OpenDaylight Project Proposal
"Dynamic Flow Management"

Ram (Ramki) Krishnan, Varma Bhupatiraju et al. (Brocade Communications)
Sriganesh Kini et al. (Ericsson)
Debo~ Dutta, Yathiraj Udupi (Cisco)

Table of Contents

1 Introduction 3
1.1. Large Flow Load Balancing 3
1.2. DDoS Mitigation 3
2 Dynamic Flow Management in OpenDaylight 4
3 Large Flow Local (LAG) Load Balancing 4
3.1 Hash based Local (LAG) load balancing 4
3.2 Large Flow Local (LAG) load balancing 5
4 Large Flow Global Load Balancing 5
4.1 Hash based ECMP load balancing 5
4.2 Large Flow Global load balancing 6
5 DDoS Mitigation 7
5.1 Current Layer 2-4 DDoS Mitigation 7
5.2 Layer 2-4 DDoS Mitigation as an SDN Application 7
6 Southbound API Requirements for OpenDaylight 8
7 Northbound API Requirements for OpenDaylight 8
8 Usage by applications of the Northbound API of OpenDaylight 9
8.1 Application Configuration Parameters 9
8.1.1 Parameters unique to Large Flow Load Balancing 9
8.2 System Configuration and Identification Parameters 10
8.2.1 Parameters unique to Large Flow Load Balancing 10
8.3 Monitoring 10
8.3.1 Parameters unique to Large Flow Load Balancing 10
9 Committed Development Resources 10
10 Potential Code Committers 11
11 References 11

1 Introduction

One of the promises of SDN is to enable the network operator have fine grained control over traffic patterns. Such fine grained control, especially at a Layer 2-4 flow level, could lead to more efficient network designs and enable network operators to offer value added services to their customers. With this background, Dynamic Flow Management addresses the following problems in an OpenDaylight framework. The first problem is that of large flow load balancing and the second one is that of DDoS mitigation.

1.1. Large Flow Load Balancing

Networks extensively deploy LAG and ECMP for bandwidth scaling. Network traffic can be predominantly categorized into two traffic types: long-lived large flows and other flows (which include long-lived small flows, short-lived small/large flows) [OPSAWG-large-flow]. Stateless hash-based techniques [ITCOM, RFC 2991, RFC 2992, and RFC 6790] are often used to distribute both long-lived large flows and other flows over the components in a LAG/ECMP. However the traffic may not be evenly distributed over the component links due to the traffic pattern.

Long-lived large flow load balancing techniques can be used for achieving the best network bandwidth utilization with LAG/ECMP. These techniques are described in detail in [OPSAWG-large-flow] and [I2RS-large-flow] and summarized here. At a high-level, the technique involves recognizing large flows and rebalancing them to achieve optimal load balancing. Large flows may be recognized within a network element, or via analysis of flow information collected from the network element using protocols such as IPFIX [RFC 7011] or sFlow [sFlow-v5]. Once a large flow has been recognized, it must be signaled to a management entity that makes the rebalancing decision. Finally, the rebalancing decision is communicated to the routers to program the forwarding plane.

1.2. DDoS Mitigation

Layer 3-4 based DDoS attacks are an ongoing problem in today's networks. Example of Layer 3-4 based DDoS attacks are [FDDOS]:

- SYN Flood Attack: Fake TCP connections are setup which result in table overflows in stateful devices.
- UDP Flood Attack: Servers are flooded with UDP packets which results in consumption of bandwidth and CPU. These can be used to target specific services by attacking, e.g., DNS servers and VOIP servers.
- Christmas Tree Flood Attack: TCP packets from non-existent connections with flags other than the SYN flag sent to servers result in consumption of more CPU than normal packets because of the effort required discarding them.

Typically, the above attacks are not from a single host or source IP address; multiple hosts with different source IP addresses working in tandem cause these attacks – hence the term Distributed DoS or DDoS.

The DDoS use case involves recognizing large flows and performing various types QoS actions on the recognized flows based on configured policies. Large flows may be recognized within a

router, or using the aid of an external management entity such as an IPFIX [RFC 7011] collector or a sFlow [sFlow-v5] collector. These techniques are described in detail in [I2RS-large-flow].

