Internet Engineering Task Force (IETF)
Request for Comments: 8593
Category: Informational
ISSN: 2070-1721

X. Zhu
S. Mena
Cisco Systems
Z. Sarker
Ericsson AB

May 2019

Video Traffic Models for RTP Congestion Control Evaluations

Abstract

This document describes two reference video traffic models for evaluating RTP congestion control algorithms. The first model statistically characterizes the behavior of a live video encoder in response to changing requests on the target video rate. The second model is trace-driven and emulates the output of actual encoded video frame sizes from a high-resolution test sequence. Both models are designed to strike a balance between simplicity, repeatability, and authenticity in modeling the interactions between a live video traffic source and the congestion control module. Finally, the document describes how both approaches can be combined into a hybrid model.

Status of This Memo

This document is not an Internet Standards Track specification; it is published for informational purposes.

This document is a product of the Internet Engineering Task Force (IETF). It represents the consensus of the IETF community. It has received public review and has been approved for publication by the Internet Engineering Steering Group (IESG). Not all documents approved by the IESG are candidates for any level of Internet Standard; see Section 2 of RFC 7841.

Information about the current status of this document, any errata, and how to provide feedback on it may be obtained at https://www.rfc-editor.org/info/rfc8593.

Copyright Notice

Copyright (c) 2019 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 (https://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 to this document. Code Components extracted from this document must include Simplified BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Simplified BSD License.

Table of Contents

1. Introduction 3
2. Terminology 3
3. Desired Behavior of a Synthetic Video Traffic Model 4
4. Interactions between Synthetic Video Traffic Source and Other Components at the Sender 5
5. A Statistical Reference Model 7
5.1. Time-Damped Response to Target-Rate Update 9
5.2. Temporary Burst and Oscillation during the Transient Period 9
5.3. Output-Rate Fluctuation at Steady State 9
5.4. Rate Range Limit Imposed by Video Content 10
6. A Trace-Driven Model 10
6.1. Choosing the Video Sequence and Generating the Traces . . 11
6.2. Using the Traces in the Synthetic Codec 13
6.2.1. Main Algorithm 13
6.2.2. Notes to the Main Algorithm 14
6.3. Varying Frame Rate and Resolution 15
7. Combining the Two Models 16
8. Reference Implementation 17
9. IANA Considerations 17
10. Security Considerations 17
11. References 17
11.1. Normative References 17
11.2. Informative References 18
Authors' Addresses 19

1. Introduction

When evaluating candidate congestion control algorithms designed for real-time interactive media, it is important to account for the characteristics of traffic patterns generated from a live video encoder. Unlike synthetic traffic sources that can conform perfectly to the rate-changing requests from the congestion control module, a live video encoder can be sluggish in reacting to such changes. The output rate of a live video encoder also typically deviates from the target rate due to uncertainties in the encoder rate-control process. Consequently, end-to-end delay and loss performance of a real-time media flow can be further impacted by rate variations introduced by the live encoder.

On the other hand, evaluation results of a candidate RTP congestion control algorithm should mostly reflect the performance of the congestion control module and somewhat decouple from peculiarities of any specific video codec. It is also desirable that evaluation tests are repeatable and easily duplicated across different candidate algorithms.

One way to strike a balance between the above considerations is to evaluate congestion control algorithms using a synthetic video traffic source model that captures key characteristics of the behavior of a live video encoder. The synthetic traffic model should also contain tunable parameters so that it can be flexibly adjusted to reflect the wide variations in real-world live video encoder behaviors. To this end, this document presents two reference models. The first is based on statistical modeling. The second is driven by frame size and interval traces recorded from a real-world encoder. This document also discusses the pros and cons of each approach, as well as how both approaches can be combined into a hybrid model.

2. Terminology

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 BCP 14 [RFC2119] [RFC8174] when, and only when, they appear in all capitals, as shown here.

3. Desired Behavior of a Synthetic Video Traffic Model

A live video encoder employs encoder rate control to meet a target rate by varying its encoding parameters, such as quantization step size, frame rate, and picture resolution, based on its estimate of the video content (e.g., motion and scene complexity). In practice, however, several factors prevent the output video rate from perfectly conforming to the input target rate.

