Internet Engineering Task Force
Internet-Draft
Intended status: Informational
Expires: November 20, 2008

Jan Novak, Ed.
Cisco Systems

May 21, 2008

IP Flow Information Accounting and Export Benchmarking Methodology
draft-janovak-bmwg-ipflow-meth-00.txt

Abstract

This document provides methodology and framework for quantifying performance implications of enabling selective monitoring of IP flows on a network device and export of this information to a collector as specified in [RFC5101].

Table of Contents

1. Introduction

Monitoring of IP flows (Flow Monitoring) on network devices is an application which has numerous usage in both service provider and enterprise segments as detailed in [RFC3917]. The question any user considering its deployment asks is - And what performance implication Flow Monitoring will have on my network ?

The network operator concern is always twofold:

a. what will be CPU usage
b. what will be the forwarding performance

when enabling Flow Monitoring on network devices. This document defines set of traffic parameters which influence most the performance of network devices and provides methodology how to measure effects of Flow Monitoring from both points of view as specified above.

IETF IPFIX working group concentrates its effort on standardising the Flow Monitoring data export to an external collecting device while the actual application providing the data inside of a network device is left to the implementors liberty. The goal of this document is to address both these aspects of Flow Monitoring, since typical implementation will allow to separately enable monitoring (for the users to query the data using command line interface or SNMP) and export of the IP flow information.

1.1. Requirements Language

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in RFC 2119 [RFC2119].

2. Terminology

2.1 Existing Terminology

Device Under Test (DUT)	[RFC2285, section 3.1.1]
IP Traffic Flow/Flow	[RFC5101, section 2]
Flow Key	[RFC5101, section 2]
Flow Record (FR)	[RFC5101, section 2]
Observation Point	[RFC5101, section 2]
Exporter	[RFC5101, section 2]

Collector	[RFC5101, section 2]
Control Information	[RFC5101, section 2]
Data Stream	[RFC5101, section 2]
Flow Expiration	[IPFIX-ARCH, section 5.1.1]
Flow Export	[IPFIX-ARCH, section 5.1.2]
Throughput	[RFC1242, section 3.17]

2.2 Defined Terminology

2.2.1 Cache

Definition:
Memory area held and dedicated by the DUT to store Flow Record information

2.2.2 Cache Size

Definition:
The size of the cache in terms of how many entries of Flow Records the cache can hold

Discussion:
This term is typically represented as a configurable option in the particular Flow Monitoring implementation. It needs to be at least known in order to define the tests circumstances properly. Its highest value will depend on the memory available in the network device.

Measurement units:
Number of entries

2.2.3 Flow Expiration Rate (FER)

Definition:
Number of Flow Records which expire (as defined by the Flow Expiration term) from the Cache within a time interval

Measurement units:
Number of Flow Records per second

2.2.4 Active Timeout

Definition:
The time interval from the time when first packet of a particular Flow was seen till the Flow will be expired while there are still

packets arriving to the DUT which belong to the Flow.

Discussion:
This term is typically represented as a configurable option in the
particular Flow Monitoring implementation. See section 5.1.1 of
[IPFIX-ARCH] for more detailed discussion.

Measurement units:
Seconds

2.2.5 Inactive Timeout

Definition:
The time interval from the time when last packet has been observed
for a particular Flow till the Flow is expired from the Cache, while
no packets which belong to the Flow are seen during the whole period.

Discussion:
This term is typically represented as a configurable option in the
particular Flow Monitoring implementation. See section 5.1.1 of
[IPFIX-ARCH] for more detailed discussion.

Measurement units:
Seconds

2.2.6 Baseline Processor Utilisation (BPU)

Definition:
The Processor (CPU) utilisation under steady traffic stream without
Flow Monitoring configured.

Discussion:
The CPU utilisation SHOULD BE collected from the DUT after sufficient
time interval under the test traffic stream to obtain reliable 1 minute
average CPU usage.
The recommended time before collecting the 1 minute average CPU usage
is 5 to 10 minutes.

Measurement units:
Percent

2.2.7 Flow Monitoring Processor Utilisation (FMPU)

Definition:
The Processor (CPU) utilisation under steady traffic stream with

Flow
      Monitoring configured.