2 Dynamic Flow Management in OpenDaylight

The goal of the Dynamic Flow Management project proposal is to implement the relevant Northbound/Southbound APIs for Large Flow Local (LAG)/Global load-balancing and DDoS SDN applications in the OpenDaylight framework. It should be noted that all references to OpenFlow in this document refer to the OpenFlow-hybrid model described in section 5.1 of [OF-1.3]. In the following sections, the SDN Applications and the relevant APIs are described in detail.

The general approach to dynamic flow management is as follows

- Determine traffic patterns based on flow classification and other techniques (plugins from vendors)
- Leverage traffic patterns to feed into a rules engine. Such a rules engine could be a simple one with if-then-else type rules. Or, it could feed into a policy manager. Another option could be to feed these patterns into an optimization framework, the solution of which would decide what actions to take. For example, it could output modified routes for a class of flows.
- Take actions. Actions could be (and not limited to) to do rate limiting of flows, re-routing of flows, Virtual Machine (VM) placement etc.

3 Large Flow Local (LAG) Load Balancing

3.1 Hash based Local (LAG) load balancing

The issues in the current hash based LAG load balancing scheme in switches/routers are summarized below and depicted in Figure 1.

- LAG Hashing unaware of large or small flows
- Large flow can penalize small flows/other large flows
- Sub-optimal LAG bandwidth utilization

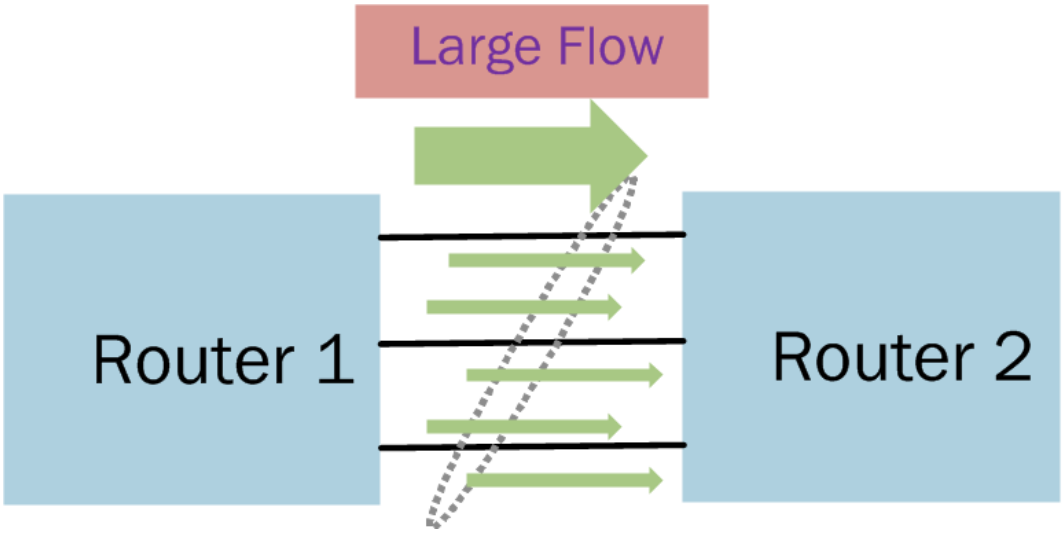

Figure 1: Hash based LAG load balancing

3.2 Large Flow Local (LAG) load balancing

The Large Flow Local (LAG) Load Balancing scheme in routers is summarized below and depicted in Figure 2.

- Identify congested egress links in the relevant LAG(s) in the network element
- Automatic large flow Recognition in the network
    - Inline line-rate large flow recognition in switches/routers; use IPFIX to convey recognized large flows to external IPFIX collector
    - Using sampling technologies like sFlow in switches/routers; large flow is recognized in an external collector like sFlow-RT
    - Automatically recognized large flow is conveyed to Large Flow Local (LAG) Load Balancing SDN Application
- Large Flow Local (LAG) Load Balancing SDN Application
    - Optimal placement of large flows (typically IP 5 tuple) in the LAG – OpenFlow PBR rule; reduce priority of large flows at edge – OpenFlow rule QoS action
    - Redistributing small flows -- adjust OpenFlow Group Table and vendor specific LAG table (assuming a non-standard Hybrid OpenFlow deployment)
- This results in Optimal bandwidth utilization for the LAG(s)

