安全性和易部署性.

关键词 SIONA;网络体系结构;面向服务和信息;路由转发;信息中心网络(ICN)

中图法分类号 TP393 **DOI号** 10.3724/SP.J.1016.2014.00289

SIONA: Service and Information Oriented Network Architecture

XIA Chun-Mei XU Ming-Wei

(Tsinghua National Laboratory for Information Science and Technology,
Department of Computer Science & Technology, Tsinghua University, Beijing 100084)

Abstract With the change of network requirements, information-centric network (ICN) has been proposed. Current ICN schemes generally have not good scalability and deployment. This paper advances a new service and information oriented network architecture—SIONA. In SIONA, we adopt the IP address plus the hash of information contents as the information name, present the routing model of cache plus multi-source multicast, and introduce the router-feedback, receiver-driven and multi-window congestion control protocol. Because it uses IP address as the prefix of the information name, SIONA not only supports the IP routing, but also solves the problems about scalability and deployment in ICN. Accordingly, this ensures that SIONA has the same scalability as IP networks. The multi-source multicast protocol and congestion control protocol ensure that SIONA has more efficiency, more stable and fairer transport. In this paper, we simulated Multi-Source Blocking Routing Algorithm (MSBRA) and Router Feedback Congestion Control Algorithm(RFCCA). SIONA has not only good scalability, but also better efficiency, security and easier deployment via analysis and simulation.

Keywords SIONA; network architecture; service and information oriented; routing and forwarding; information centric networking (ICN)

收稿日期:2012-05-18;最终修改稿收到日期:2013-08-15.本课题得到国家自然科学基金(61073166,61133015,61161140454)、国家“九七三”重点基础研究发展规划项目基金(2009CB320502,2012CB315803)和国家“八六三”高技术研究发展计划项目基金(2011AA01A101)资助.
夏春梅,女,1977年生,博士研究生,主要研究方向为网络体系结构和网络协议. E-mail: xcm19770312@sina.cn. 徐明伟,男,1971年生,博士,教授,主要研究领域为未来网络、网络路由和网络可视化.

1 引言

根据 IDC (International Data Corporation, 国际咨询数据公司) 2011 年的报告显示^①, 2010 年信息量突破了泽字节 (zettabyte = 10^{21}) 的水平, 2011 年的数量则增长到 1.8 泽字节, 达到 2006 年的 9 倍. 信息的爆炸式增长对现行软件、硬件、服务、安全、人员提出了新的挑战, 对信息的创建、存储、访问及处理等生命周期的管理成为了体系结构设计中考虑的核心因素^[1]. 尤其从近几年的 Web 和 CDN (Content Distribution Network, 内容分布网络) 应用可以看出, 用户对互联网的需求主要是海量信息的访问, 主机之间的通信需求已经成为历史. 在信息成为核心需求的情况下, 用户只关心信息本身, 不关心信息的存放位置. 现有完全以地址为中心的通信网络模型已经不能高效解决这一问题, 以信息为中心的网络成为研究热点.

在现有 IP 网络中, 为了迎合海量信息访问的需求, 主要采用了 Web Cache^[2]、P2P (Peer to Peer, 端到端)^[3]、CDN^[4] 等技术. 这些技术基于应用层, 对于采用 IP 地址的网络而言, 仍存在大量冗余数据传输, 造成了带宽浪费, 传输效率并不是很高.

目前, 以信息为中心的网络有相关研究, 如 CCN (Content Centric Network, 内容中心网络)^[5]、DONA (Data-Oriented Network Architecture, 面向数据的网络体系结构)^[6] 等, 这些研究也只是处于一种起步研究. CCN 以信息为中心, 网络层采用信息名字路由分组, 不考虑 IP 地址. CCN 由于从信息需求出发, 基于信息名字路由数据, 没有增加传输负载, 加上中间节点的缓存数据, 使得 CCN 的传输效率高于 IP 网络. 但 CCN 采用了 URL (Uniform/Universal Resource Locator, 统一资源定位符) 作为信息名字, 以 URL 路由和转发数据会带来比 IP 网络更大的路由膨胀问题、移动性问题等. 同时, CCN 采用了全新的网络体系结构, “全新”意味着不是对 IP 网络的增量部署, 而是一种全新的部署. 抛弃 IP 网络构造全新的 CCN, 应该是极度困难的工作. DONA 将数据层置于 IP 层之上, 底层仍然采用 IP 网络. IP 地址作为一种本地标识, 而数据名字作为一种全局路由标识. DONA 建立了一种层次化的 RH 结构, 将解析和路由合二为一, 提高了传输效率. 但是 DONA 由于采用扁平化名字, 顶层路由器

存在严重的可扩展问题. 还有其它的一些方案如 NetInf (Network of Information, 信息网络)^[7-8]、INS (Intentional Naming System)^[9]、TRIAD (Translating Relaying Internet Architecture integrating Active Directories)^[10]. 这些方案主要存在可扩展及部署难的问题. 本文提出了一种面向服务和信息的网络体系结构 SIONA, 这种结构不是一种新的网络体系结构, 而是 IP 网络的延展. 对比于传统的 IP 网络, SIONA 解决了传输高效性的问题. 对比于 CCN、DONA 等网络, SIONA 解决了可扩展性和易部署性问题.

在 SIONA 中, 采用混合方式^[11]来命名信息. 混合方式指将层次化命名方式和扁平化命名方式结合, 兼具了两种命名方式的优势, 既具备了良好的可扩展性, 又具备了持久性和自我验证性. SIONA 信息命名形式为 $P:N:S$, 其中 P 为信息所在主机的位置标识, 采用 IP 地址表示, 是信息名字的层次化部分; N 是信息自身标识, 采用信息内容的 Hash 值表示, 具有全局唯一性, 是信息名字的扁平化部分; S 是对信息提出的服务要求. 采用 IP 地址加信息 Hash 值的命名方式既能兼容传统的 IP 网络, 又能将信息作为第一公民考虑到网络层中. 既保证具有 IP 网络的可扩展能力, 又保证在现有 IP 网络中更易部署.

SIONA 采用多源组播树的路由模型, 支持信息的多源分块传输. SIONA 信息名字包括 IP 地址, 信息名字可聚合到 IP 前缀, FIB 表仍以 IP 前缀路由数据, 因此, 仍支持传统路由模式. 在 SIONA 的路由模型中, 我们提出了多源分块路由算法, 解决多源分块传输的问题. 在第 6 节, 我们构建了模拟环境模拟了多源分块路由算法. 通过模拟可见, 多源传输模型效率明显高于单源传输效率. SIONA 在路由器中增加缓存, 用以存储信息内容, 缓存中的信息内容作为信息源参与路由转发, 为需求端提供数据, 这种缓存技术节省了传输资源, 提高了传输效率.

SIONA 提出了路由器反馈的拥塞控制模型, 拥塞控制由接收方驱动. SIONA 的通信是由需求方发起的, 并没有三次握手协议, 服务器端只是被动的回应数据, 是一种典型的“拉”模式. 在这种模式下, 拥塞控制只能由需求方 (接收方) 完成. 路由器反馈时,

① The 2011 Digital Universe Study: Extracting Value from Chaos. <http://www.emc.com/collateral/demos/microsites/emc-digital-universe-2011/index.htm>

综合考虑了公平性和高效性,保证了每流最大利用带宽资源.在 SIONA 中,由于路由器中的缓存可作为数据源,所以在拥塞控制中存在不稳定信息源的问题,给 ICN 的拥塞控制带来新的挑战.而路由器反馈机制可以解决这类问题.第 6 节我们模拟了路由器反馈的拥塞控制算法,模拟结果显示这种机制具有较好的效率性、稳定性和公平性.

