Understanding Large Internet Service Provider Backbone Networks

Joel M. Gottlieb
IP Network Management & Performance Department
AT&T Labs – Research
Florham Park, New Jersey

joel@research.att.com

Purpose of the lecture

- You have theoretical background
  - TCP/IP, routing protocols, layer 2, layer 3…
- You've heard of the big ISP's
  - WorldCom, Sprint, AT&T, AOL…
- How do these ISP networks work?
  - How is a large network structured?
  - What routing protocols do you use?
  - How does it all fit together?
  - What are some of the challenges in operating the network?

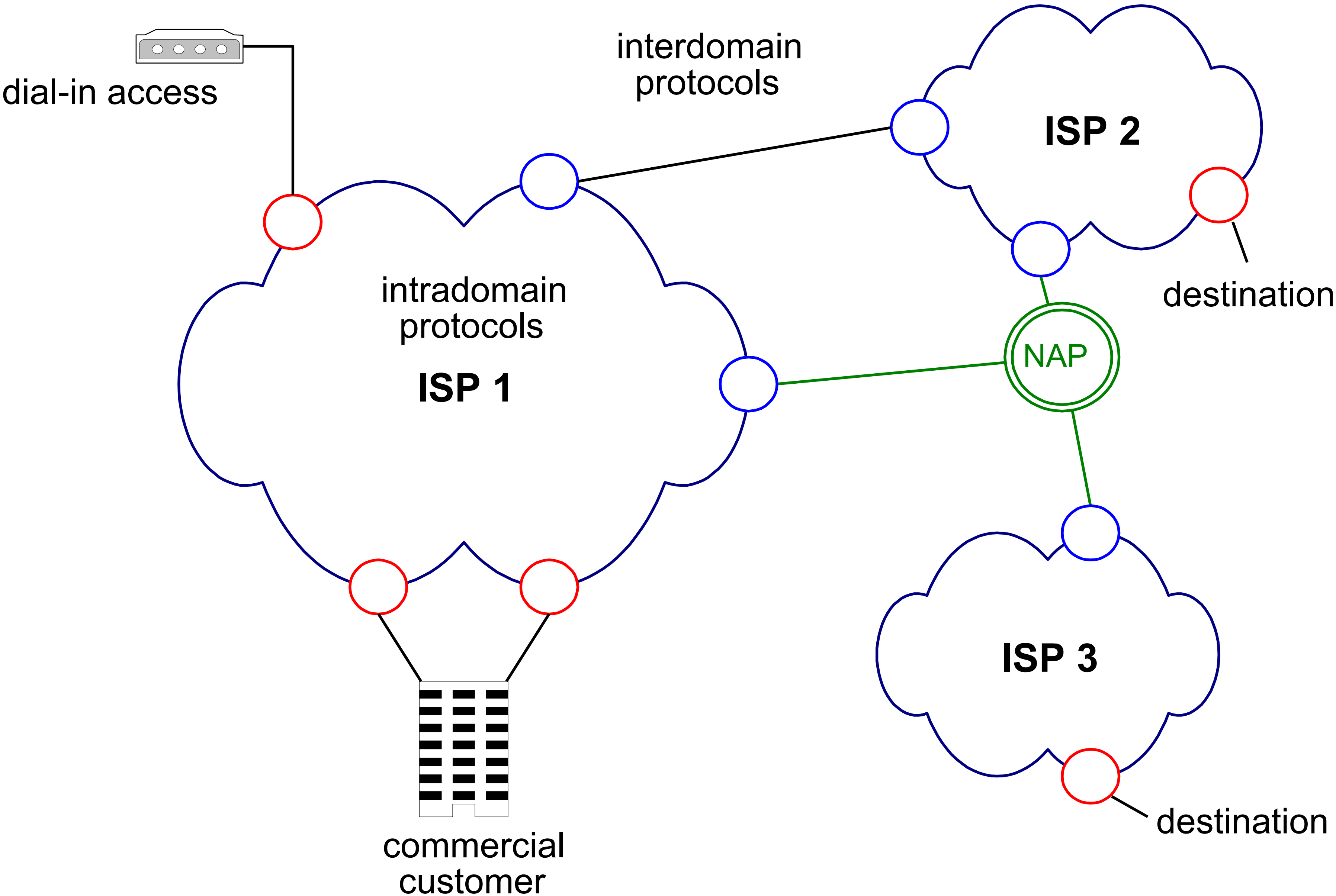
Outline

- Network Architecture
  - From a cloud to individual routers
  - Router hierarchy
  - Routing protocols
- Operational challenges
  - A variety of practical issues
- Focus on network configuration
  - The actual process of configuration
  - Configuration management and Netdb

