Light-weight Flow Accounting Protocol Specification
Version 5.0
<draft-riverstone-lfap-00.txt>

Status of this Memo

This document is an Internet-Draft and is subject to all provisions of Section 10 of RFC2026.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF), its areas, and its working groups. Note that other groups may also distribute working documents as Internet-Drafts.

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."

The list of current Internet-Drafts can be accessed at http://www.ietf.org/1id-abstracts.html

The list of Internet-Draft Shadow Directories can be accessed at http://www.ietf.org/shadow.html

Abstract

The Lightweight Flow Accounting Protocol (LFAP) is an operations and management protocol that provides detailed information about IPv4 and IPv6 packets traversing a network element. The primary function of LFAP is the reliable delivery of large volumes of packet header derived information from a network element, termed Connection Control Entity (CCE) to a Flow Accounting Server (FAS) of current or historical flows for billing, capacity planning security, or traffic engineering purposes.
1. Introduction

This document specifies Lightweight Flow Accounting Protocol (LFAP) version 5. It replaces an earlier specification LFAP version 1 as described in RFC 2124 [1]. It is intended to be deployed in a network element such as a bridge or router as well as in a passive network element such as a probe. This protocol specification and its companion document "LFAP Data Definition Specification" [2] define a method of flow export such that data collected may be used for the purposes of billing, application diagnostics, traffic-engineering, security and capacity planning.

LFAP defines two roles, one called Connection Control Entity and the other Flow Accounting Server. Network elements that employ LFAP for flow export are termed Connection Control Entities (CCE). Host systems that receive and store LFAP data are termed Flow Accounting Servers (FAS).

LFAP uses TCP [3] as its transport protocol. TCP MUST be present in all network elements and hosts where LFAP is be employed. TCP meets LFAP’s transport requirements for sequenced flow-controlled reliable delivery of large volumes of data. LFAP has been allocated TCP port 3145 [IANA-WKP] for establishing its connections. A CCE will establish a connection with a FAS in what is known as a "push" model of data delivery.

This document is the first of three documents that describe LFAP version 5. The other documents describe the encoding scheme for data transmitted (LFAP Data Definition Specification[2]), and the SNMP Management Information Base (MIB) Module to monitor LFAP at a CCE or a FAS. Please send comments to the LFAP mailing list on http://www.nmops.org (slate@nmops.org).

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 RFC 2119.
2. Summary of Operation

Upon initialization, a CCE initiates a TCP connection to a FAS. The CCE MUST be able to reach the FAS using standard IP connectivity. A CCE is configured with one or more FAS IPv4 or IPv6 addresses.

After the TCP session is opened between the CCE and FAS, the FAS enables a Message Response Timer that resets when a valid message is received. The CCE sends a Version Request (VR) message specifying the LFAP protocol version the CCE wishes to use. If the version is supported the FAS responds with a VRA and a status of SUCCESS otherwise it returns its highest supported version. Once the protocol version is established, the CCE next sends a Connection Request (CR) with the necessary information to make an authorization check. The FAS sends a Connection Accepted Notification (CAN), a Connection Rejected Notification (CRN) or a Disconnect Request (DR). A DR message may also be sent at any point after the session is established. When ready the FAS sends a Flow Export Ready message (FER).

Either the CCE or the FAS may initiate a Administrative Request (AR) message any time after the CAN message is received by the CCE. The requested entity (FAS or CCE) replies with an Administrative Request Acknowledge (ARA). The ARA is used to communicate status information or any error conditions for a given AR.

Data transfer begins with Flow Accounting Request (FAR) messages sent by a CCE to the FAS. These messages are sent when the CCE identifies a new IP microflow. Information about the flow that does not change over the life of the microflow such as the Source IP and Destination IP is sent in the FAR message. Data, which changes over the life of the flow such as the number of bytes and packets, are sent in Flow Update Notification (FAR and FUN message contents are defined in the companion document "LFAP Data Definition Specification" [2]).

The FAS sends a Keep Alive (KA) message so the CCE can determine if the connection or the end process/task is still intact. These messages are sent at a regular interval, between 1 second and MAX_INT seconds. The CCE should treat any message received from the FAS as having the same semantics as receiving the KA message. If the KA Wait Timer expires before any message is received from the FAS, then the CCE MUST terminate the TCP connection and immediately enter Connect State as described in section 6.2.
The following diagram represents which messages each side of the LFAP protocol sends.
3. Message Format


The general message format for all LFAP messages is as shown above. This is version 5 and Op Codes are as follows:

- VR = 1 # Code will never change in future revisions of LFAP
- VRA = 2 # Code will never change in future revisions of LFAP
- CR = 3
- CAN = 4
- CRN = 5
- FER = 6
- FAR = 7
- FUN = 8
- AR = 9
- ARA = 10
- KA = 11
- DR = 12

The Status field serves as a Status on the overall message.

