linux Socket-应用编程-理论概念-专题讲座

written by 王保明

## ISO的开放互联网模型（OSI）

|  |
| --- |
| **TCP回射客户/服务器模型总结**   * OSI（open system interconnection）开放系统互联模型是由ISO（International Organization for Standardization）国际标准化组织定义的网络分层模型，共七层。 |
| * 物理层(Physical Layer)：物理层定义了所有电子及物理设备的规范，为上层的传输提供了一个物理介质，本层中数据传输的单位为比特（bit）。属于本层定义的规范有EIA/TIA RS-232、EIA/TIA RS-449、V.35、RJ-45等，实际使用中的设备如网卡等属于本层。 * 数据链路层（Data Link Layer）：对物理层收到的比特流进行数据成帧。提供可靠的数据传输服务，实现无差错数据传输。在数据链路层中数据的单位为帧（frame）。属于本层定义的规范有SDLC、HDLC、PPP、STP、帧中继等，实际使用中的设备如switch交换机属于本层。 * 网络层（Network Layer）：网络层负责将各个子网之间的数据进行路由选择，分组与重组。本层中数据传输的单位为数据包（packet）。属于本层定义的规范有IP、IPX、RIP、OSPF、ICMP、IGMP等。实际使用中的设备如路由器属于本层。 |
| * 传输层（Transport Layer）：提供可靠的数据传输服务，它检测路由器丢弃的包，然后产生一个重传请求，能够将乱序收到的数据包重新排序。 * 会话层（Session Layer）：管理主机之间会话过程，包括会话建立、终止和会话过程中的管理。 * 表示层（Presentation Layer）：表示层对网络传输的数据进行变换，使得多个主机之间传送的信息能够互相理解，包括数据的压缩、加密、格式转换等。 * 应用层（Application Layer）：应用层与应用程序界面沟通，以达至展示给用户的目的。 在此常见的协定有: HTTP，HTTPS，FTP，TELNET，SSH，SMTP，POP3等 * 结论：传输层能做差错控制。。对IP包进行排序 IP路由   **OSI和邮件发送** |

## TCP/IP四层模型

|  |
| --- |
| **和OSI七层的不同**    **TCP/IP主要协议**    **运行FTP的两台主机** |
| **通过路由器连接的网络**     * **（对等通信、封装、分用、端口）**   **对等通讯**    **封装（Encapsulation）**    **分用（Demultiplexing,解封）**    **端口**   * 众所周知端口（Well Known Ports）：从0到1023，这些端口由IANA分配和控制它们紧密绑定于一些服务。通常这些端口的通讯明确表明了某种服务的协议。例如：21端口为ftp服务端口。 * 注册端口（Registered Ports）：从1024到49151。它些端口不受IANA控制，但由IANA登记并提供使用情况清单。它们松散地绑定于一些服务。也就是说有许多服务绑定于这些端口，这些端口同样用于许多其它目的。例如：1433 Microsoft SQL服务端口 * 动态或私有端口（Dynamic or Private Ports）：从49152到65535。IANA不管这些端口。实际上，机器通常从1024起分配动态端口。但也有例外：SUN的RPC端口从32768开始。 |

## TCP/IP各层报文

### 以太网帧格式

|  |
| --- |
| 说明1：链路层数据包，称为以太网帧。  说明2：链路层不识别IP地址（逻辑地址，公网IP），链路层识别物理网卡MAC地址。  说明3：需要根据IP地址找到对方的MAC地址。这个过程称为：地址解析，也就是ARP协议。（MAC—）IP地址方向地址解析 RARP）。  说明4：应用层根据对等方的IP地址进行通讯，在数据封装过程中，链路层有需要目的地址的MAC网卡地址？怎么样解决这个问题？ 需要将IP地址转换成MAC地址，也就是地址解析。 |
| ARP地址解析协议 |
|  |

### 以太网的MTU和路径MTU

|  |
| --- |
| 以太网MTU   * 以太网和IEEE 802.3对数据帧的长度都有限制，其最大值分别是1500和1492字节，将这个限制称作最大传输单元（MTU，Maximum Transmission Unit） * 如果IP层有一个数据报要传，而且数据的长度比链路层的MTU还大，那么IP层就要进行分片（Fragmentation），把数据报分成若干片，这样每一片都小于MTU。 * 当网络上的两台主机互相进行通信时，两台主机之间要经过多个网络，每个网络的链路层可能有不同的MTU，其中两台通信主机路径中的最小MTU被称作路径MTU。 |

### IP数据报

|  |
| --- |
|  |
| * 版本号，目前取值4 * 首部长度，4个字节为单位，取值范围5～15 * 服务类型，指定传输的优先级、传输速度、可靠性和吞吐量等 * 报文总长度，最大长度为65535字节 * 报文标识，唯一标识一个数据报，如果数据报分段，则每个分段的标识都一样 * 标志，最高位未使用，定义为0，其余两位为DF（不分段）和MF（更多分段） * 段偏移量，以8个字节为单位，指出该分段的第一个数据字在原始数据报中的偏移位置 |
| * 生存时间，取值0～255，每经过一个路由节点减1，为0时被丢弃 * 协议，指明该数据报的协议类型，如1为ICMP，4为IP，6为TCP，17为UDP等 * 首部校验和，每通过一次网关都要重新计算该值，用于保证IP首部的完整性 * 选项，长度可变，提供某些场合下需要的控制功能，IP首部的长度必须是4个字节的整数倍，如果选项长度不是4的整数倍，必须填充数据 |
|  |
|  |
|  |

