TCP ESTATS MIB

Matt Mathis
John Heffner
Raghu Reddy
Pittsburgh Supercomputing Center

J Saperia
JDS Consulting

Status page:
http://www.web100.org/mib/

Outline

- Overview & motivation

- Open technical issues

- Open process issues
  - IPv6MIB team has RFC2012bis

TCP Extended Statistics MIB

- Detailed, per connection TCP instruments

- When there is a problem, just ask TCP
  - TCP has an ideal vantage point
  - TCP can identify the bottleneck subsystem
  - TCP already measures the network
  - TCP can measure the application
  - TCP can adjust itself

- From the web100 project
  - www.web100.org

Why is this important?

- The hourglass hides the net from upper layers

- This is good for the growth of the 'net
  - but it also hides all bugs

- All bugs have the same symptom: less than expected performance!

- Any one bug masks all other bugs
  - TCP "tuning" is really debugging
  - Trial and error leads to a random walk
  - Can confound all but the best experts

TCP ESTATS MIB Instrument Groups

- Flags, options, state and negotiated features:
  - Window Scaling, SACK enabled, etc.

- Traffic and throughput
  - bytes & packets, in & out, etc.

- Triage - Why did TCP stop sending?
  - Receive (rwin), path (cwnd), or sender (other)

- Network path properties
  - Abstract events driving the congestion window
  - RTT, loss, ECN, reordering, etc.

- Buffering, API and tuning
  - Buffer occupancy, etc

More detail

- **About 135 Instruments in 8 tables**
  - Individual enable controls per table

- **Indexed by tcpEStatsConnectIndex**
  - Supports fast polling on one connection
  - Stats live after TCP close

Transport issues

- **Are our instruments complete?**
  - What have we missed?

- **Use bytes (instead of octets)**
  - 806 IDs use bytes
  - 697 IDs use octets
  - out of 4060 Internet Drafts

- **Change the handing of retransmitted data**
  - remove ...excluding retransmitted data
  - want load on IP layer (bytes & segs sent)
  - and progress (total ACK advance)

- **Is the table partition and control appropriate?**

SNMP/SMI issues

- Connection StartTime type
  - TimeStamp (or SysUptime?)

- Duration granularity (is micro-sec ok?)

- Counter sizes (32 v 64 bits)

- Error semantics are not sufficiently specified

Process issues

- Current IPv6 draft has non-trivial transport extensions

  - draft-ietf-ipv6-rfc2012-update-01.txt
  - New address types
  - New per connection statistics table
  - New listen table

- IPv6MIB plan - roll back to minimal RFC2012bis

  - New address types only
  - Wish to preserve "orphaned" parts of RFC2012bis

- Plan merge orphaned parts into ESTATS MIB(?)

- We need
  - Review of current ESTATS MIB
  - Review of orphaned parts of RFC2012bis