

IPFIX Working Group
Internet-Draft
Intended Status: Informational
Expires: August 25, 2008

B. Claise
P. Aitken
A. Johnson
Cisco Systems, Inc.
G. Muenz
University of Tuebingen
February 25, 2008

IPFIX Export per SCTP Stream
draft-claise-ipfix-export-per-sctp-stream-03

Status of this Memo

By submitting this Internet-Draft, each author represents that any applicable patent or other IPR claims of which he or she is aware have been or will be disclosed, and any of which he or she becomes aware will be disclosed, in accordance with [Section 6 of BCP 79](#).

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF), its areas, and its working groups. Note that other groups may also distribute working documents as Internet-Drafts.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

The list of current Internet-Drafts can be accessed at <http://www.ietf.org/ietf/1id-abstracts.txt>.

The list of Internet-Draft Shadow Directories can be accessed at <http://www.ietf.org/shadow.html>.

This Internet-Draft will expire on June, 2008.

Copyright Notice

Copyright (C) The IETF Trust (2008).

Abstract

This document specifies an improvement to the use of SCTP as specified in the IPFIX specifications in order to be able to deduce the data record loss per template record in case of partially-reliable SCTP export. This specification offers several extra advantages: immediate export of the template withdrawal message, immediate reuse of template ID within a stream, and the collecting process's job is easier.

Conventions used in this document

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](#)].

Table of Contents

1.	Terminology.....	4
1.1.	IPFIX Documents Overview.....	4
1.2.	PSAMP Documents Overview.....	4
2.	Introduction.....	5
2.1.	Applicability.....	6
3.	IPFIX Protocol Specifications Limitations and Improvements	6
3.1.	Data Record Loss per Template.....	6
3.1.1.	IPFIX Protocol Specifications Limitation.....	6
3.1.2.	IPFIX Export per SCTP Stream Advantage.....	7
3.2.	Transmission Order within a Stream.....	7
3.2.1.	IPFIX Protocol Specifications Limitation.....	7
3.2.2.	IPFIX Export per SCTP Stream Advantages.....	8
3.3.	No Transmission Order across SCTP Streams.....	8
3.3.1.	IPFIX Protocol Specifications Limitation.....	8
3.3.2.	IPFIX Export per SCTP Stream Advantages.....	9
4.	Specifications.....	9
4.1.	Template Management.....	9
4.2.	Template Withdrawal Message.....	10
4.3.	SCTP.....	11
4.4.	The Collecting Process's Side.....	12
5.	Examples.....	12
6.	IANA Considerations.....	15
7.	Security Considerations.....	15
8.	References.....	15

<Claise, et. Al>

Expires August 25, 2008

[Page 2]

8.1.	Normative References.....	15
8.2.	Informative References.....	16
9.	Acknowledgements.....	17
10.	Author's Addresses.....	18
11.	Intellectual Property Statement.....	19
12.	Copyright Statement.....	19
13.	Disclaimer.....	19

1. Terminology

IPFIX-specific terminology used in this document is defined in [section 2 of \[RFC5101\]](#). As in [\[RFC5101\]](#), these IPFIX-specific terms have the first letter of a word capitalized when used in this document.

Template Reuse Delay

The configurable timeout to allow the Collecting Process to receive and process the last Data Record using this Template information before which the Template Withdrawal Message MUST NOT be sent. [\[RFC5101\]](#) specifies a default value of 5 seconds.

1.1. IPFIX Documents Overview

The IPFIX Protocol [\[RFC5101\]](#) provides network administrators with access to IP flow information.

The architecture for the export of measured IP flow information out of an IPFIX exporting process to a collecting process is defined in the IPFIX Architecture [\[IPFIX-ARCH\]](#), per the requirements defined in [RFC 3917](#) [\[RFC3917\]](#).

The IPFIX Architecture [\[IPFIX-ARCH\]](#) specifies how IPFIX Data Records and Templates are carried via a congestion-aware transport protocol from IPFIX Exporting Processes to IPFIX Collecting Processes.