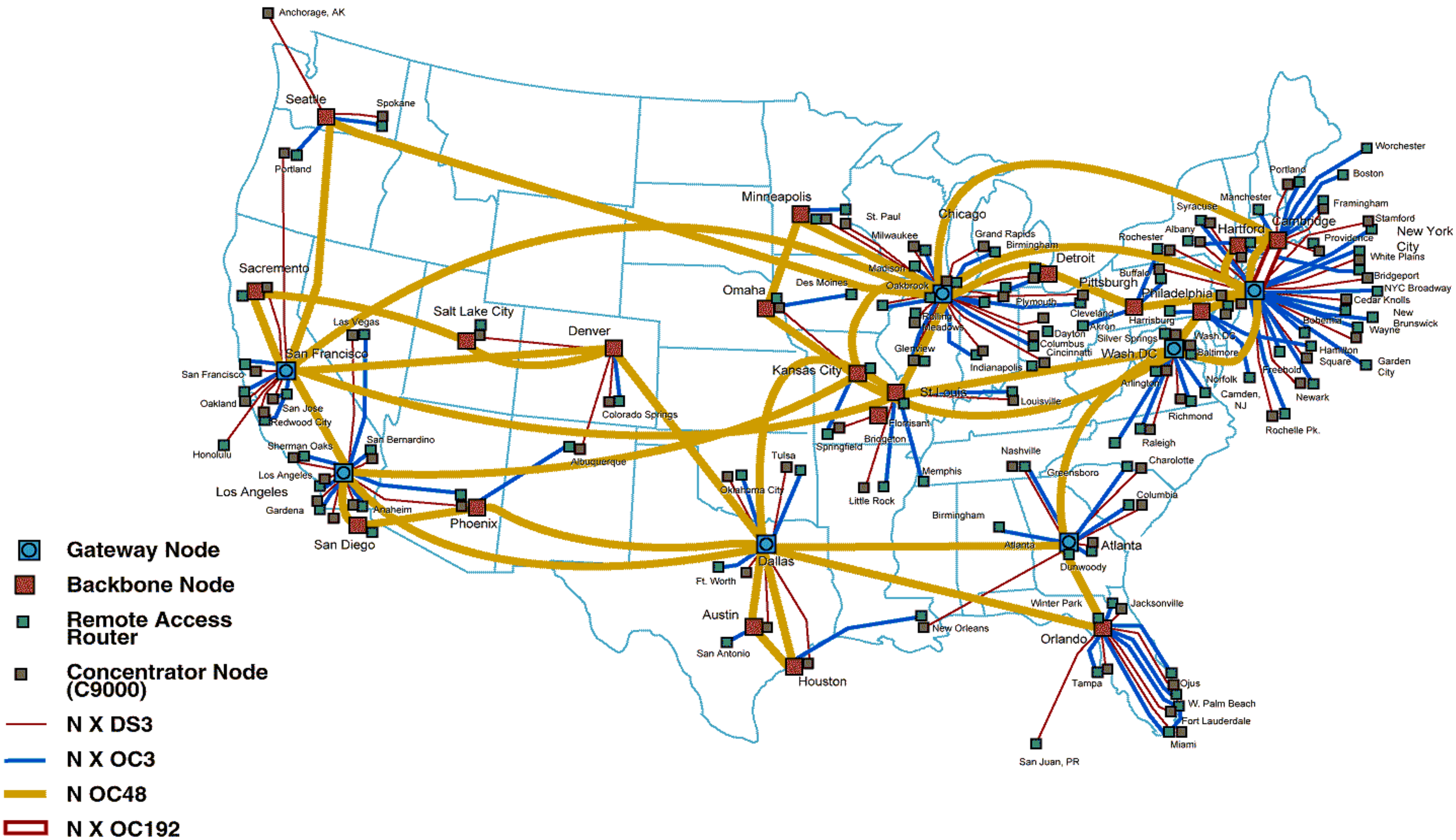
Review: Internet Architecture

- Divided into Autonomous Systems
  - Distinct regions of administrative control (~11,500)
  - Set of routers and links managed by a single "institution"
  - Service provider, company, university, …
- Hierarchy of Autonomous Systems
  - Large, tier-1 provider with a nationwide backbone
  - Medium-sized regional provider with smaller backbone
  - Small network run by a single company or university
- Interaction between Autonomous Systems
  - Internal topology is not shared between ASes
  - … but, neighboring ASes interact to coordinate routing

Review: Connections Between Providers



AT&T IP BACKBONE NETWORK



Inside the Cloud

- Multiple POPs (Points of Presence)
  - Like central offices in telephone network
  - Space in POP may be owned or rented
- Within a POP:
  - Multiple routers
  - Routers may have different responsibilities:
    - Access router
    - Backbone router
    - Internet Gateway Router
  - Routers w/different responsibilities may be same model

Internet Gateway Router

- Connections to neighboring Tier 1 providers
- Few interfaces (interface = slot, as in a PC, for plugging in cards and cables)
- Fast interfaces
- Homogeneous interface technologies
  - Mostly OC-x on Packet over SONET
- Limited filtering (filter = router feature to prevent unwanted traffic, by source or by destination)

