ConEx
Internet-Draft
Intended status: Informational
Expires: April 26, 2012

B. Briscoe
BT

October 24, 2011

Initial Congestion Exposure (ConEx) Deployment Examples
draft-briscoe-conex-initial-deploy-00

Abstract

This document gives examples of how ConEx deployment might get started, focusing on unilateral deployment by a single network.

Status of This Memo

Copyright Notice

Table of Contents

1. Introduction

This document gives examples of how ConEx deployment might get started, focusing on unilateral deployment by a single network.

2. Recap: Incremental Deployment Features of the ConEx Protocol

The ConEx mechanism document [ConEx-Abstract-Mech] goes to great lengths to design for incremental deployment in all the respects below. It should be referred to for precise details on each of these points:

- The ConEx mechanism is essentially a change to the source, in order to re-insert congestion feedback into the network.

- Source-host-only deployment is possible without any negotiation required, and individual transport protocol implementations within a source host can be updated separately.

- Receiver modification may optionally improve ConEx for some transport protocols with feedback limitations (TCP being the main example), but it is not a necessity

- Proxies for the source and/or receiver are feasible (though not necessarily straightforward)

- Queues and network forwarding do not require any modification for ConEx.

- ECN is not required in the network for ConEx. If some network nodes support ECN, it can be used by ConEx.

- ECN is not required at the receiver for ConEx. The sender should nonetheless attempt to negotiate ECN-usage with the receiver, given some aspects of ConEx work better the more ECN is deployed, particularly auditing and border measurement.

- Given ConEx exposes information for IP-layer policy devices to use, the design does not preclude possible innovative uses of ConEx information by other IP-layer devices, e.g. forwarding itself

- Packets indicate whether or not they support ConEx.

3. ConEx Components

3.1. Recap of Basic ConEx Components

[ConEx-Abstract-Mech] introduces the following components:

- The ConEx Wire Protocol
- Forwarding devices (unmodified)
- Sender (modified for ConEx)
- Receiver (optionally modified)
- Audit
- Policy Devices:
    - Rest-of-Path Congestion Monitoring Devices
    - Congestion Policers

[ConEx-Abstract-Mech] should be referred to for definitions of each of these components and further explanation.

3.2. Per-Network Deployment Concepts

Network deployment-related definitions:

Internet Ingress: The first IP node a packet traverses that is outside the source’s own network. In a domestic network that will be the first node downstream from the home access equipment. In an enterprise network this is the provider edge router.

Internet Egress: The last IP node a packet traverses before reaching the receiver’s network.

ConEx-Enabled Network: A network whose edge nodes implement ConEx policy functions.

Each network can unilaterally choose to use any ConEx information given by those sources using ConEx, independently of whether other networks use it.

Typically, a network will use ConEx information by deploying a policy function at the ingress edge of its network to monitor arriving traffic and to act in some way on the congestion information in those packets that are ConEx-enabled. Actions might include policing, altering the class of service, or re-routing. Alternatively, less direct actions via a management system might include triggering capacity upgrades, triggering penalty clauses in contracts or levying charges between networks based on ConEx measurements.

Typically, a network using ConEx info will deploy a ConEx policy function near the ingress edge and a ConEx audit function near the egress edge. The segment of the path between a ConEx policy function and a ConEx audit function can be considered to be a ConEx-protected segment of the path. Assuming a network covers all its ingresses and egresses with policy functions and audit functions respectively, the network within this ring will be a ConEx-protected network.

Of course, because each edge device usually serves as both an ingress and an egress, the two functions are both likely to be present in each edge device.

4. Example Initial Deployment Arrangements

In all the deployment scenarios below, we assume that deployment starts with some data sources being modified with ConEx code. The rationale for this is that the developer of a scavenger transport protocol like LEDBAT has a strong incentive to tell the network how little congestion it is causing despite sending large volumes of data. In this case the developer makes the first move expecting it will prompt at least some networks to move in response--so that they use the ConEx information to reward users of the scavenger protocol.

