Coarse-Grained Traffic Analysis in ISP Networks

A Router-Based Approach

Christian Martin
Verizon

Traditional ISP Traffic Analysis

- Resource Utilization
  - CPU, port utilization, memory
- Traffic Characterization
  - Flow information (src prt, dst if, src addr, etc)
  - What kind of traffic is flowing across my network? In what ratios and magnitudes? How do new application phenomena affect traffic dynamics (eg. Napster, Everquest)
- Network Performance
  - RTT, Packet Loss
  - Are there problem areas in my network that I should be addressing?

Modern ISP Traffic Analysis

- Node-Node coarse-grained volumetric statistics.
  - Traffic Engineering, metric assignment, billing
- ISP<->ISP and ISP<->{ISP_i,ISP_{i+1},…, ISP_n} traffic volumes
  - Peerseeking
- Traffic characterization
  - Topological and routing data correlation
  - Need routing table to really understand traffic patterns. Otherwise all we know is *what*, not *why* or *how*.
- Network Performance
  - Temporal Analysis – (Jitter, convergence times, etc)
  - SLAs
- This analysis is hard!

Modern Traffic Analysis (cont'd)

- Traditional Analysis is denotative (literal)
  - The information we gather has little to no implied meaning
    - Bits/sec on this interface
    - Flows/sec of http between these IPs
  - In order to infer meaning, we need one or more *heuristics*
  - One such heuristic is the routing table
- Modern Analysis is connotative (implied)
  - The information we need to gather has lots of implied or inferred meaning
    - Traffic between POP-A and POP-B
    - Relative increase in ICMP echo-reply traffic between two ASNs

Routing Data As a Heuristic

- Modern Analysis requires an understanding of the routing state of a network
  - Otherwise, the empirical data has no meaning (connotation)
- Gathering routing information offline and then correlating it to traffic data:
  - Difficult (sensor fusion)
  - Costly (high speed interfaces not cheap)
  - Disruptive (need to insert probes)
  - Inaccurate in the temporal dimension (export)
- The router has all the 'information' that is needed to perform the analysis in real-time

