

Overview:
Operations and Management of IP Networks

Based on SIGCOMM 2005 Tutorial by
Aman Shaikh and Albert Greenberg

1

**Q: Have you ever tried to configure or
troubleshoot your home network?
How do you do it?**

2

Goal

- **A robust IP network**
 - Hitless response to a plethora of “events:”
 - Failures, traffic/routing anomalies, mis-configurations, attacks, optical layer glitches, network maintenance
- **Arc of progress**
 - **Intuition-based**: test in the lab, and use trial and error in the field
 - **Managed**: integration of data and actions through scientifically well-founded tools
 - **Predictive**: system monitors, correlates, and recommends action
 - **Adaptive**: system monitors, queries as needed for additional data, and takes action
 - **Autonomic**: integrated components, dynamically managed to meet direct expression of user demands and network constraints

3

Examples

Management task	Progress
Internal traffic and network engineering; e.g., network maintenance	Managed, with inroads all the way to Adaptive
Inter-domain traffic engineering; e.g., decongesting peering links	Getting to Managed
Tuning QoS and packet handling mechanisms	Intuition-based
Multi-criteria control; e.g., DDoS and Network maintenance	Intuition-based

Q: Any other management tasks that you can think of?

4

What Do IP Networks Look Like?

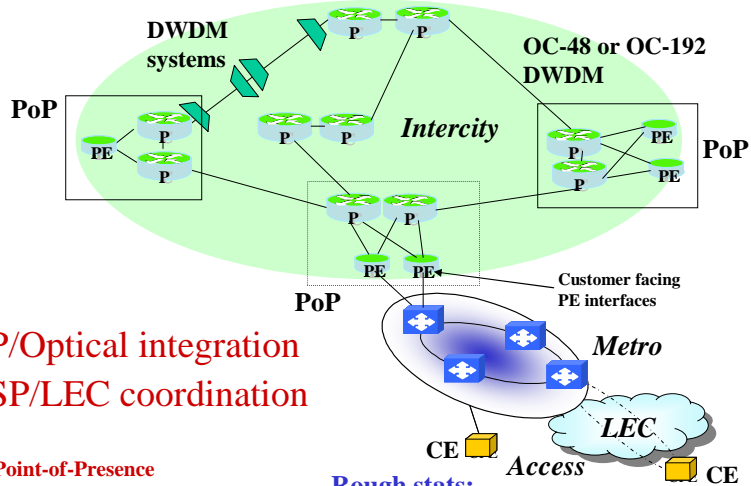
5

Commercial IP Networks

- **Enterprise networks**
 - IP networking for internal communication needs
 - Hierarchical topologies: the right structure for a small set of hubs (data centers), a huge set of spokes (remote offices)
 - Networking inside data centers is becoming an important area
- **Service provider networks**
 - IP networking as a service for a wide range of consumers, businesses, educational and government institutions
 - Mesh-like topologies: the right structure for convolving tens of thousands of enterprise networks
- **This lecture**
 - Service provider network focus
 - Yet, there are many commonalities: e.g., both have enormous geographic span, numbers of network elements, and complexity

6

Tier-1 Service Provider Network



- IP/Optical integration
- ISP/LEC coordination

PoP: Point-of-Presence
P: Backbone (core) Router
PE: Provider Edge Router
CE: Customer Edge Router