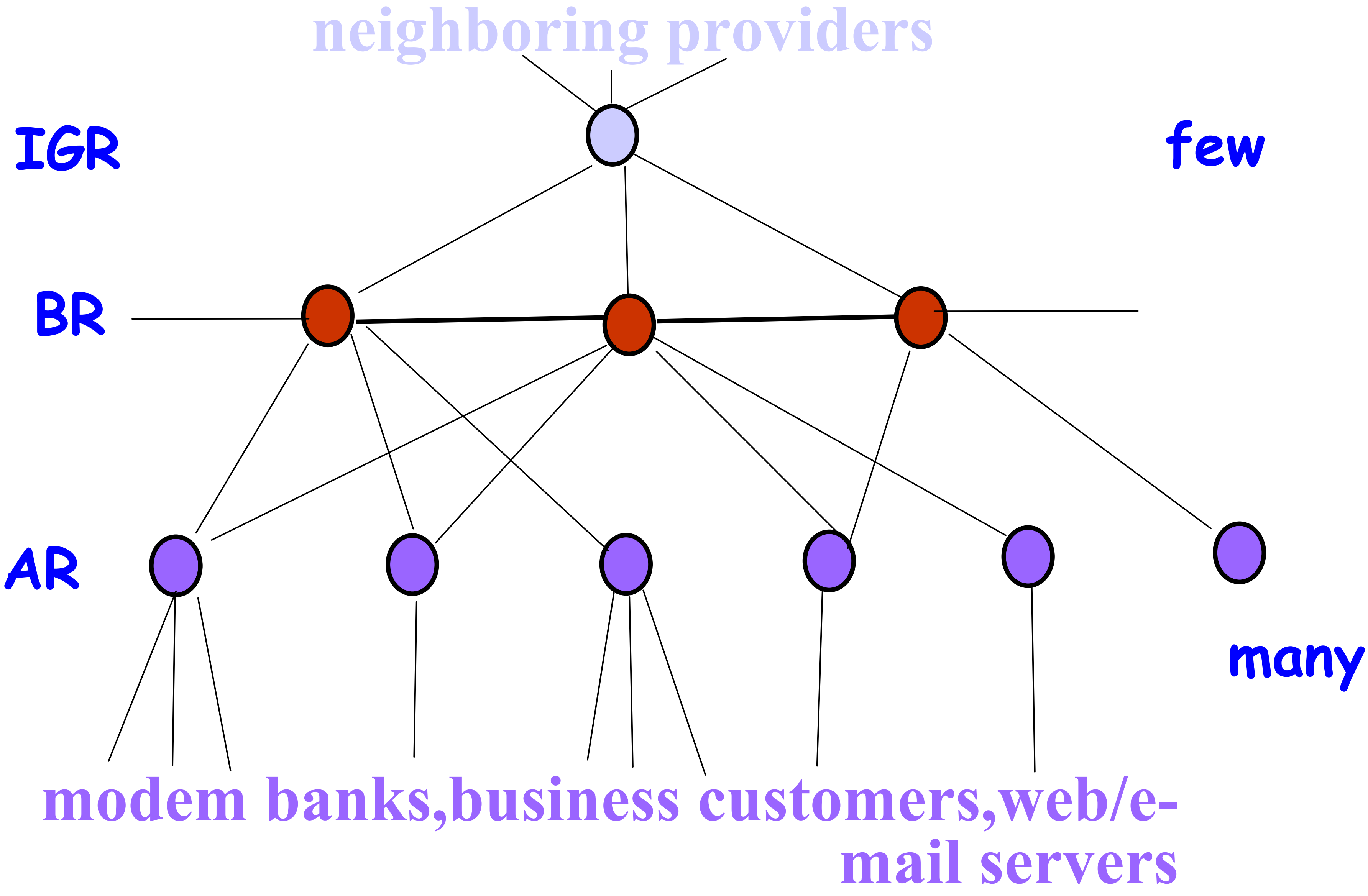
Backbone Router

- No connections outside the network
- Moderate number of interfaces
- Fastest interfaces
- Interface technologies depend on physical transport network
  - Mostly OC-x on Packet over SONET
- Very limited filtering
- Main purpose: move traffic through the network as fast and efficiently as possible
- "Big, fast and stupid"