Due to uncertainties in the captured video scene, the output rate typically deviates from the specified target. In the presence of a significant change in target rate, the encoder's output frame sizes sometimes fluctuate for a short, transient period of time before the output rate converges to the new target. Finally, while most of the frames in a live session are encoded in predictive mode (i.e., P-frames in [H264]), the encoder can occasionally generate a large intra-coded frame (i.e., I-frame as defined in [H264]) or a frame partially containing intra-coded blocks in an attempt to recover from losses, to re-sync with the receiver, or during the transient period of responding to target rate or spatial resolution changes.

Hence, a synthetic video source should have the following capabilities:

- To change bitrate. This includes the ability to change frame rate and/or spatial resolution or to skip frames upon request.

- To fluctuate around the target bitrate specified by the congestion control module.

- To show a delay in convergence to the target bitrate.

- To generate intra-coded or repair frames on demand.

While there exist many different approaches in developing a synthetic video traffic model, it is desirable that the outcome follows a few common characteristics, as outlined below.

- Low computational complexity: The model should be computationally lightweight, otherwise, it defeats the whole purpose of serving as a substitute for a live video encoder.

- Temporal pattern similarity: The individual traffic trace instances generated by the model should mimic the temporal pattern of those from a real video encoder.

- o Statistical resemblance: The synthetic traffic source should match the outcome of the real video encoder in terms of statistical characteristics, such as the mean, variance, peak, and autocorrelation coefficients of the bitrate. It is also important that the statistical resemblance should hold across different time scales ranging from tens of milliseconds to sub-seconds.

- o A wide range of coverage: The model should be easily configurable to cover a wide range of codec behaviors (e.g., with either fast or slow reaction time in live encoder rate control) and video content variations (e.g., ranging from high to low motion).

These distinct behavior features can be characterized via simple statistical modeling or a trace-driven approach. Sections 5 and 6 provide an example of each approach, respectively. Section 7 discusses how both models can be combined together.

4. Interactions between Synthetic Video Traffic Source and Other Components at the Sender

Figure 1 depicts the interactions of the synthetic video traffic source with other components at the sender, such as the application, the congestion control module, the media packet transport module, etc. Both reference models, as described later in Sections 5 and 6, follow the same set of interactions.

The synthetic video source dynamically generates a sequence of dummy video frames with varying size and interval. These dummy frames are processed by other modules in order to transmit the video stream over the network. During the lifetime of a video transmission session, the synthetic video source will typically be required to adapt its encoding bitrate and sometimes the spatial resolution and frame rate.

In this model, the synthetic video source module has a group of incoming and outgoing interface calls that allow for interaction with other modules. The following are some of the possible incoming interface calls, marked as (a) in Figure 1, that the synthetic video traffic source may accept. The list is not exhaustive and can be complemented by other interface calls if necessary.

- o Target bitrate R_v: Target bitrate request measured in bits per second (bps). Typically, the congestion control module calculates the target bitrate and updates it dynamically over time. Depending on the congestion control algorithm in use, the update requests can either be periodic (e.g., once per second), or on-demand (e.g., only when a drastic bandwidth change over the network is observed).

- o Target frame rate FPS: The instantaneous frame rate measured in frames per second at a given time. This depends on the native camera-capture frame rate as well as the target/preferred frame rate configured by the application or user.

- o Target frame resolution XY: The 2-dimensional vector indicating the preferred frame resolution in pixels. Several factors govern the resolution requested to the synthetic video source over time. Examples of such factors include the capturing resolution of the native camera and the display size of the destination screen. The target frame resolution also depends on the current target bitrate R_v, since it does not make sense to pair very low spatial resolutions with very high bitrates, and vice-versa.

- o Instant frame skipping: The request to skip the encoding of one or several captured video frames, for instance, when a drastic decrease in available network bandwidth is detected.

- o On-demand generation of intra (I) frame: The request to encode another I-frame to avoid further error propagation at the receiver when severe packet losses are observed. This request typically comes from the error control module. It can be initiated either by the sender or by the receiver via Full Intra Request (FIR) messages as defined in [RFC5104].