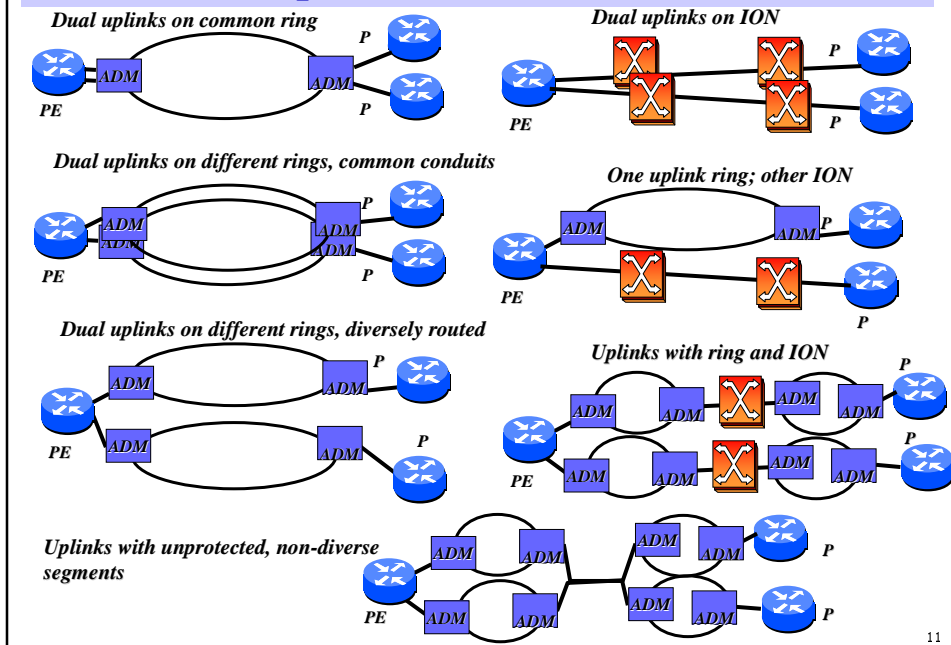
Rough stats:
 100s of offices
 100s of Ps, 1000s of PEs, 10000s of CEs
 100,000s of transport facilities

7

What Makes It Hard?

8

Complex Hardware Below IP



Emerging Applications (e.g. VoIP) Challenges

- **Architectural**
 - VoIP calls span multiple networks, server platforms and administrative domains, with signaling and media packets following different paths
- **Technological**
 - VoIP server infrastructure is relatively immature
 - Simple overload control principles are not built in (e.g., call-gapping – throttling overloads at the network edge)
 - Yet VoIP flash crowds are normal: e.g., enterprise-wide conference calls, call in votes for “American Idol”
- **Standards**
 - Competing signaling protocols (SIP, H.323, MGCP, ...), each in a state of flux, with overlapping functionality
 - E.g. normalizer for protocol parsers
 - Vendor inter-operability issues

Measurements Challenges

- **Inadequate understanding of application needs**
 - Other than for Voice
- **Inadequate implementation of the basics**
 - SNMP MIBs: e.g., packet and byte counts
- **Inadequate direct measurements of causality**
 - Cannot see the decisions (e.g., route changes), and lack the causality info to explain/diagnose problems
- **Inadequate collection and export of detailed data**
 - Flows, packet headers, routing protocol updates
 - Aggregation and sampling infrastructure
- **Inadequate ability to trace events across the network**
 - Example: no common key for Call Detail Records created at many boxes for same VoIP call

13

Wouldn't Be So Bad If It Didn't Keep Changing

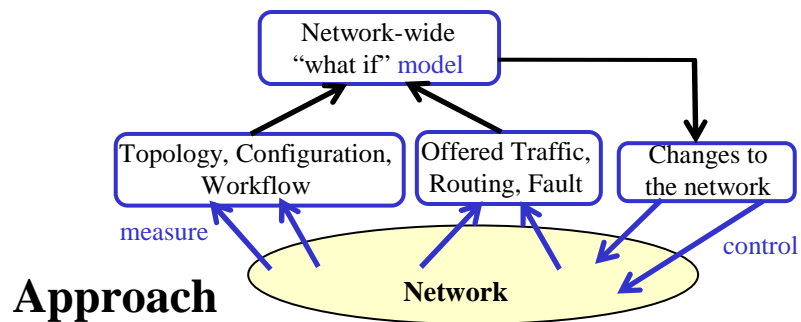
- **Technological advances, protocol evolution**
 - Example: Ultra Long Haul (ULH) in optical transport
- **Architectural change**
 - Example: executing commands to migrate to MPLS
- **Customers leave, join, change/upgrade service**
 - Example: migrating customers from Frame Relay to IP VPNs
- **“Events” will happen**
 - Failures, maintenance/software upgrades, misconfigurations, worms, viruses, DDoS, American Idol voting over VoIP

14

What To Do About It?