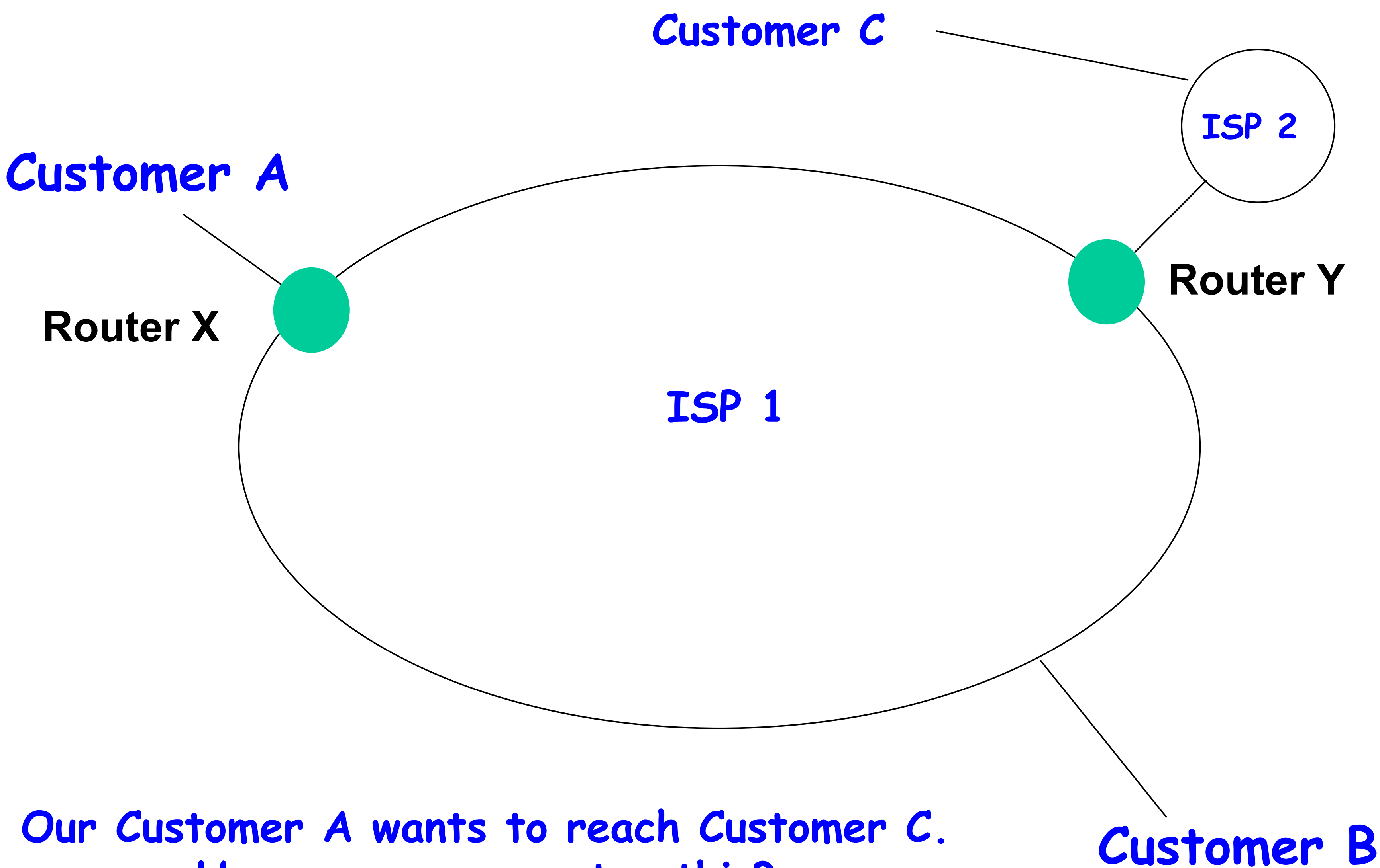
Access Router

- Many connections, to customers, modem banks…
- Connections only to backbone routers, not to each other
- Large number of interfaces
- Variety of interface speeds (depends on customer)
- Heterogeneous interface technologies (T1, T3, frame-relay…)
- Extensive filtering
- Packet marking (CoS/QoS)

The Router Hierarchy in a POP



Motivation: Routing Protocols



Our Customer A wants to reach Customer C.
How can we guarantee this?

What we need to do

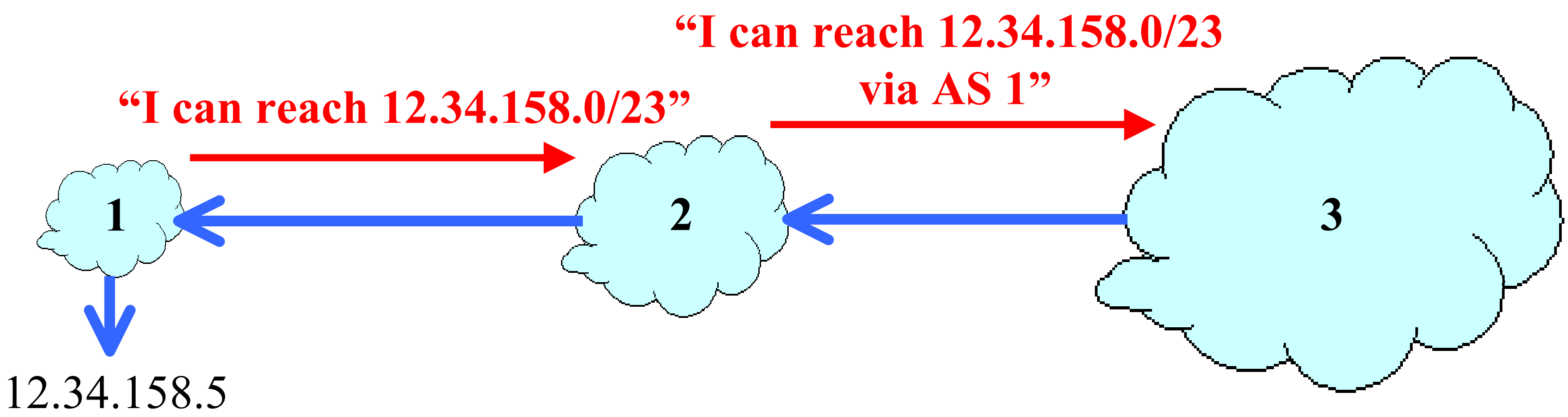
- Customer A sends us traffic destined for Customer C, which arrives at router X
- Router Y needs to know how to reach C
- Router X needs to know to go to Router Y to reach C
- Router X needs to know how to reach Router Y

Review: Routing Protocols

- BGP (Border Gateway Protocol)
  - Path-vector
  - Keep track of who knows how to reach prefixes
  - Send prefixes we know how to reach to others
- OSPF (Open Shortest Path First)
  - Link-state
  - Each router computes best paths to all destinations, based on link information it receives
  - Moy (1990)