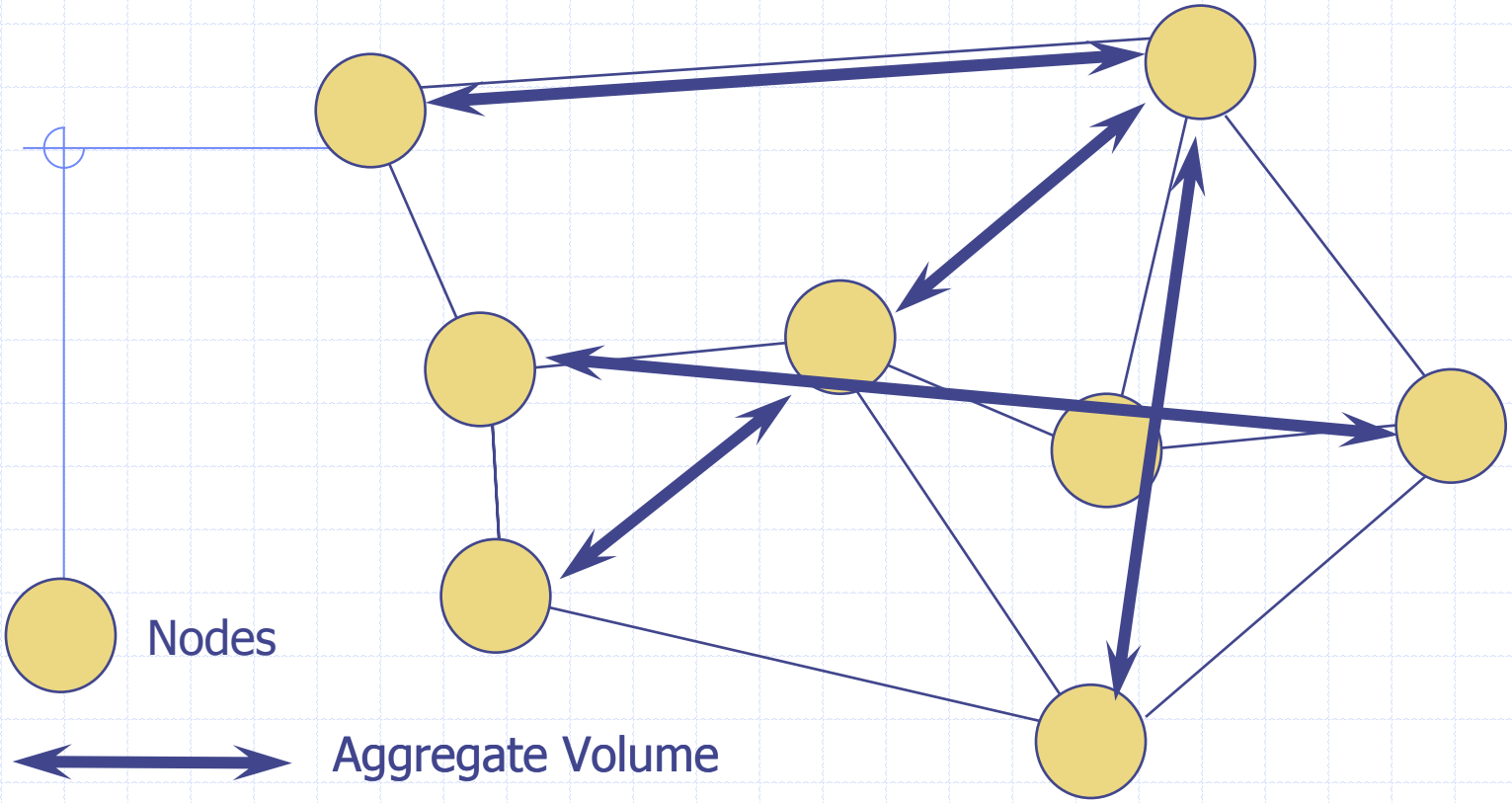
The Router as a Traffic Analyzer

- Modern routers forward data using advanced hardware
  - Legacy software routers were smart, but slow
  - Original ASIC technology was dumb and fast
  - Newer ASICs/PPEs are becoming smart and fast
- Forwarding engines are derived from routing data
  - While programming the FIB, we may be able to store additional heuristics
    - Who (Next Hop), What (POP, peer, cust), When (timestamp), Where (AS, router), How much (volume), Why (is my policy accurately applied?)
  - Create buckets, match flags, and increment counters
  - Store in memory for SNMP/other processing, rather than costly export to spinning media for analysis