   Discussion:
      The CPU utilisation SHOULD BE collected from the DUT after
sufficient
      time interval under the test traffic stream to obtain reliable 1
minute
      average CPU usage.
      The recommended time before collecting the 1 minute average CPU
usage

is 5 to 10 minutes.

Measurement units:
Percent

3. Test Set Up

3.1 Testbed Topology

The test set-up is identical to the one used by [RFC2544]:

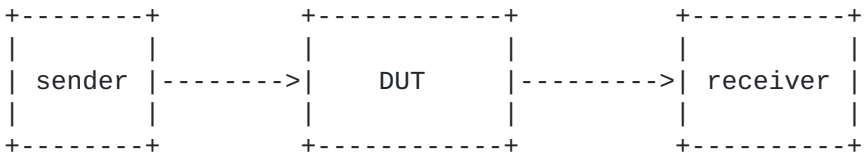

Figure 1

The ideal way to implement the test is using one tester with a sending port and a receiving port. This allows for an easy check if all the sent traffic by the sender was transmitted by the DUT and received at the receiver.

If the effects of enabling Flow Monitoring on several interfaces are of concern, the topology can be expanded with several input and output ports.

3.2 Basic Packet Forwarding Set Up

DUT needs to have very basic configuration just allowing IP packet forwarding without any use of dynamic IP routing protocols. The only objective of the configuration is to allow transmission of the test traffic with as low interference of any control (routing) traffic as possible.

3.3 Flow Monitoring Configuration

The DUT Observation Points configuration needs to be decided upon depending on the interest and scope of the testing as follows:

a. input port/ports only
b. output port/ports only
c. both input and output

The testing procedures are otherwise same for all these possible configurations.

3.4 Frame Formats

Frame formats to use are specified in [RFC2544] section 8.

3.5 Frame Sizes

Frame sizes to use are specified in [RFC2544] section 9.

3.6 Flow Records

The Flow Record definition is very implementation specific. A Flow Monitoring implementation might allow only for fixed Flow Record definition, based on the most common IP parameters in the IPv4 or IPv6 headers - like source and destination IP addresses, IP protocol numbers or transport level port numbers. Another implementation might allow the user to actually define his own completely arbitrary Flow Record to monitor the traffic. The requirement for the tests defined in this document is only the need for a large number of Flow Records in the Cache. The Flow Keys needed to achieve that will typically be source and destinations IP addresses and transport level port numbers.

3.7 Traffic Definitions

The traffic definitions in the sections below serve only as examples how to achieve the particular test objectives with certain Flow Record definition, the exact set-up will therefore always be Flow Monitoring implementation specific.

4. Processor Utilisation Metrics

Every Network Operator carefully monitors Processor (CPU) utilisation
on all network devices for two major reasons:

a. increased CPU usage can indicate unwanted network activity like
   Denial of Service attacks or faulty device or network connection
b. each network device typically runs quite large routing protocols
   tables and needs some free processing power to maintain routing
   and forwarding

There is no commonly accepted limit to the CPU usage but typically usage in the range of 70 - 80 % becomes already a critical issue.

Flow Monitoring can be run on different network device architectures from centralised software only, distributed, to fully hardware accelerated. Irrespective of its architecture, the device will have some CPU and some part of Flow Monitoring tasks will always need to be processed on that CPU even though some parts of the functionality could be off loaded to the specialised hardware. The measurements in this section can be therefore performed either on

the CPU which performs all the routing tasks or a CPU on some device linecard for a distributed system depending what represents the major concern and where are the Flow Monitoring tasks running.

The purpose of this section is to define a set of metrics and tests to

measure influence of the Flow Monitoring on the CPU of a network device.

