ISSN 1000-9825, CODEN RUXUEW
Journal of Software, Vol.18, Supplement, December 2007, pp.88−94

E-mail: jos@iscas.ac.cn
http://www.jos.org.cn
Tel/Fax: +86-10-62562563

对等网络中的路由器增强型 NAT 方法*

张建伟[1,3+], 皮人杰[2], 战晓苏[4], 郭云飞[1]

[1](国家数字交换系统工程技术研究中心,河南 郑州 450002)

[2](北京邮电大学 计算机科学与技术学院,北京 100876)

[3](郑州轻工业学院 计算机与通信工程学院,河南 郑州 450002)

[4](北京邮电大学 电子工程学院,北京 100876)

A Router Enhanced NAT Method for Peer-to-Peer Network

ZHANG Jian-Wei[1,3+], PI Ren-Jie[2], ZHAN Xiao-Su[4], GUO Yun-Fei[1]

[1](National Digital Switch System Engineering and Technological Research Center, Zhengzhou 450002, China)

[2](Department of Computer Scicence, Beijing University of Posts and Telecommunications, Beijing 100876, China)

[3](Department of Computer and Communication, Zhengzhou University of Light Industry, Zhengzhou 450002, China)

[4](Department of Electronic Engineering, Beijing University of Posts and Telecommunications, Beijing 100876, China)

+ Corresponding author: Phn: +86-371-63556633, Fax: +86-371-63556293, E-mail: ing@zzuli.edu.cn

Zhang JW, Pi RJ, Zhan XS, Guo YF. A router enhanced NAT method for peer-to-peer network. *Journal of Software*, 2007,18(Suppl.):88−94. http://www.jos.org.cn/1000-9825/18/s88.htm

Abstract: The deployment of symmetric NAT makes the communication between hosts in a peer-to-peer application very difficult, which needs many relay node to provide NAT traversal service. Traditional relay nodes are all server hosts lie in the edge of network. To reduce the duplicate traffic and excess router load caused by server relaying, a UDP switch based router enhanced NAT method is proposed in this paper, which can resolve the bottleneck of network bandwidth by enhance router function independently.

Key words: peer-to-peer network; router enhence; network address translation; transport layer switch

摘 要: 对称NAT的部署使得对等应用主机间的通信变得困难,需要大量的中转节点提供NAT穿越服务.传统的中转节点是位于网络边缘的服务器节点,为了减少服务器中转引入的重复流量和路由器交换负载,提出了一种基于UDP交换模型的路由器增强NAT方法,可以通过增强单个路由器的功能来缓解网络带宽资源压力.

关键词: 对等网络;路由器增强;网络地址转换;传输层交换

当前 P2P 的网络流量已经超过了 Web 浏览类业务和其他各类传统 C/S 类业务流量的总和,各类 P2P 网络应用成为 Internet 的主导型应用.P2P 应用的一个显著特征是节点间的流量不再经由服务器而是直接在这个成员主机之间交互,这必然要求网络能够提供节点间的直接通信能力.但是,当前的 Internet 采用的是基于 32bit 地址空间的 IPv4 网络层技术,可用的全球公网 IP 地址趋于枯竭,大量的新增 Internet 网络用户只能采用 NAT[1]方

* Supported by the National Basic Research Program of China under Grant No.2007CB307102 (国家重点基础研究发展计划(973)); the Science-Technology Supporting Project of the National 'Eleventh Five-Year-Plan' of China under Grant No.2006BAH02A03 (国家'十一'五科技支撑计划项目); the National High-Tech Research and Development Plan of China under Grant No.2006AA01Z206 (国家高技术研究发展计划(863))

Received 2007-04-15; Accepted 2007-11-25

式接入 Internet,这种情况在我国尤为普遍.NAT 技术虽然可以解决部分的节点接入问题,但是主要是针对传统的 C/S 应用模式,即网络应用是以位于公网的服务器为核心进行开展的.而 P2P 应用引出的主机之间的通信必然受到 NAT 设备存在的影响,有鉴于此,需要针对此类问题提出相应的解决方案.

本文第 1 节概述现有的 NAT 穿越技术.第 2 节详细描述一种基于路由器增强的 NAT 穿越方法.第 3 节对应用实施该方法的各个方面加以介绍.第 4 节总结全文,并提出下一步研究内容.

