Audio/Video Transport Working Group	Q. Wu, Ed.
Internet-Draft	Huawei
Intended status: Informational	G. Hunt
Expires: March 03, 2012	Unaffiliated
	P.J. Arden
	BT
	August 31, 2011

Monitoring Architectures for RTP
draft-ietf-avtcore-monarch-04.txt

Abstract

This memo proposes an architecture for extending RTCP with a new RTCP XR (RFC3611) block type to report new metrics regarding media transmission or reception quality, as proposed in RFC5968. This memo suggests that a new block should contain a single metric or a small number of metrics relevant to a single parameter of interest or concern, rather than containing a number of metrics which attempt to provide full coverage of all those parameters of concern to a specific application. Applications may then "mix and match" to create a set of blocks which covers their set of concerns. Where possible, a specific block should be designed to be re-usable across more than one application, for example, for all of voice, streaming audio and video.

Table of Contents

1. Introduction

As more users and subscribers rely on real time application services, uncertainties in the performance and availability of these services are driving the need to support new standard methods for gathering performance metrics from RTP applications. These rapidly emerging standards, such as RTCP XR [RFC3611]and other RTCP extension to Sender Reports(SR), Receiver Reports (RR) [RFC3550]are being developed for the purpose of collecting and reporting performance metrics from endpoint devices that can be used to correlate the metrics, provide end to end service visibility and measure and monitor QoE.
However the proliferation of RTP/RTCP specific metrics for transport and application quality monitoring has been identified as a potential problem for RTP/RTCP interoperability, which attempt to provide full coverage of all those parameters of concern to a specific application. Since different applications layered on RTP may have some monitoring requirements in common, therefore these metrics should be satisfied by a common design.
The objective of this document is to define an extensible RTP monitoring framework to provide a small number of re-usable QoS/QoE metrics which facilitate reduced implementation costs and help maximize inter-operability. [RFC5968] has stated that, where RTCP is to be extended with a new metric, the preferred mechanism is by the addition of a new RTCP XR [RFC3611] block. This memo assumes that any requirement for a new metric to be transported in RTCP will use a new RTCP XR block.

2. Requirements notation

This memo is informative and as such contains no normative requirements.
In addition, the following terms are defined:

Transport level metrics

A set of metrics which characterise the three transport impairments of packet loss, packet delay, and packet delay variation. These metrics should be usable by any application which uses RTP transport.

Application level metrics

Metrics relating to QoE related parameters. These metrics are measured at the application level and focus on quality of content rather than network parameters. One example of such metrics is the Multimedia Quality Metric specified in [MQ].

End System metrics

Metrics relating to the way a terminal deals with transport impairments affecting the incident RTP stream. These may include de-jitter buffering, packet loss concealment, and the use of redundant streams (if any) for correction of error or loss.

3. RTP monitoring architecture

The RTP monitoring architecture comprises the following two key functional components shown below:

