

ConEx Working Group  
Internet-Draft  
Intended status: Informational  
Expires: January 13, 2014

D. Wagner  
M. Kuehlewind  
University of Stuttgart  
July 12, 2013

**ConEx Crediting and Auditing**  
**draft-wagner-conex-credit-00**

Abstract

Congestion Exposure (ConEx) is a mechanism by which senders inform the network about the congestion encountered by previous packets on the same flow. In order to make ConEx information useful, reliable auditing is necessary to provide a strong incentive to declare ConEx information honestly. However, there is always a delay between congestion events and the respective ConEx signal at the audit. To avoid state and complex Round-Trip Time estimations at the audit, in [[draft-ietf-conex-abstract-mech](#)] it is proposed to use credit signals sent in advance to cover potential congestion in the next feedback delay duration. Unfortunately, introducing credit does not provide incentives to honestly report congestion. This document lists potential issues regarding the proposed crediting and discusses potential solutions approaches to interpret and handle credits at the audit.

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 13, 2014.

## Copyright Notice

Copyright (c) 2013 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 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

<a href="#">1.</a>	Introduction . . . . .	<a href="#">2</a>
<a href="#">1.1.</a>	Definitions . . . . .	<a href="#">3</a>
<a href="#">2.</a>	Open issues . . . . .	<a href="#">3</a>
<a href="#">2.1.</a>	No incentives to conceal congestion by sending credits .	<a href="#">3</a>
<a href="#">2.2.</a>	Handling Loss of ConEx-marked Packets . . . . .	<a href="#">4</a>
<a href="#">2.3.</a>	Independence from Audit State . . . . .	<a href="#">4</a>
<a href="#">2.4.</a>	Assumption on Distance between Congestion Events . . . .	<a href="#">4</a>
<a href="#">3.</a>	Potential Credit Definition and Auditing Approaches . . . .	<a href="#">4</a>
<a href="#">3.1.</a>	Basic Audit Reference Implementation . . . . .	<a href="#">5</a>
<a href="#">3.2.</a>	Credit As Congestion Substitute . . . . .	<a href="#">6</a>
<a href="#">3.3.</a>	Credit As Congestion Surcharge . . . . .	<a href="#">6</a>
<a href="#">3.3.1.</a>	BDP-Scaled Surcharge . . . . .	<a href="#">7</a>
<a href="#">3.4.</a>	Credit As Short-Lived Congestion Risk Compensation . . .	<a href="#">8</a>
<a href="#">4.</a>	Discussion . . . . .	<a href="#">8</a>
<a href="#">5.</a>	Security Considerations . . . . .	<a href="#">10</a>
<a href="#">6.</a>	IANA Considerations . . . . .	<a href="#">10</a>
<a href="#">7.</a>	References . . . . .	<a href="#">10</a>
<a href="#">7.1.</a>	Normative References . . . . .	<a href="#">10</a>
<a href="#">7.2.</a>	Informative References . . . . .	<a href="#">10</a>
	Authors' Addresses . . . . .	<a href="#">10</a>

## [1.](#) Introduction

In order to make ConEx information useful, reliable auditing is necessary to provide a strong incentive to declare ConEx information honestly. However, there is always a delay between congestion events and the respective ConEx signal at the audit. To avoid state and complex Round-Trip Time (RTT) estimations at the audit, in [\[draft-ietf-conex-abstract-mech\]](#) it is proposed to use credit signals sent in advance to cover potential congestion in the next feedback delay duration.



The ConEx signal is based on loss or Explicit Congestion Notification (ECN) marks [[RFC3168](#)] as a congestion indication. Following [[draft-ietf-conex-abstract-mech](#)] ([Section 4.4](#)), ConEx signaling has to encode ConEx capability, Re-Echo-Loss (L), Re-Echo-ECN (E) and credit (C). The C (credit) signal is used to build up credits at the audit in advance of congestion.

[[draft-ietf-conex-abstract-mech](#)] (currently) only states that the "transport signals sufficient credit in advance to cover congestion expected during its feedback delay". Unfortunately, introducing crediting can also provide incentives to not report congestion but send credits instead. While ConEx feedback should be provided timely and reflect the actual congestion on a path, credits should be sent at any time before the congestion event and need to cover at least the congestion 'costs'. Thus crediting might motivate to send credits instead of ConEx congestion marks (L or E). Besides this central issue, the exact meaning of these credits and their handling at the audit and therefore their usage at the sender is also left open up to now. This document presents and discusses potential concepts for crediting and auditing.

