Internet traffic measurements

Renata Teixeira (Inria)

Why measure traffic?

- Performance analysis
- Anomaly and intrusion detection
- Network engineering

Traffic at different granularities

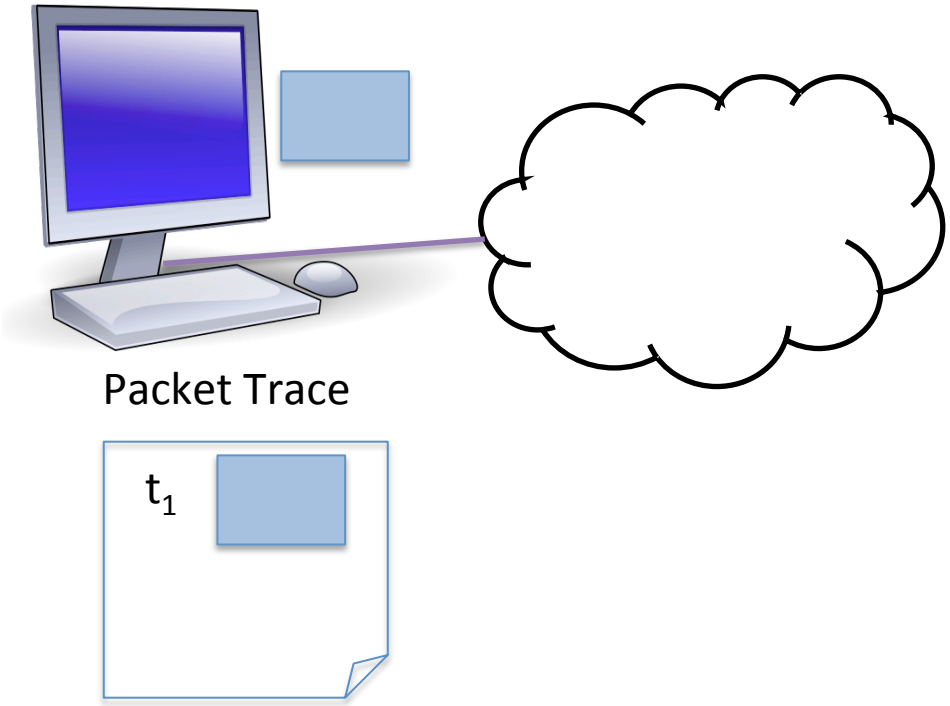
- IP-level packets
  - Capture per-packet information
- Flows
  - Statistics of packets grouped into flows
- Network interface
  - Statistics of packets that traverse a network interface

Outline

- Motivation and definitions
- Tools for measuring traffic
  - Packet capture
  - Interface counts
  - Flow capture
- Traffic matrix
- Trace anonymization
- Summary

Packet capture on end systems

- Basic method
  - Capture and record packets passing through an interface



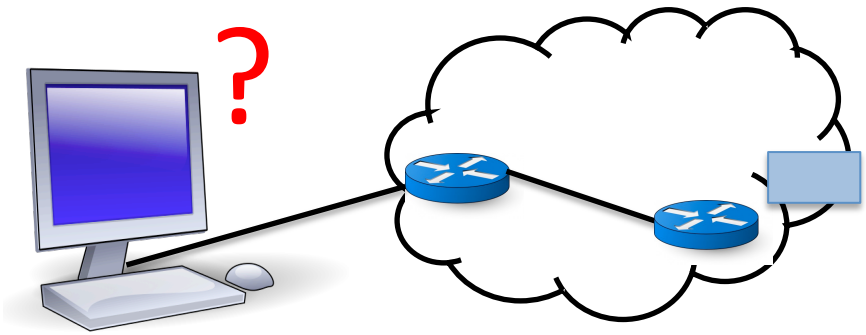
Tools

- tcpdump
  - Command-line packet capture
- libpcap
  - C/C++ library for packet capture
- Wireshark
  - Packet capture and analysis