The following CPU usage metrics are defined and measured here:

a. Cache States Maintenance CPU usage
b. Cache States Update CPU usage
c. Flow Expiration Rate CPU usage
d. Flow Export Rate CPU usage
e. Cache Overflow CPU usage

4.1 Executing the metrics measurements

The CPU tests methodology is same for all the tests specified in this section. The whole test course step by step SHOULD be executed as follows:

a. Configure DUT for base forwarding as specified in the section 3.2
b. Configure traffic streams on the Sender and configure Receiver to just sink the traffic. The Receiver SHOULD perform checks that the traffic sent by the Sender was successfully forwarded by the UUT to the Receiver

c. Perform measurements in a loop from 1 to n, where n is the number of defined streams:

1. Start traffic stream n
2. Measure BPU
3. Stop traffic, clear all packet statistics and apply Flow Monitoring configuration on the DUT as specified in the section 3.3 and as defined in the particular test section
4. Start traffic stream n
5. Wait to populate the cache and verify Flow Monitoring statistics
6. Measure FMCU
7. Stop traffic and clean all Flow Monitoring DUT configuration

4.2 Cache States Maintenance

4.2.1 Metrics Definition

CPU usage needed on the DUT to maintain the Flow Record information held in the Cache in a completely static scenario without any changes to the stored information.

4.2.2 Measurement Procedure

To measure the Cache States Maintenance CPU utilisation the presence of

a large amount of Flows Records in the Cache is needed but with no Flow
Expiration from the Cache during the test and also no counter refresh

- the test traffic is sent just once to populate the cache.

4.2.3 Measurement Configuration

Flow Keys Definition:
Needs to allow for large numbers of unique Flow Records to be created
in the Cache

Cache Size:
Maximum configurable value on the network device.

Sender Traffic Definition:

Define n traffic streams while incrementing the number of unique Flow
Keys combinations in the increments of about 1/nth of the cache size,
leaving about 10% of the cache entries free at the maximum.

The total number of created Flow Records in the Cache MUST NOT exceed
the configured Cache Size at any point of the measurement.

Number of packets sent: each stream sends just the configured number of
unique Flow Keys values in one batch to just populate the cache before
the measurement starts

Flow Monitoring Configuration:

Inactive Timeout:
Inactive Timeout must be configured in such a way, that the Flow
Records get created in the Cache and never expire. The best value is
the maximum configurable Inactive Timeout.

Active Timeout:
Active Timeout MUST be configured in such a way that the active Flow
Records never expire from the Cache during the whole measurement
period with one of the defined traffic streams.

The Flow Monitoring statistics SHOULD be checked during the measurement
execution to verify that the measurement conditions have been reached as
specified - namely the number of entries and number of added entries in
the Cache SHOULD NOT change during and after the cache has been populated.

sent.

4.2.4 Analysing the Results

The test run will produce n triplets of values as follows:
"Number of Flow Records in the Cache" "BPU" "FMCU"

The CPU usage needed to maintain the states in the Cache is represented by the values difference (FMCU - BPU) and is a function of the number of states held in the Cache.

4.3 Cache States Update

4.3.1 Metrics Definition

CPU usage needed on the DUT to update the Flow Record information held in the Cache while the number of Flow Records does not change, only the counters (typically bytes and packets numbers) corresponding to each Flow Record are updated by the flowing traffic stream.

4.3.2 Measurement Procedure

To measure the Cache States Update CPU utilisation the presence of a large amount of Flows Records in the Cache is needed but with no Flow Expiration from the Cache during the test. The traffic needs to flow steadily at certain rate while updating the Flow Record counters.

4.3.3 Measurement Configuration

Flow Keys Definition:
Needs to allow for large numbers of unique Flow Records to be created in the Cache

Flow Record Definition - MUST contain counter fields like bytes and packets which get updated with each sent packet.

Cache Size:
Maximum configurable value on the network device.

Sender Traffic Definition:

Define n traffic streams while incrementing the packet rate. The number of unique Flow Keys combinations SHOULD be at about 90% of the configured Cache Size.

The total number of created Flow Records in the Cache MUST NOT exceed the configured Cache Size at any point of the measurement.

Number of packets sent: continuous traffic stream

Flow Monitoring Configuration:

Inactive Timeout:
Inactive Timeout must be configured in such a way, that the Flow Records get created in the Cache and never expire. The best

value is
         the maximum configurable Inactive Timeout.

      Active Timeout:
         Active Timeout MUST be configured in such a way that the
active Flow
         Records never expire from the Cache during the whole
measurement
         period with one of the defined traffic streams.

The Flow Monitoring statistics SHOULD be checked during the measurement execution to verify that the measurement condition have been reached as specified - namely the number of entries and number of added entries in the Cache SHOULD NOT change after the cache is fully populated.