*Monitor

*Metric Block Structure

Monitor is a functional component defined in RFC3550 that acts as a source of information gathered for monitoring purposes. It may also collect statistics from multiple source, stores such information reported by RTCP XR or other RTCP extension appropriately as base metric or calculates composite metric. According to the definition of monitor in RFC3550, the end system that source RTP streams, an intermediate-system that forwards RTP packets to End-devices or a third party that does not participate RTP session (i.e., the third party monitor depicted in figure 1) can be envisioned to act as the Monitor within the RTP monitoring architecture.
The Metric Block exposes real time Application Quality information in the appropriate report block format to the Monitor within the RTP monitoring architecture. Both the RTCP or RTCP XR can be extended to convey such information. The details on transport protocol for metric block is described in Section 3.1.

```
                                             |---------------+
                                             | Management    |
             +-------------------+           |   System      |
             | RTP Sender        |           |  +----------+ |
             |   +-----------+   |           |  |          | |
---------------->|  Monitor  |---------5------->|  Monitor | |
|            |   |           |   |           |  |          | |
|            |   +-----------+   |           |  +----\-----+ |
|            |+-----------------+|           |       |       |
|            ||Application      ||           --------|-------+
|            ||-Streaming video ||                   |
|   |---------|-VOIP            ||                   5
|   |        ||-Video conference||                   |
|   |        ||-Telepresence    ||       +---------------+
|   |        ||-Ad insertion    ||       |  Third Party  |
5   |        |+-----------------+|       |    Monitor    |
|   |        +-------------------+       +---------------+
|   1
|   | +Intermediate------------+         |------------- ---- ----+
|   | | RTP System       Report Block    | RTP Receiver >--4-|    |
|   | |      +---------- transported over|    +-----------+  |    |
|   | |      |           RTCP extension  |    |  Monitor  |<--    |
|-------------  Monitor |<--------5------|----|           |<------|
    | |      |          |   Report Block      +----/------+      ||
    | |      +----------+   transported over       |             ||
    | |                     RTCP XR      |         |2            ||
    | | +-----------------+    |         | +-------/---------+   ||
    | | |Application      |    |         | |Application      |   ||
    | | |-Streaming video |    |         | |-Streaming video |   ||
    | | |-VOIP            |    |    1    | |-VOIP            |   3|
    ---->-Video conference|--------------->|-Video conference    ||
      | |-Telepresence    |    |         | |-Telepresence    |   ||
      | |-Ad insertion    |    |         | |-Ad insertion    |   ||
      | +-----------------+    |         | +-----------------+   ||
      | +-----------------+    |         | +-----------------+   ||
      | |Transport        |    |         | |Transport        |   ||
      | |-IP/UDP/RTP      |    |         | |-IP/UDP/RTP      >---||
      | |-IP/TCP/RTP      |    |         | | -IP/TCP/RTP     |    |
      | |-IP/TCP/RTSP/RTP |    |         | |-IP/TCP/RTSP/RTP |    |
      | +-----------------+    |         | +-----------------+    |
      +------------------------+         +------------------------+
```

1. RTP communication between real time applications.

2. Application level metrics collection.

3. Transport level metrics collection.

4. End System metrics collection.

5. Reporting Session- metrics transmitted over specified interfaces.

3.1. RTCP Metric Block Report and associated parameters

The basic RTCP Reception Report (RR) conveys reception statistics in metric block report format for multiple RTP media streams including [RFC3611] supplement the existing RTCP packets and provide more detailed feedback on reception quality in several categories:

*transport level statistics

*the fraction of packet lost since the last report

*the cumulative number of packets lost

*the highest sequence number received

*an estimate of the inter-arrival jitter

*and information to allow senders to calculate the network round trip time.

The RTCP XRs

*Loss and duplicate RLE reports

*Packet-receipt times reports

*Round-trip time reports

*Statistics Summary Reports

There are also various other scenarios in which it is desirable to send RTCP Metric reports more frequently. The Audio/Video Profile with Feedback [RFC4585]extends the standard A/V Profile[RFC3551] to allow RTCP reports to be sent early provided RTCP bandwidth allocation is respected. There are four use cases but are not limited to:

*RTCP NACK is used to provide feedback on the RTP sequence number of the lost packets. [RFC4585]

*RTCP XR is extended to provide feedback on multicast acquisition statistics information and parameters.[RFC6332]

*RTCP is extended to convey requests for full intra-coded frames or select the reference picture, and signalchanges in the desired temporal/spatial trade-off and maximum media bit rate. [RFC5104]

* RTCP or RTCP XR is extended to provide feedback on ECN statistics information. [ECN]

4. Issues with reporting metric block using RTCP XR extension

Issues that have come up in the past with reporting metric block using RTCP XR extensions include (but are probably not limited to) the following:

* Using large block. A single report block or metric is designed to contain a large number of parameters in different classes for a specific application. For example, RFC 3611 [RFC3611] defines seven report block formats for network management and quality monitoring. However some of these block types defined in [RFC3611] are only specifically designed for conveying multicast inference of network characteristics(MINC) or voice over IP (VoIP) monitoring. However different applications layered on RTP may have some monitoring requirements in common, design large block only for specific applications may increase implementation cost and minimize interoperability.