Possible measurement artifacts

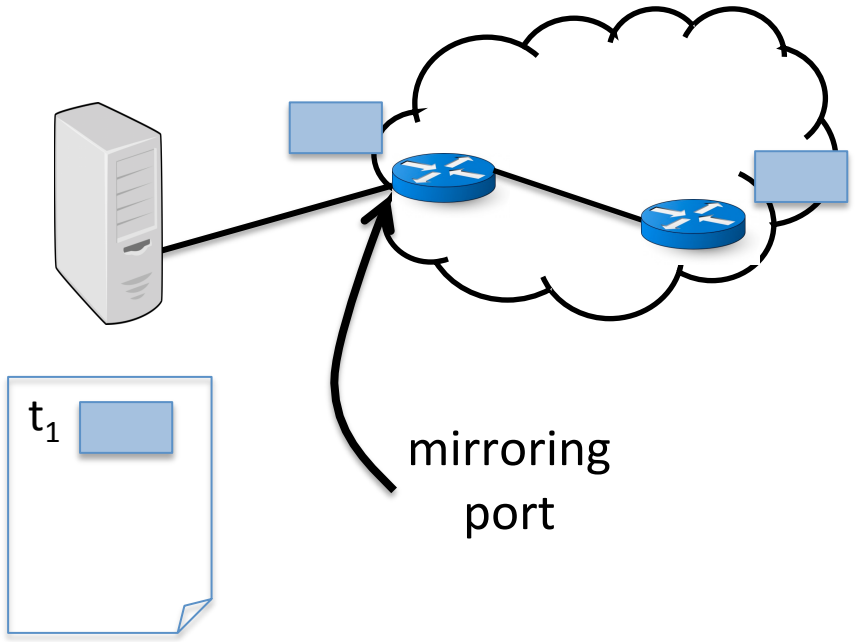
- Dropped packets are common under high utilization
  - Inspect report of dropped packets
- Other less frequent artifacts
  - Fail to report drops
  - Falsely report drops
  - Duplicate packets
  - Re-ordered packets
  - Misfilter

How to capture packets on point-to-point links?



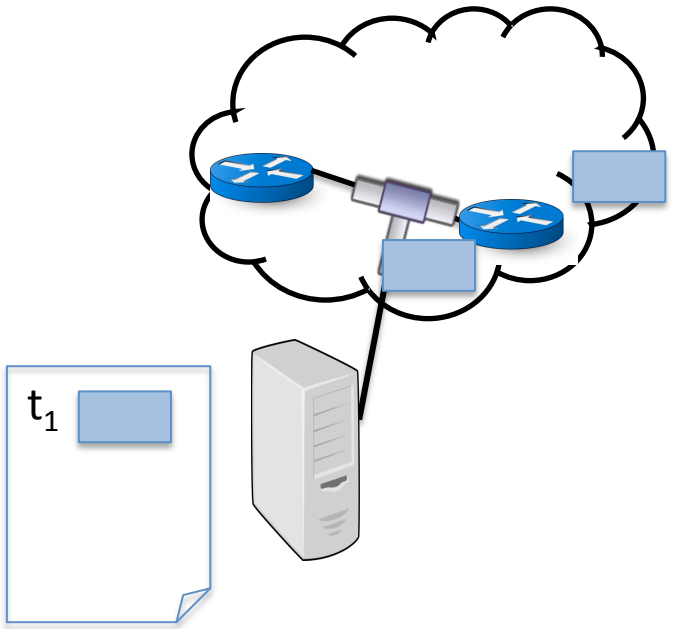
Port mirroring

- Basic method
  - Copies packets from one or more ports to a mirroring port
  - Run packet capturing tool on host connected to mirroring port



Network Tap

- Basic method
  - Electrical or optical splitter on monitored link
  - Monitoring host with specialized network interface and interface driver



Comparison

Port mirroring

- Pros
  - Easy to setup
  - Low cost
- Cons
  - Hardware and media errors are dropped
  - Packets may be dropped at high utilization

Tap

- Pros
  - Monitor all packets
  - Eliminates risk of dropped packets
- Cons
  - Expensive