### **[1.1.](#) Definitions**

Congestion occurrence

The occurrence of a signal congestion signal, which today corresponds to a packet loss or ECN-CE mark.

Congestion event

One or more congestion occurrences that happen within one RTT and therefore are perceived as one congestion event by today's congestion control algorithms.

## **[2.](#) Open issues**

A solid concept how to interpret and handle credit needs to address the issues listed in the following.

### **[2.1.](#) No incentives to conceal congestion by sending credits**

The goal of ConEx is to reveal path congestion honestly by sending the right amount of L or E marks timely after a congestion occurrence. The use of credit marks should not motivate to endanger this goal. If credits are treated equally by an audit device, there is no incentive to send additional ConEx (L or E) marks if already a sufficient large number of credits have been sent in advance of the congestion event. This is a major and fundamental issue of the credit concept in general.



## **2.2. Handling Loss of ConEx-marked Packets**

Due to the complexity of detecting loss of ConEx information and the time dependency of this information, ConEx marks should not be retransmitted. Thus ConEx marks of lost packets will never be seen at an audit. Generally, two entities could be responsible to care for this issue: the sender or the audit. To keep the audit simple, it would be preferable having this task performed by the ConEx sender. As retransmitting is not an option, the sender can only send credits as a substitute instead. It is not clear if false positives by the audit due to lost markings can be avoided by this. As, without any knowledge about lost markings and depending on the definition of credits, it is hard for the sender to determine the current number of valid credits perceived by the audit. The other option is having the audit estimating loss of ConEx marks without requiring the sender to replace them by credit.

## **2.3. Independence from Audit State**

An audit may be started with zero state information on existing flows. As credits will have been sent in advance of congestion events, it is possible that no valid credit state is available at the audit when a congestion event occurs. The credit definition and the respective implications for the audit should also consider handling this situation.

The concept of crediting should consider that existing flows may be rerouted, so that a flow may pass through different audits over time. If rerouting and thus a change of the audit happens, it is not required to have no impact at all, but starvation and finally shutting down of the connection should be avoided.

## **2.4. Assumption on Distance between Congestion Events**

Today's congestion control algorithms are designed for loss-based congestion feedback and therefore aim to get congestion feedback rather rarely, i.e. for typical BDPs there are losses only every some RTTs. Thus it could be assumed that ConEx marks will be received at the audit before the next congestion event occurs.

Nevertheless, if the congestion feedback is not based on loss, but e.g. on ECN, a more frequent signal could provide more precise information on the current congestion level and therefore allow a finer reaction of the congestion control. Since this would be a desirable situation, we should not base the definition of ConEx credit on assumptions about the distance between congestion events.

## **3. Potential Credit Definition and Auditing Approaches**



### **3.1. Basic Audit Reference Implementation**

The objective of an audit is to verify correct usage of ConEx signals and to penalize cheaters. To verify, that the congestion reported using the ConEx mechanism matches the congestion actually observed by the receiver, it has to monitor incoming and outgoing traffic close to the receiver (beyond any point of congestion). For ConEx-capable connections there are 5 types of events of interest:

- o ECN-CE (ECN codepoint set)
- o Loss (to be detected by the audit)
- o Re-Echo-ECN (ConEx signal E)
- o Re-Echo-Loss (ConEx signal L)
- o Credit (ConEx signal C)

A simple implementation could keep one counter per type of event. For a well-behaving sender, for each loss or ECN-CE signal the respective ConEx signal will follow just one RTT later, balancing both counters. Therefore, often only the balance of the counters for loss or ECN and the respective ConEx signal matter, e.g. Re-Echo-Loss - Loss. For a well behaving sender and disregarding loss of ConEx marks, at least one balance counter will be negative right after the congestion event but will recover to zero (balanced state) after one RTT (for connections with typical BDPs and today's congestion controls). Even if congestion occurs more frequently due to a fine grained congestion notification scheme, the balance counters would be negative (at about the average number of congestion occurrences per RTT) but not decline over time. Nevertheless, since ConEx marked packets can get lost and will not be re-sent, these balance counters may decline over time. Thus the balance counters can get negative or zero, but should never get positive (even when more ConEx marks are received than congestion signals are observed).

An audit could also target to estimate loss of ConEx marked packets based on an estimate of the connection's packet loss rate. It then could decide how negative the balance counters are allowed to get: if the audit additionally counts all packets of a connection, it can easily estimate the packet loss rate. It can compare the relations  $\text{Lost\_Packets}/\text{All\_Packets}$  and  $(\text{Lost\_Packets} - \text{Re-Echo-Loss})/\text{Lost\_Packets}$  and  $(\text{ECN-CE} - \text{Re-Echo-ECN})/\text{ECN-CE}$  respectively. Over time, these relations should converge, assuming that ConEx marked packets are hit with the same probability as other packets(!). Therefore, the audit may decide to penalize a flow if less ConEx marks are received than expected on that estimation.