An example of an outgoing interface call, marked as (b) in Figure 1, is the rate range [R_min, R_max]. Here, R_min and R_max are meant to capture the dynamic rate range the actual live video encoder is capable of generating given the input video content. This typically depends on the video content complexity and/or display type (e.g., higher R_max for video content with higher motion complexity or for displays of higher resolution). Therefore, these values will not change with R_v but may change over time if the content is changing.

```
                          +-------------+
                          |             |  dummy encoded
                          |  Synthetic  |   video frames
                          |    Video    | -------------->
                          |   Source    |
                          |             |
                          +--------+----+
                              /|\   |
                               |    |
                   -------------------+    +-------------------->
                      interface from          interface to
                     other modules (a)       other modules (b)
```

Figure 1: Interaction between Synthetic Video Encoder and Other Modules at the Sender

5. A Statistical Reference Model

This section describes one simple statistical model of the live video traffic source. Figure 2 summarizes the list of tunable parameters in this statistical model. A more comprehensive survey of popular methods for modeling the behavior of video traffic sources can be found in [Tanwir2013].

Notation	Parameter Name	Example Value
R_v	Target bitrate request	1 Mbps
FPS	Target frame rate	30 Hz
tau_v	Encoder reaction latency	0.2 s
K_d	Burst duration of the transient period	8 frames
K_B	Burst frame size during the transient period	13.5 KB*
t0	Reference frame interval 1/FPS	33 ms
B0	Reference frame size R_v/8/FPS	4.17 KB
SCALE_t	Scaling parameter of the zero-mean Laplacian distribution describing deviations in normalized frame interval (t-t0)/t0	0.15
SCALE_B	Scaling parameter of the zero-mean Laplacian distribution describing deviations in normalized frame size (B-B0)/B0	0.15
R_min	Minimum rate supported by video encoder type or content activity	150 kbps
R_max	Maximum rate supported by video encoder type or content activity	1.5 Mbps

* Example value of K_B for a video stream encoded at 720p and 30 frames per second using H.264/AVC encoder

Figure 2: List of Tunable Parameters in a Statistical Video Traffic Source Model

5.1. Time-Damped Response to Target-Rate Update

While the congestion control module can update its target bitrate request R_v at any time, the statistical model dictates that the encoder will only react to such changes tau_v seconds after a previous rate transition. In other words, when the encoder has reacted to a rate-change request at time t, it will simply ignore all subsequent rate-change requests until time t+tau_v.

5.2. Temporary Burst and Oscillation during the Transient Period

The output bitrate R_o during the period [t, t+tau_v] is considered to be in a transient state when reacting to abrupt changes in target rate. Based on observations from video encoder output, the encoder reaction to a new target bitrate request can be characterized by high variations in output frame sizes. It is assumed in the model that the overall average output bitrate R_o during this transient period matches the target bitrate R_v. Consequently, the occasional burst of large frames is followed by smaller-than-average encoded frames.

This temporary burst is characterized by two parameters:

- o burst duration K_d: Number of frames in the burst event, and

- o burst frame size K_B: Size of the initial burst frame, which is typically significantly larger than the average frame size at steady state.

It can be noted that these burst parameters can also be used to mimic the insertion of a large on-demand I-frame in the presence of severe packet losses. The values of K_d and K_B typically depend on the type of video codec, spatial and temporal resolution of the encoded stream, as well as the activity level in the video content.

5.3. Output-Rate Fluctuation at Steady State

The output bitrate R_o during steady state is modeled as randomly fluctuating around the target bitrate R_v. The output traffic can be characterized as the combination of two random processes that denote the frame interval t and output frame size B over time, which are the two major sources of variations in the encoder output. For simplicity, the deviations of t and B from their respective reference levels are modeled as independent and identically distributed (i.i.d) random variables following the Laplacian distribution [Papoulis]. More specifically:

- o Fluctuations in frame interval: The intervals between adjacent frames have been observed to fluctuate around the reference interval of t0 = 1/FPS. Deviations in normalized frame interval DELTA_t = (t-t0)/t0 can be modeled by a zero-mean Laplacian distribution with scaling parameter SCALE_t. The value of SCALE_t dictates the "width" of the Laplacian distribution and therefore the amount of fluctuation in actual frame intervals (t) with respect to the reference frame interval t0.

