

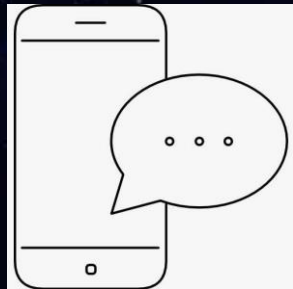


第一讲 计算机网络概述

(课程介绍)



“网”罗天下



Metaverse

数据与算力传输、共享、协作；消除信息的时空距离

网络：基础性的支撑技术



AI



Cloud Computing



AR/VR



Big Data



IoT



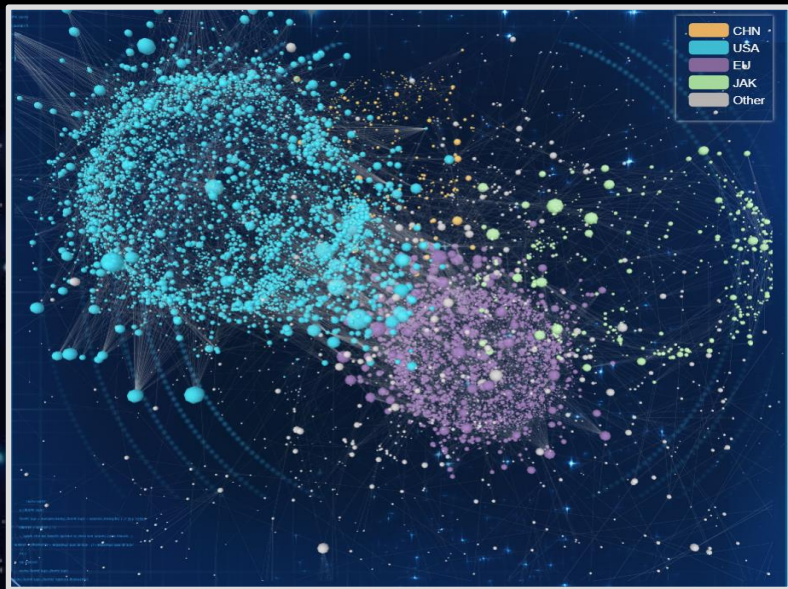
Blockchain

互联网不仅仅改变生活与社会，还推动信息技术本身发展

The future cannot be predicted, but futures can be invented.

— Dennis Gabor, Inventing the Future, 1963

网络巨大成功，为自身带来新挑战



人工构建的复杂大工程系统



Skin-integrated wireless haptic interfaces for virtual and augmented reality, Nature, Nov. 20, 2019

扩展性、兼容性、QoS、安全性；技术、工程、商业、治理

主讲



谢高岗, 博士, 研究员

方向: 网络体系结构, 分布计算系统

Email: xie@cnic.cn



孙毅, 博士, 研究员

方向: 区块链、内容分发、QoS

Email: sunyi@ict.ac.cn



李振宇, 博士, 研究员

方向: 网络体系结构、测量分析

Email: zyli@ict.ac.cn



武庆华, 博士, 副研究员

方向: 网络体系结构、路由、传输

Email: wuqinghua@ict.ac.cn



助教

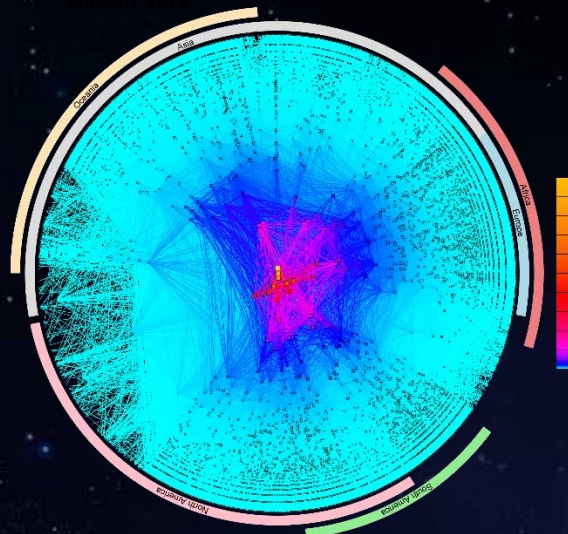
刘慧, 博士

方向: 计算机网络, 网络安全

Email: hliu@ucas.ac.cn

提纲

CAIDA'S IPv4 AS CORE GRAPH
JANUARY 2020

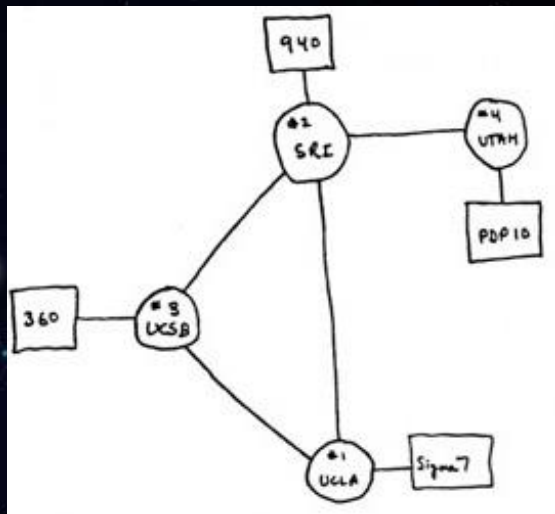


COPYRIGHT © 2020 UC REGENTS

CAIDA's IPv4 AS Core Graph, 2020

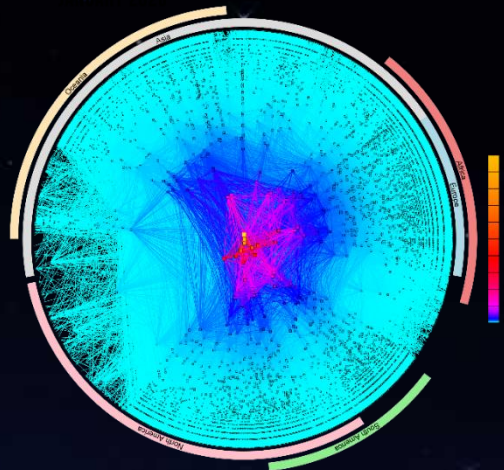
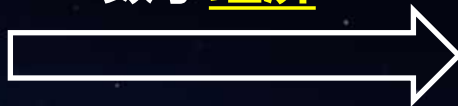
- 计算机网络(互联网)概述
- 中科院互联网前期工作
- 课程介绍

这个时代最伟大的设计之一，人类发展里程碑之一



1969

互联网时代
数字经济

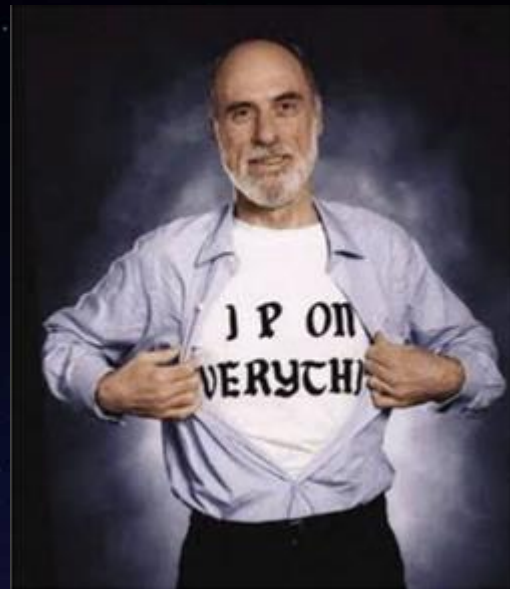


CAIDA's IPv4 AS Core Graph, 2020

协同发展：科学与工程；学术与产业；竞争与协作；技术、经济与治理

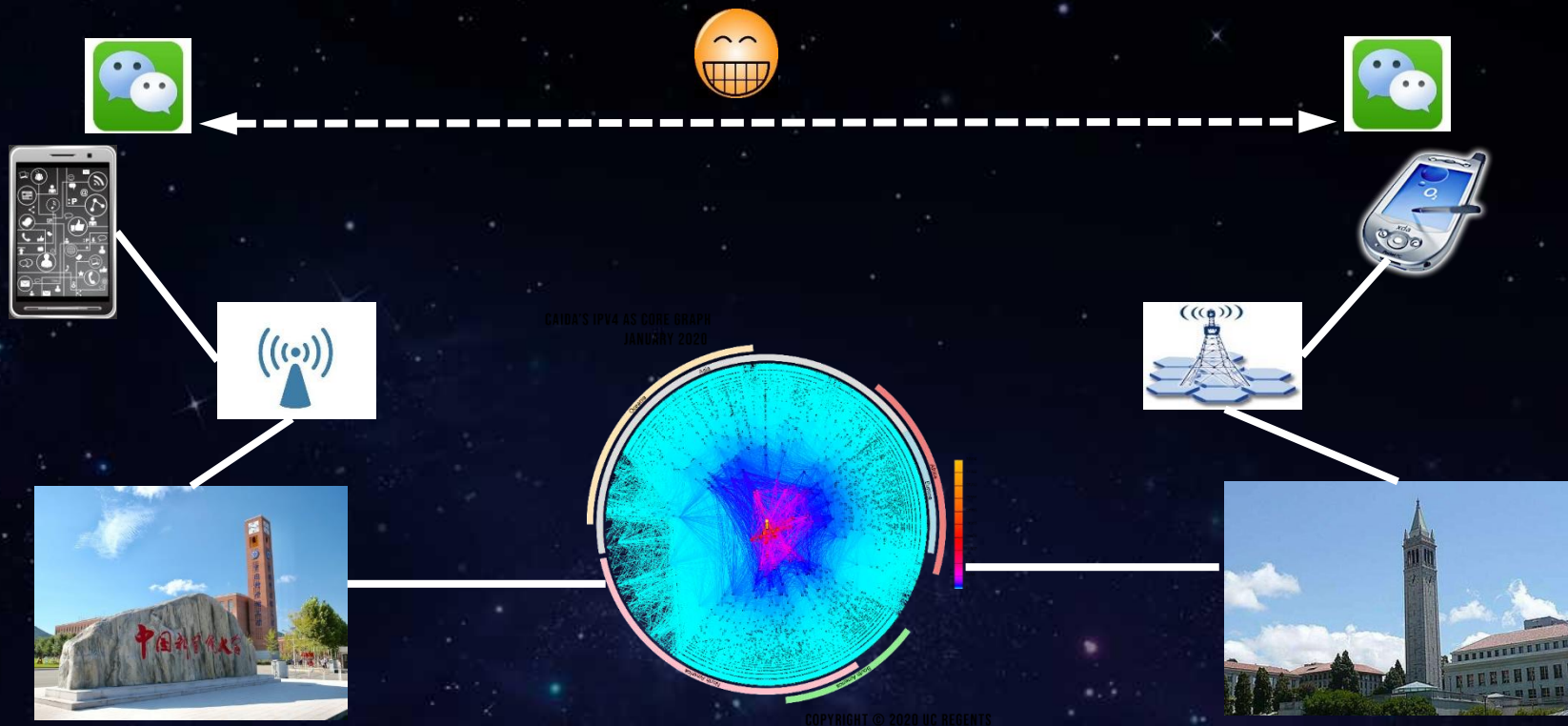
Power of the Internet Architecture, by Vint Cerf