IPFIX has a formal description of IPFIX Information Elements, their name, type and additional semantic information, as specified in the IPFIX Information Model [\[RFC5102\]](#).

Finally the IPFIX Applicability Statement [\[IPFIX-AS\]](#) describes what type of applications can use the IPFIX protocol and how they can use the information provided. It furthermore shows how the IPFIX framework relates to other architectures and frameworks.

1.2. PSAMP Documents Overview

The document "A Framework for Packet Selection and Reporting" [\[PSAMP-FMWK\]](#), describes the PSAMP framework for network elements to select subsets of packets by statistical and

<Claise, et. Al>

Expires August 25, 2008

[Page 4]

other methods, and to export a stream of reports on the selected packets to a collector.

The set of packet selection techniques (sampling, filtering, and hashing) supported by PSAMP are described in "Sampling and Filtering Techniques for IP Packet Selection" [PSAMP-TECH].

The PSAMP protocol [[PSAMP-PROTO](#)] specifies the export of packet information from a PSAMP Exporting Process to a PSAMP Collecting Process. Like IPFIX, PSAMP has a formal description of its information elements, their name, type and additional semantic information. The PSAMP information model is defined in [[PSAMP-INFO](#)].

Finally [[PSAMP-MIB](#)] describes the PSAMP Management Information Base.

2. Introduction

The IPFIX working group has specified a protocol to export IP Flow information [[RFC5101](#)]. This protocol is designed to export information about IP traffic flows and related measurement data, where a flow is defined by a set of key attributes (e.g. source and destination IP address, source and destination port, etc.). However, thanks to its template mechanism, the IPFIX protocol can export any type of information, as long as the relevant Information Element is specified in the IPFIX Information Model [[RFC5102](#)], registered with IANA, or specified as an enterprise-specific Information Element.

The IPFIX protocol [[RFC5101](#)] specifies that IP traffic measurements for flows are exported using a TLV (type, length, value) format. The information is exported using a Template Record that is sent once to export the {type, length} pairs that define the data format for one or more Data Records that are sent for a flow. The Data Records specify values for each flow.

The IPFIX protocol [[RFC5101](#)] is flexible: it foresees the usage of the multiple SCTP streams per association; allows the transmission of Data Sets, Template Sets, and/or Options Template Sets on any stream; it offers the full or partial reliability export of Data Sets; it proposes the ordered or out-of-order delivery of Data Sets. However, due the resource

constraints on the Exporter, it is not always possible to export all Data Sets in a reliable way.

The specification in this document offers several advantages such as: the data records loss per template record in case of partially-reliable SCTP export, immediate export of the Template Withdrawal Message, immediate reuse of template ID within a stream, reduce the likelihood of losing data record, and the collecting process's job is easier.

2.1. Applicability

The specification in this document applies to the IPFIX protocol specifications [[RFC5101](#)]. However, it only applies to the SCTP transport [[RFC4960](#)] and [[RFC3758](#)] protocol option of the IPFIX protocol specifications. All specifications from [[RFC5101](#)] apply unless specified otherwise in this document.

As the Packet Sampling (PSAMP) protocol specifications [[PSAMP-PROTO](#)] are based on the IPFIX protocol specifications, the specifications in this document are also valid for the PSAMP protocol. Therefore, the advantages specified by this document also apply to PSAMP.

3. IPFIX Protocol Specifications Limitations and Improvements

3.1. Data Record Loss per Template

3.1.1. IPFIX Protocol Specifications Limitation

[Section 6.3.2](#) of the Requirements for IP Flow Information Export [[RFC3917](#)] discusses the data transfer reliability issues. "Loss of flow records during the data transfer from the exporting process to the collecting process must be indicated at the collecting process." is clearly mentioned. However, in some cases, it may be important to know how many Data Records of a certain type were lost (e.g., in the case of billing), but conventionally IPFIX does not provide this information.