- o Fluctuations in frame size: The output-encoded frame sizes also tend to fluctuate around the reference frame size B0=R_v/8/FPS. Likewise, deviations in the normalized frame size DELTA_B = (B-B0)/B0 can be modeled by a zero-mean Laplacian distribution with scaling parameter SCALE_B. The value of SCALE_B dictates the "width" of this second Laplacian distribution and correspondingly the amount of fluctuations in output frame sizes (B) with respect to the reference target B0.

Both values of SCALE_t and SCALE_B can be obtained via parameter fitting from empirical data captured for a given video encoder. Example values are listed in Figure 2 based on empirical data presented in [IETF-Interim].

5.4. Rate Range Limit Imposed by Video Content

The output bitrate R_o is further clipped within the dynamic range [R_min, R_max], which in reality are dictated by scene and motion complexity of the captured video content. In the proposed statistical model, these parameters are specified by the application.

6. A Trace-Driven Model

The second approach for modeling a video traffic source is trace-driven. This can be achieved by running an actual live video encoder on a set of chosen raw video sequences and using the encoder’s output traces for constructing a synthetic video source. With this approach, the recorded video traces naturally exhibit temporal fluctuations around a given target bitrate request R_v from the congestion control module.

The following list summarizes the main steps of this approach:

1. Choose one or more representative raw video sequences.

2. Encode the sequence(s) using an actual live video encoder. Repeat the process for a number of bitrates. Keep only the sequence of frame sizes for each bitrate.

3. Construct a data structure that contains the output of the previous step. The data structure should allow for easy bitrate lookup.

4. Upon a target bitrate request R_v from the controller, look up the closest bitrates among those previously stored. Use the frame-size sequences stored for those bitrates to approximate the frame sizes to output.

5. The output of the synthetic video traffic source contains "encoded" frames with dummy contents but with realistic sizes.

Section 6.1 explains the first three steps (1-3), and Section 6.2 elaborates on the remaining two steps (4-5). Finally, Section 6.3 briefly discusses the possibility to extend the trace-driven model for supporting time-varying frame rate and/or time-varying frame resolution.

6.1. Choosing the Video Sequence and Generating the Traces

The first step is a careful choice of a set of video sequences that are representative of the target use cases for the video traffic model. For the example use case of interactive video conferencing, it is recommended to choose a sequence with content that resembles a "talking head", e.g., from a news broadcast or recording of an actual video conferencing call.

The length of the chosen video sequence is a tradeoff. If it is too long, it will be difficult to manage the data structures containing the traces. If it is too short, there will be an obvious periodic pattern in the output frame sizes, leading to biased results when evaluating congestion control performance. It has been empirically determined that a sequence 2 to 4 minutes in length sufficiently avoids the periodic pattern.

Given the chosen raw video sequence, denoted "S", one can use a live encoder, e.g., some implementation of [H264] or [H265], to produce a set of encoded sequences. As discussed in Section 3, the output bitrate of the live encoder can be achieved by tuning three input parameters: quantization step size, frame rate, and picture resolution. In order to simplify the choice of these parameters for a given target rate, one can typically assume a fixed frame rate (e.g., 30 fps) and a fixed resolution (e.g., 720p) when configuring the live encoder. See Section 6.3 for a discussion on how to relax these assumptions.

Following these simplifications, the chosen encoder can be configured to start at a constant target bitrate, then vary the quantization step size (internally via the video encoder rate controller) to meet various externally specified target rates. It can be further assumed the first frame is encoded as an I-frame and the rest are P-frames (see, e.g., [H264] for definitions of I-frames and P-frames). For live encoding, the encoder rate-control algorithm typically does not use knowledge of frames in the future when encoding a given frame.

Given the minimum and maximum bitrates at which the synthetic codec is to operate (denoted as "R_min" and "R_max", see Section 4), the entire range of target bitrates can be divided into n_s steps. This leads to an encoding bitrate ladder of (n_s + 1) choices equally spaced apart by the step length l = (R_max - R_min)/n_s. The following simple algorithm is used to encode the raw video sequence.

```
r = R_min
while r <= R_max do
    Traces[r] = encode_sequence(S, r, e)
    r = r + l
```

The function encode_sequence takes as input parameters, respectively, a raw video sequence (S), a constant target rate (r), and an encoder rate-control algorithm (e); it returns a vector with the sizes of frames in the order they were encoded. The output vector is stored in a map structure called "Traces", whose keys are bitrates and whose values are vectors of frame sizes.

