Sally Floyd
ICIR
Eddie Kohler
UCLA
29 June 2007

Profile for Datagram Congestion Control Protocol (DCCP) Congestion ID 4: TCP-Friendly Rate Control for Small Packets (TFRC-SP) draft-floyd-dccp-ccid4-01.ps

Status of This Memo

Copyright Notice

Abstract

This document contains the profile for Congestion Control Identifier 4, the Small-Packet variant of TCP-Friendly Rate Control (TFRC), in the Datagram Congestion Control Protocol (DCCP). CCID 4 is for experimental use, and uses TFRC-SP [RFC4828], a variant of TFRC designed for applications that send small packets. The goal for TFRC-SP is to achieve roughly the same bandwidth in bits per second (bps) as a TCP flow using packets of up to 1500 bytes but experiencing the same level of congestion. CCID 4 is for experimental use for senders that send small packets and would like a TCP-friendly sending rate, possibly with Explicit Congestion Notification (ECN), while minimizing abrupt rate changes.

TO BE DELETED BY THE RFC EDITOR UPON PUBLICATION:

Changes from draft-floyd-dccp-ccid4-00.txt:

* Added a subsection describing calculation of the average loss interval in TFRC-SP.

* Changed the assumed DCCP-Data header size from 12 bytes to 16 bytes, for 48-bit sequence numbers. Feedback from Ian McDonald.

* Added that the CCID4 sender can send two packets in a burst, if limited by OS granularity. From Ian McDonald.

* Added that the implementor may track Faster Restart and implement it before an explicit update to the CCID4 RFC. From Ian McDonald.

* Added an example to Section 8.4 of when errors can occur in using the Window Counter to detect loss intervals of at most two round-trip times.

Changes from draft-floyd-ccid4-00.txt:

* Added the Dropped Packets option for reporting the number of packets dropped in a loss interval.

* Added examples to Section 8.4 of the receiver incorrectly inferring whether a loss interval was short or not.

END OF SECTION TO BE DELETED.

Table of Contents

List of Tables

1. Introduction

This document contains the profile for Congestion Control Identifier 4, TCP-Friendly Rate Control for Small Packets (TFRC-SP), in the Datagram Congestion Control Protocol (DCCP) [RFC4340]. CCID 4 differs from CCID 3 in that CCID 4 uses TFRC-SP, the Small-Packet variant of TFRC, while CCID 3 [RFC4342] uses standard TFRC [RFC3448]. This document assumes that the reader is familiar with [RFC4342], instead of repeating from that document unnecessarily.

CCID 4 differs from CCID 3 only in the following respects:

- Header size: For TFRC-SP, the allowed transmit rate in bytes per second is reduced by a factor that accounts for packet header size. This is specified for TFRC-SP in Section 4.2 of [RFC4828], and described for CCID 4 in Section 5 below.
- Minimum sending rate: TFRC-SP enforces a minimum interval of 10 milliseconds between data packets. This is specified for TFRC-SP in Section 4.3 of [RFC4828], and described for CCID 4 in Section 5 below.
- Loss rates for short loss intervals: For short lost intervals of at most two round-trip times, the loss rate is computed by counting the actual number of packets lost or marked. For such a short loss interval with N data packets, including K lost or marked data packets, the loss interval length is calculated as N/K, instead of as N. This is specified for TFRC-SP in Section 4.4 of [RFC4828]. The CCID 3 Dropped Packets option [CCID3-DP] is thus mandated in addition to CCID 3's Loss Intervals option, as specified in Section 8.7 below. This section also describes the use of the Dropped Packets option in calculating the loss event rate. The computation of the loss rate by the receiver for the Loss Event Rate option is described for CCID 4 in Section 8.4 below.
- The nominal segment size: In TFRC-SP, the nominal segment size used by the TCP throughput equation is set to 1460 bytes. This is specified for TFRC-SP in Section 4.5 of [RFC3448], and described for CCID 4 in Section 5 below.

2. Conventions

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].

Additional terminology is described in Section 2 of [RFC4342].

3. Usage

Like CCID 3, CCID 4's congestion control is appropriate for flows that would prefer to minimize abrupt changes in the sending rate, including streaming media applications with small or moderate receiver buffering before playback.

