

一种 POF 环境下的内容路由架构设计

王 格,王 雷,王振法

(中国科学技术大学 自动化系,安徽 合肥 230027)

摘 要: 本文介绍了一种在协议无感知转发(POF)环境下的内容路由系统架构,其中内容路由器具备内容中心网络(CCN)提出的流处理和缓存功能,结合 POF 控制器的信息整合和控制管理能力,针对分层树状拓扑类型网络,在网络层实现依据内容名直接路由到内容源从而获取内容.实验结果表明,该设计能有效提高内容传输效率、减轻跨网链路负荷.

关键词: 内容中心网络;内容路由架构;协议无感知转发;内容路由器

中图分类号: TP393.02

文献标识码: A

文章编号: 1000-7180(2017)02-0015-05

A Design of Content-Routing System based on POF

WANG Ge, WANG Lei, WANG Zhen-fa

(Department of Automation, University of Science and Technology of China, Hefei 230027, China)

Abstract: Content Centric Networking(CCN) is a hot topic of future networking, In this paper, we propose one design of content-routing architecture, in which Content Router will be able to process packets and cache contents that CCN proposed. And with the ability of collecting information and management in POF controller, contents can be obtained by routing directly to contents sources in network layer under hierarchical tree topology. The results of experiment show that, the efficiency of content transferring can be improved and the load of inter-network links can be relieved.

Key words: content centric networking; content routing architecture; protocol oblivious protocol; content router

1 引言

为解决传统互联网的在内容分发和传输等方面的不足,近些年许多研究者提出了信息中心网络 ICN(Information-Centric Networking)^[1],与传统网络相比,ICN 的主要区别在于取消内容的名字与位置之间的硬耦合关联、以识别内容取代识别终端,其中最具有代表性地址进行路由、内容请求可以在节点上被聚合、利用全网内置的缓存空间实现内容就近响应或多源响应等理论,能够很好的解决多媒体内容尤其是流行内容的重复传输、网络拥塞等问题. CCNx^[2] 是 CCN 的一个官方原型系统,但没有物理网络设备支持,该原型采用泛洪的路由策略,内容传输效率较低下,仅适用于小规模网络. 其它基于

ICN/CCN 网络实现和应用方面的研究成果同样很少,究其原因是现有底层网络设备不支持直接的内容路由. 软件定义网络 SDN(Software Defined Networking)的出现为解决这一问题提供了一种思路,已有不少学者提出结合 SDN^[3] (Software Defined Networking)实现 ICN/CCN,如 J Suh 等提出基于 openflow 实现的 C-flow 系统^[4],以及 Hyunwoo L 等提出的 ICN-OMF 系统^[5]等,思路多是通过控制器获取内容位置,再规划数据流路径,但由于内容与位置映射的动态性、多样性,控制器需要掌握所有内容位置信息并进行路径规划,引入单点失效和性能瓶颈等可扩展性问题. 华为提出的协议无感知转发 POF(Protocol-Oblivious Forwarding)^[6] 技术突破了传统 SDN 设备只支持现有网络协议的限制,可

收稿日期: 2016-06-13; **修回日期:** 2016-07-05

基金项目: 中国科学院战略性先导科技专项(XDA06011202)

以支持上层的任意自定义协议^[7]. 借助 POF 的协议无感知特性,本文提出一种新的内容路由架构.

2 内容路由架构

2.1 内容路由架构概述

本文描述的内容路由架构如图 1 所示,基本设备组成包括 POF 控制器和内容路由器,其中 POF 控制器承担着整个系统的信息整合和控制管理功能,项目组在华为 SDN 控制器原型的基础上,添加了链路发现、拓扑管理、设备管理、数据转发等模块,实现了支持 POF 协议的控制器^[8]. 架构主要使用该 POF 控制器的链路发现和拓扑管理功能,为内容路由器的转发进程下发流表使得其处理自定义网络层数据包,通过加载链路发现、拓扑管理以及网络管理员制定的路由模块等组件为内容路由器的内容路由进程下发 FIB 表完成对数据包的路由决策.