A Collector can detect out-of-sequence, dropped, or duplicate IPFIX Messages by tracking the Sequence Number [[RFC5101](#)]. However, the Sequence Number field in the Export header

<Claise, et. Al>

Expires August 25, 2008

[Page 6]

increases with the number of IPFIX Data Records with the PR-SCTP stream.

The IPFIX protocol specification [[RFC5101](#)] specifies that Data Records associated with any Template ID may be sent on any SCTP stream. As such, if there is more than one Template IDs defined within the whole SCTP association then there is no way of knowing the which Template ID any lost Data Records are associated with. This is true, now matter what convention the Exporting Process uses to send Data Records on different SCTP streams, as the protocol makes no guarantees.

Using the specification in this document, it is guaranteed that any lost Data Records will be associated only with the Templates that are defined on that stream and by defining only one Template on a stream it is ensured that any loss is associated with that single Template.

3.1.2. IPFIX Export per SCTP Stream Advantage

By exporting each Template ID and the corresponding Data Records within a different stream, the loss pertaining to each specific Template ID can be deduced from Sequence Number field in the Export headers.

3.2. Transmission Order within a Stream

3.2.1. IPFIX Protocol Specifications Limitation

A Collecting Process must have received the Template Record associated with the Data Records to be able to decode the information in the Data Records. The IPFIX protocol specification foresees:

"The Exporting Process SHOULD transmit the Template Set and Options Template Set in advance of any Data Sets that use that (Options) Template ID, to help ensure that the Collector has the Template Record before receiving the first Data Record.",

The fact that the Collecting Process cannot decode the Data Records without the Template Record may result in the Data Records being discarded by the Collector, as specified in [[RFC5101](#)]:

"The Collecting Process normally receives Template Records from the Exporting Process before receiving Data Records. The Data Records are then decoded and stored by the Collector. If the Template Records have not been received at the time Data Records are received, the Collecting Process MAY store the Data Records for a short period of time and decode them after the Template Records are received."

In practice, Data Records without associated (Options) Template Records will probably be discarded by the Collecting Process.

3.2.2. IPFIX Export per SCTP Stream Advantages

By exporting each Template Record and the corresponding Data Records within a single stream and imposing in-order transmission, the Template will always arrive before the associated Data Records. Therefore, there is no risk that the Collecting Process discards Data Records while waiting for the Template to arrive.

Furthermore, when reusing a Template ID within a stream, the Template Withdrawal Message will be guaranteed to arrive before the new definition of the Template and therefore the Template Record may be sent directly after the Template Withdrawal Message. Therefore the Template Reuse Delay restriction can be removed for Template ID reuse with a stream.

Another advantage with the new specifications in this document is that the Collecting Process's job is now easier. Indeed, the Collecting Process doesn't have to store the Data Records while waiting for the Template Records, as the transmission order is always guaranteed. This way, extra reliability of the Data Records is achieved without extra burden on the Collecting Process.

3.3. No Transmission Order across SCTP Streams

3.3.1. IPFIX Protocol Specifications Limitation

The fact that the protocol specifications [[RFC5101](#)] are flexible in terms of stream(s) on which the Template Set, Options Template Set, and corresponding Data Sets are

exported, implies that the (Options) Template Set might be exported on a different stream than the corresponding Data Sets. This might cause Data Record loss in the Collecting Process as the ordered transmission across SCTP streams is not guaranteed.

For example, a Template may be blocked pending reliable transmission on one stream while the associated Data Records may be transmitted immediately in another stream. Also, due to different stream congestion, it is possible that even if the Template and Data Records are both sent reliably, Data Records sent on a different stream than the associated Template might still arrive before the associated Template.

3.3.2. IPFIX Export per SCTP Stream Advantages

By exporting each Template Record and the corresponding Data Records within a single stream, imposing in-order transmission, and limiting the Template ID to a single stream, the issue of ordered transmission across multiple streams is avoided.