SIONA 支持对信息提出的服务要求,如验证信息的安全性、QoS(Quality of Service,服务质量)、有效性等.

本文第 2 节介绍相关研究工作;第 3 节概览 SIONA 结构;第 4 节详细介绍 SIONA 的命名和编址结构、分组类型、路由转发模型和拥塞控制机制;第 5 节阐述 SIONA 中的多源分块路由算法和路由器反馈的拥塞控制算法;第 6 节对两种算法进行模拟和评价;第 7 节给出了结论.

2 相关工作

以信息为中心的网络目前的研究主要分为两类.一类是不改变原协议栈结构,在网络层之上增加内容层.内容层采用数据名字,路由时作为全局标识,而网络层仍然采用 IP 地址,路由时仅为本地标识,主要代表的研究成果有 DONA、INS.另一类是采用全新的网络体系结构,协议栈中网络层采用数据名字,没有 IP 层的支撑,路由采用数据名字作为标识,典型研究是 CCN/NDN.

CCN 以 URL 作为数据名字.由于 URL 具有层次性,在转发表中,数据名字可以按照前缀进行聚合,基于数据名字路由到最优服务器.服务器按照请求回应数据,同时在中间路由器缓存数据.CCN 类似于单源内容组播,大大简化了协议栈结构,提高了路由转发效率.但是 CCN 名字最终可聚合到所有主机,而主机数量最高可达到 4G,也就是说 CCN 转发表仍然存在可扩展问题.而且域名具有更显著的扁平性,对于域名的聚合研究比 IP 地址聚合研究难度更大,CCN 的可扩展问题更严重,更难以解决.CCN 采用发送一个 Interest 分组,回应一个 Data 数据的传输控制方式(Interest 分组是请求分组的名字,Data 分组是回应的数据分组的名字),形式简单,但是公平性和效率性有待提高.CCN 不支持多源到多目的的传输.

DONA 是 2007 年提出的.DONA 提出了一种

新的命名结构 $P:L$, P 是管理实体的公钥 Hash 值, L 是数据的 Hash 值,具有全局唯一性.在 DONA 中,每个域内设置 RH(Resolution Handler,解析处理器),RH 之间的关系和策略与域之间的关系和策略完全相同.RH 既支持路由又支持解析,解析和路由同时进行.采用数据名字作为全局路由标识,而 IP 作为本地路由标识仍然工作在内容层下面.利用两类分组 register(注册分组名字)和 find(查找路由分组名字)通过 RH 结构完成路由和解析功能.DONA 与 DNS 结构相比,将解析和路由过程合二为一,具有高效性.但是,由于名字和 IP 地址之间仍然存在解析过程,增加了网络负载,浪费了网络带宽.同时,DONA 也存在可扩展问题.

INS 是 1999 年提出的,采用属性值对作为数据名字.建立了名字树,基于名字树解析和路由.底层仍然采用 IP 网络.路由采用两种模式早绑定和晚绑定,早绑定基于 IP 地址路由;晚绑定基于名字路由,IP 地址只作为一种本地化标识,利用隧道方式实现通信.INS 采用了属性值对作为数据名字,这种定义方式会带来规模和性能可扩展问题.同时解析过程是一种冗余操作,浪费带宽,INS 性能不优于 IP 网络.

基于内容的网络研究成果还有 NetInf、TRIAD 等,基本思想与上述有所雷同,这里不再赘述,详见参考文献.

启动于 2010 年的 MobilityFirst 未来网络体系结构^[12-14]基于移动性设计的信息中心网络,支持名字与位置分离,名字采用内容的 Hash 值.路由采用逐跳式路由方式,基于信息名字路由.采用全局名字解析服务实现无缝隙解析和路由.采用网内缓存存储分组,防止由于带宽的急速变化和临时的链路中断所造成的分组丢失,增加了无线网络的传输效率和可靠性.

已有研究成果主要存在可扩展问题,所以本文提出了一种新的网络体系结构 SIONA.吸纳了已有研究成果的优势,尽力解决已有研究成果的不足,具备了高可扩展性、高效性和安全性.

3 SIONA 结构

在 SIONA 中,定义了形如 $P:N:S$ 的信息名字,具体名字的详细介绍在 3.2 节.SIONA 需用的分组类型有三类,即 Register、Interest、Data 三类分

组, Register 分组是服务器向路由器注册所有的信息, Interest 是信息请求分组, Data 是返回的数据分组.

图 1 显示 SIONA 的工作原理图. 虚线代表信息注册流, 实线代表通信流. SIONA 工作原理如下:

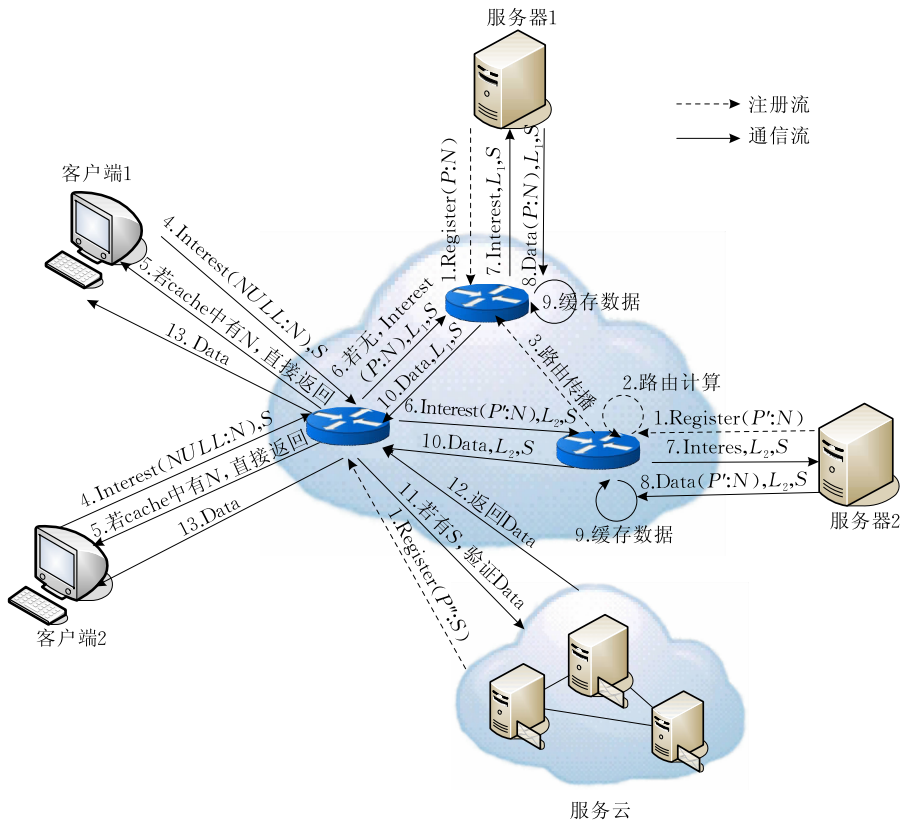


图 1 SIONA 工作原理图

当客户端 1 和 2 同时向网络请求信息名字为 N 的信息, 不考虑信息的位置 P , 同时要求对信息提供 S 服务. Interest 分组经过路由器时, 路由器首先查找缓存(在 SIONA 中叫做数据表 DT(Data Table)), 若缓存中有相应内容, 则直接返回给客户端信息. 若缓存中没有相应信息, 则合并两个 Interest 分组, 并根据 RT 表中注册信息分解 Interest 分组, 并将分解后的 Interest 分组按照地址发送. 返回信息时, 采用 Data 分组按照原路径返回. 返回的信息被中间路由器缓存. 因有 S 服务要求, 路由器将信息发送给服务云, 服务云采用云服务, 为信息提供安全、QoS 检查等服务.