STATUS:
- SUCCESS = 1
- VERSION = 2 Used by the FAS to indicate it does not support the version of LFAP Used by the CCE.

Reserved is reserved for future expansion and must be zero. The Message Length excludes the 8 octets of the message header. Message ID is used to associate each original message with its corresponding response and must be unique for the combination of sender and responder while an original message is pending.
4. Information Elements (IE’s) Formats

IE fields consist of 2-octet TYPE, 2-octet LENGTH and variable length VALUE sub-fields. All IE’s are even multiples of 4 octets in length, left aligned and zero filled if necessary. Length is computed excluding the 4 octet TYPE and LENGTH fields.

Individual IE’s necessary for data formatting or to establish and maintain a connection between the switch CCE and a FAS server are described here. Additional IE’s which may be sent are described by the LFAP Data Definition Specification[2]. IE type codes 1 - 64 are reserved for use by the LFAP Transport specification. IE type codes 65 - 60000 are used for defining additional IE’s in the LFAP Data Definition Specification. Type codes 60001 - 65535 are reserved for future use. A type code of zero "0" is invalid.

```
  0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|           TYPE =              |            LENGTH =           |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                                                               |
~                             Data                              ~
|                                                               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

4.1 FAS IP Address IE

This is the IP Address of a device running a Flow Accounting Server that may or may not be reachable from the FAS. The CCE uses this IE in the Connection Request message. IE format is:

```
  0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|           TYPE = 1            |             LENGTH            |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|     Address Family Number     |         Address Length        |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                                                               |
~                             Address                           ~
|                                                               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

The Address Length field contains the length of the address excluding any pad of zeros used to align the address field.

Address Family Numbers include:

1 = IP (IP Version 4)  as specified in RFC 1700
2 = IP6 (IP Version 6) as specified in RFC 1700
4.2 Multiple Record IE