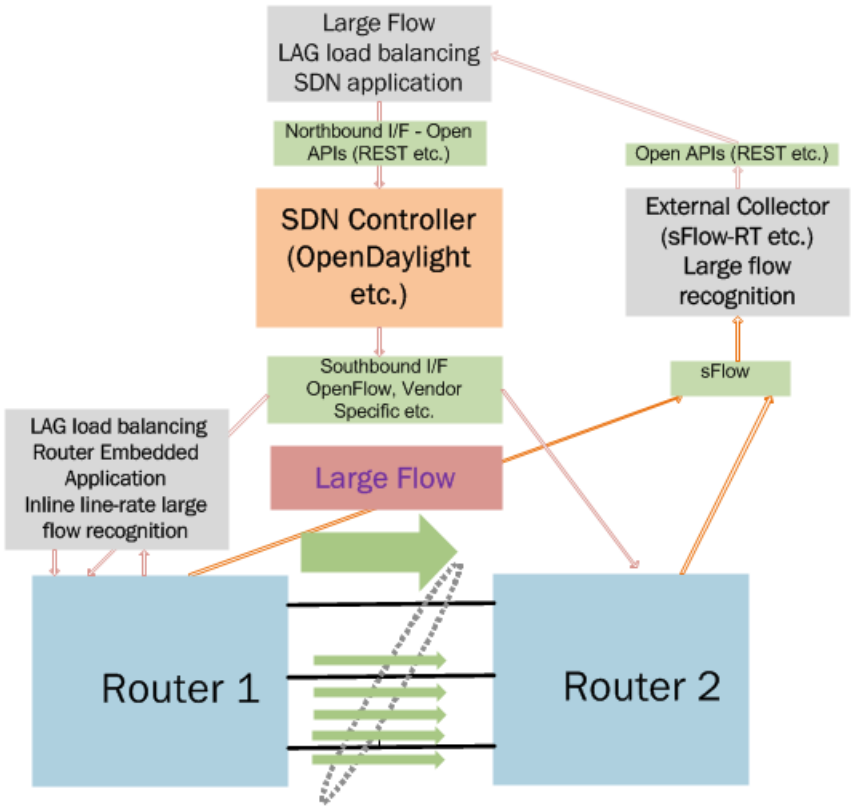

Figure 2: Large Flow Local (LAG) load balancing

4 Large Flow Global Load Balancing

4.1 Hash based ECMP load balancing

The issues in the current hash based ECMP load balancing scheme in routers are summarized below and depicted in Figure 3.

- ECMP Hashing unaware of large or small flows
- Large flow can penalize small flows/other large flows
- Sub-optimal network bandwidth utilization