* Correlating RTCP XR with the non-RTP data. CNAME [RFC3550] is an example of existing tool that allows to bind an SSRC that may change to a fixed source name in one RTP session. It is also fixed across multiple RTP sessions from the same source. However there may be situations where RTCP reports are sent to other participating endpoints using non-RTP protocol in a session. For example, as described in [RFC6035], the data contained in RTCP XR VoIP metrics reports [RFC3611] are forwarded to a central collection server systems using SIP. In such case, there is a large portfolio of quality parameters that can be associated with real time application,e.g., VOIP application, but only a minimal number of parameters are included on the RTCP-XR reports. Therefore correlation between RTCP XR and non-RTP data should be concerned if administration or management systems need to rely on the mapping RTCP statistics to non-RTCP measurements to conducts data analysis and creates alerts to the users. Without such correlation, it is hardly to provide accurate measures of real time application quality with a minimal number of parameters included on the RTCP-XR reports in such case.

* Identity Information duplication. Identity information is used to identify an instance of a metric block. The SSRC of the measured stream as part of the metric block is one example of Identity information. However in some cases, Identity information may be not part of metric and include information more than the SSRC in the metric block, e.g., when we set a metric interval for the session and monitor RTP packets within one or several consecutive metric interval, extra identity information (e.g., sequence

number of 1st packet) is expected, if we put such extra identity information into each metric block, there may be situations where an RTCP XR packet containing more than two metric blocks including the duplicated extra identity information, reports on the same streams from the same source. each block have the same extra identity information for measurement, if each metric block carry such duplicated data for the measurement, it leads to redundant information in this design since equivalent information is provided multiple times, once in *every* metric block. Though this ensures immunity to packet loss, the design may bring more complexity and the overhead is not completely trivial in some cases.

5. Guideline for reporting block format using RTCP XR

5.1. Using small blocks

Different applications using RTP for media transport certainly have differing requirements for metrics transported in RTCP to support their operation. For many applications, the basic metrics for transport impairments provided in RTCP SR and RR packets [RFC3550] (together with source identification provided in RTCP SDES packets) are sufficient. For other applications additional metrics may be required or at least sufficiently useful to justify the overheads, both of processing in endpoints and of increased session bandwidth. For example an IPTV application using Forward Error Correction (FEC) might use either a metric of post-repair loss or a metric giving detailed information about pre-repair loss bursts to optimise payload bandwidth and the strength of FEC required for changing network conditions. However there are many metrics available. It is likely that different applications or classes of applications will wish to use different metrics. Any one application is likely to require metrics for more than one parameter but if this is the case, different applications will almost certainly require different combinations of metrics. If larger blocks are defined containing multiple metrics to address the needs of each application, it becomes likely that many different such larger blocks are defined, which becomes a danger to interoperability.
To avoid this pitfall, this memo proposes the use of small RTCP XR metrics blocks each containing a very small number of individual metrics characterizing only one parameter of interest to an application running over RTP. For example, at the RTP transport layer, the parameter of interest might be packet delay variation, and specifically the metric "IPDV" defined by [Y1540]. See Section 6 for architectural considerations for a metrics block, using as an example a metrics block to report packet delay variation.