15

Measure, Model, Control via Intelligent Layer On Top



- Managing the entire network as one
- Predicting effects of actions in advance
- Making conscious trade-offs on conflicting goals
- Relying on direct correlation of real data

16

What the Network Really Needs

- **Measure: input data about the network**
 - Measurements of topology, traffic, performance
 - Ubiquity and robustness trump granularity
- **Model: accurate what-if analysis**
 - Effects of possible configuration changes
 - Predictability and robustness trump functionality
- **Control: automated reconfiguration**
 - Automated work-flow to change the configuration
 - Automation and robustness trump optimality

17

Components of the Solution

- **Network-wide views for troubleshooting, forensics, maintenance**
 - Example: traffic, configuration, fault, control plane
- **Advanced data fusion and correlation**
 - Example: OSPF + Active Performance Measurements
- **Intelligent algorithms and decisions**
 - Example: operations under impaired conditions
- **Automated and scalable network update**
 - Example: transactional update to all devices in a given role & context
- **Scalable and continuous auditing of network state**
 - Example: cyber-virus stamped out

18

Example: Maintenance

19

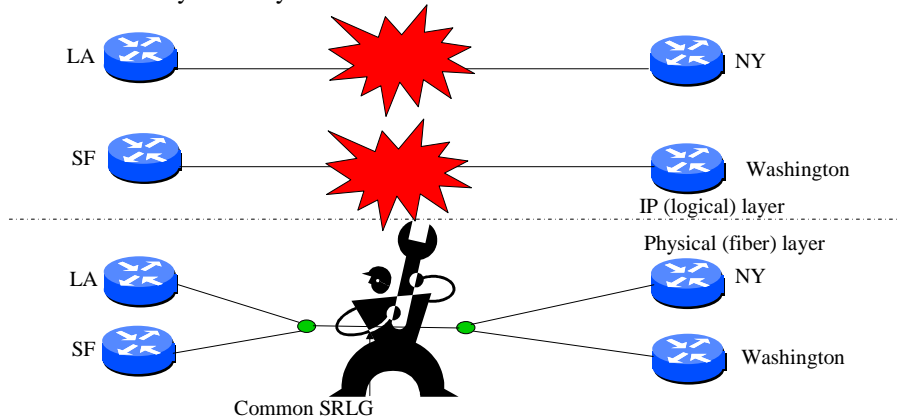
Network Maintenance

- **Why? IP networks are distributed hardware and software systems, undergoing change**
 - Hardware upgrades, faults, degrading condition
 - Software upgrades, vulnerabilities, bugs, features, executing commands leading to fundamental changes in the architecture (e.g., MPLS), ...
 - Software upgrades may require router reboot...
- **Where? A lot of boxes**
 - 1000s of routers
 - 100,000s of network elements below IP
- **When?**
 - On-site workforce available? Impact small? Customer notification required? Piggyback opportunities? Architectural exceptions? Special customer exceptions? ...

20

Example: Optical Maintenance Impacts IP

- Integrity of a simple IP link depends on a complex set of transport facilities
- Shared Risk Link Groups (SRLGs) codify cross layer dependencies (e.g., fiber conduits)
 - Not always directly available ...



21

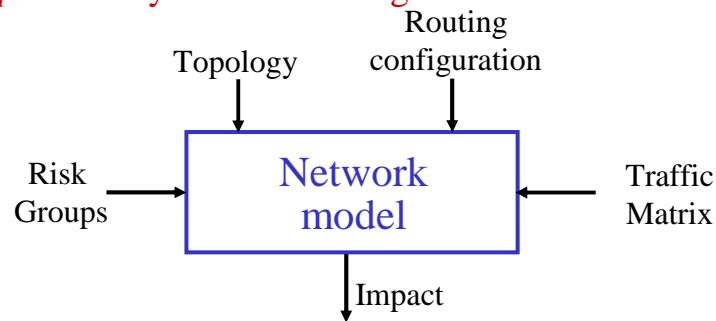
Two Tasks

- A maintenance project factors into two tasks
 - IP Decision Support (what-if analysis) to estimate and minimize impact
 - Detailed specific sequence of actions to perform maintenance
- Impact: unacceptable potential hits for customers and a growing set of applications
 - VoIP, IPTV, ...
 - SLAs with real teeth for VPNs

22

IP Decision Support

- **Goal:** through systems and tools, reliably predict the *impact* of any event involving a router



- **IP Decision Support has a plethora of other applications**
 - Risk, survivability, security vulnerability analysis; network and service evolution; capacity planning

23

Minimizing Impact: Lifecycle

- **Measure: input data about the network**
 - Measurements of topology, traffic, performance
 - Traffic matrix inference a hard problem for IP networks
- **Model: accurate what-if analysis**
 - IP route simulation
 - Effects of possible configuration changes
- **Control: hitless maintenance**
 - Cost-out/Cost-in process
 - Take router down for maintenance/ reboot
 - Router enhancements that can help
 - Router upgrades itself without reboot

24

Traffic Matrices: Big Picture

- Router Level Demand Matrices

- Granularity: router or router interface
- Killer App: IP Decision Support
- Innovation: Tomo-gravity

What has allowed robust IP decision support

Used extensively throughout the network.