- Not Designed for any Specific Applications: Just move packets
- Designed to run over any communication technology
- Permission innovation at the edges
- Design to Scale
- Open to new protocols, new technologies, new applications
- **IETF: above the wire, below the application**
 - Wire以下: 被互联的网络
 - application及以上: 交给应用
 - Area: Internet Area (int)、Routing Area (rtg)、Transport Area (tsv)、Applications and Real-Time Area (art)、Operations and Management Area (ops)、Security Area (sec)

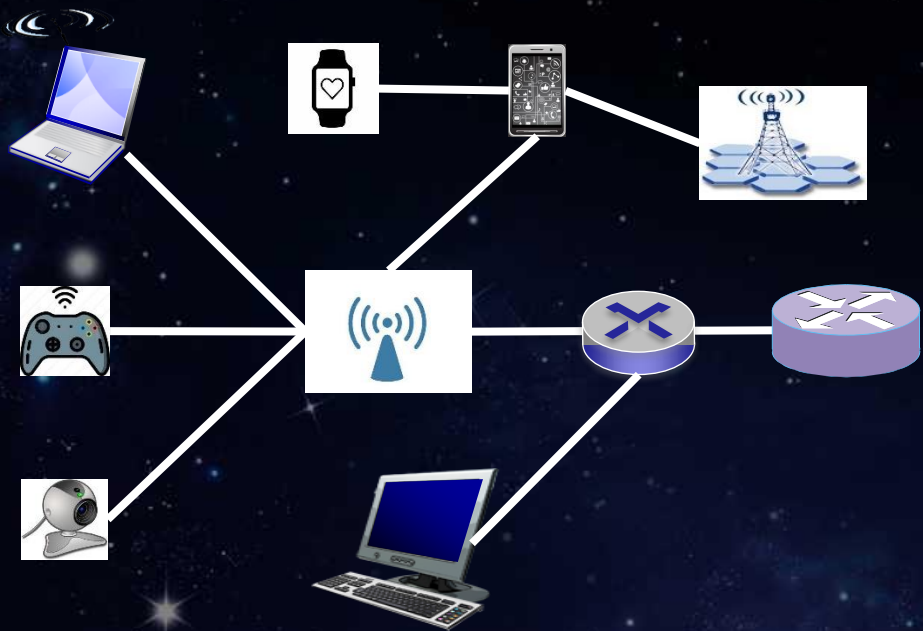




如何实现连接：不同规模范围的寻址(MAC、IP、Port)



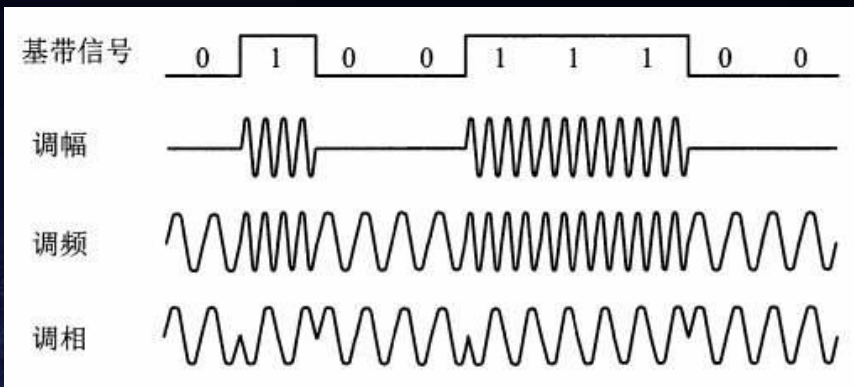
物理传输：通信



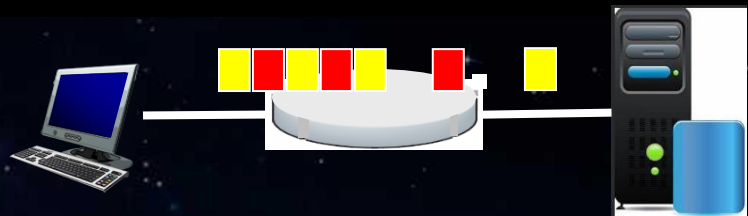
无线(蓝牙、Wifi、5G)、有线(光、电)

- 基础对象：bit
- 调制、编码、信道、同步
- 香农定理 $C = B \log_2(1 + \frac{S}{N})$

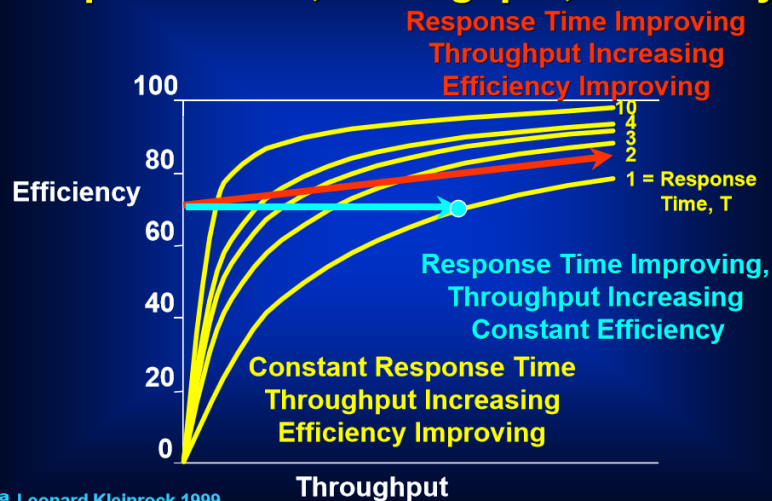
C: 传输速率; B: 信道带宽; S/N为信噪比



数据传输：帧传输（包交换）



Key Tradeoff: Response Time, Throughput, Efficiency



^a Leonard Kleinrock 1999

- 基础对象：数据帧
- 媒介访问控制
- 存储转发

— 传输延迟： $T_{\text{tras}} = 2P/R$

P: Packet Size (bits)

R: bandwidth (bps)

— 排队延迟：延迟抖动

— 缓存溢出：丢包

网络互联

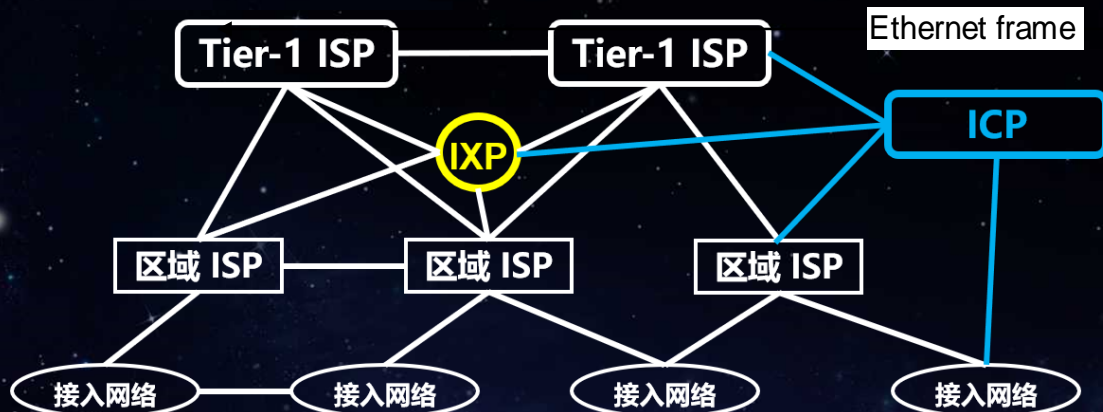
Ethernet Header

IP Header

TCP Header

Application data

Ethernet Trailer



- ISP: Internet Service Provider
- ICP: Internet Content Provider
- IXP: Internet eXchange Point

- 异构网络，不同物理寻址机制
 - 构建统一的标识空间：IP格式数据包定义
 - 兼容不同的物理寻址（链路）
 - 大规模可扩展：路由；IPv6

- 1971, Louis Pouzin, CYCLADES, Datagram
- 1983, ARPANET将NCP切换到IP; UNIX BSD

4.2实现TCP/IP

- 1986年, NSFNET TCP/IP



路由与路由查找

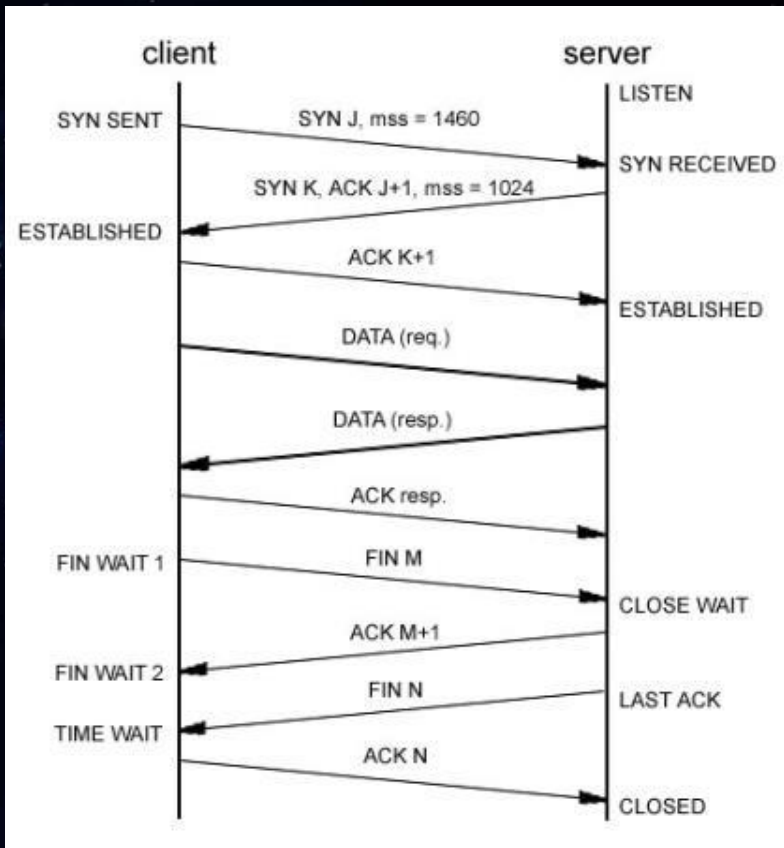
- 路由协议（控制平面）
 - 域内路由：OSPF、RIP、IS-IS
 - 域间路由：BGP
 - 无类域间路由CIDR（Classless InterDomain Routing）
- 路由查找转发（数据平面）



