Reducing Initial RTO

draft-paxson-tcpm-rfc2988bis-00.txt

H.K. Jerry Chu - hkchu@google.com
Vern Paxson – vern@icir.org
Mark Allman – mallman@icir.org

InitRTO Default

- Default value needed before RTT sample is made
- Used during 3WHS and at the beginning of the data phase
  - RFC2988 recommends against using RTT measurement from 3WHS to seed the RTO
- 3 seconds by RFC1122 and RFC2988

InitRTO - RFC1122

…
The following values SHOULD be used to initialize the estimation parameters for a new connection:

(a) RTT = 0 seconds.

(b) RTO = 3 seconds. (The smoothed variance is to be initialized to the value that will result in this RTO

...

DISCUSSION:

Experience has shown that these initialization values are reasonable, and that in any case the Karn and Jacobson algorithms make TCP behavior reasonably insensitive to the initial parameter choices.

Reducing InitRTO

- Choice of initRTO can affect performance today
  - web traffic over short lived TCP connections
  - SYN/SYN-ACK loss rate > 1%
  - ~97.5% of connections have RTT < 1 second
- More at

http://www.ietf.org/proceedings/75/slides/tcpm-1.pdf

Design Goal

- Help the vast majority (97.5% with RTT < 1sec) while not hurting the rest (2.5% with RTT > 1sec)
- Keep It Simple!

Proposal

- Set initRTO to 1 second
- Exponential backoff from 1 sec (i.e., 2secs, 4secs,…) when timer expires
- If timer expires during 3WHS, reset initRTO back to 3secs (for the initial data phase)
- See draft-paxson-tcpm-rfc2988bis-00.txt

Negative Impact during 3WHS

- One or two unnecessary SYN/SYN-ACK packets
- Prevent a valid RTT sample from being taken
  - But RFC2988 recommends against taking RTT sample during 3WHS anyway
- RFC3390:

  "… If the SYN or SYN/ACK is lost, the initial window used by a sender after a correctly transmitted SYN MUST be one segment consisting of MSS bytes."

  - Make an exception, to only slam down IW after multiple SYN/SYN-ACK retransmissions?

Negative Impact during Data Phase

- Spurious RTO, if undetected
  - causes unnecessarily-early exit from slow start and a reduced cwnd
- The following text is added to minimize spurious RTOs at the beginning of data transmission phase

  "(5.7) If the timer expires awaiting the ACK of a SYN segment and the TCP implementation is using an RTO less than 3 seconds, the RTO MUST be re-initialized to 3 seconds when data transmission begins (i.e., after the three-way handshake completes)."