

第二讲 网络基础：网络模型与直连网络

中国科学院计算技术研究所

网络技术研究中心

个人简介

- 武庆华，博士，副研究员
- 研究方向：
 - 互联网体系结构
 - 互联网测量与优化
- Email: wuqinghua@ict.ac.cn

课程大纲

- 第1讲 计算机网络概述
- 第2讲 网络基础：直连网络
- 第3讲 网络基础：网络互连
- 第4讲 网络基础：网络路由
- 第5讲 网络基础：网络传输
- 第6讲 网络基础：网络应用、协议栈实现
- 第7/8讲 网络基础：服务质量
- 第9讲 网络基础：网络安全
- 第10+讲 网络专题



我讲的内容

本讲提纲

- 计算机网络体系结构模型
 - 分层网络的形成机理与演化
- 直连网络 (Direct Link Networks)
 - 更关注数据链路层
 - 如何发送数据
 - 如何共享链路发送数据

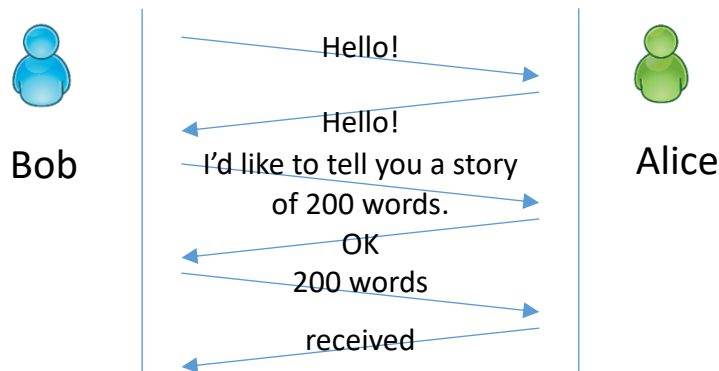
计算机网络体系结构模型

分层网络模型(Layered Network Model)

- 为什么需要分层网络模型
- 如何定义分层网络模型
- 分层模型的不足和改进

协议是计算机网络的基本组成部分

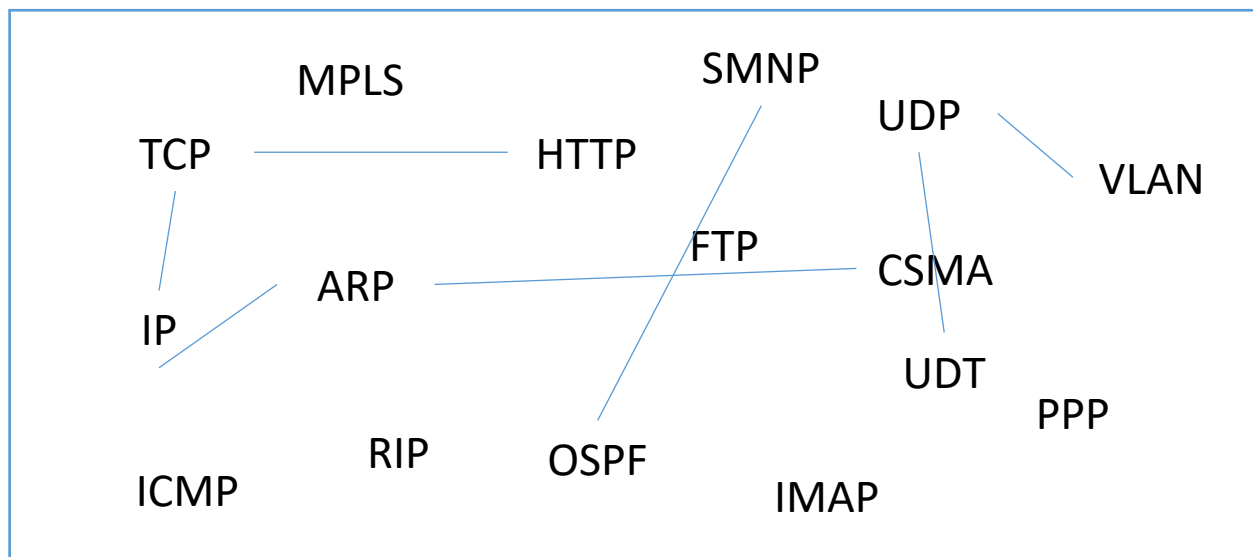
- 网络通信中需要很多很多协议、功能、组件的协作
 - 传输纠错，网络路由，地址解析，流量控制，应用代理，
 - RFC协议规范文档已超过8000件
- 什么是协议(Protocol) ?
 - 定义通信参与方如何交互的一种协定和规范



- 如何使用这些协议构建计算机网络？

方案1 – 模块化

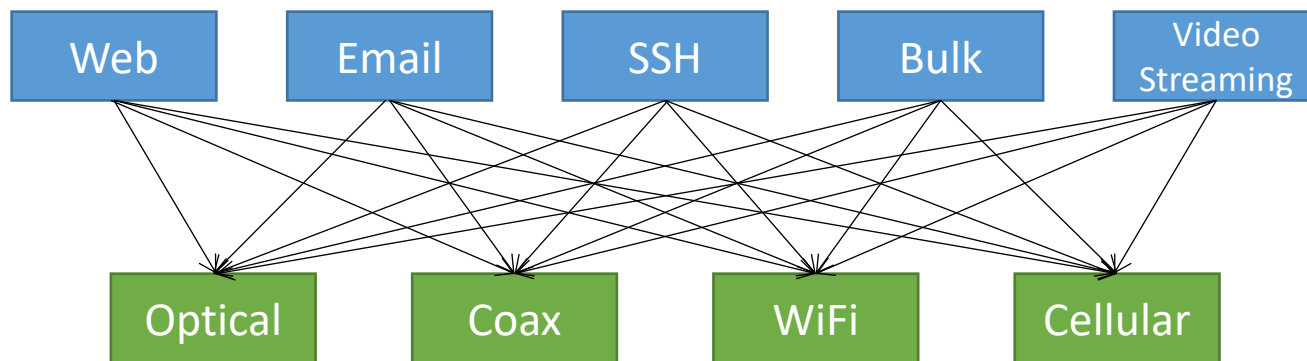
- 模块化的协议栈 + 良好定义的模块接口
- 优点: 独立性强, 功能简单, 模块易于实现
- 缺点: 适应性差, 难以维护, 对系统实现是场灾难



模块化的网络协议栈

方案2 – 两层结构

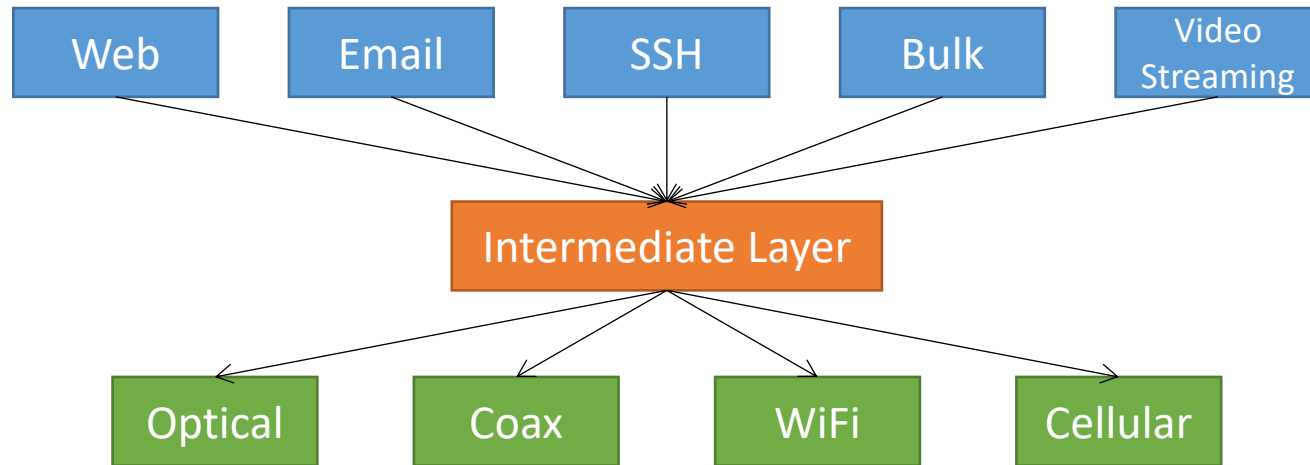
- 将计算机网络分为两层
 - 应用层与物理层，**两层的每种类型一对一适配**
- 优点: 相比于方案1，结构更清晰，更易于维护
- 缺点: 每有一种新的应用，都需要与物理层的每种类型去适配



两层结构的计算机网络体系结构模型

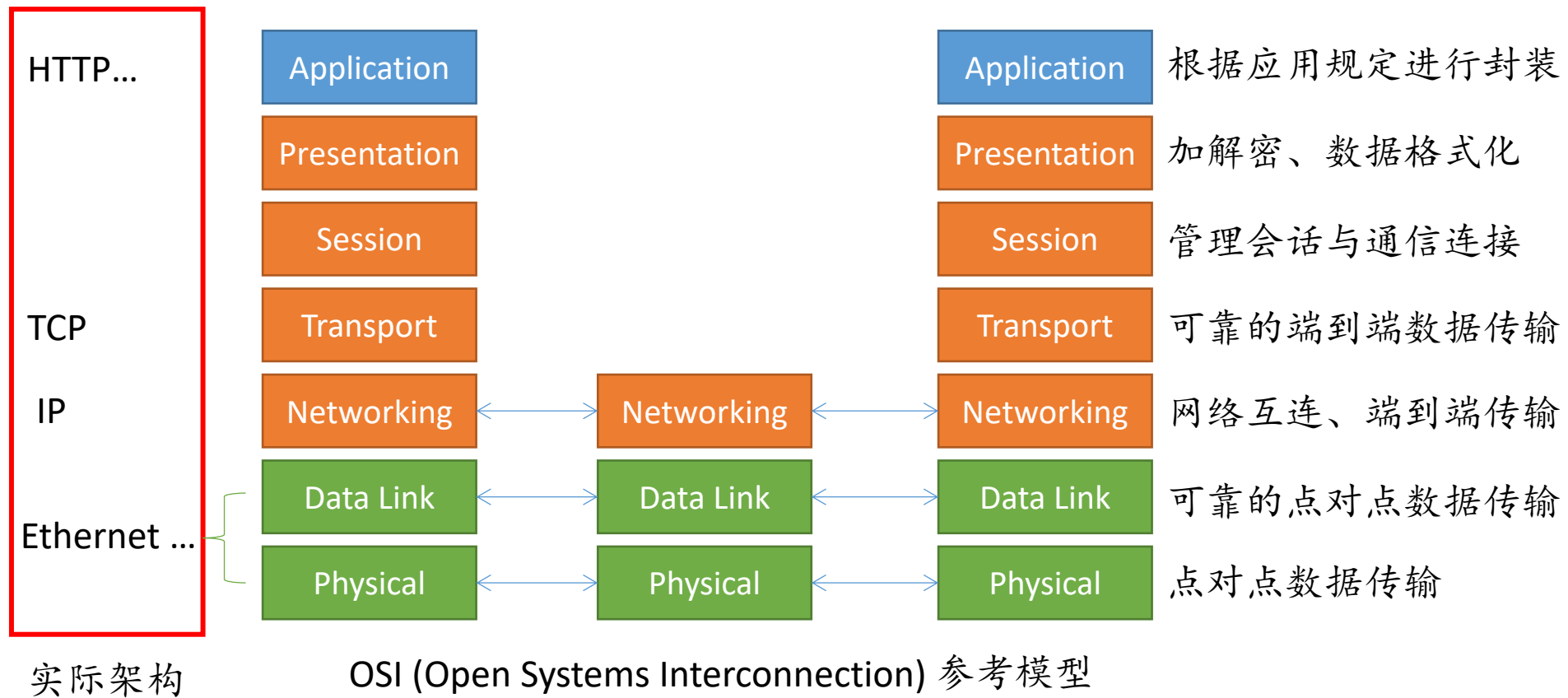
方案3 – 三层结构

- 引入中间层，分别与应用层和物理层适配
 - 中间层实现对物理层的抽象，对应用层提供统一接口
- 优点: 极大降低了适配的工作量， $M*N \rightarrow M+N$
- 缺点: 引入适配层增加了设计难度 (可忽略不计)