5.2. Correlating identity information with the non-RTP data

When more than one media transport protocols are used by one application to interconnected to the same session (in gateway),e.g., one RTCP XR Packet is sent to the participating endpoints using non-RTP-based media transport (e.g., using SIP) in a VOIP session, one crucial factor lies in how to handle their different identities that are corresponding to different media transport.
This memo proposes an approach to facilitate the correlation of the RTCP Session with other session-related non-RTP data, i.e., if there is a need to correlate RTP sessions with non-RTP sessions, then the correlation information needed should be conveyed in RTCP SDES packets since such correlation information describes the source, rather than providing a quality report. An example use case is for a participant endpoint may convey a call identifier or a global call identifier associated with the SSRC of measured RTP stream . In such case, the participant endpoint uses SSRC of source to bind the call identifier in each chunk of the SDES RTCP packet and send such correlation using the chunk containing SDES item to the network management system. A flow measurement tool that is not call-aware then forward the RTCP XR reports along with SSRC of the measured RTP stream which is included in the XR Block header to the network management system. Network management system can then correlate this report using SSRC with other diagnostic information such as call detail records.

6. An example of a metric block

This section uses the example of an existing proposed metrics block to illustrate the application of the principles set out in Section 5.1.
The example [PDV] (work in progress) is a block to convey information about packet delay variation (PDV) only, consistent with the principle that a metrics block should address only one parameter of interest. One simple metric of PDV is available in the RTCP RR packet as the "jit" field. There are other PDV metrics which may be more useful to certain applications. Two such metrics are the IPDV metric ([Y1540], [RFC3393]) and the MAPDV2 metric [G1020]. Use of these metrics is consistent with the principle in Section 5 of [RFC5968] that metrics should usually be defined elsewhere, so that RTCP standards define only the transport of the metric rather than its nature. The purpose of this section is to illustrate the architecture using the example of [PDV] (work in progress) rather than to document the design of the PDV metrics block or to provide a tutorial on PDV in general.
Given the availability of at least three metrics for PDV, there are design options for the allocation of metrics to RTCP XR blocks:

*provide an RTCP XR block per metric

*provide a single RTCP XR block which contains all three metrics *provide a single RTCP block to convey any one of the three metrics, together with a identifier to inform the receiving RTP system of the specific metric being conveyed

In choosing between these options, extensibility is important, because additional metrics of PDV may well be standardized and require inclusion in this framework. The first option is extensible but only by use of additional RTCP XR blocks, which may consume the limited namespace for RTCP XR blocks at an unacceptable rate. The second option is not extensible, so could be rejected on that basis, but in any case a single application is quite unlikely to require transport of more than one metric for PDV. Hence the third option was chosen. This implies the creation of a subsidiary namespace to enumerate the PDV metrics which may be transported by this block, as discussed further in [PDV] (work in progress).

7. Application to RFC 5117 topologies

The topologies specified in [RFC5117] fall into two categories. The first category relates to the RTP system model utilizing multicast and/or unicast. The topologies in this category are specifically Topo-Point-to-Point, Topo- Multicast, Topo-Translator (both variants, Topo-Trn-Translator and Topo-Media-Translator, and combinations of the two), and Topo-Mixer. These topologies use RTP end systems, RTP mixers and RTP translators defined in [RFC3550]. For purposes of reporting connection quality to other RTP systems, RTP mixers and RTP end systems are very similar. Mixers resynchronize audio packets and do not relay RTCP reports received from one cloud towards other cloud(s). Translators do not resynchronize packets and SHOULD forward certain RTCP reports between clouds. In this category, the RTP system (end system, mixer or translator) which originates, terminates or forwards RTCP XR blocks is expected to handle RTCP, including RTCP XR, according to [RFC3550]. Provided this expectation is met, an RTP system using RTCP XR is architecturally no different from an RTP system of the same class (end system, mixer, or translator) which does not use RTCP XR. The second category relates to deployed system models used in many H. 323 [H323] video conferences. The topologies in this category are Topo-Video-Switch-MCU and Topo-RTCP-terminating-MCU. Such topologies based on systems do not behave according to [RFC3550].
Considering the translator and MCU are two typical topologies in the two categories mentioned above, this document will take them as two typical examples to explain how RTCP XR report works in different RFC5117 topologies.