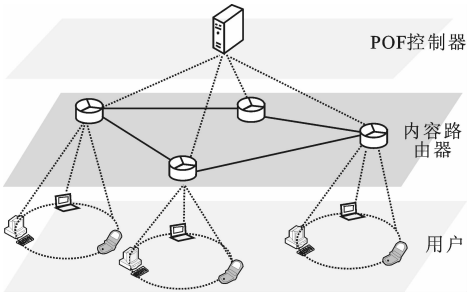


图 1 内容路由架构总图

内容路由器是该架构下的网络核心设备,多个内容路由器相互连接构成内容路由网络. 其设计是结合了华为 POF 软交换机 POFSwitch 和内容路由进程而实现,前者通过 POF 协议与控制器直接通信、接收物理端口数据流并按照流表进行处理,其中内容路由网协议数据流将被转发给内容路由进程,充分利用了 POF 的协议无感知协议转发特性;后者参照 CCN 框架中 PIT、CS、FIB 设计以及缓存空间而实现,承担网络层中路由和缓存等功能. 而用户应用通过内容路由器接入网络中完成对内容的发布、请求等.

2.2 内容路由架构中的模块设计

POF 控制器中的链路发现模块 (Discovery Manager) 负责发现内容路由网络中的链路并维护链路状态,内容路由器一启动,转发进程 (POFSwitch) 就会连接 POF 控制器的 TCP6633 号端口,该连接事件触发 POF 控制器链路侦测操作,即开始对所连接的设备下发链路层发现包 (LLDPs) 和广播包 (BBDPs) 等过程.

POF 控制器中的拓扑管理模块 (Topology Manager) 为控制器维护整个网络的拓扑信息,为网络中的设备间路由路径规划提供服务. 计算拓扑路径的算法一般是基于链路拓扑,对整个 SDN 域使用最短路径算法得出所有节点间的最短路径,但在本架构的设计中并不需要让节点路由表维护到域内所有其它节点的路由信息,对域内的任意节点 a ,其计算路由路径的集合 S 满足 $S[a]=\{a \text{ 的父节点、} a \text{ 的子节点、} a \text{ 的兄弟节点}\}$,算法如下所示. 最后 POF 控制器使用 PacketOut 消息将节点路由表信息发送给相应的节点,即构成内容路由器路由进程中 FIB 表.

```
拓扑路径算法:
Input:  $r[0:n]$  // 所有节点的节点标识
Output: ShortPathTable & D // 输出所有节点的路由表
Begin:
  for  $i = r_0:r_n\{$ 
    for  $j = r_0:r_n\{$ 
      if ( $j \in \{ \text{parent}(i), \text{sibling}(i), \text{child}(i) \}$ )
        insert  $j$  to  $S[i]$ ;
    }
    Insert Dijkstra( $S[i]$ ) to  $D[i]$ ;
  }
  return  $D$ ;
}
```

内容路由器的设计和实现目前是采用 POF 软交换机和内容路由进程互相协助配合方式实现,两者通过将数据包发送给对方的物理或者虚拟端口的方式实现进程通信. 考虑现有 SDN 理念中使交换机只能做到简单的完全/掩码匹配^[9],而且流表项的动态增删只能借助控制器完成,要实现 CCN 框架而且不扩展 SDN 技术框架,采用了多进程协作是最简便有效的办法,其模块关系图如图 2 所示. 内容路由器启动后,POF 控制器将为转发进程 (POFSwitch) 下发流表,对于 LLDP 等消息,转发进程根据流表直接处理;对于路由表消息,转发进程根据流表转发给路由进程以生成路由进程中的 FIB 表;对于其中的内容路由协议中的公告包、兴趣包和内容包,转发进程也将其转发给内容路由进程,其处理过程同 CCN 框架基本处理流程类似.