CCID 4 is designed to be used either by applications that use a small fixed segment size, or by applications that change their sending rate by varying the segment size. If CCID 4 is used by an application that varies its segment size in response to changes in the allowed sending rate in bps, we note that CCID 4 doesn't dictate the segment size to be used by the application; this is done by the application itself. The CCID 4 sender determines the allowed sending rate in bps, in response to on-going feedback from the CCID 4 receiver, and the application can use information about the current allowed sending rate to decide whether to change the current segment size.

We note that in some environments there will be a feedback loop, with changes in the packet size or in the sending rate in bps affecting congestion along the path, therefore affecting the allowed sending rate in the future.

3.1. Relationship with TFRC

The congestion control mechanisms described here follow the TFRC-SP mechanism specified in [RFC4828]. As with CCID 3, conformant CCID 4 implementations MAY track updates to the TCP throughput equation directly, as updates are standardized in the IETF, rather than waiting for revisions of this document. However, conformant implementations SHOULD wait for explicit updates to CCID 4 before implementing other changes to TFRC congestion control.

3.2. Example Half-Connection

This example shows the typical progress of a half-connection using CCID 4's TFRC Congestion Control, not including connection initiation and termination. The example is informative, not normative. This example differs from that for CCID 3 in [RFC4342] only in that the allowed transmit rate is determined by [RFC4828] as well as by [RFC3448].

1. The sender transmits DCCP-Data packets, where the sending rate is governed by the allowed transmit rate as specified in [RFC4828]. Each DCCP-Data packet has a sequence number, and the DCCP header's CCVal field contains the window counter value, used by the receiver in determining when multiple losses belong in a single loss event.

   In the typical case of an ECN-capable half-connection, each DCCP-Data and DCCP-DataAck packet is sent as ECN-Capable, with either the ECT(0) or the ECT(1) codepoint set. The use of the ECN Nonce with TFRC is described in Section 9.

2. The receiver sends DCCP-Ack packets at least once per round-trip time acknowledging the data packets, unless the sender is sending at a rate of less than one packet per round-trip time, as indicated by the TFRC specification [RFC3448] (Section 6). Each DCCP-Ack packet uses a sequence number, identifies the most recent packet received from the sender, and includes feedback about the recent loss intervals experienced by the receiver.

3. The sender continues sending DCCP-Data packets as controlled by the allowed transmit rate. Upon receiving DCCP-Ack packets, the sender updates its allowed transmit rate as specified in [RFC3448] (Section 4.3) and [RFC4828]. This update is based upon a loss event rate calculated by the sender, based on the receiver's loss intervals feedback. If it prefers, the sender can also use a loss event rate calculated and reported by the receiver.

4. The sender estimates round-trip times and calculates a nofeedback time, as specified in [RFC3448] (Section 4.4). If no feedback is received from the receiver in that time (at least four round-trip times), the sender halves its sending rate.

4. Connection Establishment

The connection establishment is as specified in Section 4 of [RFC4342].

5. Congestion Control on Data Packets

CCID 4 uses the congestion control mechanisms of TFRC [RFC3448] and TFRC-SP [RFC4828]. [RFC4828] should be considered normative except where specifically indicated.

Loss Event Rate

As with CCID 3, the basic operation of CCID 4 centers around the calculation of a loss event rate: the number of loss events as a fraction of the number of packets transmitted, weighted over the last several loss intervals. For CCID 4, this loss event rate, a round-trip time estimate, and a nominal packet size of 1460 bytes are plugged into the TCP throughput equation, as specified in RFC 3448 (Section 3.1) and [RFC4828].

Because CCID 4 is intended for applications that send small packets, the allowed transmit rate derived from the TCP throughput equation is reduced by a factor that accounts for packet header size, as specified in Section 4.2 of [RFC4828]. The header size on data packets is estimated as 36 bytes (20 bytes for the IP header, and 16 bytes for the DCCP-Data header with 48-bit sequence numbers). If the DCCP sender is sending N-byte data packets, the allowed transmit rate is reduced by N/(N+36). CCID 4 senders are limited to this fair rate. The header size would be 32 bytes instead of 36 bytes when 24-bit sequence numbers were used in the DCCP-Data header.

The loss event rate itself is calculated in CCID 4 using recent loss interval lengths reported by the receiver. Loss intervals are precisely defined in Section 6.1 of [RFC4342], with the modification in [RFC4828] (Section 3) for loss intervals of at most two round-trip times. In summary, a loss interval is up to 1 RTT of possibly lost or ECN-marked data packets, followed by an arbitrary number of non-dropped, non-marked data packets. The CCID 3 Loss Intervals option is used to report loss interval lengths; see Section 8.6.