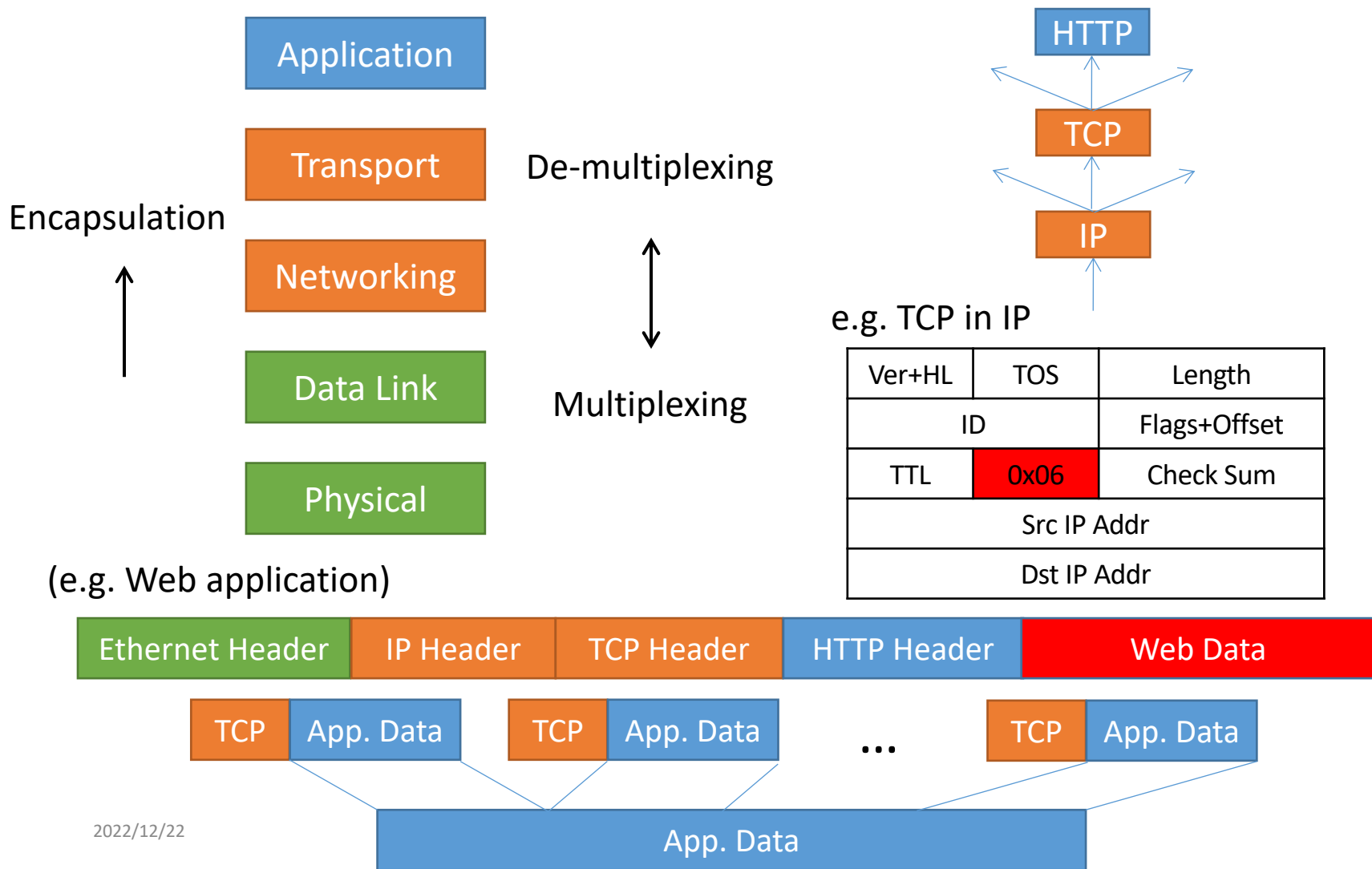


三层结构的计算机网络体系结构模型

分层模型 – 理论模型和实际架构

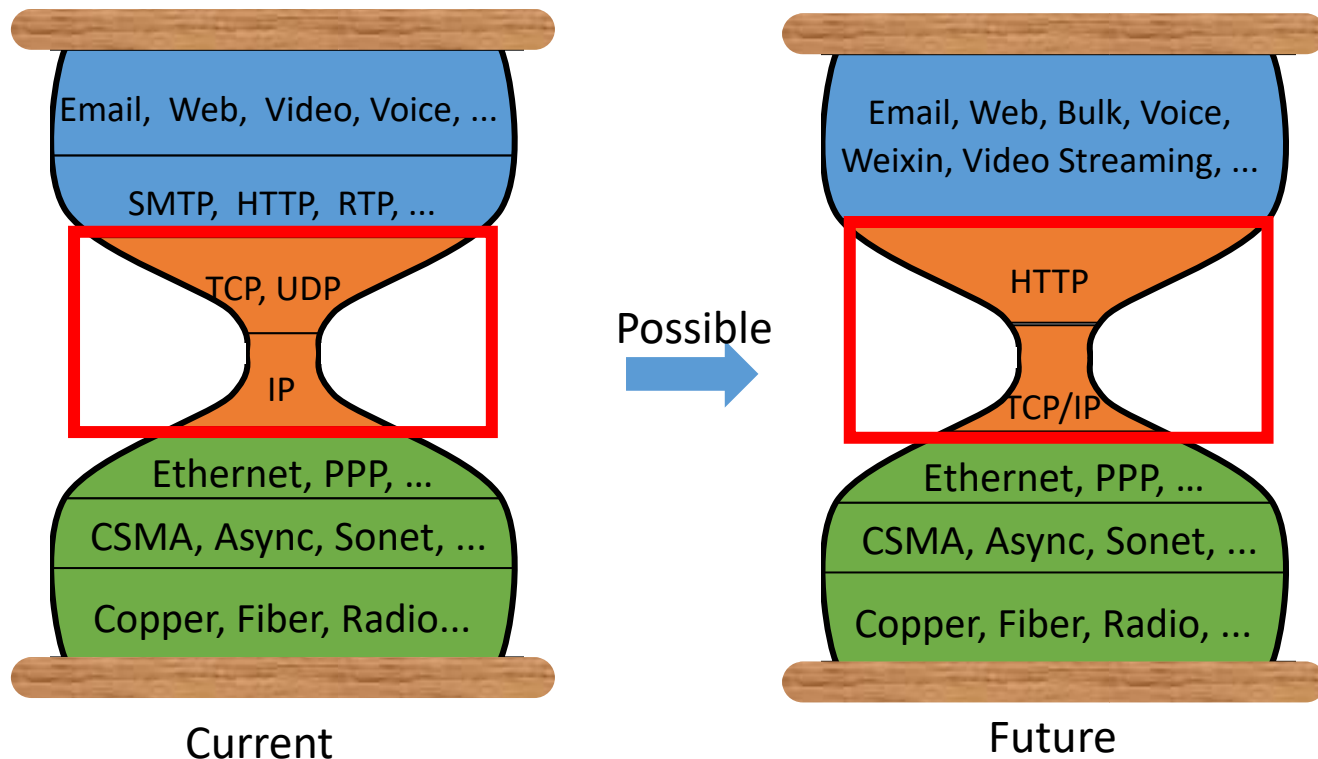


分层模型工作机制 – 封装和复用



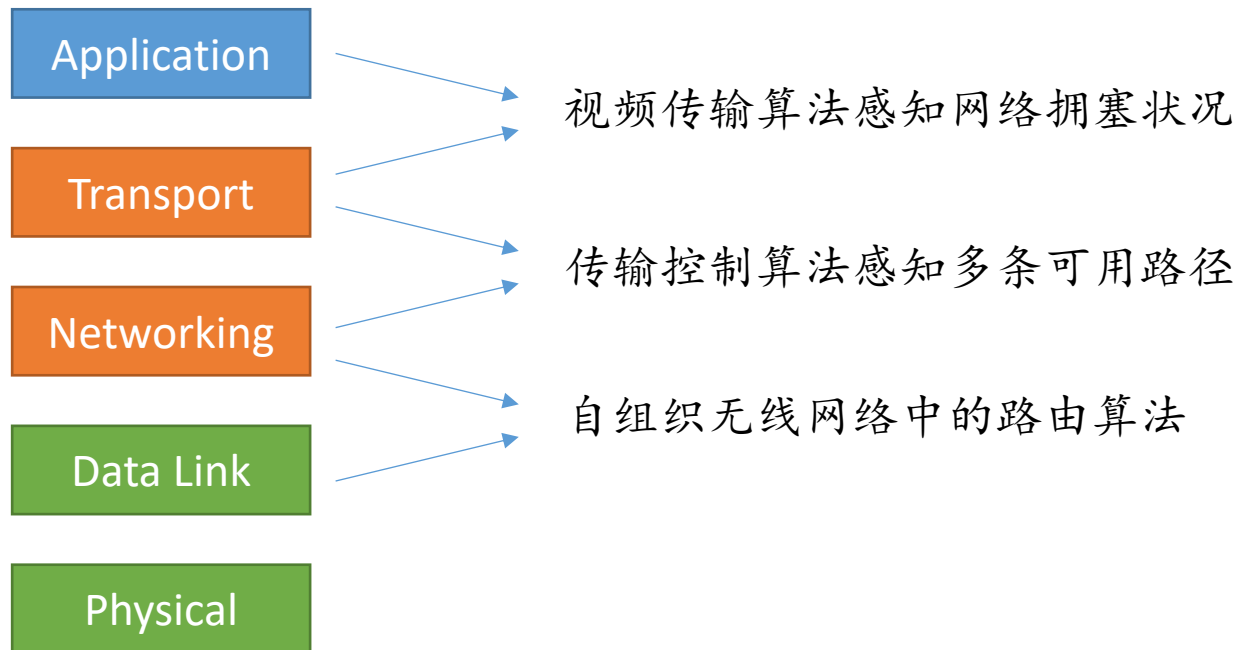
网络体系结构的细腰特征

- 细腰结构(narrow-waist)是互联网体系结构模型中最典型的特征
 - 研究表明，分层的体系结构最终会演化成细腰模型 [1]
 - 互联网体系结构一直在演进中，现有结构可能会演化成新的细腰模型 [2]