1 NAT 技术概述

总结起来,目前解决 NAT 穿越的方法主要分为 3 类:NAT 受控方式、中转方式和非中转方式.

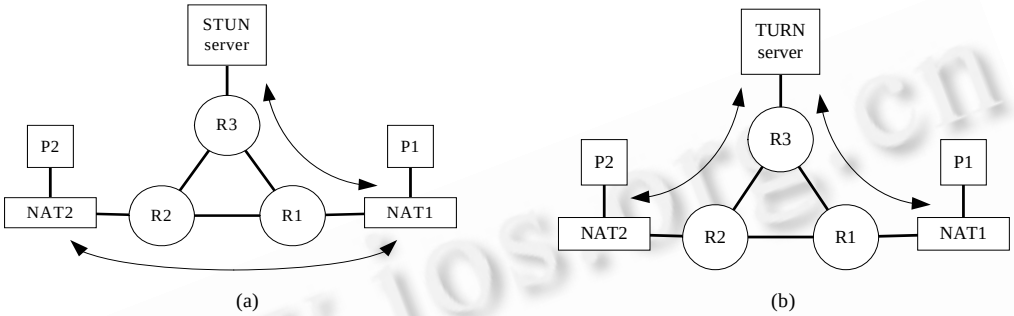

Fig.1 STUN and TURN traffic model

图 1 STUN 和 TURN 流量示意图

1.1 NAT受控方式

这是最直接的解决方案,即增强 NAT 设备的功能,使其支持 P2P 类应用.例如,RSIP[2],UPnP[3]和 ALG(application level gateway).ALG 本身受限于具体的网络应用类型,无法解决不断出现的新型 P2P 网络应用的问题,所以此类方法只能做到对主机软件的透明性,除此之外无其他优势.虽然 UPnP 作为一种通用的 NAT 设备动态应对方法已经被很多软件所采用,但是由于 NAT 设备已经存在了很长时间,很难对已经大量部署的原有非 UPnP 设备进行升级和改造,而且 UPnP 也无法解决多重 NAT 的问题.另外,更为重要的是,由于考虑网络安全的因素,很多即使本身支持 P2P 应用的 NAT 设备都人为地被网络管理员关闭了此项功能.RSIP 与 UPnP 非常类似.

1.2 非中转方式

此类方式以 STUN[4]为代表,如图 1(a)所示,即通过探测出设备的公网地址后,在后续的操作中通过通告其公网地址使得其他主机可以主动进行数据发送,从而建立 NAT 映射关系.但是这类方法只能应对各类锥型 NAT(cone NAT),对于目前网络中大量使用的对称 NAT(symmetric NAT),当通信双方主机都位于对称 NAT 内网时,非中转方式无法完成 NAT 映射关系的建立.

1.3 中转方式

由于非中转方式不能解决对称型 NAT 的问题,所以出现了像 SOCKS 代理、TURN[5]中继这样的流量中转处理方法.如图 1(b)所示,TURN 服务器在通信双方主机直接提供双向的流量中继,所以可以解决多重 NAT 和对称 NAT 问题,但其采用了 TCP 可见的协议直接中继处理,对于大规模存在的 NAT 内网主机其扩展性较差.这里所说的中转方式是基于专门服务器方式的,而文中介绍的 RUS(router enhenced UDP switch)方法则将中转节点的位置放在通信双方主机的单播传输路径上.

ICE[6]是一个建立在 STUN 和 TURN 之上的解决方案框架,用于让节点能够根据当前的情况自由选择 NAT 穿越方式,尽可能地降低网络负载,但不可避免地受到 STUN 和 TURN 的限制.

2 算法描述

目前 IPv4 网络中的 NAT 设备只支持 TCP、UDP、ICMP 这 3 种基本协议的地址/端口转换,而相对于 ICMP 需要在 raw socket 一级进行网络编程而言,TCP 和 UDP 提供了更加友好的 socket 编程接口,但是 TCP 是一种有状态的传输协议(分为 client 和 server 两种工作模式).一般的 NAT 设备缺省的动态 NAT 映射只处理 TCP 的 client 模式,即只允许内网主机发起连接,该连接的 NAT 映射具有 TCP 的状态字段,不能用于外网主机通过这个映射向内网主机发起新的连接,所以采用 TCP 方式很难解决双向连接的问题.由于 UDP 本身是无状态的,任何网络节点对 UDP 关系的维护都要比 TCP 容易得多,而且由内网主机触发的 NAT 映射可以被外网主机用来随意传送数据,所以选择采用 UDP 方式作为交换的传输层协议.