Rapid R → D → Ops.

- Flow Level Demand Matrices

- Granularity: TCP/IP headers
- Killer App: Traffic Analysis with Drill-down
- Innovation: Priority Sampling

- Path Matrices

- Granularity: TCP/IP headers
- Killer App: Passive Performance Measurement
- Innovation: Trajectory Sampling

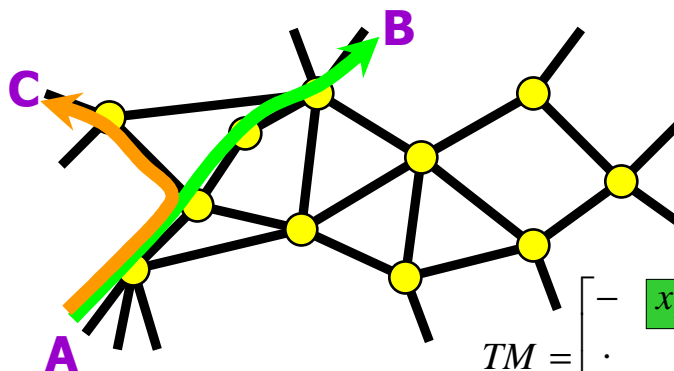
25

Traffic Matrix Estimation

Have link traffic measurements

Want to know demands from source to destination

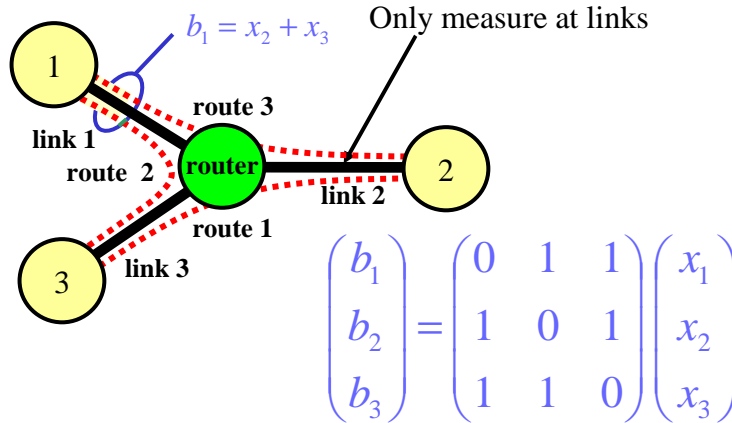
Approach: network tomography



$$TM = \begin{bmatrix} - & x_{A,B} & x_{A,C} & \cdots \\ \cdot & \cdot & \cdot & \cdots \\ \cdot & \cdot & \cdot & \cdots \\ \cdot & \cdot & \cdot & \cdots \end{bmatrix}$$

26

Problem: $b = Ax$



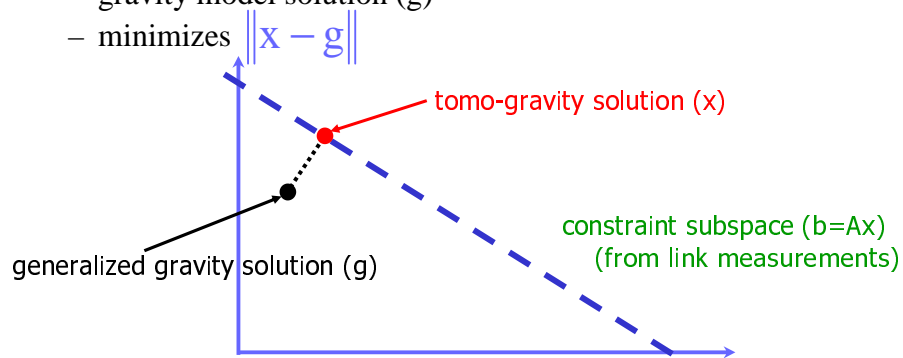
Problem: Estimate traffic matrix (x 's) from the link measurements (b 's)

