Applying TCP User Timeout Parameters to BGP Sessions

draft-chen-idr-tcp-user-timeout-01
IETF IDR WG Interim-2023

Enke Chen
Robert Raszuk
TCP User Timeout

- Defined in the TCP base spec [RFC0793] [RFC9293]
  - Terminate the connection if TCP is not progressing within certain time

- Application data delivery by TCP
  - Transmitted data: waiting for acknowledgment
  - Buffered data: waiting for transmission (due to 0-window size)

- No difference to an application between the two TCP cases
  - Application data not delivered
Implementations

- Linux implementation as an example
  - TCP_USER_TIMEOUT option
    - setsockopt(s, IPPROTO_TCP, TCP_USER_TIMEOUT, &timeout, sizeof(timeout));
  - Return ETIMEDOUT to the app in read/write (just like EPIPE)
  - Recent bug fixes (v5.4.x, 01/2021)
    - 9d9b1ee0b2d1 tcp: fix TCP_USER_TIMEOUT with zero window
    - 344db93ae3ee tcp: make TCP_USER_TIMEOUT accurate for zero window probes

- Straightforward for timeouts
  - Timer start, reset, measure
Applying TCP User Timeout to BGP

- **BGP “stuck” session**
  - BGP messages are not delivered for an extended period of time
  - Can result in “stale” routes in the routing system
  - Desirable to detect and terminate the session

- **TCP User Timeout Option**
  - Specifically defined for such data delivery issues
  - Deterministic detection than by BGP itself (zero-window in particular)
  - Minimal code changes in BGP
Recommendations

- Use TCP User Timeout
  - Default timeout value: 5 x holdtimer, and no less than 10 minutes
  - Enable after EOR, or 30 minutes post session establishment
  - Follow the GR procedures for TCP ETIMEDOUT error
  - Log the event, and work with the peer admin for resolution
Next Steps

- Welcome feedback
- Request IDR WG Adoption