SIONA 的主要贡献如下:

(1) SIONA 定义了一种面向信息和服务的网络体系结构, 并不是一种全新的网络体系结构, 而是 IP 网络的扩展.

(2) 采用混合方式定义信息名字, 形式如 $P:N:S$.

首先是初始化过程. 服务器采用 Register 分组向路由器注册信息, 路由器将注册信息存储在注册 RT (Registration Table) 中, 注册表中记录了信息名字、所在服务器的 IP 地址及传输延时等参数. 同时路由器向全网广播信息.

其中 P 是位置标识, 采用信息所属主机的 IP 地址表示, 是信息名字的层次化部分. P 是信息名字的前缀, 可实现聚合. N 是信息标识, 采用信息内容的 Hash 值表示, 具有全局唯一性, 是信息名字的扁平化部分. N 具有自我验证性. S 是服务要求.

(3) 采用多源组播树分发数据, 支持数据的分块传输.

(4) 采用接收端驱动、路由器反馈的拥塞控制模型, 保证了每流传输的高效性、公平性和稳定性.

(5) 支持对数据的服务要求, 如要求数据的安全性、QoS、有效性等等.

对比于传统的 IP 网络, SIONA 具有高效性、高安全性和交互性更好的优势. SIONA 采用了多源组播路由算法和中间缓存技术, 保证了传输效率高于 IP 网络. SIONA 安全构建在信息上而不是容器上, 其安全性高于 IP 网络. SIONA 面向信息和业务, 更符合目前网络需求, 其交互性更强.

对比于已有的 ICN 网络,SIONA 具有高效性、高可扩展性和易部署性的优势. SIONA 采用多源组播路由算法,对比于目前 ICN 中的单源组播路由算法,SIONA 传输效率更高. SIONA 是 IP 网络的扩展,信息名字采用 IP 地址作为前缀,可采用 IP 地址的聚合方式,因此,可扩展能力与目前的 IP 网络相同. 而已有的 ICN 网络中,无论是扁平化名字还是层次化名字,其聚合能力都不如 IP 网络,因此,SIONA 可扩展性更好. SIONA 是 IP 网络的扩展,在路由器处部署 SIONA 时,只需要做一些增量部署即可. 而如 CCN 等 ICN 网络是一种全新的面向信息的网络体系结构,与 IP 网络没有必然联系,推翻现有 IP 网络的部署,构建 CCN 等 ICN 网络,并不是一朝一夕能够完成的. 因此,SIONA 在部署性方面有明显的优势.

4 SIONA 设计

4.1 命名及分组

4.1.1 数据命名

在 SIONA 中,我们提出了一种新的命名结构,如图 2 所示. 其中 $P:N:S$ 是信息名字部分, O,L 是分组中的属性字段.

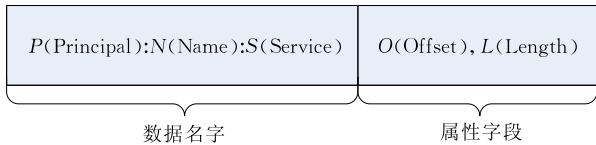


图 2 信息命名结构

$P:N:S$ 是一种混合式命名. 其中 P 代表数据所属实体的名字,可采用多种形式,如 IP 地址、某种位置标识等等,本文采用 IP 地址表示. P 是信息名字的前缀,具有多样化和层次化的特性,可实现 IP 地址的聚合能力. N 是信息内容的 Hash 值,由 P 来管理. N 是信息名字的扁平化部分,具有持久性、全球唯一性和自我验证性的特点. 返回的信息采用 (data, key) 的形式表示. 需求者可根据 data 和 key 的 Hash 值,判断是否与 N 一致. 若一致,可确定 data 是正确的信息,没有被攻击者篡改. S 字段是服务字段,主要定义了服务属性,如数据的安全性、QoS 等,定义 S 字段主要提供服务需求.

在属性字段中,定义了 O (Offset)、 L (Length) 两个字段. O 字段是偏移量字段,定义了与文件第一个字节的偏移量,以字节为单位,如文件的第一个字

节其偏移量是 0. 定义该字段保证了数据可以分布式存储,并且协助实现信息多点分发的路由机制. L 字段是长度字段,表示数据的长度, L 字段与 O 字段共同定义了文件中的一块数据.

新的信息命名结构符合了网络的信息访问需求. 不仅保证了唯一性、兼容性,而且可以采用 IP 网络的聚合方法聚合 ICN 网络,直接采用现有 IP 网络的可扩展性方面的研究成果. 同时,新的命名结构保证了大容量数据的分布式存储,提高了存储空间的利用率和路由的命中率. 并支持多点分发路由机制,提高了传输效率.

4.1.2 数据分组

SIONA 中定义了三类分组类型,即 Register、Interest 和 Data 分组,三类分组格式如图 3 所示.

O (16 bit)		L (16 bit)	
PF (2 bit)	S (6 bit)	签名信息(24 bit)	
签名(160 bit)			
P(128 bit)			
N(160 bit)			

Register 分组

O (16 bit)		L (16 bit)	
PF (2 bit)	S (6 bit)	选择器(12 bit)	随机数(12 bit)
P(128 bit)			
N(160 bit)			

Interest 分组

O (16 bit)		L (16 bit)	
PF (2 bit)	S (6 bit)	签名信息(24 bit)	
签名(160 bit)			
P(128 bit)			
N(160 bit)			
数据			

Data 分组

图 3 三类分组格式

在三类分组中,定义了分组标识 (Package Flag, 简称为 PF) 字段. 路由器通过 PF 字段区分不同的分组类型,在这里,定义 00、01、10、11 分别为 Register、Interest、Data(无 S 需求)、Data(有 S 需求) 分组.