4.1. Single Receiving Network Scenario

The name ’Receiving Network’ for this scenario merely emphasises that most data is arriving from connected networks and data centres and being consumed by residential customers on this access network. Some data is of course also travelling in the other direction.

```
                                          DSLAMs __
                                              /|/     ,-.Home-a
                                        __/__| |-----(   )
                    ,-----.            /  \  | |---   ‘-’
       ,---.       /       \  ,------P/       \|\__
      /     \     ’  Core   ’/| BRAS |          __
     ( Peer  )-->-|P        | ’------’       /|/
      \     /     |         |          _____| |---
       ’---’      ’         ’\,------./     | |---
                   \ M     /  |BRAS  |       \|\__
                    ‘-----’   ’------A\          __
                     |          P|     \      /|/
                    /|\         /|\     \__\_| |---   ,-.
                   ,---.        ,---.      / | |-----(   )
                  /Data \      /     \        \|\__   ‘-’Home-b
                 ( Centre)    (  CDN  )
                  \     /      \     /  Access Network
                   ’---‘        ’---‘  <-------------->
```

P=Congestion-Policer; M=Congestion-Monitor; A=Audit function

Figure 1: Single Receiving Network Scenario

Figure Figure 1 is an attempt to show the salient features of a ConEx deployment in a typical broadband access provider’s network (within the constraints of ASCII art). Broadband remote access servers (BRASs) control access to the core network from the access network and vice versa. Home networks (and small businesses) connect to the access network, but only two are shown.

In this diagram, all data is travelling towards the access network of Home-b, from the Peer network, the Data centre, the CDN and Home-a. Data actually travels in both directions on all links, but only one direction is shown.

The data centre, core and access network are all run by the same network operator, but each is the responsibility of a different department with internal accounting between them. The content distribution network (CDN) is operated by a third party CDN provider, and of course the peer network is also operated by a third party.

This operator of the data centre, core and access network is the only one in the diagram to have deployed ConEx monitoring and policy devices at the edges of its network. However, it has not enabled ECN on any of its network elements and neither has any other network in the diagram. The operator has deployed a congestion policing function (P) on the provider-edge router where the peer attaches to its core, on the BRAS where the CDN attaches and on the other BRAS where each of the residential customers like Home-a attach. On the provider-edge router where the data centre attaches it has deployed a congestion monitoring function (M). Each of these policing and monitoring functions handles the aggregate of all traffic traversing it, for all destinations.

The operator has deployed an audit function on each logical output port of the BRAS for each end-customer site like Home-b. The Audit function handles the aggregate of all traffic for that end-customer from all sources. For traffic in the opposite direction (e.g. from Home-b to Home-a, there would be equivalent policing (P) and audit (A) functions in the converse locations to those shown.

Some content sources in the CDN and in the data centre are using the ConEx protocol, but others are not. There is a similar situation for hosts attached to the Peer network and hosts in home networks like Home-a: some are sending ConEx packets at least for bulk data transports, while others are not.

4.1.1. ConEx Functions in the Single Receiving Network Scenario

Within the BRAS there are logical ports that model the rate of each access line from the DSLAM to each home network [TR-059]. They are fed by a shared queue that models the rate of the downstream link from the BRAS to the DSLAM (sometimes called the backhaul network). If there is congestion anywhere in the set of networks in Figure Figure 1 it is nearly always:

o either self-congestion in the queues into the logical ports representing the access lines

o or shared congestion in the shared queue on the BRAS that feeds them.

Any ConEx sources sending data through this BRAS will receive feedback about these losses from the destination and re-insert it as ConEx markings into the data. Figure 2 shows an example plot of the loss levels that might be seen at different monitoring points along a path between the data centre and home-b, for instance. The top half of the figure shows the loss probability within the BRAS consists of 0.1% at the shared queue and 0.2% self-congestion in the logical output port that models the access line, making 0.3% in total. This upper diagram also shows whole path congestion as signalled by the ConEx sender, which remains unchanged along the whole path at 0.3%.