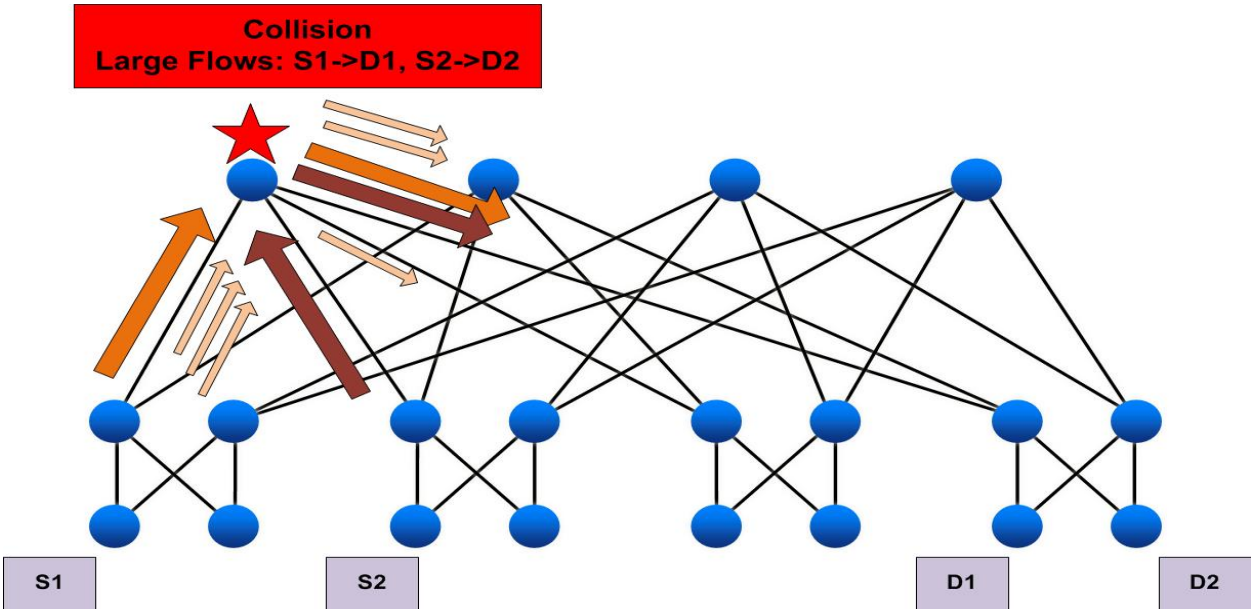

Figure 3: Hash based ECMP load balancing

4.2 Large Flow Global load balancing

The Large Flow Global Load Balancing scheme in routers is summarized below and depicted in Figure 4. More details are in [I2RS-large-flow].

- Identify congested egress links in all network elements
- Automatic large flow Recognition in the network
    - Inline line-rate large flow recognition in switches/routers; use IPFIX to convey recognized large flows to external IPFIX collector
    - Using sampling technologies like sFlow in switches/routers; large flow is recognized in an external collector like sFlow-RT
    - Automatically recognized large flow is conveyed to Large Flow Global Load Balancing SDN Application
- Application based signaling of large flows
    - The end use application, e.g. backup, signals the large flow to the Large Flow Global Load Balancing SDN Application
- Large Flow Global Load Balancing SDN Application
    - Globally optimal placement of large flows (typically IP 5 tuple) – hop-by-hop OpenFlow PBR rule; reduce priority of large flows at edge – OpenFlow rule QoS action
    - Redistributing small flows -- adjust OpenFlow ECMP table and vendor specific ECMP table (assuming a non-standard Hybrid OpenFlow deployment)
- This results in Optimal network bandwidth utilization