2.3 内容路由原理和协议

由于本系统中主要的设备与数据处理都依赖于内容路由器,相当于 CCN 中的节点概念,因而在本节主要介绍由内容路由器组成的节点网络模型. 本系统通过在网络层中根据内容标识与节点标识的关联性将内容映射到节点,解决海量内容对象的分布和定位问题,并基于 POF 控制器的集中控制作用设计节点网络模型,解决如何保证路由效率、控制路由

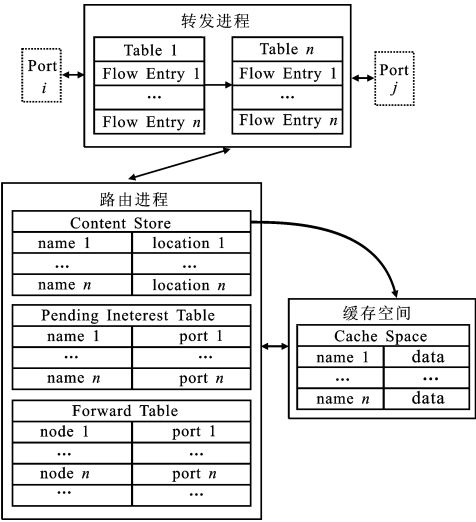


图 2 内容路由器模块构成

表规模以及在定位内容副本过程中让已缓存内容的节点就近响应等问题。

为实现内容与节点的映射,需要两者具有同一命名空间下的标识以进行匹配操作. 由于内容名是在网元设备间路由的必需信息,采用定长 hash 值这种二进制扁平化命名方式更能提高命名空间利用效率,因而本架构使用的内容标识,是使用特定哈希函数对内容相关描述计算得出的哈希值。

如果对节点同样采用一致性哈希的命名方式,由于节点标识与节点物理位置不具有关联性,节点标识难以被聚合,因而任意节点需要维护到其它所有节点的路由信息,造成节点路由表规模随着网络规模的增大而增大. 此外,实际环境中网络的拓扑总是复杂的树或图结构,用户或小型网络总是通过汇聚后连入上层网络中,这样铺设通信链路符合管理需要和信息传输模型,分层的网络结构也有利于提高网络规模扩展性. 基于上述考虑,本架构使用分层树状模型来组织网络,因为允许兄弟节点互联互通,其物理拓扑不是严格树结构. 该模型要求每个节点至多有一个父节点(顶层域节点无父节点)、个数不限的兄弟节点和子节点。

基于上述分层树状模型,节点有如下原则:(1)、每个节点拥有独一无二的节点标识,该节点标识与内容标识等长;(2)、节点标识按照网络模型的层数被划分成多段,且各段的子标识产生满足继承关系;(3)、任意节点的 FIB 表无需维护到全网其它所有节点的路由表项,而是让控制器根据该节点的实际拓扑位置和连接关系,只下发其父/子/兄弟节点的路由表项. 以图 3 的三层网络模型为例,节点 ID 被

划分成三段:{网络 ID/子网 ID/叶节点 ID},为方便说明,本例用两位十进制数来表示节点的网络号、子网号、叶节点号. 顶层域的每个节点具有顶层域中唯一网络 ID、且子网 ID 和叶节点 ID 为 0(如 R_1 的网络 ID 为 15,其余为全 0);各顶层域节点下属的子网域节点继承相应的网络 ID、具有本子网域中唯一子网 ID 并且叶节点 ID 为 0(如 R_2 、 R_5 、 R_6 的网络 ID 继承自 R_1 的 15,具有本域唯一子网 ID,叶节点 ID 为全 0);叶节点的网络 ID、子网 ID 也继承自其父节点且具有本底层域中唯一的叶节点 ID(如 R_3 、 R_4 、 R_{10} 、 R_8 网络 ID 均为 15,子网 ID 继承其相应父节点,具有本域唯一叶节点 ID). 节点 ID 的合理分配会关系到内容是否均匀分布、数据传输是否高效等问题,但在这里我们不做深入讨论——正如 IP 地址的具体分配并不是 TCP/IP 协议部分一样,只是给出了以上的一般划分原则。