The lower half of the figure shows (downstream congestion) = (whole path) - (upstream congestion). Upstream congestion can only be monitored locally where the loss actually happens (within the BRAS output queues). Nonetheless, given there is rarely loss anywhere else but within the BRAS, this limitation is not significant in this scenario. The lower half of the figure also shows the location of the policing and audit functions. Policing anywhere within or upstream ofthe BRAS will be based on the downstream congestion level of 0.3%. While Auditing within the BRAS but after all the queues can check that the whole path congestion signalled by ConEx is no less than the loss levels experienced within the BRAS itself.

```
      Data centre-->|<--core-->|<------BRAS--------->|<--Home--
                               |                     |
    ^loss                      |<-Shared->|<-Access->|
    |probability                 backhaul
    |
0.3%|- - - - - - - - - - - - - - - - - - - - +-----------------
    |      whole path congestion            |
    |                                       |
    |                                       |upstream
0.1%|                             +---------+congestion
    |                             |
   -O=============================+----------------------------->
                                                monitoring point
    ^loss
    |probability   Policing                   Audit
    |                |                           |
    |                V                           |
0.3%|----------------O-------------+             |
    |                              |downstream   |
0.2%|                              +---------+    |
    |                              congestion|    |
    |                                        |    |
    |                                        |    V
   -O----------------------------------------+====O============-->
                                                monitoring point
```

Figure 2: Example plot of loss levels along a path

4.1.2. Incentives to Unilaterally Deploy ConEx in a Receiving Network

Even a sending application that is modified to use ConEx can choose whether to send ConEx or Not-ConEx packets. Nonethelss, ConEx packets bring information to a policer about congestion expected on the rest of the path beyond the policer. Not-ConEx packets bring no such information. Therefore a network that has deployed ConEx policers will tend to rate-limit not-ConEx packets conservatively in order to manage the unknown risk of congestion. In contrast, a network doesn’t normally need to rate-limit ConEx-enabled packets unless they reveal a persistently high contribution to congestion. This natural tendency for networks to favour senders that provide ConEx information encourages senders to choose to use the ConEx protocol whenever they can.

{ToDo: complete this section}

4.2. Mobile Network Scenario

Placeholder for summary of the scenario in a mobile network described in [conex-mobile]

In mobile networks, both mobile terminals and mobile network equipment are standardised by the 3GPP. If the 3GPP were to adopt the ConEx protocol, it might mandate ConEx implementation for compliant equipment.

{ToDo: Describe how a central traffic management box can arrange to remotely view upstream congestion as it would be seen from the interface with the mobile terminal.}

4.3. Scenario Internal to a Multi-Tenant Data Centre

A number of companies offer hosting of virtual machines on their data centre infrastructure--so-called infrastructure as a service (IaaS). A set amount of processing power, memory, storage and network are offered. Although processing power, memory and storage are relatively simple to allocate on the 'pay as you go' basis that has become common, the network is less easy to allocate given it is a naturally distributed system.

{ToDo: Complete this section.}

5. Security Considerations

6. IANA Considerations

This document does not require actions by IANA.

7. Conclusions

{ToDo}

8. Acknowledgments

9. Informative References

[ConEx-Abstract-Mech] Mathis, M. and B. Briscoe, "Congestion Exposure (ConEx) Concepts and Abstract Mechanism", draft-ietf-conex-abstract-mech-02 (work in progress), July 2011.

[Seawall] Shieh, A., Kandula, S., Greenberg, A., and C. Kim, "Seawall: Performance Isolation in Cloud Datacenter Networks", Proc 2nd USENIX Workshop on Hot Topics in Cloud Computing , June 2010, <http://research.microsoft.com/en-us/projects/seawall/>.

[TR-059] Anschutz, T., Ed., "DSL Forum Technical Report TR-059: Requirements for the Support of QoS-Enabled IP Services", September 2003.

[conex-mobile] Kutscher, D., Mir, F., Winter, R., Krishnan, S., and Y. Zhang, "Mobile Communication Congestion Exposure Scenario", draft-kutscher-conex-mobile-00 (work in progress), March 2011.