4.3.4 Analysing the Results

The test run will produce n triplets of values as follows:
"Packet Rate" "BPU" "FMCU"

The CPU usage needed to update and maintain the states in the Cache is represented by the values difference (FMCU - BPU) and is a function of the number of updates per second - in this particular set-up it is equal to the packet rates.

4.4 Flow Expiration Rate

4.4.1 Metrics Definition

CPU usage needed on the DUT to expire at certain rate the Flow Record information held in the Cache while the number of Flow Records in the Cache never exceeds the configured Cache Size.

4.4.2 Measurement Procedure

To measure the Flow Expiration Rate CPU utilisation the traffic needs to populate the Cache at certain steady rate. The Flow Record needs to be expired from the Cache before it can be refreshed by the next packet with the same Flow Key parameters combination. The total amount of Flow Records in the Cache MUST NOT reach the configured Cache Size.

4.4.3 Measurement Configuration

Flow Keys Definition:
Needs to allow for large numbers of unique Flow Records to be created in the Cache

Cache Size:
Maximum configurable value on the network device.

Sender Traffic Definition:

Define n traffic streams while incrementing the packet rate. The number of unique Flow Keys combinations MUST be many multiples of the configured Cache Size.

The total number of created Flow Records in the Cache MUST NOT exceed the configured Cache Size at any point of the measurement. This can be achieved by the Flow Monitoring timers configuration as specified below.

Number of packets sent: continuous traffic stream

Flow Monitoring Configuration:

Inactive Timeout:
Inactive Timeout must be configured in such a way, that the Flow
Records get created in the Cache and expire before they can be refreshed by the next packet in the stream with the same parameters.
The Inactive Timeout value MUST be smaller than (90% configured Cache
Size) divided by the stream packet rate. This way the number of
active Flow Records in the Cache will never exceed the Cache Size.
The same thing can be achieved if the Flow Monitoring implementation
allows to configure Inactive Timeout equal to 0 seconds (e.g. immediate expiration).

The Flow Monitoring statistics SHOULD be checked during the measurement
execution to verify that the measurement conditions have been reached as
specified - namely the number of entries MUST NOT exceed the Cache Size
during the whole measurement with one of the defined traffic streams.

The number of Flow Records in the Cache entries SHOULD oscillate around
the value:

(Inactive Timeout * packet rate)

or ideally be stable and equal to that value.

4.4.4 Analysing the Results

The test run will produce N triplets of values as follows:
"Packet rate" "BPU" "FMCU"

Due to the Flow Monitoring and the test traffic configuration the packet
rate represents also the Flow Expiration Rate - each packet of the test
traffic streams creates one Flow and the Flows later expire at the same
rate the packets were sent and the Flows created.
The triplets can therefore be interpreted as:
"Flow Expiration Rate" "BPU" "FMCU"

The measured FMCU represents the CPU usage of three components:

a. BPU
b. Cache state updates and maintenance
c. Flow Expiration Rate

4.5 Flow Export Rate

4.5.1 Metrics Definition

CPU usage needed on the DUT to export at certain rate the Flow Record
information held in the Cache while the number of Flow Records in the Cache
never exceeds the configured Cache Size.

4.5.2 Measurement Procedure

Same as in 4.4.2, the DUT is in addition configured with Flow Export

4.5.3 Measurement Configuration

Flow Monitoring Configuration:

The DUT needs to be configured for Flow Export to one or more Collectors.
The Collectors do not need to exist neither any export data analysis needs
to be done. The DUT MUST have a route and forwarding adjacency to reach
all the Collectors. The export packets SHOULD exit the DUT on another
interface than the one used to connect the Receiver so that the test
traffic statistics on that interface are not polluted by the export
packets. The Flow Export statistics SHOULD be checked on the DUT and
compared to the expected Flow Expiration Rate. The Flow Export packets
exit interface SHOULD be checked for the packets statistics to make sure
the export packets indeed leave the DUT.

All the other measurement components need to be configured exactly same way as
in the section 4.4.

