Overcoming Precision Limitations in Adaptive Bandwidth Measurements

Prepared for the 3rd
New York Metro Area
Networking Workshop
NYMAN 2003

James Alberi
Allen McIntosh
Marc Pucci
Thomas Raleigh

{jla, mcintosh, marc, tom}
@research.telcordia.com
September 12, 2003

An SAIC Company

Summary

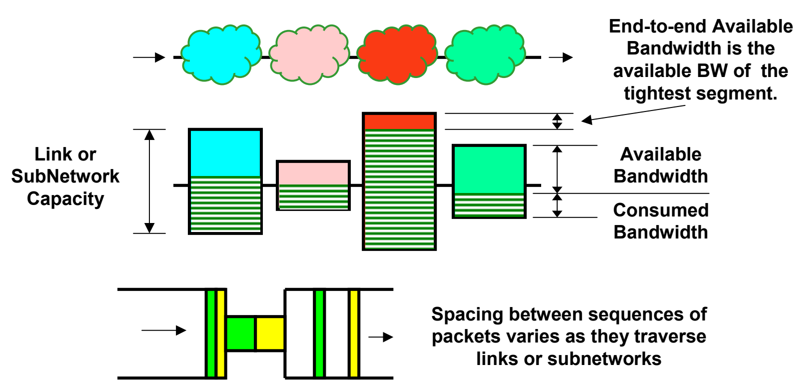
- IMP, a system for network measurement
- Support of active and passive measurement
- Provides a measurement infrastructure
  - More accurate timing
  - Packet train specification
  - Ease the introduction of new measurement techniques

Motivation

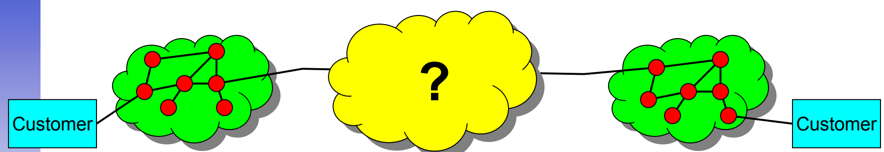
- Needs of Service Providers and Enterprises
  - Accurate measurements for operational support
    - Management of Quality of Service
    - Identification of service-affecting conditions
  - Accurate measurements for Service Level Agreement support
    - Delay, loss, delay variance and availability are basic requirements
  - Validation of Service Level Agreement claims
    - How does an enterprise know if their agreement is being met?
- Needs of the Research Community
  - Improve the accuracy of measurements
    - Especially relevant for new measurement techniques
      - Adaptive Dispersion Techniques to measure available bandwidth
  - Simplify the addition of a new measurement technique
    - Provide basic functionality in a platform environment
    - Study the multitude of Adaptive Dispersion Techniques
    - Leave the plumbing to us

What are Adaptive Dispersion Techniques?

- Active measurement of link capacity and available bandwidth
- Measures internal links not under the control of the measurer
- Can verify available bandwidth without necessarily saturating link
- Inject traffic, measure dispersion, feedback to injector



Application of Adaptive Dispersion Technique



Local Regional Network A (with internal visibility)

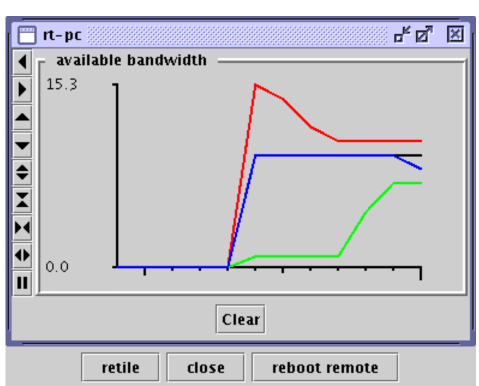
Unaffiliated Transit Carrier Network (No internal visibility)

Local Regional Network B (with internal visibility)

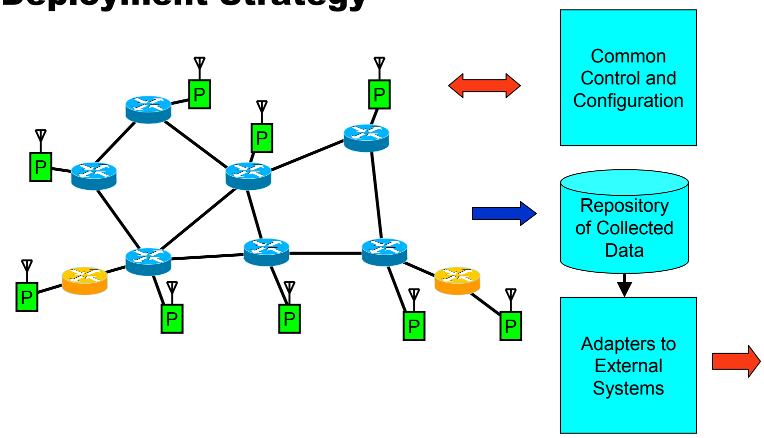
- To support an end-to-end VPN service, a geographically distributed provider uses a transit carrier to link regional networks
- The transit carrier agrees to provide a transit bandwidth of X Mb/s across its network
- How does the regional carrier verify that the transit carrier is really providing the agreed upon bandwidth?

