文章编号：2096-1472(2021)-10-14-04
DOI:10.19644/j.cnki.issn2096-1472.2021.010.004

点对点传输协议虚拟专用网在大数据中的应用

刘邦桂

(广东开放大学人工智能学院，广东 广州 510091)
✉liubanggui@qq.com

摘 要：网络空间安全博弈日益激烈，先进的网络安全技术已成为主动应对安全威胁、及时打破安全攻防不对称局面的关键因素。针对复杂的网络环境和多样的网络应用需求，分析对比了多种虚拟专用网技术，设计了基于点对点传输协议的虚拟专用网方案。方案选用MS CHAP(微软质询握手身份验证协议)来验证客户端身份，采用MPPE(点对点加密算法)对公网地址封装私网地址后的IP数据包进行加密，确保了数据的机密性、完整性和可靠性。最后，利用免费的开源代码在Linux平台进行了验证。实验证明，该方案实现了大数据处理过程中数据从采集、传输、应用到存储和分析的安全保证，部署灵活，应用范围广，为大数据中心与数据源采集点，以及各企事业单位跨地区之间搭建专用网提供了参考。

关键词：网络空间安全；大数据；点对点传输协议；虚拟专用网
中图分类号：TP393.2 **文献标识码**：A

Application of Virtual Private Network based on Point-to-Point Tunneling Protocol in Big Data

LIU Banggui

(*School of Artificial Intelligence*, *The Open University of Guangdong*, *Guangzhou* 510091, *China*)
✉liubanggui@qq.com

Abstract: With the increasingly fierce security game in cyberspace, advanced network security technology has become a key factor in actively responding to security threats and breaking the asymmetry of security offense and defense in time. Aiming at the complex network environment and diverse network application requirements, this paper proposes to design a virtual private network solution based on point-to-point tunneling protocol by comparing a variety of virtual private network technologies. MS CHAP (Microsoft Challenge Handshake Authentication Protocol) is used to verify clients' identity, and MPPE (Point-to-Point Encryption Algorithm) is used to encrypt the IP data packets after the public network address encapsulates the private network address, so to ensure the confidentiality, integrity and reliability of the data. Finally, free open source code is used to verify the proposed solution on Linux platform. Experiments have proved that the security guarantee is realized from data collection, transmission, application to storage and analysis in the process of big data processing. The solution is flexible in deployment and has a wide range of application, providing a reference for building a private network between the big data center and data source collection points, as well as various cross-regional enterprises and institutions.

Keywords: cyberspace security; big data; point-to-point tunneling protocol; virtual private network

1 引言(Introduction)

随着全球信息化建设的快速发展，网络应用不断深入，网络规模逐渐变大，网络结构日趋复杂，加上网络空间具有开放性、虚拟性、资源丰富性与时空压缩化等优势[1]，对网络基础设施的功能和可延伸性提出了新的要求。例如，一些跨地区组织的各分支机构之间需要进行远距离互联；一些单位员工需要远程接入内部网络进行移动办公。为了解决各分支机构局域网之间的互联问题，早期只能直接铺设网络线路或

租用运营商的专线，不但成本高，而且实现困难。对于移动办公用户，一般采用拨号方式接入内部网络，在需要支付较高的网络流量通信费用的同时[2]，却无法保证数据的机密性、完整性和可靠性。

大数据处理过程中从采集点获取数据时，除了必要的身份认证以外，还需要对数据进行网络安全保护，对机密等级较高的数据使用加密算法进行加密处理，以确保数据不被篡改或非法获取，提高数据的正确性和完整性，减少冗余干扰信息，提升数据采集质量。为此在数据安全传输过程中，虚拟专用网技术拓宽了网络环境的应用，大数据传输可以采用虚拟专用网技术来建立安全通道，为数据中心和采集点之间远程连接提供廉价的方式，将需要保护的数据样本通过加密和封装处理后作为载荷嵌套在其他协议数据报文中传输[3]。本文在Linux平台中使用PPTP(点对点传输协议)实现虚拟专用网来满足安全需求。