The problem is massively under-constrained

27

A successful approach: Tomo-gravity

- Tomo-gravity = tomo-graphy + gravity modeling
- Reduce problem size
 - Exploit topological equivalence
- Find a solution x , which
 - satisfies the constraints, and is closest to the generalized gravity model solution (g)
 - minimizes $\|x - g\|$



28

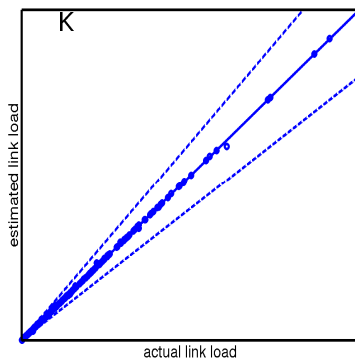
Foundation in Information Theory

- **Minimize Mutual Information I(S,D)**
 - Information gained about source (S) from destination (D)
 - Assume no information beyond the link load constraints $b=Ax$
- **Framework for tomo-gravity**
 - Gravity model = independence (between S and D)
 - Generalized gravity model = conditional independence
- **Explains tomo-gravity's success with** $\|x - g\| = \sum_p \left\{ \frac{(x[p] - g[p])}{\sqrt{g[p]}} \right\}^2$
 - First-order approx. to **Kullback-Leibler** divergence from independence for I(S,D)

$$\begin{aligned} K(x \| g) &= \sum_p x[p] \cdot \log \frac{x[p]}{g[p]} \approx \sum_p x[p] \left(\frac{x[p]}{g[p]} - 1 \right) \\ &= \sum_p x[p] \left(\frac{x[p]}{g[p]} - 1 \right) - \sum_p (x[p] - g[p]) = \sum_p \left(\frac{x[p] - g[p]}{\sqrt{g[p]}} \right)^2 \end{aligned}$$

29

Tomo-gravity Works



- **Simple, and quick: A few seconds for large IP backbone**
- **Accurate: average ~11% error**
 - Including netflow significantly improves this! Errors become a few percent.
- **Uses widely available SNMP data**
 - Highly robust → can work within the limitations of SNMP data

30

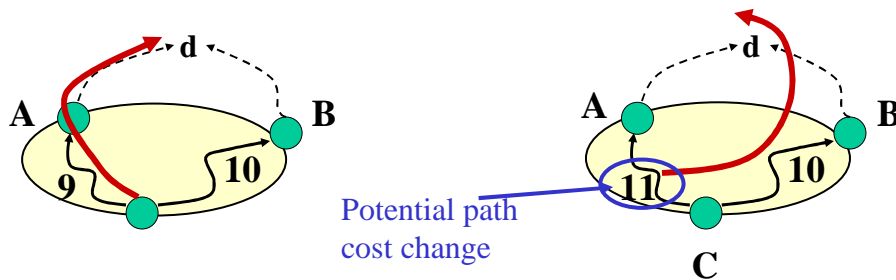
Minimizing Impact: Lifecycle

- Measure: input data about the network
 - Measurements of topology, traffic, performance
 - Traffic matrix inference a hard problem for IP networks
- Model: accurate what-if analysis
 - IP route simulation
 - Effects of possible configuration changes
- Control: hitless maintenance
 - Cost-out/Cost-in process
 - Take router down for maintenance/ reboot
 - Router enhancements that can help
 - Router upgrades itself without reboot

31

IP Route and Risk Modeling

- Inputs
 - Physical topology, IP topology, SRLGs
 - IP routing, IP traffic matrix
 - Workflow: maintenance schedule
- Output
 - Metrics relating to congestion
 - Suggested routing adjustments



32

Back To Router Maintenance...

- **Example Scenario**
 - Failure in LA, maintenance scheduled in NY
- **Can the network “survive” the maintenance?**
 - Performance metrics within the design envelope
- **Apply IP Decision support tools**
 - No impact \Rightarrow perform maintenance
 - Impact \Rightarrow reschedule if possible, or if not then re-optimize routing for least impact during maintenance