分层模型的改进

- 理论上，每层只需要实现**对上层提供接口，对下层复用**
 - 大多数情况下都可以取得不错的性能结果
- 现实中，网络环境的复杂性使得分层独立设计难以达到最优性能
 - 跨层(Cross Layer)感知可以弥补这一缺陷



直连网络

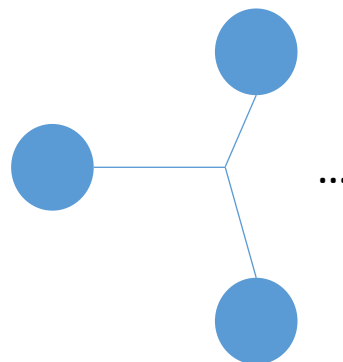
直连网络

- 如何发送数据？
 - 数据帧封装
 - 差错检测
 - 可靠传输
- 节点间如何共享链路？
 - 面向固定带宽分配的多路复用机制
 - 争用式多路复用机制
- 不同网络接入方式
 - 以太网(Ethernet)
 - 无线局域网(WiFi)
 - 蜂窝通信网络(5G)

直连网络模型与性能指标

- 网络模型

- 点线模型，包括有线网、无线网

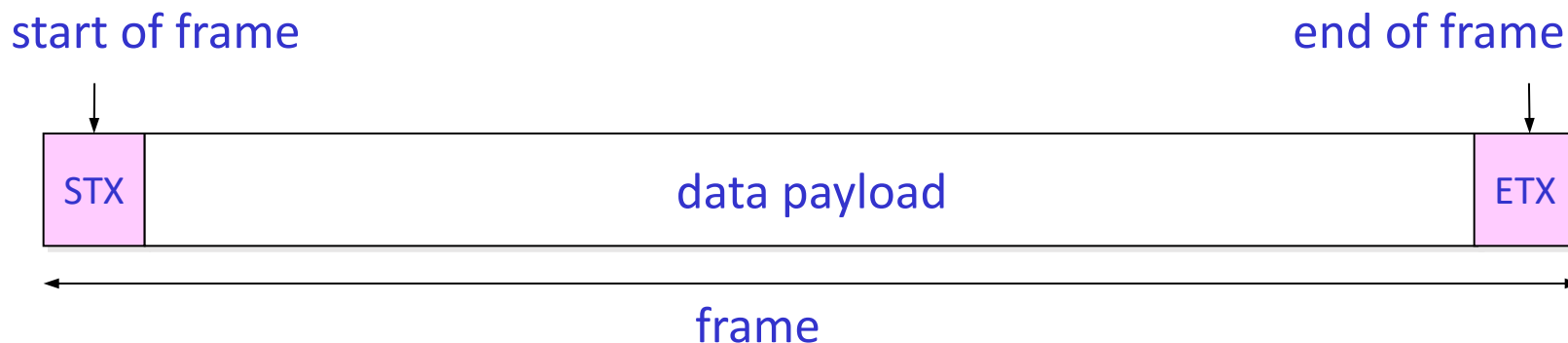


- 性能指标

- **带宽(Bandwidth)**: 单位时间内通过链路所能传输的比特数
 - 用户网络接入带宽: 200Mbps; WiFi接入带宽: 150Mbps
 - 应用层更关心单位时间传输的字节数
 - **时延(Latency)**: 消息从网络一端传到另一端所需的时间
 - 传播时延 + 处理时延 + 排队时延
 - 大多时候更关注往返时间 (RTT)
 - 北京到广州的往返时间为30ms

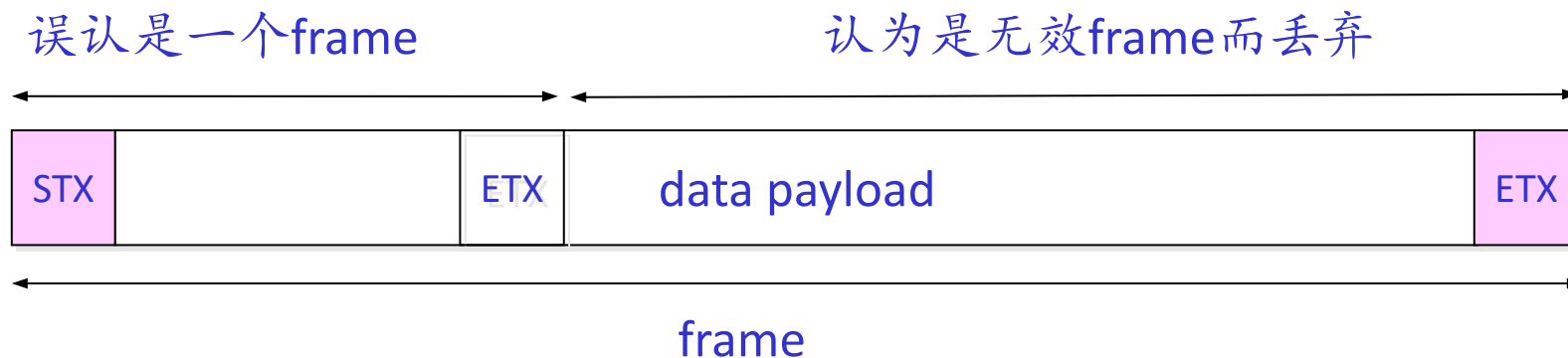
数据帧封装

- 在一段数据的前后分别添加首部和尾部，就构成了一个帧
 - 不同协议对数据长度有上、下限规定
- 首部和尾部的作用是界定帧的范围
 - 不同的数据链路协议有不同的格式定义，这里以BISYNC为例



数据帧的透明传输

- 如果数据负载中也包含STX或ETX字符如何处理？



- 引入转义符DLE，将数据负载中所有的STX, ETX, DLE进行转义

发送方数据负载:

STX -> DLE STX

ETX -> DLE ETX

DLE -> DLE DLE

接收方数据负载:

DLE STX -> STX

DLE ETX -> ETX

DLE DLE -> DLE

差错检测

- 数据在传输过程中可能会产生比特差错
 - 数据传输过程中，受电磁干扰等，比特位发生反转
 - 传输错误的比特占传输比特总数的比率称为误码率(Bit Error Rate, BER)
- 差错检测的基本思想
 - 在数据帧中加入冗余信息来确定是否存在差错
 - 一个极端的例子：每份数据，传输两个相同的副本；接收时发现两者不相等，则认为传输数据有差错
- 当接收方检测到差错时，可以：
 1. 通知对方数据有差错，使其重传数据副本（重传机制）
 2. 通过加入的冗余信息，重新构造正确的数据（纠错码）

差错检测方法举例

- 奇偶校验：

- 一个包含 m 比特的数据帧，假设最多含有1比特差错
- 将数据帧中所有比特异或求和，所得比特结果附加于数据帧后面
- 接收方收到数据帧后，检查前 m 位的异或和是否与 $m+1$ 位相等，可检测1比特差错

- 互联网校验和 (checksum)：

- 一个包含 m 比特的数据帧
- 将数据帧以 k 比特为单位，进行异或求和，所得结果为 k 比特，附加于数据帧后面
- 接收方收到数据帧后，检验前 m 位以 k 比特位单位的异或和是否与最后 k 位相等，如果不相等，则存在传输差错

循环冗余校验 (Cyclic Redundancy Check, CRC)

- 差错检测算法设计目标
 - 使用最少的冗余比特检测最多的错误
- CRC: 本质上是Hash函数
 - 对于n比特数据D, k比特CRC码C, 计算校验和S如下:
 - $S = (D \ll k) \% C$ // 求模计算基于有限域
 - 发送方发送: 附加校验和的数据 $D' = (D \ll k) \oplus S$
 - 接收方校验: $D' \% C == 0$

$$\begin{aligned} D' \% C &= (D \ll k) \oplus S) \% C \\ &= ((D \ll k) \% C) \oplus (S \% C) \\ &= S \oplus S \\ &= 0 \end{aligned}$$

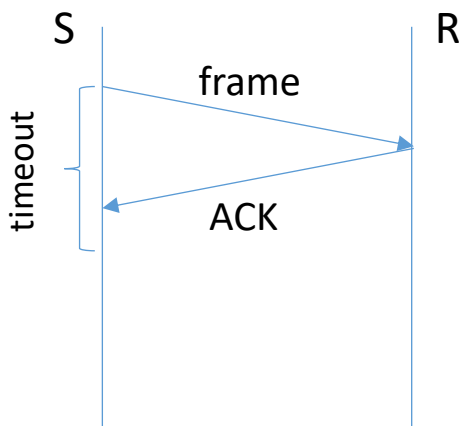
- CRC码选择
 - 经过特殊挑选, 保证较低的碰撞概率
 - 位数越多, 碰撞概率越低; 目前主要用CRC32

可靠传输

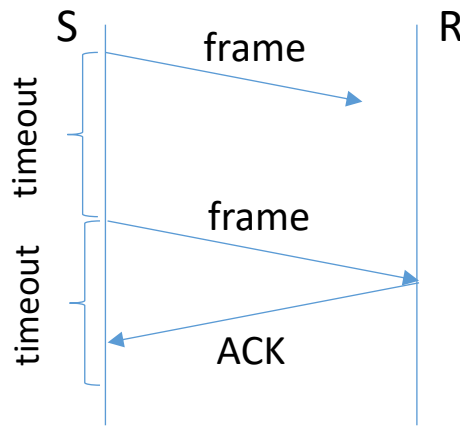
- 循环冗余校验等差错检测技术只能做到无差错接受
 - 凡是接受的帧（即没有丢弃），我们能以非常接近于1的概率认为这些帧在传输过程中没有产生差错
- CRC等检测出的错误帧被接收方丢弃
 - 有些纠错码技术可以恢复轻微的误码，但实际计算开销太大
 - 无线链路中的高误码率使得即使纠错码也不适用
- 可靠传输基本思想
 - 确认(acknowledgment, ACK)：接收方接受数据帧，回复ACK帧
 - 超时(timeout)：发送方在规定时间内没收到ACK，重传该数据

停等(stop-and-wait)协议

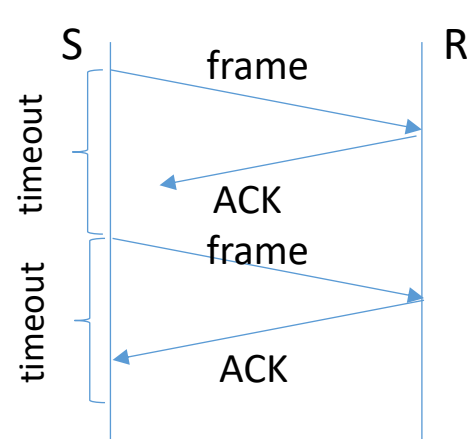
- 停等协议：最简单的可靠传输协议
 - 发送方每传输一个数据帧，在传输下一数据帧之前等待确认



