Internet Engineering Task Force (IETF)
Request for Comments: 7121
Updates: 5810
Category: Standards Track
ISSN: 2070-1721

K. Ogawa
NTT Corporation
W. Wang
Zhejiang Gongshang University
E. Haleplidis
University of Patras
J. Hadi Salim
Mojatatu Networks

February 2014

High Availability within a Forwarding and Control Element Separation (ForCES) Network Element

Abstract

This document discusses Control Element (CE) High Availability (HA) within a Forwarding and Control Element Separation (ForCES) Network Element (NE). Additionally, this document updates RFC 5810 by providing new normative text for the Cold Standby High Availability mechanism.

Status of This Memo

This is an Internet Standards Track document.

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). Further information on Internet Standards is available in Section 2 of RFC 5741.

Information about the current status of this document, any errata, and how to provide feedback on it may be obtained at http://www.rfc-editor.org/info/rfc7121.

Copyright Notice

Table of Contents

1. Introduction

Figure 1 illustrates a ForCES Network Element (NE) controlled by a set of redundant Control Elements (CEs) with CE1 being active and CE2 and CEn being backups.

```
                           -----------------------------------------
                           | ForCES Network Element                |
                           |                        +-----------+  |
                           |                        |  CEn      |  |
                           |                        |  (Backup) |  |
     --------------   Fc   | +------------+      +------------+ |  |
     | CE Manager |--------+-|    CE1     |------|    CE2     |-+  |
     --------------        | |  (Active)  |  Fr  |  (Backup)  |    |
           |               | +-------+--+-+      +---+---+----+    |
           | Fl            |         |  |    Fp      /   |         |
           |               |         |  +---------+ /    |         |
           |               |       Fp|            |/     |Fp       |
           |               |         |            |      |         |
           |               |         |      Fp   /+--+   |         |
           |               |         |  +-------+    |   |         |
           |               |         |  |            |   |         |
     --------------    Ff  | --------+--+--      ----+---+----+    |
     | FE Manager |--------+-|    FE1     |  Fi  |     FE2    |    |
     --------------        | |            |------|            |    |
                           | --------------      --------------    |
                           |   |  |  |  |          |  |  |  |      |
                           ----+--+--+--+----------+--+--+--+-------
                               |  |  |  |          |  |  |  |
                               |  |  |  |          |  |  |  |
                                 Fi/f                   Fi/f
```

       Fp: CE-FE interface
       Fi: FE-FE interface
       Fr: CE-CE interface
       Fc: Interface between the CE manager and a CE
       Ff: Interface between the FE manager and an FE
       Fl: Interface between the CE manager and the FE manager
       Fi/f: FE external interface

Figure 1: ForCES Architecture

The ForCES architecture allows Forwarding Elements (FEs) to be aware of multiple CEs but enforces that only one CE be the master controller. This is known in the industry as 1+N redundancy. The master CE controls the FEs via the ForCES protocol operating on the Fp interface. If the master CE becomes faulty, i.e., crashes or loses connectivity, a backup CE takes over and NE operation continues. By definition, the current documented setup is known as cold standby. The set of CEs controlling an FE is static and is passed to the FE by the FE Manager (FEM) via the Ff interface and to each CE by the CE Manager (CEM) in the Fc interface during the pre-association phase.

From an FE perspective, the operational parameters for a CE set are defined as components in the FEPO LFB in [RFC5810], Appendix B. In Section 2.1 of this document, we discuss further details of these parameters.

It is assumed that the reader is aware of the ForCES architecture to make sense of the changes being described in this document. This document provides background information to set the context of the discussion in Section 3.

At the time of writing, the Fr interface is out of scope for the ForCES architecture. However, it is expected that organizations implementing a set of CEs will need to have the CEs communicate to each other via the Fr interface in order to achieve the synchronization necessary for controlling the FEs.

The problem scope addressed by this document falls into two areas:

1. To update the description of [RFC5810] with more clarity on how the current cold standby approach operates within the NE cluster.

2. To describe how to evolve the [RFC5810] cold standby setup to a hot standby redundancy setup to improve the failover time and NE availability.

1.1. Quantifying Problem Scope

NE recovery and availability is dependent on several time-sensitive metrics:

1. How fast the CE plane failure is detected by the FE.

2. How fast a backup CE becomes operational.

3. How fast the FEs associate with the new master CE.

4. How fast the FEs recover their state and become operational. Each FE state is the collective state of all its instantiated LFBs.

The design intent of [RFC5810] as well as this document to meet the above goals is driven by desire for simplicity.

To quantify the above criteria with the current prescribed ForCES CE setup in [RFC5810]:

1. How fast the FE side detects a CE failure is left undefined. To illustrate an extreme scenario, we could have a human operator acting as the monitoring entity to detect faulty CEs. How fast such detection happens could be in the range of seconds to days. A more active monitor on the Fp interface could improve this detection. Usually, the FE will detect a CE failure either by the TML if the Fp interface terminates or by the ForCES protocol by utilizing the ForCES Heartbeat mechanism.

2. How fast the backup CE becomes operational is also currently out of scope. In the current setup, a backup CE need not be operational at all (for example, to save power), and therefore it is feasible for a monitoring entity to boot up a backup CE after it detects the failure of the master CE. In Section 3 of this document, we suggest that at least one backup CE be online so as to improve this metric.