The choice of a value for the number of bitrate steps n_s is important, since it determines the number of vectors of frame sizes stored in the map Traces. The minimum value one can choose for n_s is 1; the maximum value depends on the amount of memory available for holding the map Traces. A reasonable value for n_s is one that results in steps of length l = 200 kbps. Section 6.2.2 will discuss further the choice of step length l.

Finally, note that, as mentioned in previous sections, R_min and R_max may be modified after the initial sequences are encoded. Henceforth, for notational clarity, we refer to the bitrate range of the trace file as [Rf_min, Rf_max]. The algorithm described in Section 6.2.1 also covers the cases when the current target bitrate is less than Rf_min or greater than Rf_max.

6.2. Using the Traces in the Synthetic Codec

The main idea behind the trace-driven synthetic codec is that it mimics the rate-adaptation behavior of a real live codec upon dynamic updates of the target bitrate request R_v by the congestion control module. It does so by switching to a different frame-size vector stored in the map Traces when needed.

6.2.1. Main Algorithm

The main algorithm for rate adaptation in the synthetic codec maintains two variables: r_current and t_current.

- o The variable r_current points to one of the keys of map Traces. Upon a change in the value of R_v, typically because the congestion controller detects that the network conditions have changed, r_current is updated based on R_v as follows:

```
R_ref = min (Rf_max, max(Rf_min, R_v))

r_current = r
such that
    (r in keys(Traces)  and
     r <= R_ref  and
    (not(exists) r' in keys(Traces) such that r <r'<= R_ref))
```

- o The variable t_current is an index to the frame-size vector stored in Traces[r_current]. It is updated every time a new frame is due. It is assumed that all vectors stored in Traces have the same size, denoted as "size_traces". The following equation governs the update of t_current:

```
if t_current < SkipFrames then
    t_current = t_current + 1
else
    t_current = ((t_current + 1 - SkipFrames)
                 % (size_traces-SkipFrames)) + SkipFrames
```

where operator "%" denotes modulo, and SkipFrames is a predefined constant that denotes the number of frames to be skipped at the beginning of frame-size vectors after t_current has wrapped around. The point of constant SkipFrames is avoiding the effect of periodically sending a large I-frame followed by several smaller-than-average P-frames. A typical value of SkipFrames is 20, although it could be set to 0 if one is interested in studying the effect of sending I-frames periodically.

The initial value of r_current is set to R_min, and the initial value of t_current is set to 0.

When a new frame is due, its size can be calculated following one of the three cases below:

a) Rf_min <= R_v < Rf_max: The output frame size is calculated via linear interpolation of the frame sizes appearing in Traces[r_current] and Traces[r_current + l]. The interpolation is done as follows:

```
size_lo = Traces[r_current][t_current]
size_hi = Traces[r_current + l][t_current]
distance_lo = (R_v - r_current) / l
framesize = size_hi*distance_lo + size_lo*(1-distance_lo)
```

b) R_v < Rf_min: The output frame size is calculated via scaling with respect to the lowest bitrate Rf_min in the trace file, as follows:

```
w = R_v / Rf_min
framesize = max(fs_min, factor * Traces[Rf_min][t_current])
```

c) R_v >= Rf_max: The output frame size is calculated by scaling with respect to the highest bitrate Rf_max in the trace file, as follows:

```
w = R_v / Rf_max
framesize = min(fs_max, w * Traces[Rf_max][t_current])
```

In cases b) and c), floating-point arithmetic is used for computing the scaling factor "w". The resulting value of the instantaneous frame size (framesize) is further clipped within a reasonable range between fs_min (e.g., 10 bytes) and fs_max (e.g., 1 MB).

6.2.2. Notes to the Main Algorithm

Note that the main algorithm as described above can be further extended to mimic some additional typical behaviors of a live video encoder. Two examples are given below:

- I-frames on demand: The synthetic codec can be extended to simulate the sending of I-frames on demand, e.g., as a reaction to losses. To implement this extension, the codec's incoming interface (see (a) in Figure 1) is augmented with a new function to request a new I-frame. Upon calling such function, t_current is reset to 0.