FIB

prefix	next-hop
* / 0	6
1 * / 1	4
01 * / 2	3
001 * / 3	3
111 * / 3	7
0011 * / 4	1
1110 * / 4	8
11100 * / 5	2
001011 * / 6	9

可靠性与传输控制



● 如何在尽力而为的不可靠路径上实现可靠数据传输？

— 无差错、不丢失、不重复、顺序

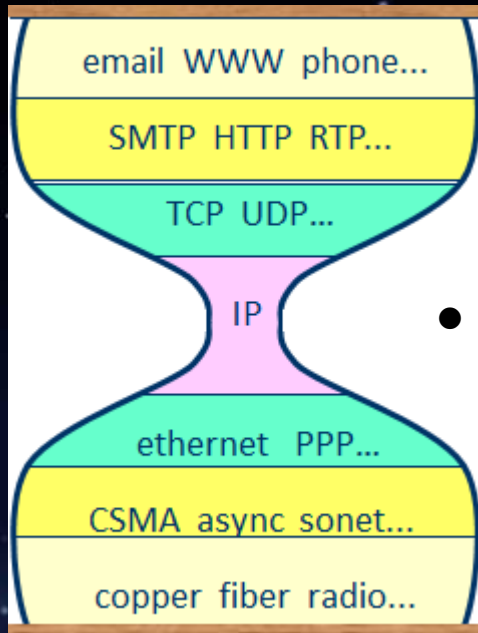
● 传输控制

— 传输控制：发送端与接收端

— 拥塞控制：发送端与网络

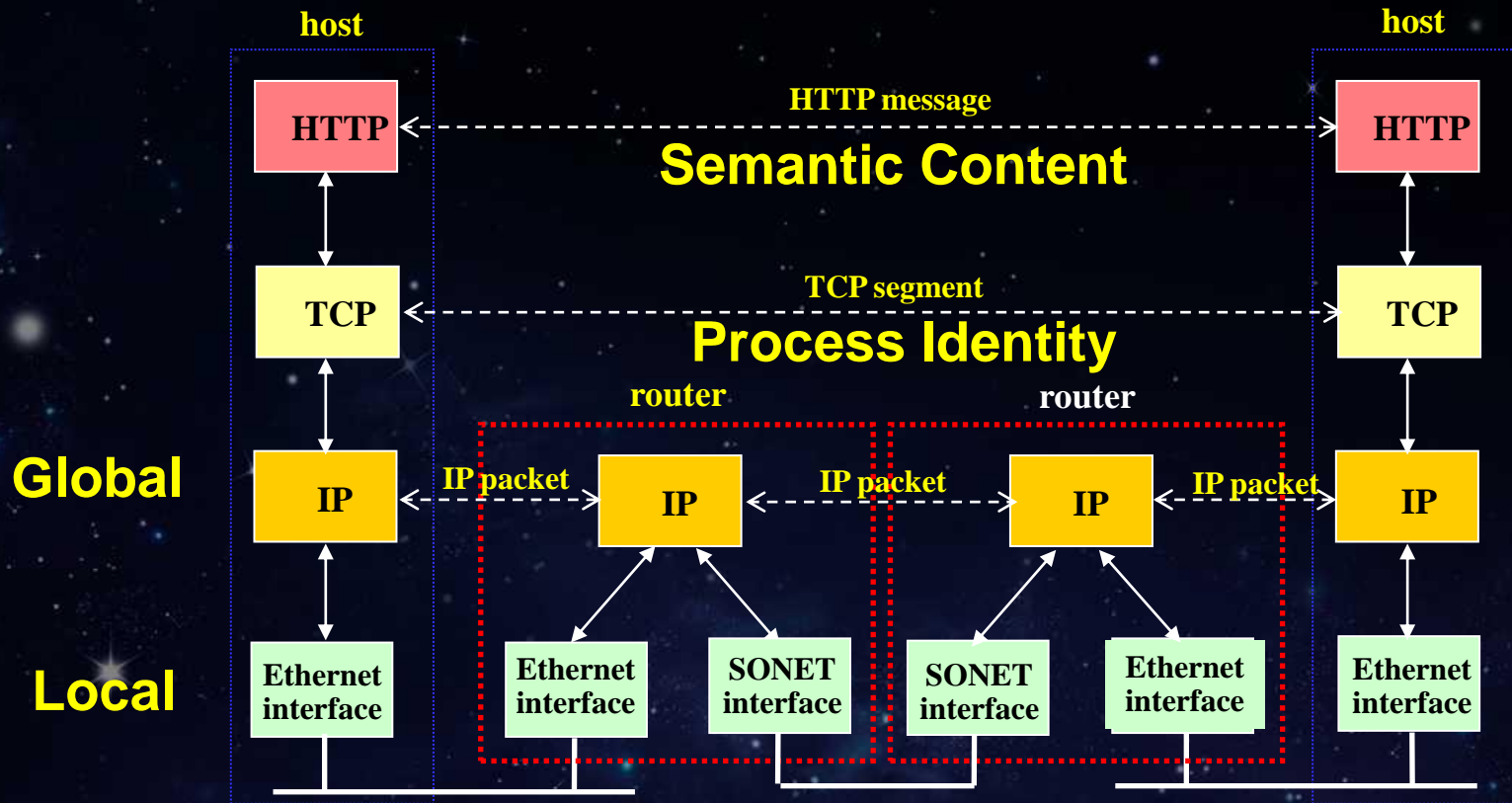
层次结构与实现模块化

- 模块化是复杂系统通用的设计原则，例如软件工程
- 模块化导致协议层次化
- 层次化是互联网创新的基础条件
 - 对其他层次影响与并行创新
- 为什么网络层创新少? IPv6 例子
- 模块实现位置
 - 路由器: L1~L3; 主机: L1~L5

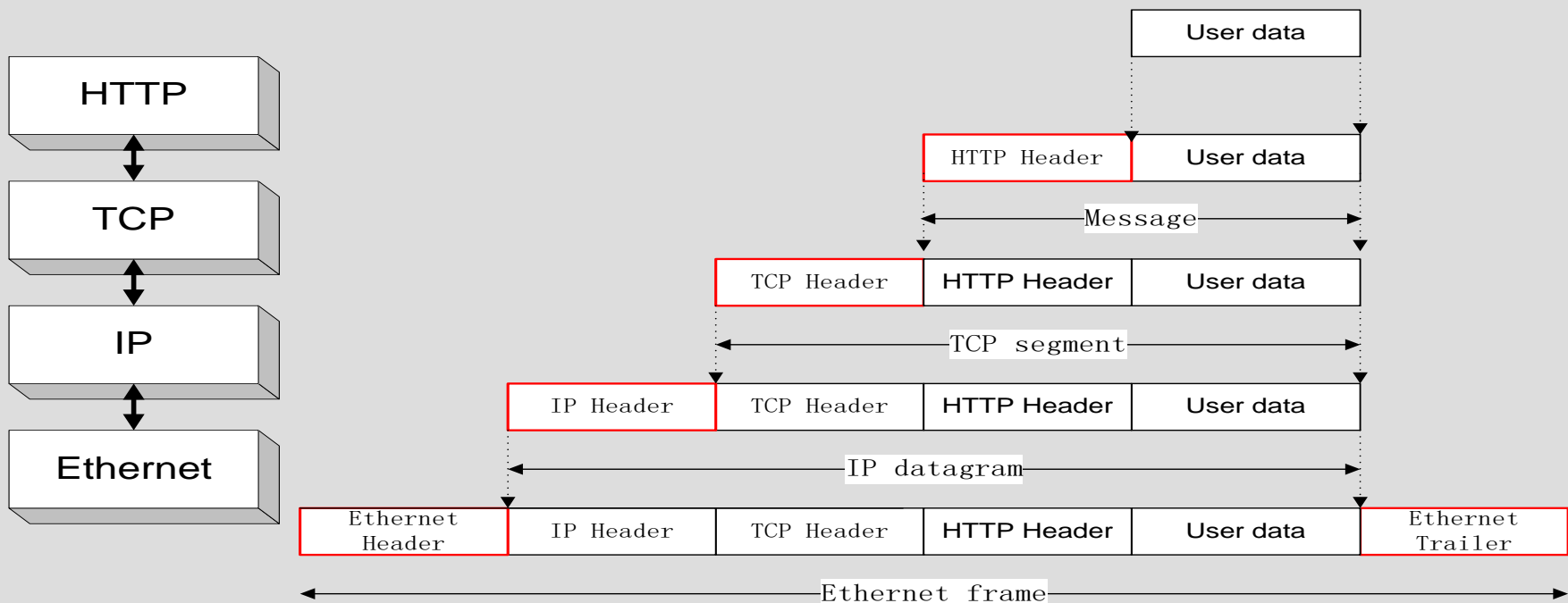


- 应用层: Message
- 传输层: Segment, 可靠传输数据流的逻辑连接
- 网络层: Datagram/packet, 数据报文全局网络不可靠传输
- 数据链路层: Frame, 本地网络数据帧逐跳转发
- 物理层: Bit