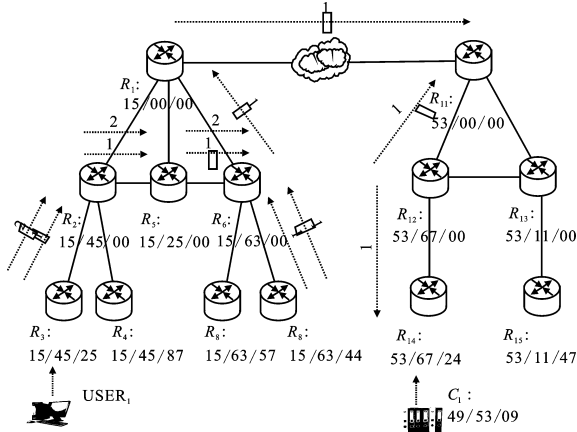


图 3 三层树状模型示例

基于上述命名规则,由于内容和节点存在于相同的名空间,内容标识与节点标识即可进行映射. 这里我们定义绝对后继节点为全网范围内内容标识的直接后继节点(节点标识不小于但最为接近内容标识),定义相对后继节点是在某一网层中,易知节点标识中继承部分长度,屏蔽节点标识和内容标识该长度部分、剩余部分比较得到的直接后继节点. 例如在图 3 中,内容 c_1 的内容标识为 49/53/09,其绝对后继节点是 R_{14} ;在包含 R_2 、 R_5 、 R_6 的子网层中,节点标识前两位标识 15 是由继承得到,计算相对后继节点是屏蔽节点标识和内容标识的前两位后、比较剩下四位得到 C_1 在该子网层的相对后继节点是 R_6 。

用户应用所发布的内容,系统根据内容标识把内容副本推送到对应的绝对后继节点中,完成内容

发布即内容副本部署操作. 用户应用对内容请求的缺省操作,是节点根据内容标识匹配 FIB 表中查找相对后继节点查询是否有缓存;若该相对后继节点的缓存空间中找不到相应内容的缓存后,将推送给上一层节点查找,直至推到最顶层域时,则为该内容查找绝对后继节点请求内容副本. 内容被传输过程中,可被缓存在各层的相对后继节点中. 上述内容路由网协议包的详细处理流程参见图 4、图 5 和图 6 所示.

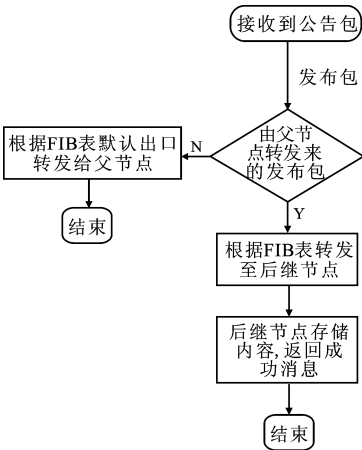


图 4 内容网公告包处理流程

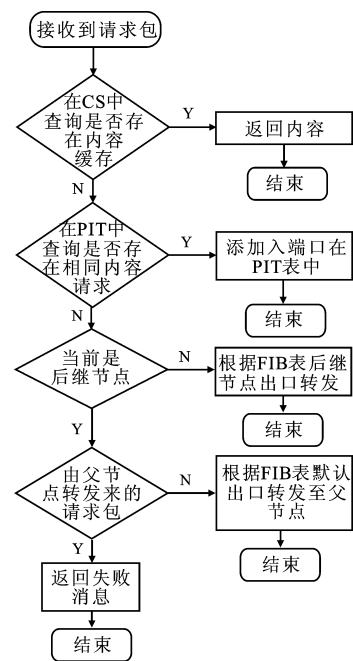


图 5 内容网请求包处理流程

CCN 框架中节点内置缓存,并提倡优先使用靠近请求者的内容缓存来响应请求,故框架采用相对/绝对后继节点的概念,提出优先进行区域性的缓存查找、保有全局性的副本路由可达的理念. 显而易见