(a) 正常传输



(b) 数据帧丢失

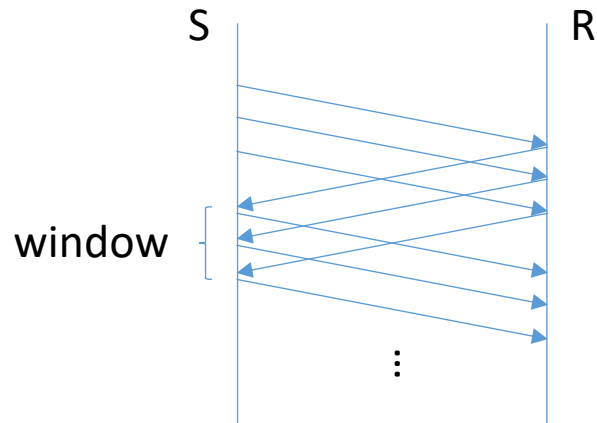


(c) ACK帧丢失

- 缺点：链路每次只能传输一个数据帧，带宽利用率低
 - 10Gb/s以太网链路，往返时延10us，帧大小1500B
 - 停等协议传输速率： $1500\text{Bytes}/10\mu\text{s} = 1.2\text{Gb/s}$

提升传输速率

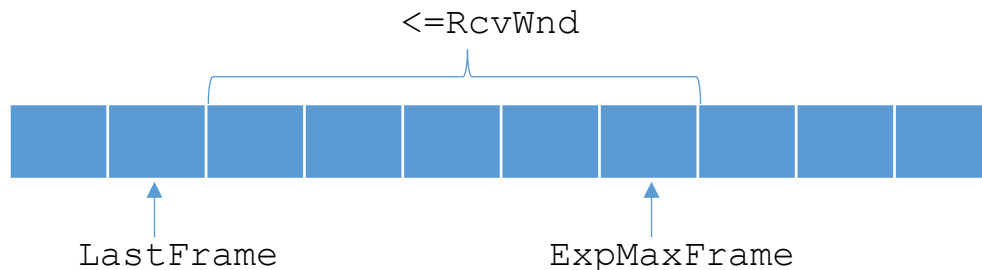
- 提升传输速率的根本方法是增加单位时间传输数据帧的数目
- 并发传输数据
 - 允许多个在途传输(未收到ACK)的数据帧
 - 通过窗口大小(window)限制在途传输的数据帧个数



- 每个数据帧赋予一个序列号 (Seq)，数据传输和确认都基于Seq

滑动窗口算法(sliding-window) – 接收端

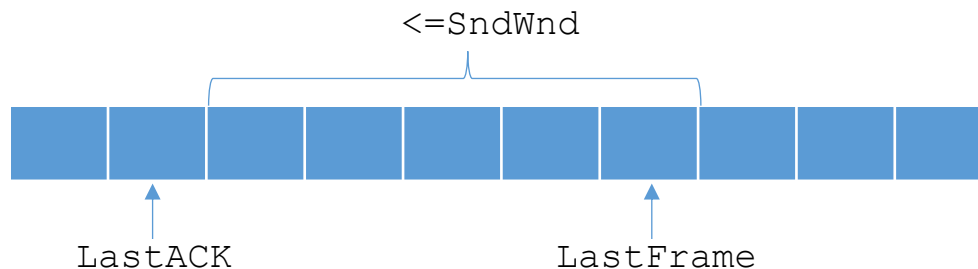
- 接收方维护三个状态变量：
 - 接收窗口大小：RcvWnd
 - 最后一个收到的数据帧：LastFrame
 - 期望收到的最大数据帧：ExpMaxFrame
- 数据接收过程中，保证 $\text{ExpMaxFrame} - \text{LastFrame} \leq \text{RcvWnd}$



- 对于每个新到达的数据帧Seq：
 - 如果 $\text{LastFrame} < \text{Seq} \leq \text{ExpMaxFrame}$ ，则接受；否则，丢弃
 - 接受数据帧后，将收到的最大连续数据帧Seq作为ACK回复

滑动窗口算法 – 发送端

- 发送方维护三个状态变量：
 - 发送窗口大小：SndWnd
 - 最后一个连续确认的ACK：LastACK
 - 最后一个发送的数据帧：LastFrame
- 传输数据过程中，保证： $\text{LastFrame} - \text{LastACK} \leq \text{SndWnd}$

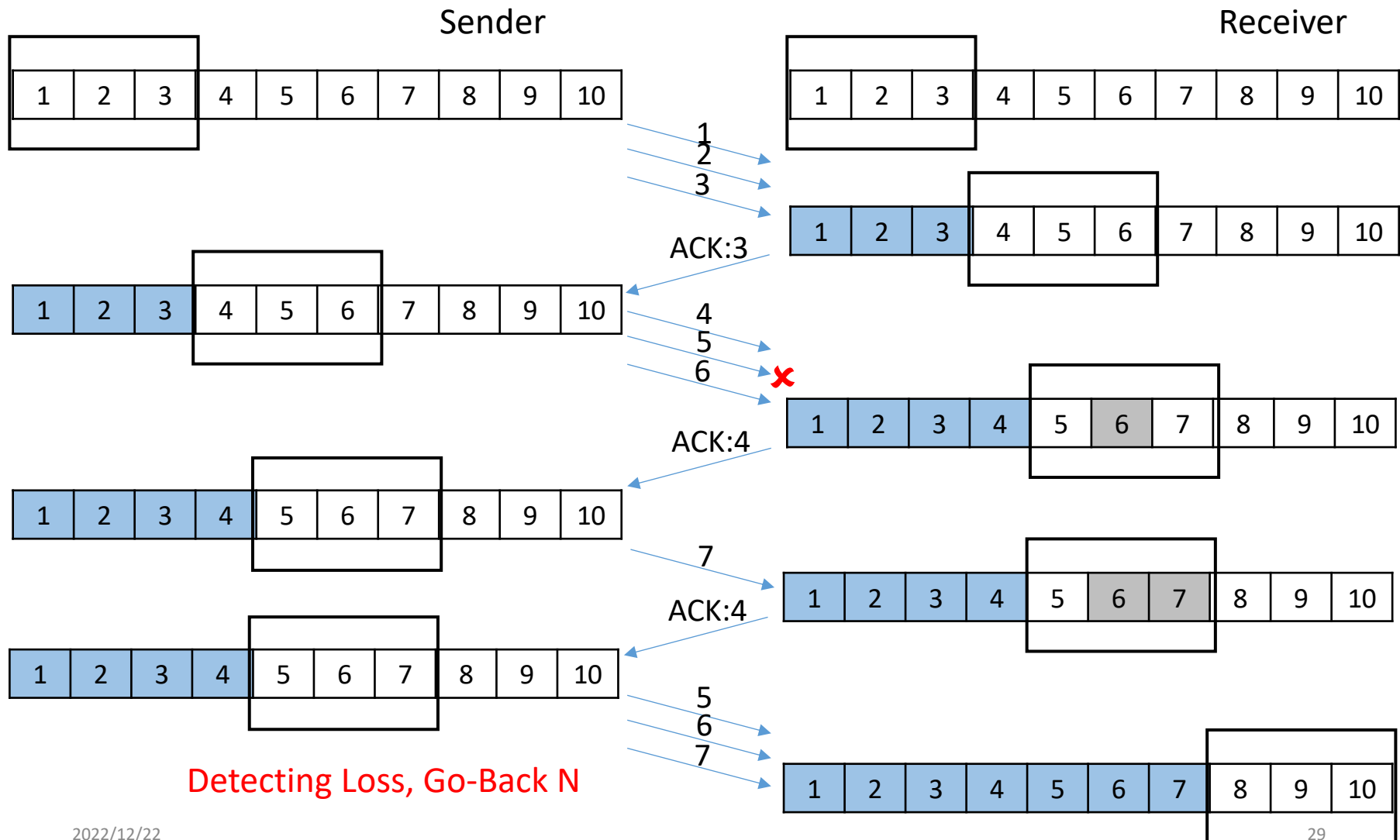


- 收到新的ACK：
 - 更新LastACK，如果窗口允许，发送新的数据帧，更新LastFrame

滑动窗口算法（续）

- 当数据帧丢失时，使用回退N机制(Go-Back-N)恢复丢包
 - 接收方只对连续收到的数据帧回复ACK
 - 例如，接收方收到数据2, 4, 5，只对数据帧2回复ACK
 - 由于接收不到新的ACK，发送方超时后重传LastACK+1与LastFrame之间的数据帧
- 滑动窗口其他细节
 - 受字段大小限制，Seq不能无限增大，因此Seq必须能够回绕
 - 只需要保证 $\text{MaxSeq}/2 \geq \text{Wnd}$ ，就可以准确区分已接收和等待确认的数据帧
 - 选择确认(Selective ACK)可为发送方提供更多信息
 - 接收方准确的确认每个已接受的数据帧，发送方根据这些信息更快重传
 - 传输效率更高，但实现更复杂

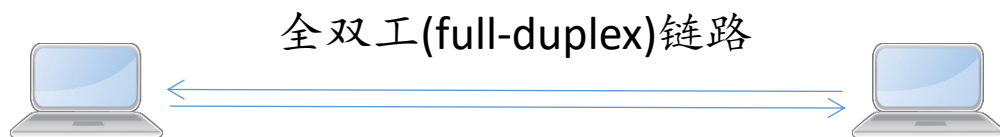
滑动窗口例子



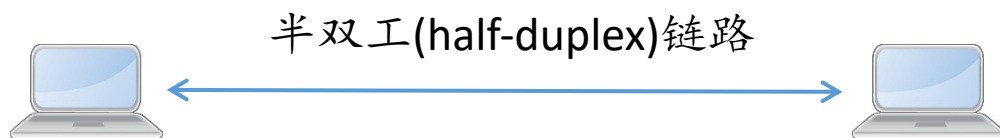
滑动窗口小结

- 滑动窗口是一种高效的可靠传输机制，包含4部分：
 1. 可靠传输
 - 数据确认和超时重传机制
 2. 高效传输
 - 通过并发提升传输性能
 3. 按序传送
 - 接收方将连续的数据交给上层，不连续的数据存放在buffer中
 4. 流控功能
 - 通信双方通过设定Window大小表达自己的发送/接收能力