2 VPN与隧道(VPN and tunneling)

2.1 VPN基本概念

VPN(Virtual Private Network)，中文名称为虚拟专用网络[4]。虚拟专用网络是一种虚拟通道，是一组通信协议，并不是一种独立的组网技术，目的是在公共网络基础设施上，通过隧道技术提供两个节点之间进行安全传输的专用通道。这条隧道可以有选择地对数据进行必要加密和认证，从而保证传输数据的机密性和完整性，只有通过了接入认证并且获得服务端授权的用户才能使用隧道。虚拟专用网技术通过建立可信的安全连接，运用在公司外出员工、合作伙伴、公司分支机构等有内部网访问需求的任何环境。这条承载在公用网络上的私有信息隧道在客户端看来，与在局域网内互访没有太大区别，加上配备有永久和临时隧道的不同特点，成为多种网络应用的优选“承载协议”。

VPN包含认证、授权和加密。认证是鉴定用户的真伪和通信数据的不可抵赖；授权是检查客户端拨入后可以访问哪些资源；加密是访问资源过程中对通信数据采用加密算法进行处理，以此来保证数据的机密性、完整性、可靠性和不可抵赖性。

VPN服务安全通道有主动和被动连接两种。主动连接一般首先需要客户端主动请求拨入VPN服务器，经过VPN服务器与认证中心对客户端进行身份验证并得到授权后，VPN服务器与客户端建立连接才开始传送数据，适合在少量或者单个用户不定时不固定位置与VPN服务连接的情况。被动连接主要是在连接两个远程局域网的网关位置进行，内网客户端无须配置便可与对端内网用户通信，适合在大量用户固定或者暂时固定的场合使用。

2.2 隧道协议

VPN采用隧道技术进行通信。待转发的数据经过源局域网与公网的接口时，用设定的隧道协议将数据包作为有效载荷封装进行封装，封装后的数据包经过公网路由转发到达目的局域网与公网接口时，由同样的隧道协议解封装，再取出载荷中的源局域网中传输的数据包转发到目的主机。被封装数据包在公网上传输时经过的逻辑路径称为“隧道”。

常用隧道协议有PPTP、L2TP、GRE、IPSec等。其中PPTP、L2TP是网络参考模型第二层的隧道协议，GRE、IPSec是第三层的隧道协议。

(1)PPTP(Point-to-Point Tunneling Protocol)点对点传输协议[5]，实现的前提是通信双方有连通且可用的IP网络，服务器监听端口号为1723，有认证、加密等功能。

(2)L2TP(Layer 2 Tunneling Protocol)第二层隧道协议，使用UDP协议封装，协议端口号为1701，默认无加密算法，若想使用加密算法，可结合IPsec。

(3)GRE(Generic Routing Encapsulation)通用路由封装协议是由思科公司提出的，目前版本是GREv2，可以应用于多种承载和载荷协议，为解决IPv4和IPv6技术孤岛提供了解决方案，提供了对建立隧道双方用户的认证，没有对数据的加密功能。

(4)IPsec(Internet Protocol Security)Internet协议安全[6]，其基本思想是把与密码学相关的安全机制引入IP协议，通过现代密码学所创立的方法来支持保密和验证服务，使用户可以有选择地使用所提供的功能，并得到所要求的安全服务。原本IPv6的制定才产生了安全性较高的IPsec，但目前IPv4协议还在广泛应用，所以在IPsec标准制定过程中也增加了对IPv4的支持，引入了IKE等密钥交换协议，工作模式有多种，数据加密和验证算法多样，能嵌套在GRE、l2tp中，满足大部分使用的场景，为此部署也相对复杂，对设备的要求也较高，一般在路由器和防火墙设备上实现。