2.1 需要解决的问题

在有 NAT 设备的网络环境中,只有在通信双方都处于对称型 NAT 后面的情况下,才会导致非中转型方案(STUN)的失败,此时才必须引入中转型方法,而 RUS 是一种基于路由器的中转型方法,与其他中转型方法相比,RUS 的优势是所消耗的带宽更少,对单个路由器的增强可以避免传统中转型方法对一组路由器造成的带宽影响.

2.2 算法思想

RUS 的基本思路是让处于内网主机之间的单播路径上的路由器承担数据的中转功能,而不再需要由路径之外的服务器节点承担中转,这样就避免了本来应该由路由器转发的流量旁路到了非路径节点上.以图 2 中的 P1 和 P2 间的流量为例,如果采用由 P3 作为中转节点,流经 R3 的流量变成了 P1 和 P2 直接交互时的 2 倍(P1 与 P3 之间、P3 与 P2 之间),当网络中具有大量的 P2P 类应用,而且节点大多数处于 NAT 内网的时候,这种路由器负载增大的情况将覆盖整个网络,这样必然导致网络性能受到影响.如果此时 R3 本身能够提供 NAT 中转的功能,则节点的交换负载不会成倍增加,当然,RUS 的引入必然会增加一些额外的处理消耗,但是这种处理是纯计算型的,可以被节点通过并行处理进行分担的,并不像流量增加那样直接对路由器链路造成负载增加.

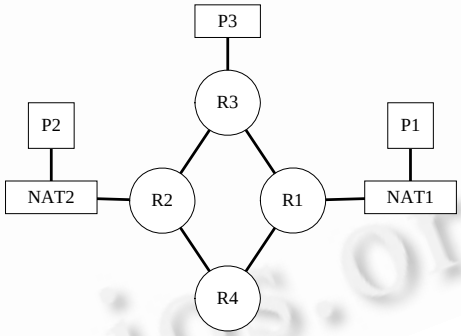

Fig.2 P2P application nodes distribution sample
图 2 P2P 应用中节点分布图示例

2.3 节点探测

RUS 中的交换路由器(即增强型路由器)的信息来源是由通信主机主动探测得出的.RUS 采用了发送具有特定 UDP 目的端口信息的探测包来触发沿路各个路由器启动 RUS 功能.每个通信主机独立地向通信对方之前通告的公网 IP 地址发送 PING 消息(UDP 包),该 UDP 包的目的端口值为 RUS 定义的保留值(在本文中假设为 56 789),沿途的每个路由器都会受到该 UDP 包,包括可能存在的 RUS 路由器,对于非 RUS 路由器而言,这是一个普通的 UDP 单播数据包,所以不会做任何额外处理,继续其单播路由进程.当路径上的某个 RUS 路由器收到该 UDP 包时,由于其目的端口为 56 789,所以对该 PING 消息响应一个 PONG 消息(也是 UDP 包),其源地址为 PING 消息中的目的地址,源端口为 56 789,这样,PONG 消息就能通过发送主机一侧的 NAT 检查回到主机.PONG 消息

中含有 PING 消息中所没有的信息:RID,RID 即该 RUS 路由器的访问 IP 地址,用于告知主机可以通过哪个 IP 地址进行 RUS 的后继操作,因为 UDP 字段中的信息并没有涉及路由器.考虑每个路由器的性能门限,即使是支持 RUS 功能的路由器也可以根据当前的运行负荷决定是否要对 PING 消息进行响应,如果不作响应,则对 PING 消息的处理与普通路由器一样,继续其单播路由进程,即传递给下一跳路由器.