The Multiple Record IE is composed of 4 parts. The record descriptor, fixed information, record format IEs and individual records. The record descriptor consist of two fields the first field is the length of the fixed information field. The second is the length of the Record format section. The fixed information is the IE’s that apply to all the records that follow. The Record Format is the list of IE’s that make up each record. The individual record section contains the individual records that are being reported in the format given by the Record Format section.

```
0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|           TYPE = 2            |          LENGTH               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|  Length of fixed Information  |   Length of Record Format     |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                                                               |
| Fixed Information                                              |
|                                                               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                                                               |
| Record Format                                                 |
|                                                               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                                                               |
| Individual Record (1)                                         |
| Individual Record (2)                                         |
| Individual Record (3)                                         |
| .                                                               |
| .                                                               |
| Individual Record (n)                                         |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

5. Individual Message contents

5.1 Version Request (VR) Message

VR messages are sent only by the CCE to request a session with the FAS under a specific version of the LFAP protocol. The VR message may only be sent by a CCE while is in the Connect State. It must be sent within the time specified by the Message Response Timer. A CCE MUST NOT request the same version within a single TCP session.
otherwise the FAS MUST disconnect the TCP session and the counter Session Establishment Errors will be incremented.

5.2 Version Request Acknowledge (VRA) Message

VRA messages are sent only by a FAS in response to a VR when a new TCP connection has been established and the FAS is in the Connect State. The Message ID field is set to the value of the received VR message. If a VR is not received within the given Message Response Timer the TCP session is dropped and counter Session Establishment Errors is incremented.

If the FAS supports the version requested in the VR message it also copies the version number from the VR message to the VRA message and status is set to SUCCESS in the VRA message. Upon the CCE receiving this message protocol version negotiation is complete and the CCE and FAS both transition to Connected State.

If the FAS does not support the version requested in the VR message, the FAS places the highest version it supports in the message header and sets the STATUS field to VERSION, which indicates a version mismatch. The Version Mismatch Errors is incremented.

The CCE MUST NOT request a version higher than the one returned in the VRA otherwise the FAS MUST disconnect the TCP session and the counter Session Establishment Errors will be incremented.

5.3 Connection Request (CR) Message

A single CR message is sent only by the CCE to request an authorized connection. CR messages contain a mandatory IE that is the FAS IP Address IE. There is one IE for each FAS the CCE has been configured with in the order that the CCE will attempt to connect with. Only one CR message is sent by a CCE and only when it is in the Connected State. Status field must be set to SUCCESS.

If the FAS does not receive a CR message within the timeout specified by the Message Response Timer, the TCP session MUST be disconnected and the Session Establishments Rejected counter incremented.

Mandatory IE’s

FAS IP Address IE - the multiple record IE MUST be used to list more than one FAS IP address
5.4 Connection Accepted Notification (CAN) Message

A single CAN message is sent only by a FAS when it is in the Connected State and only in response to a CR message. A CAN message is sent by the FAS to indicate the CCE is authorized to connect to this FAS. CAN messages do not contain any mandatory IE’s. Status is set to SUCCESS.

If the CCE does not receive a CAN (or CRN/DR) message within the timeout specified by the Message Response Timer, the TCP session MUST be disconnected and the Session Establishments Rejected counter incremented.

5.5 Connection Rejected Notification (CRN) Message

A single CRN message is sent only by a FAS when it is in the Connected State and only in response to a CR message. A CRN message is sent by the FAS to indicate the CCE is not authorized to connect to this FAS. CRN messages do not contain any mandatory IE’s. Status is set to SUCCESS. A CCE will terminate the TCP connection upon receipt. The FAS will terminate the TCP connection after the message was transmitted successfully.