不同协议层次的视角



报文封装



例子：IP数据包

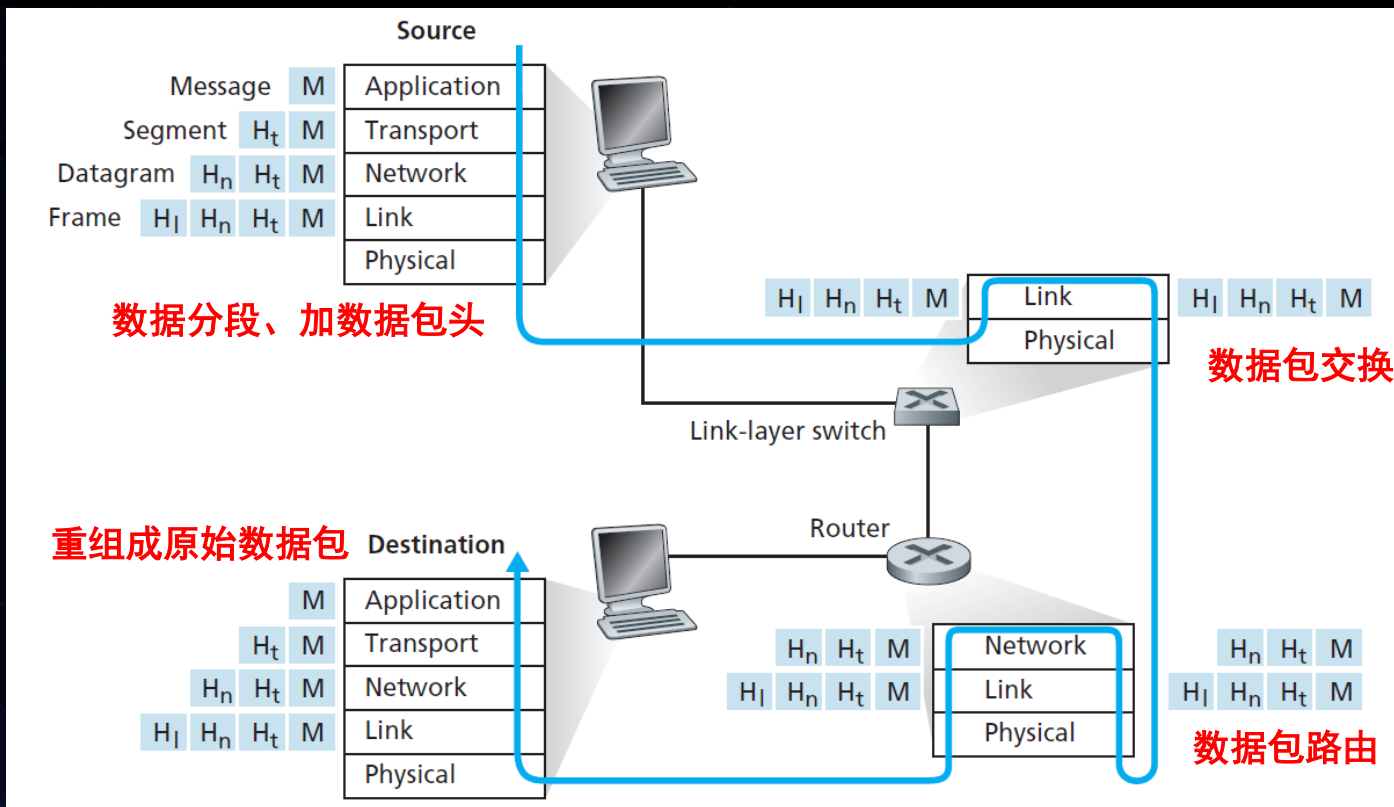
32 bits

version (4 bits)	header length	DS	ECN	Total Length (in bytes) (16 bits)	
Identification (16 bits)				flags (3 bits)	Fragment Offset (13 bits)
TTL Time-to-Live (8 bits)		Protocol (8 bits)		Header Checksum (16 bits)	
Source IP address (32 bits)					
Destination IP address (32 bits)					



Ethernet frame

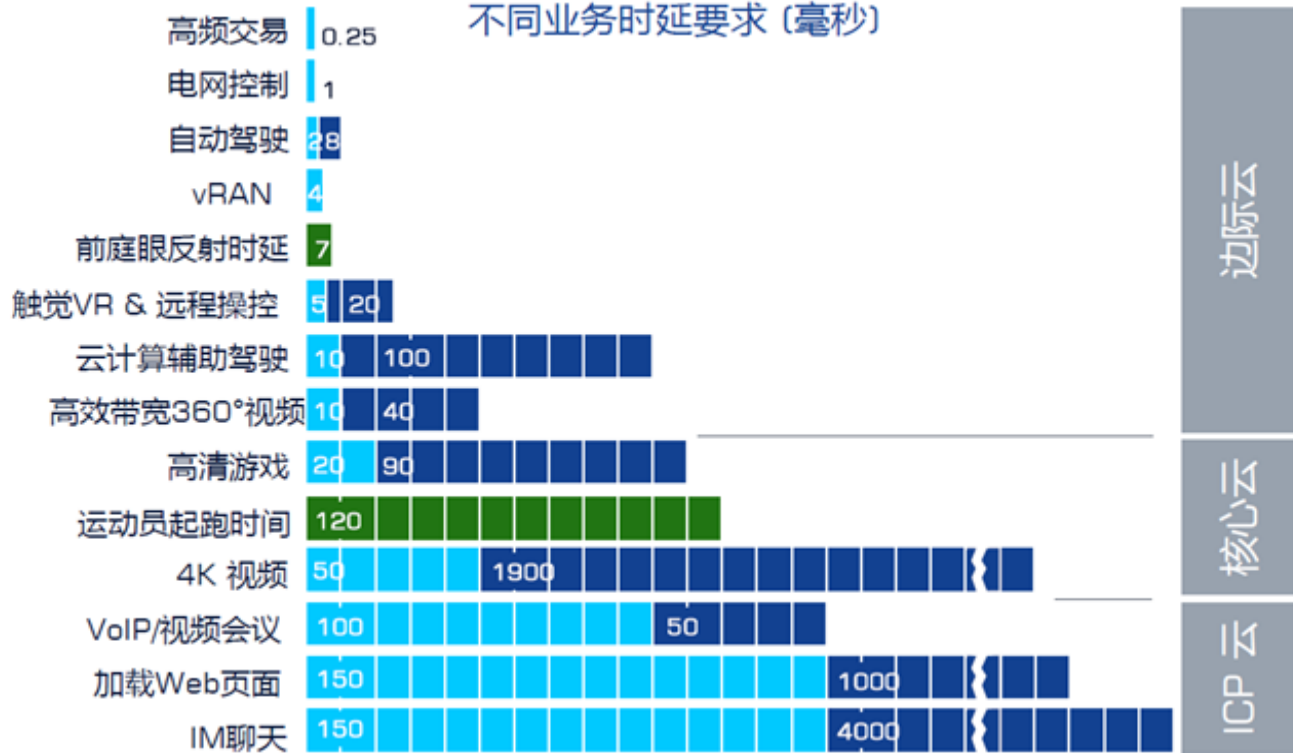
报文传输路径



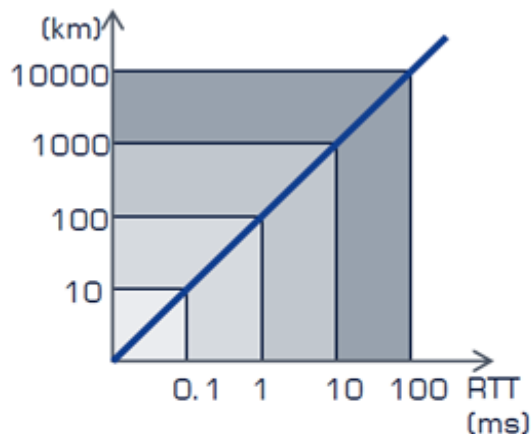
新需求：延迟



不同业务时延要求 (毫秒)



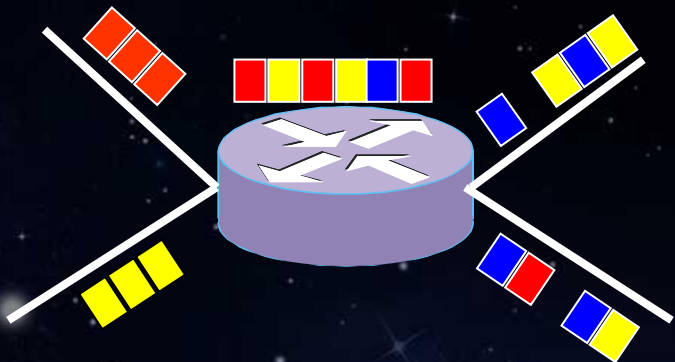
在光纤的光传播



- 最大允许网络传送时延
- 最大应用层处理和缓存时延
- 人类反应时间(参考)

From Nokia

延迟构成



- $T = T_{\text{tras}} + T_{\text{proc}} + T_{\text{prop}} + T_{\text{queue}}$

- $T_{\text{tras}} = 2P/R$: 传输延迟

- T_{proc} : 查找、处理

- T_{prop} : 信号传播延迟

- 卫星网络 vs. 光纤传输?

- T_{queue} : 排队延迟


- 堪萨斯-纽约证券交易所

- 光纤, 17ms → 16ms

- 微波, 11ms



新需求：吞吐量

	~Kbps 2G	~100kbps 3G	~10Mbps 4G	~Gbps 5G	~Tbps B5G
					
					
					
					
					
文本					
图像					
音频					
视频					
VR/AR					
全息					

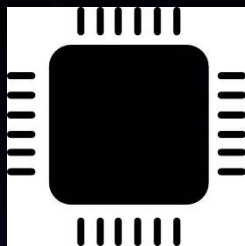
4K/8K: 35~140Mbps; VR/AR: 25Mbps~5Gbps; 全息: ~4Tbps

丢包; 吞吐量; 延迟

驱动性能提升的技术因素



通信
香农定理



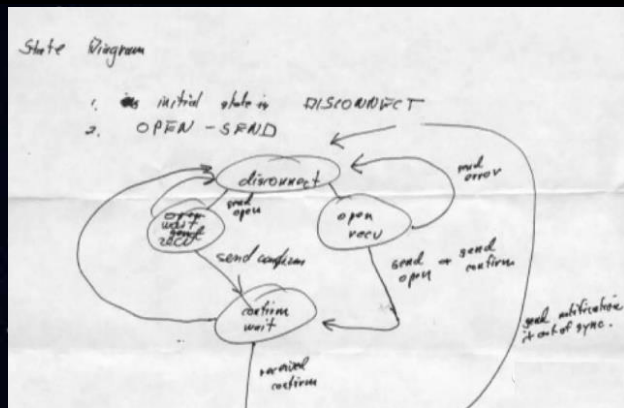
微电子
摩尔定律



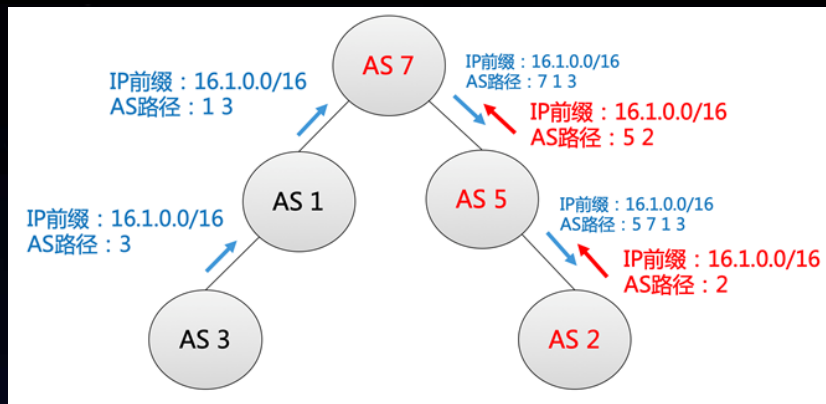
算法

体系结构与组网技术创新?

新需求：网络安全



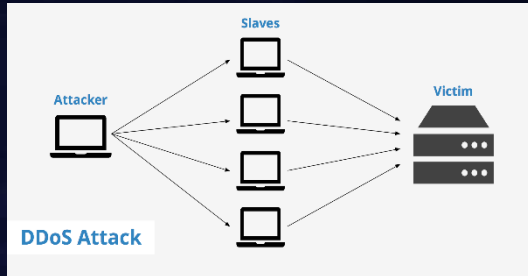
1989, 3张餐巾, BGP



路由劫持



信息安全



网络攻击



安全漏洞

网络发展方向?