7.1. Applicability to MCU

Topo-Video-Switch-MCU and Topo-RTCP-terminating-MCU, suffer from the difficulties described in [RFC5117]. These difficulties apply to systems sending, and expecting to receive, RTCP XR blocks as much as to systems using other RTCP packet types. For example, a participant RTP end system may send media to a video switch MCU. If the media stream is not selected for forwarding by the switch, neither RTCP RR packets nor RTCP XR blocks referring to the end system's generated stream will be received at the RTP end system. Strictly the RTP end system can only conclude that its RTP has been lost in the network, though an RTP end system complying with the robustness principle of [RFC1122] should survive with essential functions unimpaired.

7.2. Applicability to Translators

Section 7.2 of [RFC3550] describes processing of RTCP by translators. RTCP XR is within the scope of the recommendations of [RFC3550]. Some RTCP XR metrics blocks may usefully be measured at, and reported by, translators. As described in [RFC3550] this creates a requirement for the translator to allocate an SSRC for the monitor collocated with itself so that the monitor may populate the SSRC in the RTCP XR packet header as packet sender SSRC and send it out(although the translator is not a Synchronisation Source in the sense of originating RTP media packets). It must also supply this SSRC and the corresponding CNAME in RTCP SDES packets.

In RTP sessions where one or more translators generate any RTCP traffic towards their next-neighbour RTP system, other translators in the session have a choice as to whether they forward a translator's RTCP packets. Forwarding may provide additional information to other RTP systems in the connection but increases RTCP bandwidth and may in some cases present a security risk. RTP translators may have forwarding behaviour based on local policy, which might differ between different interfaces of the same translator.

For bidirectional unicast, an RTP system may usually detect RTCP XR from a translator by noting that the sending SSRC is not present in any RTP media packet. However there is a possibility of a source sending RTCP XR before it has sent any RTP media (leading to transient mis-categorisation of an RTP end system or RTP mixer as a translator), and for multicast sessions - or unidirectional/streaming unicast - there is also a possibility of a receive-only end system being permanently mis-categorised as a translator sending XR report, i.e.,the monitor sending XR report within the translator. Hence it is desirable for a translator that sends XR report to have a way to declare itself explicitly.

8. IANA Considerations

None.

9. Security Considerations

This document itself contains no normative text and hence should not give rise to any new security considerations, to be confirmed.

10. Acknowledgement

The authors would also like to thank Colin Perkins, Graeme Gibbs, Debbie Greenstreet, Keith Drage, Dan Romascanu, Ali C. Begen, Roni Even for their valuable comments and suggestions on the early version of this document.

11. References

[RFC1122]	Braden, R., "Requirements for Internet Hosts -- Communication Layers", RFC 1122, October 1989.
[RFC3393]	Demichelis, C., "IP Packet Delay Variation Metric for IP Performance Metrics (IPPM)", RFC 3393, November 2002.
[RFC3550]	Schulzrinne, H., "RTP: A Transport Protocol for Real-Time Applications", RFC 3550, July 2003.
[RFC3611]	Friedman, T. (Ed), "RTP Control Protocol Extended Reports (RTCP XR)", RFC 3611, November 2003.
[RFC5117]	Westerlund, M., "RTP Topologies", RFC 5117, January 2008.
[RFC5968]	Ott, J. and C. Perkins, "Guidelines for Extending the RTP Control Protocol (RTCP)", RFC 5968, September 2010.
[RFC4585]	Ott, J. and S. Wenger, "Extended RTP Profile for Real-time Transport Control Protocol (RTCP)-Based Feedback (RTP/AVPF)", RFC 4585, July 2006.
[RFC3551]	Schulzrinne , H. and S. Casner, "Extended RTP Profile for Real-time Transport Control Protocol (RTCP)-Based Feedback (RTP/AVPF)", RFC 3551, July 2003.
[RFC6035]	Pendleton, A., Clark, A., Johnston, A. and H. Sinnreich, "Session Initiation Protocol Event Package for Voice Quality Reporting", RFC 6035, November 2010.
[RFC5104]	Wenger, S., Chandra, U., Westerlund, M. and B. Burman, "Session Initiation Protocol Event Package for Voice Quality Reporting", RFC 5104, February 2008.
[RFC6332]	Begen, A. and E. Friedrich, "Multicast Acquisition Report Block Type for RTP Control Protocol (RTCP) Extended Reports (XRs) ", RFC 6332, July 2011.
[PDV]	Hunt, G., "RTCP XR Report Block for Packet Delay Variation Metric Reporting", ID draft-ietf-avt-rtcp-xr-pdv-03, May 2009.
[MQ]	Wu, Q., Zorn, G., Schott, R. and K. Lee, "RTCP XR Blocks for multimedia quality metric reporting", ID draft-wu-xrblock-rtcp-xr-quality-monitoring-02, May 2011.
[G1020]	ITU-T, "ITU-T Rec. G.1020, Performance parameter definitions for quality of speech and other voiceband applications utilizing IP networks", July 2006.
[Y1540]	