对于得到 PONG 消息的主机,如果希望进行 RUS 下一步操作,则需要向所得到的 RID 地址的 56 789 端口发送 OPEN 消息,该消息中不携带任何参数,由于 RUS 主要用于解决对称 NAT 问题,所以无须提供任何主机方面的地址或端口信息,收到 OPEN 消息的 RUS 路由器如果允许此次映射关系的建立,则响应 ACK 消息,ACK 消息中携带有为此次会话分配的 UDP 端口(一对端口:一个用于探测主机发送数据,另外一个用于对方主机发送数据),还包括此次映射的会话标识 SID,此 SID 可为可变长度的任意字节串,用于防止非会话主机的流量干扰.

由于没有其他方法可以让 RUS 路由器获知自己是否处于通信双方的单播路径上,所以探测主机必须保持定时地发送 OPEN 消息,不过在收到 ACK 消息后的 OPEN 消息中携带有合法的 RID、UDP 端口、SID,拥有 RID 的路由器在检查了 UDP 端口和 SID 有效的情况下不需要发送 ACK,这种定时发送的 OPEN 消息是用于让路由器确认所处的单播路径情况没有改变.如果路由器在一定时间内没有收到合法的 OPEN 消息,则认为此次会话失效,转换关系被删除.

主动探测的主机在完成了映射关系之后需要通过非 RUS 途径将此次映射的 RID 和 SID 告知通信对方的主机,这样对方才能通过向 RID 位置发送 UDP 分组激活自己一侧的 NAT 映射关系,数据才能通过 NAT 传递给被告知一方的主机.非 RUS 途径告知必须得到确认,否则主动探测一方可能在对方 NAT 映射建立之前发送数据,导致数据丢失.

转换关系的终止可以是由探测主机停止发送 OPEN 消息触发,也可以通过探测主机主动地发送 SHUT 消息来关闭,SHUT 消息携带有与定时发送的 OPEN 消息一样的信息,收到 SHUT 消息后路由器无须响应,直接关闭该转换关系即可.

2.4 消息类型

RUS 采用 UDP 作为数据的传输方式,与普通的 NAT 设备不同,RUS 不仅依赖地址和端口来维护转换的映射关系,而是增加了一个用于判定数据是否合法的会话标识字段(SID),这是因为 RUS 交换路由器处于网络的核心位置,较处于网络边缘的 NAT 设备更容易受到非法流量的攻击,因而很容易干扰转换映射关系.所以,通过增加一个位于 UDP 数据字段内的 SID,可以避免随机的流量导致的干扰,而且每个会话的 SID 长度和内容都可以是不同的,SID 的长度以字节为单位,在 RUS 探测时由交换路由器在 ICMP 响应报文中指定.

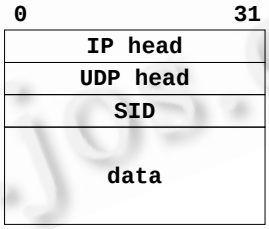

Fig.3 RUS carrier message format

图 3 RUS 中用于承载用户数据的消息格式

2.5 转发算法

当转换关系建立之后,通信双方主机就可以通过 RUS 的 UDP 数据承载进行数据转发,与节点探测阶段不同,数据转发过程中的 UDP 数据是直接建立在主机和路由器之间的,即 UDP 包的 IP 地址为 RUS 路由器的 RID,而不是节点探测阶段的对方通告 IP 地址.只是因为转发使用了路由器的 UDP 端口来区分会话,从而减少了 UDP 负荷.

```
int RusForward(Msg)
{
  if(Msg.SID is Valid)
  {
    if(Msg.DstPort == Tab[Msg.SID].SelfPort)
    {
       Tab[Msg.SID].SrcAddr := Msg.SrcAddr;
       Tab[Msg.SID].SrcPort := Msg.SrcPort;
       Msg.SrcAddr := RID;
       Msg.SrcPort := Tab[Msg.SID].PeerPort;
       Msg.DstAddr := Tab[Msg.SID].DstAddr;
       Msg.DstPort := Tab[Msg.SID].DstPort;
    } else
    if(Msg.DstPort == Tab[Msg.SID].PeerPort)
    {
       Tab[Msg.SID].DstAddr := Msg.SrcAddr;
       Tab[Msg.SID].DstPort := Msg.SrcPort;
       Msg.SrcAddr := RID;
       Msg.SrcPort := Tab[Msg.SID].SelfPort;
       Msg.DstAddr := Tab[Msg.SID].SrcAddr;
       Msg.DstPort := Tab[Msg.SID].SrcPort;
    } else
      return 0;
  } else
    return 0;
  return 1;
}
```