If the audit detects misbehavior or cheating e.g. due to permanent negative counters, the audit shall penalize the connection. The only actually possible penalty would be dropping packets (with a certain probability). The actual drop rate should provide a tangible disadvantage to the sender but should not render the connection unusable. E.g. it could depend on how negative the counter is. This could motivate the sender to just open a new connection as replacement. Moreover, false positives probably can't be avoided completely. The actual definition of penalties requires further research.

In the following different proposals for crediting are presented and the handling in the audit based on this general principle is discussed.

### **3.2. Credit As Congestion Substitute**

Credit marks may be understood by the audit function as an equal substitute for congestion marks. This means, that an audit device will count and keep credit marks the same way as congestion marks. Usually, as credits should cover the risk of causing congestion, a large number of credits will be sent during Slow Start phase of TCP congestion control (as the sending rate is increased quite aggressively), e.g. at the start-up of a new TCP flow. Later the sending rate is adjusted more slowly, thus usually no further credits are needed, if the initially send credits remain valid for the lifetime of a flow. During the connection the number of lost or ECN-(CE)-marked packets indicating congestion should be balanced by ConEx L or E marks. So at to the end of a flow's lifetime, there is an amount of credits "sitting in the network" that is finally discarded.

Audit implementation:

An audit maintains three counters per flow: one for credit, one for loss balance and one for ECN balance (see [Section 3.1](#)). Whenever a marked packet is seen, the respective counter is increased. When a loss or ECN-(CE)-marked packet is observed, the respective counter is decreased.

If the sum of the counters is negative, the flow is penalized.

### **3.3. Credit As Congestion Surcharge**



Another option is to interpret credit as compensation for late arrival of congestion marks, or surcharge on (following) congestion marks. This would basically mean that the sender pays twice for congestion: first in advance by sending credit marks, and one RTT after the congestion event by sending the respective number of ConEx L or E marked packets.

Audit implementation:

An audit maintains three counters per flow, one for credit, one for loss events and one for ECN events. Whenever a marked packet is seen, the respective counter is increased. When a loss or ECN-(CE)-marked packet is observed, the respective counter and also the credit counter is decreased.

If the credit counter is negative, the flow is penalized.

### **3.3.1. BDP-Scaled Surcharge**

For the sake of completeness and mainly motivated by the audit implementation below we introduce a variation of the Congestion Surcharge definition. In this approach the charge that needs to be provided by credits per congestion event scales with the BDP (as the maximum congestion risk) and additionally the longest delay between two congestion occurrences within the congestion event. More precisely, the proposed auditing scheme requires the sender to react on a congestion event by sending credits until there was one RTT without congestion events. This means that the sender pays for a single congestion occurrence at least one RTT of credit marks.. If several congestion signals occur within one RTT, the sender sends credit marks until one RTT after the last signal. Thus the credit cost of two congestion occurrences within one RTT varies from BDP to  $2 \cdot \text{BDP} - 1$ .

Audit implementation:

An audit maintains three counters per flow, one for credit, one balance counter for loss events and one for ECN events. Whenever a L- or C-marked packet is seen, the respective balance is increased. When a loss or ECN-(CE)-marked packet is observed, the respective counter and also the credit counter is decreased. For any packet seen while the balance is negative, the credit counter is decreased. If the credit counter is negative, the flow is penalized.



### **3.4. Credit As Short-Lived Congestion Risk Compensation**

This approach tries to provide an incentive for the sender to send correct congestion feedback and not sending credits instead. One basic property of the approaches presented above is the infinite validity of credit. The expiry of credits could depend on the RTT or other channel characteristics, but we deem the reasoning in [\[draft-ietf-conex-abstract-mech\]](#) valid, so the audit should not need to estimate channel characteristics per flow. However, credits could also expire after a fixed time, e.g. 300 seconds. This expiration time must be fixed and known to all senders so that they can replace vanishing credit in time.

This timeout must at least be large enough to cover the length of a feedback cycle of TCP congestion control. A feedback cycle is the time between to congestion events. Today's congestion control algorithms result in quite low loss rate and thus feedback rate for large BDPs and therefore rather long feedback cycles. Nevertheless, even for NewReno [\[RFC5681\]](#) being used for a single connection on a 1Gbps link with 100ms RTT and 1500Byte segment size, a feedback cycle is about  $(RTT * BDP) / 2 = 41.6$  seconds. Since even occasional packet errors also discourage from using congestion controls with that low probing frequency, we deem 300 seconds a safe proposal for the expiration timeout.