Author’s Address

Bob Briscoe
BT
B54/77, Adastral Park
Martlesham Heath
Ipswich IP5 3RE
UK

Phone: +44 1473 645196
EMail: bob.briscoe@bt.com
URI: http://bobbriscoe.net/

ConEx
Internet-Draft
Intended status: Informational
Expires: April 26, 2012

B. Briscoe
BT

October 24, 2011

Initial Congestion Exposure (ConEx) Deployment Examples
draft-briscoe-conex-initial-deploy-00

Abstract

This document gives examples of how ConEx deployment might get started, focusing on unilateral deployment by a single network.

Status of This Memo

Copyright Notice

Table of Contents

1. Introduction

This document gives examples of how ConEx deployment might get started, focusing on unilateral deployment by a single network.

2. Recap: Incremental Deployment Features of the ConEx Protocol

The ConEx mechanism document [ConEx-Abstract-Mech] goes to great lengths to design for incremental deployment in all the respects below. It should be referred to for precise details on each of these points:

- The ConEx mechanism is essentially a change to the source, in order to re-insert congestion feedback into the network.

- Source-host-only deployment is possible without any negotiation required, and individual transport protocol implementations within a source host can be updated separately.

- Receiver modification may optionally improve ConEx for some transport protocols with feedback limitations (TCP being the main example), but it is not a necessity

- Proxies for the source and/or receiver are feasible (though not necessarily straightforward)

- Queues and network forwarding do not require any modification for ConEx.

- ECN is not required in the network for ConEx. If some network nodes support ECN, it can be used by ConEx.

- ECN is not required at the receiver for ConEx. The sender should nonetheless attempt to negotiate ECN-usage with the receiver, given some aspects of ConEx work better the more ECN is deployed, particularly auditing and border measurement.

- Given ConEx exposes information for IP-layer policy devices to use, the design does not preclude possible innovative uses of ConEx information by other IP-layer devices, e.g. forwarding itself

- Packets indicate whether or not they support ConEx.

3. ConEx Components

3.1. Recap of Basic ConEx Components

[ConEx-Abstract-Mech] introduces the following components:

- The ConEx Wire Protocol
- Forwarding devices (unmodified)
- Sender (modified for ConEx)
- Receiver (optionally modified)
- Audit
- Policy Devices:
  * Rest-of-Path Congestion Monitoring Devices
  * Congestion Policers

[ConEx-Abstract-Mech] should be referred to for definitions of each of these components and further explanation.

3.2. Per-Network Deployment Concepts

Network deployment-related definitions:

Internet Ingress: The first IP node a packet traverses that is outside the source’s own network. In a domestic network that will be the first node downstream from the home access equipment. In an enterprise network this is the provider edge router.

Internet Egress: The last IP node a packet traverses before reaching the receiver’s network.

ConEx-Enabled Network: A network whose edge nodes implement ConEx policy functions.

Each network can unilaterally choose to use any ConEx information given by those sources using ConEx, independently of whether other networks use it.

Typically, a network will use ConEx information by deploying a policy function at the ingress edge of its network to monitor arriving traffic and to act in some way on the congestion information in those packets that are ConEx-enabled. Actions might include policing, altering the class of service, or re-routing. Alternatively, less direct actions via a management system might include triggering capacity upgrades, triggering penalty clauses in contracts or levying charges between networks based on ConEx measurements.

Typically, a network using ConEx info will deploy a ConEx policy function near the ingress edge and a ConEx audit function near the egress edge. The segment of the path between a ConEx policy function and a ConEx audit function can be considered to be a ConEx-protected segment of the path. Assuming a network covers all its ingresses and egresses with policy functions and audit functions respectively, the network within this ring will be a ConEx-protected network.

Of course, because each edge device usually serves as both an ingress and an egress, the two functions are both likely to be present in each edge device.

4. Example Initial Deployment Arrangements

In all the deployment scenarios below, we assume that deployment starts with some data sources being modified with ConEx code. The rationale for this is that the developer of a scavenger transport protocol like LEDBAT has a strong incentive to tell the network how little congestion it is causing despite sending large volumes of data. In this case the developer makes the first move expecting it will prompt at least some networks to move in response--so that they use the ConEx information to reward users of the scavenger protocol.