3. How fast an FE associates with a new master CE is also currently undefined. The cost of an FE connecting and associating adds to the recovery overhead. As mentioned above, we suggest having at least one backup CE online. In Section 3, we propose to remove the connection and association cost on failover by having each FE associate with all online backup CEs after associating to an active/master CE. Note that if an FE pre-associates with at least one backup CE, then the system will be technically operating in hot standby mode.

4. Finally, how fast an FE recovers its state depends on how much NE state exists. By the ForCES current definition, the new master CE assumes zero state on the FE and starts from scratch to update the FE. So, the larger the state, the longer the recovery.

1.2. Definitions

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].

The following definitions are taken from [RFC3654], [RFC3746], and [RFC5810]. They are repeated here for convenience as needed, but the normative definitions are found in the referenced RFCs:

Logical Functional Block (LFB): A template that represents fine-grained, logically separate aspects of FE processing.

Forwarding Element (FE): A logical entity that implements the ForCES protocol. FEs use the underlying hardware to provide per-packet processing and handling as directed by a CE via the ForCES protocol.

Control Element (CE): A logical entity that implements the ForCES protocol and uses it to instruct one or more FEs on how to process packets. CEs handle functionality such as the execution of control and signaling protocols.

ForCES Network Element (NE): An entity composed of one or more CEs and one or more FEs. An NE usually hides its internal organization from external entities and represents a single point of management to entities outside the NE.

FE Manager (FEM): A logical entity that operates in the pre-association phase and is responsible for determining to which CE(s) an FE should communicate. This process is called CE discovery and may involve the FE manager learning the capabilities of available CEs.

CE Manager (CEM): A logical entity that operates in the pre-association phase and is responsible for determining to which FE(s) a CE should communicate. This process is called FE discovery and may involve the CE manager learning the capabilities of available FEs.

ForCES Protocol: The protocol used for communication between CEs and FEs. This protocol does not apply to CE-to-CE communication, FE-to-FE communication, or to communication between FE and CE managers. The ForCES protocol is a master-slave protocol in which FEs are slaves and CEs are masters. This protocol includes both the management of the communication channel (e.g., connection establishment and heartbeats) and the control messages themselves.

ForCES Protocol Layer (ForCES PL): A layer in the ForCES protocol architecture that defines the ForCES protocol messages, the protocol state transfer scheme, and the ForCES protocol architecture itself (including requirements of ForCES Transport Mapping Layer (TML) as shown below). Specifications of ForCES PL are defined in [RFC5810].

ForCES Protocol Transport Mapping Layer (ForCES TML): A layer in the ForCES protocol architecture that specifically addresses the protocol message transportation issues, such as how the protocol messages are mapped to different transport media (like Stream

Control Transmission Protocol (SCTP), IP, TCP, UDP, ATM, Ethernet, etc.), and how to achieve and implement reliability, security, etc.

2. RFC 5810 CE HA Framework

To achieve CE High Availability (HA), FEs and CEs MUST interoperate per the definition in [RFC5810], which is repeated for contextual reasons in Section 2.1. It should be noted that in this default setup, which MUST be implemented by CEs and FEs requiring HA, the Fr plane is out of scope (and if available, is proprietary to an implementation).

2.1. RFC 5810 CE HA Support

As mentioned earlier, although there can be multiple redundant CEs, only one CE actively controls FEs in a ForCES NE. In practice, there may be only one backup CE. At any moment in time, only one master CE can control an FE. In addition, the FE connects and associates to only the master CE. The FE and the CE are aware of the primary and one or more secondary CEs. This information (primary and secondary CEs) is configured on the FE and the CE during pre-association by the FEM and the CEM, respectively.

This section includes a new normative description that updates [RFC5810] for the Cold Standby High Availability mechanism.

Figure 2 below illustrates the ForCES message sequences that the FE uses to recover the connection in the currently defined cold standby scheme.

```
         FE                       CE Primary         CE Secondary
         |                           |                     |
         | Association Establishment |                     |
         |   Capabilities Exchange   |                     |
       1 |<------------------------->|                     |
         |                           |                     |
         |       State Update        |                     |
       2 |<------------------------->|                     |
         |                           |                     |
         |                        FAILURE                  |
         |                                                 |
         | Association Establishment, Capabilities Exchange|
       3 |<----------------------------------------------->|
         |                                                 |
         |         Event Report (primary CE down)          |
       4 |------------------------------------------------>|
         |                                                 |
         |                  State Update                   |
       5 |<----------------------------------------------->|
```

Figure 2: CE Failover for Cold Standby

2.1.1. Cold Standby Interaction with the ForCES Protocol

HA parameterization in an FE is driven by configuring the FE Protocol Object (FEPO) LFB.

The FEPO Control Element ID (CEID) component identifies the current master CE, and the component table BackupCEs identifies the configured backup CEs. The FEPO FE Heartbeat Interval (FEHI), CE Heartbeat Dead Interval (CEHDI), and CE Heartbeat policy help in detecting connectivity problems between an FE and CE. The CE failover policy defines how the FE should react on a detected failure. The FEObject FEState component [RFC5812] defines the operational forwarding status and control. The CE can turn off the FE's forwarding operations by setting the FEState to AdminDisable and can turn it on by setting it to OperEnable. Note: Section 5.1 of [RFC5812] has been updated by an erratum ([Err3487]) that describes the FEState as read-only when it should be read-write.

