近期重要时间节点

1、下周（16周）调课
时间：2020-12-27（周日），19：00-21：25
地点： 主M102
2、提交小组大作业（参见课程中心截止时间，
2020-12-24 20:00）
3、每人提交大作业技术报告（参见课程中心截止时间，
2021-1-7 20:00）
4、期末考试（开卷）：
19周，待定

主要内容

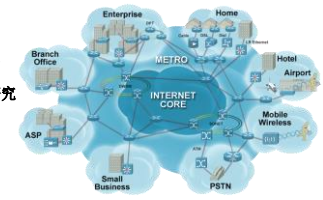
- 网络测量
  - 被动测量
  - 主动测量
  - 网络遥测（Network Telemtry）

网络测量

Network Measurement

Internet的现状

- 应用 &互联网络
  - 没有中央管理机构
  - 自治系统对本地通信的影响
  - 网络各个组成部分的深入研究
  - 网络的整体特征描述
- 实际部署的网络：
  - 复杂网络 complex-networks
- 全局属性难以推断
  - 多样性：Engineered with large technical diversity
  - 范围：Range from local campuses to transcontinental backbone providers



Internet 测量

- 为什么需要测量：The Internet is man-made
  - Because we still don't really understand it
    - Sometimes things go wrong
    - Malicious users
- 网络管理和运维的需要
  - Billing customers
  - Detecting, diagnosing, and fixing problems
  - Planning cost of new equipment
- 网络研究和科学发现的需要
  - Characterizing traffic, topology, performance
  - Understanding protocol performance and dynamics

动机

- 解决问题
  - Internet是如何运行的？
  - 效率如何?
  - 网络特性和趋势对运维的影响？
  - 未来的协议如何设计？
- 难点
  - 如何模拟和分析？
    - Need to understand how Internet is being used!
    - Too difficult to analyze or simulate parts we do understand

测量什么?

- 流量（Traffic）
  - Load statistics
  - Packet or flow traces
- 路径属性（Performance of paths）
  - Application 应用性能, e.g,. Web download time
  - Transport 传输性能 e.g., TCP bulk throughput
  - Network 网络性能 e.g., packet delay and loss
- 网络拓扑结构
  - Topology, and paths on the topology
  - Dynamics of the routing protocol

网络性能参数

- 时延 Latency
- 吞吐量 Throughput
- 响应时间 Response time
- 到达速率 Arrival rate
- 利用率 Utilization
- 带宽 Bandwidth
- 丢包率 Loss
- 路由 Routing
- 可靠性 Reliability
- 。。。。。。

在哪里进行测量?

- 端主机（End hosts）
  - 日志：Application logs, e.g., Web server logs
  - 探针：Sending active probes to measure performance
- 链路/路由器
  - Load statistics, packet traces, flow traces
  - Configuration state
  - Routing-protocol messages or table dumps
  - Alarms

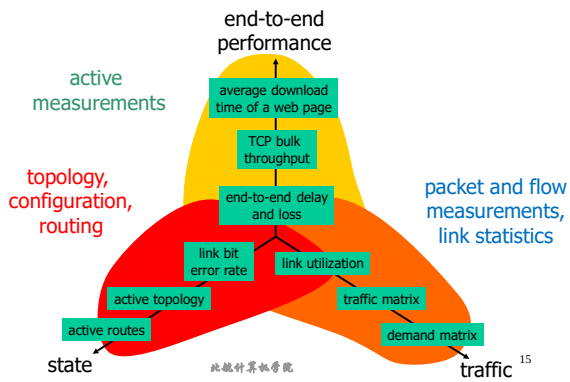
网络测量的挑战

- 无状态的路由器
  - Routers do not routinely store packet/flow state
  - Measurement is an afterthought, adds overhead
- 违背端到端原则
  - 中间盒：E.g., firewalls, address translators, and proxies
  - 不可见：Not directly visible, and may block measurements
- 去中心化控制（decentralized control）
  - 自治系统的限制：Autonomous Systems may block measurements
  - 缺乏全局时钟：No global notion of time

网络测量研究内容

- 端到端测量
  - 网络用户角度：用户使用网络存在的问题
  - Internet端到端路由特性
  - 端到端分组传输特性：延迟、丢包、带宽等
- 网络结构测量
  - 网络运营商角度：ISP内部网络运行中存在的问题
  - 流量矩阵估计
  - 流采样方法（如netflow）
  - 流量识别和分类
- 主要国际会议
  - Sigcomm， IMC， PAM， Sigmetrics