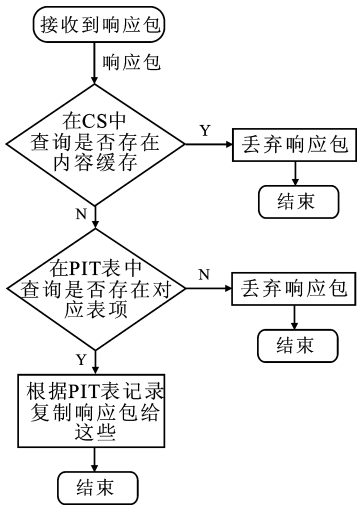


图 6 内容网响应包处理流程

的是,该架构下请求已经缓存在附近节点的内容,该路由策略可以实现就近缓存响应,可以有效减少内容响应时间、提高传输效率,即便对于非临近缓存内容的请求,会额外增加走最短路径去相对后继节点的路径,但该路由过程也不会增加太多冗余路径. 以图 3 的三层简单网络模型为例,请求内容 C_2 副本的请求过程如①箭头所示,而在首次请求 C_2 副本请求后可被缓存在 R_6 ,则接入 R_4 的用户 $User_2$ 发起对 C_2 的请求,其路径将如②箭头所示,该案例中由于省去了跨网路由路径,请求到达内容源的跳数由 9 跳以上缩减为 3 跳,路由效率明显提高.

3 性能分析与仿真

为了验证系统的性能,我们构造了一个三层结构、各层节点度分别为 10、10、5 的一个具有 610 个节点的网络拓扑,首先在网络中随机发布 200 条内容在其绝对后继节点中,赋予子网节点缓存能力(单点缓存容量与内容总大小比重 0.08),所有请求发自叶节点,响应包在返回途中在子网节点缓存(使用 LRU 缓存置换策略). 采用如下三个指标对性能进行分析:1、响应时延(Request Latency)——请求内容至数据首次到达请求者间的平均跳数;2、传输时间(Transfer Time)——数据回传给请求者的平均跳数;3、跨网请求次数(Cross-Net Times)——请求经过顶层节点至其它域的请求总次数. 最终得出在对随机内容的不同请求总数(100 用户数 * 平均每个用户请求的内容数)下,传统的 IP 式路由(只获取内容不使用缓存空间)和本框架的内容路由(为简化实验未使用 PIT 功能)的性能对比图.

图 7 显示两种路由方式在响应时延和传输时间

上的性能对比,可以看出在初始请求内容较少时,IP 式路由和本架构内容路由(以 POF-CRN 表示)在数据响应和数据传输的平均跳数上接近,这是因为请求的内容重复性不高,多数内容在缓存不足情况下都需要去内容副本源获取,而随着请求内容总数的增大,用户请求的内容重叠,部分重叠内容可在区域的节点缓存中可以获取,因而响应时延和传输时间明显减少。

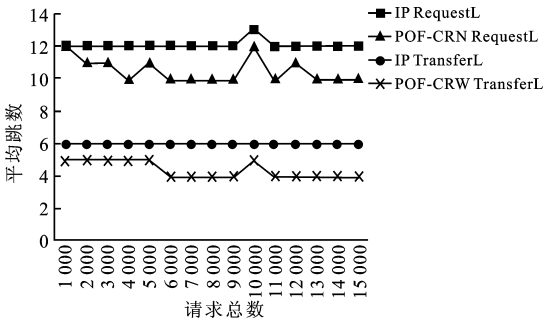


图 7 传统路由请求和本架构内容路由平均跳数对比

图 8 显示两种路由方式在跨网请求次数上的性能对比,可以看出,相比传统路由方式,使用了缓存的内容路由的跨网流量显著减少,减少比例随着请求总数增大而显著增加,从 20%增加到了 40%。由于该实验尚未使用 CCN 内置 PIT 表的请求汇聚功能,因而完整 POF-CRN 系统中平均跳数和跨网流量将进一步减少。