### ICMP协议

|  |
| --- |
| ICMP协议说明   * ICMP协议用于传递差错信息、时间、回显、网络信息等控制数据 * 当你ping 一台主机想看它是否运行时，就产生了一条ICMP 信息。远程主机将用它自己的ICMP 信息对ping 请求作出回应。   **ICMP协议**   * 本身是IP的一部分。 * 用途:网关或者目标机器利用ICMP与源通讯，当出现问题时，提供反馈信息 * 在IP协议栈中必须实现 * 特点：   + 其控制能力并不用于保证传输的可靠性   + 它本身也不是可靠传输的 * ICMP包含在IP数据包的净荷数据中，IP头中协议类型为1 * ICMP数据的第一个字节代表ICMP报文的类型，它决定了后续数据的格式   **ICMP数据包**   * ICMP包含在IP数据包的净荷数据中，IP头中协议类型为1 * ICMP数据的第一个字节代表ICMP报文的类型，它决定了后续数据的格式       **ICMP报文类型**      报文类型说明   * 0 Echo Reply * 3 Destination Unreachable * 4 Source Quench * 5 Redirect * 8 Echo * 11 Time Exceeded * 12 Parameter Problem * 13 Timestamp * 14 Timestamp Reply * 15 Information Request * 16 Information Reply * 17 Address Mask Request * 18 Address Mask Reply   报文分类   * ICMP差错报文   + 目的不可达报文（类型3）   + 超时报文（类型11）   + 参数出错报文（类型12） * ICMP控制报文   + 报源抑制报文（类型4）   + 重定向（类型5） * ICMP请求/应答报文   + 回送请求和响应报文（类型0和8）   + 时间戳请求和响应报文（类型13和14）   + 地址掩码请求和响应报文（类型17和18） * 其他 |
| **ICMP Echo报文**     * 类型：0表示Echo Reply，8表示Echo * 代码：0 * 标识符：标识一个会话，例如，用进程ID * 序号：例如每个请求增1 * 选项数据：回显 |
| **ICMP 目的不可达报文**     * 类型：3 * 代码：0表示网络不可达，1表示主机不可达； 2表示协议不可达；3表示端口不可达；等等 * 出错的IP包的IP首部+原始IP数据包中前8个字节 |

**UDP数据报**

|  |
| --- |
|  |

### TCP报文段

|  |
| --- |
| **协议描述**     * 源端口号和目的端口号：源和目的主机的IP地址加上端口号构成一个TCP连接 * 序号和确认号：序号为该TCP数据包的第一个数据字在所发送的数据流中的偏移量；确认号为希望接收的下一个数据字的序号； * 首部长度，以4个字节为单位，通常为20个字节 * 6个标志位：   + URG：如果使用了紧急指针，URG置1，紧急指针为当前序号到紧急数据位置的偏移量   + ACK：为1表示确认号有效，为0表示该TCP数据包不包含确认信息   + PSH：表示是带有PUSH标志的数据，接收到数据后不必等缓冲区满再发送 |
| * + RST：用于连接复位，也可用于拒绝非法的数据或拒绝连接请求   + SYN：用于建立连接，连接请求时SYN＝1，ACK=0；响应连接请求时SYN=1，ACK=1   + FIN：用于释放连接，表示发送方已经没有供发送的数据 * 窗口大小：表示在确认字节后还可以发送字节数，用于流量控制 * 校验和：覆盖了整个数据包，包括对数据包的首部和数据 * 选项：常见的选项是MSS(Maximum Segment Size) |
| **连接建立三次握手** |
| **连接终止四次握手** |
| **TCP如何保证可靠性**   * 应用数据被分割成TCP认为最适合发送的数据块，称为段传递给IP层。 * 当TCP发出一个段后，它启动一个定时器，等待目的端确认收到这个报文段。如果不能及时收到一个确认，将重发这个报文段。 * 当TCP收到发自TCP连接另一端的数据，它将发送一个确认。这个确认不是立即发送，通常将推迟几分之一秒。 * TCP将保持它首部和数据的校验和。这是一个端到端的校验和，目的是检测数据在传输过程中的任何变化。如果收到段的校验和有差错，TCP将丢弃这个报文段并且不确认（导致对方超时重传） * TCP承载于IP数据报来传输，而IP数据报的到达可能会失序，因此TCP报文段的到达也可能会失序。TCP将对收到的数据进行重新排序。 * IP数据报会发生重复，TCP的接收端必须丢弃重复的数据。 * TCP还能提供流量控制。TCP连接的每一方都有一定大小的缓冲空间。 |

### 综合案例

|  |
| --- |
| **数据在网络中传输过程**     * 步骤a：应用程序ping会判断发送的是主机名还是IP地址，调用函数gethostbyname()解析主机机B，将主机名转换成一个32位的IP地址。这个过程叫做DNS域名解析 * 步骤b：ping程序向目的IP地址发送一个ICMP的ECHO包 * 步骤c：将目标主机的IP地址转换为48位硬件地址，在局域网内发送ARP请求广播，查找主机B的硬件地址。 * 步骤d：主机B的ARP协议层接收到主机A的ARP请求后，将本机的硬件地址填充到应答包，发送ARP应答到主机A。 * 步骤e：发送ICMP数据包到主机B * 步骤f：主机B接收到主机A的ICMP包，发送响应包。 * 步骤g：主机A接收到主机B的ICMP包响应包。 |

### 其他补充

|  |
| --- |
| **主要协议及其关系**    **TCP/IP协议栈** |
|  |