Internet Traffic Characterization

George C. Polyzos

Computer Systems Laboratory
Department of Computer Science & Engineering
University of California, San Diego
La Jolla, CA 92093-0114

`polyzos@cs.ucsd.edu`
`http://www-cse.ucsd.edu/~polyzos`
`Tel.: 619-534-3508, FAX: 619-534-7029`

Outline

- history
- NSFNET traffic
- IP flows
- current architecture
- outlook

Collaborators

- Hans-Werner Braun
- Kimberly Claffy
- Nancy Groschwitz
- Henry Sariowan
- Jarom Smith

polyzos@cs.ucsd.edu

Historical Overview

- ARPANET (1969-89)
- ... Internet ... chaos
- NSFNET backbone (1987-1995)
  - 56 kb/s
  - T1: 07/88-12/92
  - T3: 1990-04/95
- ISPs, NAPs, vBNS, etc.
- Note: historical response to growth and resource contention
  - ignore until utilization limit approached
  - then increase bandwidth...
    - often, uniformly over the infrastructure

Network Characterization Taxonomy

- traffic characteristics
  - at various levels/layers
  - network layer
    - locality ("geographic")
    - burstiness
    - packet size
  - transport/application
    - traffic cross section
- performance metrics
  - reachability
  - capacity/max. throughput
  - utilization
  - delay, jitter
  - loss
  - router uptime

NSFNET Operational Statistics

- goals
  - reports to NSF
  - (very coarse granularity) performance tracking
- considerations
  - impact of measurements on network performance
  - post-capture resource demands (analysis, storage)
  - privacy

polyzos@cs.ucsd.edu

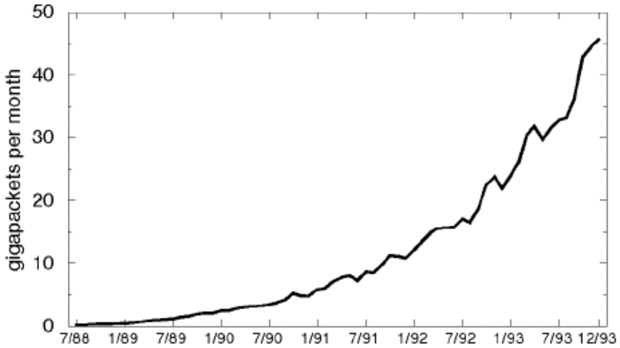
NSFNET Collected Data

- collection mechanisms
  - SNMP
  - NNStat/ARTS
- example data
  - src-dst net traffic matrix (pck, bytes)
  - protocol above IP distribution (pck, bytes)
  - TCP/UDP port distribution (well known subset+: pck, bytes)
  - pck length histogram (50 B granularity, T1 only)
- 15 min is finest granularity
- workload modeling/forecasting
- no usable data for modeling
  - round-trip delays
  - interface error counts (lumped)

polyzos@cs.ucsd.edu

NSFNET Backbone Traffic Characteristics

- overall, long-term traffic volume growth
- faster growth of particular applications/protocols (http)
- geographic distribution of traffic
  - popular src, dst, src/dst pairs
  - source dependent favoritism
  - international distribution



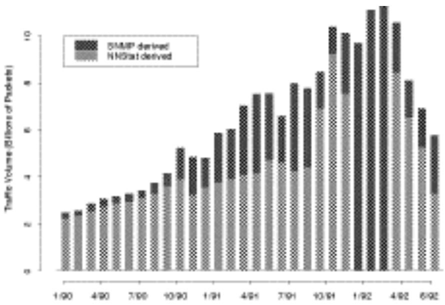
polyzos@cs.ucsd.edu

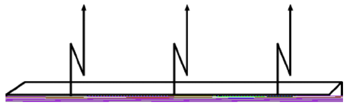
Feedback to Infrastructural Efforts

- integrity of statistics collection process
  - definitions of measured objects
  - impact of sampling
- network reachability
  - better management of misbehaving networks, e.g., leaky, inactive
- traffic cross-section tracking method (TCP/UDP ports) needs improvement
- flow locality (traffic exchange hot-spots)
  - net-net, country-country, node-node, ISP-ISP, etc.