Register 是注册分组. 若主机拥有信息,则可通过 Register 分组完成网络注册. 网络注册主要通过洪泛 Register 分组完成,注册完成后,网络上的路由器可识别不同的信息,确定信息源. 在 Register

分组中,签名保证了注册服务器的合法性, P 、 N 、 S 、 O 、 L 确定了信息 $P:N:S$ 中的从 O 开始 L 长度的字节。

Interest 是信息请求分组.若用户有需求时,通过 Interest 分组向网络请求信息.在 Interest 分组中,选择器字段加入限制条件如优先级、范围等.Nonce 是随机数,该字段主要避免出现 Interest 分组闭环现象,通过 Nonce 区分 Interest 分组是否相同,若相同则抛弃。

Data 分组是 Interest 分组的回应分组.在 Data 分组中, P 、 N 、 S 、 O 、 L 字段与 Interest 分组相同.签名字段保证了信息源的合法性、信息的完整性等. Data 分组按照原始路径返回,若 Interest 分组没有闭环问题,Data 分组也不存在闭环问题。

4.2 路由转发

4.2.1 具有缓存的多源组播树

CCN 提出了基于数据名字路由的机制.在路由转发中,CCN 选择最优性能的服务器传输数据,不支持多源到多目的的传输方式.实际上,在网络中存在多个符合需求的服务器,采用多个服务器同时传输性能更优、效率更高.所以在 SIONA 中,我们扩展了 CCN 的路由转发技术,采用具有缓存的多源组播树模型传输数据。

图 4 示意了 SIONA 路由转发模型——具有缓存的多源组播树.当 $client_1$ 发送 Interest($P:N:S$) 数据时,中间节点获取到 $server_1$ 、 $server_2$ 和 $server_3$ 具有 $P:N:S$ 数据,则沿着路径转发请求到 3 个服务器,每个服务器承担一部分数据的传输.在 Data($P:N:S$) 返回时,路径上的中间节点缓存 $P:N:S$ 数据.当 $client_2$ 发起 Interest($P:N:S$) 时,Interest 经过中间节点 N 到达了 I ,在缓存中获取到 $P:N:S$ 数据,则将数据返回给 $client_2$,并在节点 N 中缓存 $P:N$ 数据.图中用黑色节点代表缓存了数据 $P:N:S$ 。

SIONA 的路由转发模型采用了多点同时传输的方式和缓存机制大大提高了网络传输效率.多点同时传输效率一般高于单点传输效率,如果参数设置合理,效率可以提高几十倍.同时,多点同时参与传输提高了可靠性.缓存机制保证了客户端可以获取最近的副本,不需到服务器处获取数据,缩短了传输时间,减少了网络负担,提高了网络效率.在两种机制共同作用下,SIONA 路由转发模型具有高效性、可靠性,可获取更好的路由转发服务。

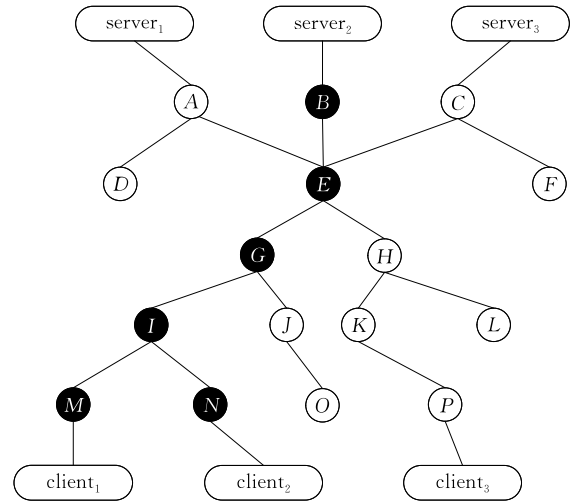


图 4 具有缓存的多源组播树

4.2.2 路由转发引擎

(1) 四大存储结构

在路由转发引擎中,存在四类数据结构,即 DS (Data Storage)、PIT (Pending Interest Table)、RT (Registration Table) 和 FIB. 在转发引擎中,Interest 的转发顺序是 DS、PIT、RT、FIB. DS 是数据存储结构,对数据进行缓存,当 Data($P:N:S$) 经过节点时,根据不同的存储策略存储 $P:N:S$ 的数据.如流行数据存储生命长,而即时数据可不存储. DS 采用分布式存储模式,可选择多个节点存储一个数据文件,保证 DS 的高利用率,缓存 DS 保证了某些数据传输的高效性. PIT 记录 Interest 分组的轨迹,为 Data 分组返回提供路径. PIT 表与 CCN 中的 PIT 表功能相同.当请求分组经过中间路由器时,PIT 记录该分组的输出接口、需求的数据名字等信息,为 Data 分组返回需求端提供依据. RT 是 Register 分组洪泛过程中,记录在节点中的存储结构,RT 主要记录数据和位置的对应关系.当 PIT 中没有 Interest 轨迹信息时,查看 RT 获取位置信息. RT 为路由器获取下载源信息、分解 Interest 分组、选取最优下载路径提供依据. FIB 表的功能与 IP 网络中的 FIB 表基本相同,前缀匹配,按照接口进行转发。

(2) 路由器处理流程

在 SIONA 中,路由器根据分组不同处理的流程不同.图 6 显示了路由器处理流程.分组到达路由器后,首先判断分组标识,按照标识码 00、01、10、11 分别处理 Register 分组、Interest 分组和 Data 无 S 请求和有 S 请求分组。