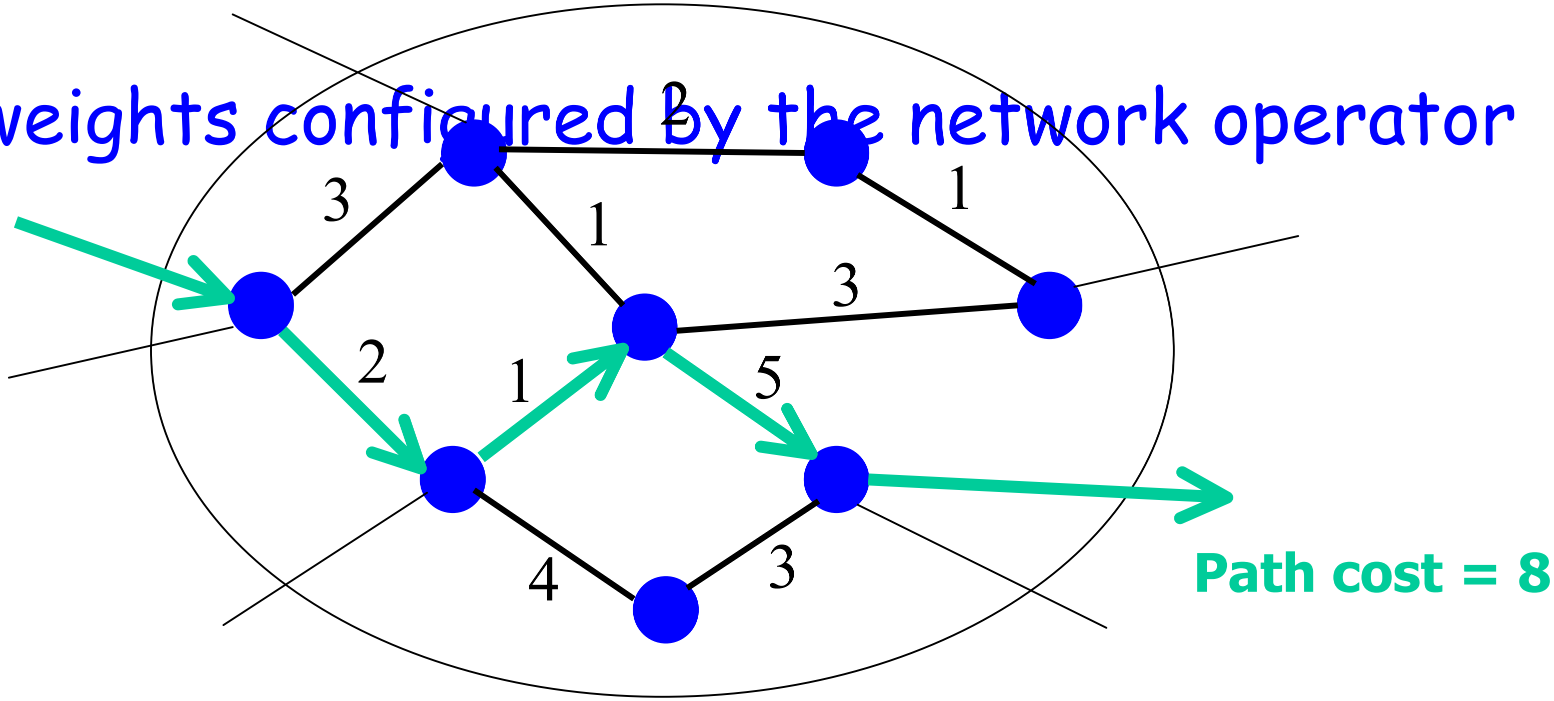
Border Gateway Protocol (BGP)

- ASes announce info about prefixes they can reach
- Local policies for path selection (which to use?)
- Local policies for route propagation (who to tell?)
- Policies configured by the AS's network operator



OSPF

- Routers flood information to learn the topology
- Routers determine "next hop" to reach other routers
- Path selection based on link weights (shortest path)
- Link weights configured by the network operator



Routing Design – first guess

- Use E-BGP as external routing protocol
  - Send and receive advertisements about specific prefixes
  - This will tell us Router Y how to reach C
- Use OSPF as internal routing protocol
  - All destinations inside the network are known and controlled by us, so full table can be built based on a common strategy
  - Provides lowest-cost path through the network
  - This tells Router X how to reach Router Y
- Inject all E-BGP routes into OSPF
  - Would then know how to reach C
  - Would this work for large ISP?

Modify Routing Design

- OSPF would be flooding a huge amount of E-BGP information
  - but would work on enterprise network, with small BGP table
- Use I-BGP to distribute E-BGP routes:
  - E-BGP information obtained from other networks
  - I-BGP relationship pairs (sessions) set up between routers inside network
  - OSPF continues to provide lowest-cost paths through the network

Getting to Customer C

- Router X receives packets with destination customer C
- eBGP information tells Router Y the external path for customer C
- iBGP information tells router X that router Y is an exit point from our network to C
- OSPF tells Router X which outgoing link is on the shortest path through the network to router Y
- Forwarding table combines BGP and OSPF to map destination directly to outgoing link on Router X