Figure 3 illustrates the defined state machine that facilitates the recovery of the connection state.

The FE connects to the CE specified on the FEPO CEID component. If it fails to connect to the defined CE, it moves it to the bottom of table BackupCEs and sets its CEID component to be the first CE retrieved from table BackupCEs. The FE then attempts to associate with the CE designated as the new primary CE. The FE continues through this procedure until it successfully connects to one of the CEs or until the CE Failover Timeout Interval (CEFTI) expires.

```
                             FE tries to associate
                                   +-->-----+
                                   |        |
      (CE changes master ||        |        |
      CE issues Teardown ||    +---+--------v----+
        Lost association) &&   | Pre-association |
       CE failover policy = 0  | (Association    |
           +------------>-->-->|   in            +<----+
           |                   | progress)       |     |
           |                   |                 |     |
           |                   +--------+--------+     |
           |  CE Association        |                  | CEFTI
           |       Response         V                  | timer
           |     +------------------+                  | expires
           |     |FE issues CEPrimaryDown              ^
           |     V                                     |
         +-+-----------+                        +------+-----+
         |             |  (CE changes master || | Not        |
         |             |  CE issues Teardown || | Associated |
         |             |  Lost association) &&  |            +->---+
         | Associated  | CE failover policy = 1 |(May        | FE  |
         |             |                        | Continue   | try v
         |             |-------->------->------>| Forwarding)| assn|
         |             |   Start CEFTI timer    |            |-<---+
         |             |                        |            |
         +-------------+                        +-------+----+
              ^                                         |
              |            Successful                   V
              |            Association                  |
              |            Setup                        |
              |            (Cancel CEFTI timer)         |
              +_________________________________________+
                       FE issues CEPrimaryDown event
```

Figure 3: FE State Machine Considering HA

There are several events that trigger mastership changes. The master CE may issue a mastership change (by changing the CEID component), it may tear down an existing association, or connectivity may be lost between the CE and FE.

When communication fails between the FE and CE (which can be caused by either the CE or link failure but is not FE related), either the TML on the FE will trigger the FE PL regarding this failure or it will be detected using the Heartbeat messages between FEs and CEs. The communication failure, regardless of how it is detected, MUST be considered to be a loss of association between the CE and corresponding FE.

If the FE's FEPO CE failover policy is configured to mode 0 (the default), it will immediately transition to the pre-association phase. This means that if association is later re-established with a CE, all FE states will need to be re-created.

If the FE's FEPO CE failover policy is configured to mode 1, it indicates that the FE will run in HA restart recovery. In such a case, the FE transitions to the not associated state and the CEFTI timer [RFC5810] is started. The FE may continue to forward packets during this state, depending upon the value of the CEFailoverPolicy component of the FEPO LFB. The FE recycles through any configured backup CEs in a round-robin fashion. It first adds its primary CE to the bottom of table BackupCEs and sets its CEID component to be the first secondary retrieved from table BackupCEs. The FE then attempts to associate with the CE designated as the new primary CE. If it fails to re-associate with any CE and the CEFTI expires, the FE then transitions to the pre-association state and the FE will operationally bring down its forwarding path (and set the [RFC5812] FEObject FEState component to OperDisable).

If the FE, while in the not associated state, manages to reconnect to a new primary CE before the CEFTI expires, it transitions to the associated state. Once re-associated, the CE may try to synchronize any state that the FE may have lost during disconnection. How the CE re-synchronizes such a state is out of scope for the current ForCES architecture but would typically constitute the issuing of new Config messages and queries.

An explicit message (a Config message setting the primary CE component in the ForCES Protocol Object) from the primary CE can also be used to change the primary CE for an FE during normal protocol operation. In this case, the FE transitions to the not associated state and attempts to associate with the new CE.

2.1.2. Responsibilities for HA

TML Level:

1. The TML controls logical connection availability and failover.

2. The TML also controls peer HA management.

At this level, control of all lower layers, for example, the transport level (such as IP addresses, Media Access Control (MAC) addresses, etc.), and associated links going down are the role of the TML.

PL Level:
All other functionality, including configuring the HA behavior during setup, Control Element IDs (CE IDs) used to identify primary and secondary CEs, protocol messages used to report CE failure (event report), Heartbeat messages used to detect association failure, messages to change the primary CE (Config), and other HA-related operations described in Section 2.1, are the PL's responsibility.

To put the two together, if a path to a primary CE is down, the TML would help recover from a failure by switching over to a backup path, if one is available. If the CE is totally unreachable, then the PL would be informed and it would take the appropriate actions described before.

3. CE HA Hot Standby

In this section, we describe small extensions to the existing scheme to enable hot standby HA. To achieve hot standby HA, we aim to improve the specific goals defined in Section 1.1, namely:

- How fast a backup CE becomes operational.

- How fast the FEs associate with the new master CE.

As described in Section 2.1, in the pre-association phase, the FEM configures the FE to make it aware of all the CEs in the NE. The FEM MUST configure the FE to make it aware of which CE is the master and MAY specify any backup CE(s).

