January 2020

INTENT-AWARE FLOW REDIRECT USING SINGLE-HOP INTERNET CONTROL MESSAGE PROTOCOL (ICMP) EXTENSION

Carlos M. Pignataro

Nagendra Kumar Nainar

INTENT-AWARE FLOW REDIRECT USING SINGLE-HOP INTERNET CONTROL MESSAGE PROTOCOL (ICMP) EXTENSION

AUTHORS:

Carlos M. Pignataro
Nagendra Kumar Nainar

ABSTRACT

Techniques presented herein provide for a Control Agent that monitors flow distribution and Smart Network Interface Card (sNIC) capability and utilizes a simple Internet Control Message Protocol (ICMP) extension to instruct upstream node(s) to redirect traffic over other sNIC connected link(s). The techniques may be utilized for both Layer 2 (L2) and Layer 3 (L3) links and may provide a novel mechanism to leverage an sNIC for flow-based/intent-based/capability-aware load balancing and/or policy-balancing.

DETAILED DESCRIPTION

Smart NIC (sNIC) is a recent advancement in the virtualization arena that introduces intelligence and programmability into a NIC. The programming capability and the architecture provides for the ability to offload cloud native services state entries and accelerate packet forwarding and treatment. Smart NICs can be implemented using technologies such as Application-Specific Integrated Circuit (ASIC) technologies, Field Programmable Gate Array (FPGA) technologies, chip-based technologies, etc. Recent advancements demonstrate that sNICs can be equipped with hypervisor capabilities that can allow users to instantiate Ubuntu® virtual machines (VMs) or execute vector packet processing (VPP) for acceleration.

As illustrated in Figure 1, below, an upstream node (switch) will load share based on flow entropy and may not have visibility into the offload capability of the sNIC connected to the host.

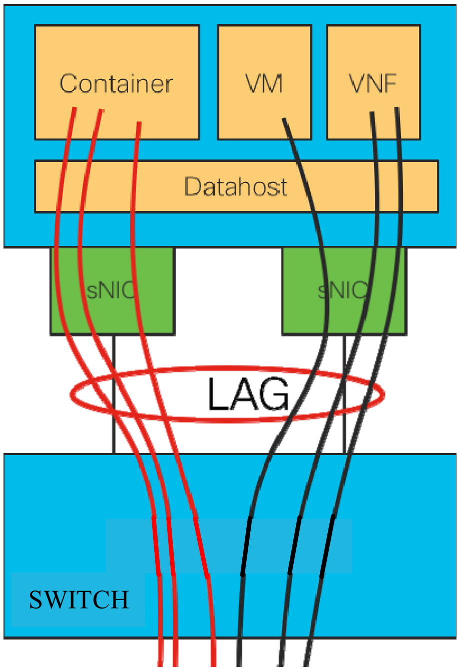
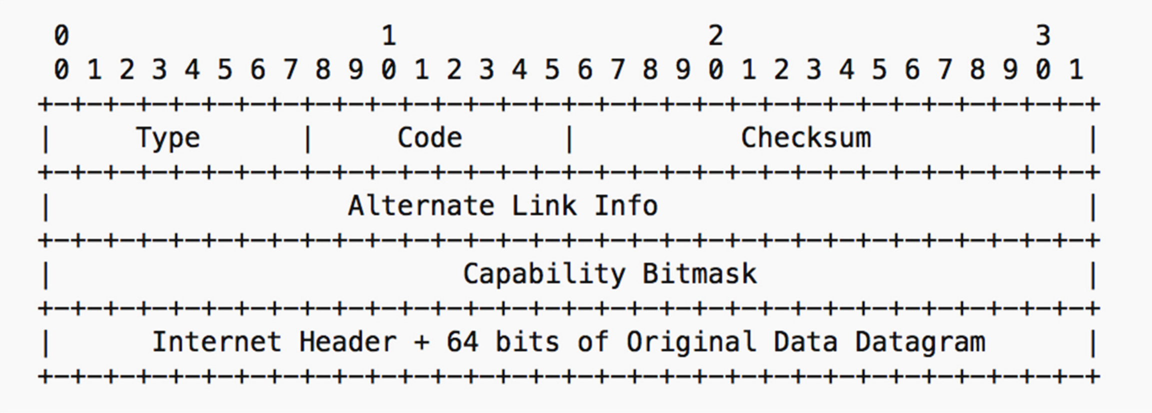


Figure 1

Techniques of this proposal provide a simple extension to ICMP that can be utilized to instruct an upstream node to redirect traffic over other sNIC connected links, including both L2 and L3 links. To implement these techniques, sNIC connected hosts can be enabled with a Control Agent that collects load data from different sNICs for analytical purposes. The Control Agent will have basic topology visibility (e.g., how sNICs are connected) that can be leveraged for load analysis.