By exporting all corresponding Data Records within the same ordered stream as the Template Definition then each stream is independent and self-contained and the interaction between streams is limited to that of Options Data interactions. This has several advantageous consequences, including order preservation does not result in the blocking of unrelated data and that the Collector's job is simplified as the Template Records are guaranteed to be delivered before the associated Data Records.

4. Specifications

4.1. Template Management

This section introduces modifications compared to the Template Management [section 8 in \[RFC5101\]](#).

As specified in [\[RFC5101\]](#), Template Sets and Options Template Sets MUST be sent reliably. In other words, any IPFIX Message containing an (Options) Template Set MUST be sent reliably.

Any Data Sets associated with a Template Record MUST be sent on the same stream on which the Template Record was sent.

By sending only a single Template ID and associated Data Sets within a single stream, any Data Record loss can be calculated on a per Template basis.

If there is some Options Data Records required to interpret the Data Records, the Options Data Records MUST be sent reliably within the same stream. If not sent reliably, the Collector could consider that any loss might be associated with the Options Template Record, even if the Exporter is always sending them reliably.

The Option Data Record SHOULD be sent before the Data Record that needs it so that it arrives first and is available for the Collector to use. By sending the Options Data reliably, any loss will be limited to the non-option Data Record and loss can still be calculated on a per Template basis.

Furthermore, the Exporter MAY group related Template IDs and their associated Data Sets within a single stream so loss statistics are calculated for the group. This may be suitable in cases where there is insufficient SCTP streams to send each Template on its own stream and/or the case where there are slight variations on a single Template to show that some fields were unavailable at the time of monitoring.

4.2. Template Withdrawal Message

This section introduces Template Withdrawal Message-related modifications compared to the Template Management [section 8 in \[RFC5101\]](#).

Templates that are not used anymore SHOULD be deleted. Before reusing a Template ID, the Template MUST be deleted. In order to delete an allocated Template, the Template is withdrawn through the use of a Template Withdrawal Message. The Template Withdrawal Message MUST be sent on the same stream as the Template ID to be removed.

As all IPFIX Messages are sent in order within a stream and reliably per [\[RFC5101\]](#), the IPFIX Message containing the Template Withdrawal Message will not arrive at the Collecting Process before any associated and previously sent Data Record.

As a consequence, no Data Records will be lost due to delayed arrival at the Collector.

The Template ID from a withdrawn Template MAY be reused on the same stream immediately after the Template Withdrawal Message is sent. This case is equivalent to the use of a Template Reuse Delay value of 0.

If the new definition of the Template ID is to be reused on a different stream, the Template Withdrawal Message MUST NOT be sent before the Template Reuse Delay.

A Template Withdrawal Message to withdraw all Templates for the Observation Domain ID specified in the IPFIX Message header MUST NOT be used.

Multiple Template IDs MAY be withdrawn with a single Template Withdrawal Message at the condition that all the Template IDs in the Template Withdrawal Message are used on the same SCTP stream.

4.3. SCTP

This section introduces modifications compared to the "SCTP" [section 10.2](#) (and subsections) in [[RFC5101](#)]. More specifically the "Stream" [section 10.2.4.3](#)

PR-SCTP [[RFC3758](#)] MUST be implemented by all compliant implementations.

All IPFIX Messages MUST be sent in order within a stream.

Depending on the application requirement, the Exporting Process MAY send Data Sets with full or partial reliability. Unreliable data transfer MAY be used where the application does not require reliable transmission and the use of a retransmission queue is impractical.

If the Exporting Process requires to export a new Template but there are no more free SCTP streams available, it SHOULD attempt to increase the number of outbound streams it is able to send to, per [[SCTP-RESET](#)]. Alternatively, the Exporting Process MAY add the Template Set and Data Records to an existing stream at the cost of diluting the granularity of Data Records loss.

4.4. The Collecting Process's Side

This refers to the Collecting Process's Side [section 9 in \[RFC5101\]](#).

