Internet Draft
Document: draft-quittek-ipfix-middlebox-00.txt
Expires: August 2004

J. Quittek
M. Stiemerling
NEC Europe Ltd.

January 2004

Guidelines for IPFIX Implementations on Middleboxes

<draft-quittek-ipfix-middlebox-00.txt>

Status of this Memo

Copyright Notice

Abstract

This memo gives recommendations for the implementation of IP Flow Information eXport (IPFIX) metering processes and IPFIX exporting processes on middleboxes, such as firewalls, network address translators, tunnel endpoints, packet classifiers, etc. Middlebox functions potentially change properties of traffic flows passing the box, for example NATs change addresses in header fields and firewalls change the numbers of packets and bytes belonging to a traffic flow. An IPFIX implementation on a middlebox should reflect this by the way it selects and reports the observation point and by the way it measures and reports traffic flows.

Table of Contents

1. Introduction

The IP Flow Information eXport (IPFIX) protocol is defined in [IPFIX-PR] and [IPFIX-IM]. The protocol describes how information about traffic flows observed at an observation point can be exported via the Internet. The protocol can transfer information about traffic flow properties as well as information about where and how a certain traffic flow was measured.

The IPFIX architecture description gives insight in how to measure traffic flows at hosts, routers and passive probes, but at middleboxes more complicated situations may occur. Middleboxes change properties of IP traffic flows: NATs change addresses in header fields, firewalls change the numbers of packets and bytes belonging to a traffic flow, traffic shapers drop or delay packets, etc.

2. Terminology

The terminology used in this document is fully aligned with the terminology defined in [IPFIX-PR].

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

3. Middleboxes

The term middlebox is defined in RFC 3234 [RFC3234] by:

> "A middlebox is defined as any intermediary device performing functions other than the normal, standard functions of an IP router on the datagram path between a source host and destination host."

The list of middleboxes discussed in RFC 3234 contains:

1. NAT,
2. NAT-PT,
3. SOCKS gateway,
4. IP tunnel endpoints,
5. packet classifiers, markers, schedulers,
6. transport relay,
7. TCP performance enhancing proxies,
8. load balancers that divert/munge packets,
9. IP firewalls,

10. application firewalls,

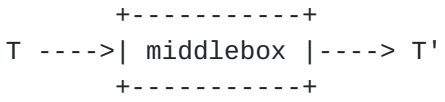
11. application-level gateways,
12. gatekeepers / session control boxes,
13. transcoders,
14. proxies,
15. caches,
16. modified DNS servers,
17. content and applications distribution boxes,
18. load balancers that divert/munge URLs,
19. application-level interceptors,
20. application-level multicast,
21. involuntary packet redirection,
22. anonymizers.

It is likely that since the publication of RFC 3234 new kinds of middleboxes have been added.

4. Traffic Flow Scenarios at Middleboxes

Middleboxes may delay, re-order, drop, or multiply packets, they may change packet header fields and change the payload. All these action have an impact on traffic flow properties.

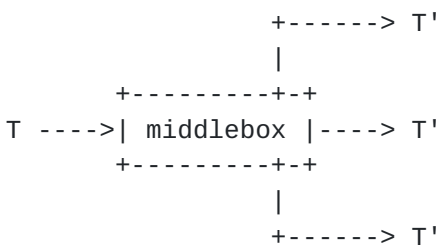
In general, a middlebox transforms a uni-directional original traffic flow T that arrives at the middlebox into a transformed traffic flow T' that leaves the middlebox.



Uni-directional traffic flow traversing a middlebox

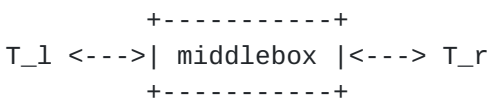
Note that in an extreme case, T' may be an empty traffic flow (a flow with no packets), for example, if the middlebox is a firewall and blocks the flow.

In case of a middlebox performing a multicast function, a single original traffic flow may be transformed into a more than one transformed traffic flow.



Uni-directional traffic flow traversing a middlebox

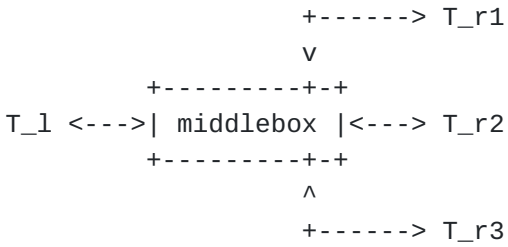
For bi-directional traffic flows we can not identify original and transformed traffic flow, we can just identify flows on different sides of the middlebox, say T_l on the left side and T_r on the right side.