3.1. Changes to the FEPO Model

In order for the above to be achievable, there is a need to make a few changes in the FEPO model. Appendix A contains the xml definition of the new version 1.1 of the FEPO LFB.

Changes from version 1 of the FEPO are:

1. Added four new datatypes:

   1. CEStatusType -- an unsigned char to specify the status of a connection with a CE. Special values are:

      + 0 (Disconnected) represents that no connection attempt has been made with the CE yet

      + 1 (Connected) represents that the FE connection with the CE at the TML has completed successfully

      + 2 (Associated) represents that the FE has successfully associated with the CE

      + 3 (IsMaster) represents that the FE has associated with the CE and is the master of the FE

      + 4 (LostConnection) represents that the FE was associated with the CE at one point but lost the connection

      + 5 (Unreachable) represents that the FE deems this CE unreachable, i.e., the FE has tried over a period to connect to it but has failed

   2. HAModeValues -- an unsigned char to specify a selected HA mode. Special values are:

      + 0 (No HA Mode) represents that the FE is not running in HA mode

      + 1 (HA Mode - Cold Standby) represents that the FE is in HA mode cold standby

      + 2 (HA Mode - Hot Standby) represents that the FE is in HA mode hot standby

   3. Statistics -- a complex structure representing the communication statistics between the FE and CE. The components are:

      + RecvPackets, representing the packet count received from the CE

      + RecvBytes, representing the byte count received from the CE

        + RecvErrPackets, representing the erroneous packets received from the CE. This component logs badly formatted packets as well as good packets sent to the FE by the CE to set components whilst that CE is not the master. Erroneous packets are dropped (i.e., not responded to).

        + RecvErrBytes, representing the RecvErrPackets byte count received from the CE

        + TxmitPackets, representing the packet count transmitted to the CE

        + TxmitErrPackets, representing the error packet count transmitted to the CE. Typically, these would be failures due to communication.

        + TxmitBytes, representing the byte count transmitted to the CE

        + TxmitErrBytes, representing the byte count of errors from transmit to the CE

    4. AllCEType -- a complex structure constituting the CE IDs, statistics, and CEStatusType to reflect connection information for one CE. Used in the AllCE's component array.

2. Appended two new components:

    1. Read-only AllCEs to hold the status for all CEs. AllCEs is an array of the AllCEType.

    2. Read-write HAMode of type HAModeValues to carry the HA mode used by the FE.

3. Added one additional event, PrimaryCEChanged, reporting the new master CE ID when there is a mastership change.

Since no component from FEPO v1 has been changed, FEPO v1.1 retains backwards compatibility with CEs that know only version 1.0. These CEs, however, cannot make use of the HA options that the new FEPO provides.

3.2. FEPO Processing

The FE's FEPO LFB version 1.1 AllCEs table contains all the CE IDs with which the FE may connect and associate. The ordering of the CE IDs in this table defines the priority order in which an FE will connect to the CEs. This table is provisioned initially from the configuration plane (FEM). In the pre-association phase, the first CE (lowest table index) in the AllCEs table MUST be the first CE with which the FE will attempt to connect and associate. If the FE fails to connect and associate with the first listed CE, it will attempt to connect to the second CE and so forth, and it cycles back to the beginning of the list until there is a successful association. The FE MUST associate with at least one CE. Upon a successful association, a component of the FEPO LFB, specifically the CEID component, identifies the current associated master CE.

While it would be much simpler to have the FE not respond to any messages from a CE other than the master, in practice it has been found to be useful to respond to queries and heartbeats from backup CEs. For this reason, we allow backup CEs to issue queries to the FE. Configuration messages (SET/DEL) from backup CEs MUST be dropped by the FE and logged as received errors.

Asynchronous events that the master CE has subscribed to, as well as heartbeats, are sent to all associated CEs. Packet redirects continue to be sent only to the master CE. The Heartbeat Interval, the CE Heartbeat (CEHB) policy, and the FE Heartbeat (FEHB) policy are global for all CEs (and changed only by the master CE).

Figure 4 illustrates the state machine that facilitates connection recovery with HA enabled.

```
                           FE tries to associate
                                +-->-----+
                                |        |
   (CE changes master ||        |        |
   CE issues Teardown ||    +---+--------v----+
     Lost association) &&   | Pre-association |
    CE failover policy = 0  | (Association    |
        +------------>-->-->|   in            +<----+
        |                   | progress)       |     |
        |                   |                 |     |
        |                   +--------+--------+     |
        |  CE Association       |                   | CEFTI
        |       Response        V                   | timer
        |     +-------------------+                 | expires
        |     |FE issues CEPrimaryDown              ^
        |     |FE issues PrimaryCEChanged           ^
        |     V                                     |
      +-+-----------+                        +------+-----+
      |             |  (CE changes master || | Not        |
      |             |  CE issues Teardown || | Associated |
      |             |  Lost association) &&               +->----------+
      | Associated  | CE failover policy = 1 |(May        | find first |
      |             |                        | Continue   | associated v
      |             |-------->------->------>| Forwarding)| CE or retry|
      |             |   Start CEFTI timer    |            | associating|
      |             |                        |            |-<----------+
      |             |                        |            |
      +----+--------+                        +-------+----+
           |                                         |
           ^                                   Found | associated CE
           |                                or newly | associated CE
           |                                         V
           |            (Cancel CEFTI timer)         |
           +_________________________________________+
                    FE issues CEPrimaryDown event
                    FE issues PrimaryCEChanged event
```