The Collecting Process SHOULD listen for a new association request from the Exporting Process. The Exporting Process will request a number of streams to use for export: the number of streams SHOULD be equivalent to the number of simultaneous Template IDs used in the association.

A Collecting process SHOULD support the procedure for the addition of a SCTP stream [[SCTP-RESET](#)].

The IPFIX protocol has a Sequence Number field in the Message header that increases with the number of IPFIX Data Records in the IPFIX Message. A Collector may detect out-of-sequence, dropped, or duplicate IPFIX Messages by tracking the Sequence Number. As this Sequence Number is per SCTP stream, the loss for the Data Records sent in that stream can be calculated in case of partially-reliable export. If the SCTP stream only contains Data Records from a single Template ID, the Data Records for that Template ID can be calculated.

If the Collecting Process receives a Template Withdrawal Message on a different stream than the one on which the Template ID is used, then the Collecting Process MUST reset the association and SHOULD log an error message.

The following sentence from [[RFC5101](#)] is not applicable in this specification:

"The Collecting Process normally receives Template Records from the Exporting Process before receiving Data Records. The Data Records are then decoded and stored by the Collector. If the Template Records have not been received at the time Data Records are received, the Collecting Process MAY store the Data Records for a short period of time and decode them after the Template Records are received."

5. Examples

Figure 1 shows an example where the stream 10 carries a Template with the Template ID 256 transmitted with full reliability (FR), together with associated Data Records transmitted with partial reliability (PR). Note that, because all IPFIX Messages are

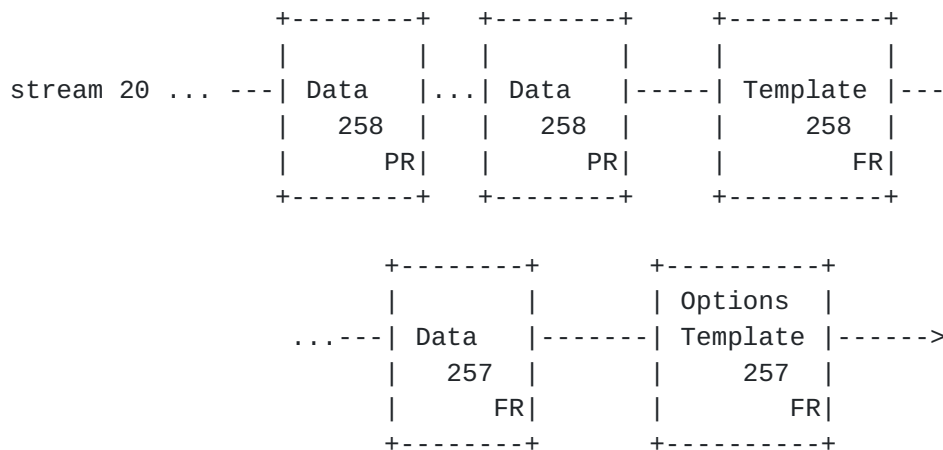


Figure 2

Figure 2 shows an example where stream 20 carries an Options Template with Template ID 257 transmitted with full reliability (FR), an associated Data Record transmitted with full reliability (FR), a Template with Template ID 258 transmitted with full reliability (FR), and associated Data Records transmitted with partial reliability (PR). In this example the Option Template Record contains information required to decode the latter Data Records, such as Common Properties information [[IPFIX-RED-RED](#)]. So it makes sense to export the Data Sets 257 reliably. If some Data Record loss is observed from the Sequence Number, the loss can only stem from the Data Sets with the Template ID 258, as these are the only Sets not exported reliably. Therefore, the calculation of loss per Template ID 258 is possible.

Note that, because all IPFIX Message must be sent in order within a stream, the Options Template 257 will always arrive before its associated Data Records, and that the Template 259 will always arrive before the its associated Data Records.