Bi-directional unicast traffic flow traversing a middlebox

In case of a NAT T_l might be a traffic flow in a private address realm and T_r the translated traffic flow in the public address realm. If the middlebox is a NAT-PT, then T_l may be an IPv4 traffic flow and T_r the translated IPv6 traffic flow.

At tunnel endpoints, flows are multiplexed or de-multiplexed. In general, tunnel endpoints can deal with bi-directional traffic flows.



Bi-directional traffic flow traversing a tunnel endpoint

An example is a traffic flow T_l of a tunnel and flows T_rx that are multipled into or de-multiplexed out of a tunnel.

According to the IPFIX definiton of traffic flows in [IPFIX-PR] T and T' or T_l and T_ri, respectively, are different flows in general. However, from an application point of view, they might be considered as closely related or even as the same flow, for example if the payloads they carry are identical.

5. Location of the Observation Point

Middleboxes might be integrated with other devices. An example is a router with a NAT or a firewall at a line card. If an IPFIX observation point is located at the line card, then the measured properties of measured traffic flows may depend on the side of the integrated middlebox at which packets were captured for traffic flow measurement.

Consequently, an exporting process reporting traffic flows measured

at a device that hosts one or more middleboxes MUST clearly indicate

to collecting processes the location of the used observation point(s) with respect to the middlebox(es). Otherwise, processing the measured flow data could lead to wrong results.

At the first glance, choosing an observation point that covers the entire middlebox looks like an attractive choice for the location of the observation point. But this leads to ambiguities for all kinds of middleboxes. Within the middlebox properites of packets are modified and it MUST be clear at a collecting process whether packets were observed and measured before or after modification, for example it must be clear whether a reported source IP address was observed before or after a NAT changed it or whether a reported packet count was measured before or after a firewall dropped packets.

Only in case of composed middleboxes with well defined and well separated internal middlebox functions, for example a combined NAT and firewall, an observation point MAY be inside a middlebox, but in any case it MUST be located in between the middlebox functions.

6. Reporting Flow-related Middlebox Internals

While this document requests IPFIX implementations using observations points outside of middlebox functions, there are cases, where reporting flow-related internals of a middlebox is of interest.

For many application that use traffic measurement results it is desirable to get more information than can be derived from just observing packets on one side of a middlebox. If, for example, packets are dropped by the middlebox acting as firewall, NAT or traffic shaper, then information about how many packets of the observed packets are dropped may be of high interest.

This section gives recommendations on middlebox internal information that SHOULD or MAY be reported if the IPFIX observation point is co-located with one or more middleboxes. Since the internal information to be reported depends on the kind of middlebox, it is discussed per kind.

The recommendations cover middleboxes that act per packet and that do not modify the application level payload of the packet (except by dropping the entire packet) and that do not insert additional packets into an application level or transport level traffic stream.

Covered are the packet level middleboxes of kind 1 - 6, 8 - 10, 21, and 22 (according to the enumeration given in Sestion 3). Not covered are 7 and 11 - 20. TCP performance enhancing proxies (7) are not covered because they may add ACK packets to a TCP connection. Still, if possible, IPFIX implementation co-located with not covered

middleboxes MAY follow the recommendations given in this section if

they can be applied in a way that reflects the intention of the recommendations.

6.1. Packet Dropping Middleboxes

If an IPFIX observation point is co-located with one or more middleboxes that potentially drop packets, then the corresponding IPFIX exporter SHOULD be able to report the number of packets that were dropped per reported flow.

Concerned kinds of middleboxes are NAT (1), NAT-PT (2), SOCKS gateway (3), packet schedulers (5), IP firewalls (9) and application level firewalls (10).

6.2. Middleboxes Changing the DSCP

If an IPFIX observation point is co-located with one or more middleboxes that potentially modify the DiffServ Code Point (DSCP, see [RFC2474]) in the IP header, then the corresponding IPFIX exporter SHOULD be able to report besides the observed value of the DSCP also the value of the DSCP on the 'other' side of the middlebox if this is a constant value for the particular traffic flow.

Note that the 'other' side of the middlebox can be before or after changing the DSCP value depending on the location of the middlebox.