Figure 4: FE State Machine Considering HA

Once the FE has associated with a master CE, it moves to the post-association phase (associated state). It is assumed that the master CE will communicate with other CEs within the NE for the purpose of synchronization via the CE-CE interface. The CE-CE interface is out of scope for this document. An election result amongst CEs may result in the desire to change the mastership to a different associated CE; at which point, the current assumed master CE will instruct the FE to use a different master CE.

```
         FE                          CE#1         CE#2 ... CE#N
         |                           |            |        |
         | Association Establishment |            |        |
         |   Capabilities Exchange   |            |        |
       1 |<------------------------->|            |        |
         |                           |            |        |
         |      State Update         |            |        |
       2 |<------------------------->|            |        |
         |                           |            |        |
         |      Association Establishment         |        |
         |        Capabilities Exchange           |        |
      3I |<-------------------------------------->|        |
        ...                         ...          ...      ...
         |Association Establishment, Capabilities Exchange |
      3N |<----------------------------------------------->|
         |                           |            |        |
       4 |<------------------------->|            |        |
         .                           .            .        .
      4x |<------------------------->|            |        |
         |                        FAILURE         |        |
         |                           |            |        |
         |    Event Report (LastCEID changed)     |        |
       5 |--------------------------------------->|------->|
         |    Event Report (CE#2 is new master)   |        |
       6 |--------------------------------------->|------->|
         |                                        |        |
       7 |<-------------------------------------->|        |
         .                           .            .        .
      7x |<-------------------------------------->|        |
         .                           .            .        .
```

Figure 5: CE Failover for Hot Standby

While in the post-association phase, if the CE failover policy is set to 1 and the HAMode is set to 2 (hot standby), then the FE, after successfully associating with the master CE, MUST attempt to connect and associate with all the CEs of which it is aware. Figure 5, steps #1 and #2 illustrates the FE associating with CE#1 as the master, and then proceeding to steps #3I to #3N, it shows the association with backup CEs CE#2 to CE#N. If the FE fails to connect or associate with some CEs, the FE MAY flag them as unreachable to avoid continuous attempts to connect. The FE MAY try to re-associate with unreachable CEs when possible.

When the master CE, for any reason, is considered to be down, then the FE MUST try to find the first associated CE from the list of all CEs in a round-robin fashion.

If the FE is unable to find an associated FE in its list of CEs, then it MUST attempt to connect and associate with the first from the list of all CEs and continue in a round-robin fashion until it connects and associates with a CE or the CEFTI timer expires.

Once the FE selects an associated CE to use as the new master, the FE issues a PrimaryCEDown Event Notification to all associated CEs to notify them that the last primary CE went down (and what its identity was); a second event, PrimaryCEChanged, identifying the new master CE is sent as well to identify which CE the reporting FE considers to be the new master.

In most HA architectures, there exists the possibility of split brain. However, in our setup, since the FE will never accept any configuration messages from any other than the master CE, we consider the FE to be fenced against data corruption from the other CEs that consider themselves as the master. The split-brain issue becomes mostly a CE-CE communication problem, which is considered to be out of scope.

By virtue of having multiple CE connections, the FE switchover to a new master CE will be relatively much faster. The overall effect is improving the NE recovery time in case of communication failure or faults of the master CE. This satisfies the requirement we set to fulfill.

4. IANA Considerations

Following the policies outlined in "Guidelines for Writing an IANA Considerations Section in RFCs" [RFC5226], the "Logical Functional Block (LFB) Class Names and Class Identifiers" namespace has been updated.

A new column, LFB version, has been added to the table after the LFB Class Name. The table now reads as follows:

LFB Class Identifier	LFB Class Name	LFB Version	Description	Reference

Logical Functional Block (LFB) Class Names and Class Identifiers

The rules defined in [RFC5812] apply, with the addition that entries must provide the LFB version as a string.

Upon publication of this document, all current entries are assigned a value of 1.0.

New versions of already defined LFBs MUST NOT remove the previous version entries.

It would make sense to have LFB versions appear in sequence in the registry. The table SHOULD be sorted, and the sorting should be done by Class ID first and then by version.

This document introduces the FE Protocol Object version 1.1 as follows:

LFB Class Identifier	LFB Class Name	LFB Version	Description	Reference
2	FE Protocol Object	1.1	Defines parameters for the ForCES protocol operation	[RFC7121]

Logical Functional Block (LFB) Class Names and Class Identifiers

5. Security Considerations

Security considerations, as defined in Section 9 of [RFC5810], apply to securing each CE-FE communication. Multiple CEs associated with the same FE still require the same procedure to be followed on a per-association basis.

It should be noted that since the FE is initiating the association with a CE, a CE cannot initiate association with the FE and such messages will be dropped. Thus, the FE is secured from rogue CEs that are attempting to associate with it.

CE implementers should have in mind that once associated, the FE cannot distinguish whether the CE has been compromised or has been malfunctioning while not losing connectivity. Securing the CE is out of scope of this document.