From <https://singularityhub.com/>

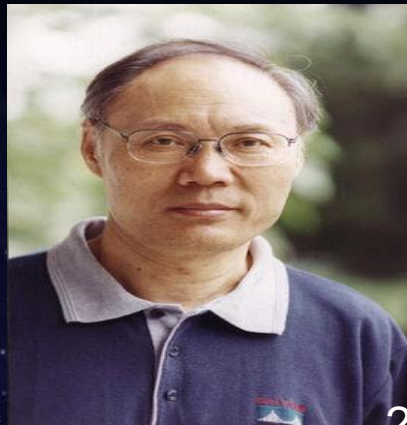
人机物三元融合

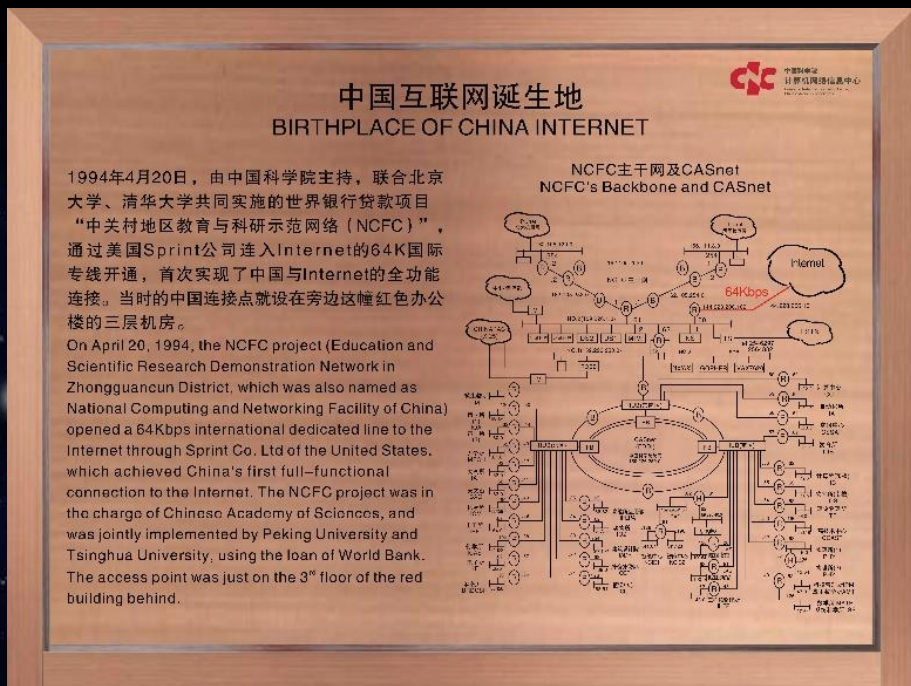
- 测量、分析、模型化
- 性能、扩展性、安全性
- 体系结构与组网技术
- 芯片与设备
- 创新应用

中科院与中国互联网发展

Internet Hall of Fame

- 1975年，计算所开始关注网络研究
- 1981年，计算所成立了网络研究室(十室)，国内最早从事互联网研究的实验室
- 1983年，中科院与德国弗朗霍夫信息与生物技术研究合作
研制了X.25分组交换网络，十室承担了该项目
 - 国内最早开始与国外合作从事的网络研究项目
 - 开发了ISO X.25底层软件，网络通信和管理软件
- 1989：NCFC
- 1992年6月开始，由十室研究中国域名体系
- 1995年3月，成立计算机网络信息中心





- IPv6网络关键技术和城域网示范系统，知识创新工程重大项目，2001
- 新型网络体系与机制研究，973，2011



为什么开设研究生网络课？



What

体系结构

协议



How, Why, Next

算法, 实现

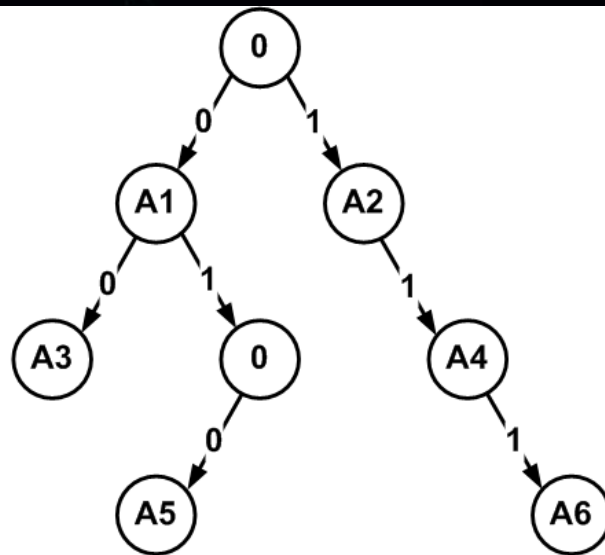
挑战, 研究

路由器：不仅仅是路由协议



前缀	下一跳
0*	A1
1*	A2
00*	A3
11*	A4
010*	A5
111*	A6

(a)

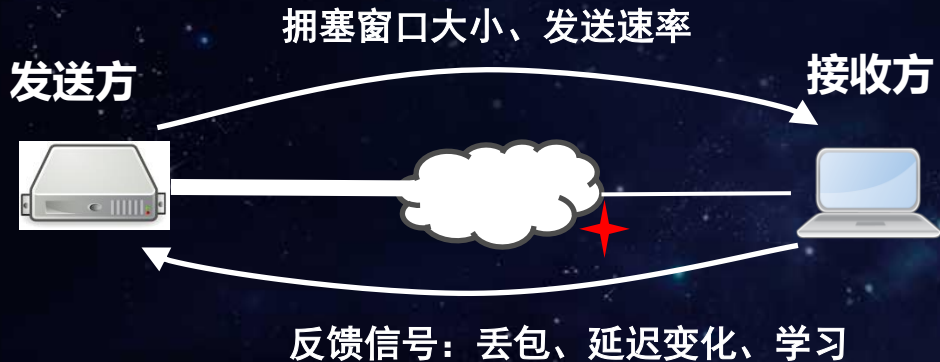
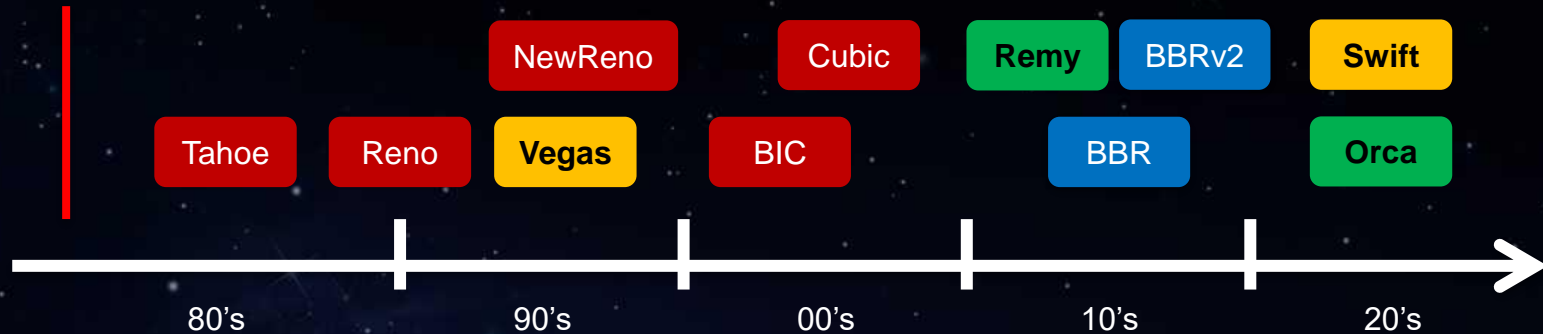


(b)

时间复杂度、空间复杂度、更新；SRAM、TCAM、Pipeline

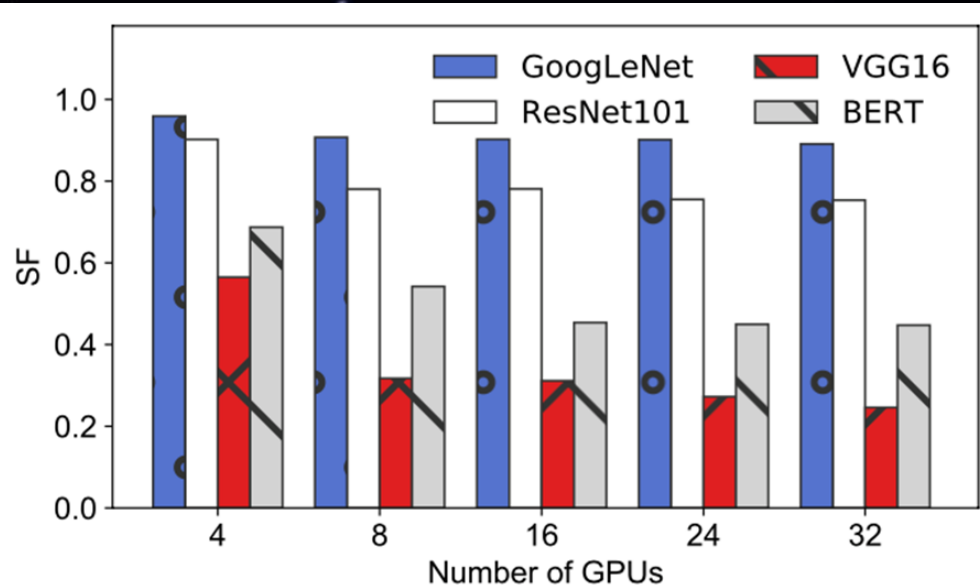
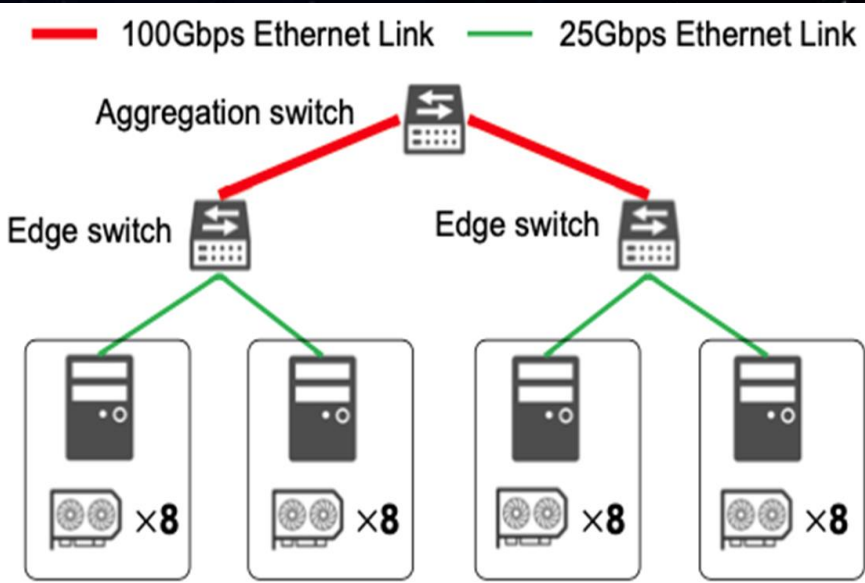
TCP: 控制理论

