Path MTU discovery
11/09/04
draft-ietf-pmtud-method-03.txt

Matt Mathis <mathis@psc.edu>

John Heffner <jheffner@psc.edu>

Kevin Lahey <kml@patheticgeek.net>

Packetization Layer path MTU discovery

- Start with 1kByte MTU
- Probe with larger packets to test MTUs
  - Provisionally raise MTU if successful
  - (Optional) Process RFC1191 ICMP
  - Do not reduce TCP window on lost probe
- Verify provisional MTU for 1 RTT
  - Additional losses imply problems

Layered Implementation

- State kept in path information cache in IP layer
  - Probing state and timers
  - Recent successful and unsuccessful probe sizes
- Algorithm runs in the Packetization Layer
  - PL cuts the data into packets
  - Probing and verification are intrinsically PL specific
  - New description facilitates sharing the rest of the code
    - The search heuristic and error logic can be shared

Key Properties

- Robust
  - Tolerates ICMP delivery problems
  - Verification phase addresses spurious delivery
- Progressive interoperation with classical pMTUd
  - Start large and process all ICMP
  - Start small and ignore all ICMP
- Parallel to congestion control
  - End to end algorithm: use loss as the feedback to adjust window or packet size
  - Well understood limitations

Robust

- Primary design goal: Do no harm
- Avoid problems with RFC 1191 pMTUd
  - Not affected by ICMP delivery problems
  - Not affected by tunnels and encapsulation
  - Not exposed to RFC 2923 problems
- Minimal new exposure
  - Spurious delivery of oversized packets
  - Verification phase provides protection

Progressive deployment

- Enhance RFC1191 pMTUd
  - Start with large MTU and process ICMP
  - Use PLPMTUD iff repeated timeouts
  - Maximally robust from a deployment perspective
- Replace RFC1191 pMTUd
  - Start with small MTU, ignore all ICMP PTB messages
  - Search upwards to raise MTU
  - Maximally robust from a security perspective

Parallel to Congestion Control

- End-to-end algorithm
- Adjust data stream parameters:
  - Packet or window size
- Use packet loss for feedback
  - Interactions with Congestion Control are specified in RFC2119 standards language
- Better fit with end-to-end principle(?)

New with -03 draft

- Generalized to be PL protocol independent
  - Requirements for PL protocols
    - Bi-directional, timely and accurate delivery reports
    - Mechanisms for probing and supporting provisional MTU
  - Distilled descriptions for selected PL protocols
    - TCP, SCTP, IP fragmentation, UDP/application
- Clarified interactions between PLPMTUD and congestion control

What next?

- Implementations
- A MIB
  - AUGMENT the IP (routing) MIB?
- All known document holes are fairly minor
  - Better support for short/small flows
  - Add more PL protocols
    - RTP: the variable length payload