Computation of Available Bandwidth

- Pathload: Tries to bracket available BW
- Plot drawn by infrastructure shows convergence progress
  - Use for developing algorithm, debugging

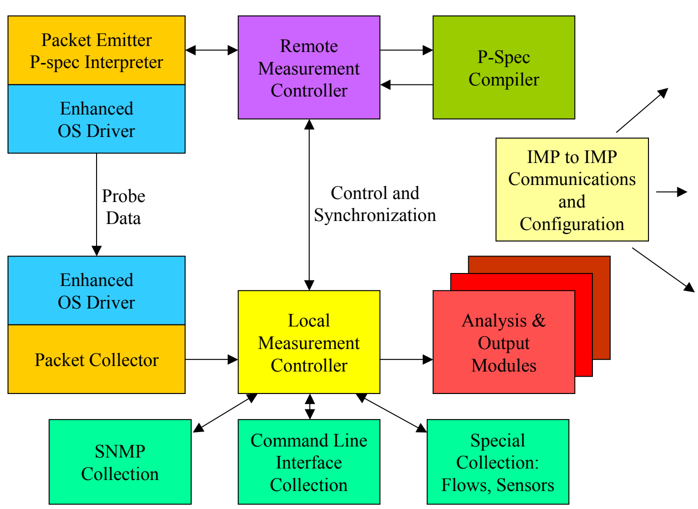


Deployment Strategy



- Measurement platforms cover network
- System includes scaleable configuration, initialization, deployment, reporting, etc.

Platform Architecture

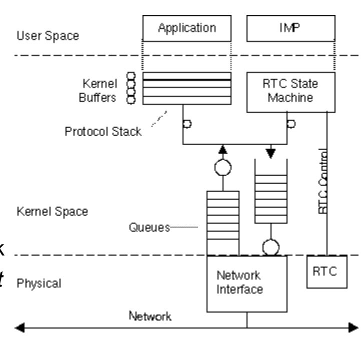


Timing Support

- Measurements can be supported at several layers
  - Application Software Based
    - Easiest to deal with, but subject to lack of control
  - Kernel Interrupt Driven
    - Significantly greater accuracy
    - Can be difficult to use
      - Unless a platform provides the details
    - Still interruptible
  - Exploitation of Network Interface Card characteristics
    - Even greater accuracy for higher speeds
    - No interrupts
    - Difficult to use
  - Dedicated Hardware
    - Cost
    - You get what you pay for

IMP Packet Generation Design

- Common hardware – custom software
  - Real-time clock with 122 μsec granularity
  - Network interface card
  - No impact on existing functionality
- Design components
  - Standard system calls for standard packets
  - Schedule IMP packet transmission via RTC interrupt-driven state machine
  - IMP packets bypass the kernel protocol stack
  - IMP packets formatted via an array of *Packet Descriptors* copied into kernel space (/proc).
- IMP packet trains have kernel generated
  - Programmable timing profiles
  - Programmable packet size profiles
  - Driven by P-spec



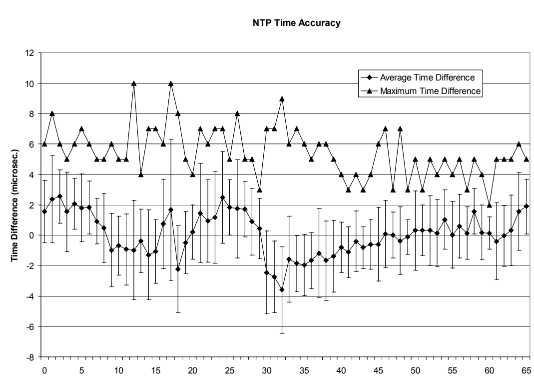
Clock Synchronization

- Intersystem synchronization of absolute time is critical to the success of network measurements
  - Precision and accuracy a function of network speed
  - Use GPS-based time reference for 10/100 MHz
  - Currently implemented using NTP (GPS “bug” + PPSkit)
- Drawbacks of NTP
  - Unstable against brief outages of the reference time base
  - Heavyweight mechanism at application and kernel
- Plan to replace entirely in a kernel module

Telcordia Technologies

Synchronizing Clocks

- Test NTP implementation against a hardware-based commercial time reference
- μsec error very good for NTP