拥塞崩溃 (1986)



- 流量控制: 防止发送方 > 接收方速率
- 拥塞控制: 防止发送方 > 网络处理能力
- 差错控制: 发送方推断丢包并重传

分布式深度学习系统：网络与计算



$$\text{scaling factor} = \frac{T_1}{T_N \cdot N}$$

课程内容设置

- 涵盖体系结构与工程实现
 - 回顾，体系结构：设计原则、功能、功能放置、接口
 - 重点，工程实现：协议、算法、实现机制
- 涵盖基本原理与最新进展
 - 5G、FIA、NetAI、SDN/NFV、DCN、云计算、边缘计算、区块链
- 涵盖网、边、端
- 涵盖设备、路径、系统

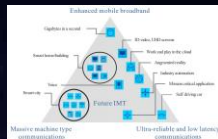


内容大类	课程题目	主要内容
网络基础	网络(课程)概述	网络概述、课程介绍
	网络直连	数据链路层机制, 5G等不同数据链路
	网络互联	交换、路由、队列管理
	路由机制	路由计算、协议、MIP
	网络传输	TCP、MPTCP
网络专题	网络应用与课程实验	网络应用协议、课程实验
	服务质量	服务质量原理、流量调度
	传输机制与优化	TCP测量、优化机制及BBR、QUIC等实现
	网络测量	测量方法、可用带宽测量、拓扑测量、测量应用
	网络安全	密码、网络攻击、移动互联网安全
	路由器设计与实现	路由器原理、路由查找算法
	SDN/NFV	SDN/NFV概念、流表查找
	数据中心网络	拓扑、传输控制、故障定位
	分布式机器学习系统	分布式机器学习框架、测量、优化
	内容分发	CDN、P2P
	区块链技术与应用	区块链技术、应用与挑战
	未来网络体系结构及应用	体系结构演进, 代表性体系结构、实现与测试
	前沿学术讨论	分组讨论

课程内容

直连网络(链路传输)

- 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
- 典型网络接入方式
 - 以太网、无线局域网、蜂窝通信网络(5G)



网络互连

● 交换网络

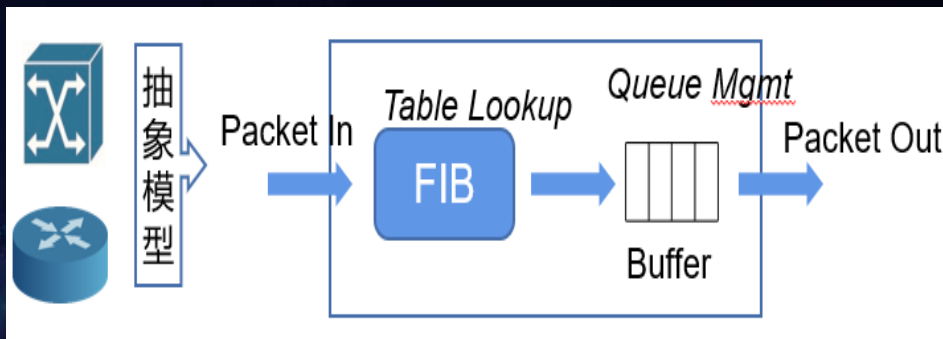
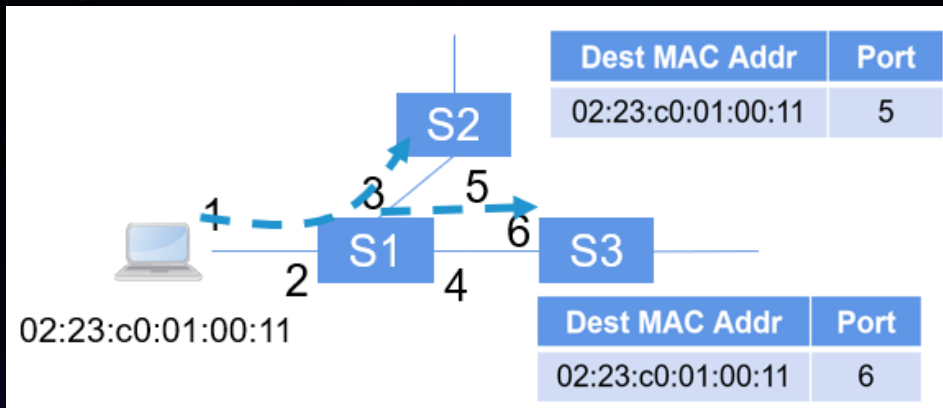
- 交换机学习：自动学习转发规则
- 生成树协议：构建树状逻辑拓扑

● 网络互连

- IPv4协议、数据包转发
- IPv6协议、IPv6过渡机制

● 数据包队列

- 数据包队列管理机制



网络路由

● 路由路径计算

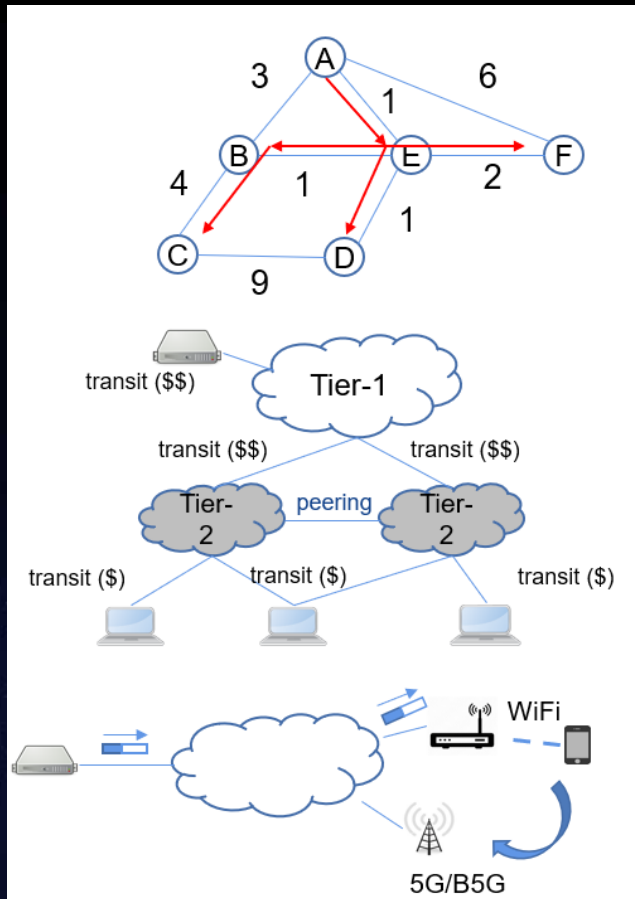
- 距离向量、链路状态

● 路由协议

- 域内路由协议 RIP, OSPF
- 域间路由协议 BGP

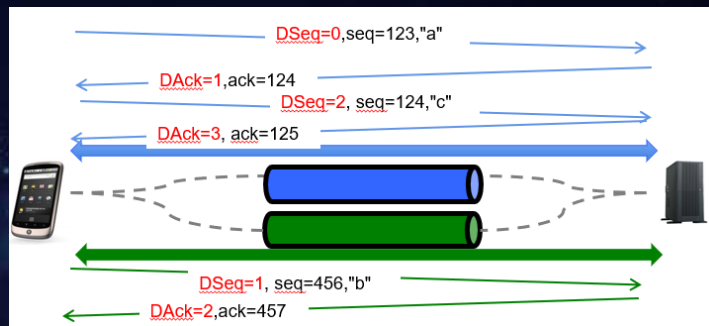
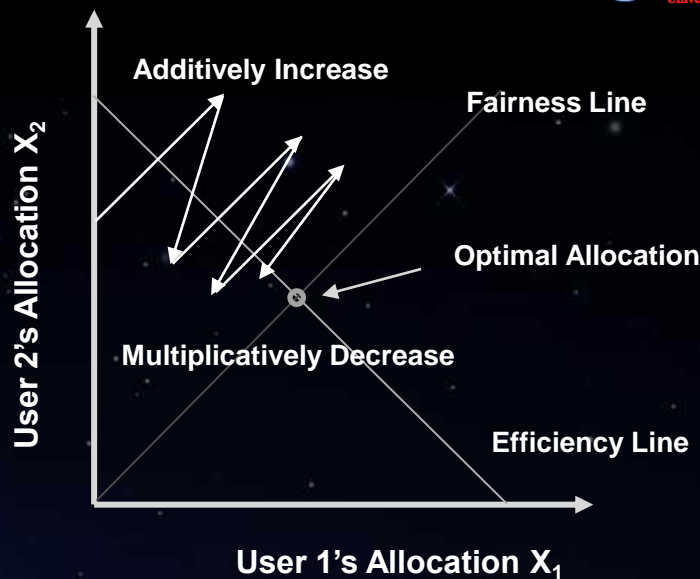
● 其他路由机制

- Mobile IP、基于扁平化标识的路由
- Segment Routing



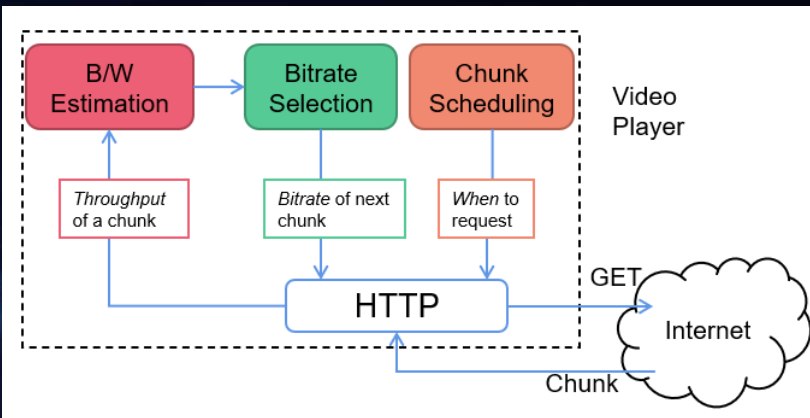
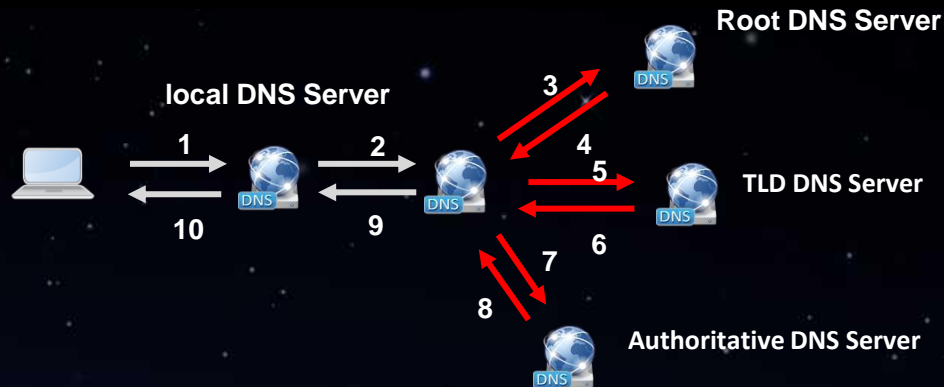
网络传输