I-BGP scaling

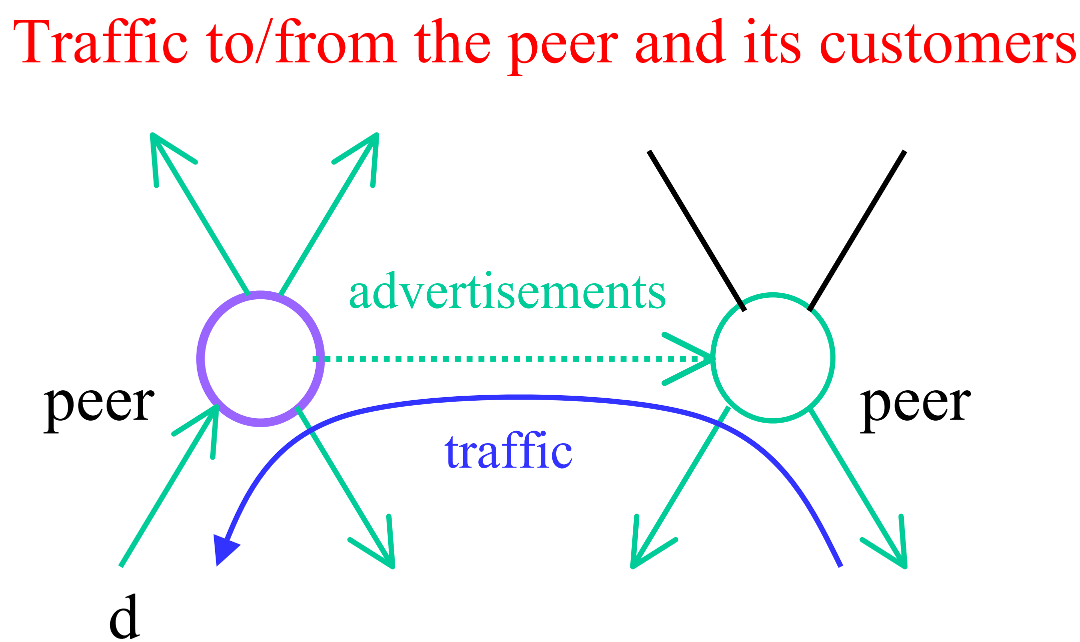
- Scaling problems with full-mesh IBGP
  - Must have sessions between each pair of routers
  - Configuration overhead; touch every router when adding a new router
  - Session overhead: N TCP connections per router with associated state
  - Memory overhead: storing routes learned from N sessions
- Solution: Route Reflector hierarchy
  - Route reflector: learns best routes from edge routers
  - Route reflector client: have I-BGP peering with one or two route reflectors

Peering options

- Peering
  - Tier I providers mutually exchange information about their sets of customers
  - Reduces cost and improves performance
  - Normally a contractual business relationship
- Public peering
  - Network Access Points (e.g., MAE East and MAE West)
  - Public location for connecting routers from different ISPs
  - Routers exchange both data and routing information
- Private peering
  - Private connections between two peers
  - Private peers exchange traffic between their customers
  - Private peers must exchange similar traffic volumes

Peer-Peer Relationship

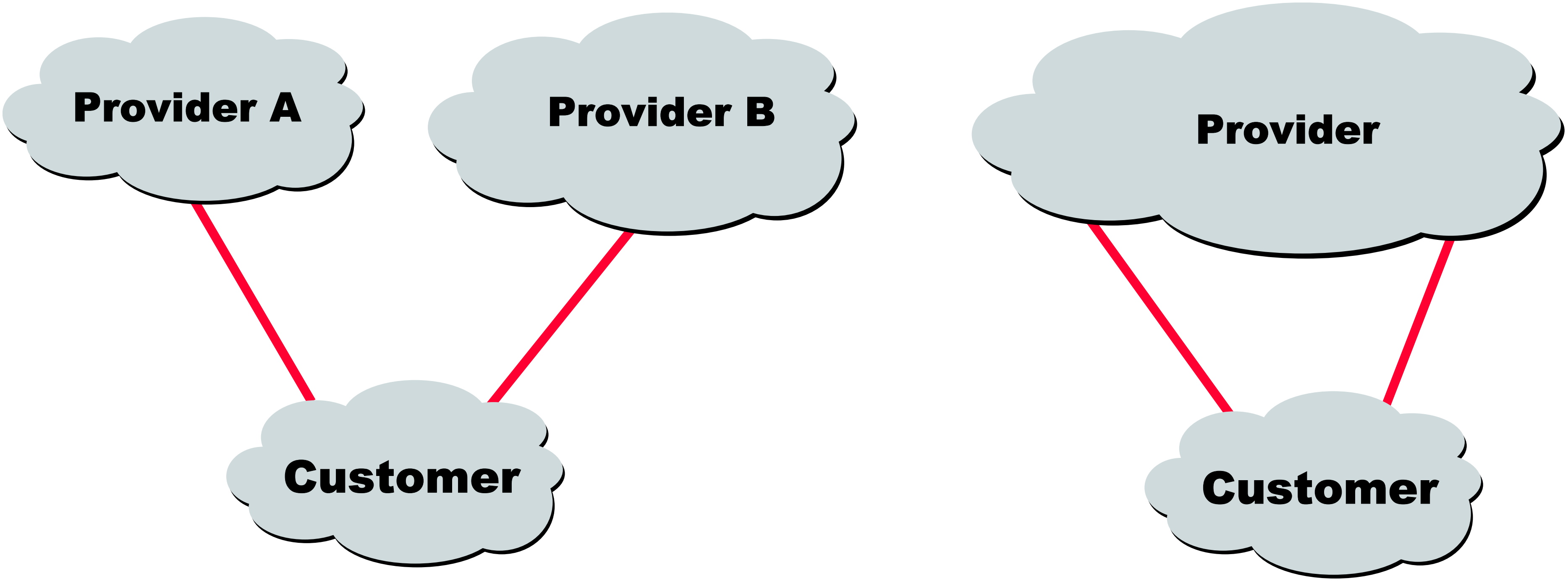
- Peers exchange traffic between their customers
- Free of charge (assumption of even traffic load)
- AS exports a peer's routes *only* to its customers



Also: how do we get customer routes?

- Customer pays us for service; forwarding traffic to them, sending their traffic out
- How do we know what their routes are?
- One option: static route
  - Customer's IP addresses are kept in static route on provider's access router
  - All traffic coming through provider's network uses static route to reach customer
  - Static routes are redistributed into BGP
  - Customer always sends traffic to Provider AR
- Is this good enough?