Applicability

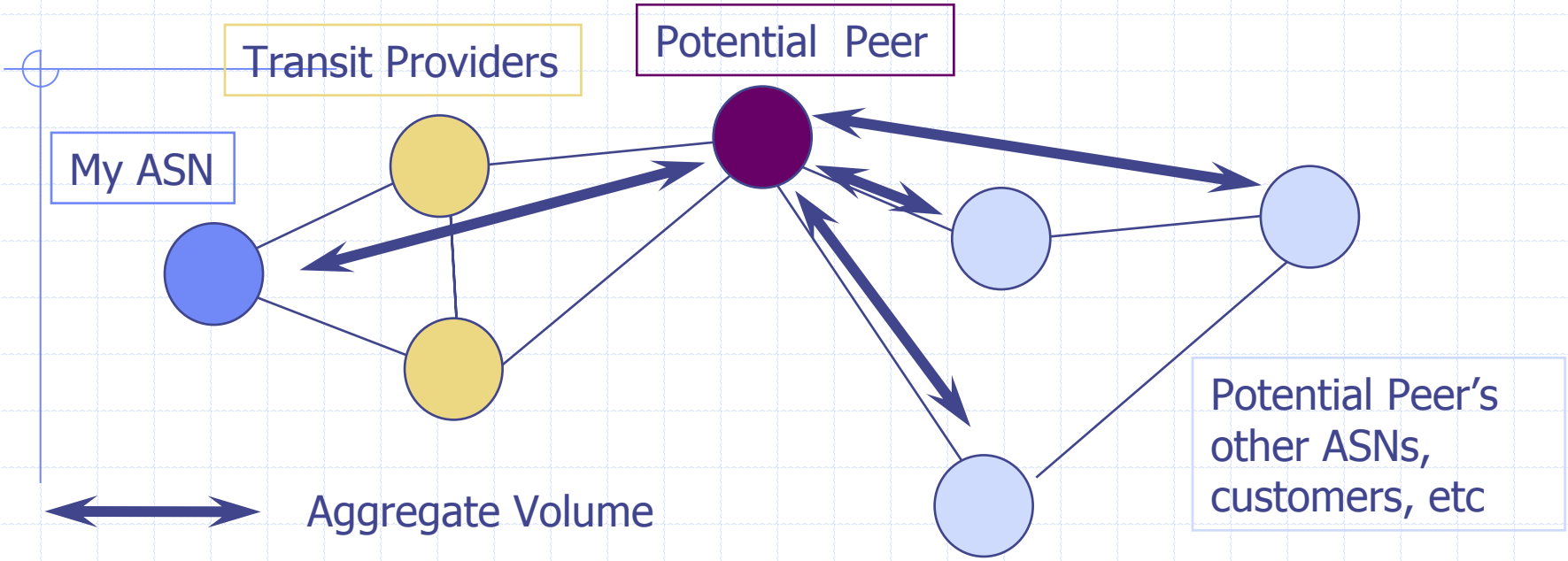
- Coarse-Grained Node-to-Node Analysis
  - How much traffic of a certain class to a certain class, derived from BGP (or other) information
- Peer-Seeking
  - How much traffic to and from a *set* of AS-PATHs, with additional flags if needed
- Fine Grained Analysis (future)?
  - Real-time analysis of data patterns, router-based heuristics, inferencing, reacting?
  - Anything that requires routing information for correlation
  - Match BGP info and port (NNTP between a set of AS's)

Coarse-Grained Node-Node Analysis



- Here, a node is a router, or cluster, or POP, or an entire AS
- Use BGP Next-Hop or AS-PATH to define class
- Program FIB with class, increment when a match occurs
- Can be done for source match or destination match
- Simple way to derive arbitrary demand matrices

Peerseeking



- Determine heuristically (or thru flow info) what ASNs you are interested in
- Create AS-PATH matches, bucketize and count
- Netflow processing for AS information is expensive
  - Must Sample - sampling accuracy is questionable

Fine Grained Analysis

- With new imbedded hardware being developed, much of our analysis will be performed on the router
  - Similar to Aggregated Netflow
- With real time data and robust inferencing software, we may have:
  - Anomaly detection (DoS Attack)
  - Traffic pattern recognition and adjustment
- By offloading analysis to hardware in the router, we ensure temporal accuracy and reduce the system processing and storage burden on offline tools
  - Data is transient and self-similarities exist. Why store it?

Caveats

- BGP Asymmetry
  - Intradomain should not be a problem
  - Interdomain is a problem, particularly as connectivity becomes rich
  - More useful for transit buyers looking to offload some transit expense
- No *a priori* approach to rule setting
  - Need to know what to look for beforehand
- Requires hardware/software upgrades
  - Ask your vendor when/where this capability is supported

Conclusion

- Routers have all the data in real time
- ASICs/PPEs are advancing along with traffic volume on (most) networks
- Inline probes are intrusive and expensive
- Offline collection and analysis of large data sets is costly and difficult
- Features are available today on modern hardware
- Complimentary to flow-based tools, and of course, MIB-II and ifMIB polling