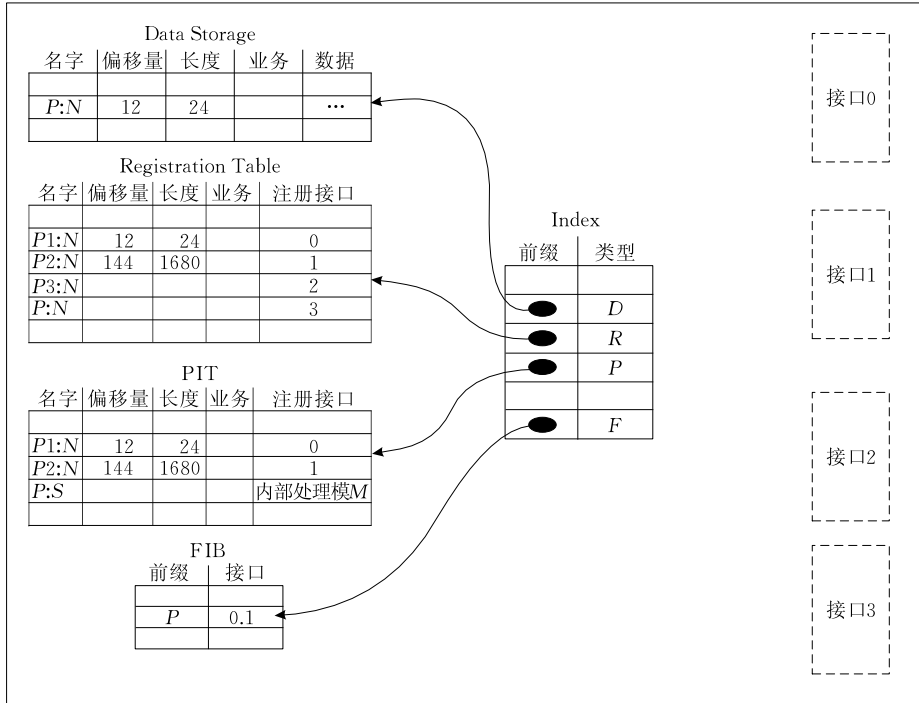


图 5 路由转发引擎模型

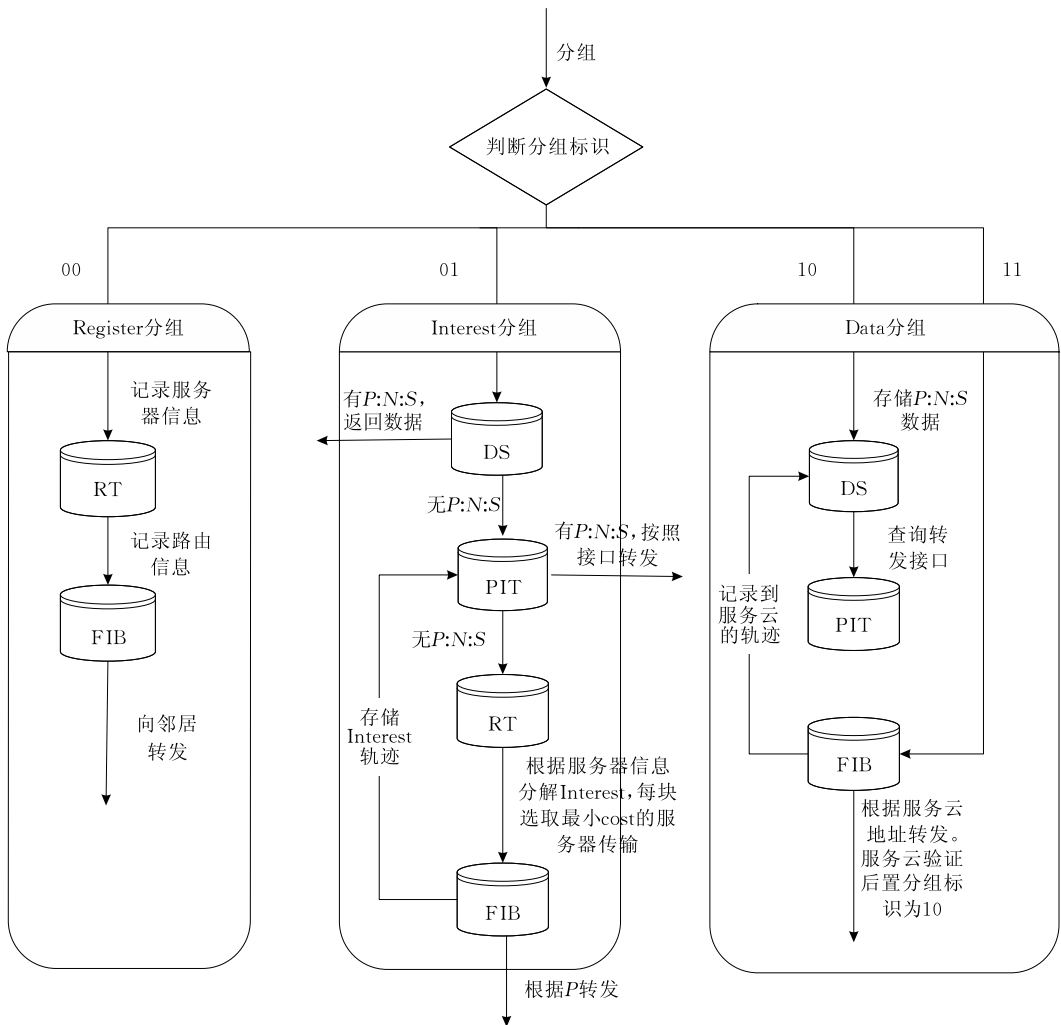


图 6 路由器处理流

① Register 分组

Register 分组向网络中注册信息,其主要路由模式是洪泛,一般应用于网络初始化阶段.对于路由器而言,主要是通过洪泛的信息获取路由信息,并将服务器信息和路由信息分别记录在 RT 和 FIB 中.

② Interest 分组

当路由器收到 Interest 分组时,首先查询 DS.若 DS 中有 $P:N:S$ 数据时,路由器直接向需求端返回数据.若 DS 中没有数据,则查询 PIT.若 PIT 中有相应的记录,表示已经有相同的数据请求,记录中所对应的接口是路由到 $P:N:S$ 信息的接口,路由器按照该接口转发 Interest 分组.若没有记录,则查询 RT. RT 中存储了各种信息的名字、对应的注册接口、存储的信息块、服务器性能等.查询 RT 后,可以获得拥有 $P:N:S$ 信息的服务器情况,根据服务器情况确定下载策略.在 SIONA 中,采用了多源分块路由算法,按照数据源的数量、性能、信道带宽等分配下载资源,保证需求端以最优的性能、最小的代价获取信息.确定好最优的下载策略后,按照 FIB 表转发信息,同时在 PIT 中存储 Interest 的转发轨迹.

③ Data 分组

对 Data 分组的处理根据 S 分为两种情况.当没有 S 要求时,路由器只需要将 Data 分组返回给需求端,并在 DS 中缓存数据.若有 S 请求,路由器则需要将 Data 分组发送给服务云,由服务云负责处理 S 请求.同时,为了 Data 分组能够返回路由器, PIT 表中记录向服务云的发送路径.服务云验证后,将分组标识置成 10,并按照无 S 请求的 Data 分组处理.

4.3 传输控制

SIONA 的传输与 IP 网络的传输有本质的区别.首先,SIONA 的传输并不局限于一条信道,而是采用多信道传输信息的不同部分.由于 SIONA 采用多信道传输,所以 SIONA 应采用多窗口机制,每个信道采用一个滑动窗口,从而控制不同信道的传输,避免拥塞.其次,IP 网络作为“推”的传输模式,其网络传输速率受到服务器控制,即网络传输控制由服务器端决定;而 SIONA 的传输采用“拉”模式,客户端作为主动端请求数据,内容服务器作为被动端传送所需求的数据,服务器端根据需求端的请求速率回送数据,即一个需求分组回应一个数据分组.所以 SIONA 的传输控制由需求端(即接收端)确定.最后,SIONA 中间路由器具有信息缓存器,用以缓存经过路由器的信息,可作为信息源传输信息,从而保证传输速率.但是,缓存的信息随着时间在不停改变,具有不稳定性,会给信息的传输控制增加不确

定性,是 SIONA 传输控制需要重点解决的问题.

既然 SIONA 的传输不同于 IP 网络传输,那么,IP 网络的拥塞控制机制不能直接运用于 SIONA 中.本文中我们提出一种接收端驱动的拥塞控制机制,即保证了 SIONA 传输的高效性,也保证了公平性.

图 7 示意了 SIONA 中的拥塞控制机制.

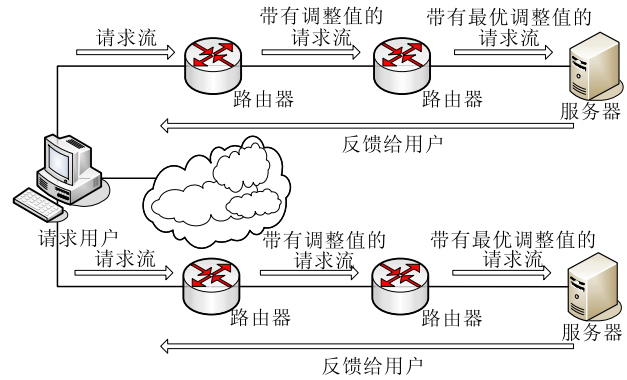


图 7 接收端驱动路由器反馈的拥塞控制机制

由于在传输过程中服务器只是被动的发送数据,对客户端的请求并不知晓,拥塞控制只能由接收端一方完成.在 SIONA 中,由接收端来驱动完成拥塞控制机制,接收端定义滑动窗口 $CWND$,每一条传输信道拥有一个窗口 ID,即 $CWND_i$,则 $CWND$ 是各个信道 $CWND_i$ 之和.