4.1. Single Receiving Network Scenario

The name ’Receiving Network’ for this scenario merely emphasises that most data is arriving from connected networks and data centres and being consumed by residential customers on this access network. Some data is of course also travelling in the other direction.

```
                                          DSLAMs __
                                              /|/     ,-.Home-a
                                        __/__| |-----(   )
                    ,-----.            /  \  | |---   ‘-’
       ,---.       /       \  ,------P/       \|\__
      /     \     ’  Core   ’/| BRAS |          __
     ( Peer  )-->-|P        | ’------’       /|/
      \     /     |         |          _____| |---
       ’---‘      ’         ’\,------./     | |---
                   \ M     /  |BRAS  |       \|\__
                    ‘-----’   ’------A\          __
                     |          P|     \      /|/
                    /|\         /|\     \__\_| |---   ,-.
                   ,---.       ,---.       / | |-----(   )
                  /Data \     /      \        \|\__   ‘-’Home-b
                 ( Centre)    (  CDN  )
                  \     /     \      /  Access Network
                   ’---‘       ’---‘   <------------->
```

P=Congestion-Policer; M=Congestion-Monitor; A=Audit function

Figure 1: Single Receiving Network Scenario

Figure Figure 1 is an attempt to show the salient features of a ConEx deployment in a typical broadband access provider’s network (within the constraints of ASCII art). Broadband remote access servers (BRASs) control access to the core network from the access network and vice versa. Home networks (and small businesses) connect to the access network, but only two are shown.

In this diagram, all data is travelling towards the access network of Home-b, from the Peer network, the Data centre, the CDN and Home-a. Data actually travels in both directions on all links, but only one direction is shown.

The data centre, core and access network are all run by the same network operator, but each is the responsibility of a different department with internal accounting between them. The content distribution network (CDN) is operated by a third party CDN provider, and of course the peer network is also operated by a third party.

This operator of the data centre, core and access network is the only one in the diagram to have deployed ConEx monitoring and policy devices at the edges of its network. However, it has not enabled ECN on any of its network elements and neither has any other network in the diagram. The operator has deployed a congestion policing function (P) on the provider-edge router where the peer attaches to its core, on the BRAS where the CDN attaches and on the other BRAS where each of the residential customers like Home-a attach. On the provider-edge router where the data centre attaches it has deployed a congestion monitoring function (M). Each of these policing and monitoring functions handles the aggregate of all traffic traversing it, for all destinations.

The operator has deployed an audit function on each logical output port of the BRAS for each end-customer site like Home-b. The Audit function handles the aggregate of all traffic for that end-customer from all sources. For traffic in the opposite direction (e.g. from Home-b to Home-a, there would be equivalent policing (P) and audit (A) functions in the converse locations to those shown.

Some content sources in the CDN and in the data centre are using the ConEx protocol, but others are not. There is a similar situation for hosts attached to the Peer network and hosts in home networks like Home-a: some are sending ConEx packets at least for bulk data transports, while others are not.

4.1.1. ConEx Functions in the Single Receiving Network Scenario

Within the BRAS there are logical ports that model the rate of each access line from the DSLAM to each home network [TR-059]. They are fed by a shared queue that models the rate of the downstream link from the BRAS to the DSLAM (sometimes called the backhaul network). If there is congestion anywhere in the set of networks in Figure Figure 1 it is nearly always:

o either self-congestion in the queues into the logical ports representing the access lines

o or shared congestion in the shared queue on the BRAS that feeds them.

Any ConEx sources sending data through this BRAS will receive feedback about these losses from the destination and re-insert it as ConEx markings into the data. Figure 2 shows an example plot of the loss levels that might be seen at different monitoring points along a path between the data centre and home-b, for instance. The top half of the figure shows the loss probability within the BRAS consists of 0.1% at the shared queue and 0.2% self-congestion in the logical output port that models the access line, making 0.3% in total. This upper diagram also shows whole path congestion as signalled by the ConEx sender, which remains unchanged along the whole path at 0.3%.