除此之外，还有多种VPN实现技术，比如OpenVPN。它是一种基于OpenSSL库和SSL/TSL协议的应用层VPN，属于免费开源软件，加密强度高，信息的机密性和完整性保护效果好，还可以配置在任意端口运行，具有NAT穿越等功能，但安装和配置过程复杂，连接速度和传输效率相对较低。

3 点对点传输协议(Point-to-point tunneling protocol)

PPTP协议是常用的一种协议，是在PPP协议和TCP/IP协议上开发的二层隧道协议，通过加密、认证、压缩算法增强了安全性和可靠性。PPTP将PPP帧封装成IP数据包，在互联网上进行传输。PPTP运用TCP通过三次握手实现隧道创建、维护与终止，使用GRE(通用路由封装)封装隧道数据的PPP帧。PPTP通信与FTP类似，需要建立控制连接和数据连接。控制连接主要进行连接的维护，数据连接主要负责数据的通信。

3.1 PPTP工作过程

对于基于PPTP的VPN，由于客户端通过拨号方式接入VPN服务器，因此该VPN服务器也称为VPDN虚拟专用拨号网(Virtual Private Dial-Up Network)服务器。其实现过程如下：

(1)发送建立连接请求。在VPDN服务器为PPTP客户端

预先建立好用户账户(包括登录账号和密码)。PPTP客户端利用VPN连接软件(Windows操作系统自带)向VPDN服务器发起连接请求。在客户端输入VPDN服务器对外公开的IP地址后，首先将分配的用户账号和密码发送到服务器进行认证和授权。PPTP对用户进行认证的方法有：密码身份验证协议PAP(Password Authentication Protocol)、质询握手身份验证协议CHAP(Challenge Handshake Authentication Protocol)、微软质询握手身份验证协议MS CHAP(Microsoft Challenge Handshake Authentication Protocol)等。

(2)如果用户认证通过，正式建立VPN连接，返回连接完成信息。

(3)VPN安全隧道建立后，将进行正常数据传输。为保证数据传输的机密性，可以选用数据加密算法DES(Data Encryption Standard)、点对点加密算法MPPE(Microsoft Point-to-Point Encryption)对IP数据进行加密。如果我们的客户端是Windows操作系统，默认使用MPPE。

(4)数据传输至目的地。VPDN服务器收到发往内网的PPTP数据包后,将按流程对其进行解封装,解封装后从PPTP数据包中取出目的地址为本地内网主机的网络地址,然后按照原先设定的私网路由进行转发，完成传输。

3.2 数据封装与解封装过程

PPTP协议数据采用多层封装的方式。具体封装和解封装的过程如图1所示。

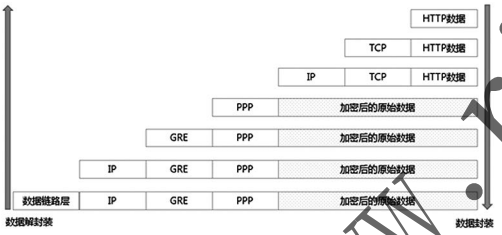

图1 数据封装与解封装过程

Fig.1 Data encapsulation and decapsulation process

封装过程：(1)初始应用层数据封装成IP数据包；(2)利用内网私有路由将数据包发送到VPN虚拟接口；(3)在虚拟接口中使用设置的安全协议，对其压缩和加密，添加PPP头部信息，封装成PPP帧后发送给PPTP协议；(4)为PPTP帧添加GRE头部信息后提交给TCP/IP协议；(5)TCP/IP协议为GRE报头添加公网源IP地址和目的IP地址；(6)为IP数据包进行数据链路层封装后通过物理层与普通数据包一样选择合适路由进行转发。

解封装过程：(1)物理层收到数据包；(2)链路层剥掉外层帧后交给TCP/IP协议；(3)TCP/IP协议剥掉IP头；(4)IP协议剥掉GRE头后通过内网私有路由，将PPP帧发给VPN虚拟接口；(5)VPN虚拟接口剥掉PPP头对有效载荷进行解压缩或解密后提交给上层应用；(6)对数据进行普通处理。