	, , "ITU-T Rec. Y.1540, IP packet transfer and availability performance parameters", November 2007.
[H323]	, , "ITU-T Rec. H.323, Packet-based multimedia communications systems", June 2006.
[ECN]	Westerlund, M., Johansson, I., Perkins, C., O'Hanlon, P. and K. Carlberg, "Explicit Congestion Notification (ECN) for RTP over UDP", ID draft-ietf-avtcore-ecn-for-rtp-04, July 2011.

Appendix A. Change Log

Note to the RFC-Editor: please remove this section prior to publication as an RFC.

Appendix A.1. draft-ietf-avtcore-monarch-00

The following are the major changes compared to draft-hunt-avtcore-monarch-02:

- *Move Geoff Hunt and Philip Arden to acknowledgement section.

Appendix A.2. draft-ietf-avtcore-monarch-01

The following are the major changes compared to 00:

- *Restructure the document by merging section 4 into section 3.
- *Remove section 4.1,section 5 that is out of scope of this document.
- *Remove the last bullet in section 6 and section 7.3 based on conclusion of last meeting.
- *Update figure 1 and related text in section 3 according to the monitor definition in RFC3550.
- *Revise section 9 to address monitor declaration issue.
- *Merge the first two bullet in section 6.
- *Add one new bullet to discuss metric block association in section 6.

Appendix A.3. draft-ietf-avtcore-monarch-02

The following are the major changes compared to 01:

- *Deleting first paragraph of Section 1.

* Deleting Section 3.1, since the interaction with the management application is out of scope of this draft.

* Separeate identity information correlation from section 5.2 as new section 5.3.

* Remove figure 2 and related text from section 5.2.

* Editorial changes in the section 4 and the first paragraph of section 7.

Appendix A.4. draft-ietf-avtcore-monarch-03

The following are the major changes compared to 02:

* Update bullet 2 in section 4 to explain the ill-effect of Identity Information duplication.

* Update bullet 3 in section 4 to explain why Correlating RTCP XR with the non-RTP data is needed.

* Update section 5.2 to focus on how to reduce the identity information repetition

* Update section 5.3 to explain how to correlate identity information with the non-RTP data

Appendix A.5. draft-ietf-avtcore-monarch-04

The following are the major changes compared to 03:

* Update section 5.2 to clarify using SDES packet to carry correlation information.

* Remove section 5.3 since additional identity information goes to SDES packet and using SSRC to identify each block is standard RTP feature.

* Swap the last two paragraphs in the section 4 since identity information duplication can not been 100% avoided.

* Other editorial changes.

Authors' Addresses

Qin Wu editor Wu Huawei 101 Software Avenue, Yuhua District Nanjing, Jiangsu 210012 China EMail: sunseawq@huawei.com

Geoff Hunt Hunt Unaffiliated EMail: r.geoff.hunt@gmail.com

Philip Arden Arden BT Orion 3/7 PP4 Adastral Park Martlesham Heath
Ipswich, Suffolk IP5 3RE United Kingdom Phone: +44 1473 644192
EMail: philip.arden@bt.com