Transport Considerations for IP and UDP Tunneling

Magnus Westerlund
Mirja Kühlewind
Marcus Ihlar
Zaheduzzaman Sarker

draft-westerlund-masque-transport-issues-00
Outline

- Connect-UDP is NOT HTTP Connect
- Terminology
- IP Header
  - ECN
  - DSCP
  - Fragmentation & MTU
- Options
  - Hop Limit
- UDP Header
- ICMP
- Conclusions
Why Connect-UDP is not Connect

- HTTP Connect
  - Provide end-to-end byte stream semantics
  - Back to back TCP/QUIC sessions
  - Consumption of IP header fields for transport by individual transport connections
  - TCP/QUIC support only single QoS Level

- Connect-UDP
  - Providing end-to-end datagram transport
  - Network to end-point signals need to reach endpoints
    - ECN
    - ICMP
  - Endpoint to network signals
    - DSCP
Terminology

- **Proxy** = Masque Server
- **Address**
  - IP address + UDP Port
  - Domain names + UDP Port
- **IP Flow** = 3-tuple (source, dest. address, IP vers)
- **UDP Flow** = 5-tuple (Src + Dest Addr + Port, UDP)

- **Datagram flow**
- **Signaling flow**
  - Flow establishment
  - Asynchronous Events
IPv4 Header

- **Version** = Fixed
- **IHL** = Derived
- **Type of Service** = DSCP + ECN
  - Needs discussion
- **Total Length** = Per Packet
- **Identification** = Sender assigned
- **Flags** = Fragmentation including Don’t Fragment
- **Fragment Offset** = Derived when fragmenting
- **Time to Live** = Sender assigned
- **Protocol** = Per Packet / Fixed (UDP)
- **Header Checksum** = Derived
- **Source Address** = Flow Static
- **Destination Address** = Flow Static
- **IP Options** = Discussion needed

### IPv4 Header Fields

- **Version**: Fixed, Flow Static or Derived
- **IHL**: Derived
- **Type of Service**: DSCP + ECN
- **Total Length**: Per Packet
- **Identification**: Sender assigned
- **Flags**: Fragmentation including Don’t Fragment
- **Fragment Offset**: Derived when fragmenting
- **Time to Live**: Sender assigned
- **Protocol**: Per Packet / Fixed (UDP)
- **Header Checksum**: Derived
- **Source Address**: Flow Static
- **Destination Address**: Flow Static
- **IP Options**: Discussion needed

### Notes

- **Version**: Fixed
- **IHL**: Derived
- **Type of Service**: DSCP + ECN
  - Needs discussion
- **Total Length**: Per Packet
- **Identification**: Sender assigned
- **Flags**: Fragmentation including Don’t Fragment
- **Fragment Offset**: Derived when fragmenting
- **Time to Live**: Sender assigned
- **Protocol**: Per Packet / Fixed (UDP)
- **Header Checksum**: Derived
- **Source Address**: Flow Static
- **Destination Address**: Flow Static
- **IP Options**: Discussion needed
IPv6 Header

- **Version** = Fixed
- **Type of Service** = DSCP + ECN
  - Needs Discussion
- **Flow Label** = Per ULP flow / Flow Static (UDP)
- **Payload Length** = Per Packet
- **Next Header** = Per Packet / Fixed (UDP)
  - Extension Headers (Needs Discussion)
- **Hop Limit** = Sender Assigned
- **Source Address** = Signaled Static
- **Destination Address** = Signaled Static
Type Of Service - ECN

- ECN on Path#1 is dealt with MASQUE’s QUIC connection
- CE marks impacts congestion state
- To enable ECN on Path#2
  - On request of Upper Layer using MASQUE
    - Set ECT(0/1) in Proxy to Client direction
    - Change during flow lifetime
  - Proxy read incoming ECN value in Target to Proxy direction
  - Propagate per packet ECN value to MASQUE client to upper layer
Type Of Service - DSCP

- **Path #1**
  - RFC 7656 applies to QUIC tunnel
    - Use multiple QUIC connection, one per forwarding behavior
  - Setting DSCP to use in Proxy to Client
  - Not necessary to support

- **Path #2**
  - Setting Proxy to Target DSCP to send with
    - Authorization to use DSCP
    - DSCP to forwarding behavior mapping
    - Target to Proxy packets
    - Upper Layer Protocol needs to know Forwarding behavior applied
    - To ensure different CC state
  - Priorities in Packet scheduling in Client and Proxy of aggregated flows
Fragmentation and MTU

- Control IP fragmentation
  - DPLPMTUD requires Don’t Fragment
  - Some upper layer application may lack fragmentation support and require relying on IP fragmentation mechanism

- MASQUE Tunnel has its MTU
  - Using Streams up to 64K
  - Using QUIC Datagrams Path MTU – Overhead
  - Needs API to upper layer
  - Proxy’s external interface has its MTU
    - Should be conveyed to client
  - Control over DF bit on Proxy to Client path
    - Per Packet in some cases
    - Per Default value
IP Options / Extension Headers

- Are there IP Options or Extension headers that upper layer request be add to the IP header?
  - IPv6 minimum Path MTU HBH ([draft-ietf-6man-mtu-option])
  - Network Tokens ([draft-yiakoumis-network-tokens])
  - Other Options / Extensions?