4 方案设计(Solution design)

4.1 方案背景

为了数据采集点能在外网环境下通过公网访问内部数据中心，需要内网VPN服务网关给采集点提供一个内网IP地址，通过访问侧和公网侧的网关对数据进行封装和解封装，以实现数据通信，为此整个过程均需要对应的公网和私网路由。对于采集点较多，而且需要同一时间访问内部服务的公司来说，需要购置一台专门支持PPTP VPN的路由器或者防火墙，资金允许的情况下还可以购置专门用于用户认证的服务器。对于采集点访问量不是很大的情况，可以使用Windows或者Linux服务器操作系统来实现，既节约成本又方便管理。本文用一台安装有Red Hat Linux 7.4的服务器来开展研究。

4.2 方案需求分析

为了保证采集点和数据中心实现安全数据传输，加上数据交换不是很多，在不增加购买设备的情况下，利用公司企业内网配置一台双网卡的Linux服务器来做VPN服务器，由于经过解包、配置、编译和安装四个步骤就能进行配置和使用，开源软件pptpd源码包成为方案的首选。方案网络拓扑图如图2所示。

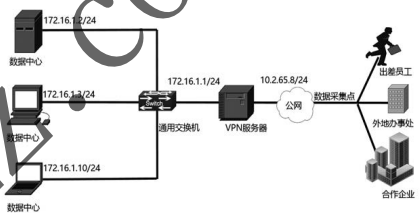

图2 网络拓扑图

Fig.2 Diagram of the network topology

此方案需要内网配置有数据中心，通过图2中的通用交换机互联，VPN配置在安装有Linux 7.4的软网关上，该网卡上的IP地址也是处在外网的采集点，包括员工、办事处和合作企业，均采用通过拨号获得网关分配的内网地址后连入内外的唯一可识别公网地址，内网数据中心的地址全部不对外公开，方案验证之前需要确保内网到VPN内外地址、外网到VPN网关外网地址的路由连通性。

5 方案验证(Solution verification)

5.1 VPN服务器端的配置

VPN服务器安装在Linux 7.4中，在系统中配置VPN服务，基于PPTP的VPN服务名称是pptpd。pptpd是PPTP的守护进程，用来管理基于PPTP隧道协议的VPN连接。当pptpd接收到用户的VPN接入请求后，会自动调用PPP协议的pptpd程序来完整验证，然后建立VPN连接，需要安装ppp和pptpd两个软件包。

(1)安装ppp和pptpd。

ppp安装ppp服务：

[root@vpnserver/]# yum install

解压缩pptpd源码文件：

[root@vpnserver vpn]# tar zxfv pptpd-1.4.0.tar.gz

使用源码目录中的configure脚本将程序安装到指定目录：

[root@vpnserver pptpd-1.4.0]#./configure--prefix=/usr/local/pptpd

使用make命令将源代码文件变为二进制的可执行程序：

[root@vpnserver pptpd-1.4.0]#make

将上一步编译好的程序文件复制到系统中：

[root@vpnserver pptpd-1.4.0]#make install

(2)基于pptpd协议的VPN服务主配置文件是/etc/pptpd.conf。使用模板生成主配置文件，主要配置分配给内网的地址。

[root@vpnserver etc]# vi pptpd.conf

设置pppd程序的位置：

ppp/usr/sbin/pppd

设置options文件的位置：

option/etc/ppp/options.pptpd

局域网分配给客户机的网关地址：

localip 192.168.1.1

局域网分配给客户机的IP地址：

remoteip 192.168.1.7-177

(3)由于pptpd在接收到用户VPN接入请求后，会自动调用ppp服务来完成验证过程，以建立VPN连接，因此，要使pptpd服务正常工作，还必须在ppp配置文件中对VPN连接验证服务等进行相关的配置。ppp选项文件由pptpd.conf文件中的option参数指定默认为/etc/ppp/options.pptpd，文件中可以设置身份验证方式、加密长度，以及为VPN客户端指定的DNS服务器和WINS服务器的IP地址。

