Framework for Real-time Media Congestion Avoidance Techniques

draft-zhu-rmcat-framework-00

Abstract

Congestion control is an essential element in ensuring fair bandwidth usage and preventing congestion collapse for traffic sharing the Internet. For interactive real-time media traffic such as video conferencing, design of congestion control solution also needs to account for many other factors such as the requirement for low latency packet delivery and interactions with live video encoder. This document describes a common framework with the core functional building blocks for a real-time media congestion solutions.

Status of This Memo

This Internet-Draft is submitted in full conformance with the provisions of BCP 78 and BCP 79.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF). Note that other groups may also distribute working documents as Internet-Drafts. The list of current Internet-Drafts is at http://datatracker.ietf.org/drafts/current/.

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."

This Internet-Draft will expire on January 8, 2017.

Copyright Notice

Copyright (c) 2016 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to BCP 78 and the IETF Trust’s Legal Provisions Relating to IETF Documents (http://trustee.ietf.org/license-info) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect
1. Introduction

Given increasing amount of interactive real-time media traffic over the Internet, such as video conferencing, it is important that these applications employ proper congestion control mechanisms to avoid congestion collapse. [I-D.ietf-rmcat-cc-requirements] specifies the list of requirements of a viable solution.

This document outlines a common framework for designing a congestion control mechanism for real-time interactive communication, so that individual drafts on specific solutions follow a consistent set of terminologies in describing their respective components. The next section (Section 3) describes common functional modules in this framework, whereas Section 4 provides examples on how these modules build together to support single and multiple media streams.

[ Editor’s note : This document does not describe the interaction between application, codec and congestion control system. The interaction among application, codec and congestion control system are defined in other documents. There is a possibility to merge all the documents into one single document. ]

2. Key Words for Requirements

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 [RFC2119].
3. Functional Modules

A viable solution for real-time media congestion control needs to comprise of several common modules. This section provides a brief description of them and their respective functionalities. A congestion control solution for real-time media should comprise of the described functional modules. This should help understanding different congestion control solutions.

- Network Congestion Controller: this is the core module for estimating available bandwidth over the network based on periodic RTCP feedback reports [RFC3550] from the receiver. This module contains key functions and calculations required to detect congestion and estimate available bandwidth on the transmission path based on the reception quality of the media traffic. Different congestion control solutions employ different algorithms in detecting congestion and estimating available bandwidth for its media flow. It is also possible that multiple media streams are multiplexed over a single transport, hence share a common congestion control module in aggregation.

- Transmission Queue: this module is needed to absorb the instantaneous mismatch between output from a live video encoder and regulated outgoing media flow. The transmission queue schedules outgoing traffic according to sending rate recommended by the rate controller module. It reports back its occupancy level to the rate controller module to assist future rate control decisions on target video rate, sending rate, and probing rate.

- Rate Controller: this module takes the estimated available bandwidth from the network congestion controller, shared states of other flows, as well as occupancy level of the transmission queue as input. It makes holistic decisions on: a) target video rate for the live video encoder; b) sending rate for regulating outgoing media flow(s) for the transmission queue; and c) rate of probing packets when needed. In the case where multiple media streams share a single transport and a common network congestion controller (for estimating available bandwidth in aggregation), the rate controller is also responsible for distributing available bandwidth amongst different media streams according to their relative priorities as well as share state information. When losses occur over the network and some previous media packets need to be retransmitted, the rate controller should also account for the bandwidth needed for retransmission.

- Network Probe Generator: A congestion control solution can actively probe to estimate the available bandwidth on the media transmission path by sending more than what the live video encoder
produces. Such an approach can be especially effective during the ramp up period of media and transmission rates, when no congestion has been observed over the network yet. The network probe generator is responsible for generating probing packets according to the probing rate specified by the rate controller. It can employ different techniques in doing so -- for example by generating simple dummy packets with unknown payload type or by generating Forward Error Correction (FEC) packets. While this document does not specify what probing technique to use or how those packets should be generated, a complete congestion control solution needs should specify total rate of the probe packets via the rate controller module.

- **Live Video Encoder:** The sender typically also contains a live video encoder, which adjusts its encoding parameters according to the target video rate set by the rate controller. The output rate from the video encoder may deviate from this target due to uncertainty in the captured video content characteristics and the encoder rate control process. The output encoded media packets are fed to the transmission queue. Note that internal operations of the live video encoder (i.e., how video encoder rate control works) is out of scope for this document.

- **Shared State:** In the case of multiple media streams sharing a common sender hence a common network congestion controller, the sender should also contain a shared state module for storage and exchange of congestion control states [Editor’s Note from Xiaoqing: examples of congestion control states??] amongst the multiple flows.

### 4. Example Configurations

#### 4.1. Example Configurations for a Single Stream
Figure 1: RMCAT Solution Framework at the Sender: Single Stream

Figure 1 shows an example configuration at the sender for supporting a single media stream. The Network Congestion Controller estimates available bandwidth based on periodic RTCP feedback reports. The Rate Controller takes as input the estimated available bandwidth (1) and the current occupancy level of the Transmission Queue (2). It calculates as output sending rate (3) for the Transmission Queue, video target rate (4) for the Live Video Encoder, and probing rate (5) -- if they are needed -- for the Network Probe Generator. The Transmission Queue holds packets generated by both the Live Video Encoder and the Network Probe Generator; it paces transmission of its outgoing packets according to the sending rate (3) specified by Rate Controller.

Obviously, it is possible for a congestion control solution to contain alternative configurations between these functional modules. [TODO: add one quick example on alternative wiring.] It is required that the candidate solution draft specify how their internal functional modules align to this framework.
4.2. Example Configurations for Multiple Streams

Figure 2 shows an example configuration for multiple video streams sharing a common Network Congestion Controller. The Network Congestion Controller calculates an aggregated estimated available bandwidth (1) based on periodic RTCP feedback reports. The Rate Controller divides up the aggregate estimated bandwidth (1) from the Network Congestion Controller amongst sub-streams based on their relative priority levels, Shared States, as well as current occupancy level of the Transmission Queue. It subsequently determines the per-flow sending rate (3) as regulated by the Transmission Queue and target video rate (4) for each flow.

In this specific example, the transmission queue is envisioned as a logical entity. For instance, this transmission queue can be implemented priority-based scheduling and one physical queue per stream. For sake of simplicity the role of Network Probe Generator is omitted in the above figure.
5. Acknowledgements

The RMCAT design team discussions contributed to this memo.

6. IANA Considerations

This memo includes no request to IANA.

7. Security Considerations

TBD

8. References

8.1. Normative References


8.2. Informative References


Authors’ Addresses

Xianqing Zhu  
Cisco Systems  
12515 Research Blvd., Building 4  
Austin, TX  78759  
USA  
Email: xiaoqzhu@cisco.com

Zaheduzzaman Sarker  
Ericsson AB  
Luleå  
Sweden  
Phone: 00 46 10 717 37 43  
Email: zaheduzzaman.sarker@ericsson.com