- 最简单的传输协议 UDP
- 可靠传输协议 TCP
 - 连接管理
 - 数据传输
 - 拥塞控制
 - 传输性能优化
- 多路径TCP



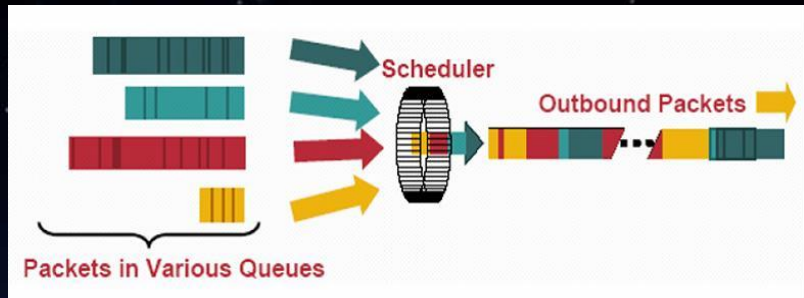
网络应用

- DNS /域名解析
- Web应用
 - HTTP、性能和安全
- 互联网视频
 - 系统设计、性能测量和优化



服务质量 (QoS)

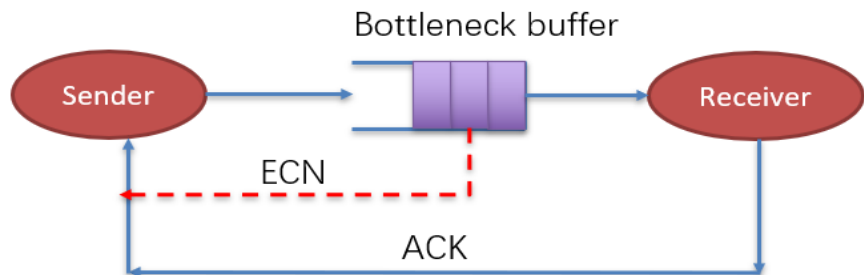
- 基本概念
- 确定性网络
- 流量分类和标记
- 队列管理
- 拥塞管理和调度策略



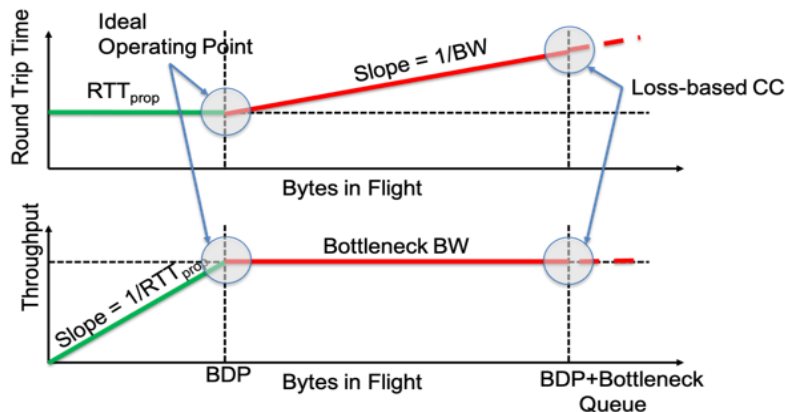
拥塞管理-队列调度

传输协议性能分析和优化

- TCP详解
- TCP实现及测量分析
- 拥塞控制机制研究进展
- 新型传输控制协议
 - BBR、QUIC、MPQUIC



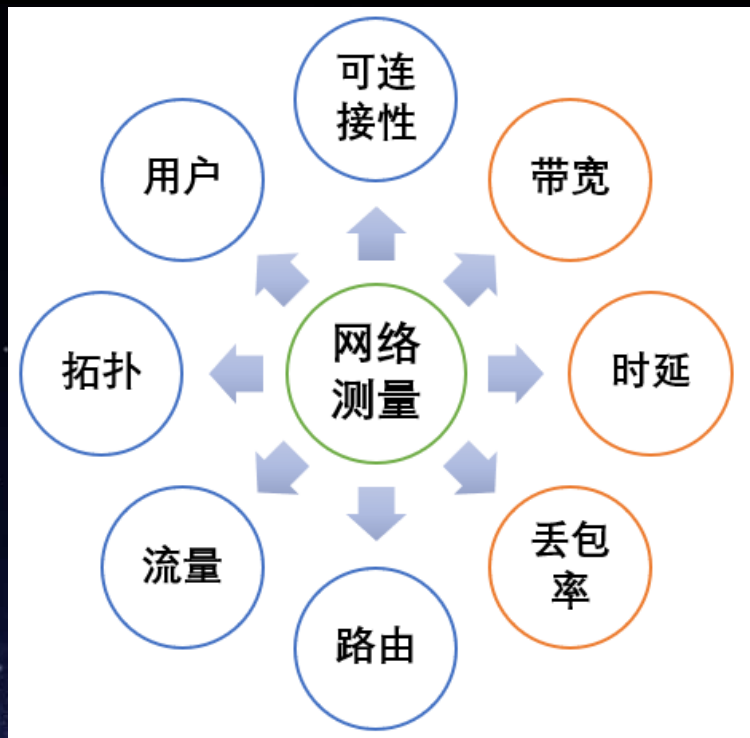
拥塞控制反馈



窗口(inflight)与RTT & 吞吐关系

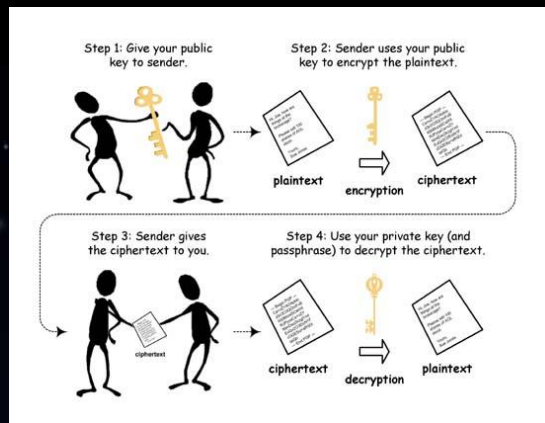
网络测量

- 测量概述
- 测量分类
 - 主动测量、被动测量
 - 拓扑测量、性能测量、流量测量
- 带宽测量
 - 可用带宽测量、瓶颈带宽测量
- 案例：利用视频流量进行网络测量
 - 绘制网络流量地图、预测链路带宽

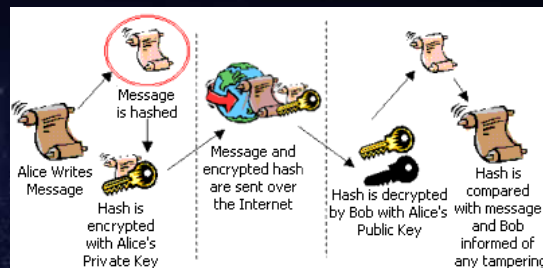


网络安全

- 网络安全概述
- 密码技术
 - 哈希、单密钥加密、公开密钥加密
 - RSA算法
 - 数字签名
- 网络攻击
 - 口令入侵、特洛伊木马、web欺骗等
- 移动互联网安全
- 新型网络安全威胁与防护