在每个信道中,采用路由器反馈机制确定滑动窗口的大小,从而控制 Interest 分组的发送间隔.当 Data 分组经过路由器时,路由器根据剩余带宽、队列情况、平均 RTT 延时计算每流的最大利用带宽,并将计算结果放置在 Data 分组中的反馈字段中,路径中的每个路由器都要进行计算,并累积叠加到反馈字段中,最后反馈给接收端. Data 分组的拥塞头部定义如图 8. 在反馈机制中,采用效率控制模块和公平控制模块保证了拥塞控制机制具有高效性和公平性.为了保证两大模块发挥最优性能,SIONA 分离了效率控制模块和公平控制模块.效率控制模块采用 MIMD,公平控制模块采用 AIMD,效率控制模块针对聚合流量和队列,尽最大能力使用完队列;而公

H_{cwnd}
H_{cwnd_i}
$Channel_i$
$H_{feedback}$
Type (sent channel or received channel)

图 8 Data 分组的拥塞头部

平控制模块针对每流, 每流根据比例分配剩余带宽.

由于 SIONA 系统基于数据传输, 更易发生拥塞. 除了采用反馈机制, SIONA 设置了可靠性机制保证丢失的 Data 分组重新传输, 即设置超时机. 若在超时时间内, 相同的 Interest 分组到达路由器会被抛弃, 而在超时时间外, 相同的 Interest 分组可以重新转发, 重新请求数据.

采用接收端驱动路由器反馈的拥塞控制机制可解决的问题如下:

- (1) 适用于“拉”模式的 ICN 网络.
- (2) 解决了多信道的传输模式.
- (3) 区分了错误重传、拥塞重传等问题, 不单纯依靠超时笼统地解决重传问题.
- (4) 反馈机制解决了 Cache 中存在不稳定信息来源问题.
- (5) 保证了公平性和效率性.

5 SIONA 实现

在 SIONA 中, 主要实现了多源分块路由算法和路由器反馈拥塞控制算法.

5.1 多源分块路由算法

多源分块路由算法主要应用于路由器分解 Interest 分组的过程中. 多源分块算法复杂, 本文中, 我们只给出框架, 优化算法是下一步研究的重点. 算法如算法 1 所示, 根据下载源数目等分数据, 每块数据选取最小代价值(与最短距离、服务器负载和链路带宽等有关)的下载源下载, 被分解的 Interest 分组并行请求每块数据, 获取数据后并行返回给客户端, 客户端组合各块数据.

算法 1. 多源分块路由算法.

```

N ← 请求的数据名字
m ← 具有 N 的服务器数目
n ← 数据 N 的大小(字节)
S ← 具有数据 N 的服务器集
对于确定的 N
S = LookupRT(N)
m = Sizeof(S), n = Sizeof(N), i = 0
FOR i ≤ n,
  FOR s ∈ S,
    cost(s) // 是 S 中开销代价最小的
    Forward(Interest(N, i, n/m), s)
    // Interest(N, i, n/m) 请求名字为 N, 从 i 开始 n/m
    字节长度的信息块, 转发 Interest 给 s
RETURN i = i + n/m

```

5.2 路由器反馈拥塞控制算法

路由器反馈拥塞控制算法的核心在于计算反馈

值. 路由器设置效率控制模块和公平控制模块, 并且采用不同的算法处理两大模块, 即 MIMD(乘法增加乘法减少)和 AIMD(加法增加乘法减少). 效率控制模块保证 SIONA 拥塞控制机制具有高效性, 计算时充分考虑剩余带宽及队列数, 保证最大限度利用剩余资源, 提高网络利用率. 为了保证稳定性, 计算时应考虑一些系统参数. 公平控制模块具体到每一个流, 若反馈值为正值, 则给每个流平均分配反馈值, 否则, 按照每流吞吐量成比例的减少反馈值. 算法如算法 2 所示.

算法 2. 路由器反馈拥塞控制算法.

```

CWND ← 接受端总滑动窗口大小
CWNDi ← 第 i 信道的滑动窗口大小
H-feedbacki ← 第 i 信道的反馈值
davg ← 平均延时
Spare ← 剩余带宽
Q ← 队列长度
D ← 聚合反馈值
Δd ← 每流反馈值增量
m ← 信道上输入的流数
n ← 传输信道数
s ← 一个数据段大小
对于一个流
FOR i ≤ n // 对于一个信道 i
  Di = α × davgi × Sparei - β × Qi
  IF Di > 0, Δdi = Di / m
  ELSE IF Di < 0, Δdi = h × Di // h 为该流吞吐量的比例
  H-feedbacki = H-feedbacki + Δdi
  CWNDi = max(CWNDi + H-feedbacki, s)
  CWND = CWND + CWNDi
RETURN i = i + 1

```

6 模拟与评价

6.1 多源分块路由算法

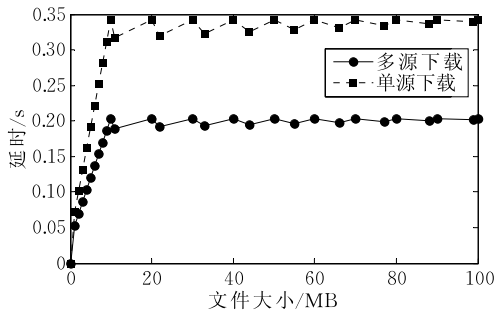
6.1.1 网络设置

为了验证 SIONA 的高效性, 本文采用了两种方式模拟多源分块路由算法. 第 1 种方式, 利用 brite 构建拓扑图, 节点为 100 个, 拓扑模型为 Waxman. 随机定义了 15 个节点作为客户端, 10 个节点作为服务器端. 每个服务器中拥有 100 个相同的文件, 文件大小从 1 MB 到 10 MB 不等, 服务器的处理性能相同. 每条路径延时为 1 ms, 带宽为 100 Mb/s. 客户端向服务器端连续不断的请求数据, 请求时间间隔为 1 s. 几个客户端可以同时请求相同的文件, 选择服务器时, 通过计算路径长度来确定代价, 本文不考虑服务器处理文件的延时. 第 2 种方式, 在真实环境中, 采用多进程机制模拟多源分块路由算法, 根据服务器数均分数

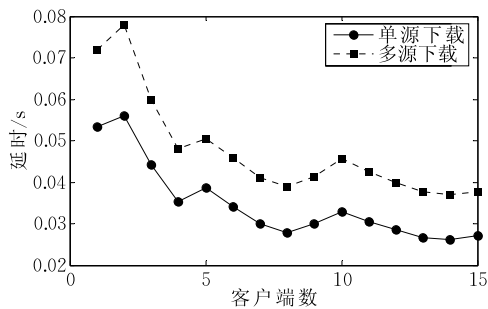
据,进程数等于服务器数,每一个进程向一个服务器获取对应部分的数据,下载完各自的部分后,合并下载文件.模拟程序采用 VC,基于 Windows 平台.