Figure 3 show an example where stream 30 carries an Template with Template ID 259 transmitted with full reliability (FR), an associated Data Record transmitted with partial reliability (PR), a Template Withdrawal Message, followed by a redefinition of the Template ID 259, and finally the new definition of Data Record transmitted with partial reliability. The Template Withdrawal Message and the new definition of the Template ID 259 are sent immediately, without waiting for the Template Reuse Delay.

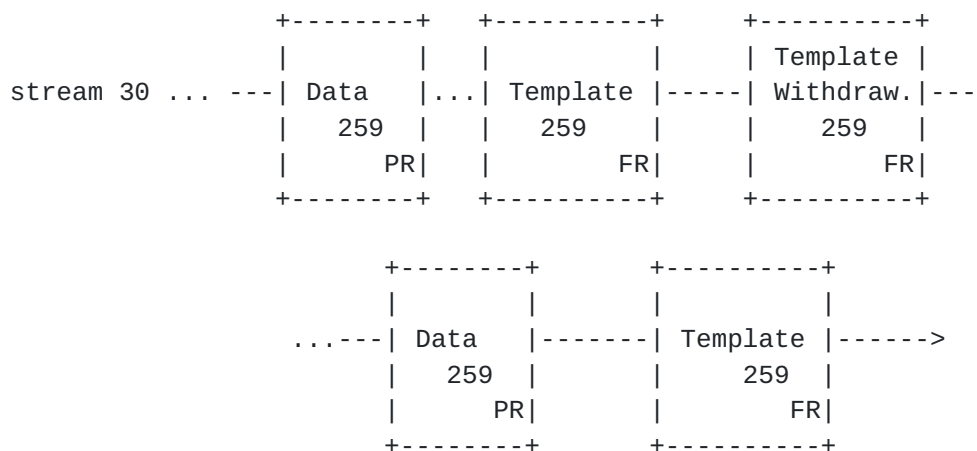


Figure 3

6. IANA Considerations

This document has no actions for IANA.

7. Security Considerations

The same security considerations as for the IPFIX Protocol [[RFC5101](#)] apply.

8. References

8.1. Normative References

- [RFC2119] S. Bradner, Key words for use in RFCs to Indicate Requirement Levels, [BCP 14](#), [RFC 2119](#), March 1997
- [RFC3758] Stewart, R., Ramalho, M, Xie, Q., Tuexen, M., Conrad, P., "Stream Control Transmission Protocol (SCTP), Partial Reliability Extension", May 2004
- [RFC4960] Stewart, R., Ed., "Stream Control Transmission Protocol", [RFC 4960](#), September 2007.
- [RFC5101] Claise, B., Ed., "Specification of the IP Flow Information Export (IPFIX) Protocol for the Exchange of IP Traffic Flow Information", [RFC 5101](#), January 2008.

[PSAMP-TECH] T. Zseby, M. Molina, N. Duffield, S. Niccolini, F. Raspall, "Sampling and Filtering Techniques for IP Packet Selection" [draft-ietf-psamp-sample-tech-10.txt](#), Internet-Draft work in progress, June 2007

[SCTP-RESET] Stewart, R., Lei, P., Tuexen, M, "Stream Control Transmission Protocol (SCTP) Stream Reset", [draft-stewart-sctpstrrst-05.txt](#), Internet-Draft work in progress, June 2007

8.2. Informative References

[RFC3917] Quittek, J., Zseby, T., Claise, B. Zander, S, Requirements for IP Flow Information Export, [RFC 3917](#), October 2004

[IPFIX-ARCH] Sadasivan, G., Brownlee, N., Claise, B., Quittek, J., "Architecture Model for IP Flow Information Export" [draft-ietf-ipfix-architecture-12](#), Internet-Draft work in progress, September 2006

[IPFIX-AS] Zseby, T., Boschi, E., Brownlee, N., Claise, B., "IPFIX Applicability", [draft-ietf-ipfix-as-12.txt](#), Internet-Draft work in progress, February 2007

[RFC5102] Quittek, J., Bryant, S., Claise, B., Aitken, P., and J. Meyer, "Information Model for IP Flow Information Export", [RFC 5102](#), January 2008.