- o Variable step length l between R_min and R_max: In the main algorithm, the step length l is fixed for ease of explanation. However, if the range [R_min, R_max] is very wide, it is also possible to define a set of intermediate encoding rates with variable step length. The rationale behind this modification is that the difference between 400 and 600 kbps as target bitrate is much more significant than the difference between 4400 kbps and 4600 kbps. For example, one could define steps of length 200 kbps under 1 Mbps, then steps of length 300 kbps between 1 Mbps and 2 Mbps, then 400 kbps between 2 Mbps and 3 Mbps, and so on.

6.3. Varying Frame Rate and Resolution

The trace-driven synthetic codec model explained in this section is relatively simple due to the choice of fixed frame rate and frame resolution. The model can be extended further to accommodate variable frame rate and/or variable spatial resolution.

When the encoded picture quality at a given bitrate is low, one can potentially decrease either the frame rate (if the video sequence is currently in low motion) or the spatial resolution in order to improve quality of experience (QoE) in the overall encoded video. On the other hand, if target bitrate increases to a point where there is no longer a perceptible improvement in the picture quality of individual frames, then one might afford to increase the spatial resolution or the frame rate (useful if the video is currently in high motion).

Many techniques have been proposed to choose over time the best combination of encoder-quantization step size, frame rate, and spatial resolution in order to maximize the quality of live video codecs [Ozer2011] [Hu2012]. Future work may consider extending the trace-driven codec to accommodate variable frame rate and/or resolution.

From the perspective of congestion control, varying the spatial resolution typically requires a new intra-coded frame to be generated, thereby incurring a temporary burst in the output traffic pattern. The impact of frame-rate change tends to be more subtle: reducing frame rate from high to low leads to sparsely spaced larger encoded packets instead of many densely spaced smaller packets. Such difference in traffic profiles may still affect the performance of congestion control, especially when outgoing packets are not paced by the media transport module. Investigation of varying frame rate and resolution are left for future work.

7. Combining the Two Models

It is worthwhile noting that the statistical and trace-driven models each have their own advantages and drawbacks. Both models are fairly simple to implement. It takes significantly greater effort to fit the parameters of a statistical model to actual encoder output data. In contrast, it is straightforward for a trace-driven model to obtain encoded frame-size data. Once validated, the statistical model is more flexible in mimicking a wide range of encoder/content behaviors by simply varying the corresponding parameters in the model. In this regard, a trace-driven model relies, by definition, on additional data-collection efforts for accommodating new codecs or video contents.

In general, the trace-driven model is more realistic for mimicking the ongoing steady-state behavior of a video traffic source with fluctuations around a constant target rate. In contrast, the statistical model is more versatile for simulating the behavior of a video stream in transient, such as when encountering sudden rate changes. It is also possible to combine both methods into a hybrid model. In this case, the steady-state behavior is driven by traces during steady state and the transient-state behavior is driven by the statistical model.

```
                                 transient +---------------+
                                   state   | Generate next |
                                   +------>| K_d transient |
            +-----------------+  /         |    frames     |
       R_v  | Compare against | /          +---------------+
     ------>|   previous      |/
            | target bitrate  |\
            +-----------------+ \          +---------------+
                                 \         | Generate next |
                                  +------>|  frame from   |
                                   steady |    trace      |
                                    state +---------------+
```

Figure 3: A Hybrid Video Traffic Model

As shown in Figure 3, the video traffic model operates in a transient state if the requested target rate R_v is substantially different from the previous target; otherwise, it operates in a steady state. During the transient state, a total of K_d frames are generated by the statistical model, resulting in one (1) big burst frame with size K_B followed by K_d-1 smaller frames. When operating at steady state, the video traffic model simply generates a frame according to the trace-driven model given the target rate while modulating the frame interval according to the distribution specified by the

statistical model. One example criterion for determining whether the traffic model should operate in a transient state is whether the rate change exceeds 10% of the previous target rate. Finally, as this model follows transient-state behavior dictated by the statistical model, upon a substantial rate change, the model will follow the time-damping mechanism as defined in Section 5.1, which is governed by parameter tau_v.