For loss intervals of at most two round-trip times, CCID 4 calculates the loss event rate for that interval by counting the number of packets lost or marked, as described in Section 4.4 of [RFC4828]. Thus, for such a short loss interval with N data packets, including K lost or marked data packets, the loss interval length is calculated as N/K, instead as N. The CCID 3 Dropped Packets option is used to report K, the count of lost or marked data packets.

Unlike CCID 3, the CCID 4 sender enforces a minimum interval of 10 ms between data packets, regardless of the allowed transmit rate. If operating system scheduling granularity makes this impractical, up to one additional packet MAY be sent per timeslice, providing that no more than three packets are sent in any 30 ms interval.

Other Congestion Control Mechanisms

The other congestion control mechanisms such as slow-start, feedback packets, and the like are exactly as in CCID 3, and are described in the subsection on "Other Congestion Control Mechanisms" of Section 5 in [RFC4342].

5.1. Response to Idle and Application-limited Periods

This is described in Section 5.1 of [RFC4342]. If Faster Restart is standardized in the IETF for TFRC [KFS07], then Faster Restart MAY be implemented in CCID4 without having to wait for an explicit update to this document.

5.2. Response to Data Dropped and Slow Receiver

This is described in Section 5.2 of [RFC4342].

5.3. Packet Sizes

CCID 4 is intended for applications that use a fixed small segment size, or that vary their segment size in response to congestion.

The CCID 4 sender uses a segment size of 1460 bytes in the TCP throughput equation. This gives the CCID 4 sender roughly the same sending rate in bytes per second as a TFRC flow using 1460-byte segments but experiencing the same packet drop rate.

6. Acknowledgements

The acknowledgements are as specified in Section 6 of [RFC4342] with the exception of the Loss Interval lengths specified below.

6.1. Loss Interval Definition

The loss interval definition is as defined in Section 6.1 of [RFC4342], except as specified below. Section 6.1.1 of RFC 4342 specifies that for all loss intervals except the first one, the data length equals the sequence length minus the number of non-data packets the sender transmitted during the loss interval, with a minimum data length of one packet. For TFRC-SP, for short loss intervals of at most two round-trip times, the loss interval length is computed not as the data length of the loss interval, but instead as the data length divided by the number of dropped or marked data packets.

Section 5.4 of RFC 4342 described when to use the most recent loss interval in the calculation of the average loss interval. [RFC4828] adds to this procedure the restriction that the most recent loss interval is only used in the calculation of the average loss interval if the most recent loss interval is greater than two round-trip times. The pseudocode is given in Section 3 of [RFC4828].

6.2. Congestion Control on Acknowledgements

The congestion control on acknowledgements is as specified in Section 6.2 of [RFC4342].

6.3. Acknowledgements of Acknowledgements

Procedures for the acknowledgement of acknowledgements are as specified in Section 6.3 of [RFC4342].

6.4. Quiescence

The procedure for detecting that the sender has gone quiescent is as specified in Section 6.4 of [RFC4342].

7. Explicit Congestion Notification

Procedures for the use of Explicit Congestion Notification (ECN) are as specified in Section 7 of [RFC4342].

8. Options and Features

CCID 4 can make use of DCCP's Ack Vector, Timestamp, Timestamp Echo, and Elapsed Time options, and its Send Ack Vector and ECN Incapable features. CCID 4 also imports the currently defined CCID 3-specific options and features [RFC4342], augmented by the Dropped Packets options and features [CCID3-DP]. Each CCID 4-specific option and feature contains the same data as the corresponding CCID 3 option or feature, and is interpreted in the same way, except as specified elsewhere in this document.

Type	Option Length	Meaning	DCCP-Data?	Section Reference
128-191		Reserved		
192	6	Loss Event Rate	N	8.5
193	variable	Loss Intervals	N	8.6
194	6	Receive Rate	N	8.3
195	variable	Dropped Packets	N	8.7
196-255		Reserved		

Table 1: DCCP CCID 4 Options

The "DCCP-Data?" column indicates that all currently defined CCID 4-specific options MUST be ignored when they occur on DCCP-Data packets.