While the CE-CE plane is outside the current scope of ForCES, we recognize that it may be subjected to attacks that may affect the CE-FE communication.

The following considerations should be made:

1. Secure communication channels should be used between CEs for coordination and keeping of state to at least avoid connection of malicious CEs.

2. The master CE should take into account DoS and Distributed Denial-of-Service (DDoS) attacks from malicious or malfunctioning CEs.

3. CEs should take into account the split-brain issue. There are currently two fail-safes in the FE: Firstly, the FE has the CEID component that denotes which CE is the master. Secondly, the FE does not allow BackupCEs to configure the FE. However, backup CEs that consider that the master CE has dropped should, as masters themselves, first do a sanity check and query the FE CEID component.

6. References

6.1. Normative References

[RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, March 1997.

[RFC5226] Narten, T. and H. Alvestrand, "Guidelines for Writing an IANA Considerations Section in RFCs", BCP 26, RFC 5226, May 2008.

[RFC5810] Doria, A., Hadi Salim, J., Haas, R., Khosravi, H., Wang, W., Dong, L., Gopal, R., and J. Halpern, "Forwarding and Control Element Separation (ForCES) Protocol Specification", RFC 5810, March 2010.

[RFC5812] Halpern, J. and J. Hadi Salim, "Forwarding and Control Element Separation (ForCES) Forwarding Element Model", RFC 5812, March 2010.

6.2. Informative References

[Err3487] RFC Errata, Errata ID 3487, RFC 5812, <http://www.rfc-editor.org>.

[RFC3654] Khosravi, H. and T. Anderson, "Requirements for Separation of IP Control and Forwarding", RFC 3654, November 2003.

[RFC3746] Yang, L., Dantu, R., Anderson, T., and R. Gopal, "Forwarding and Control Element Separation (ForCES) Framework", RFC 3746, April 2004.