8. Reference Implementation

The statistical, trace-driven, and hybrid models as described in this document have been implemented as a stand-alone, platform-independent synthetic traffic source module. It can be easily integrated into network simulation platforms such as [ns-2] and [ns-3], as well as testbeds using a real network. The stand-alone traffic source module is available as an open-source implementation at [Syncodecs].

9. IANA Considerations

This document has no IANA actions.

10. Security Considerations

The synthetic video traffic models as described in this document do not impose any security threats. They are designed to mimic realistic traffic patterns for evaluating candidate RTP-based congestion control algorithms so as to ensure stable operations of the network. It is RECOMMENDED that candidate algorithms be tested using the video traffic models presented in this document before wide deployment over the Internet. If the generated synthetic traffic flows are sent over the Internet, they also need to be congestion controlled.

11. References

11.1. Normative References

[RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, DOI 10.17487/RFC2119, March 1997, <https://www.rfc-editor.org/info/rfc2119>.

[RFC8174] Leiba, B., "Ambiguity of Uppercase vs Lowercase in RFC 2119 Key Words", BCP 14, RFC 8174, DOI 10.17487/RFC8174, May 2017, <https://www.rfc-editor.org/info/rfc8174>.

11.2. Informative References

[H264] ITU-T, "Advanced video coding for generic audiovisual services", Recommendation H.264, April 2017, <https://www.itu.int/rec/T-REC-H.264>.

[H265] ITU-T, "High efficiency video coding", Recommendation H.265, February 2018, <https://www.itu.int/rec/T-REC-H.265>.

[Hu2012] Hu, H., Ma, Z., and Y. Wang, "Optimization of Spatial, Temporal and Amplitude Resolution for Rate-Constrained Video Coding and Scalable Video Adaptation", Proc. 19th IEEE International Conference on Image Processing (ICIP), DOI 10.1109/ICIP.2012.6466960, September 2012.

[IETF-Interim] Zhu, X., Mena, S., and Z. Sarker, "Update on RMCAT Video Traffic Model: Trace Analysis and Model Update", IETF RMCAT Virtual Interim, April 2017, <https://www.ietf.org/proceedings/interim-2017-rmcat-01/slides/slides-interim-2017-rmcat-01-sessa-update-on-video-traffic-model-draft-00.pdf>.

[ns-2] "The Network Simulator - ns-2", December 2015, <https://nsnam.sourceforge.net/wiki/index.php/User_Information>.

[ns-3] "NS-3 Network Simulator", <https://www.nsnam.org/>.

[Ozer2011] Ozer, J., "Video Compression for Flash, Apple Devices and HTML5", Galax: Doceo Publishing, ISBN-13: 978-0976259503, 2011.

[Papoulis] Papoulis, A. and S. Pillai, "Probability, Random Variables and Stochastic Processes", London: McGraw-Hill Europe, ISBN-13: 978-0071226615, 2002.

[RFC5104] Wenger, S., Chandra, U., Westerlund, M., and B. Burman, "Codec Control Messages in the RTP Audio-Visual Profile with Feedback (AVPF)", RFC 5104, DOI 10.17487/RFC5104, February 2008, <https://www.rfc-editor.org/info/rfc5104>.

[Syncodecs]
"Syncodecs: Synthetic codecs for evaluation of RMCAT work", commit a92d6c8, May 2018, <https://github.com/cisco/syncodecs>.

[Tanwir2013]
Tanwir, S. and H. Perros, "A Survey of VBR Video Traffic Models", IEEE Communications Surveys and Tutorials, Volume 15, Issue 4, p. 1778-1802, DOI 10.1109/SURV.2013.010413.00071, January 2013.

Authors' Addresses

Xiaoqing Zhu
Cisco Systems
12515 Research Blvd., Building 4
Austin, TX 78759
United States of America

Email: xiaoqzhu@cisco.com

Sergio Mena
Cisco Systems
EPFL, Quartier de l'Innovation, Batiment E
Ecublens, Vaud 1015
Switzerland

Email: semena@cisco.com

Zaheduzzaman Sarker
Ericsson AB
Torshamnsgatan 23
Stockholm, SE 164 83
Sweden

Phone: +46 10 717 37 43
Email: zaheduzzaman.sarker@ericsson.com