4.5.4 Analysing the Results

The test run will produce n triplets of values as follows:
"Flow Expiration Rate" "BPU" "FMCU"

The measured FMCU represents now the CPU usage of four components

a.) BPU
b.) Cache state update and maintenance
c.) Flow Expiration Rate
d.) Flow Export Rate

4.6 Cache Overflow

The common factor of sections 4.2 to 4.4 was that the number of Flow Records
created in the Cache never exceeded the configured Cache Size. This represents
a normal and very healthy network status. This picture can easily change in
the environment of a more busy network device - either having less resources
available for Flow Monitoring or simply more active Flow Monitoring due to
different traffic patterns. The Flow Monitoring implementation

specific
   processes which handle Cache maintenance can significantly change
under the
   condition where the amount of created Flow Records exceeds the
available Cache
   Size to hold them.

4.6.1 Metrics Definition

   CPU usage needed on the DUT to expire at certain rate the Flow
Record
   information held in the Cache while the number of Flow Records
created in the
   Cache always reaches and overflows the configured Cache Size.

4.6.2 Measurement Procedure

To measure the Cache Overflow CPU utilisation the traffic needs to fill the Cache at the beginning of the measurement. The number of unique Flow Key values combinations must be very large as compared to the configured Cache Size so that each packet always creates a new Flow Record and forces another Flow Record to be expired from the Cache.

4.6.3 Measurement Configuration

Flow Keys Definition:
Needs to allow for large numbers of unique Flow Records to be created in the Cache

Cache Size:
Configured to such a value so that the Cache gets filled up within short initial interfval of the stream sending.

Sender Traffic Definition:

Define n traffic streams while incrementing the packet rate. The number of unique Flow Keys combinations MUST be many multiples of the configured Cache Size.

The total number of created Flow Records in the Cache MUST exceed the configured Cache Size before the test starts.

Flow Monitoring Configuration:

Inactive Timeout:
Inactive Timeout SHOULD be configured to the largest possible value so that the flows do not expire from the Cache using ordinary expiration processes.
Under this Sender and Flow Monitoring configuration the number of Flow Records created in the Cache will exceed the configure Cache Size in the first seconds of sending the traffic.

Active Timeout:
Active Timeout SHOULD be configured equal or larger than Inactive Timeout.

The Flow Monitoring statistics SHOULD be checked during the test execution
to verify that the test conditions have been reached as specified - namely the
number of Flow Records in the Cache needs to be always very close or equal to
the configured Cache Size.

4.6.4 Analysing the Results

The test run will produce N triplets of values as follows:
"Packet rate" "BPU" "FMCU"

Due to the Flow Monitoring and the test traffic configuration the packet rate
represents also the Flow Expiration Rate - each packet of the test traffic stream
creates one Flow and the Flows later expire at the same rate the packets were sent
and the Flows created. The triplets can therefore be interpreted as:
"Flow Expiration Rate" "BPU" "FMCU"

The measured FMCU represents the CPU usage of three components:

a. BPU
b. Cache state updates and maintenance
c. Flow Expiration Rate under the condition of Cache overflow

5. Throughput Tests

The throughput tests can use the methodology as defined in [RFC2544] but in the presence of Flow Monitoring configuration on the DUT. This represents certain challenge to create controlled and well defined test environment also from the point of view of Flow Monitoring.
[RFC2544] defines frames formats, sizes and rates for the Throughput tests. From the prospective of Flow Monitoring these test streams can be used in two ways to create Cache as specified in the following sections.

5.1 Single Traffic Component

Section 12 of [RFC2544] discusses the use of protocol source and destination addresses for the Throughput tests. In order to perform Throughput measurement with Flow Monitoring enabled the defined Flow Keys MUST contain IP source and destination address. The IP Flow
Monitoring measurements 4.4, 4.5 and 4.6 can be executed using [RFC2544]
Throughput test methodology under these additional conditions:

a. the test traffic does not use just single unique pair of source and
destination address

b. the Sender allows to define test traffic as follows

1) define the test streams exactly as specified in the sections 4.4,
4.5 and 4.6
2) allow for a parameter to say send N (where N is an integer number
starting at 1 and incremented in small steps) packets with IP addresses A and B before changing both IP addresses to the next
value

This test traffic definition allows to execute the above defined IP FLow
Monitoring tests with one defined Flow Expiration rate while measuring at
the same time the DUT Throughput as defined in [RFC2544].