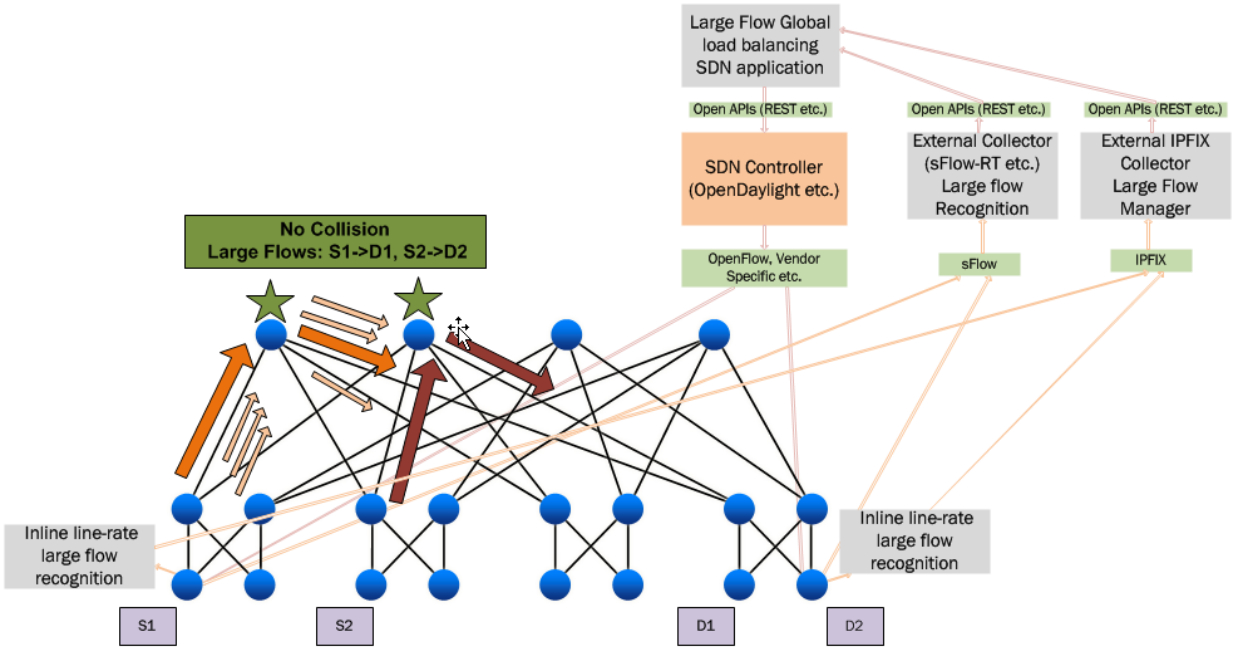

Figure 4: Large Flow Global load balancing

5 DDoS Mitigation

5.1 Current Layer 2-4 DDoS Mitigation

The issues in the current DDoS mitigation schemes are summarized below

- Need dedicated L7 DDoS appliance for Layer 2-4 DDoS detection/mitigation
- Switches /Routers support only static firewall rule (L2-L4 ACL) configuration – scalability issues in protecting all servers

5.2 Layer 2-4 DDoS Mitigation as an SDN Application

The Layer 2-4 DDoS mitigation scheme in routers is summarized below and depicted in Figure 5. More details are in [I2RS-large-flow].

- Automatic large flow Recognition in the network
    - Inline line-rate large flow recognition in switches/routers; use IPFIX to convey recognized large flows to external IPFIX collector
    - Using sampling technologies like sFlow in switches/routers; large flow is recognized in an external collector like sFlow-RT
    - Automatically recognized large flow is conveyed to Layer 2-4 DDoS mitigation SDN Application
- Layer 2-4 DDoS Mitigation SDN Application
    - Apply QoS policies such as dropping, rate-limiting re-marking to mitigate the effect of DDoS attacks
    - Apply nexthop redirection policies to redirect traffic to a scrubber appliance for further examination
- This results in DDoS mitigation