polyzos@cs.ucsd.edu

Sampling





polyzos@cs.ucsd.edu

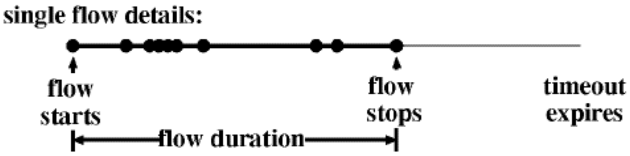
IP Flows

- Definition of IP Flows
- IP Flow Data Collection
- Alternatives to Implicit Definition of IP Flows
- Parameterizable Definition of IP Flows
- Applications
- Results
- Observations

polyzos@cs.ucsd.edu

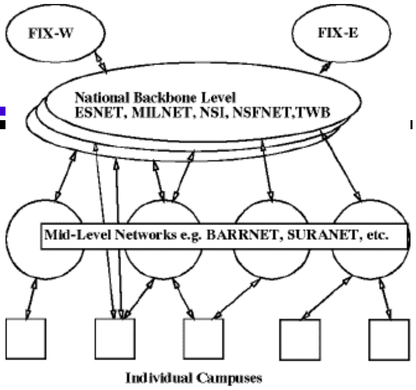
Definition of IP Flows

- IP-layer
  - network-centric perspective
  - pure layering: no access to transport and higher layer information within the (inter)network
    - but it might be worthwhile to cheat if you can...
  - significant/increasing non-TCP uses of the network
  - routing, accounting, network management
  - SYN/FINs are problematic for routers (datagram network)
  - implicit (IPv4), timeout based



polyzos@cs.ucsd.edu

IP Flow Data Collection



- data collection
  - non operational
    - at multiple points (router interfaces)
    - at different levels of the network "hierarchy"
    - ... "sampling" (non-scientific) over the topology
- captured data
  - timestamp (µs)
  - src/dst IP adddress
  - protocol id
  - TCP flags
  - src/dst TCP/UDP port

polyzos@cs.ucsd.edu

Alternatives to Implicit Definition of IP Flows

- IPv6 flows
  - explicit, if flow id used
  - network layer, but subset of traffic
- TCP flows (explicit)
  - SYN, FIN markers
  - only subset of traffic
    - major loss: no UDP, DNS, IPIP, etc.
    - minor loss: routing changes

polyzos@cs.ucsd.edu

Parameterizable Definition of IP Flows

- directionality
- one vs. two endpoint aggregation of traffic
  - e.g., dst net vs. net pair
- endpoint granularity
  - IP layer
    - backbone, multi-backbone, backbone node/interface
    - administrative domain
    - IP network number
    - host
  - higher layers
    - users
    - applications
- functional (protocol) layer

polyzos@cs.ucsd.edu

Applications

- route caching
- analysis of requirements for QoS provisioning
- datagram - virtual circuit interplay
  - IP over ATM
  - VC holding time characterization

polyzos@cs.ucsd.edu

Results

- individual flow metrics
  - flow duration
  - flow volume (packets, bytes)
- aggregate flow metrics
  - new, active, deleted flows per second
  - flow interarrival time distribution

polyzos@cs.ucsd.edu

Observations

- "natural" IP flow timeout ~ 16 - 128 s
  - 2 s timeout: flows are re-created 10 times on average; 64 s: 1.2 (UC-NSF data)
- most IP flows are short
  - 60% had duration < 1 s, 40% were <= 1 packet long
- volume/duration correlated with high level protocol
  - single packet flows (64 s timeout):
    10% of TCP, 60% of UDP, 52% of ICMP flows
- dns flows: 2.4% of packets, but 27.6% of flows
- most short flows are: dns, gopher, ntp, finger
  - not caching/opening new VC for them could save over 1/3 of the space
- # of host pair flows is only 2-3 times # of net pair or dst net flows
- significant short term locality
  - 60% of packets across UC-NSF and SD-NSF were destined to netsreferenced within 30 and 90 previous address references, respectively

polyzos@cs.ucsd.edu

Outlook

- current status
  - congested ISPs, NAPs
- better analysis of WAN traffic and performance characteristics may allow more efficient network evolution
- market driven model alternative might be more appropriate for the chaotic Internet environment
- collaboration among ISPs?
  - need standard
    - service model
    - metrics
    - measurement techniques
- NLANR real-time web based tools

polyzos@cs.ucsd.edu

Traffic Characterization Perspectives

- user's/customer's
  - performance, cost/performance
  - no concern about other traffic if QoS guaranteed
  - if no QoS guarantee, possible interest to avoid contention
- network operator's
  - look for trends, drastic changes
    - to set policies, possibly modify parameters
  - cares about utilization at the minute granularity
- network/protocol/equipment designer's
  - maybe interested in statistical profiles of traffic
    - burstiness at different time granularities, etc.
  - but ... not realy
- performance analyst's
  - Poisson, fractal, ...

polyzos@cs.ucsd.edu

Conclusion

- operational statistics useful, but not to the performance analyst
- provided feedback to infrastructural efforts
  - Merit, ANS, MCI, NSF
- developed methodology for characterizing Internet flows from a network centric perspective
  - with applications to: route caching, IP over ATM...
- new environment trickier
- impact of QoS guarantees/contracts

polyzos@cs.ucsd.edu