33

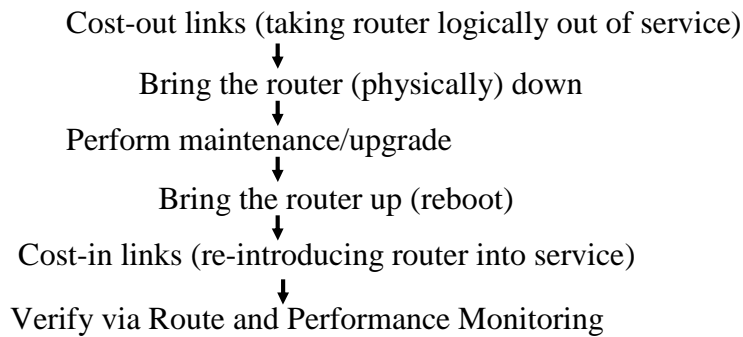
Minimizing Impact: Lifecycle

- **Measure: input data about the network**
 - Measurements of topology, traffic, performance
- **Model: accurate what-if analysis**
 - IP route simulation
 - Effects of possible configuration changes
- **Control: hitless maintenance**
 - Cost-out/Cost-in process
 - Take router down for maintenance/ reboot
 - Router enhancements that can help
 - Router upgrades itself without reboot

34

Cost-out/Cost-in Process

- **Challenge:** Perform maintenance with minimal impact on customer; upgrades etc. require router reboots!
- **Solution:** move traffic off the router / link under maintenance, perform maintenance, restore the router to service
- **Achieved through configuration changes**
 - Example: OSPF link weight changes, where “cost-out” means setting the link weight to a high value, to shift traffic off



35

Router Cost-out Options

- **Option 1: Cost-out all outgoing links of a router**
 - Based on IETF RFC 3137
 - Configuration changes only at the router in question
 - Cisco ‘max-metric router-LSA’ command allows entire cost-out in one atomic operation
- **Option 2: Cost-out all incoming links of a router**
 - Configuration changes (cost-out links) at the routers neighboring the one in question
- **Which is better?**
 - Option 1 is operationally simpler
 - Which has least impact on traffic? It depends

36

Summary

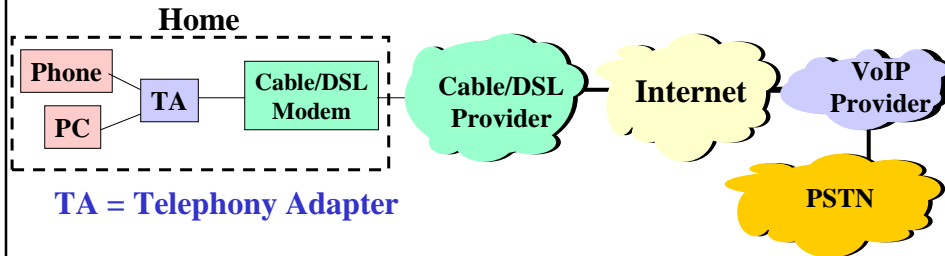
- **A Wild Wild World!**
 - Full of messy, hairy, challenging, important real-world problems
- **Tons of opportunities for science**
 - Need scientific network engineering
 - Measure, model and control
 - Inter-disciplinary by nature: statistics, algorithms, software engineering, visualization, machine learning, data mining, automation

37

Additional Slides

38

What Does a Service Look Like? Example: BYOA – Consumer VoIP

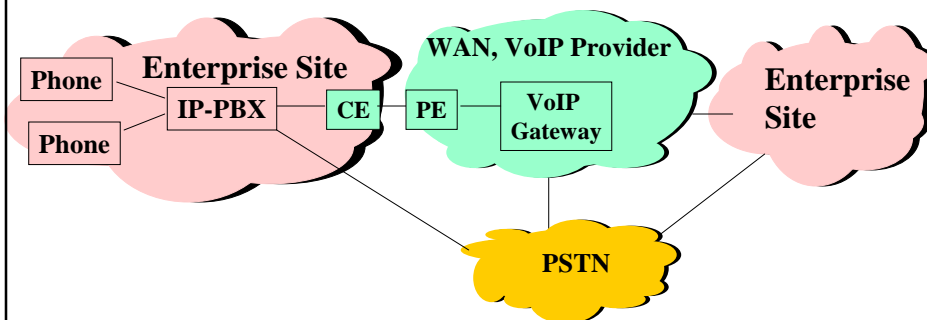


TA = Telephony Adapter

- **QoS**
 - In the Telephony Adapters (TA) and in the VoIP provider's application servers
 - *no access* to QoS in access or core networks
- **Commercial offers**
 - AT&T CallVantage, Vonage, 8x8, Skype

39

What Does a Service Look Like? Example: Business VoIP



- **QoS**
 - CE to CE, backed by an SLA
- **Commercial offers**
 - AT&T, Sprint, ... , often integrated with VPNs