When one sNIC within a pair connected to the same upstream node is overloaded, the Control Agent can identify a flow that can be redirected and can trigger an ICMP redirect with link information to redirect the flow over another sNIC connected link. Thus, techniques of this proposal may include providing a new link aware ICMP/ICMP version 6 (ICMPv6) redirect message that is capable of instructing an alternate link within an aggregation. Figure 2, below is a schematic diagram illustrating an example message format that may be associated with the new link aware ICMP/ICMPv6 redirect message.



```
  0                   1                   2                   3
  0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
 +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
 |     Type      |     Code      |          Checksum             |
 +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
 |                 Alternate Link Info                           |
 +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
 |                        Capability Bitmask                     |
 +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
 |      Internet Header + 64 bits of Original Data Datagram      |
 +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 2

For the new link aware redirect message, the "Alternate Link Info" field is used to carry link parameters, such as Link identifier (ID), etc., that can be used by an upstream node to redirect traffic over another link within the aggregation. The "Capability Bitmask" field is an optional field that can carry additional capability information.

In some implementations, a load aware redirect can be provided, as shown below in Figure 3.

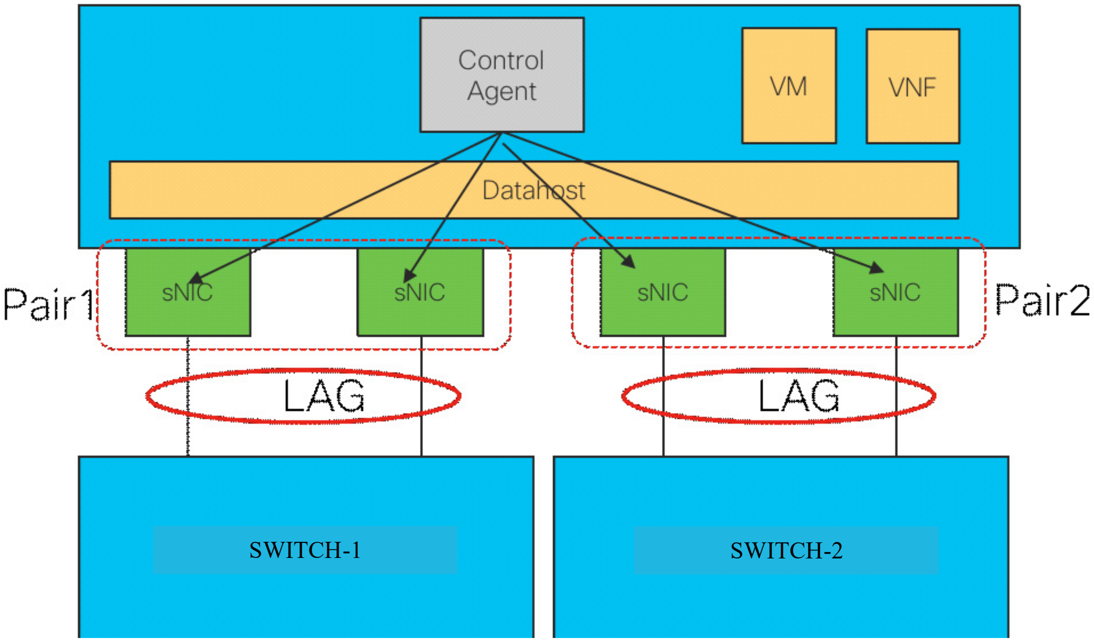


Figure 3

For such implementations, each sNIC connected host can be enabled with a Control Agent that collects load data from different sNICs for analytical purposes. The Control Agent will have basic topology visibility that can be leveraged for load analysis. As illustrated in Figure 3, a first pair (Pair1) of sNICs may be connected to a first switch (switch-1), while a second pair (Pair2) of sNICs may be connected to a second switch (switch-2).

During operation, the Control Agent can collect flow statistics from each sNIC for analytical purposes. When it detects any inefficiency, the Control Agent will generate (or instruct the relevant sNIC) via an ICMP redirect message that includes flow details and the link to which the flow should be redirected. Thus, this simple yet useful technique introduces load awareness at a host level and utilizes redirect messages to efficiently load share between available sNICs.

In another implementation, these techniques can be extended to scenarios that may include a mix of NIC/sNIC. As shown Figure 4, a host may be connected with a combination of NIC/sNIC for various reasons (e.g., cost, incremental deployment, etc.).

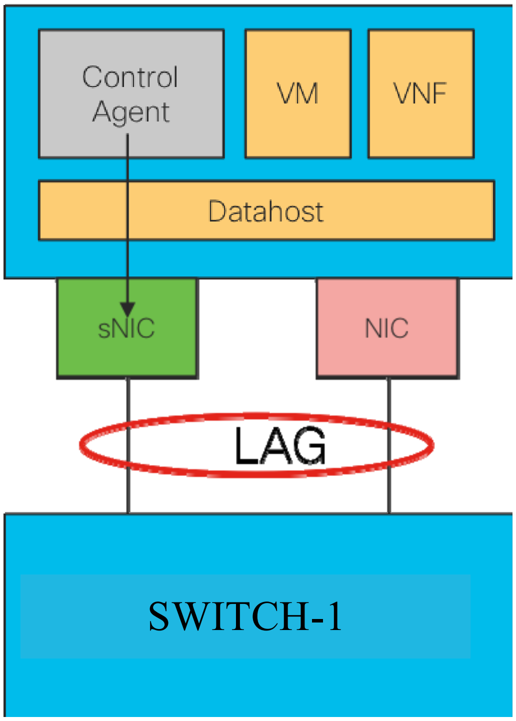


Figure 4

In at least one example for such an implementation, the Control Agent can monitor flows over different NICs. When the Control Agent detects any flow received by a traditional NIC, the Control Agent can determine if the flow is capable of offloading. Upon determining that the flow is capable of offloading, the Control Agent can triggers a redirect message to the upstream node with the flow details.

In summary, techniques herein provide for a Control Agent that monitors flow distribution and sNIC capability and utilizes a simple ICMP extension to instruct upstream node(s) to redirect traffic over other sNIC connected link(s). The techniques leverage a new extension of ICMP redirect messaging to achieve the novel features discussed herein. The techniques may be utilized for both L2 and L3 links and may provide a novel mechanism to leverage an sNIC for flow-based/intent-based/capability-aware load balancing and/or policy-balancing. As industry continues to evolve, it is possible that different sNIC/Top-of-Rack (ToR) combinations may be possible; thus, standardizing the new ICMP extension provided by this proposal may be beneficial to facilitate interoperability for different implementations.