Appendix A. New FEPO Version

The xml has been validated against the schema defined in [RFC5812].

```
<LFBLibrary xmlns="urn:ietf:params:xml:ns:forces:lfbmodel:1.0"
   xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
   xsi:noNamespaceSchemaLocation="lfb-schema.xsd" provides="FEPO">
   <!-- XXX -->
   <dataTypeDefs>
      <dataTypeDef>
         <name>CEHBPolicyValues</name>
         <synopsis>
            The possible values of the CE Heartbeat policy
         </synopsis>
         <atomic>
            <baseType>uchar</baseType>
            <specialValues>
               <specialValue value="0">
                  <name>CEHBPolicy0</name>
                  <synopsis>
              The CE will send heartbeats to the FE
              every CEHDI timeout if no other messages
              have been sent since.
                  </synopsis>
               </specialValue>
               <specialValue value="1">
                  <name>CEHBPolicy1</name>
                  <synopsis>
              The CE will not send heartbeats to the FE
                  </synopsis>
               </specialValue>
            </specialValues>
         </atomic>
      </dataTypeDef>
      <dataTypeDef>
         <name>FEHBPolicyValues</name>
         <synopsis>
            The possible values of the FE Heartbeat policy
         </synopsis>
         <atomic>
            <baseType>uchar</baseType>
            <specialValues>
               <specialValue value="0">
                  <name>FEHBPolicy0</name>
                  <synopsis>
        The FE will not generate any heartbeats to the CE
                  </synopsis>
               </specialValue>
               <specialValue value="1">
                  <name>FEHBPolicy1</name>
                  <synopsis>
        The FE generates heartbeats to the CE every FEHI
        if no other messages have been sent to the CE.
                  </synopsis>
               </specialValue>
            </specialValues>
         </atomic>
      </dataTypeDef>
      <dataTypeDef>
         <name>FERestartPolicyValues</name>
         <synopsis>
            The possible values of the FE restart policy
         </synopsis>
         <atomic>
            <baseType>uchar</baseType>
            <specialValues>
               <specialValue value="0">
                  <name>FERestartPolicy0</name>
                  <synopsis>
                     The FE restarts its state from scratch
                  </synopsis>
               </specialValue>
            </specialValues>
         </atomic>
      </dataTypeDef>
      <dataTypeDef>
         <name>HAModeValues</name>
         <synopsis>
            The possible values of HA modes
         </synopsis>
         <atomic>
            <baseType>uchar</baseType>
            <specialValues>
               <specialValue value="0">
                  <name>NoHA</name>
                  <synopsis>
                     The FE is not running in HA mode
                  </synopsis>
               </specialValue>
               <specialValue value="1">
                  <name>ColdStandby</name>
                  <synopsis>
                     The FE is running in HA mode cold standby
                  </synopsis>
               </specialValue>
               <specialValue value="2">
```

```
                  <name>HotStandby</name>
                  <synopsis>
                     The FE is running in HA mode hot standby
                  </synopsis>
               </specialValue>
            </specialValues>
         </atomic>
      </dataTypeDef>
      <dataTypeDef>
         <name>CEFailoverPolicyValues</name>
         <synopsis>
            The possible values of the CE failover policy
         </synopsis>
         <atomic>
            <baseType>uchar</baseType>
            <specialValues>
               <specialValue value="0">
                  <name>CEFailoverPolicy0</name>
                  <synopsis>
        The FE should stop functioning immediately and
        transition to the FE OperDisable state
                  </synopsis>
               </specialValue>
               <specialValue value="1">
                  <name>CEFailoverPolicy1</name>
                  <synopsis>
        The FE should continue forwarding even without an
        associated CE for CEFTI. The FE goes to FE
        OperDisable when the CEFTI expires and there is no
        association. Requires graceful restart support.
                  </synopsis>
               </specialValue>
            </specialValues>
         </atomic>
      </dataTypeDef>
      <dataTypeDef>
         <name>FEHACapab</name>
         <synopsis>
            The supported HA features
         </synopsis>
         <atomic>
            <baseType>uchar</baseType>
            <specialValues>
               <specialValue value="0">
                  <name>GracefullRestart</name>
                  <synopsis>
                     The FE supports graceful restart
                  </synopsis>
               </specialValue>
               <specialValue value="1">
                  <name>HA</name>
                  <synopsis>
                     The FE supports HA
                  </synopsis>
               </specialValue>
            </specialValues>
         </atomic>
      </dataTypeDef>
      <dataTypeDef>
         <name>CEStatusType</name>
         <synopsis>Status values. Status for each CE</synopsis>
         <atomic>
            <baseType>uchar</baseType>
            <specialValues>
               <specialValue value="0">
                  <name>Disconnected</name>
                  <synopsis>No connection attempt with the CE yet
                  </synopsis>
               </specialValue>
               <specialValue value="1">
                  <name>Connected</name>
                  <synopsis>The FE connection with the CE at the TML
                     has been completed
                  </synopsis>
               </specialValue>
               <specialValue value="2">
                  <name>Associated</name>
                  <synopsis>The FE has associated with the CE
                  </synopsis>
               </specialValue>
               <specialValue value="3">
                  <name>IsMaster</name>
                  <synopsis>The CE is the master (and associated)
                  </synopsis>
               </specialValue>
               <specialValue value="4">
                  <name>LostConnection</name>
                  <synopsis>The FE was associated with the CE but
                     lost the connection
                  </synopsis>
               </specialValue>
               <specialValue value="5">
                  <name>Unreachable</name>
                  <synopsis>The CE is deemed as unreachable by the FE
                  </synopsis>
               </specialValue>
            </specialValues>
         </atomic>
      </dataTypeDef>
      <dataTypeDef>
         <name>StatisticsType</name>
         <synopsis>Statistics Definition</synopsis>
         <struct>
            <component componentID="1">
               <name>RecvPackets</name>
               <synopsis>Packets received</synopsis>
               <typeRef>uint64</typeRef>
            </component>
            <component componentID="2">
               <name>RecvErrPackets</name>
               <synopsis>Packets received from the CE with errors
               </synopsis>
               <typeRef>uint64</typeRef>
            </component>
            <component componentID="3">
               <name>RecvBytes</name>
               <synopsis>Bytes received from the CE</synopsis>
               <typeRef>uint64</typeRef>
            </component>
            <component componentID="4">
               <name>RecvErrBytes</name>
               <synopsis>Bytes received from the CE in Error</synopsis>
               <typeRef>uint64</typeRef>
            </component>
            <component componentID="5">
               <name>TxmitPackets</name>
               <synopsis>Packets transmitted to the CE</synopsis>
               <typeRef>uint64</typeRef>
            </component>
            <component componentID="6">
               <name>TxmitErrPackets</name>
               <synopsis>
                  Packets transmitted to the CE that
                  incurred errors