40

Provisioning

41

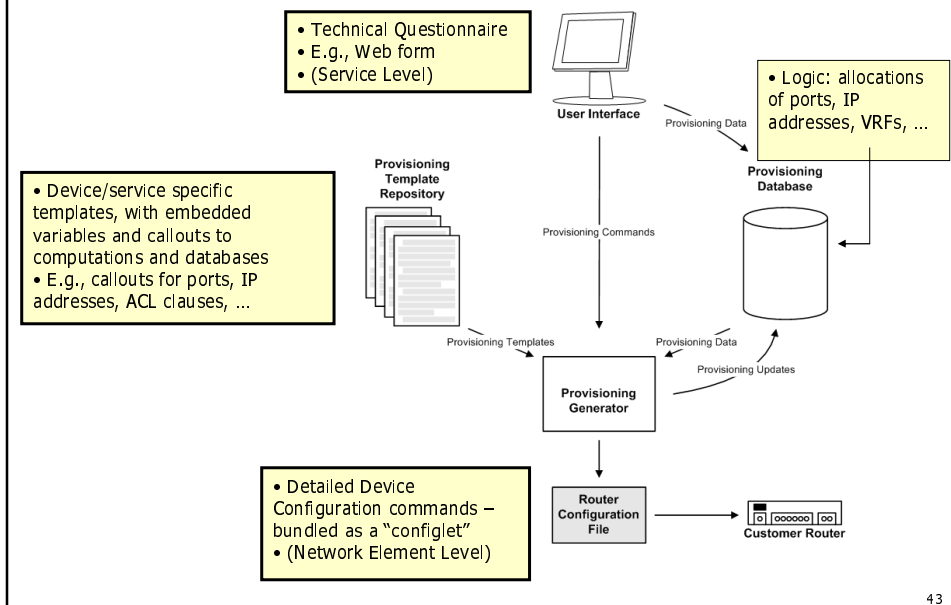
Provisioning

Transforming Service Intent to Network Reality

- **Speed and “flow-through” are the priorities**
 - Customers want “on demand” service
 - Providers want “on demand” revenue (which flows after provisioning)
- **When it’s slow, why?**
 - Physical provisioning of circuits can take time
 - Help on the way?
 - Market mechanisms, such as optical exchange (meet) points
 - Intelligent optical technologies for bandwidth on demand
- **Logistics**
 - Getting hardware, circuits to the right place at the right time
 - Updating databases, routers to establish and verify the service

42

Automated Router Provisioning



Provisioning Example CE and PE Access Interfaces

- **Basic interface configuration**
 - Media and location in router (POS7/3, ATM5/0.1)
 - IP address and network address (mask)
 - Capacity (bandwidth)
- **Rich configurable parameters at layer 3**
 - Packet marking and scheduling (differentiated services)
 - Buffer management (memory size, WRED parameters)
 - Access control (inbound and outbound packet filters)
- **Diverse communication media at layer 2**
 - Serial link, ATM, Frame Relay, packet over SONET, etc.
 - Various low-level, media-specific parameters

Provisioning Example CE and PE BGP Configuration

- **Determine customer's AS number**
 - Some customers have their own AS number
 - Example: customers multi-homed to multiple providers (e.g., 88)
 - Some customers cannot get their own AS number
 - Example: single-homed customers
 - Assign private ASN (64,512 to 65,535) or use provider's ASN
- **Establish BGP session with the customer**
 - Determine interface(s) connected to the customer
 - Associate BGP session with the interfaces
 - Set authentication parameters
- **Enforce provider's routing policies while taking customer's routing intent into account**
 - BGP import and export policies
- **Configure other BGP session parameters**
 - Example: timer settings