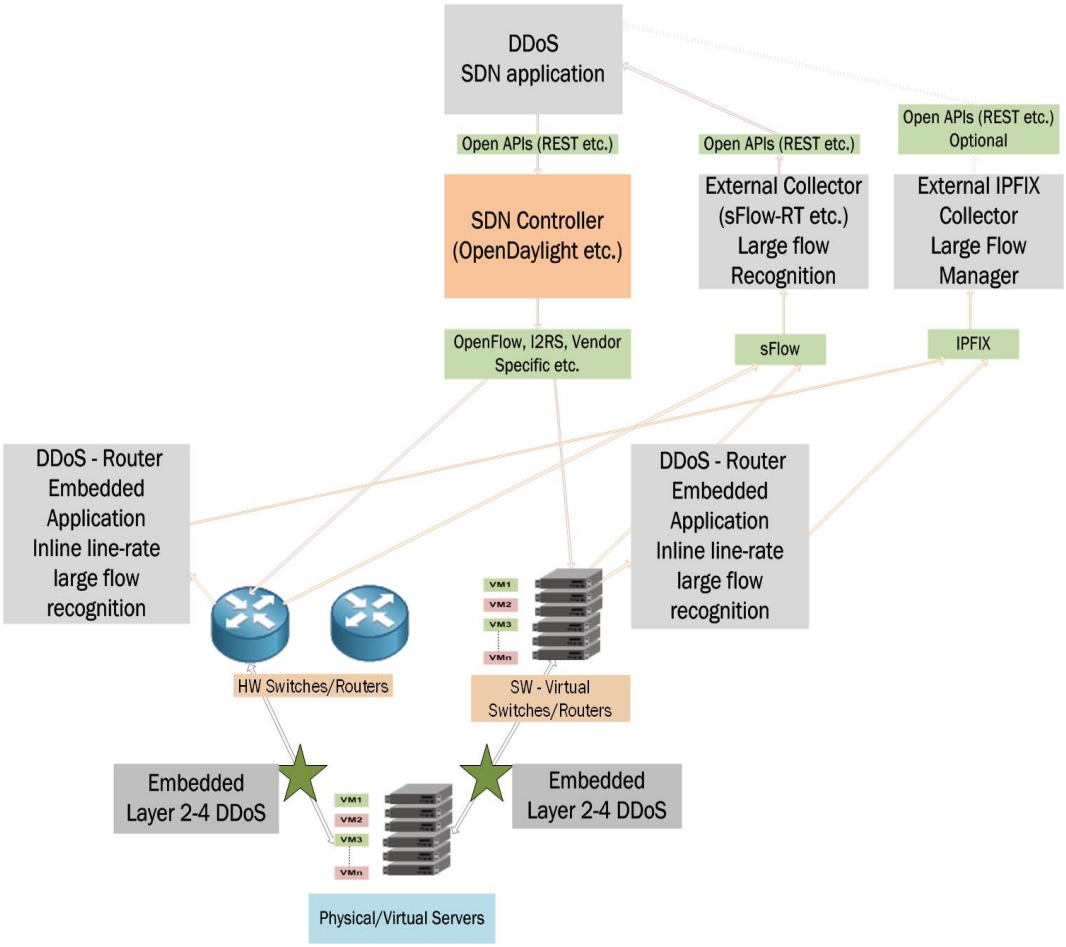

Figure 5: Layer 2-4 DDoS Mitigation as an SDN Application

6 Southbound API Requirements for OpenDaylight

All the Southbound API requirements are explained in detail in IETF Working group draft [OPSAWG-large-flow] -- Section 5, Information Model. The new Southbound API requirements for OpenDaylight are detailed below.

Link Utilization

- For normal speed links, use Interface table (iftable) MIB [RFC 1213]. For high speed links, use etherStatsHighCapacityTable MIB [RFC 3273]. For link utilization per priority, use Remote Network Monitory MIB Extensions [RFC 2613]. These are needed in an IPFIX implementation [RFC 7011]. These are not needed in a sFlow based implementation [sflow-v5] [sFlow-LAG] where the counter push mechanism for the interface counters can be leveraged.
    - Consider doing this as part of SNMP4SDN

Vendor Specific extensions to OpenFlow (section B.6.6 of [OF-1.3]) for manipulating

- Group table [OF1.3 sec 6.5]
- LAG [802.1ax] /ECMP [RFC 2992] table

7 Northbound API Requirements for OpenDaylight

This would be very similar to the Southbound API requirements explained in section 6 above, with relevant abstractions as needed

Link Utilization

- REST APIs for the requirements in section 6 above.

Vendor Specific extensions

- Abstracted to Group/LAG/ECMP table manipulation based on weights for the requirements in Section 6.

8 Usage by applications of the Northbound API of OpenDaylight

The Application Northbound API requirements are explained in detail in IETF Working group draft [OPSAWG-large-flow] -- Section 5, Information Model. These are detailed below.

8.1 Application Configuration Parameters

The following parameters are required for the configuration of large flow load balancing and DDoS mitigation applications

- Large flow recognition parameters:
  - Observation interval: The observation interval is the time period in seconds over which the packet arrivals are observed for the purpose of large flow recognition.
  - Minimum bandwidth threshold: The minimum bandwidth threshold would be configured as a percentage of link speed and translated into a number of bytes over the observation interval. A flow for which the number of bytes received, for a given observation interval, exceeds this number would be recognized as a large flow.
  - Minimum bandwidth threshold for large flow maintenance: The minimum bandwidth threshold for large flow maintenance is used to provide hysteresis for large flow recognition. Once a flow is recognized as a large flow, it continues to be recognized as a large flow until it falls below this threshold. This is also configured as a percentage of link speed and is typically lower than the minimum bandwidth threshold defined above.
- Match field selection:
  - Flexible selection of Layer 2/3/4 fields in the packet header
- Action selection:
  - Drop
  - Re-mark: IP DSCP, 802.1p
  - Rate-limit: Meter id
  - Redirection: IP Nexthop, Tunnel id, Physical Port, LAG id

8.1.1 Parameters unique to Large Flow Load Balancing

Configuration parameters unique to large flow load balancing application are listed below.

- Imbalance threshold: the difference between the utilization of the least utilized and most utilized component links. Expressed as a percentage of link speed.
- Rebalancing interval: the minimum amount of time between rebalancing events. This parameter ensures that rebalancing is not invoked too frequently as it impacts frame ordering.

These parameters may be configured on a system-wide basis or it may apply to an individual LAG.