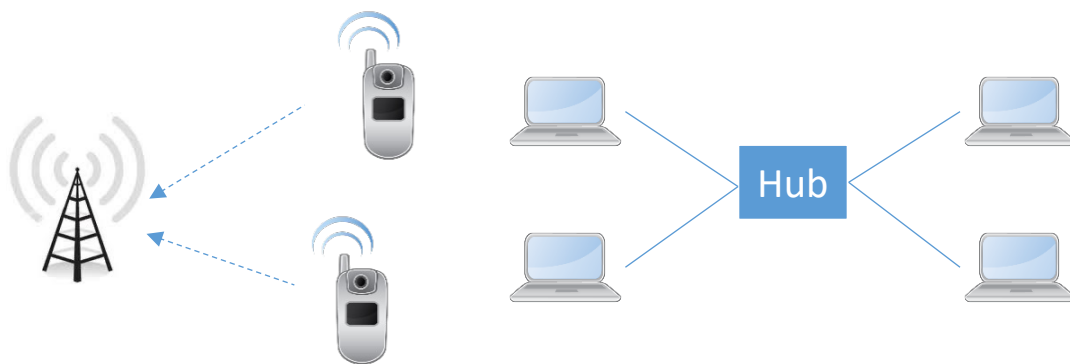
链路模型



节点可同时收发数据



节点不能同时收发数据



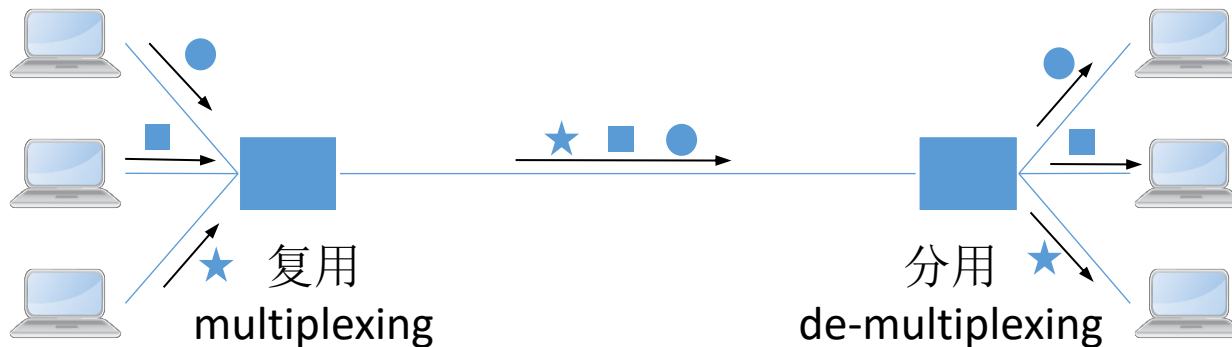
无线链路

星型链路

节点收发数据受其他节点影响

多路复用技术

- 复用(multiplexing)是通信网络技术中的基本概念

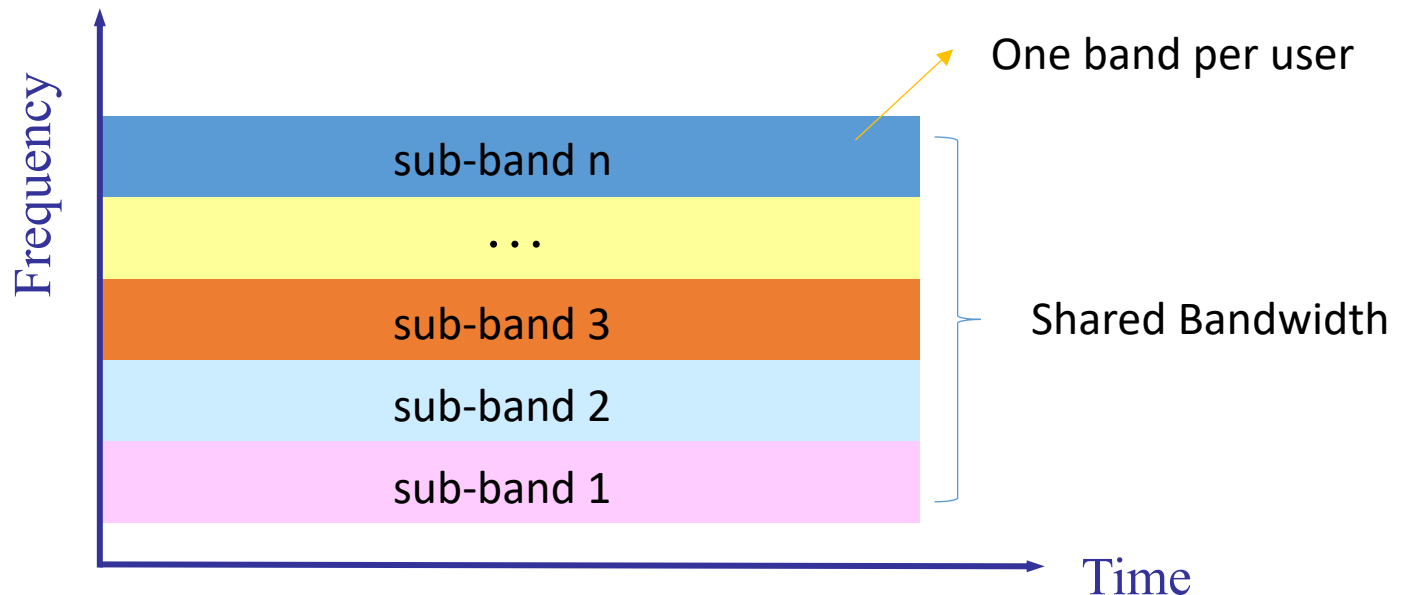


- 复用技术

- 频分复用(FDM)、时分复用(TDM)、统计时分复用(STATDM)、码分复用(CDM)、载波侦听多路访问(CSMA)

频分复用(Frequency Division Multiplexing, FDM)

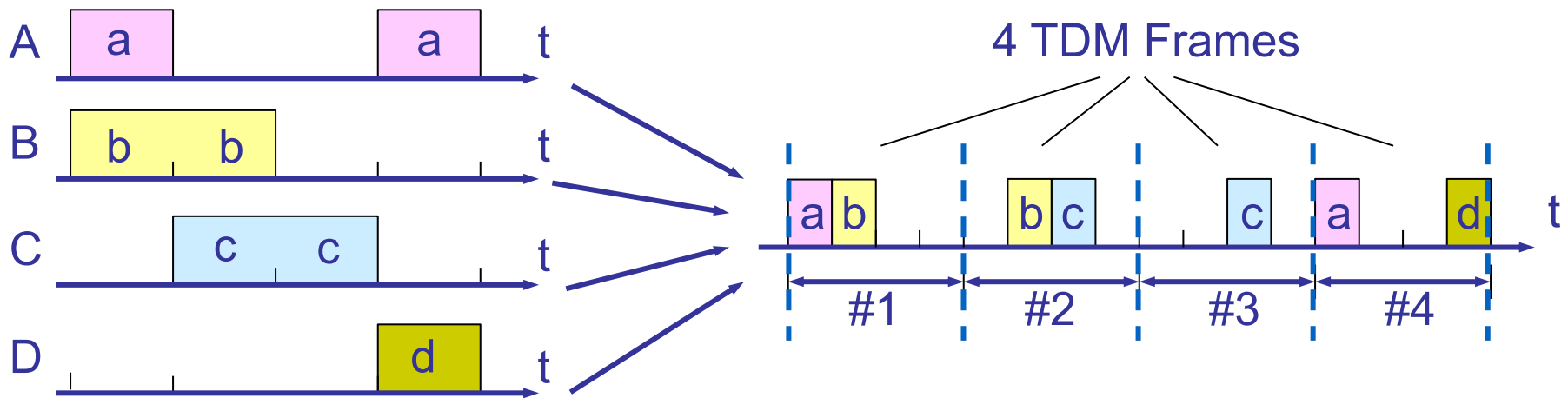
- 用户分配到一定的频带后，在通信过程中自始至终都占用这个频带
- 频分复用的所有用户在同样的时间占用各自的带宽资源



- 应用：
 - 早期的有线电视网络、光纤通信网络、模拟电话系统等

时分复用(Time Division Multiplexing, TDM)

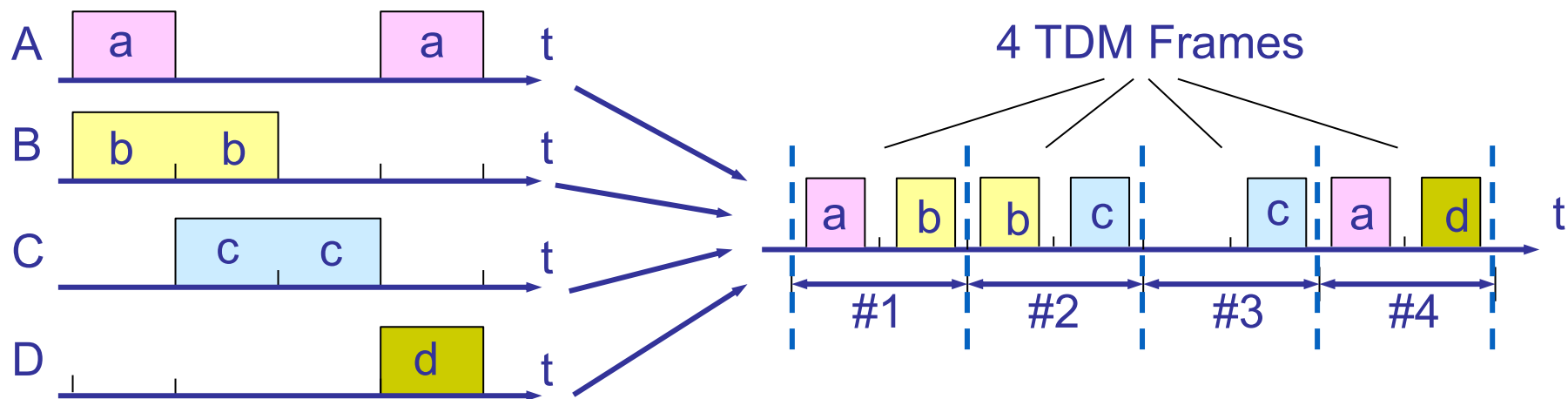
- 时分复用将时间划分为一段段等长的时分复用帧(TDM 帧)
- 每个用户在每一个TDM帧中占用固定序号的时隙



- 计算机网络中，由于数据传输具有突发性，TDM总带宽利用率不高

统计时分复用 (Statistic TDM, STDIM)

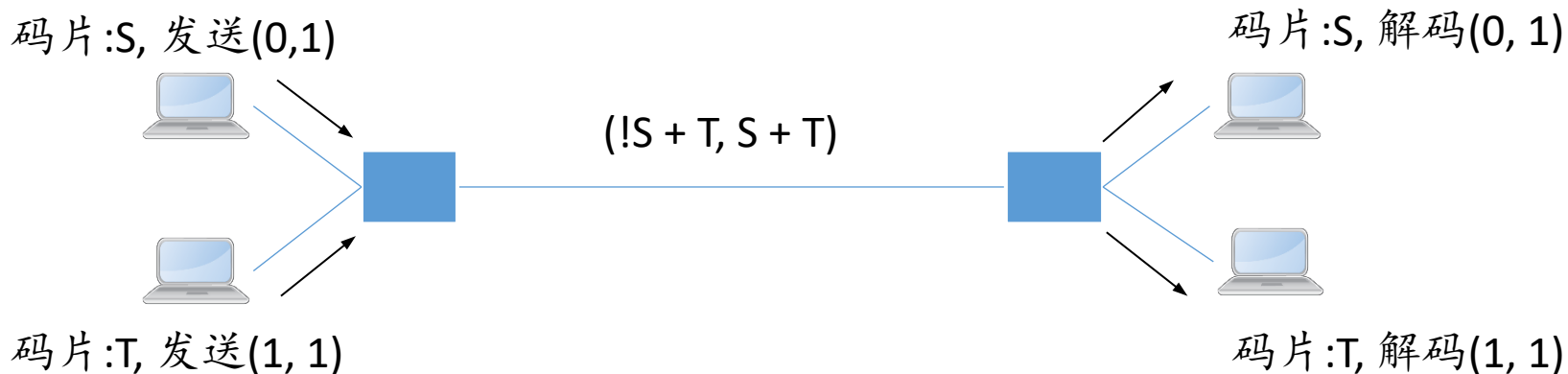
- 基于动态带宽分配的思想，STDIM将信号按需分成可变数目的子信道



- 相比于TDM，STDIM子信道浪费较少，因此有更高的带宽利用率

码分复用(Code Division Multiplexing, CDM)