If the CCE does not receive a CRN (or CAN/DR) message within the timeout specified by the Message Response Timer, the TCP session MUST be disconnected and the Session Establishments Rejected counter incremented.

5.6 Flow Export Ready (FER) Message

A single FER message is sent only by a FAS and only when it is in the Data Negotiation state. FER messages are sent by the FAS to indicate the FAS is ready to accept FAR and FUN messages. Only AR, ARA and FER messages are allowed while in the Data Negotiation state. FER messages do not contain any mandatory IE’s. Status is set to success.

If the CCE does not receive AR, ARA or an FER message within the timeout specified by the Message Response Timer, the TCP session MUST be disconnected and the Session Establishments Rejected counter incremented.

5.7 Flow Accounting Request (FAR) Message

FAR message contents are described in the LFAP Data Definition RFC. FAR messages are only sent by a CCE when the CCE is in Send State. The Status field is set to SUCCESS in FAR messages.
5.8 Flow Update Notification (FUN) Message

FUN message contents are described in the LFAP Data Definition RFC. FUN messages are only sent by a CCE when the CCE is in Send state and are sent only after a FAR has been sent for any data contained within the FUN. Status is set to SUCCESS in FUN messages.

5.9 Administrative Request (AR) Message

AR’s can be sent by CCE or FAS at any time when in the Send or Data Negotiation State. Status is set to SUCCESS in AR messages. AR message contents are described in the LFAP Data Definition RFC.

5.10 Administrative Request Acknowledge (ARA) Message

ARA’s can be sent by CCE or FAS at any time when in the Send or Data Negotiation State in response to a specific AR. Status reflects the result of the corresponding AR. Message ID is copied from the AR message. ARA message contents are described in the LFAP Data Definition RFC.

5.11 Keep Alive (KA) Message

KA messages do not contain any IE’s. The FAS sends a Keep Alive (KA) message so the CCE can determine if the connection is still intact while in the Send State. These messages are sent at a regular interval, between 1 second and MAX_INT seconds. The CCE should treat any message received from the FAS as having the same semantics as receiving the KA message. Status is set to SUCCESS in KA messages.

5.12 Disconnect Request (DR) Message

A DR message is sent only by the FAS to request the CCE disconnect. It can be sent when the FAS is in the Send State or Connected State. DR messages may contain one optional IE which is the FAS IP Address IE. The FAS may suggest an alternative FAS by supplying the FAS IP Address IE. Status is set to SUCCESS in DR messages.

The Optional IE that can be transmitted in a DR message is a:
- FAS IP Address IE
5.13 Message ID Error

If the same Message ID is received in consecutive messages, all but the first one is ignored and the counter Invalid Messages Received is incremented.

6. Session Establishment

The CCE initiates a TCP connection to a FAS. Session establishment happens in three phases following TCP connection establishment. In the first phase the protocol version is exchanged and verified. In the second phase, the FAS allows or denies the connection. The FAS may deny the connection for a variety of reasons. For example, the CCE may not be authorized to connect to this FAS. The third phase allows setup information to be exchanged before flow export begins. Messages received outside their specified state cause the session to be aborted.

6.1 Protocol Negotiation

In this phase the FAS and the CCE negotiate the version of the LFAP protocol to be used. The CCE initiates the process by sending a VR message with the highest version it supports in the message header. If the FAS supports the version it MUST respond with a VRA with the same version number and a status of success. For example, a CCE may request version 3 and the FAS may support version 3, 4, and 5. In this case the FAS will respond with a VRA message with a version of 3 and status of success.

If the FAS does not support the version it MUST send a VRA message with the highest version it supports and the status set to VERSION. At this point the CCE MAY disconnect and try another FAS. Or, if the CCE supports the version indicated by the FAS, the CCE MAY send a new VR message with the indicated version number. For example, if the CCE supports version 5 and 6, it will send a VR message to the FAS with version 6. Since the FAS supports version 3, 4, and 5, it responds with a VRA message with a version of 5 and a status of VERSION (indicating it does not support version 6). The CCE may now send another VR message, this time with version 5.

If the CCE supports a lower version than indicated in the ARA, it MAY also send a VR with its lower version. The same version MUST NOT be requested twice in a single TCP session. Additionally, once the CCE receives an ARA it MUST NOT request a version higher than the one received in the ARA.
6.2 Connection Establishment

Once the protocol version negotiation is complete, the FAS then will either accept or reject the connection. First, the FAS MUST receive from the CCE a Connection Request (CR) message with the necessary data. In LFAP Version 5 the data consists of a list of FAS IP addresses. After receiving the CR message the FAS MUST send a Connection Accepted Notification, a Connection Rejected Notification or a Disconnect Request. The DR message MAY contain a server that should be used instead.