Multi-Homed Customer Speaking BGP



- Motivations for using BGP
  - Load balancing and fault tolerance
  - Send updates to influence how traffic enters
  - Apply routing policy to select outbound paths

Customers speaking BGP

- Some customers have their own ASN
  - Customer multi-homed to multiple providers
  - Customer with an ASN for historical reasons
- Some customers cannot get their own ASN
  - Limited # of ASNs available (16-bit number)
  - No new ASNs for single-homed customers
- Speaking BGP when customer has no ASN
  - Private ASNs in the range of 64,512 to 65,535
  - Dedicated ASN owned by provider

Some operational challenges

- Now you have a rough idea of the structure
- Operational challenges – what is it really like to operate a large ISP?
- Some topics to touch on briefly
  - Management issues
  - Provisioning issues
  - Capacity planning issues
  - Performance issues
  - Configuration issues (we'll focus here in a moment)

Practical Operational Challenges

- Increase in the scale of the network
  - Link speeds, # of routers/links, # of peering points
  - Large network has 100s of routers and 1000s of links (already discussed managing routing protocols)
- Significant traffic fluctuations
  - Time-of-day changes and addition of new customers/peers
  - Special events (Olympics) and new applications (Napster)
  - Difficult to forecast traffic load before designing topology
- Market demand for stringent network performance
  - Service level agreements (SLAs), high-quality voice-over-IP

Additional Practical Operational Challenges

- Increase in network capability & feature complexity
  - New services (Quality of Service, Virtual Private Networks)
  - New routing protocols (MPLS, multicast)

Practical Capacity Planning Issues

- Deciding whether to buy/install new equipment
  - What? Where? When?
- Examples
  - Where to put the next backbone router
  - When to upgrade a peering link to higher capacity
  - Whether to add/remove a particular private peer
  - Whether the network can accommodate a new customer
  - Whether to install a caching proxy for cable modems
- Requirements
  - Projections of future traffic patterns from measurements
  - Cost estimates for buying/deploying the new equipment
  - Model of the potential impact of the change (e.g., latency reduction and bandwidth savings from a caching proxy)

Focus on router configuration

- Uncertainty
  - Decentralized manual router configuration (telnet)
  - Databases of record must be kept accurate
- Complexity
  - Network policy not widely available/understood
  - Very deep subject matter (e.g., interfaces)
- Limited commercial tools for CM/debugging
  - Tools do not cover local conventions and policies
  - Tools typically lag behind product releases

Cisco Router Configuration Language (IOS)

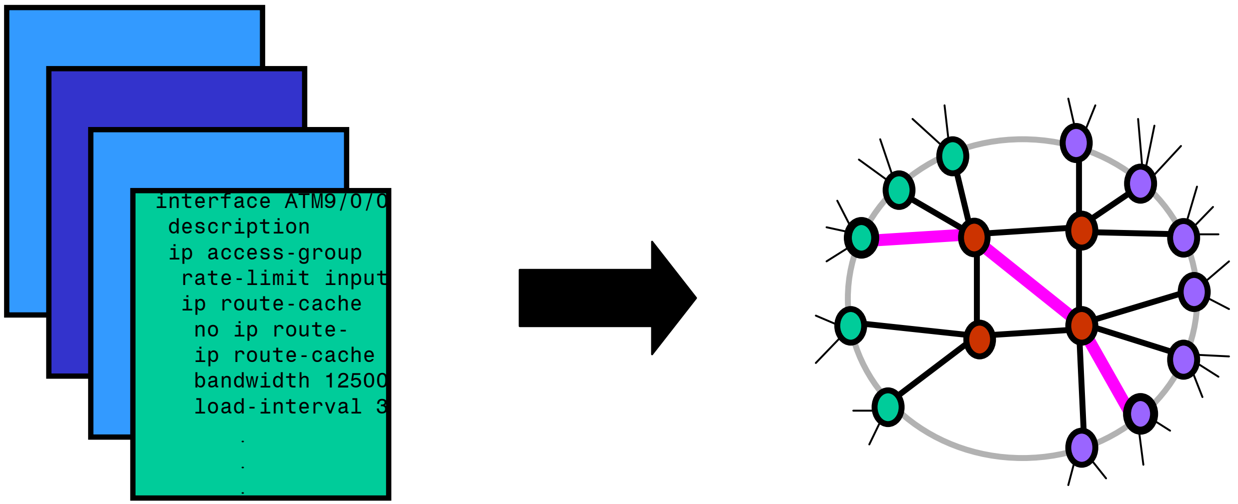
- Not user-friendly
  - Certifications offered (CCNE etc.)
  - Requires knowledge of low-level details (“assembly language”)
  - Many options for arguments
- Not a formal language
  - Simple grammar (keywords mixed with optional args)
  - Generally unstructured - very specific parsing required
- Presents a moving target
  - Multiple versions in marketplace (and in single network)
  - Command-set extended very often
- Substantial expertise required
  - 900+ unique statements in single network
  - Long files (AR 1000's of lines; BR and IGR 100's)

Example: Cisco Router Configuration File

- Language with hundreds of different commands
- Cisco IOS is a de facto standard config language
- Sections for interfaces, routing protocols, filters, etc.

```
version 12.0
hostname MyRouter
!
interface Loopback0
 ip address 12.123.37.250 255.255.255.255
!
interface Serial9/1/0/4:0
 description MyT1Customer
 bandwidth 1536
 ip address 12.125.133.89 255.255.255.252
 ip access-group 10 in
!
interface POS6/0
 description MyBackboneLink
 ip address 12.123.36.73 255.255.255.252
 ip ospf cost 1024
!
router ospf 2
 network 12.123.36.72 0.0.0.3 area 9
 network 12.123.37.250 0.0.0.0 area 9
!
access-list 10 permit 12.125.133.88 0.0.0.3
access-list 10 permit 135.205.0.0  0.0.255.255
ip route 135.205.0.0  255.255.0.0 Serial9/1/0/4:0
```