测量的分类



主动测量和被动测量

- **主动测量– 探测网络并分析其响应**
  - 如何测量到需要的信息？(and without bias)
  - 路由行为 routing behavior （如：路由变化对BGP行为的影响）
  - 分组动态性 packet dynamics（如： 执行传输，并记录其行为）
  - 延迟delay 和丢包 loss（如：记录UDP探针的行为）
- **被动测量– 测量现有的行为**
  - 不干扰网络信息的私有性
  - BGP异常（如：记录所有BGP交换的消息）
  - 路由行为（如：记录主机之间的路径信息）
  - 流量的自相似性 （如：记录以太网的流量）

测量的数据类型

主动测量

- traceroute
- ping
- UDP probes
- TCP probes
- Application-level "probes"
  - Web downloads
  - DNS queries

被动测量

- Packet traces
  - Complete
  - Headers only
  - Specific protocols
- Flow records
- Specific data
  - Syslogs …
  - HTTP server traces
  - DHCP logs
  - Wireless association logs
  - DNSBL lookups
  - …
- Routing data
  - BGP updates / tables, ISIS, etc.

测量工具

- **点对点测量工具**
  - Ping及其变种，“往返时间”和“丢包率”
- **路由信息测量工具**
  - traceroute及其变种，测量从源端到目的端的路由信息
- **路由协议交互（BGP，OSPF等）：Routeviews, RIPE RIS**
- **吞吐量测量工具**
  - 通过测量带宽的方式来获得吞吐量
- **流量监控工具**
  - TcpDump、Netflow、OCXMON、DAG等流量采集工具
- **Tcptrace等离线流量分析工具**
- **统计分析工具**
  - mrtg,rrdtool等，与流量监视工具不同，主要是给出网络状态消息数据，并进行一定的统计分析

Passive measurement

被动测量

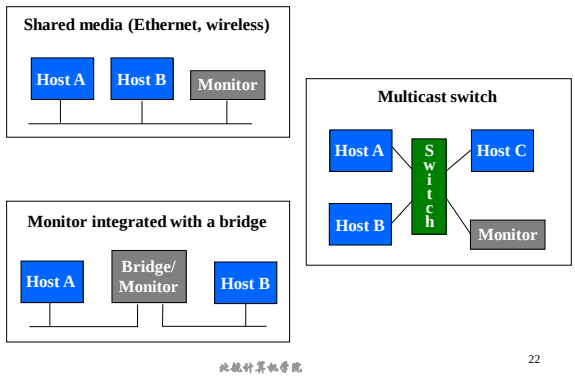
“Passive” Traffic Measurement

- 分组: Packet-level:
  - Tcpdump: software based
  - Special hardware packet collectors
- 流: Flow-level:
  - Cisco Netflow; other vendors have similar facility
  - 5-tuple flow: srcIP, dstIP, scrPort, dstPort, protocol
    - use a time-out value to “terminate” a flow
    - statistics collected: start/end time, packet/byte counts
  - Sampling may be used for scalability
- 链路: Link-level:
  - SNMP traffic statistics, often over 5-min interval
  - IETF MIB (management information base)
    - Byte counts, packet counts, etc.

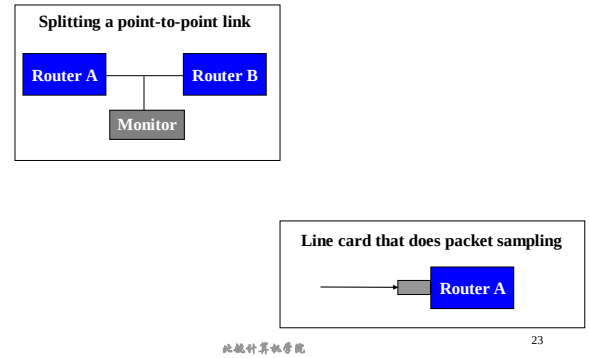
分组监测Packet Monitoring

- 什么是分组监测?
  - Passively collecting IP packets on one or more links
  - Recording IP, TCP/UDP, or application-layer traces
- 适用范围
  - Fine-grain information about user behavior
  - Passively monitoring the network infrastructure
  - Characterizing traffic and diagnosing problems