As with CCID 3, the following CCID-specific features are also defined.

Number	Meaning	Rec'n Rule	Initial Value	Req'd	Section Reference
128-191	Reserved				
192	Send Loss Event Rate	SP	0	N	8.4
193-194	Reserved				
195	Send Dropped Packets	SP	0	N	
196-255	Reserved				

Table 2: DCCP CCID 4 Feature Numbers

More information is available in Section 8 of [RFC4342] and in [CCID3-DP].

8.1. Window Counter Value

The use of the Window Counter Value in the DCCP generic header's CCVal field is as specified in Section 8.1 of [RFC4342]. In addition to their use described in CCID 3, the CCVal counters are used by the receiver in CCID 4 to determine when the length of a loss interval is at most two round-trip times. None of these procedures require the receiver to maintain an explicit estimate of the round-trip time. However, Section 8.1 of [RFC4342] gives a procedure that implementors may use if they wish to keep such an RTT estimate using CCVal.

8.2. Elapsed Time Options

The use of the Elapsed Time option is defined in Section 8.2 of [RFC4342].

8.3. Receive Rate Option

The Receive Rate option is as specified in Section 8.3 of [RFC4342].

8.4. Send Loss Event Rate Feature

The Send Loss Event Rate feature is as defined in Section 8.4 of [RFC4342].

See [RFC3448], Section 5 and [RFC4828], Section 4.4 for a normative calculation of the loss event rate. Section 4.4 of [RFC4828] modifies the calculation of the loss interval size for loss intervals of at most two round-trip times.

If the CCID 4 receiver is using the Loss Event Rate option, the receiver needs to be able to determine if a loss interval is short, of at most two round-trip times. The receiver can heuristically detect a short loss interval by using the Window Counter in arriving data packets. The sender increases the Window Counter by 1 every quarter of a round-trip time, with the caveat that the Window Counter is never increased by more than five, modulo 16, from one data packet to the next. Using the Window Counter to detect loss intervals of at most two round-trip times could result in some false positives, with some longer loss intervals incorrectly identified as short ones. For example, if the loss interval contained data packets with only two Window Counter values, say, k and k+5, then the receiver could not tell if the loss interval was at most two round-trip times long or not. Similarly, if the sender sent data packets with Window Counter values of 4, 8, 12, 0, 5, but the packets with Window Counter values of 8, 12, and 0 were lost in the network, then the receiver would only receive data packets with Window Counter values of 4 and 5, and would incorrectly infer that the loss interval was at most two round-trip times.

8.5. Loss Event Rate Option

The Loss Event Rate option is as specified in Section 8.5 of [RFC4342].

See [RFC3448] (Section 5) and [RFC4828] for a normative calculation of the loss event rate.

8.6. Loss Intervals Option

The Loss Intervals option is as specified in Section 8.6 of [RFC4342].

8.7. Dropped Packets Option

The Dropped Packets option is as specified in [CCID3-DP]. CCID 4 receivers MUST always include Dropped Packets options on their feedback packets, regardless of the value of the Send Dropped Packets feature. If, nevertheless, a feedback packet does not include a relevant Dropped Packets option, a CCID 4 sender MUST act as if the relevant loss intervals' Drop Counts equal the corresponding Loss Lengths, as specified in [CCID3-DP].

8.8. Send Dropped Packets Feature

The Send Dropped Packets feature is as specified in [CCID3-DP].

9. Verifying Congestion Control Compliance With ECN

Verifying congestion control compliance with ECN is as discussed in Section 9 of [RFC4342].

9.1. Verifying the ECN Nonce Echo

Procedures for verifying the ECN Nonce Echo are as specified in Section 9.1 of [RFC4342].

9.2. Verifying the Reported Loss Intervals and Loss Event Rate

Section 9.2 of [RFC4342] discusses the sender's possible verification of loss intervals and loss event rate information reported by the receiver.

10. Implementation Issues

10.1. Timestamp Usage

The use of the Timestamp option is as discussed in Section 10.1 of [RFC4342].

10.2. Determining Loss Events at the Receiver

The use of the window counter by the receiver to determine if multiple lost packets belong to the same loss event is as described in Section 10.2 of [RFC4342].

10.3. Sending Feedback Packets

The procedure for sending feedback packets is as described in Section 10.3 of [RFC4342].