Audit implementation:

An audit maintains three counters per flow, one for credit, one for loss events and one for ECN events. Whenever a marked packet is seen, the respective counter is increased. When a loss or CE-event is detected, the respective counter is decreased and the credit counter is decreased additionally. Incoming credits set a timer upon which's timeout the credit counter is decreased.

Note: Since credit expiring a little later than expected does not harm the overall function of the audit, it might aggregate expiration timeouts, e.g. for 10 seconds so for each flow only 31 bin counters would be needed.

If the credit counter is negative, the flow is penalized.

## **4. Discussion**

This section shall provide an initial list of arguments as the basis for further discussions.

For all proposals, honest congestion marks can be replaced with credit marks without limitation. Moreover, for the Substitute approach the sender has to the end of an connection no motivation to provide any ConEx signals because he can assume that the balance at



the audit is far in his favor. This aspect may concern a significant time, depending on which congestion rate the used congestion control algorithm implements.

Introducing a limitation for sending credit could limit the impact of this fact but first a good definition of credit limits is not obvious and second it would not work for short flows.

Any sender-based approach to handle loss of ConEx-marked packets requires to allow replacing ConEx L and E signals by credit to a some extend. This is basically contradicting to hindering concealing of congestion by using credits. Note: the same calculation as proposed for loss handling at the audit (see [Section 3.1](#)) can also be performed by the sender, allowing him to send compensational credit in advance. Another advantage of sender-based loss handling is that the sender may use that mechanism to compensate false positives of the audit. Of course this a drawback at the same time, as it opens for abuse.

A main advantage of handling loss at the audit is that no compensation by credit is necessary, so issue#1 can be avoided. The main disadvantage is that the outlined mechanism only works for longer flows since the statistical deviation of the observed loss rate needs be acceptable small. It must also be noted, that the loss probability changes over time during a flow's life time. For today's congestion control algorithms, ConEx marked packets will be sent one RTT after the congestion event when the sender reduced its sending rate, thus the loss rate of ConEx marked packets should be smaller than the total average loss rate. So more complex estimators might be necessary, further increasing the audit complexity.

If ConEx loss is handled by the sender, re-routing or restarting audits can be expected to be handled in a similar but this definitely requires further research.

The BDP-Scaled Surcharge-approach has several properties which we deem undesirable: although the RTT is out of control of the end user, for this definition he has to pay more for connections with longer RTTs. Moreover, the distribution of congestion occurrences affects the credit cost of one congestion event significantly.

Approach [Section 3.4](#) with limited life time of credit at least solves the issue of large amounts of credits being available at the audit to the end of a connection. Yet the issue of cheating by sending credit instead of congestion marks remains unsolved. Maybe the proposed credit definition could be used with a modified audit algorithm limiting the decrease of the balance counters.





## **5. Security Considerations**

This document has no security considerations.

## **6. IANA Considerations**

This document has no IANA considerations.

## **7. References**

### **7.1. Normative References**

[[draft-ietf-conex-abstract-mech](#)]

Mathis, M. and B. Briscoe, "Congestion Exposure (ConEx) Concepts and Abstract Mechanism", [draft-ietf-conex-abstract-mech-06](#) (work in progress), October 2012.

[[draft-ietf-conex-destopt](#)]

Krishnan, S., Kuehlewind, M., and C. Ucendo, "IPv6 Destination Option for ConEx", [draft-ietf-conex-destopt-04](#) (work in progress), March 2013.

### **7.2. Informative References**

[RFC5681] Allman, M., Paxson, V., and E. Blanton, "TCP Congestion Control", [RFC 5681](#), September 2009.

[RFC3168] Ramakrishnan, K., Floyd, S., and D. Black, "The Addition of Explicit Congestion Notification (ECN) to IP", [RFC 3168](#), September 2001.

## Authors' Addresses

David Wagner  
University of Stuttgart  
Pfaffenwaldring 47  
70569 Stuttgart  
Germany

Email: [david.wagner@ikr.uni-stuttgart.de](mailto:david.wagner@ikr.uni-stuttgart.de)



Mirja Kuehlewind  
University of Stuttgart  
Pfaffenwaldring 47  
70569 Stuttgart  
Germany

Email: [mirja.kuehlewind@ikr.uni-stuttgart.de](mailto:mirja.kuehlewind@ikr.uni-stuttgart.de)