- IPv6 minimum Path MTU HBH Option
  - Data center targeted initially
  - To be added in outgoing packets
  - Target can echo value in the option in next returning packet, needing propagation

- Network Tokens for Path #2 would need MASQUE support
  - Indicate which packets the HBH Extension header is to be added and its value

- What functionality is required here?
Time to Live / Hop Limit

- The usage of Time to Live / Hop Limit
  - Primary Prevent ever living packets in routing loops
  - Path change indicator
    - Currently no transport appear to use it

- Path #1 TTL is directly readable by Client and Proxy on reception
  - Sent value is not known and usually OS specific
  - No special function needed

- Path #2
  - If Client want to know received value
    - Would require MASQUE to propagate
  - Appear to be insufficient need to motivate functionality

Client → Path #1 → Proxy ← Path #2 → Target
# UDP Header

<table>
<thead>
<tr>
<th>Source Port</th>
<th>Destination Port</th>
</tr>
</thead>
<tbody>
<tr>
<td>Length</td>
<td>Checksum</td>
</tr>
</tbody>
</table>

- Source Port = Flow Static
- Destination Port = Flow Static
- Length = Per Packet
- UDP Options ([draft-ietf-tsvwg-udp-options](https://datatracker.ietf.org/doc/draft-ietf-tsvwg-udp-options/)) utilize UDP Length to indicate options area
- How to support UDP Options?
- Checksum = Calculated per Packet

<table>
<thead>
<tr>
<th>Fixed, Flow Static or Derived</th>
</tr>
</thead>
<tbody>
<tr>
<td>+ Per Packet</td>
</tr>
<tr>
<td>* Discussion needed</td>
</tr>
</tbody>
</table>
ICMP

- ICMP provides network to endpoint signaling
- Primary case to consider is the case depicted
- IP/UDP packets on Path #2 from proxy to target that triggers ICMP message transmission
  - On Path: Packet Too Big, Destination Unreachable, TTL Expired
  - End-point (Target): Port unreachable
- Proxy matches ICMP to a tunneled flow
  - Needs asynchronous signal from proxy to client to convey the ICMP information
Categorization IP/UDP flow

- Static - Flow Establishment
  - IP version
  - IP Source address
  - IP Destination address
  - ECN to send
  - Default Don’t Fragment
  - Default DSCP
  - IPv6 Flow Label
  - TTL / Hop Limit
  - UDP source port
  - UDP destination port

- Per Packet
  - ECN
  - IP Packet length
  - TTL received
  - DSCP to send
  - DSCP received
  - IP Options / Extension headers
  - UDP Length
  - UDP Options

- Asynchronous Signaling
  - ICMP Received
    - Packet to Big (size)
    - Destination (Port, Address) unreachable
  - ECN to send (change)
  - Don’t fragment bit (change)
  - DSCP to send (change)
Categorization IP Tunneling

- Static - Flow Establishment
  - IP version
  - IP Source address
  - IP Destination address
  - Default DSCP
  - ECN value to send
  - Default Don’t Fragment
  - IPv6 Flow Label
  - TTL / Hop Limit

- Per Packet
  - ECN
  - Don’t Fragment bit
  - IP Packet length
  - IP TTL received
  - IP Protocol / Next Header
  - IPv6 Flow Label
  - DSCP to send
  - DSCP received

- Asynchronous Signaling
  - ICMP Received
    - Packet to Big (size)
    - Destination (Port, Address) unreachable
  - ECN to send (change)
  - Don’t fragment bit (change)
  - DSCP to send (change)
Conclusions

Summary

- IP and UDP headers include fields that needs to be consumed by the upper layer protocol, e.g.
  - ECN
- IP and UDP headers include fields where the upper layer can set them on per packet basis
  - DSCP
  - Don’t fragment
- There exist asynchronous feedback from the network that can speed up endpoint behaviors
  - ICMP Packet to Big and Destination / Port Unreachable

Requirements

- Need per flow bi-directional signaling
  - Asynchronous signaling for events
  - Change of default used values when sending flow
- Need Flexible and Extensible inclusion of per packet information for the encapsulation
  - Different from Default Value
  - Extension headers or UDP Options
- Flow Establishment need Extensibility
  - Enable extension of functionality