Handling incoming fragments in IPsec

IPsec Workshop 2025, Madrid

Tero Kivinen <kivinen@iki.fi>

Problem description

- When security gateway receives a fragmented IP packet that needs to be sent to the VPN tunnel, only one of the fragments contains the transport protocol information
- In IPv4 the first fragment always contains the transport information as if packet is fragmented to be so short that transport header does not fit to first fragment, then that packet can be considered as attack
- In IPv6 it is possible to have much more hop-by-hop etc headers before the actual transport header, and it is possible that transport header is not in the first fragment, but in some of the later fragments.

IPsec Architecture RFC4301 solutions

- IPsec Architecture (RFC4301) offers three solutions:
  1) One SA that carries both initial and non-initial fragments
  2) Separate tunnel mode SAs for non-initial fragments
  3) Stateful fragment checking
- First option is mandatory to implement
- Other two are optional

Option 1: One SA that carries both initial and non-initial fragments

- Simple to implement, just create one SA with protocol and port ANY, all traffic goes to this SA.
- Does not allow per protocol or per port policies.
- Example:
  - SA: IP protocol ID: 0, Start Port: 0, End Port: 65535
    - All traffic including initial and non-initial fragments.

Option 2: Separate tunnel mode SAs for non-initial fragments

- Separate SA to transport all non-initial fragments.
  - Uses OPAQUE traffic selectors (<65535, 0>).
- Another SA for each per port/protocol traffic allowed.
- MUST make sure that non-initial fragments does not overwrite selector data for IPv4.
- Cannot be done for IPv6, so should not be used for IPv6.
- Example:
  - SA1: IP protocol ID: 0, Start Port: 65535, End Port: 0
    - Traffic for all non-initial fragments
  - SA2: IP protocol ID: UDP, Start Port: 53, End Port: 53
    - Initial fragments (or full packets) for DNS.

Option 3: Stateful fragment checking

- One SA to include both the initial and non-initial fragments.
  - Uses normal traffic selectors, but includes NON_FIRST_FRAGMENTS_ALSO notification to indicate that the SA will also include non-initial fragments.
- Implementation will need to wait for the first fragment, and after finding the first fragment, and an SA where it is sent to, it can use same SA for all non-initial fragment following the first fragment.
  - It needs to make sure that non-first fragments can't overwrite the traffic selector parts of the first fragment.
    - i.e., it can remember in which offset it found the traffic selector information and reject all fragments which have offset less than that.
  - Can also work with IPv6, but in that case it might not find the traffic selector information from first fragments, so it might need to wait for later fragments before deciding the SA to be used.

Option 3: Stateful fragment checking

- This is mostly same that needs to be done when doing reassembly on the packet anyways, except when the traffic selector information has been found the SGW can send all fragments received until that point to that SA, and then only remember IP numbers, fragment id, and minimum offset.