局域网：Monitoring a LAN Link



广域网：Monitoring a WAN Link



流量选择与过滤

- 过滤器设置： Filter to focus on a subset of the packets
  - IP addresses/prefixes (e.g., to/from specific sites)
  - Protocol (e.g., TCP, UDP, or ICMP)
  - Port numbers (e.g., HTTP, DNS, BGP, Napster)
- 数据采集：Collect first n bytes of packet
  - Medium access control header (if present)
  - IP header (typically 20 bytes)
  - IP+UDP header (typically 28 bytes)
  - IP+TCP header (typically 40 bytes)
  - Application-layer message (entire packet)

测量目标

- 理解路由器负载模型
  - Distribution of packet sizes
- 定量分析web传输的包大小
  - Number of packets/bytes per connection
- 分析服务器的访问特征：Know which servers are popular & who their heavy clients are
  - Collect source/destination IP address (on port 80)
  - Collection application URLs (harder!)
- 安全监测：Know if a denial-of-service attack is underway
  - SYN flooding (spoofable)
  - Unusual # requests to particular (potentially expensive) page

Analysis of IP Header Traces

- 源/目的地址
  - Identity of popular Web servers & heavy customers
- 路由器上的分组时延分布
  - Identification of typical delays and anomalies
- 分组大小分布
  - Workload models for routers
- 链路上流量的突发性
  - Provisioning rules for allocating link capacity
- 端到端的吞吐量
  - Detection and diagnosis of performance problems

TCP Header Analysis

- 源和目的端口号
  - Popular applications; parallel connections
- Sequence/ACK 序号，分组时间戳
  - Out-of-order/lost packets; throughput and delay
- 每个连接上的分组/字节数
  - Web transfer sizes; frequency of bulk transfers
- 同步标志位：SYN flags from client machines
  - Unsuccessful requests; denial-of-service attacks
- 终止标志位：FIN/RST flags from client machines
  - Frequency of Web transfers aborted by clients

Active measurement

主动测量

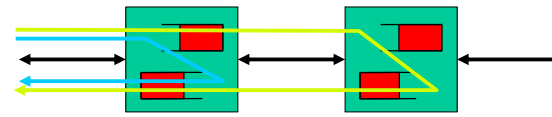
主动测量

- 方法：向网络中发送探针，并测量响应时间
  - Ping: RTT and loss
    - Zing: one way Poisson probes
  - Traceroute: path and RTT
  - Nettimer : latest bottleneck bandwidth using packet pair method
  - Pathchar: per-hop bandwidth, latency, loss measurement
    - Pchar, clink: open-source reimplementation of pathchar
- 问题：measurement timescales vary widely

Ping

- 向网络中注入流量
  - 折中：Trade-offs between accuracy and overhead
  - 干扰：Need careful methods to avoid introducing bias
- Ping
  - Host sends an ICMP ECHO packet to a target
  - ... and captures the ICMP ECHO REPLY
  - Useful for checking connectivity, and RTT
  - Only requires control of one of the two end-points
- 问题
  - Round-trip rather than one-way delays
  - Some hosts might not respond

Pathchar for Links



$rtt(i+1) = rtt(i) + d + L/c + \varepsilon$ 如何推导传播时延d, 链路容量c?

i : initial TTL value

c : link capacity

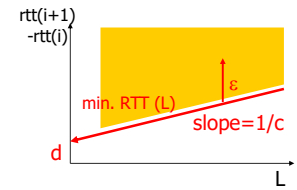
L : packet size

Three delay components:

d : propagation delay

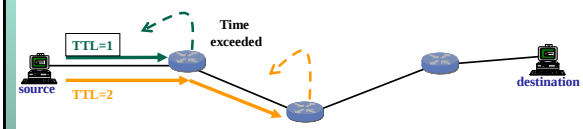
L/c : transmission delay

ε : queueing delay + noise



Traceroute

- Time-To-Live field in IP packet header
  - Source sends a packet with a TTL of *n*
  - Each router along the path decrements the TTL
  - "TTL exceeded" sent when TTL reaches *0*
- Traceroute tool exploits this TTL behavior



Send packets with TTL=1, 2, 3, … and record source of "time exceeded" message

Traceroute 局限性

- 多重路径问题：Measuring multiple paths
  - Successive probes may traverse different paths
- 非协作节点：Non-participating network elements
  - Some routers and firewalls don't reply
- 延迟不准确：Inaccurate delay information
  - Includes processing delays on the router CPU
- 双向Round-trip vs. 单向one-way measurements
  - Paths may have asymmetric properties
- 多接口设备：Interfaces, not routers
  - Returns IP address of interfaces, not routers