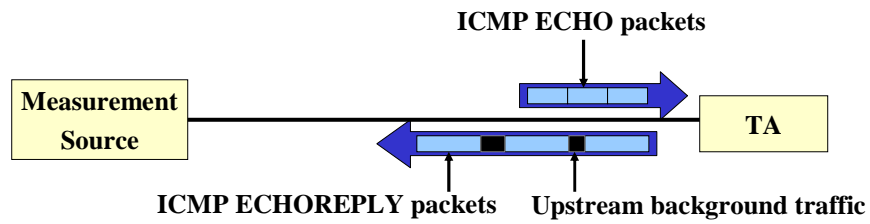
45

Provisioning BYOA VoIP TAs

- **Goal: plug and play**
 - Bootstrapping workflow – coordinating in-band with the VoIP service provider
- **Problem: provisioning *depends* on the environment**
 - Users expect to surf (as fast as before adding VoIP) and talk simultaneously
 - Upstream (home → Internet) bandwidth too small ⇒ VoIP infeasible
 - TAs that do not measure upstream bandwidth and self-configure
- **Solution: TA config dependent on *measured* upstream bandwidth**
 - Data: set MSS (max segment size) and QoS parameters
 - MSS too small ⇒ web download speed degrades (potential dissatisfier)
 - Voice: set CODEC and associated parameters
 - Provided TA supports multiple CODECs

46

Estimating Upstream Bandwidth of VoIP Customer



- **Methodology**
 - Measurement source sends ICMP ECHO packets to the TA
 - Estimate the customer's upstream bandwidth by measuring the arrival rate of ICMP ECHOREPLY packets from the TA
- **Assumptions:**
 - Upstream link of the customer is the bottleneck
 - TA replies to ICMP ECHO packets

47

Challenges

48

Challenges on Maintenance

- **How to achieve robustness in measurement, modeling?**
 - Insensitivities to constraints that are hard to track and model
 - Example: link utilizations should be < 80% except for links involved in that new VoIP trial in Phoenix with vendor X equipment, where utilizations should be < 50%, except for ...
- **How to achieve robustness in control?**
 - How to design mechanisms and checks to reduce the chance that automation goes wrong (and if it does, how to rapidly recover or rollback)
 - Guard rails provide a metaphor
- **How can we re-factor router design so that maintenance and repair is less needed and has less impact?**
 - E.g., router farms, router virtualization



49

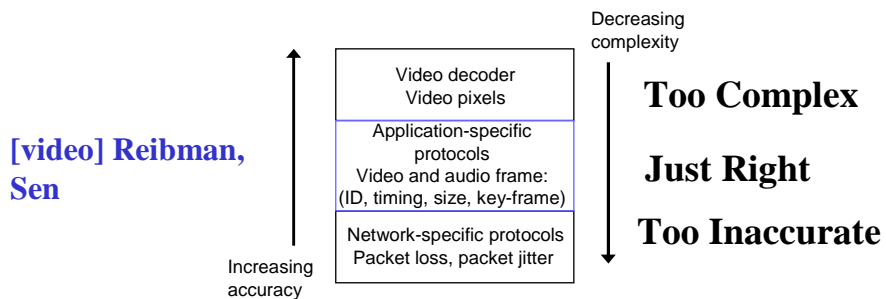
Challenges on Troubleshooting

- **Robust design**
 - How to create protocols and systems that are robust by design
 - How to build diagnosis capabilities into the protocols – revealing what the routers were thinking when deciding ...
- **Measurement: data data everywhere and not a thought to think**
 - How to find (possibly large) needles in (massive) haystacks – anomalies, stepping stones
 - How to design a minimal set of experiments that maximize information gained about the network
 - How to overcome proprietary data models, systems thwarting forensics
- **Modeling/Correlation**
 - One size fits all anomaly detector? How to trade domain expertise for statistics to scale across hundreds of data sources
 - How to use statistical inference to best reconstruct global state based on partial measurements
- **Rapid, automated response**
 - How to get to sub-minute response? How to assure every step of the detect, localize, diagnose, fix, verify lifecycle is very fast

50

Challenges in Provisioning

- **Scalability**
 - How can network providers measure with the same agility as customers?
 - N^2 problem for path measurements and for CE-CE SLAs
- **Application-oriented performance monitoring and SLAs**
 - How to reliably detect/repair at application level (across servers, networks)
 - Example: Video “R-Factor”



51

Challenges on Data Fusion

- **How to re-factor and dramatically shrink configuration?**
 - Router configuration state represents a huge un-normalized database -- a big problem for manageability and security
- **How to improve quality within and across databases?**
 - Robust matching in the presence of data errors (e.g., customer names)
- **How to rapidly integrate evolving schemas?**
 - Each new service creates its own schema (abstraction of reality)
- **How to reduce the drag from systems on networks?**
 - Updating systems (inventory, fault/performance, care) a formidable barrier to network feature and architectural change
- **How to raise the level of abstraction?**
 - Today’s networks data models tend to be “flat,” dealing with the minutia: router config templates, parameter lists

52