IPFIX
Internet Draft
Intended status: Informational
Expires: August 2013
February 23, 2013

R. Krishnan
D. Meyer
Brocade Communications
Ning So
Tata Communications

Flow Aware Packet Sampling Techniques

draft-krishnan-ipfix-flow-aware-packet-sampling-02.txt

Status of this Memo

Copyright Notice

Abstract

The demands on the networking infrastructure and thus the switch/router bandwidths are growing exponentially; the drivers are bandwidth hungry rich media applications, inter data center communications etc. Using sampling techniques, for a given sampling rate, the amount of samples that need to be processed is increasing exponentially. This draft suggests flow aware sampling techniques for handling various scenarios with minimal sampling overhead.

Table of Contents

1. Introduction

Packet sampling techniques in switches and routers provide an effective mechanism for approximate detection of various types of flows -- long-lived large flows and other flows (which include long-lived small flows, short-lived small/large flows) with minimal packet replication bandwidth overhead. A large percentage of the packet samples comprise of long-lived large flows and a small percentage of the packet samples comprise of other flows. The long-lived large flows aka top-talkers consume a large percentage of the bandwidth and small percentage of the flow space. The other flows, which are the typical cause of security threats like Denial of Service (DOS) attacks, Scanning attacks etc., consume a small percentage of the

bandwidth and a large percentage of the flow space. This draft explores light-weight techniques for automatically detecting the top-talkers in real-time with a high degree of accuracy and sampling only the other flows -- this makes security threat detection more effective with minimal sampling overhead.

1.1. Acronyms

DOS: Denial of Service

GRE: Generic Routing Encapsulation

MPLS: Multi Protocol Label Switching

NVGRE: Network Virtualization using Generic Routing Encapsulation

TCAM: Ternary Content Addressable Memory

STT: Stateless Transport Tunneling

VXLAN: Virtual Extensible LAN

1.2. Terminology

Large flow(s): long-lived large flow(s)

Small flow(s): long-lived small flow(s) and short-lived small/large flow(s)

2. Flow Aware Packet Sampling

The steps in flow aware packet sampling are described below

1) Large Flow Recognition in switches and routers:

From a bandwidth and time duration perspective, in order to identify large flows in switches and routers, we define an observation interval and observe the bandwidth of the flow over that interval. A flow that exceeds a certain minimum bandwidth threshold over that observation interval would be considered a large flow. For identifying large flows, use the techniques described in Section 2.1. This helps in identifying the large flows aka top-talkers in real-time with a high degree of accuracy in switches and routers.

2) Large Flow Classification:

The identified large flows can be broadly classified into 2 categories as detailed below.

a. Well behaved (steady rate) large flows, e.g. video streams

b. Bursty (fluctuating rate) large flows e.g. Peer-to-Peer traffic

The large flows can be sampled at a low rate for further analysis or need not be sampled. If desired, the large flows could be exported to a central entity, for e.g. sFlow Collector, for further analysis.

3) Small Flow Processing:

The small flows (excluding the large flows) can be sampled at a normal rate. The small flows can be examined for determining security threats like DOS attacks (for e.g. SYN floods), Scanning attacks etc. [LANCOPE]

Thus, we can see that, security threat detection is possible with minimal sampling overhead.

For packet sampling, it is recommended to use PSAMP -- [RFC 5474], [RFC 5475], [RFC 5476], [RFC 5477] or sFlow -- [sFlow-v5].

2.1. Large Flow Recognition

2.1.1. Flow Identification

A flow (large flow or small flow) can be defined as a sequence of packets for which ordered delivery should be maintained. Flows are typically identified using one or more fields from the packet header from the following list:

. Layer 2: source MAC address, destination MAC address, VLAN ID.

. IP header: IP Protocol, IP source address, IP destination address, flow label (IPv6 only), TCP/UDP source port, TCP/UDP destination port.

. MPLS Labels.

For tunneling protocols like GRE, VXLAN, NVGRE, STT, etc., flow identification is possible based on inner and/or outer headers. The above list is not exhaustive. The mechanisms described in this

document are agnostic to the fields that are used for flow identification.

2.1.2. Criteria for Identifying a Large Flow

From a bandwidth and time duration perspective, in order to identify large flows we define an observation interval and observe the bandwidth of the flow over that interval. A flow that exceeds a certain minimum bandwidth threshold over that observation interval would be considered a large flow.