参考论文： Soule *et al.*, "Avoiding Traceroute Anomalies with Paris Traceroute", *IMC 2006*

Traceroute 应用

- 网络故障诊断 Network troubleshooting
  - Identify forwarding loops and black holes
  - Identify long and convoluted paths
  - See how far the probe packets get
- 网络拓扑结构推断 Network topology inference
  - Launch traceroute probes from many places
  - Join together to fill in parts of the topology

流量匿名化Anonymization

- Researchers always want full packet captures with payloads
  - ...but many questions can be answered without complete information
- Privacy / de-anonymization issues

带宽测量
(Bandwidth Measurement)

（补充一）

带宽测量（Bandwidth Measurement）

- 带宽Bandwidth
  - Amount of data the network can transmit per unit time
- 三种类型
  - *Capacity容量*: max throughput a link can sustain,
  - *available bandwidth 可用带宽*:
    - capacity – used bandwidth
  - *bulk transfer capacity块传输容量*: rate that a new single long-lived TCP connection would obtain over a path

定义

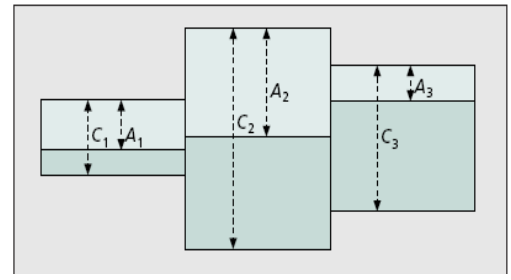
Available bandwidth可用带宽

- Let u_i be the average utiliztion of the link i over a period of time
- Let C_i be the capacity of the hop i
- Then the available bandwidth during that period
  - $A_i = (1 - u_i) C_i$
- Available bandwidth along the path

$$A = \min_{i=1,2,\ldots H} A_i$$

瓶颈链路Tight link(bottleneck): minimum avail-bw link

管道模型



带宽测量方法

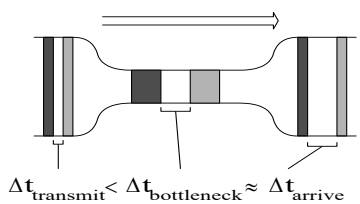
- Packet Dispersion technology
  - packet pair and packet train
  - Self-Loading Periodic streams (SLOPS)
- Variable Packet Size (VPS) technology
  - VPS even/odd
  - Tailgating technique

Packet Dispersion

- Sender sends two same-size packets back-to-back from source to sink.
- The packets will reach the sink dispersed by the transmission delay of the bottleneck links if there is no cross traffic.
- Measuring the dispersion can infer the bottleneck link bandwidth capacity.

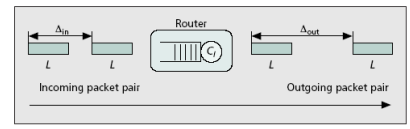
注: 瓶颈链路（Bottleneck link）可以是指具有最小数据率的链路，也可以是指具有最小可用带宽（available bandwidth）的链路。这里指前者。

Packet Pair Bandwidth Estimation



- Bandwidth = Packet Size / Separation

Packet Pair Dispersion (PPD)



- Measures end-to-end capacity

$$\Delta_{out} = \max\left(\Delta_{in}, \frac{L}{C_i}\right)$$

$$\Delta_r = \max_{i=0,1,\ldots H}\left(\frac{L}{C_i}\right) = \frac{L}{\min_{i=0,\ldots H}(C_i)} = \frac{L}{C}$$

Packet Pair存在的问题

- Competing traffic:
  - Time compression: Other packet queue ahead of the first probe packet when it is downstream of the bottleneck link. This leads to high estimates.
  - Time extension: Other packet delay the second probe packet and extends the spacing between the two probe packets. This leads to low estimates.
- Lack of queuing at bottleneck link
  - The probe packets were not sent fast enough to cause queuing at the bottleneck link. (Transmitting the packets slower than the bottleneck bandwidth would cause this)

Packet Pair存在的问题(续)

- Packet drops
- Multiple routes:
  - Out of order packet delivery
  - Multi-channel bottleneck links
- Clock resolution: Can't measure bandwidth higher than the one limited by the clock resolution.
- Changing bottleneck bandwidth: Routing changes or ISDN channel activating a second channel.
- Asymmetric Bandwidth: For methods that measures round trip time instead of one way transit time.