- 每一个比特时间划分为 m 个短的间隔，称为码片(chip)
- 每个用户被指派一个唯一的 m bit 码片序列
 - 发送比特 1，则发送自己的 m bit 码片序列
 - 发送比特 0，则发送该码片序列的二进制反码
- 不同用户分配的码片序列是正交的(orthogonal)
 - 两个正交的码片序列 S, T : $p(S, S) = 1, p(S, T) = 0, p(S, !S) = 0$



载波帧听多路访问 (Carrier Sense Multiple Access, CSMA)

- 核心思想：先侦听，后发送(sense before transmit)
- 根据具体的侦听/发送策略，CSMA可以分为如下3类：

		发送策略	
		可发送时1概率发送	可发送时p概率发送
侦听策略	忙时q概率侦听	非持续CSMA	X
	忙时1概率侦听	1-持续CSMA	p-持续CSMA

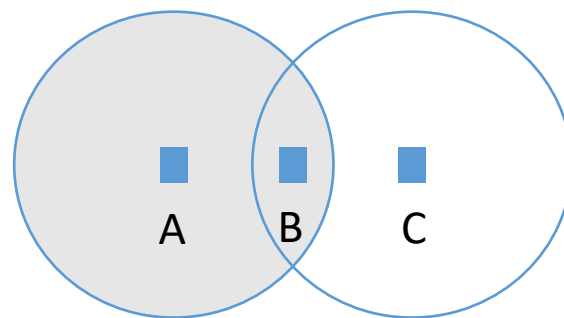
- 非持续CSMA：可减少碰撞，会导致信道利用率降低，较长的延迟
- 1-持续CSMA：会导致较多的碰撞，导致性能降低
- p-持续CSMA：通过调节p在减少碰撞和高信道利用率之间取得平衡

带碰撞检测(Collision Detection)的CSMA (CSMA/CD)

- 带碰撞检测的载波帧听多路访问
 - 核心思想：1-持续CSMA + 碰撞检测
 - 侦听/发送：
 1. 发送数据时，侦听链路中是否有其他数据在传输。
 2. 如果有数据传输，则继续侦听等待。
 3. 如果没有检测到，则传输数据。
 4. 在数据传输过程中，如果遇到碰撞，执行碰撞处理操作。
 - 碰撞处理：
 1. 停止发送数据，发送特殊阻塞信息通告其他设备有碰撞发生。
 2. 根据发生碰撞的次数，计算**指数退避时间**(exp backoff)，并在时间内随机等待。
- CSMA/CD用于Ethernet（半双工链路和星型链路）中

带碰撞避免(Collision Avoidance)的CSMA (CSMA/CA)

- 无线网络中存在隐藏节点问题(hidden node problem), CSMA/CD难以准确检测碰撞，因此不再适用



- 带碰撞避免的载波帧听多路访问
 - 核心思想：非持续CSMA + 碰撞避免
 - 侦听/发送：
 1. 发送数据时，先侦听信道是否空闲。如果信道忙，随机等待后再次侦听。
 2. 如果信道空闲，执行碰撞避免操作，发送数据。
 - 碰撞避免：
 1. 发送方向接收方发送Request to Send帧，请求对方开始接受数据。
 2. 接收方回复Clear to Send帧，表示现在可以接收数据。
- CSMA/CA用于无线局域网络(例如WiFi)

多路复用机制对比

非争用多路复用	争用性多路复用
FDM, TDM, CDM	CSMA/CD, CSMA/CA
使用控制器/仲裁器	不使用控制器/仲裁器
使用已分配的固定带宽	使用的带宽是变动的
强调公平性	强调自组织和带宽利用率
带宽利用率较低	带宽利用率较高
2G(GPRS, Edge), 3G(CDMA2000, WCDMA), 4G+	Ethernet, WiFi
多用于传统语音、视频系统, 以及移动蜂窝网络	用于互联网数据传输

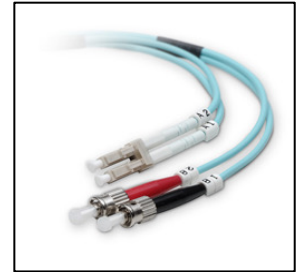
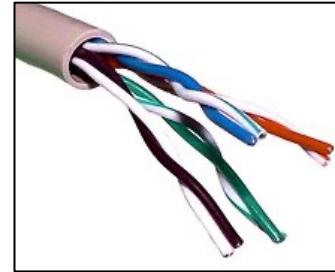
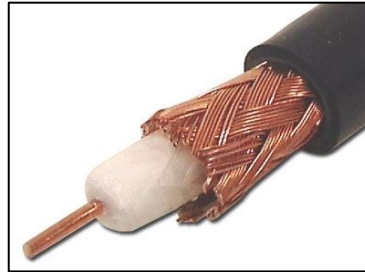
以太网(Ethernet)简介

- 1970年代中期，诞生于Xerox PARC实验室
- 是一种局域网技术，服从802.3标准
- 带宽范围：10Mb/s ~ 10+Gb/s
- 工作距离：2.5Km(双绞线: 100m, 同轴电缆: 500m)以内
- 组网方式：直连型、总线型、星型
- 链路共享机制：CSMA/CD

以太网网络基本组成

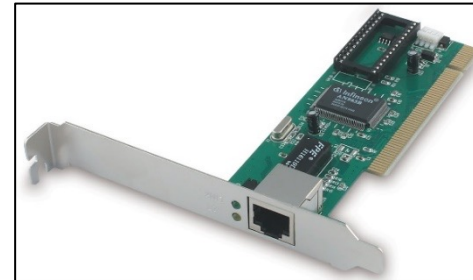
- 电缆

- 同轴电缆 (Coaxial Cable) ,
- 双绞线 (Twisted Pair) ,
- 光纤 (Optical Fiber)



- 网络适配器

- 网卡 (NIC)



- 中继设备

- 集线器 (Hub)

- 交换设备

- 网桥 (Bridge) , 交换机 (Switch)

- 路由设备

- 路由器 (Router)



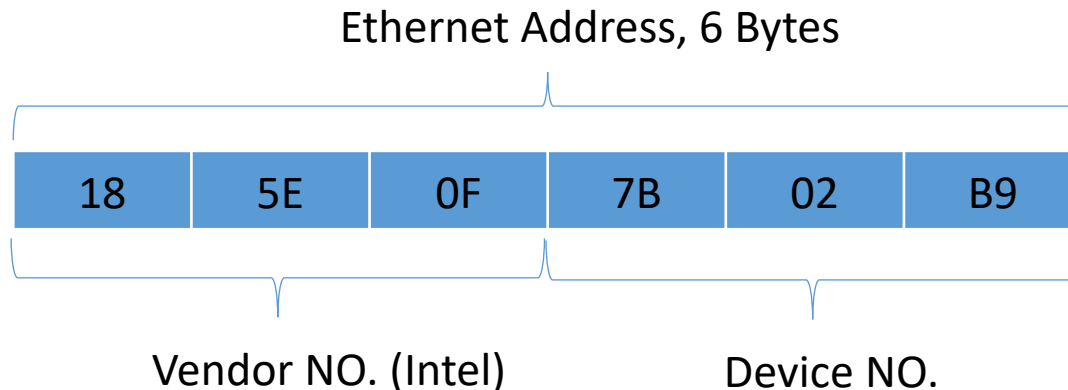
以太网规格

- 以太网技术规格由802.3文档规定，由3部分组成
- 链路速率 (10, 100, 1000, 10G)
 - 10Mbps, 100Mbps, 1000Mbps, 10Gbps
- 信号制式 (base, broad)
 - 基带数字信号
 - 宽带模拟信号 (几乎不再使用)
- 电缆类型 (2, 5, T, F, ...)
 - 2 – 同轴细缆 (200m)
 - 5 – 同轴粗缆 (500m)
 - T – 双绞线
 - TX – 5类以上双绞线
 - F – 光纤

以太网规格举例：
100BASE-T, 1000BASE-TX

以太网地址

- 为了标识以太网上的每台主机，需要给每台主机上的网络适配器分配一个唯一的通信地址
 - 以太网地址或称为网卡的物理地址、**MAC地址**
- IEEE为每个网络设备制造厂商分配一个唯一的厂商代码，各厂商为自己生产的每块网络适配器分配一个唯一的设备编码
 - 每块网卡的Ethernet地址就是这两者的结合，所以该地址是唯一的



以太网帧格式

8 Bytes	6 Bytes	6 Bytes	2 Bytes	46-1500 Bytes	4 Bytes
Preamble	Dst Addr	Src Addr	Type	Payload	FCS

- Preamble – 前导码，用来标识Ethernet帧的开始 ($0x55 * 7 + 0xD5$)
- DstAddr, SrcAddr – 目的以太网地址和源以太网地址
- Type – 所封装的三层协议类型 (IP: 0x0800)
- Payload – 所封装的三层协议数据，有效长度范围是46~1500字节（如果不足46字节，则用0补齐）
- FCS – 帧校验序列，包含32比特的循环冗余校验值，用来校验损坏的帧

以太网评价

- 以太网基本上统治了有线局域网
- 以太网的优点
 - 便宜，高速，易于使用、管理和扩展
- 以太网的缺点
 - 数据包越小，传输单位数据的代价越大
 - 用户数增加时，更容易发生碰撞
 - 链路变长时，需要更长时间来检测拥塞
 - 网络负载越大，传输性能越低