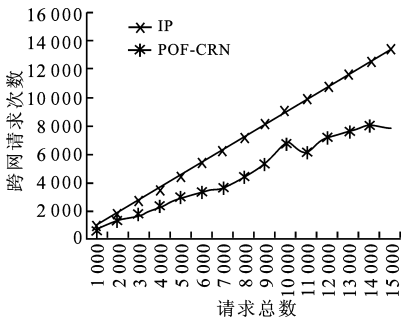


图 8 传统路由请求和 POF-CRN 内容路由跨网请求次数对比

4 结束语

本文提出一种基于内容路由系统设计架构,结合 POF 和 CCN 技术,在网络层可实现依据内容名直接路由到内容源。实验数据表明,在本架构中,对于有较高请求频率内容,传输效率有明显提高;CCN 内置的流汇聚和缓存功能,能显著降低跨网流量。但值得注意的是,缓存空间的大小和分布是一个

影响内容传输效率的因素;此外由于节点分层产生的节点标识集群效应会引发内容分布不均,我们将继续在之后的研究中对这些问题进行研究和优化。

参考文献:

[1] Ahlgren B, Dannewitz C, Imbrenda C, et al. A survey of information-centric networking[J]. Communications Magazine, IEEE, 2012, 50(7): 26-36.

[2] Moskon M, Solis I, Uzun E, et al. CCNx 1.0 protocol architecture[R/OL]. [2015-12-25]. [http://www. cc-nx.org/pubs/CCNxProtocolArch itecture. pdf](http://www.cc-nx.org/pubs/CCNxProtocolArchitecture.pdf).

[3] Software-defined networking: the new norm for network[EB/OL]. [2015-05-29]. [https://www. open-networking.org/images/stories/download/openflow/wp-sdn-newnorm. pdf](https://www.open-networking.org/images/stories/download/openflow/wp-sdn-newnorm.pdf).

[4] C-flow: content-oriented-networking over OpenFlow [EB/OL]. [2015-05-29]. [http://www. opennetsummit.org/archives/april2/site/pdf/snu. pdf](http://www.opennetsummit.org/archives/april2/site/pdf/snu.pdf).

[5] Hyunwoo L, Kim D, Suh J, et al. ICN-OMF: A control, management framework for Information-Centric Network testbed [C]// Information Networking (ICOIN), 2015 International Conference on, Cambodia, Siem Reap, IEEE, 2015: 416-417.

[6] Huawei Technologies. Protocol-oblivious forwarding [EB/OL]. [2015-05-30]. [http://www. poforwarding.org/](http://www.poforwarding.org/).

[7] Song H. Protocol-oblivious forwarding: Unleash the power of SDN through a future-proof forwarding plane [C]// Proceedings of the second ACM SIGCOMM workshop on Hot topics in software defined networking. China, HongKong, ACM, 2013: 127-132.

[8] 高翔宇,王雷,王振法,等. 支持协议无感知转发的 Floodlight 控制器实现[J]. 微电子学与计算机, 2015, 32(11): 92-96.

[9] Song Haoyu. Principle and implementation of protocol Oblivious forwarding[R/OL]. [2016-06-10]. [http://www. poforwarding.org/document/Principle_and_Implementation_of_POF. pdf](http://www.poforwarding.org/document/Principle_and_Implementation_of_POF.pdf)

作者简介:

王 格 男,(1990-),硕士研究生,研究方向为未来网络。E-mail: wangge@mail. ustc. edu. cn.

王 雷 男,(1972-),博士,副教授,副主任。研究方向为未来网络的架构与应用以及计算机仿真与优化、视频传输与控制。