6.1.2 模拟结果

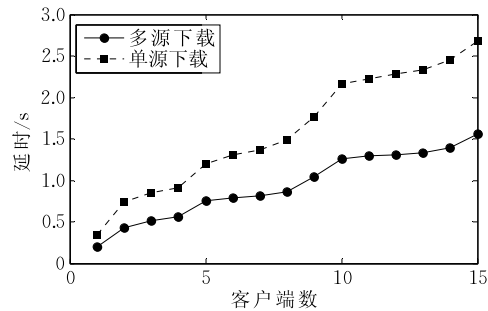
图 9(a)、(b)、(c)显示了第 1 种模拟方式的结果.



(a) 一个客户端请求不同文件数的延时



(b) 不同客户端数请求一个相同文件的延时



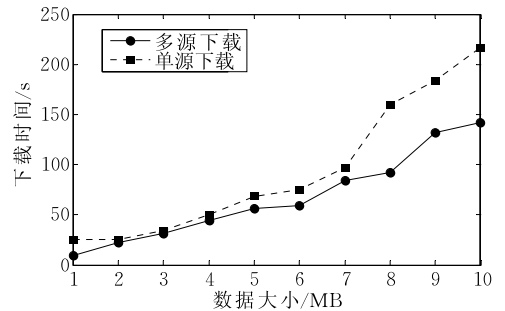
(c) 不同客户端数同时请求100个相同的文件的延时

图 9 第 1 种模拟方式的结果

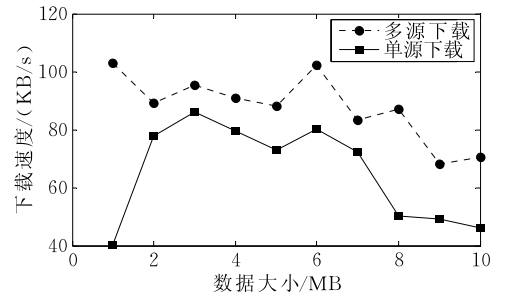
图 9 显示了不同场景下的多源下载和单源下载的延时比较.图 9(a)说明了对于一个客户端而言,无论请求文件数的多少,多源下载的延时性能优于单源下载.图 9(b)说明了多源下载在并发请求方面的性能优于单源下载.图 9(c)则考虑了多个客户端并发请求大量文件的场景,多源下载的性能优于单源下载,通过观察两条曲线的斜率可见,并发请求的客户端数越多,请求的文件数越多,多源下载的性能明显更加好于单源下载.

图 10 显示了第 2 种方式的模拟结果.图 10(a)模拟了不同大小(1 MB 到 10 MB)的数据采用单源下载和多源下载所获得的下载时间.单源下载时间明显高于多源下载时间.而且,随着数据大小的增

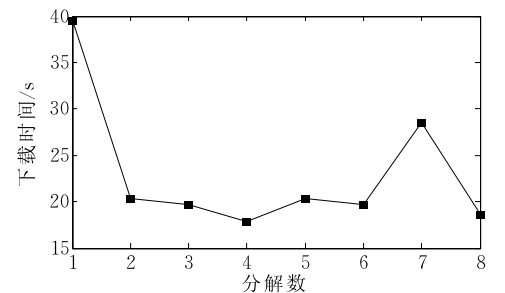
加,单源下载时间明显增大,可扩展性不好,而多源下载可扩展性较好.图 10(b)展示了不同大小的数据采用单源下载和多源下载所获得的下载速度.多源下载速度明显高于单源下载速度.图 10(c)模拟了相同数据采用不同多源分块路由算法的下载时间.从图中可以看出采用不同的算法获取的下载时间不同.其中,当采用 4 个服务器下载时,下载时间最小,性能最优,也就是说,在不同的数据下载任务中有一个最优的多源分块路由算法,保证性能达到最优.研究最优算法是下一步工作的重点.



(a) 下载不同大小文件的下载时间



(b) 下载不同大小文件的下载速率



(c) 采用不同分块算法的下载时间

图 10 第 2 种模拟方式的结果

6.2 路由器反馈的拥塞控制算法

6.2.1 网络设置

为了评估路由器反馈拥塞控制算法,本文在 NS2 中修改了 XCP 协议,增加了总滑动窗口和各个信道滑动窗口的计算,拥塞控制窗口的计算由接收端完成,并由接收端按照拥塞控制窗口发送请求数据,从而控制发送端的传输速率.

评估拓扑如图 11 所示. C 是请求端, S_1 、 S_2 、 S_3 是服务器端,其它是路由器.其中, R_0 、 R_1 之间的路

径为瓶颈路径。 C 向网络中请求文件, S_1 、 S_2 、 S_3 同时具有文件,则按照相差 10 s 的时间回应相等的数据流,即 S_1 在 0 s 时回应 1/3 大小的数据, S_2 在 10 s 时回应 1/3 大小的数据, S_3 在 20 s 时回应 1/3 大小的数据,模拟中采用了 FTP 流。在 C 中设有总发送窗口和 3 个各信道发送窗口。整个模拟持续了 30 s。表 1 列出了模拟参数。模拟中没有考虑节点中的 Cache。

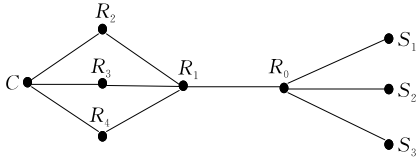


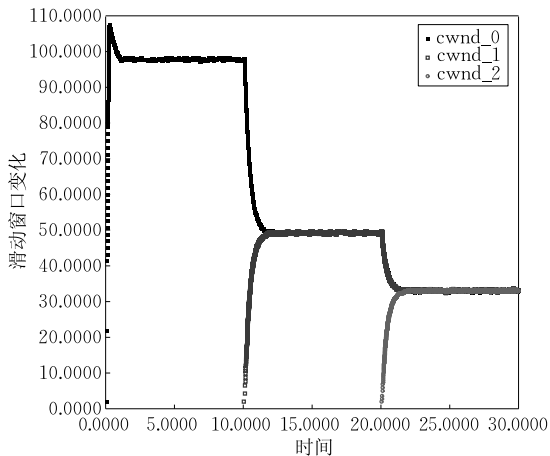
图 11 模拟拓扑图

表 1 网络模拟参数

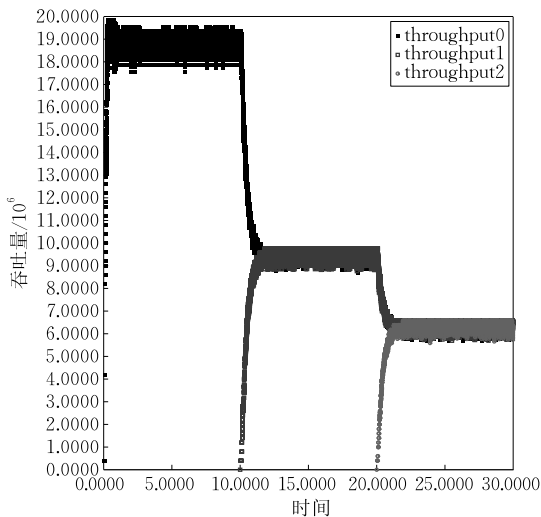
带宽	链路延时	队列大小	模拟时间	分组大小	分段大小
10 MB/s	10 ms	100 KB	30 s	1 KB	1 KB