               </synopsis>
               <typeRef>uint64</typeRef>
            </component>
            <component componentID="7">
               <name>TxmitBytes</name>
               <synopsis>Bytes transmitted to the CE</synopsis>
               <typeRef>uint64</typeRef>
            </component>
            <component componentID="8">
               <name>TxmitErrBytes</name>
               <synopsis>
                  Bytes transmitted to the CE that
                  incurred errors
               </synopsis>
               <typeRef>uint64</typeRef>
            </component>
         </struct>
      </dataTypeDef>
      <dataTypeDef>
         <name>AllCEType</name>
         <synopsis>Table type for the AllCE component</synopsis>
         <struct>
            <component componentID="1">
               <name>CEID</name>
               <synopsis>ID of the CE</synopsis>
               <typeRef>uint32</typeRef>
            </component>
            <component componentID="2">
               <name>Statistics</name>
               <synopsis>Statistics per the CE</synopsis>
               <typeRef>StatisticsType</typeRef>
            </component>
            <component componentID="3">
               <name>CEStatus</name>
               <synopsis>Status of the CE</synopsis>
               <typeRef>CEStatusType</typeRef>
            </component>
         </struct>
      </dataTypeDef>
   </dataTypeDefs>
   <LFBClassDefs>
      <LFBClassDef LFBClassID="2">
         <name>FEPO</name>
         <synopsis>
            The FE Protocol Object, with new CEHA
         </synopsis>
         <version>1.1</version>
         <components>
            <component componentID="1" access="read-only">
               <name>CurrentRunningVersion</name>
               <synopsis>Currently running the ForCES version</synopsis>
               <typeRef>uchar</typeRef>
            </component>
            <component componentID="2" access="read-only">
               <name>FEID</name>
               <synopsis>Unicast FEID</synopsis>
               <typeRef>uint32</typeRef>
            </component>
            <component componentID="3" access="read-write">
               <name>MulticastFEIDs</name>
               <synopsis>
                  The table of all multicast IDs
               </synopsis>
               <array type="variable-size">
                  <typeRef>uint32</typeRef>
               </array>
            </component>
            <component componentID="4" access="read-write">
               <name>CEHBPolicy</name>
               <synopsis>
                  The CE Heartbeat policy
               </synopsis>
               <typeRef>CEHBPolicyValues</typeRef>
            </component>
            <component componentID="5" access="read-write">
               <name>CEHDI</name>
               <synopsis>
                  The CE Heartbeat Dead Interval in milliseconds
               </synopsis>
               <typeRef>uint32</typeRef>
            </component>
            <component componentID="6" access="read-write">
               <name>FEHBPolicy</name>
               <synopsis>
                  The FE Heartbeat policy
               </synopsis>
               <typeRef>FEHBPolicyValues</typeRef>
            </component>
            <component componentID="7" access="read-write">
               <name>FEHI</name>
               <synopsis>
                  The FE Heartbeat Interval in milliseconds
               </synopsis>
               <typeRef>uint32</typeRef>
            </component>
            <component componentID="8" access="read-write">
               <name>CEID</name>
               <synopsis>
                  The primary CE this FE is associated with
               </synopsis>
               <typeRef>uint32</typeRef>
            </component>
            <component componentID="9" access="read-write">
               <name>BackupCEs</name>
               <synopsis>
                  The table of all backup CEs other than the
                  primary
               </synopsis>
               <array type="variable-size">
                  <typeRef>uint32</typeRef>
               </array>
            </component>
            <component componentID="10" access="read-write">
               <name>CEFailoverPolicy</name>
               <synopsis>
                  The CE failover policy
               </synopsis>
               <typeRef>CEFailoverPolicyValues</typeRef>
            </component>
            <component componentID="11" access="read-write">
               <name>CEFTI</name>
               <synopsis>
                  The CE Failover Timeout Interval in milliseconds
               </synopsis>
               <typeRef>uint32</typeRef>
            </component>
            <component componentID="12" access="read-write">
               <name>FERestartPolicy</name>
               <synopsis>
                  The FE restart policy
               </synopsis>
               <typeRef>FERestartPolicyValues</typeRef>
            </component>
            <component componentID="13" access="read-write">
               <name>LastCEID</name>
               <synopsis>
                  The primary CE this FE was last associated
                  with
               </synopsis>
               <typeRef>uint32</typeRef>
            </component>
            <component componentID="14" access="read-write">
               <name>HAMode</name>
               <synopsis>
                  The HA mode used
               </synopsis>
               <typeRef>HAModeValues</typeRef>
            </component>
            <component componentID="15" access="read-only">
               <name>AllCEs</name>
               <synopsis>The table of all CEs</synopsis>
               <array type="variable-size">
```

```
                  <typeRef>AllCETYPE</typeRef>
               </array>
            </component>
         </components>
         <capabilities>
            <capability componentID="30">
               <name>SupportableVersions</name>
               <synopsis>
                  The table of ForCES versions that FE supports
               </synopsis>
               <array type="variable-size">
                  <typeRef>uchar</typeRef>
               </array>
            </capability>
            <capability componentID="31">
               <name>HACapabilities</name>
               <synopsis>
                  The table of HA capabilities the FE supports
               </synopsis>
               <array type="variable-size">
                  <typeRef>FEHACapab</typeRef>
               </array>
            </capability>
         </capabilities>
         <events baseID="61">
            <event eventID="1">
               <name>PrimaryCEDown</name>
               <synopsis>
                  The primary CE has changed
               </synopsis>
               <eventTarget>
                  <eventField>LastCEID</eventField>
               </eventTarget>
               <eventChanged/>
               <eventReports>
                  <eventReport>
                     <eventField>LastCEID</eventField>
                  </eventReport>
               </eventReports>
            </event>
            <event eventID="2">
               <name>PrimaryCEChanged</name>
               <synopsis>A new primary CE has been selected
               </synopsis>
               <eventTarget>
                  <eventField>CEID</eventField>
               </eventTarget>
               <eventChanged/>
               <eventReports>
                  <eventReport>
                     <eventField>CEID</eventField>
                  </eventReport>
               </eventReports>
            </event>
         </events>
      </LFBClassDef>
   </LFBClassDefs>
</LFBLibrary>
```

Authors' Addresses

Kentaro Ogawa
NTT Corporation
3-9-11 Midori-cho
Musashino-shi, Tokyo 180-8585
Japan

EMail: k.ogawa@ntt.com

Weiming Wang
Zhejiang Gongshang University
18 Xuezheng Str., Xiasha University Town
Hangzhou 310018
P.R. China

Phone: +86 571 28877751
EMail: wmwang@zjsu.edu.cn

Evangelos Haleplidis
University of Patras
Department of Electrical and Computer Engineering
Patras 26500
Greece

EMail: ehalep@ece.upatras.gr

Jamal Hadi Salim
Mojatatu Networks
Suite 400, 303 Moodie Dr.
Ottawa, Ontario K2H 9R4
Canada

EMail: hadi@mojatatu.com