Generic Network Virtualization Encapsulation

draft-ietf-geneve-00

Jesse Gross, VMware

T. Sridhar, VMware

Pankaj Garg, Microsoft

Chris Wright, Red Hat

Ilango Ganga, Intel

Puneet Agarwal, Innovium

Ken Duda, Arista

Dinesh Dutt, Cumulus

Jon Hudson, Brocade

IETF 93 Prague   July 22, 2015

Geneve Overview

- Extensible encapsulation format to allow for future innovation

- Decouple control plane and data plane components to allow different rates of evolution

- Continue to use standard IP fabrics as an underlay

- Support for multiple encapsulated protocols and OAM

Geneve combines a UDP shim, small base header, and TLV options to achieve these goals.

Draft Progress

Adopted as working group draft since Dallas.

What's needed before Last Call?
- Congestion control
- Header checksum

➔ Based on encapsulation considerations design team and experience of other recent protocols such as GRE/UDP and MPLS/UDP.

Congestion Control

Issue: Encapsulation protocols running over UDP enabled non-congestion controlled frames to potentially leak to larger domain (i.e. Internet).

Solution: Borrow applicability statement tsvwg has worked on:

- Encapsulated IP is OK.
- Traffic engineering within operator's network.
- Block potential bad traffic at edge.

Conditions should be OK for network virtualization applications.

Header Checksum

Issue: IPv6 addresses are protected by only the L4 checksum over the pseudoheader. Zero UDP checksum gives no protection.

Possible solutions:

- Applicability statement
- Use UDP checksums
- Use reserved space for checksum over Geneve header
- TLV option to carry checksum

Header Checksum (2)

Possible design requirements to consider:

- Integrity of Geneve metadata
- Protection of encapsulated L2 header
- Strength of checksum
- Implementation complexity and compatibility
- Addition of telemetry data by underlay
- Consistency with future encryption/authentication

Implementations

Controller:
- Open Virtual Networking (OVN)

Software Endpoint:
- Open vSwitch
- Linux

Debugging Tool:
- Wireshark
- tcpdump
- libpcap

NIC:
- Intel XL710
- Mellanox ConnectX-4
- Netronome NFP-6xxx

Switching ASIC:
- Broadcom Trident 2+/DNX
- Cavium XPliant
- Mellanox Spectrum
- Centec GoldenGate

Option Class Assignments

- Focus up to now has been on establishing the base header
- Implementations are similarly agnostic to option definition for maximum flexibility
- Beginning of option definition:
  - (Possibly) Checksum
  - Encryption/authentication
  - Open source projects

- Plenty of room for everyone

Summary

- A few outstanding issues to be addressed before progressing the draft

- End-to-end set of implementations give a good indication of real-world practicality

- Addition of checksums shows the value of extensibility