The two parameters -- the observation interval, and the minimum bandwidth threshold over that observation interval -- should be programmable in a switch or a router to facilitate handling of different use cases and traffic characteristics. For example, a flow which is at or above 10 Mbps for a time period of at least 30 minutes could be declared a large flow.

An optional parameter is a policy specification (for e.g. identify flows only from a given IP source and/or destination address)

2.1.3. Automatic Recognition

Implementations can perform automatic recognition of large flows in a switch or a router -- it is an inline solution and would be expected to operate at line rate.

The advantages and disadvantages of automatic recognition are:

Advantages:

- Accurate and performed in real-time.

Disadvantages:

- Not supported in many switches and routers.

As mentioned earlier, the observation interval for determining a large flow and the bandwidth threshold for classifying a flow as a large flow should be programmable parameters in a switch or a router.

The implementation of automatic recognition of large flows is vendor dependent. Below is a suggested technique.

This technique uses a counting Bloom filter using thresholding and periodic reset. This technique requires a few tables -- a flow table, and multiple hash tables.

The flow table comprises entries which are programmed with packet fields for flows that are already known to be large flows and each entry has a corresponding byte counter. It is initialized as an empty table (i.e. none of the incoming packets would match a flow table entry).

The hash tables each have a different hash function and comprise entries which are byte counters. The counters are initialized to zero and would be modified as described by the algorithm below.

Step 1) If the large flow exists in the flow table (for e.g. TCAM), increment the counter associated with the flow by the packet size. Else, proceed to Step 2.

Step 2) The hash function for each table is applied to the fields of the packet header and the result is looked up in parallel in corresponding hash table and the associated counter corresponding to the entry that is hit in that table is incremented by the packet size. If the counter exceeds a programmed byte threshold in the observation interval (this counter threshold would be set to match the bandwidth threshold) in the entries that were hit in all of the hash tables, a candidate large flow is learnt and programmed in the flow table and the counters are reset.

Additionally, the counters in all of the hash tables must be reset every observation interval.

There may be some false positives due to multiple small flows masquerading as a large flow. The number of such false positives is reduced by increasing the number of parallel hash tables using different hash functions. There will be a design tradeoff between size of the hash tables, the number of hash tables, and the probability of a false positive.

This technique for automatic recognition is also suggested in [draft-krishnan-opsawg-large-flow-load-balancing] -- please refer to the draft for more details on the algorithm.

2.1.3.1. Applicability of suggested technique

The suggested technique for automatic recognition works well for standard applications generating large flows, for e.g. video content like movies and catch-up episodes, backup transactions etc. with a detection time of approximately 30-60 seconds. These detection times ensure that short-lived large flows, for e.g. HD video clips, are not unnecessarily recognized.

2.1.3.2. Enhancements to suggested technique

If faster flow recognition times are desired (much shorter than 30s), the suggested technique may pose the following problem that the effective filtered flow size is phase-dependent: that is, relatively smaller constant-rate flows, for e.g. HD video clips, beginning early within a counting Bloom filter reset interval would be unnecessarily detected with the same probability as relatively larger flows beginning toward the interval.

[VRM] suggests techniques for addressing the above problem using rotating conservative counting Bloom filters with periodic decay.

2.1.3.1. Handling Inactive Large Flows

Once a flow has been recognized as a large flow, it should continue to be recognized as a large flow as long as the traffic received during an observation interval exceeds some fraction of the bandwidth threshold, for example 80% of the bandwidth threshold. If the traffic received during an observation interval falls below a fraction of the bandwidth threshold, the large flow should be removed from the flow-table.

2.1.4. Simulation

Simulation results for flow aware packet sampling are presented in Appendix A. The goal of the simulation is to demonstrate the effectiveness of flow aware packet sampling in a multi-tenant video streaming data center.

3. Acknowledgements

The authors would like to thank Juergen Quittek, Brian Carpenter, Michael Fargano, Michael Bugenhagen, Jianrong Wong and Brian Trammell for all the support and valuable input.

4. IANA Considerations

This memo includes no request to IANA.