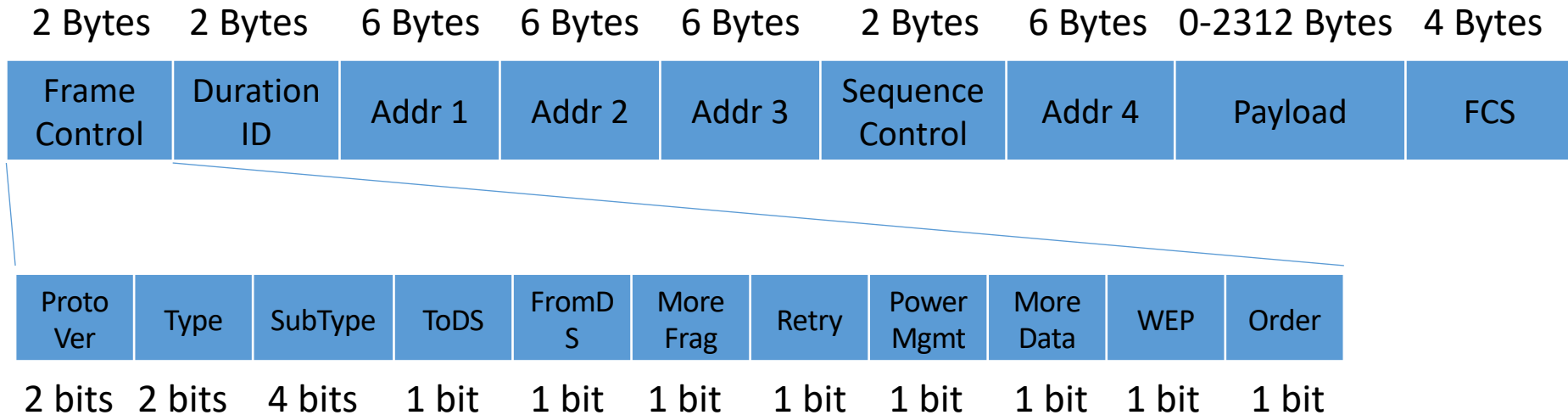
WiFi简介

- 1990年代，由Vic Hayes带领IEEE 802.11工作组制定标准
- WiFi (Wireless Fidelity) \approx Wireless LAN
- 工作频率：2.4GHz, 5GHz
- 带宽范围：54Mb/s \sim 600+Mb/s
- 工作距离： $< 50\text{m}$ (室内)
- 组网方式：自组织网络、面向接入点的网络
- 链路共享机制：CSMA/CA

WiFi标准修订(举例)

- IEEE 802.11a OFDM 5GHz
- IEEE 802.11b DSSS 2.4GHz
- IEEE 802.11e QoS
- IEEE 802.11g OFDM 2.4GHz
- IEEE 802.11k Radio Resource Measurement
- IEEE 802.11n MIMO
- IEEE 802.11ac
- IEEE 802.11ax

WiFi数据帧格式



- Frame Control标识数据帧类型和其他行为
 - Type + SubType标识帧类型；ToDS和FromDS决定4个地址的含义
- 4个地址是MAC地址格式，含义由ToDS和FromDS标志位决定
 - 当ToDS和FromDS都为0时，Addr 1表示目的地址，Addr 2表示源地址
- FCS为帧校验序列，由32位CRC校验码构成

接入WiFi热点

- 当节点移动到一个WiFi热点的覆盖范围时，通过如下步骤接入该热点：

1. 扫描：节点搜寻一个可用的信道

- A. 节点发送一个Probe帧
- B. 所有可达的热点发送Probe Response帧来应答

2. 关联：节点选择一个热点进行关联

1. 节点选择一个热点，并向其发送一个Association Request帧
2. 对应热点发送Associate Response帧来应答

3. IP地址分配：通过DHCP方式获得IP地址

4. 认证：如果热点配置认证选项，则进行认证

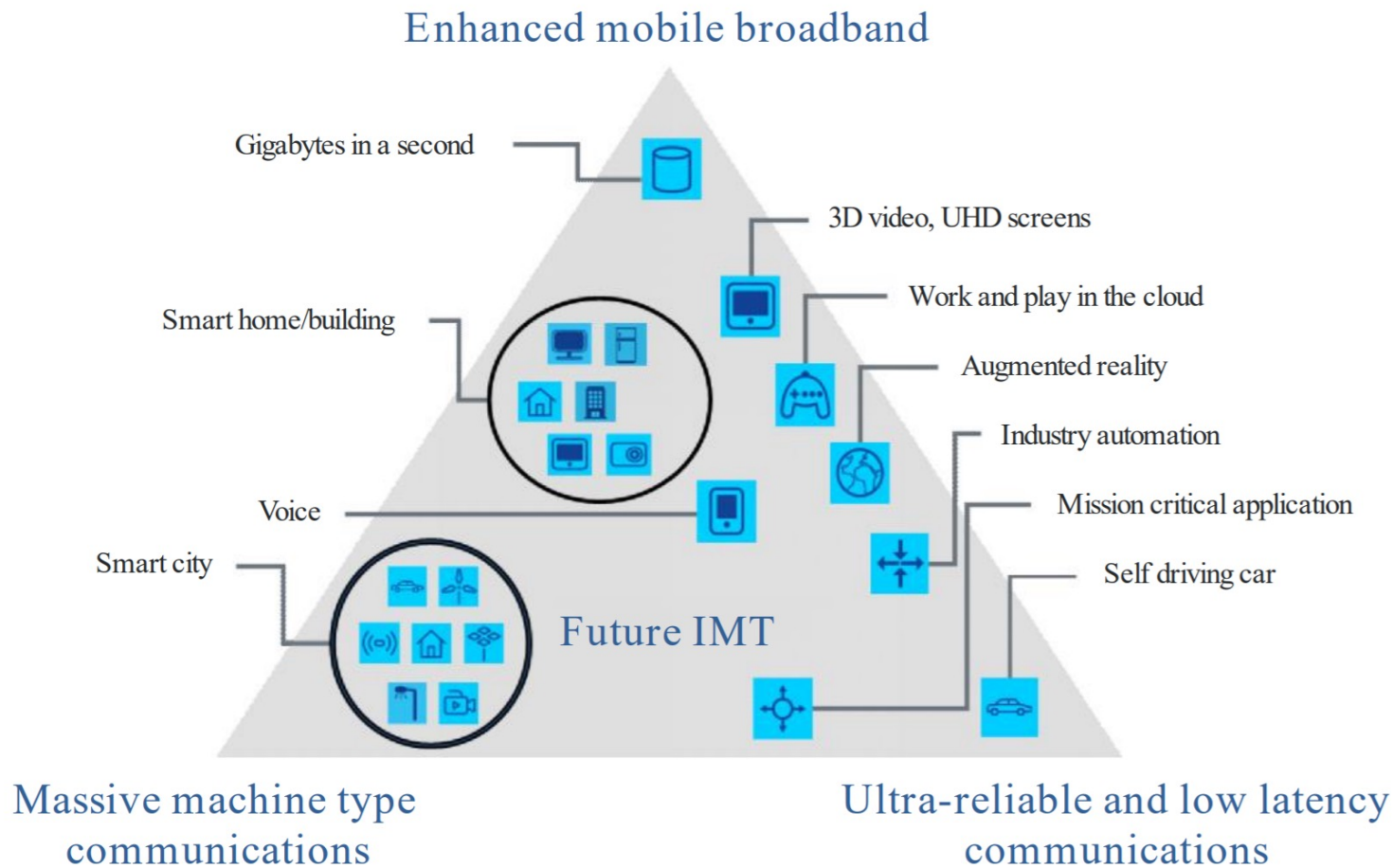
WiFi评价

- WiFi已经广泛应用于家庭网络、办公网络、公共场所等
 - 是最成功的无线局域网技术
- WiFi优点：
 - 不需要使用电缆
 - 可以稍微移动位置
- WiFi缺点：
 - 不同热点间信号相互干扰造成传输性能急剧下降
 - 信号质量随着距离增大显著下降
 - 不同热点间切换造成上层连接终止

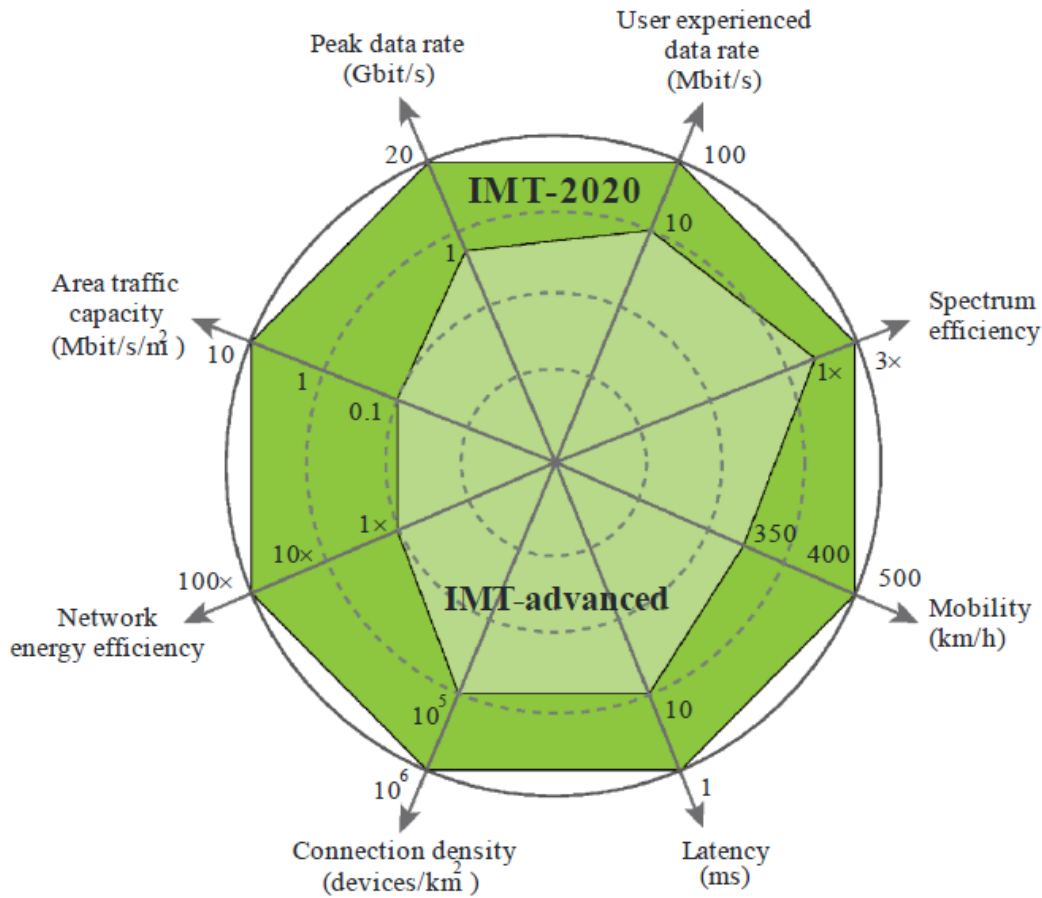
蜂窝通信技术简介

- 1G 蜂窝通信是为语音通信设计的模拟FDM系统，几乎不支持数据传输
- 2G 蜂窝通信提供低速数字通信（短信服务），代表制式为GSM
 - 2.5G 技术是从 2G 向 3G 过渡的扩展技术，如 GPRS 和 EDGE 等，支持低速数据传输
- 3G 蜂窝通信使用混合的交换机制，可支持移动宽带多媒体业务，主要制式为WCDMA和CDMA2000
- 4G 蜂窝通信网络完全基于IP技术，语音通话也基于IP数据，主要制式为TD-LTE和FDD-LTE
- 5G通信网络完全基于IP技术，使用毫米波、大规模MIMO、高密微基站大幅提升网络性能