The lower half of the figure shows (downstream congestion) = (whole path) - (upstream congestion). Upstream congestion can only be monitored locally where the loss actually happens (within the BRAS output queues). Nonetheless, given there is rarely loss anywhere else but within the BRAS, this limitation is not significant in this scenario. The lower half of the figure also shows the location of the policing and audit functions. Policing anywhere within or upstream ofthe BRAS will be based on the downstream congestion level of 0.3%. While Auditing within the BRAS but after all the queues can check that the whole path congestion signalled by ConEx is no less than the loss levels experienced within the BRAS itself.

```
      Data centre-->|<--core-->|<------BRAS--------->|<--Home--
                               |                     |
    ^loss                      |<-Shared->|<-Access->|
    |probability                 backhaul
    |
0.3%|- - - - - - - - - - - - - - - - - - - -  +-----------------
    |      whole path congestion              |
    |                                         |
    |                                         |upstream
0.1%|                             +---------+congestion
    |                             |
   -O==============================+------------------------------>
                                                        monitoring point
    ^loss
    |probability   Policing                  Audit
    |                |                          |
    |                V                          |
0.3%|----------------O-------------+            |
    |                              |downstream  |
0.2%|                              +---------+  |
    |                              congestion|  |
    |                                        |  |
    |                                        |  V
   -O----------------------------------------+====O=============-->
                                                        monitoring point
```

Figure 2: Example plot of loss levels along a path

4.1.2. Incentives to Unilaterally Deploy ConEx in a Receiving Network

Even a sending application that is modified to use ConEx can choose whether to send ConEx or Not-ConEx packets. Nonethelss, ConEx packets bring information to a policer about congestion expected on the rest of the path beyond the policer. Not-ConEx packets bring no such information. Therefore a network that has deployed ConEx policers will tend to rate-limit not-ConEx packets conservatively in order to manage the unknown risk of congestion. In contrast, a network doesn’t normally need to rate-limit ConEx-enabled packets unless they reveal a persistently high contribution to congestion. This natural tendency for networks to favour senders that provide ConEx information encourages senders to choose to use the ConEx protocol whenever they can.

{ToDo: complete this section}

4.2. Mobile Network Scenario

Placeholder for summary of the scenario in a mobile network described in [conex-mobile]

In mobile networks, both mobile terminals and mobile network equipment are standardised by the 3GPP. If the 3GPP were to adopt the ConEx protocol, it might mandate ConEx implementation for compliant equipment.

{ToDo: Describe how a central traffic management box can arrange to remotely view upstream congestion as it would be seen from the interface with the mobile terminal.}

4.3. Scenario Internal to a Multi-Tenant Data Centre

A number of companies offer hosting of virtual machines on their data centre infrastructure--so-called infrastructure as a service (IaaS). A set amount of processing power, memory, storage and network are offered. Although processing power, memory and storage are relatively simple to allocate on the ’pay as you go’ basis that has become common, the network is less easy to allocate given it is a naturally distributed system.

{ToDo: Complete this section.}

5. Security Considerations

6. IANA Considerations

This document does not require actions by IANA.

7. Conclusions

{ToDo}

8. Acknowledgments

9. Informative References

[ConEx-Abstract-Mech] Mathis, M. and B. Briscoe, "Congestion Exposure (ConEx) Concepts and Abstract Mechanism", draft-ietf-conex-abstract-mech-02 (work in progress), July 2011.

[Seawall] Shieh, A., Kandula, S., Greenberg, A., and C. Kim, "Seawall: Performance Isolation in Cloud Datacenter Networks", Proc 2nd USENIX Workshop on Hot Topics in Cloud Computing , June 2010, <http://research.microsoft.com/en-us/projects/seawall/>.

[TR-059] Anschutz, T., Ed., "DSL Forum Technical Report TR-059: Requirements for the Support of QoS-Enabled IP Services", September 2003.

[conex-mobile] Kutscher, D., Mir, F., Winter, R., Krishnan, S., and Y. Zhang, "Mobile Communication Congestion Exposure Scenario", draft-kutscher-conex-mobile-00 (work in progress), March 2011.

Author’s Address

Bob Briscoe
BT
B54/77, Adastral Park
Martlesham Heath
Ipswich IP5 3RE
UK

Phone: +44 1473 645196
EMail: bob.briscoe@bt.com
URI: http://bobbriscoe.net/