5. Security Considerations

This document does not directly impact the security of the Internet infrastructure or its applications. In fact, it proposes techniques which could help in identifying a DOS attack pattern.

6. Data Model Considerations

In Section 2, for exporting the identified large flows to an external entity, it is recommended to use IPFIX protocol [RFC 5101].

Section 2.1.2 defines programmable parameters in switches and routers for automatic identification. IETF could potentially consider a standards-based activity around defining a data model for moving this information from a central management entity to the switch/router.

7. References

7.1. Normative References

7.2. Informative References

[RFC 5474] N. Duffield et al., "A Framework for Packet Selection and Reporting", March 2009.

[RFC 5475] T. Zseby et al., "Sampling and Filtering Techniques for IP Packet Selection", March 2009.

[RFC 5476] B. Claise, Ed. et al., "Packet Sampling (PSAMP) Protocol Specifications", March 2009.

[RFC 5477] T. Dietz et al., "Information Model for Packet Sampling Exports", March 2009.

[RFC 5101] B. Claise, "Specification of the IP Flow Information Export (IPFIX) Protocol for the Exchange of IP Traffic Flow Information", January 2008

[LANCOPE] "Benefits of Flow Analysis Using sFlow: Network Visibility, Security and Integrity" http://www.lancope.com/files/Lancope_Generic_sFlow_WP.pdf

[draft-krishnan-opsawg-large-flow-load-balancing] R. Krishnan et al., "Best Practices for Optimal LAG/ECMP Component Link Utilization in Provider Backbone Networks", February 2013

[VRM] G. Bianchi et al., "Measurement Data Reduction through Variation Rate Metering", INFOCOM 2010

Appendix A: Simulation of Flow aware packet sampling

Goal:

Demonstrate the effectiveness of flow aware packet sampling in a practical use case, for e.g. multi-tenant video streaming in a data center.

Test Topology:

Multiple virtual servers (server hosted on a virtual machine) connected to a virtual switch (vSwitch) which in turn connects to the data center network using a 10Gbps ethernet interface.

2 virtual servers are active.

First virtual server

- Traffic types
  - HD MPEG-4 video streams (bit rate 10Mbps) - 100 - 1Gbps
  - SD MPEG-2 video streams (bit rate 4Mbps) - 300 - 1.2Gbps
  - Other traffic - 500Mbps (Video clips, DOS attacks (for e.g. SYN floods), Scanning attacks etc.)
- Aggregate traffic - 2.7Gbps

Second virtual server

- Traffic types
  - HD MPEG-4 video streams (bit rate 10Mbps) - 50 - .5Gbps
  - SD MPEG-2 video streams (bit rate 4Mbps) - 500 - 2.0Gbps
  - Backup transaction - 100Mbps
  - Other traffic - 500Mbps (Video clips, DOS attacks (for e.g. SYN floods), Scanning attacks etc.)
- Aggregate traffic - 3.1Gbps

Total traffic on 2 servers - 5.8Gbps

Existing techniques:

Normal sampling rate - 1:1000

Total sampled traffic = 5.8Gbps/1000 = 5.8Mbps

Flow aware sampling technique:

Large flow recognition parameters

- Observation interval for large flow - 60 seconds
- Minimum bandwidth threshold over the observation interval - 2Mbps

Aggregate bit rate of large flows = 4.8Gbps

Aggregate bit rate of small flows = 1Gbps

Low sampling rate of large flows - 1:10000

Normal sampling rate of small flows - 1:1000

Total sampled traffic = 4.8Gbps/10000 + 1Gbps/1000 = 1.48Mbps

Percentage improvement in sampling (most of the samples are only small flows) = (5.8 - 1.48)/5.8 ~= 78%

The small flows can be examined in a central entity like sFlow Collector for determining security threats like DOS attacks, Scanning attacks etc. Thus, we can see that, security threat detection is possible with minimal sampling overhead.

Authors' Addresses

Ram Krishnan
Brocade Communications
San Jose, 95134, USA

Phone: +001-408-406-7890
Email: ramk@brocade.com

David Meyer
Brocade Communications
San Jose, 95134, USA

Phone: +001-408-333-4193
Email: dmm@1-4-5.net

Ning So
Tata Communications
Plano, TX 75082, USA

Phone: +001-972-955-0914
Email: ning.so@tatacommunications.com