Telcordia Technologies

P-SPEC – Specification Language for Packet Sequences

- Example P-Spec for an Adaptive Dispersion Technique packet sequence (pathload)
- The probe sends fleets of packets, with the packet size, inter packet gap and inter burst delay adjusted by the collecting process according to the analysis algorithm.

```
fleets_per_burst=3              probe parameters
packets_per_fleet=20
burst=3
*:(
    WAIT(fleet++)               send fleet index to collector
                                      and await response
    gap=arg0                    reset new values for gap,
                                      size and delay
    size=arg1
    delay=arg2
    p1=0
    p3=packets_per_fleet
    fleets_per_burst:(          generate multiple packet fleets
        p0=burst                packet word 0: burst number
        p2=0                    packet word 2: packet in fleet
        packets_per_fleet:(
            PACKET()
            p2+=1
        )
        p1+=1                   packet word 1: fleet in burst
        DELAY(delay)            inter-fleet delay
    )
)
```

Telcordia Technologies

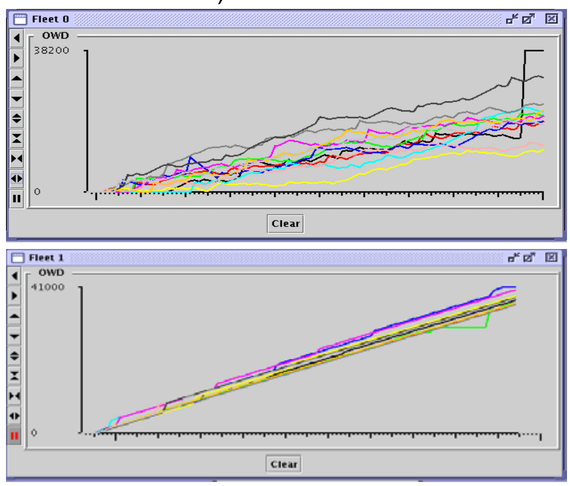
P-SPEC Language

Packet variables	p0 - pN	Sets the values of the first N words of a packet
Assignment ops	=, *=, etc.	Binary operators for generating simple progressions such as sequence numbers or varying delays or sizes
Synchronization commands	wait, sync, delay,	Generates the notation for a synchronization point in the stream for communicating back to the measurement collector
Synchronization variables	arg0 - argN	Control arguments set by “start” or “continue” commands from the collector
Packet descriptor variables	size, gap, tos,	Define packet constraints such as size and gap before next packet
Packet generator	packet	Forces a Packet Descriptor to be generated with the current set of values for content, size, etc.
Looping control	N:(…)	Repeat internal instructions N times; for generating repeated subsequences of packets
General purpose variables	a - z	General purpose variables for processing, but which have no special meaning to the packet generator.

Telcordia Technologies

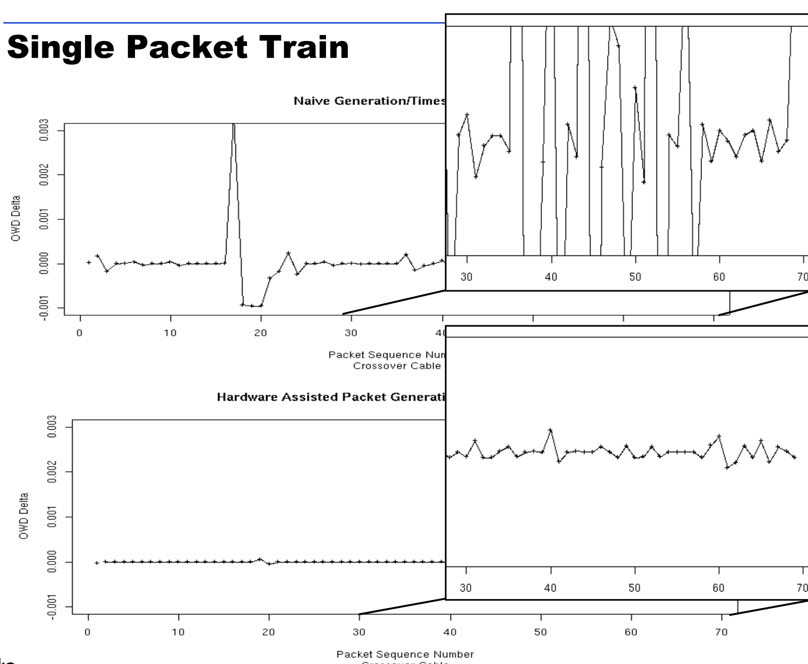
IMP: Measurement Plots

- Send 12 trains of 100 packets through network
- Simple Java implementation vs. kernel support
  - (Most code the same)



Telcordia Technologies

Single Packet Train



Telcordia Technologies

Summary

- IMP, a system for network measurement
  - Commodity hardware + Linux
- Active and passive measurements
- Provides measurement infrastructure
  - More accurate timing
  - Packet train specification
  - Data storage
  - Co-ordination
- Next
  - Push timing into hardware
  - Real-time OS?
  - Comparing algorithms
- Looking for friendly users