Note also that a classifier may change the same DSCP value of packets from the same flow to different values depending on the packet or other conditions. Also it is possible that packets of a single uni-directional arriving flow contain packets with different DSCP values that are all set to the same value by the middlebox. In both cases there is a constant value for the DSCP field in the IP packets header to be observed on one side of the middlebox, but on the other side the value may vary. In such a case reliable reporting of the DSCP value on the 'other' side of the middlebox is not possible by just reporting a single value.

This recommendation concerns packet markers (5).

6.3. Middleboxes Changing IP Addresses and Port Numbers

If an IPFIX observation point is co-located with one or more middleboxes that potentially modify the

- IP version field,
- IP source address header field,
- IP destination header field,
- TCP source port number,
- TCP destination port number,

- UDP source port number and/or

- UDP destination port number

in one of the headers, then the corresponding IPFIX exporter SHOULD be able to report besides the observed value of the particular header fields also the 'translated' value of these fields, as far as they have constant values for the particular traffic flow.

Note that the 'translated' values of the fields can be the fields values before or after the translation depending on the flow direction and the location of the observation point with respect to the middlebox. We alway call the value that is not the one observed at the observation point the translated value.

This paragraph needs to be adapted from DSCP to addresses and port numbers: Note also that a classifier may change the same DSCP value of packets from the same flow to different values depending on the packet or other conditions. Also it is possible that packets of a single uni-directional arriving flow contain packets with different DSCP values that are all set to the same value by the middlebox. In both cases there is a constant value for the DSCP field in the IP packets header to be observed on one side of the middlebox, but on the other side the value may vary. In such a case reliable reporting of the DSCP value on the 'other' side of the middlebox is not possible by just reporting a single value.

Concerned kinds of middleboxes are NAT (1), NAT-PT (2), SOCKS gateway (3) and involuntary packet redirection (21).

This recommendation MAY also be applied to anonymizers (21), but it should be noted that this includes the risk of loosing the effect of anonymisation.

6.4. Tunnel Endpoints

If an IPFIX observation point is co-located with one or more tunnel endpoints such that it observes packets that will be multiplexed into a tunnel or that have been de-multiplexed out of a tunnel, then the corresponding IPFIX exporter SHOULD be able to report the corresponding tunnel ID.

7. Security Considerations

Section 6 recommends that IPFIX exporting processes report internals about middleboxes. These internals may be security-relevant and the reported information needs to be protected appropriately for reasons given below.

Reporting the packets dropped by firewalls and other packet dropping

middleboxes imply the risk that this information is used by attackers

for analyzing the configuration of the packet dropper and for developing attacks that pass the middlebox.

Address translation may be used for hiding the network structure behind an address translator. If an IPFIX exporting process reports the translations performed by an address tranlator, then parts of the network structure may get uncovered.

If an IPFIX exporting process reports the translations performed by an anonymizer, the main function of the anonymizer might get lost.

Also information about which packet enters of leaves which tunnel may need protection.

8. Acknowledgements

Many thanks to Reinaldo Penno who raised the issue of IPFIX observation points co-located with middleboxes by a contribution to an earlier version of the IPFIX applicability statements.

9. Open Issues

- Do NATs (1-3) change DSCP?
- Are reports on DSCP modifications relevant for security?

10. Normative References

[IPFIX-PR] Claise, B., Fullmer, M., Calato, P. and R. Penno, "IPFIX Protocol Specifications", work in progress, <draft-ietf-ipfix-protocol-02.txt>, January 2003.

[IPFIX-IM] Calato, P., Meyer, J. and J. Quittek, "Information Model for IP Flow Information Export", work in progress, <draft-ietf-ipfix-info-02>, November 2003.

[RFC2474] Nichols, K., Blake, S., Baker, F. and D. Black, "Definition of the Differentiated Services Field (DS Field) in the IPv4 and IPv6 Headers", RFC 2474, December 1998.

11. Informative References

[RFC3415] Carpenter, B., and S. Brim, "Middleboxes: Taxonomy and Issues", RFC 3234, February 2002.

12. Authors' Addresses

Juergen Quittek
NEC Europe Ltd.
Network Laboratories
Kurfuersten-Anlage 36
69115 Heidelberg
Germany

Phone: +49 6221 90511-15
EMail: quittek@netlab.nec.de

Martin Stiemerling
NEC Europe Ltd.
Network Laboratories
Kurfuersten-Anlage 36
69115 Heidelberg
Germany

Phone: +49 6221 90511-13
Email: stiemerling@netlab.nec.de

13. IPR Notices

14. Full Copyright Statement