5G应用场景



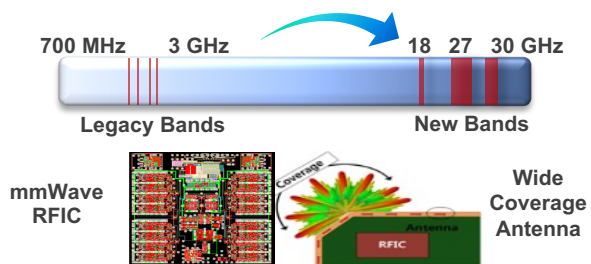
5G性能需求



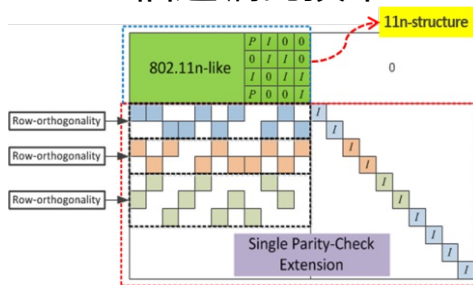
- 10x bandwidth per connection
- 100x bandwidth/area
- 10 year battery life
- >10x connections
- Low-ms latency
- Five 9's reliability
- 100% coverage
- 50Mbps per connection

5G核心技术

毫米波技术

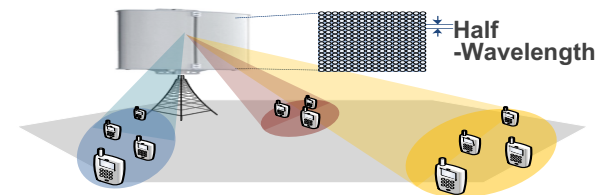


信道编码技术

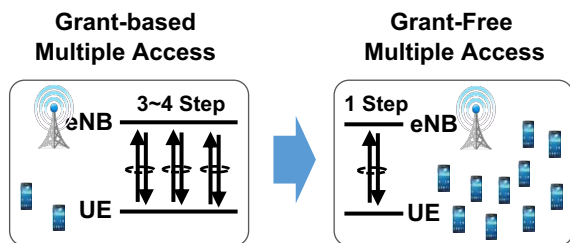


LDPC (Low-Density Parity-Check)

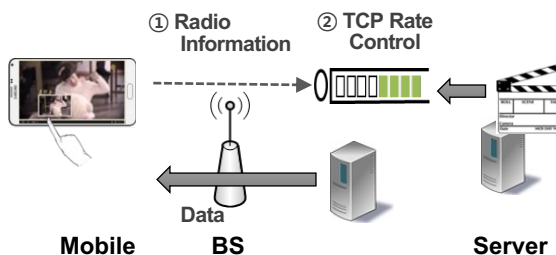
大规模MIMO



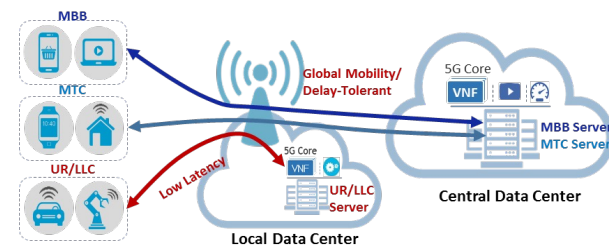
海量连接



低延迟技术



网络切片

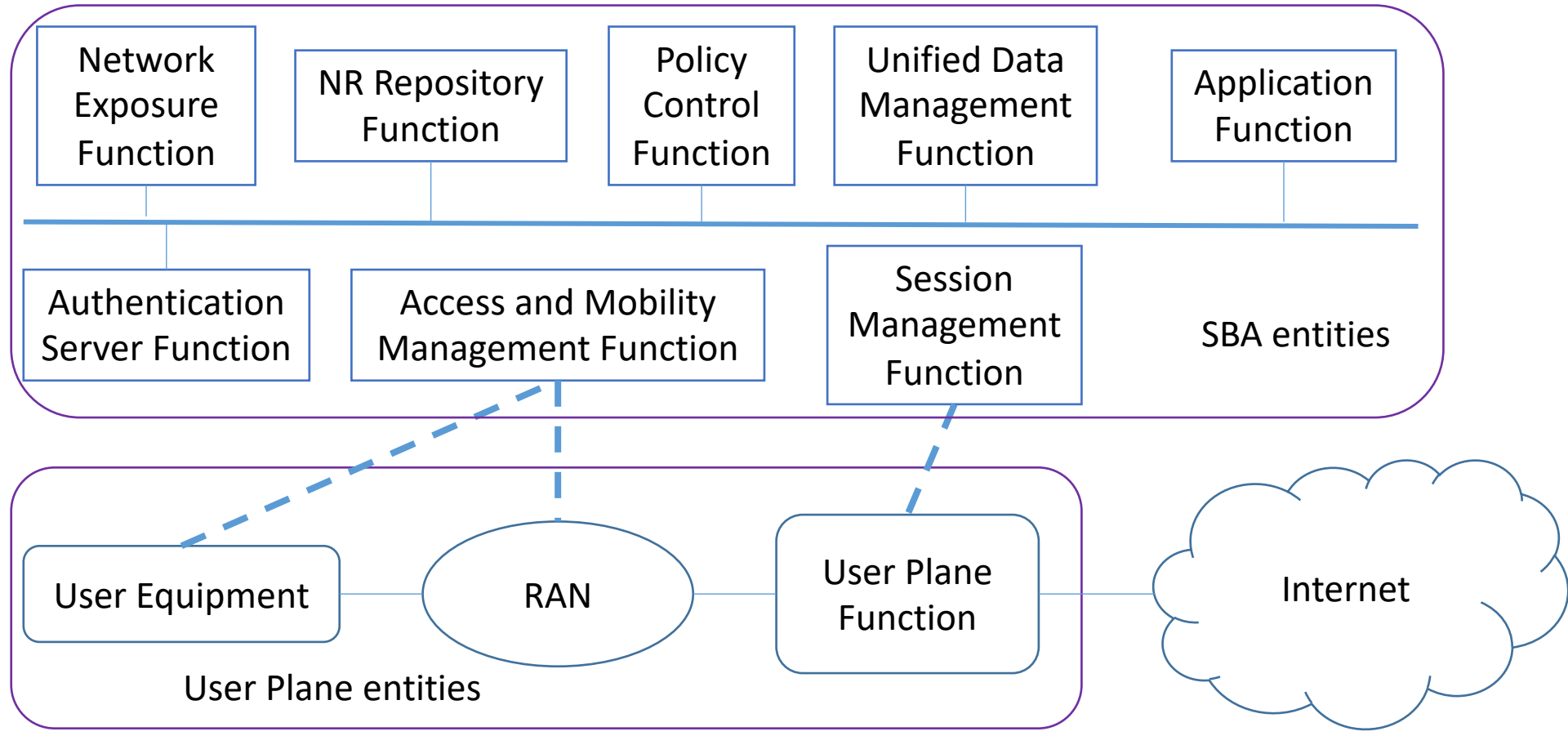


5G核心网架构

- 基于服务的体系结构
- 网络切片
- 控制平面与用户平面的分离

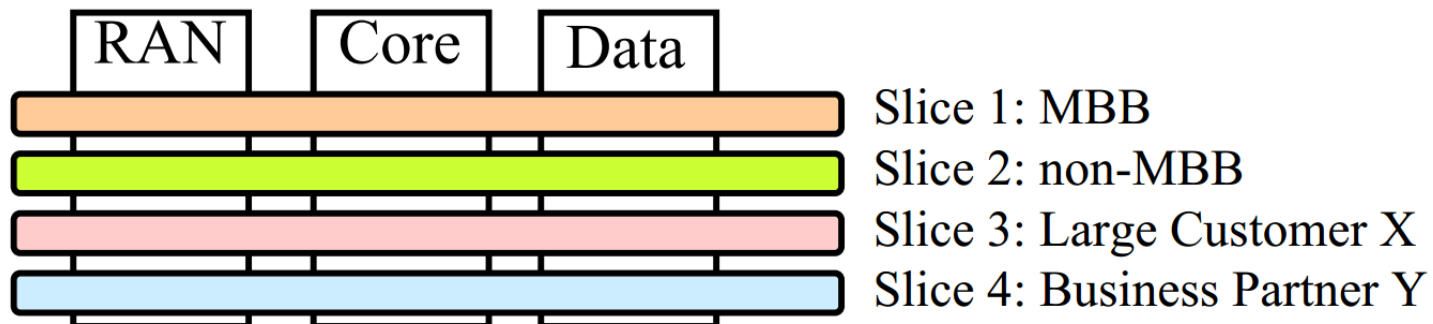
基于服务的体系结构

- 每个服务(Service)是一个功能(Function)，功能可以在物理节点上实现，也可以在虚拟节点上实现



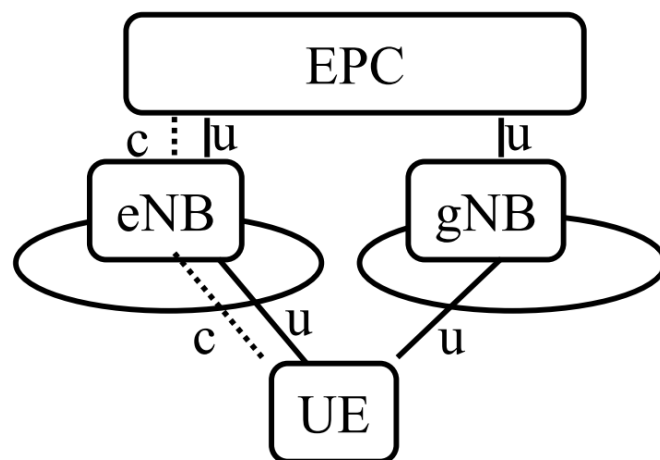
网络切片 (Network Slicing)

- 网络切片对应一个逻辑上服务于某一特定用户或应用的网络
- 一个网络可以划分成多个网络切片
- 对于用户来说，其网络切片是一个有隔离资源的独立的网络



控制平面与用户平面的分离

- 控制平面的功能：会话管理、IP地址分配、移动管理、认证、安全等
- 用户平面的功能：数据包路由与转发、数据包过滤、数据包检测、服务质量等
- 控制平面和用户平面的接口完全独立
 - 例如，在4G/5G混合部署下，控制平面数据使用4G，用户平面数据使用4G和5G



蜂窝通信网络评价

- 蜂窝通信网络是覆盖范围最广的通信机制之一
- 蜂窝网络优点：
 - 接入范围广
 - 很好的支持移动性
 - 资源分配更公平
- 蜂窝网络缺点：
 - 费用高
 - 网络性能相对较差
 - 当用户接入数增加时，网络性能急剧下降

课后阅读

- 《计算机网络 —— 系统方法》第1、2章
- 计算机网络系统设计准则
 - Jerome Saltzer et al., **End-to-end arguments in system design**, ACM Transactions on Computer Systems (TOCS), 1984, 2(4): 277-288.
- 互联网细腰模型及演化
 - Saamer Akhshabi et al., **The Evolution of Layered Protocol Stacks Leads to an Hourglass-Shaped Architecture**, ACM SIGCOMM 2011
 - Lucian Popa et al., **HTTP as the Narrow Waist of the Future Internet**, ACM HotNets 2010
- 5G
 - Mansoor Shafi et al., **5G: A Tutorial Overview of Standards, Trials, Challenges, Deployment, and Practice**, IEEE Journal on Selected Areas in Communications, 2017.

Any
Questions?

谢谢！