6.2.2 模拟结果

图 12(a)、(b)显示了模拟结果。



(a) 30 s 内各个分滑动窗口的变化情况



(b) 吞吐量的变化情况

图 12 拥塞控制算法模拟结果

图 12(a)图显示了各个信道滑动窗口的变化情况。从图中可以看出,0 s 到 10 s 中,由于拓扑中只有一个流,所以该流很快占据了最大的窗口,充分利用网络资源。而到了 10 s 时,由于第 2 个流的加入,第 1 个流的滑动窗口迅速减小,第 2 个流的窗口迅速增加,大约 1 s 后,达到新的平衡,每个流占有一半的窗口,并一直稳定的保持在这个窗口数,稳定性较好。直到 20 s 时,新的流加入后破坏了这种平衡,但是 1 s 后又稳定下来,并各占 1/3 窗口数。图 12(b)显示了模拟时间内网络中吞吐量的变化。结果与图 12(a)很相似。在图 12(b)中,每个流公平的均分带宽,并且每个流吞吐量之和基本等于带宽,可见路由器反馈的拥塞控制算法兼具高效性和公平性。通过图 12(a)、(b)可见,我们提出的拥塞控制算法不仅具有高效性,而且公平性和稳定性都很好。

7 结论

互联网需求由资源共享演进为数据访问,SIONA 正是在这种背景下应运而生的,SIONA 是以信息和服务为中心的网络体系结构,采用了具有缓存的多源组播树的路由转发模型,具备了高效性、可扩展性和安全性。本文主要阐述了 SIONA 的命名框架、分组类型、路由转发模型和拥塞控制模型,并通过模拟验证了多源分块路由算法的高效性和路由器反馈的拥塞控制算法的高效性、稳定性和公平性。在 SIONA 的研究中,仍存在诸多问题,如多源分块路由算法的优化、拥塞控制算法的优化、故障恢复、模拟应考虑底层的传输机制等等,将作为下一步的研究重点。

参考文献

- [1] Gantz J F, et al. IDC-The expanding digital universe: A forecast of worldwide information growth through 2010. MA: IDC, An IDC White Paper, 2007
- [2] Davison B D. A web caching primer. IEEE Internet Computing, 2001, 5(4): 38-45
- [3] Rowstron A, Druschel P. Pastry: Scalable, distributed object location and routing for largescale peer-to-peer systems //Proceedings of the 18th IFIP/ACM International Conference on Distributed Systems Platforms (Middleware). Heidelberg, Germany, 2001: 329-350
- [4] Krishnamurthy B, Wills C, Zhang Yin. On the use and performance of content distribution networks//Proceedings of the 1st ACM SIGCOMM Workshop on Internet Measurement (IMW'01). New York, USA, 2001: 169-182
- [5] Jacobson V, Smetters D K, Thornton J D, et al. Networking named content//Proceedings of the 5th International

Conference on Emerging Networking Experiments and Technologies (CoNEXT'09). New York, USA, 2009; 1-12

- [6] Koponen T, Chawla M, Chun Byung-Gon, et al. A data-oriented (and beyond) network architecture//Proceedings of the 2007 Conference on Applications, Technologies, Architectures, and Protocols for Computer Communications (SIGCOMM'07). New York, USA, 2007; 181-192
- [7] Ahlgren B, D' Ambrosio M, Dannewitz C, et al. Design considerations for a network of information//Proceedings of the 2008 ACM CoNEXT Conference (CoNEXT'08). New York, USA, 2008; 66;1-66;6
- [8] Ohlman B. First NetInf architecture description. Europe: 4WARD, Technical Report: fp7-ict-2007-1-216041-4ward/D6.1, 2009
- [9] Adjie-Winoto W, Schwartz E, Balakrishnan H, Lilley J. The design and implementation of an intentional naming system //Proceedings of the 17th ACM Symposium on Operating Systems Principles (SOSP'99). New York, USA, 1999; 186-201
- [10] Gritter M, Cheriton D R. An architecture for content routing

support in the Internet//Proceedings of the 3rd Conference on USENIX Symposium on Internet Technologies and Systems-Volume 3 (USITS'01). Berkeley, USA, 2001, 1; 4

- [11] Ghodsi A, Koponen T, Rajahalme J, et al. Naming in content-oriented architectures//Proceedings of the ACM SIGCOMM Workshop on Information-Centric Networking (ICN'11). New York, USA, 2011; 1-6
- [12] Seskar I, Nagaraja K, Nelson S, Raychaudhuri D. Mobility-First future Internet architecture project//Proceedings of the 7th Asian Internet Engineering Conference (AINTEC'11). New York, USA, 2011; 1-3
- [13] Baid A, Raychaudhuri D. Wireless access considerations for the MobilityFirst future Internet architecture//Proceedings of the IEEE Sarnoff Symposium 2012. Newark, USA, 2012; 1-5
- [14] Baid A, Vu Tam, Raychaudhuri D. Comparing alternative approaches for networking of named objects in the future Internet//Proceedings of the Computer Communications Workshops (INFOCOM WKSHPS). Orlando, USA, 2012; 298-303



XIA Chun-Mei, born in 1977, Ph.D. candidate. Her research interests include network protocol and network architecture.

XU Ming-Wei, born in 1971, Ph. D. , professor. His research interests include future Internet, Internet routing and network virtualization.

Background

Network requirements evolved from data communications between two hosts to the access to a large amount of information through networks, which gives a new challenge to the current address-centric network architecture. Under the requirements based on accessing a large amount of information, the users show concern for the contents, not for the hosts' addresses. The current address-centric network has no high efficiency and produces a lot of redundant data due to unsuitable for the requirement. Given this fact, Information-Centric Networking (ICN) becomes the hot spot to meet the new requirements. The existing ICN schemes, such as CCN and DONA, generally had not good scalability and deployment. This paper advances a new service and information oriented network architecture—SIONA. SIONA ensures not only high efficiency, but also good scalability and deployment. There are mainly two ways to name the current ICN schemes. One is the hierarchical way of naming, such as CCN whose FIB aggregation is 4G at the best. The other is the flat way, such as DONA and INS, whose FIB aggregation is worse than the hierarchical one. SIONA was named by the IP address plus the hash of the contents of the information. The IP address becomes the prefix of the information name.

This aggregates the information names into the IP prefix

eventually so that it ensures ICN to have the same scalability as that of IP networks. In IP networks, IP aggregation can achieve 0.45% ~ 2.88% residual ratio at the best which is better than the FIB aggregation in CCN. ICN is hard to deploy, because ICN changes the communication between the hosts and adopts the communication from the host to the network, and the routing label changes from the IP address to the information name. However, because SIONA extends the current IP networks, SIONA has easier deployment than the current ICN schemes.

The research is a part of the National High Technology Research and Development Program (863 Program) of China: The Integration of Three Networks Oriented Novel Network Architecture (2011AA01A101). The project integrates not only the current telecommunication networks, cable TV network and Internet, but also all the network resources effectively. It also further improves the network coverage, significantly decreases the power consumption and realizes the fast, convenient, and green broadband information service. What's more, it improves the overall level of the information industry. The research mainly solves the one of the contents in the project; the integration of three networks oriented novel network architecture.