以上是转发算法的伪码描述,返回 1 表示转发成功,返回 0 表示失败,其中 Msg 为接收到的 UDP 数据包,首先判断所携带的 SID 是否有效,若有效,再判断 UDP 目的端口是否与该 SID 对应的转换关系表(Tab)中的 SelfPort 或 PeerPort 相等,SelfPort 和 PeerPort 为路由器为此次会话分配的 UDP 交换端口对,其中 SelfPort 接收探测主机发送的数据,PeerPort 接收对方主机发送的数据.处理中首先保存 Msg 的源地址和端口,因为很可能发送一方的 NAT 地址已经发生了变化,然后将 Msg 的源地址/端口和目的地址/端口都进行转换,使其能够穿过对方的 NAT 检查.

3 应用实施

在具体的实施中,RUS 需要处理各种不同的网络和节点状况.

3.1 并行处理

RUS 从某种程度上看就是一种(源+目的)地址转换,与普通的服务器方式的中转节点不同,其处理完全是在交换路由器内部完成的,对于资源的占用也只限于路由器内部,并不会影响外部链路带宽资源的竞争.所以,关键路径上单个路由器的功能增强可以解决传统服务器模式中原来需要增强整体网络节点容量的问题.为了不影响既有的路由业务流量,RUS 功能部件可以采用在路由器内部进行旁路处理的方式,即将入接口中的 RUS 流量旁路到额外的处理模块中,完成了地址转换后再进入正常的路由交换系统中.

3.2 MTU问题

由于 RUS 是一种建立在 UDP 之上的隧道传输机制,所以在该虚拟链路上的最大传输单元 MTU 要比普通的二层链路小至少(40+n)个字节,其中 40 个字节为 IP 头和 UDP 头字段负荷,n 为 RUS 的会话标识字段(SID)的长度.结合现有的 MTU 探测机制,RUS 也应该尽量避免 IP 分组在传输中进行分片.

3.3 非对称路由

实际的 Internet 中存在着非对称路由的情况,RUS 本身是基于主机主动探测的,所以每个主机探测得到的交换路由器只是该主机作为源节点时的单向中转节点,很可能该路由器并不在相反流量的单播路径上.如果通信双方主机都能探测出各自的交换路由器,则两个方向的流量各自进行交换.如果只有一方的探测得到了结果,此

时流量只能在一个方向进行最优化传输.基于这种情况,RUS 可以在两种模式下工作,一种是路径内模式,即该模式中的流量本身就是经过单播路由,另外一种是路径外模式,此时,原来充当 NAT 映射保持功能的定时 KEEPALIVE 消息就变为了承载消息,只不过此时的一侧流量是由非单播路由传输的.

3.4 加入ICE框架

RUS 作为一种介于非中转和中转之间的 NAT 穿越方法,可以加入到 ICE 的框架中.对于支持 STUN 的非中转情况,RUS 需要额外的路由器支持和客户端软件修改,但是与需要增加网络整体负载的中转型解决方案(例如 TURN)相比,RUS 具有明显的性能优势,所以可以作为 ICE 框架中的穿越选择方法.而且 RUS 路由器是作为被动呈现方式提供服务的,所以并不会影响现有的 NAT 解决方法的实施.

3.5 应用示例

RUS 作为一种独立与具体应用之外的 NAT 穿越方法,可以应用于各种 P2P 网络应用场景.SIP[7]作为一种典型的 P2P 应用协议,目前主要用于多媒体会话(特别是音视频会话)的建立,其实 SIP 本身可以作为任意 P2P 应用的会话发起信令.本文以 SIP 为例说明 RUS 在会话中的使用.