Packet Train Dispersion

- Packet train
  - sender sends the packets as one observation sample more than two
- Statistical filtering techniques
  - find valid samples.
- Interferring cross traffic
- Measure the multi-channel bottleneck link
- Reduce the limitation of clock resolution.

Self-Loading Periodic Streams(SLOPS)

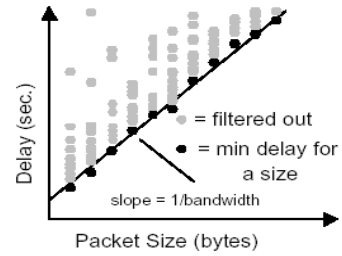
- Sender sends series of packets to the sink at the rate of larger than the bottleneck link available bandwidth.
- Every packets get a timestamp at sender side.
- Compare the difference of successive packets timestamp and their arrival times to infer the available bandwidth.
- Rate-adjustment adaptive algorithm to converge to the available bandwidth.

Variable Packet Size (VPS) Technique

- Step1. Sender set TTL=1, send out the packet, and wait for the ICMP TTL-exceeded packet back.
- Step2. Upon receiving ICMP, estimate the RTT.
  - Estimate the RTT multiple times for various size packets.
  - The minimum RTT of various packets are believed to be the valid sample.
- Step3. The first link capacity is C=1/β , β is slope of RTT graph.

Set the TTL=2,3...n, repeat the process of step1 to 3, to Calculate the C=1/ βi – βi-1

VPS 例子



VPS改进：Even-odd VPS

- 目标：improve reliability
- 方法：
  - For each of the probing sizes, divide the set of samples into even and odd numbers.
  - Calculation is based on even-odd samples. i.e. the even sample of link i, the odd sample of link i+1.

Tailgating Technique

- A deterministic model of packet delay
  - Unifies one packet and packet pair models
- Measuring link bandwidth using packet tailgating
- two phrase
  - Like VPS probing, but for entire path instead of per link.
  - The largest possible non-fragmented packet followed by a tailgater which is the smallest possible packet size (i.e 40 bytes). This causes the smaller packet always queue behind the larger packet.

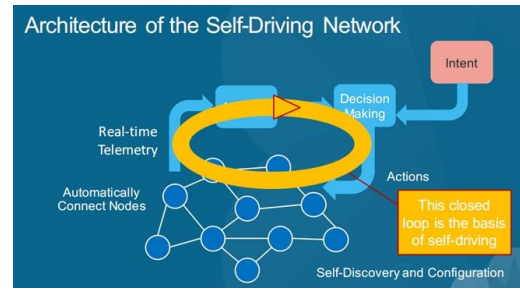
Ref.
- Lai, K., and Baker, M. Measuring Link Bandwidth Using a. Deterministic Model of Packet Delay, In Proceedings of the SIGCOMM (SIGCOMM'00)

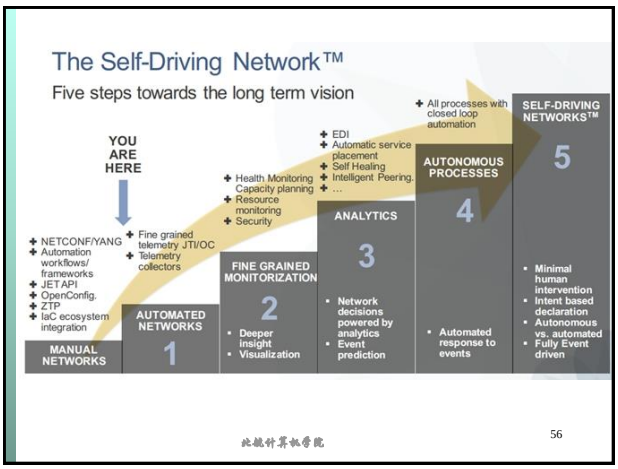
网络遥测

Network Telemtry

自动驾驶网络(Self-Driving Network)

- Juniper: 自动驾驶网络(Self-Driving Network)
  - 可预测并具有自主运行能力的网络





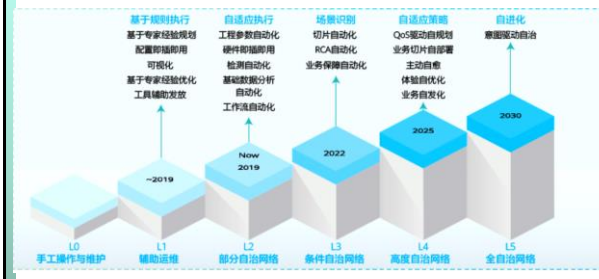
华为核心网自动驾驶网络

- Autonomous Driving Network（ ADN,自动驾驶网络）
  - 从客户体验、解放人力的程度和网络环境复杂性等方面，定义了通信网络的自动驾驶分级标准
- 智简网络（Intent-Driven Network，以下简称 IDN）