This set-up is the better option since it best simulates the life network

traffic scenario with Flows containing more than just one packet.

5.2 Two Traffic Components

The test traffic set-up in the section 5.1 might be difficult to achieve with commercial traffic generators. The way around it is to define two traffic components in the test traffic - one to populate Flow Monitoring Cache and the second one to execute the throughput test.

Flow Monitoring Test Traffic Component - the exact traffic definition as specified in the sections 4.4, 4.5 and 4.6.

Throughput Test Traffic Component - test traffic as specified by [RFC2544] but under the condition it MUST create just one Flow Record in the DUT Cache. In the particular set-up discussed here this would mean a traffic stream with just one pair of unique source and destination IP address (but could be avoided if Flow Keys were for example UDP/TCP source and destination ports).

The first traffic component will exercise the DUT CPU in terms of IP Flow activity while the second traffic component will measure the Throughput under the conditions of [RFC2544]. The traffic rates of first component can be simply added to the achieved Throughput value of the second component.

6. Evaluating Flow Monitoring Applicability

The results obtained for certain DUT allow for a preliminary analysis of a Flow Monitoring deployment based on Internet traffic analysis data provided by organisations like [CAIDA]. The data needed to make an estimate of a network device CPU usage spent on Flow Monitoring are as follows:

Average packet size: 350 bytes
Number of packets per IP Flow: 20

Expected data rate on the network device: 1 Gbit/s

This results in:

Expected packet rate: 357 000 pps

being (1 Gbit/s divided by 350 bytes/packet)

Flows per second: : 18 000

being (packet rate 357 000 pps divided by 20 packets per IP Flow)

Under constant load of traffic with the parameters above the network device will always run in the Cache Overflow mode (if the network is large enough the Flows will hardly ever repeat itself within few seconds needed to fill up lets say 300 000 entries Cache) and from pure Flow Monitoring point of view the CPU usage will be around 14 % - see

This needs to be add up on top of the Base CPU usage measurement for the corresponding traffic rate and packet sizes.

7. Acknowledgements

This work could have been performed thanks to the patience and support of Cisco System Netflow development team, namely Paul Aitken, Paul Atkins and Andrew Johnson. Thanks belong also to Aamer Akhter for initiating this work.

8. IANA Considerations

This document requires no IANA considerations.

9. Security Considerations

Documents of this type do not directly affect the security of the Internet or corporate networks as long as benchmarking is not performed on devices or systems connected to operating networks.

10. References

10.1. Normative References

[RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, March 1997.

[RFC5101] Claise B., "Specification of the IP Flow Information Export (IPFIX) Protocol for the Exchange of IP Traffic Flow Information", Standards Track, RFC 5101, January 2008.

10.2. Informative References

[RFC1242] Bradner, S., "Benchmarking Terminology for Network Interconnection Devices", RFC 1242, July 1991.

[RFC2285] Mandeville R., "Benchmarking Terminology for LAN Switching Devices", Informational, RFC 2285, May 21998.

[RFC2544] Bradner, S., "Benchmarking Methodology for Network Interconnect Devices", Informational, RFC 2544, March 1999

[RFC3917] Quittek j., "Requirements for IP Flow Information Export (IPFIX)", Informational, RFC 3917, October 2004.

[IPFIX-ARCH] Sadasivan, G., Brownlee, N., Claise, B., and J. Quittek,
"Architecture Model for IP Flow Information Export", Work in Progress, September 2006.

[CAIDA] Claffy, K., "The nature of the beast: recent traffic measurements
from an Internet backbone",
http://www.caida.org/publications/papers/1998/Inet98/Inet98.html

Author's Addresses

Jan Novak (editor)
Cisco System
Edinburgh,
UK
Phone: +44 7740 925889
Email: janovak@cisco.com

Full Copyright Statement

Intellectual Property

at
http://www.ietf.org/ipr.

The IETF invites any interested party to bring to its attention any copyrights, patents or patent applications, or other proprietary rights that may cover technology that may be required to implement this standard. Please address the information to the IETF at ietf-ipr@ietf.org.