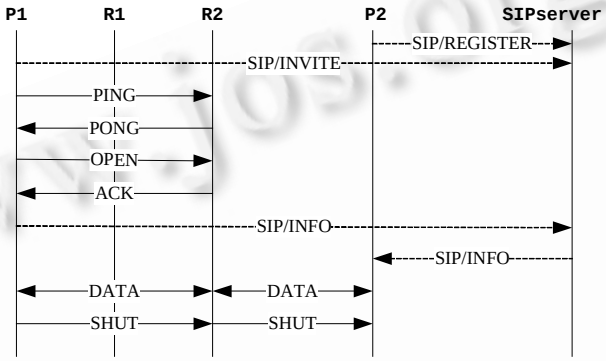

Fig.4 RUS usage sample in SIP application

图 4 RUS 在 SIP 会话中的应用示例

图 4 显示了 RUS 在 SIP 应用中的使用.P1 和 P2 为 SIP 通信双方节点,它们都位于对称 NAT 后面(图中省略了双方的 NAT 设备).P2 作为被叫一方向 SIP 服务器(SIPserver)注册了自己的信令地址,支持 NAT 环境的 SIP 服务器可以从接收到的 UDP 包的源地址获取 P2 的通告地址,当主叫一方 P1 节点发起呼叫时,首先从 SIP 服务器处获取了 P2 的通告地址(通过 SIP/INVITE 消息).由于 P1 和 P2 都位于对称 NAT 后面(这一点需要从 RUS 之外的途径获取,例如 STUN),所以 P1 一方发起 RUS 探测,途中的 R2 路由器支持 RUS 操作,P1 在通过 ACK 消息得到合法的 SID 之后将该信息通过 SIP/INFO 消息告知 P2 节点,该 RUS 交换会话可以被用于 P1 和 P2 间的媒体通道,也可以有其他用途.

```
INFO sip:100@20.1.2.3 SIP/2.0
Via: SIP/2.0/UDP 10.3.4.5
From: <sip:100@20.1.2.3>
To: <sip:200@20.1.2.3>
Call-ID: 321253
CSeq: 5 INFO
Content-Length: 20
Content-Type: application/x-udp-switch
SID=9277xyzNgWe
port=7008
```

以上是一个对 SIP/INFO 消息的扩展让其支持 RUS 的 SID 和 PeerPort 的通知,这样对方主机能够使用这些信息向 RUS 路由器发送 NAT 映射激活报文,让探测主机的流量能够经由该 NAT 映射到达对方主机.

4 总 结

NAT 穿越作为目前 IPv4 网络中 P2P 类应用必须解决的问题,已经显得越来越重要.传统的 NAT 解决方法大多采用服务器方式进行流量中转,但是这样必然导致网络整体负载的加大,因为中转服务器与路由器之间的流量重复使得沿途的所有路由器都必须对同一流量进行重复处理.RUS 作为一种基于路由器增强的解决方案,建立在通信双方主机间的原有单播路径传输之上,避免了重复处理对整个网络资源的消耗,使得局部的功能增强可以解决整体的瓶颈问题.文中还提出了在网络中部署和应用 RUS 的一些关注事项.下一步的工作主要集中在实现一个可集成的 RUS 路由器处理模块用于商业路由器的增强,并完善面向主机用户的编程接口库,同时,在实际网络环境中验证该方法的有效性.

References:

[1] Egevang K, Francis P. The IP network address translator. RFC1631, 1994.

[2] Borella M, Grabelsky D, Lo J, Taniguchi K. Realm specific IP: Protocol specificatoin. RFC3103, 2001.

[3] Internet gateway device (IGD) standardized device control protocol V 1.0. UPnP Forum. 2001. http://www.upnp.org

[4] Rosenberg J, Weinberger J, Huitema C, Mahy R. STUN—Simple traversal of user datagram protocol (UDP) through network address translators (NATs). RFC3489, 2003.

[5] Rosenberg J, Mahy R, Huitema C. Traversal using relay NAT (TURN). IETF Individual Draft, 2005.

[6] Rosenberg J. Interactive connectivity establishment (ICE): A protocol for network address translator (NAT) traversal for offer/answer protocols. IETF Draft, 2007.

[7] Rosenberg J, Schulzrinne H, Camarillo G, Johnston A, Peterson J, Sparks R, Handley M, Schooler E. SIP: Session initiation protocol. RFC3261, 2002.

张建伟(1971－),男,河南方城人,博士生,副教授,主要研究领域为下一代网络关键技术,网络安全.

战晓苏(1964－),男,博士,教授,博士生导师,主要研究领域为网络融合,网络安全管理,网格计算,协同计算.

皮人杰(1977－),男,博士,讲师,主要研究领域为分布式网络计算,嵌入式软件环境.

郭云飞(1964－),男,教授,博士生导师,主要研究领域为下一代网络关键技术.