High-speed capture with commodity hardware

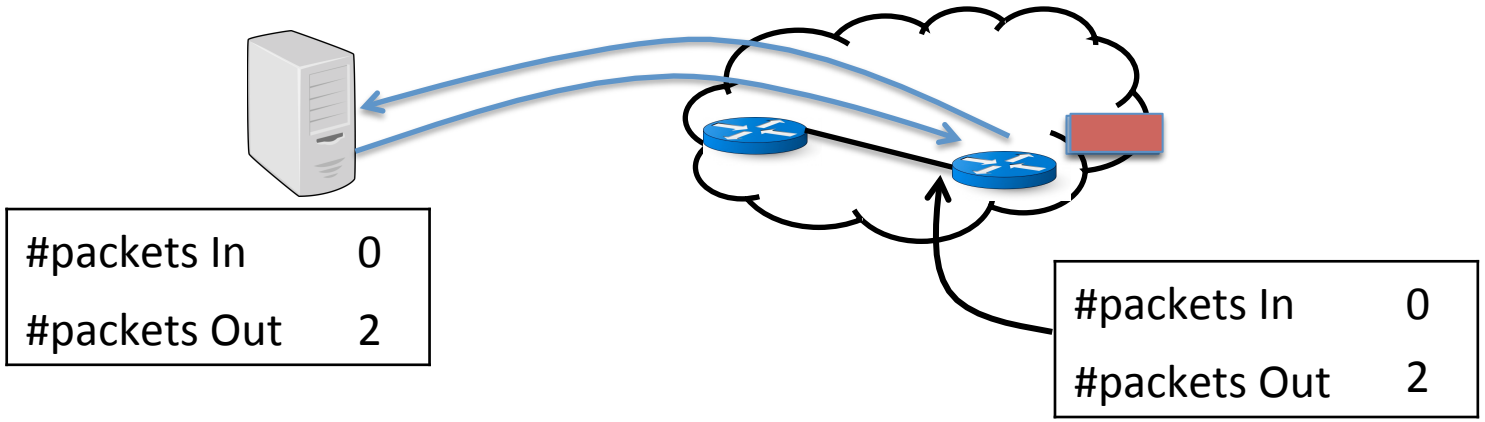
- Key idea
  - Direct access to NIC (i.e., bypass kernel)
  - Parallelism
- Tools
  - TStat
  - ntop
  - WAND

Outline

- Motivation and definitions
- Tools for measuring traffic
  - Packet capture
  - Interface counts
  - Flow capture
- Traffic matrix
- Trace anonymization
- Summary

Interface counts

- Basic method
  - Routers log simple statistics (bytes/packets)
    - Total values since interface initialized
  - Request statistics using SNMP (MIB-II MIB)



Example properties

- Number of In/Out bytes (total, unicast, non-unicast)
- Number of In/Out packets (total, unicast, non-unicast)
- Number of In/Out discarded/corrupted packets

Interface counts: Pros and Cons

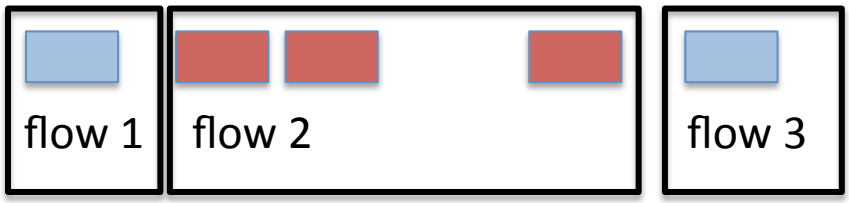
- Pros
  - Supported on all networking equipment
  - Little performance impact on routers
  - Little storage needs
- Cons
  - Missing data (SNMP uses UDP)
  - Polling makes it hard to synchronize data from multiple interfaces
  - Coarse-grained measurements

Outline

- Motivation and definitions
- Tools for measuring traffic
  - Packet capture
  - Interface counts
  - Flow capture
- Traffic matrix
- Trace anonymization
- Summary

IP Flows

- Set of packets with common properties
  - Definition can vary
    - Traditional 5-tuple: src IP, dst IP, src port, dst port, protocol
    - Packets from one ingress to an egress point
- Packets that are "close" together in time
  - Maximum spacing between packets (e.g., 15 sec, 30 sec)

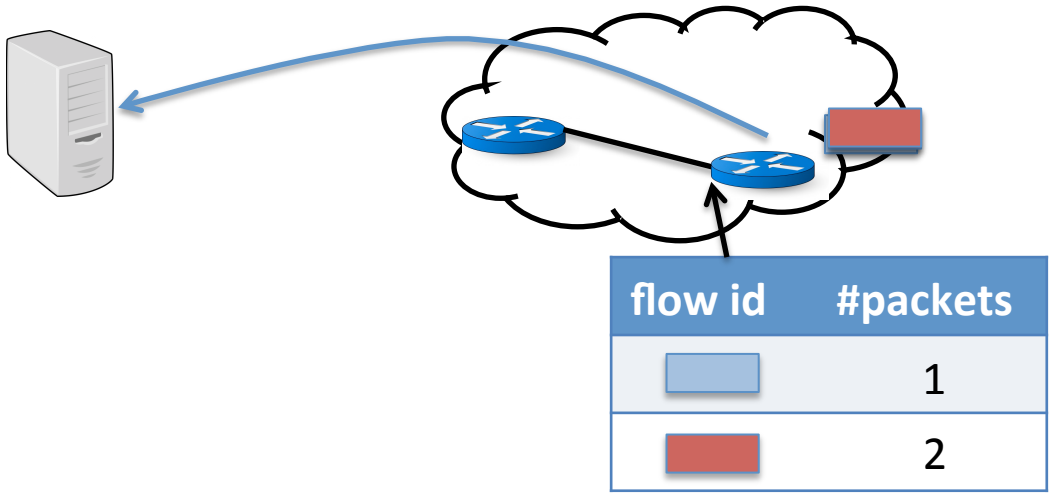


Flow ≠ application session

- Application session may be composed of multiple flows
- Packets in application session may not follow same links
- Hard to measure application session inside the network