6.3 Data Negotiation

During this phase only AR and ARA messages may be exchanged. The LFAP Data Definition specification [2] describes the message contents. When complete, a FER message is sent from the FAS to the CCE and the LFAP session is then fully established.

The normal sequence of events when a CCE and FAS start to talk would be as follows:

- The CCE opens a TCP socket
- The CCE sends a VR message with the desired protocol version
- The FAS sends a VRA response with its desired version.
- The CCE sends a CR message assuming the two agree on a version
- The FAS sends a CAN message, Assuming the CCE is authorized
- The FAS and CCE exchange AR and ARA messages
- The FAS sends a FER

At this point the session is established and FAR/FUN messages will begin along with Keep Alive messages.

6.4 State Diagrams

Two state diagrams, one from the CCE perspective and one from the FAS perspective follow which describe the three phases of the connection setup:
CCE State Machine for Connection Setup

Begin  +-------------+  * master on/off switch
       |             |  * FAS IP address list
Init State  * server time-out
             * keep alive
       +-------------+  * server retry interval
       +-------------+  * select first FAS IP address

       +-----+-----+  * if needed, TCP connect
       |     |     |  +-------V-------+
       |     |     |  |               |  * Send VR
       |     |     |  |               |  +------>+-------+-----+
       |     |     |  |               |  | (A)  |   |         - receive VRA from FAS
       |     |     |  |               |  +-------V------V------+
       |     |     |  |               |  | Protocol |  * evaluate VERSION
       |     |     |  |               |  | Negotiation |  (B) |  State
       |     |     |  |               |  |               |  +-------V-------+
       |     |     |  |               |  * VRA Version matched
       |     |     |  +-----+-----+  * Send CR
       |     |     |  |     |     |     |  +-------+-------+  * Received CAN
       |     |     |  |  (C)  |               |  +-------V-------+
       |     |     |  +-----+-------+  * Exchange AR/ARA
       |     |     |  |     |     |     |  +-------V-------+
       |     |     |  |  (D)  | Negotiation |  * Received FER
       |     |     |  +-----+-------+
       |     |     |  |     |     |     |  +-------V-------+
       |     |     |  |  (E)  |               |  * Send FAR/FUN
       |     |     |  +-----+-------+  * Send/Receive AR/ARA
       |     |     |  |               |  * Receive KA/DR
       +-----+-----+-----+-----+
(A) The error conditions that occur while the CCE is in the Connect state include:
- No response to VR message, message Response Timer expired
- TCP connection failed to be established.
- Protocol violation (the wrong message type was received), disconnect and increment counter.
- Corrupted message received. Increment counter.

(B) The error conditions that occur while the CCE is in the Protocol Negotiation State include:
- NOT Supported Version - FAS and CCE do not agree on LFAP version number. CCE can either try sending a VR with the version the FAS sent if it is supported or it can disconnect the TCP session and try another unique IP address of a FAS.
- TCP session disconnects - try next server in list
- Protocol violation (the wrong message type was received), disconnect and increment counter.
- Corrupted message received. Increment counter.

(C) The error conditions that occur while the CCE is in the Connected state include:
- No response to CR message, Message Response Timer expired then drop connection, increment counter.
- Receive CRN message, disconnect current TCP session, select next FAS IP address. If next FAS IP address in list is the first one, wait for period specified in Retry FAS List Timer.
- DR message received from FAS. Disconnect and try next IP server in list or use the server specified if in current list
- TCP session disconnects
- Protocol violation (the wrong message type was received), disconnect and increment counter.
- Corrupted message received. Increment counter.

(D) The error conditions that occur while the CCE is in the Data Negotiation state include:
- No AR, ARA or FER message, Message Response Timer expired then drop connection, increment counter.
- TCP session disconnects
- Protocol violation (the wrong message type was received), disconnect and increment counter.
- Corrupted message received. Increment counter.

(E) The error conditions that occur while CCE is in Send State include:
- DR message received from FAS. Disconnect and try next IP server or use server specified if in current list.
- KA Timer expired or ARA message not received to last CCE issued AR, disconnect and try next server in list.
- TCP session disconnects - try next server in list
- Protocol violation (the wrong message type was received), disconnect and increment counter.
- Corrupted message received. Increment counter.
FAS State Machine

Begin +-------------+
    |             |
    | Init State  |
    +-------------+

    +------>+-------V-------+
    | +---->|               |  * if needed, Await TCP connection
    | |  +->| Connect State |  * Await VR message