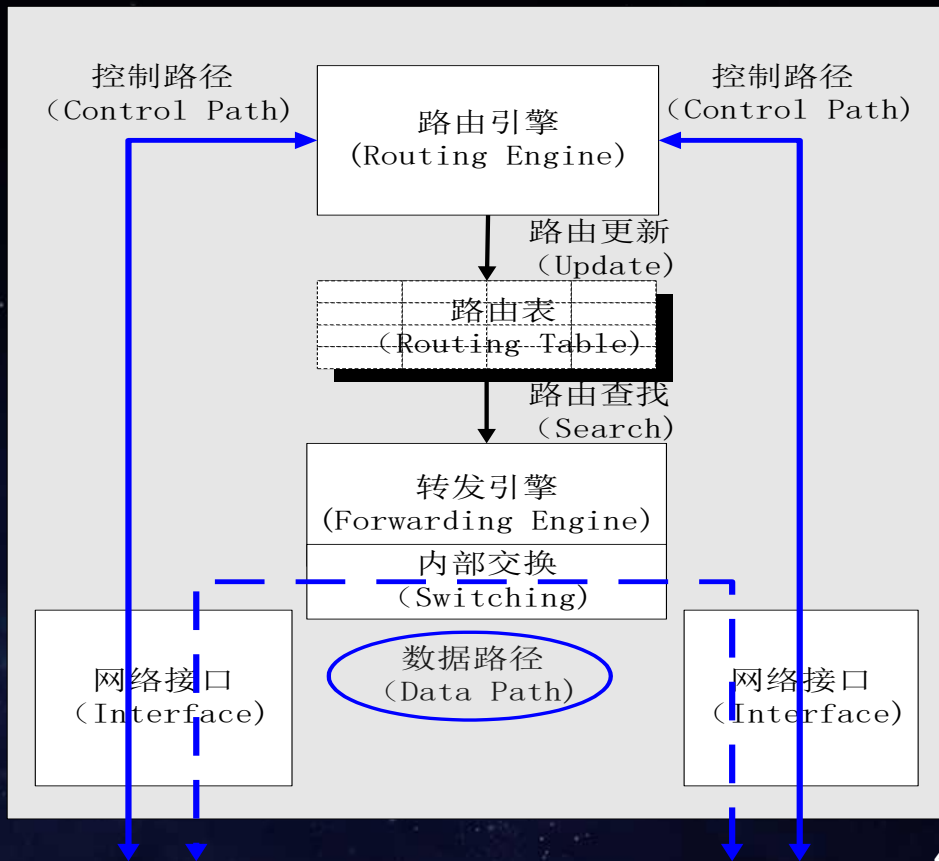
公开密钥加密认证



数字签名认证

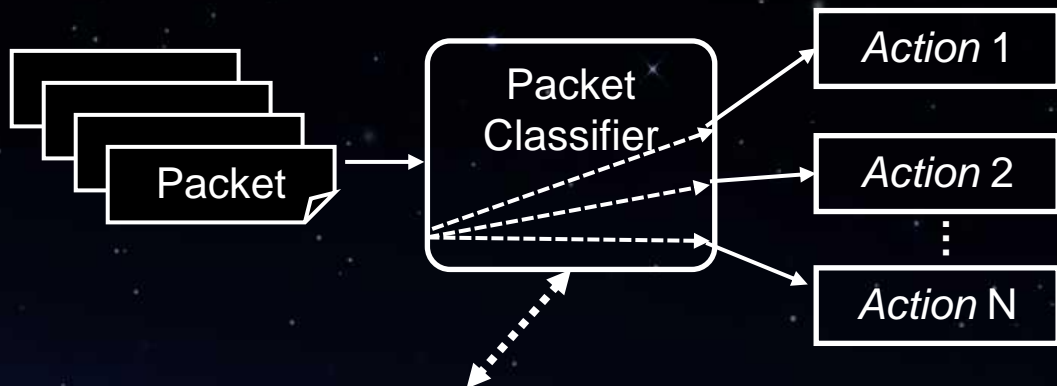
路由器设计与实现

- 路由器基本概念
- 路由器结构
- 路由查找算法



SDN/NFV

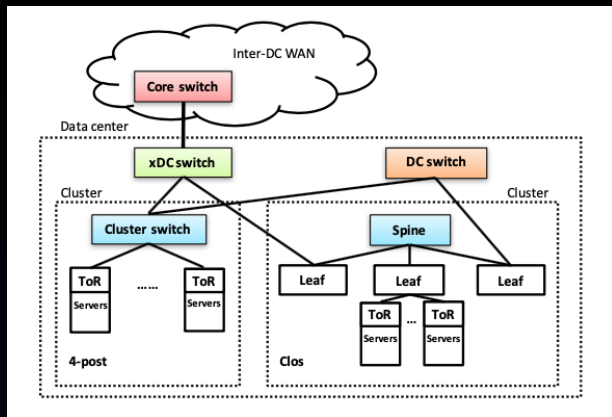
- SDN基本概念
- NFV基本概念
- 数据包查找转发算法
 - 虚拟路由查找
 - 数据包分类
 - 流表查找转发



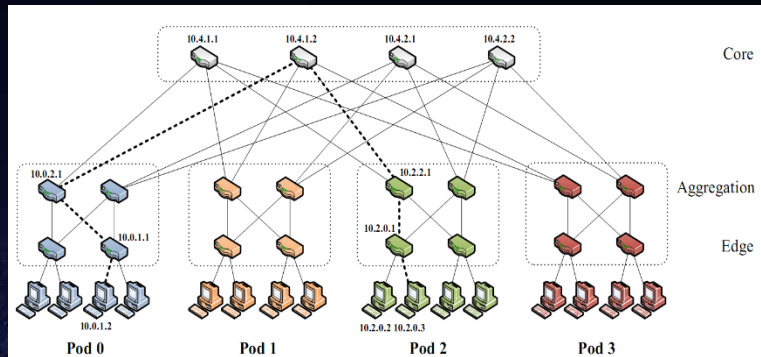
#	SA	DA	SP	DP	Prot.	Pri.	Act.
1	10.4.30.0/24	159.226.39.0/24	*	80	TCP	1	Fwd 2
2	49.123.0.0/16	188.0.20.128/30	233	53	UDP	2	Fwd 1
...
N	0.0.0.0/0	73.9.52.95/32	*	22	TCP	3	Drop

数据中心网络

- 数据中心网络概览
- 拓扑结构
 - Fat-Tree, VL2
- 流量模式
 - Facebook、DC-WAN (Baidu)
- 传输控制协议
 - TCP In-cast, DCTCP
- 基于INT的故障定位



数据中心网络 (DCN)



DC内部拓扑: Fat-Tree

分布式机器学习系统性能优化

- 分布式机器学习框架
- 数据并行模式
- 分布式通信机制
- 分布式机器学习系统性能分析与优化
 - 模型分析
 - 通信框架性能分析
 - 基础设施性能分析

Models

GoogleNet ResNet101 VGG16 BERT

Communication Framework

Communication Phase

NEGOTIATE ALLREDUCE WAIT

Communication Scheduling

Tensor Fusion

Communication Infrastructure

Transport Protocol

TCP/IP RDMA

Communication Topology

GPU Selection GPU Topology

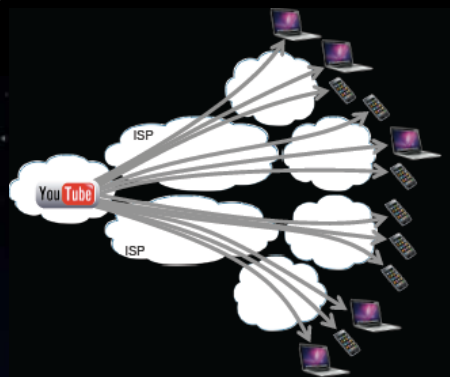
内容分发: CDN & P2P

- CDN (内容分发网络)

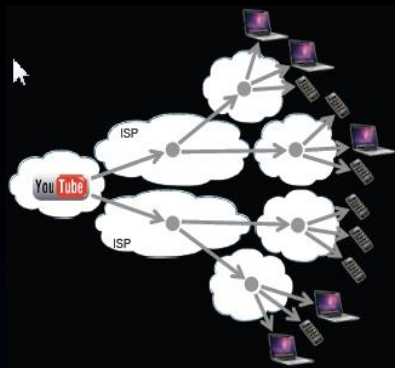
- CDN基本原理
- CDN缓存与放置
- 视频流媒体QoE优化

- P2P (对等网络)

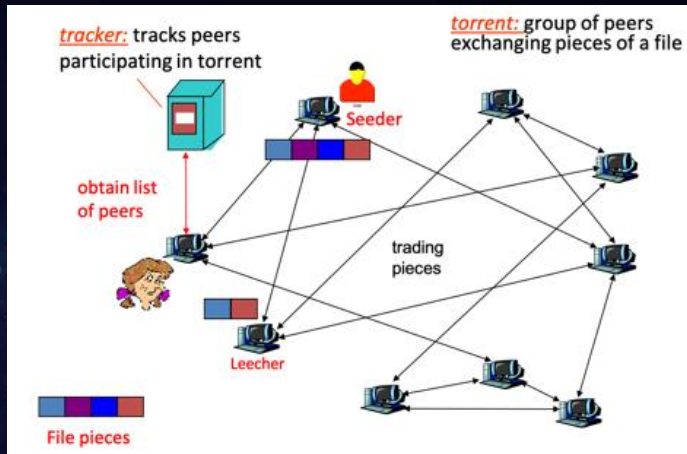
- 对等网络传输基本原理
- 典型应用



W/O CDN



W/ CDN



P2P文件分发

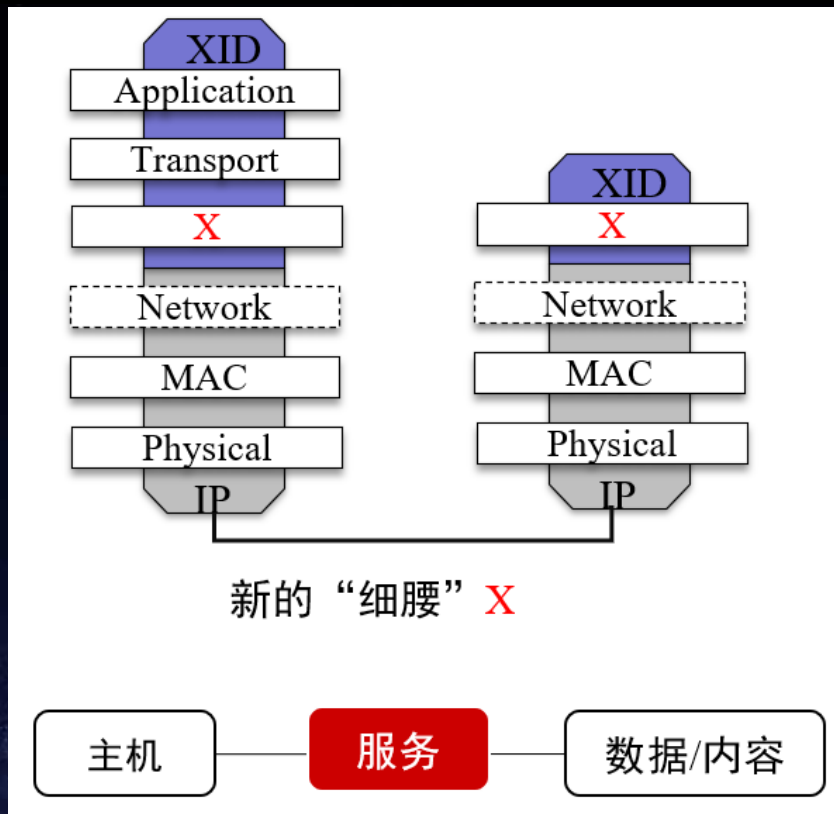
区块链技术与应用

- 区块链历史脉络
- 区块链技术
 - 分布式记账、拜占庭将军问题
 - 分类：公共、私有、联盟
- 区块链发展与应用
 - 供应链金融、跨境贸易、政务
- 区块链挑战与趋势



未来互联网体系结构

- TCP/IP体系结构
- “细腰”结构形成
- 典型未来互联网体系结构
- 演进机制和网络内缓存
- 应用和展望



理解问题，解决问题

独立思考

We reject kings, presidents and voting.
We believe in: rough consensus and
running code.

— David D. Clark

动手实践

Talk is cheap. Show me the code.

— Linus Torvalds

- 大作业30%
- 前沿学术讨论20%
- 平时课堂10%
- 考试40%

参考资料

- 教科书(任选)

- Larry L. Peterson, Bruce S. Davie, Computer Networks: A Systems Approach
- James Kurose, Keith Ross, Computer Networking, A Top-Down Approach

- 参考资料

- Sigcomm、NSDI、CoNEXT、ANCS、IMC、ICNP、Infocom、OSDI
- ToN、JSAC、TPDS、ToC
- IETF RFCs
- George Varghese, Network Algorithmics

- Opensource

- TCP/IP, Click, Open vSwitch, Snort等

课后阅读

- Barry M. Leiner, Vinton Cerf, et. Al., Brief History of the Internet, internetociety.org
- David D. Clark, The Design Philosophy of the DARPA Internet Protocols, ACM Sigcomm 1988
- Brian E.Carpenter, Where to Discuss a New Internet?, 2020*
- <https://www.wireshark.org/>



*<https://www.cs.auckland.ac.nz/~brian/InterOmnesNovasRete.html?from=singlemessage&isappinstalled=0>