Capturing flow statistics in routers

- Basic method
  - Specify set of properties that define a flow
  - Router log statistics per flow (flow records)
  - Push flow records to collecting process (IPFIX)



Flow records: Flow identifier

- Packet header information
  - Source and destination IP addresses
  - Source and destination TCP/UDP port numbers
  - Other IP & TCP/UDP header fields (e.g., protocol, ToS bits)
- Routing information
  - Input and output interfaces
  - Source and destination IP prefix (mask length)
  - Source and destination autonomous system numbers

Flow records: Flow properties

- Aggregate traffic information
  - Start and finish time of the flow (time of first & last packet)
  - Total number of bytes and number of packets in the flow
  - TCP flags (e.g., logical OR over the sequence of packets)

Packet Sampling

- Packet sampling before flow creation
  - 1-out-of-m sampling of individual packets (e.g., m=100)
  - Creation of flow records over the sampled packets
- Reducing overhead
  - Avoid per-packet overhead on (m-1)/m packets
  - Avoid creating records for a large number of small flows
- Increasing overhead (in some cases)
  - May split some long transfers into multiple flow records
  - … due to larger time gaps between successive packets

Tools

- In-router capture
  - Cisco NetFlow
  - Juniper JFlow
- Collection and post-processing
  - Flow-tools
  - ntop

Flow monitoring: Pros and Cons

Pros

- More details about traffic compared to counters
- Lower measurement volume than full packet traces
- Available on high-end line cards (Netflow, Jflow)
- Control over overhead via aggregation and sampling

Cons

- Less details than packet capture
  - No individual packet arrival times
  - No information on packet content
- Not uniformly supported (getting better with IPFIX)
- Computation/memory requirements for the flow cache

Using the traffic data in network operations

- Interface counts: everywhere
  - Tracking link utilizations and detecting anomalies
  - Generating bills for traffic on customer links
  - Inference of the offered load (i.e., traffic matrix)
- Packet monitoring: selected locations
  - Analyzing the small time-scale behavior of traffic
  - Troubleshooting specific problems on demand
- Flow monitoring: selective, e.g,. network edge
  - Tracking the application mix
  - Direct computation of the traffic matrix
  - Input to denial-of-service attack detection

Outline

- Motivation and definitions
- Tools for measuring traffic
  - Packet capture
  - Interface counts
  - Flow capture
- Traffic matrix
- Trace anonymization
- Summary