- L0手工运维：具备辅助监控能力，所有动态任务都依赖人执行。
- L1辅助运维：系统基于已知规则重复性地执行某一子任务，提高重复性工作的执行效率。
- L2部分自治网络：系统可基于确定的外部环境，对特定单元实现闭环运维，降低对人员经验和技能的要求。

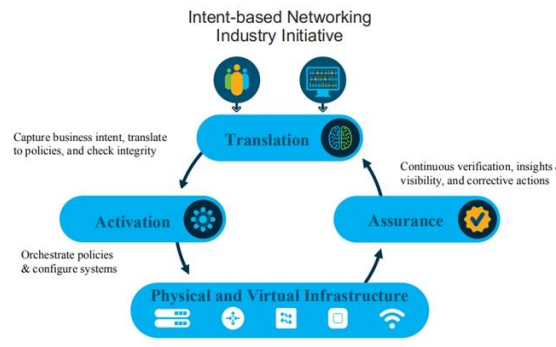
华为核心网自动驾驶网络（续）

- L3有条件自治网络：在L2的能力基础上，系统可以实时感知环境变化，在特定领域内基于外部环境动态优化调整，实现基于意图的闭环管理。
- L4高度自治网络：在L3的能力基础上，系统能够在更复杂的跨域环境中，面向业务和客户体验驱动网络的预测性或主动性闭环管理，早于客户投诉解决问题，减少业务中断和客户影响，大幅提升客户满意度。
- L5完全自治网络：这是电信网络发展的终极目标，系统具备跨多业务、跨领域的全生命周期的闭环自动化能力，真正实现无人驾驶。



基于意图的网络（IBN）

- Gartner：IBN定义包括四个部分
  - 转译和验证
  - 自动化实施
  - 网络状态感知
  - 保障和自动化优化/补救
- 客户的业务需求自动转换为网络配置策略



INT(Inband Network Telemetry)带内网络遥测

◆一种混合测量技术

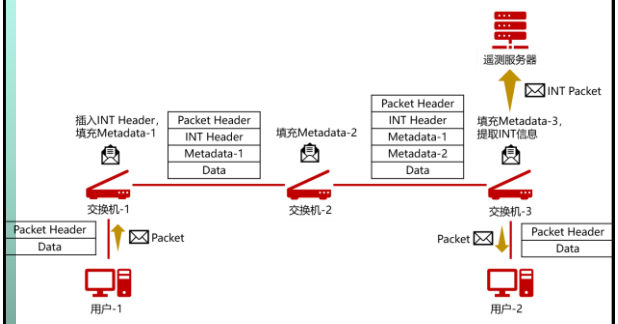
❖通过数据平面业务进行网络状况的收集、携带、整理、上报的技术，不使用单独的控制面管理流量进行上述信息收集。

◆特点

❖Inband（带内），借助数据平面的业务流量

❖Telemetry（遥测），为测量网络的数据并远程上报的特点

分类	常见协议或方案
主动测量	PING、Traceroute、Iperf、IPMP、OWAMP、TWAMP、MPLS L/DM、Pingmesh 等
被动测量	Netflow、sFlow、IPFIX、PSAMP 等
混合测量	Reactive Measurement、In-band Measurement、AM-PM、Postcard Based Telemetry、In-band Flow Analyzer、Hybrid Two Steps 等



大作业提交要求

◆ 小组提交大作业ppt和源码

❖ 每个小组 提交 2个 文档（小组中一个同学提交即可）：
（1）一份课程设计 ppt（总结大作业工作，汇报时间约5 分钟）；
（2）源代码（有必要的注释及编译运行环境说明），并压缩成 rar 或 zip 文件 。

◆ 每个同学独立提交大作业技术报告

❖ 每个同学独立提交大作业技术报告，对自己承担的工作内容以及相关技术进行综述。参考期刊论文的格式撰写。

提交文件：

（1）技术报告（word格式）

（2）主要参考资料

提交时间：具体时间参见课程中心的截止期。