TCP Persistence Techniques

Chakchai So-In
Department of Computer Science & Engineering
Washington University in St. Louis

Outline

- Introduction
- TCP Persistence Techniques
- Process Migration
- TCP Failover and Fault tolerant
- Miscellaneous
- TCP Multi-homing
- Conclusions

Introduction

- Persistence/ Mobility
- Goals
- Layer Persistence
- Transport Persistence
  - Breaking TCP connection
  - UDP Persistence
  - TCP Persistence

Persistence/ Mobility

- Persistence: The property of a continuous and connected period of time
- Mobility: The ability and willingness to move or change
- Persistence Mobility : To achieve the continuous operation due to movement or lost of connectivity

Goals

- Users manually states the requirement of having continuous connection for their applications regardless of user movement and link disconnection.
  - Unexpected disconnection
    - Roaming (Handoff issue: Change of IP)
    - Long period of link disconnection
    - Server down
  - Expected disconnection
    - Laptop Hibernation
- Policy based control
  - Pre-negotiation disconnection time from User, Server, and Network.

Layer Persistence

- L1 issue: Low power RF and DSP
- L2 issue: FEC/ARQ
- VPN issue:
  - IKE/IPsec Re-Keying
  - Allow for Dead-Spot time
  - IKE Keep-alive: Spoof Keep Alive (need to check)
  - Set keep-alive to 0 and account for DPD
- L3 issue: Mobile IP
- L4 issue: Maintain transport state
- L7 issue: Suspend/Resume state/ Buffering

Transport persistence

- Breaking TCP connection
  - Retransmission Timeout (RTO) is the first main factor. For example, under Window 95, TCP retransmits the dropped segments up to five times. RTO is increased exponentially and the upper bound of Window 95's RTO is 260 seconds.
  - There is no TCP keep-alive response back within time period.
  - The TCP timeout is expired.
    - There is no standard, but usually this factor is based on the predefined time units or a number of retransmissions.

Transport persistence (cont)

- UDP Persistence (UDP -> Connectionless)
  - Resent previous UDP packet (like UDP keepalive)
  - MMS and Share Drive
    - Should be ok if control message is alive
- TCP Persistence
  - Hardcode all parameters: RTO, Keepalive, Connection Timeout (How long? Congestion?)
  - Create TCP Suspend/Resume State (TCP State modification)
  - Other mechanisms such as TCP freeze

TCP Persistence Techniques

- Indirect Layer Model
- Location Management
- Buffer Management
- TCP State Modification and Migration
- TCP Freeze Option
- Link Detection
- Connection Authentication

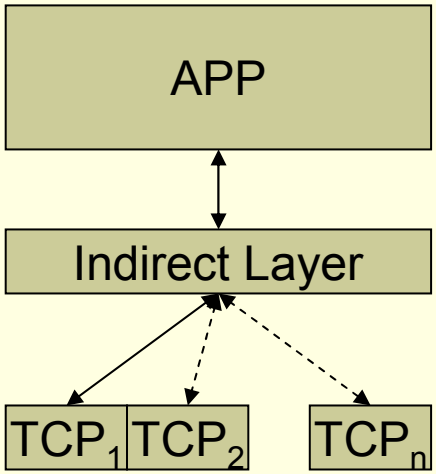
Indirect Layer Model

- An always connected logical end point mapped to physical endpoint.
  - An additional layer provides a redirection (between application layer and transport layer).
- Indirect Layer addressing:
  - *Virtual Address*: Socket descriptor, persistence process identification, IP address, Port number, and connection identification such as <*s, 0.0.0.15:3*>.
  - *Connection Identifier* The public key 32 bits random number is used to identify the connection

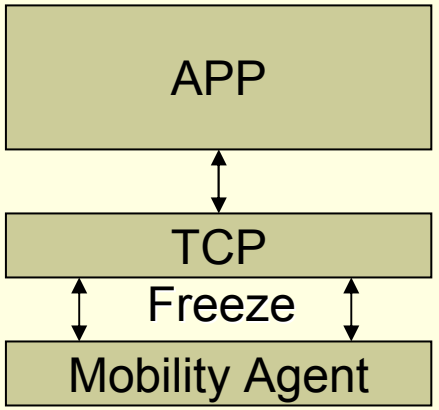
Indirect Layer Model (cont)

- *"MSOCK"*, is the redirection mechanism to redirect the traffic through a split-connection proxy server and that proxy maintains all connection states.
  - The proxy can transparently splice the connection to alternate server where ensure the connection consistency.

ILM and TCP Freeze Option



Indirect Layer Model



TCP Freeze Option

Location Management

- A central server is used to track the movement and also maintain the connection identifier.
- With Mobile IP, the home agent automatically update the change of IP address.
- Nodes keep updating the IP address regularly (Client resolves server name before making a reconnection)
- The use of Session Initiation Protocol (SIP) is applied to identify the unique name.

Buffer Management

- Buffer in which all non-blocking write sockets are sent (block all blocking write sockets
- Fixed buffer size and having the acknowledgement signal sending regularly to confirm the received packet. Those packets can be removed from the user buffer space and re-sending the lost data when reconnection.