11. Security Considerations

Security considerations include those discussed in Section 11 of [RFC4342]. There are no new security considerations introduced by CCID 4.

12. IANA Considerations

This specification defines the value 4 in the DCCP CCID namespace managed by IANA.

CCID 4 also uses three sets of numbers whose values should be allocated by IANA, namely CCID 4-specific Reset Codes, option types, and feature numbers. This document makes no particular allocations from the Reset Code range, except for experimental and testing use [RFC3692]. We refer to the Standards Action policy outlined in [RFC2434].

12.1. Reset Codes

Each entry in the DCCP CCID 4 Reset Code registry contains a CCID 4-specific Reset Code, which is a number in the range 128-255; a short description of the Reset Code; and a reference to the RFC defining the Reset Code. Reset Codes 184-190 and 248-254 are permanently reserved for experimental and testing use. The remaining Reset Codes -- 128-183, 191-247, and 255 -- are currently reserved, and should be allocated with the Standards Action policy, which requires IESG review and approval and standards-track IETF RFC publication.

12.2. Option Types

Each entry in the DCCP CCID 4 option type registry contains a CCID 4-specific option type, which is a number in the range 128-255; the name of the option, such as "Loss Intervals"; and a reference to the RFC defining the option type. The registry is initially populated using the values in Table 1, in Section 8. This document allocates option types 192-195, and option types 184-190 and 248-254 are permanently reserved for experimental and testing use. The remaining option types -- 128-183, 191, 196-247, and 255 -- are currently reserved, and should be allocated with the Standards Action policy, which requires IESG review and approval and standards-track IETF RFC publication.

12.3. Feature Numbers

Each entry in the DCCP CCID 4 feature number registry contains a CCID 4-specific feature number, which is a number in the range 128-255; the name of the feature, such as "Send Loss Event Rate"; and a reference to the RFC defining the feature number. The registry is initially populated using the values in Table 2, in Section 8. This document allocates feature numbers 192 and 195, and feature numbers 184-190 and 248-254 are permanently reserved for experimental and testing use. The remaining feature numbers -- 128-183, 191, 193-194, 196-247, and 255 -- are currently reserved, and should be allocated with the Standards Action policy, which requires IESG review and approval and standards-track IETF RFC publication.

13. Thanks

Normative References

[RFC2119] S. Bradner. Key Words For Use in RFCs to Indicate Requirement Levels. RFC 2119.

[RFC2434] T. Narten and H. Alvestrand. Guidelines for Writing an IANA Considerations Section in RFCs. RFC 2434.

[RFC3448] M. Handley, S. Floyd, J. Padhye, and J. Widmer, TCP Friendly Rate Control (TFRC): Protocol Specification, RFC 3448, Proposed Standard, January 2003.

[RFC3692] T. Narten. Assigning Experimental and Testing Numbers Considered Useful. RFC 3692.

[RFC4340] Kohler, E., Handley, M., and S. Floyd. Datagram Congestion Control Protocol (DCCP), RFC 4340, March 2006.

[RFC4342] Floyd, S., Kohler, E., and J. Padhye. Profile for Datagram Congestion Control Protocol (DCCP) Congestion Control ID 3: TCP-Friendly Rate Control (TFRC), RFC 4342, March 2006.

Informative References

[RFC4828] S. Floyd and E. Kohler. TCP Friendly Rate Control (TFRC): the Small-Packet (SP) Variant. RFC 4828, April 2007.

[CCID3-DP] Kohler, E., Datagram Congestion Control Protocol (DCCP) Congestion Control ID 3 Dropped Packets Option, Internet-draft draft-kohler-dccp-ccid3-drops-01.txt, work-in-progress, June 2007. URL "http://www.read.cs.ucla.edu/dccp/".

[KFS07] Kohler, E., S. Floyd, and A. Sathiaseelan, Faster Restart for TCP Friendly Rate Control (TFRC), Internet-draft draft-ietf-dccp-tfrc-faster-restart-02.txt, work-in-progress, March 2007.

Authors' Addresses

Sally Floyd <floyd@icir.org>
ICSI Center for Internet Research
1947 Center Street, Suite 600

Berkeley, CA 94704
USA

Eddie Kohler <kohler@cs.ucla.edu>
4531C Boelter Hall
UCLA Computer Science Department
Los Angeles, CA 90095
USA

Full Copyright Statement

Intellectual Property