    | |  |+>|               |
    | |  || +---+---+-------+
    | |  ||(A)  |   |         - receive VR from CCE
    | |  |+-----+   |
    | |  |          |
    | |  |  +-------V-------+
    | |  |  |  Protocol     | * evaluate VERSION
    | |  +--+  Negotiation  |
    | |     +-------+-------+
    | |             |         - Send VRA (version matched)
    | +-----+ Connected State |
    |    (C) |               |
    |       +-------+-------+
    |               |         - Send CAN

    +------>+-------V-------+
    | +---->|               |  * Await CR
    | |  +->| Connected State |
    | |  |+>|               |
    | |  || +-------V-------+
    | |  |  |  Data        |  * Exchange AR/ARA messages
    | |  +--+  Negotiation |
    | |     +-------+-------+
    | |             |         - send FER

    +------>+-------V-------+
    | +---->| Send State    |  * Receive FAR/FUN
    | |  +->|               |  * Send/receive AR/ARA
    | |  |+>|               |  * Send KA messages
    | |  +-----+           |  * Can send DR

Calato, MacFaden [Page 16]
(A) The error conditions that occur while the FAS is in the Connect State include:
- No VR message, Message Response Timer expired then drop connection, increment counter.
- TCP session disconnects
- Protocol violation (the wrong message type was received), disconnect and increment counter.
- Corrupted message received. Increment counter.

(B) The error conditions that occur while the FAS is in the Protocol Negotiation State include:
- NOT Supported Version - The FAS sends a VRA with its highest supported version.
- TCP session disconnects
- Protocol violation (the wrong message type was received), disconnect and increment counter.
- Corrupted message received. Increment counter.

(C) The error conditions that occur while the FAS is in the Connected state include:
- No CR message, Message Response Timer expired then drop connection, increment counter.
- Send CRN message, then disconnect TCP session.
- Send DR message, then disconnect TCP session.
- TCP session disconnects
- Protocol violation (the wrong message type was received), disconnect and increment counter.
- Corrupted message received. Increment counter.

(D) The error conditions that occur while the FAS is in the Data Negotiation state include:
- No ARA message in response to AR, Message Response Timer expired then drop connection, increment counter.
- TCP session disconnects
- Protocol violation (the wrong message type was received), disconnect and increment counter.
- Corrupted message received. Increment counter.

(E) The error conditions that occur while FAS is in Send State include
- DR message sent, disconnect TCP session.
- No ARA message in response to AR, Message Response Timer expired then drop connection, increment counter.
- TCP session disconnects.
- Protocol violation (the wrong message type was received), disconnect and increment counter.
- Corrupted message received. Increment counter.
6.5  Protocol Errors

Any message received by either the CCE or FAS outside its specified state will cause the TCP connection to be disconnected and the Protocol Error counter will be incremented.

<table>
<thead>
<tr>
<th>State</th>
<th>Messages Allowed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Connect</td>
<td>VR, VRA</td>
</tr>
<tr>
<td>Protocol Negotiation</td>
<td>VR, VRA</td>
</tr>
<tr>
<td>Connected</td>
<td>CR, CAN, CRN, DR</td>
</tr>
<tr>
<td>Data Negotiation</td>
<td>AR, ARA, FER</td>
</tr>
<tr>
<td>Send</td>
<td>FAR, FUN, AR, ARA, KA, DR</td>
</tr>
</tbody>
</table>

7.  Error Handling

The receiver will ignore corrupted message contents — receipt of a LFAP message that cannot be parsed. An error should be logged if possible and the Corrupted Message counter should be incremented. Other error handling is naturally associated with each message and is listed in the following sections.

7.1  VRA Related Error Handling

If the FAS does not support the version requested in the VR message, the FAS places the highest version it supports in the message header and sets the STATUS field to VERSION, which indicates a version error.

7.2  ARA Related Error Handling

Any AR messages with commands that result in an error may have the error information returned via the ARA.

8.  Statistics

Each implementation of the protocol must maintain various statistics on the operation of the protocol. Statistics common to both the CCE and FAS perspective are described first followed by statistics that are role dependent.

8.1  Common Statistics

A FAS or CCE will maintain the following 32 bit counters for each of the following messages that were sent and received correctly: VR, VRA, CR, CAN, CRN, FER, FAR, FUN, AR, ARA, KA
The following 32 bit counters will be kept:

Version Mismatch Errors
Number of times CCE or FAS did not agree on LFAP version.

Session Establishments Accepted
Number of times either a CCE or FAS reached the Send State as indicated in the state diagrams. In summary, a CCE receives a FER message or a FAS sends an FER message.

Session Establishments Rejected
Number of times a CCE has received a CRN message or a FAS has sent a CRN message.

Lost Contact
Number of times a CCE or FAS lost the session while in the Send State. For a CCE, this includes Message timeout, Lost TCP session, or KA timeout. For a FAS, it includes Message timeout and Lost TCP.