TCP State Modification/ Migration

- Add SUSPEND and RESUME state directly to TCP state
- TCP Migration Option (in-band signaling)
- State Synchronization

TCP Freeze Option

- *"rack"* and *"STCP-F"*, the idea is to apply the use of TCP Freeze technique to preserve the connection. (Up to application timeout)
  - The agent sends zero window packets back to the application due to the disconnection. Once the application receives those packets, it sends probe packet back to check if it can send more packets. The agent keeps sending the probe packet back, and this process keeps going back and fourth until the link is resumed.
  - When link is resumed (exchanging window update packets), it goes back to the normal operation.

Link Detection

- DHCP: Change of IP address or network prefix
- TCP Keep-alive mechanism
- ICMP
- *heartbeat probe* mechanism which the out-band UDP signaling is used to constantly probe the disconnection/ service inactivity
- Trade-off (frequency/ traffic)

Connection Authentication

- Secure Reconnection (connection hijacking):
  - Global naming or unique connection identifier (the combination of IP address, port number, and unique connection identifier).
  - Pre-Share key (randomly select)
    - Store ID in both end or central server
  - Public key (Elliptical Curve Diffie Hellman key exchange)
    - issue on the re-keying with the long session

Process Migration

- Unless process is down and without TCP failover/ fault tolerant, most TCP persistence techniques can function properly.
  - Process checkpoint and process backup and migration (Operating System) supported
  - Socket migration;
    - *"MIGSOCK"* library in Linux kernel to make the existing process migration system (*CRAK*) to support a socket migration.
    - *"tcpcp"*, TCP connection passing for Linux

TCP Failover and Fault tolerant

- Server is no longer in service.
  - Redundant server is required, and usually it runs the same application as in the primary server and execute the same operation but not sending the ack back to the client.
  - The backup server is running in promiscuous mode and does the same process as that in the primary server
  - TCP states are synchronized between both of them.

Miscellaneous

- *"SLM" :* Session Layer Mobility Management
  - Maintain each connection as a single session separately.
  - A unique session identification is provided for that which all session numbers are stored in the session table.
- *"A reliable and secure connection migration mechanism for mobile agents"*
  - *"suspend()"* and *"resume()"* for connection migration
  - The delay and priority (giving one side higher priority and delay the other and also the use of priority is to prevent the deadlock of simultaneous migration) techniques is used if the agent moves simultaneously.

Miscellaneous (cont)

- *"Application-layer Mobility support for Streaming Real-time Media"*
  - a UDP socket extension for application layer mobility (socket translation: NAT-liked),
- *"WebPod"*,
  - *WebPod* is built on the small storage device such as USB. It stores the user's plug-ins, bookmarks, browser web content, open browser windows, and browser configuration options and preferences.
  - The checkpoint and restart techniques are used to suspense and resume the web browsing session.

Miscellaneous (cont)

- *"TCP User Timeout Option"*,
  - Simply allow users to advertise the user timeout for each connection
  - Assumption: users can predict how long the disconnection would be; however, it might bring the issue on congestion control and this technique directly absorbs the problem on TCP idle timeout.

TCP Multi-homing

- *" SCTP"*: multi-homing: allow a single SCTP endpoint to support multiple IP address within a single association",
  - The mobile node can have at least two IP addresses in the existing association. The mobile node uses one of the IP addresses to prepare for actual handoff and tells the corresponding server using another transport layer connection that it can be reachable by another IP address.
- *"mSCTP" , " M-SCTP",* and *" Mobile SCTP"*
  - Use of the SCTP and ADDIP extension (Dynamic Address Reconfiguration), so it can dynamically delete and insert IP address during the operation.

TCP Multi-homing (cont)

- Applied SCTP concept, multiple connections are allowed at the same time, so one might be a backup of the other.
  - Like TCP Migrate option, some kinds of STCP option can be added to make a smooth internal migration (at least one connection is active)
  - The migration mechanism is needed to be added.
  - The internal connection splicing technique can be used. for example, the virtual address and sockets layer can be added to hide the change of TCP association

Conclusions

- Persistence Mobility
- TCP Persistence Techniques
- Process Migration
- TCP Failover and Fault tolerant
- TCP Multi-homing

Conclusions (cont)

- How long should we keep a connection alive?
  - What if users have never terminated a connection?
  - Given priority of the application persistence (The server predefines the maximum number of persistence connection to <> for each user), (*?DoS?*)
    - 00 Critical: do not end the connection for any reasons *(Server may not grant this service)*
    - 10 Suspension: keep the connection alive for at least <> minutes
    - 11 Temporarily (Default): keep the connection as long as possible (allow the server drop the connection for some reasons such as TCP descriptors are full)

Conclusions (cont)

- Congestion versus Disconnection/ Suspension
  - Users specify if the connection is needed to be terminated (ABORT command)
- Compatibility
  - Header bits (TCP or MobileIP) are used to identify if the persistence feature is supported. If not, it works as a normal operation

Conclusions (cont)

- Application Idle Timeout/ TCP Timeout
  - Link Disconnection/ Suspension
    - Freeze APP timeout once disconnection/ suspend is detected and resume the timeout after link is resumed.
  - What if link is just idle/ inactivity?
    - Basic Keepalive mechanism
    - Users manually define idle timeout (best effort)