Traffic matrix: Definition

– Representation of traffic volume flowing from sources to destinations

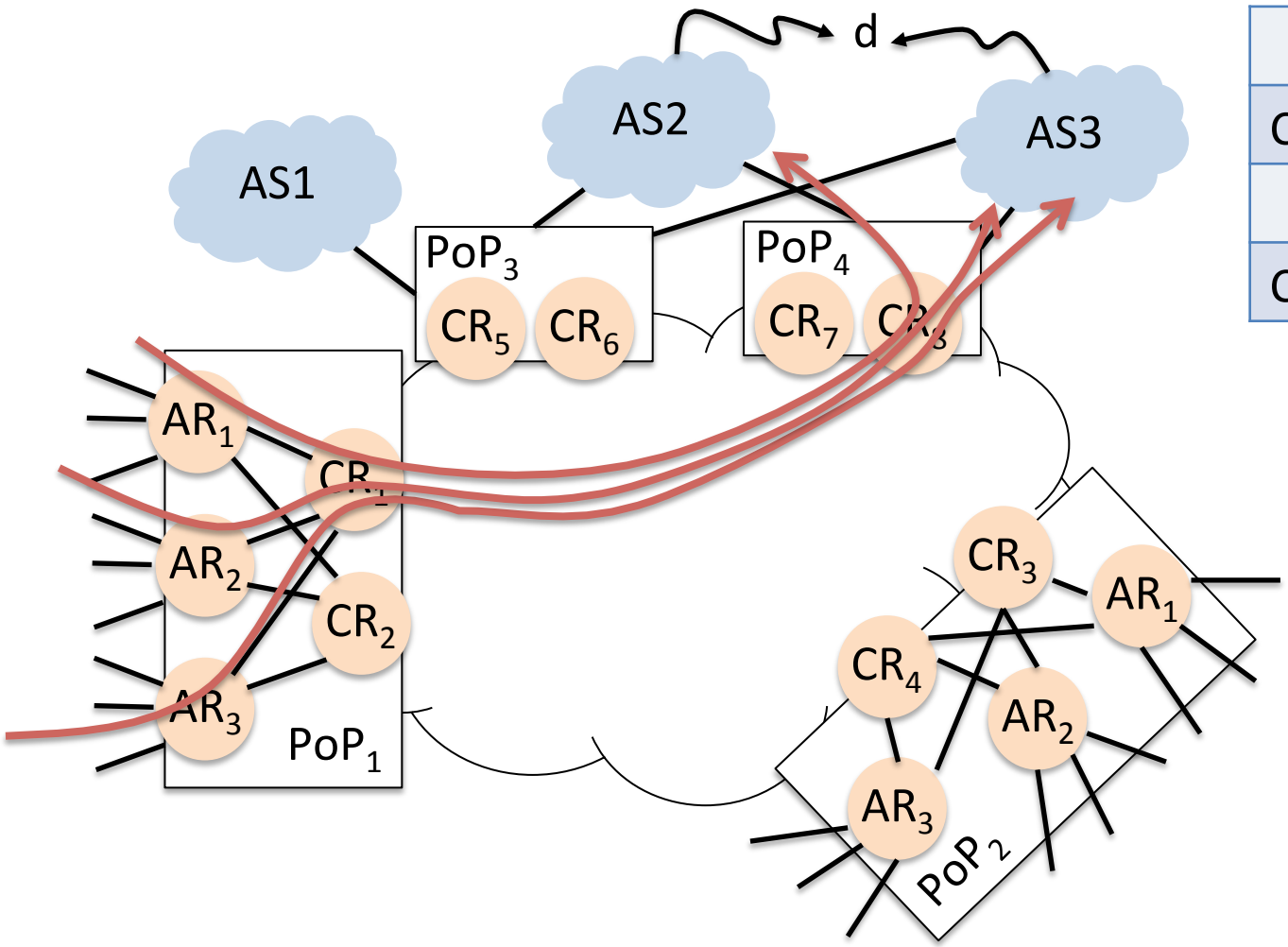
- Links
- Routers
- Points of Presence (PoPs)
- Networks

- Bytes
- Packets
- Flows, etc.

Usage

- Capacity planning
- Traffic engineering (IGP and BGP)
- Billing
- Peering analysis
- Anomaly detection
- Design of new protocols

Ingress router to egress router matrix



	CR_1	...	CR_8
CR_1			
...			
CR_8			

Measuring the traffic matrix

- Packet capture
  - Gives the most detailed view of traffic
  - But, expensive and high collection overhead
- Flow capture
  - Enough to build traffic matrix
  - Lower collection overhead (in particular with sampling)
- Interface counts
  - Cannot directly measure traffic matrix, must estimate
  - Lowest overhead, widely available

Outline

- Motivation and definitions
- Tools for measuring traffic
  - Packet capture
  - Interface counts
  - Flow capture
- Traffic matrix
- Trace anonymization
- Summary

Benefits of sharing data

- Good scientific practice
- Get others to work on relevant problems
- Learn from analysis of others
- Get broader view

But, packet traces contain lots of sensitive information

- Headers
  - Connection endpoints: who is talking to who; sites visited
  - Protocol, ports: applications used
- Payload
  - Visited content
  - Passwords, etc.

Solution: Anonymization

- Process to sanitize data to ensure anonymity
  - Absence of identity
  - Prevent others from linking identity to actions of an individual
- Packet trace anonymization tools
  - tcpdpriv, ipsumdump, ip2anonip, Crypto-PAn, PktAnon

Anonymizing payload

- Payload contains most sensitive information
  - Better if removed completely
  - If not possible, get minimum necessary
    - E.g., HTTP host better than full URL

Anonymizing packet headers

- Packet headers can be shared with care
  - MAC addresses
    - Potential to link records with the same MAC across datasets
  - IP addresses often need to be anonymized
  - IP addresses appear in other parts of the packet
    - IP options (e.g., record route)
    - ICMP/DNS packets

Outline

- Motivation and definitions
- Tools for measuring traffic
  - Packet capture
  - Interface counts
  - Flow capture
- Traffic matrix
- Trace anonymization
- Summary

Summary

- Packet capture
  - Detailed per-packet measurements; high overhead
- Interface counts
  - Coarse measurements per link; low overhead
- Flow capture
  - More details than link counts, less than packet captures
  - Medium collection overhead controlled with sampling
- Traffic matrix
  - Measured from flow capture
- Trace aonymization is key for data sharing

Questions?