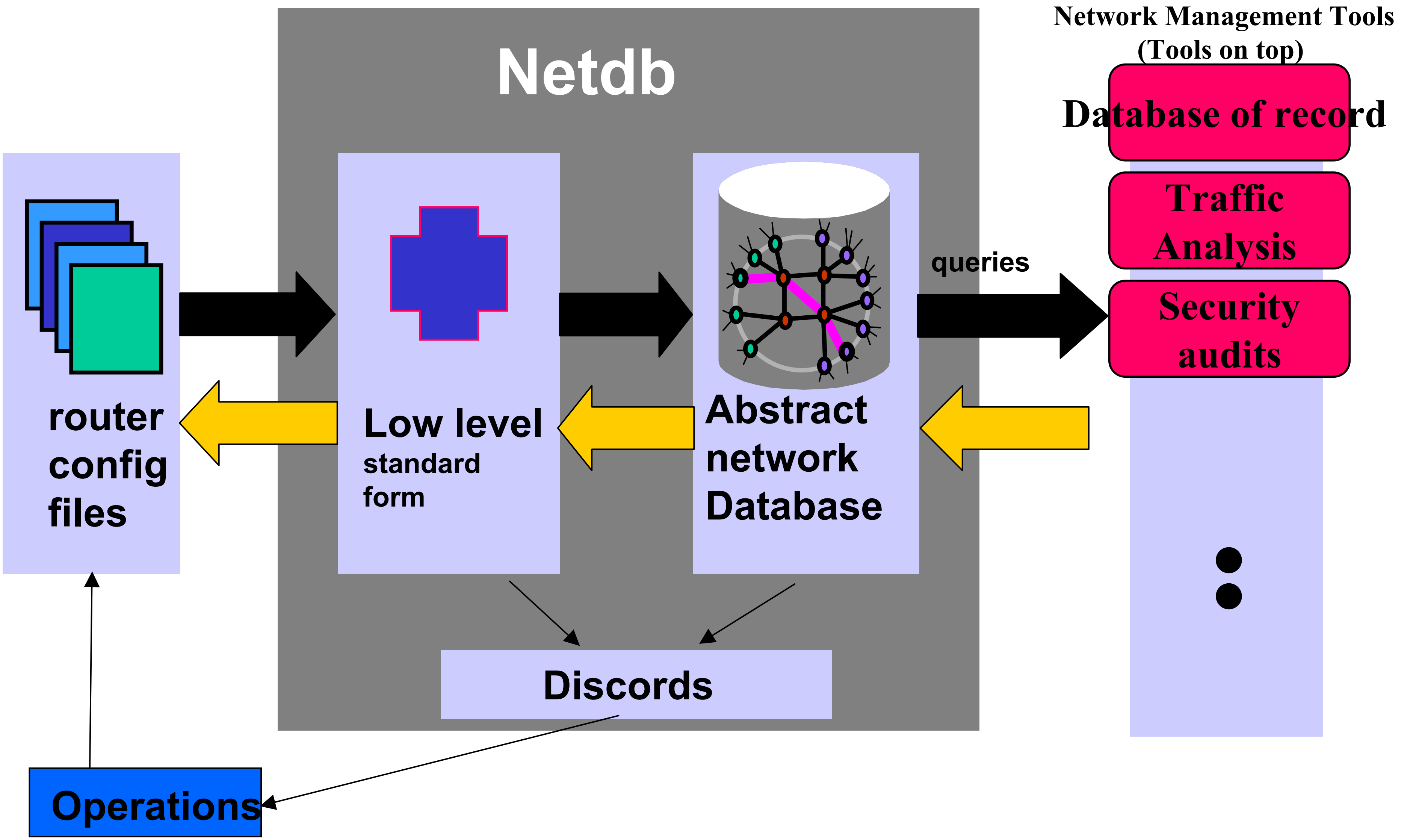
Netdb: Router configuration files to Network abstraction



Router configuration files | **Network abstraction**

Using Netdb, an accurate network view can be stored in a database, permitting querying, error checking, and specialized reporting

Netdb Architecture



Netdb and the CBB network

- Queries developed for specific topics
  - Which router cards? Which router models?
  - Are security features configured properly?
  - Are BGP relationships configured properly?
- Results available daily
  - Operations groups note *discords*, may fix them
  - Capacity Planning may use topology queries
- Many research efforts enhanced
  - Traffic engineering, traffic analysis
  - Developing expertise base in configuration management

Tracking the State of the Network

- Network management groups
  - Tier 1: Customer care
  - Tier 2: Individual network elements
  - Tier 3: Network-wide view
- Databases
  - Customers (name, billing info, IP addresses, service,...)
  - Network assets (routers, links, configuration,...)
- Data from the operational network
  - Router configuration files (commands applied to router)
  - Fault data (link/router failures, BGP session failures,...)
  - Routing tables (dumps of BGP and forwarding tables)

Conclusions

- Large IP network is very complex; requires both expertise and personpower to manage carefully
- Router configuration is a very big subject (High schools now teaching Cisco configuration)
- Management issues often as difficult as technical issues
- Small changes in network configuration can have serious consequences; work must be done very expertly
- Performance management is a recent subject and just developing
- Large IP networks are *very* complicated!
- New technologies coming all the time (MPLS - Tim Griffin on Thursday)