[PSAMP-INFO] T. Dietz, F. Dressler, G. Carle, B. Claise, "Information Model for Packet Sampling Exports", [draft-ietf-psamp-info-07.txt](#), Internet-Draft work in progress, October 2007

[PSAMP-PROTO] Claise, B., Quittek, J., and A. Johnson, "Packet Sampling (PSAMP) Protocol Specifications", [draft-ietf-psamp-protocol-09](#), Internet-Draft work in progress, December 2007.

[PSAMP-FMWK] D. Chiou, B. Claise, N. Duffield, A. Greenberg, M. Grossglauser, P. Marimuthu, J. Rexford, G. Sadasivan, "A Framework for Passive Packet Measurement" [draft-ietf-psamp-framework-12.txt](#), Internet-Draft work in progress, June 2007

[IPFIX-RED-RED] Boschi, E., Mark, L., Claise, B. "Reducing Redundancy in IPFIX and PSAMP Reports", Internet-Draft work in progress, [draft-ietf-ipfix-reducing-redundancy-04.txt](#), May 2007

[PSAMP-MIB] Dietz, T., Claise, B. "Definitions of Managed Objects for Packet Sampling", Internet-Draft work in progress, June 2006

9. Acknowledgements

The authors would like to thank Brian Trammell for expert feedback, Randall Stewart and Peter Lei for their SCTP-related feedback.

10. Author's Addresses

Benoit Claise
Cisco Systems Inc.
De Kleetlaan 6a b1
Diegem 1813
Belgium

Phone: +32 2 704 5622
Email: bclaise@cisco.com

Paul Aitken
Cisco Systems (Scotland) Ltd.
96 Commercial Quay
Commercial Street
Edinburgh, EH6 6LX, United Kingdom

Phone: +44 131 561 3616
Email: paitken@cisco.com

Andrew Johnson
Cisco Systems (Scotland) Ltd.
96 Commercial Quay
Commercial Street
Edinburgh, EH6 6LX, United Kingdom

Phone: +44 131 561 3641
Email: andrjohn@cisco.com

Gerhard Muenz
University of Tuebingen
Computer Networks and Internet
Sand 13
Tuebingen D-72076
DE

Phone: +49 7071 29-70534
Email: muenz@informatik.uni-tuebingen.de
URI: <http://net.informatik.uni-tuebingen.de/~muenz>

11. Intellectual Property Statement

The IETF takes no position regarding the validity or scope of any Intellectual Property Rights or other rights that might be claimed to pertain to the implementation or use of the technology described in this document or the extent to which any license under such rights might or might not be available; nor does it represent that it has made any independent effort to identify any such rights. Information on the procedures with respect to rights in RFC documents can be found in [BCP 78](#) and [BCP 79](#).

Copies of IPR disclosures made to the IETF Secretariat and any assurances of licenses to be made available, or the result of an attempt made to obtain a general license or permission for the use of such proprietary rights by implementers or users of this specification can be obtained from the IETF on-line IPR repository 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.

12. Copyright Statement

Copyright (C) The IETF Trust (2008).

This document is subject to the rights, licenses and restrictions contained in [BCP 78](#), and except as set forth therein, the authors retain all their rights.

13. Disclaimer

This document and the information contained herein are provided on an "AS IS" basis and THE CONTRIBUTOR, THE ORGANIZATION HE/SHE REPRESENTS OR IS SPONSORED BY (IF ANY), THE INTERNET SOCIETY, THE IETF TRUST AND THE INTERNET ENGINEERING TASK FORCE DISCLAIM ALL WARRANTIES, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO ANY WARRANTY THAT THE USE OF THE INFORMATION HEREIN WILL NOT INFRINGE ANY RIGHTS OR ANY IMPLIED WARRANTIES OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE.