8.1.2 Parameters unique to DDoS Mitigation

Configuration parameters unique to DDoS Mitigation application are listed below.

- Specific attacks
    - Flood attacks:
        - UDP Flood: Destination IP, IP Protocol UDP, UDP Destination Port
        - SYN Flood: Destination IP, IP Protocol TCP, TCP SYN Flag
        - Ping Flood: Destination, IP Protocol ICMP
    - Reflection attacks:
        - NTP Reflection Attack: Destination IP, IP Protocol UDP, UDP Source Port NTP
        - DNS Reflection Attack: Destination IP, IP Protocol UDP, UDP Source Port DNS

8.2 System Configuration and Identification Parameters

The following system configuration parameters are required for the configuration of large flow load balancing and DDoS mitigation:

- IP address: The IP address of a specific router that the feature is being configured on, or that the large flow placement is being applied to.

8.2.1 Parameters unique to Large Flow Load Balancing

System configuration parameters unique to large flow load balancing are listed below.

- LAG ID: Identifies the LAG. The LAG ID may be required when configuring this feature (to apply a specific set of large flow identification parameters to the LAG) and will be required when specifying flow placement to achieve the desired rebalancing.

8.3 Monitoring

8.3.1 Parameters unique to Large Flow Load Balancing

The following monitoring parameters are required. Applications like EMS can take advantage of this.

- Number of times rebalancing was done
- Time since the last rebalancing event

8.3.2 Parameters unique to DDoS Mitigation

The following monitoring parameters are required. Applications like EMS can take advantage of this.

- Attacks which have been detected and the actions which have been applied
- Time duration for which the detected attack lasted

9 Committed Development Resources

Company: Brocade Communications

- Ram (Ramki) Krishnan
- Varma Bhupatiraju

Company: Ericsson

- Sriganesh Kini

Organization: Amrita University

- Krishnakumar Rajagopal
- Naveen Narayan

10 Potential Code Committers

Company: Brocade Communications

- Muhammad Durrani

Company: Ericsson

Company: Cisco

- Debo Dutta
- Yathiraj Udupi

11 References

[OPSAWG-large-flow] Krishnan, R. et al., “Mechanisms for Optimal LAG/ECMP Component Link Utilization in Networks,” February 2014.

[I2RS-large-flow] Krishnan, R. et al., “I2RS Large Flow Use Case,” April 2014.

[802.1AX] IEEE Standards Association, "IEEE Std. 802.1AX-2008 IEEE Standard for Local and Metropolitan Area Networks – Link Aggregation", 2008.

[ITCOM] Jo, J., et al., “Internet traffic load balancing using dynamic hashing with flow volume,” SPIE ITCOM, 2002.

[RFC 2991] Thaler, D. and C. Hopps, "Multipath Issues in Unicast and Multicast," November 2000.

[RFC 6790] Kompella, K. et al., “The Use of Entropy Labels in MPLS Forwarding,” November 2012.

[RFC 1213] McCloghrie, K., “Management Information Base for Network Management of TCP/IP-based internets: MIB-II,” March 1991.

[RFC 2992] Hopps, C., "Analysis of an Equal-Cost Multi-Path Algorithm," November 2000.

[RFC 3273] Waldbusser, S., “Remote Network Monitoring Management Information Base for High Capacity Networks,” July 2002.

[RFC 7011] Claise, B., “Specification of the IP Flow Information Export (IPFIX) Protocol for the Exchange of IP Traffic Flow Information,” September 2013.

[sFlow-LAG] Phaal, P. and A. Ghanwani, “sFlow LAG counters structure,” http://www.sflow.org/sflow_lag.txt, September 2012.

[sFlow-v5] Phaal, P. and M. Lavine, “sFlow version 5,” http://www.sflow.org/sflow_version_5.txt, July 2004.

[ATIS-SDN-PAPER] Krishnan, R. et al., “ATIS SDN FOCUS GROUP White Paper - Operational Opportunities and Challenges of SDN/NFV Programmable Infrastructure”, November 2013

[OF-1.3] OpenFlow Switch Specification version 1.3.0 June 25, 2012 - openflow-spec-v1.3.0

[RFC 2613] Waterman, R., "Remote Network Monitoring MIB Extensions for Switched Networks Version 1.0, " June 1999