[root@vpnserver ppp]# vi options.pptpd

设置DNS服务器地址：

ms-dns 8.8.8.8

默认使用/etc/ppp/chap-secrets文件来进行VPN用户身份验证：

auth

(4)在ppp选项文件中，已通过auth选项指定默认使用/etc/ppp/chap-secrets安全验证文件进行身份验证，所以创建VPN用户和密码可以通过直接编辑该文件来完成。

[root@vpnserver ppp]#vi/etc/ppp/chap-secrets

#client　server　secret　IP addresses

wgsu1　pptpd　123456

启动PPTP对应的服务：

[root@vpnserver~]#/usr/local/pptpd/sbin/pptpd

[root@vpnserver~]#systemctl stop firewalld

5.2 客户端连接

客户端通过新建网络连接，输入服务器地址和登录用的账号和密码，登录成功后可以查看客户端获取了服务器分配的内网地址，同时查看服务端的日志发现客户端详细信息。客户端通过分配的内网地址成功访问内网服务群中的数据中心，查看连接情况可以了解使用的身份验证和数据加密算法，如图3和图4所示。

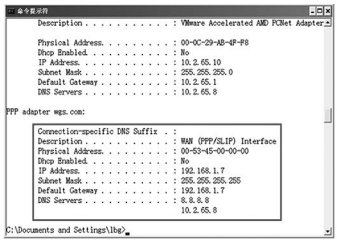

图3 客户端的地址

Fig.3 Client address

图4 连接状态

Fig.4 Connection status

6 结论(Conclusion)

安全与应用相伴而生，随着大数据应用不断发展，安全问题也随之出现。该方案在Linux平台上建立了PPTP VPN，实现了外网对内网的私有访问，并对客户端进行了MS CHAP认证，采用MPPE对数据加密，保证了PPTP客户端与服务器之间的安全通信。客户端无须下载任何软件便可连接，能够满足安全等级中等的适用场合，也可以推广运用于大部分中小企事业单位。由于组网技术的原因，PPTP VPN 无法穿越NAT[7]，如果服务器的网关在防火墙内部，客户端将无法建立VPN连接，为此需要结合第三层的IPsec VPN来混合使用[8]，通过防火墙来增强VPN系统的安全性和实用性。

参考文献(References)

[1] 岳少博,王清河,王晓春,等.基于虚拟专用技术的网络空间防御方法仿真[J].计算机仿真,2020,37(5):273-277.

[2] 陈良臣,高曙,刘宝旭,等.网络流量异常检测中的维数约简研究[J].计算机工程,2020,46(2):11-20.

[3] 张尧,刘笑凯.基于国密算法IPSec VPN设计与实现[J].信息技术与网络安全,2020,39(6):49-52.

[4] 季征南,马立国,吴英梁,等.计算机网络信息安全中虚拟专用网络技术的应用研究[J].冶金管理,2021(1):185-186.

[5] SAITO S, UEHARA T, IZUMI Y, et al. Implementation and performance evaluation of pass-through PPTP relay system with authentication at each gateway[J]. Electrical Engineering in Japan, 2010, 154(2):40-51.

[6] 段翠华.计算机网络安全中虚拟网络技术的应用[J].网络安全技术与应用,2021(1):8-9.

[7] 倪洁,徐志伟,李鸿志.PPTP VPN与L2TP/IPSec VPN的实现与安全测试[J].电子技术与软件工程,2019(12):192.

[8] 彭治湘.L2TP over IPSec VPN NAT穿越技术研究与实验仿真[J].信息技术与信息化,2020,247(10):210-212.

作者简介：

刘邦桂(1983-)，男，硕士，讲师.研究领域：服务器技术，网络安全技术.