Active flows
On a CCE, number of current flows being tracked. On the FAS this counter is kept per CCE and is the number of flows known to the FAS for that CCE (note - on failover the FAS does not know about any active flows on the CCE). This number is computed as follows. Increment when FAR messages are created on the CCE or received by a FAS. Decrement when a FUN with state of inactive is created on a CCE or received by a FAS. If a FAS receives a FUN for a flow which is unknown, it increments the active counter first and adds it to the known list.

Bytes Sent
Number of bytes excluding written to TCP session since TCP session was opened. FAS keeps this counter per CCE.

Packets Sent
Number of bytes excluding written to TCP session since TCP session was opened. FAS keeps this counter per CCE.

Peak active flows
Watermark of most flows managed concurrently by the CCE. The FAS keeps one counter for all currently connected CCEs.

Corrupted messages received
A count of any message received that could not be decoded.

Protocol Violations
A count of valid messages received but considered invalid in the current CCE or FAS state.
8.2 CCE Statistics

Each of the following statistics must be maintained by a CCE as a 32 bit counter. These must be kept for the device and optionally per configured LFAP server.

Session Establishment Errors
Number of times TCP connection was tried but failed.

Failed to establish a session to any server
A count of the number of times the CCE attempted to connect to each of the FAS it was configured with in succession without success. It includes cases where a DR with an alternate FAS was also tried.

Session establishment failed due to version mismatch
A count of the number of times the CCE negotiated LFAP version with a FAS without agreement.

Dropped messages
A count of the number of FAR or FUN messages the CCE could not be transmitted to the FAS while not in Send State.

Dropped messages while connected
A count of the number of FAR or FUN messages the CCE could not be transmitted to the FAS while in Send State.

Number of bytes not accounted for due to dropped messages
A count of the number of bytes accounted for in FUN messages being dropped that CCE could not transmit to the FAS while in Send State.

Number of packets not accounted for due to dropped messages
A count of the number of packets accounted for on the wire that CCE could not transmit to the FAS while in Send State.

A timestamp will be kept for the following events: This timestamp is the time in hundredths of a second since the system was last restarted.

Time of last session status change in our out of the Send State
Time of last dropped message
Time of last dropped message while connected
Time of peak active flows

8.3 FAS Statistics

Each of the following statistics is a 32 bit counter
Number of CCE’s connected
Number of active Flows being tracked by all CCEs
Timestamp of last Error to store any received FAR or FUN
Number of accounting bytes received but not stored
Number of accounting packets received but not stored

These stats are kept per CCE
Number of active flows
Timestamp when connection entered Send State
Version of protocol selected
Time left till next Keep Alive

9. Security Considerations

The LFAP protocol carries two types of sensitive data.
1) Micro-flow information gleaned from datagram headers such as IP addresses, protocol, and TCP/UDP source/dest port numbers.
2) Information used for billing purposes, which in addition to flow information, includes bytes and packets transmitted.

Four levels of security are defined.

Level I - No Security

The intent of level I is to allow the protocol to run with no security overhead when there is no perceived threat or when some other mechanism is in place.

Level II - No Unauthorized Sessions

The intent of level II is to protect against the establishment of a session with an impersonated CCE or FAS.

Level III - Secure Transport

The intent of level III to provide a secure transport for LFAP messages. Types of attacks to be prevented are listed below.

Level IV - Encrypted Transport

The intent of level IV is to prevent the reading of LFAP messages by a third party.

The following attacks should be considered when evaluating any security solution.

1) Message altering
2) Message replay
3) Message Insertion
4) Message reading (snooping)
5) Impersonation of a CCE or FAS
9.1 Lower Layer Security

Since LFAP runs over TCP, a lower layer protocol may provide the necessary security. Two such lower layer protocols are IPSec[4] and Transport Layer Security [5](TLS a.k.a. Secure Socket Layer – SSL). Either of these security protocols operating under the LFAP protocol can adequately accomplish the defined security levels and address the attacks mentioned.

9.2 LFAP Security

In addition to IPSec and TLS to provide protection against the attacks outlined, the base protocol has provisions to accomplish a minimal base level of security within the protocol itself. LFAP implementations MAY implement one or more levels of security via the LFAP protocol. If a lower layer protocol is in use the CCE and FAS can be set to level I.

9.2.1. Extended LFAP Header

Additional fields are needed in the LFAP message header to provide the necessary security mechanisms. The enhanced header format is as follows:

```
+-----------------+-----------------+-----------------+-----------------+-----------------+
| Version         | Op Code         | Reserved        | A | E | Status        |
+-----------------+-----------------+-----------------+-----------------+-----------------+
| Message ID      | Message Length  |                 |
+-----------------+-----------------+-----------------+-----------------+-----------------+
| Authentication Data |
+-----------------+-----------------+-----------------+-----------------+-----------------+
| Sender ID       |
+-----------------+-----------------+-----------------+-----------------+-----------------+
| Session ID      | Sequence ID     |
+-----------------+-----------------+-----------------+-----------------+-----------------+
| Sequence ID     |
```

Version, Op code, Status, Message ID and Message Length are the same as described previously. Note - message Length excludes the first 8 octets of the header, but includes the remaining header octets when the extended header is used.

A = 1 – the message is authenticated
E = 1 – the message is encrypted

Authentication Data - a 96 bit field containing the Authentication

Sender ID - the router ID for the CCE or the IP address for the FAS. This value is used to determine which key to use for authentication and encryption.

Session ID - starting number is picked at random, subsequent IDs are increased by 1. Zero "0" is invalid.

Sequence ID - picked at random using the lower 36 bits of sequence ID. Each subsequent message increases it by 1. Zero "0" is invalid.

When calculating Authentication Data, the field is set to zero in the message and the value is calculated over the entire message. Thus, all fields are protected. Messages with an invalid Authentication Data field are discarded.

The sender ID and Session ID must match the corresponding IDs established for the session or the message is discarded.

The Sequence field is incremented by the sender, the receiver discards messages with an ID not equal to the last sequence ID + 1. Only a successfully authenticated message causes the last sequence ID to be updated.

An invalid Authentication Data field or incorrect ID fields (sender ID, session ID and sequence ID) is considered to be a "Security Violation". Messages that result in a security violation are discarded. A new statistic is added called Security Violations. It is incremented each time a Security violation is encountered.

The sequence ID is large enough that should the value roll over, the amount of time elapsed would make any replay attack unlikely. Note - on rollover, zero "0" MUST NOT be not used.

9.2.2. Extended Header ID Exchange

Some minor changes are necessary to exchange the IDs needed for the extended LFAP header. The changes are outlined below.

When the CCE sends a VR with A = 1 it fills in the session ID and sequence ID with the values the FAS can use in subsequent messages to the CCE.

When the FAS sends a VRA with A = 1 it fills the session ID and
sequence ID with the values the CCE can use in subsequent messages to the FAS. If the version requested is not supported by the FAS it returns the VRA as if the VR did not have the "A" flag set.

At this point, the CCE sets its Last Received Sequence ID to the value it, the CCE, set in the VR message. Similarly the FAS sets its Last Received Sequence ID to the value it, the FAS, set in the VRA.

From this point on when the FAS wishes to send a message which is to be authenticated it sets the "A" flag in the header to 1 and it sets the sender ID to its own ID. It sets the Session ID to the ID supplied in the VR message. It then sets the Sequence ID to 1 greater than the one it received in the VR message. Subsequent messages continue to increment the last Sequence ID by 1.

Similarly when the CCE wishes to send a message that is to be authenticated it sets the "A" flag in the header to 1 and sets the sender ID to its own ID. It sets the Session ID to the ID supplied in the VRA. It then sets the Sequence ID to 1 greater than the one it received in the VRA message. Subsequent messages continue to increment the last Sequence ID by 1.

9.2.3.  Level I Security

Level I security is the default. All implementation have this level of security by definition.

9.2.4.  Level II Security

To achieve level II security the LFAP Session establishment is modified as follows.

CCE sends VR as described in the previous section
FAS sends VRA as described in the previous section
CCE sends CR with A = 1 and the Session ID from the VRA. The Sequence ID from the VRA is incremented and used
FAS sends CAN with A = 1 and the Session ID from the VR. The Sequence ID from the VR is incremented and used
CCE sends FAR/FUNs with A = 0
FAS sends message with A = 0

This will approach will not allow impersonation of a FAS or CCE since they will not have the necessary secret key to compute the Authentication Data field. It is however subject to attacks such as replay of FAR/FUN messages after session establishment is complete.

Note - Messages received before the session is established cause the connection to be aborted. Thus an attacker cannot simply start by sending FAR/FUN messages.
9.2.5. Level III Security

Level III security is built on top of Level II security. The only change is that with Level III security, all messages set the "A" flag to 1 and fill in the extended LFAP header. This level of security protects against each form of attack listed except for message reading.

9.2.6. Level IV Security

Level IV security SHOULD use Level II security. To enable Level IV security the "E" flag is set to 1 in the LFAP header. If the "A" flag is 0 encryption begins after the LFAP header. If the "A" flag is set to 1 then encryption begins after the Authenticated Data field in the extended header. Note - if the "E" flag is set in the VR and the FAS does not support the version, it sends the VRA as of the "E" flag was not set.

Encryption is done using DES-CBC [FIPS-46-3][FIPS-74][FIPS-81].

9.2.7. Security Configuration

Both the FAS and CCE must be configured with the expected security level, the default is level I. If the CCE or FAS does not receive the expected level of security, it MUST disconnect the TCP session and increment the Security Violations counter. Additionally the way the FAS and CCE obtain the necessary keys is undefined.

9.3 Denial of Service

Since LFAP runs over TCP, both the CCE and FAS need to be protected from TCP denial of service attacks. Defining how to protect against TCP attacks is outside the scope of this document.

Additional steps can be taken to help protect against flooding of invalid messages on an existing connection. If more than "n" Security Violations occur or "n" corrupted messages are received, the TCP connection MAY be terminated. In this case the previous state diagrams are modified to include one extra error event triggered when "n" is reached. "n" MUST be configurable from 1 to MAX_INT.

9.4 Message Loss

There is one additional type of attack that should be considered and that is an attack that results in message loss.

There are a variety of ways that an attacker can cause messages to be lost. For example, the message may be altered in such a way that the application can no longer make use of the message when it is
delivered. The attacker may cause TCP to abort the connection or cause the connection to get backed up. In such cases the data must be stored by the application and re-sent if necessary.

There are two issues that make Message Loss a more difficult problem to solve.

1) Identifying duplicate flow information when data is re-sent.  
2) Limited storage space on most CCE devices.

The first problem poses technical challenges given the distributed nature of the LFAP protocol. A single CCE can fail-over to several servers and thus those servers must coordinate to identify and eliminate any duplicate records. The solution must also take into consideration the enormous volume of data generated by each CCE and the practical implications of implementing any given solution.

Assuming the first problem is solved, the second problem will limit its effectiveness under a sustained attack. Once the device runs out of storage space, subsequent messages are not recoverable. This makes the cost benefit ratio for solving the duplicate data problem even less attractive.

Given the above consideration, there are no provisions in the LFAP protocol to solve the Message Loss issue. However, there is support to aide in quantifying the amount of loss sustained. Counters are defined for the CCE and FAS which track the amount of data counted and the amount of data lost. This will allow a management tool to analyze the information and determine the scope of loss.

9.5 Runtime Configuration Changes

If a session is established when the initial CCE configuration is changed, the session should only be dropped if the new FAS IP list does not contain the existing IP address or the master off switch has been applied.

10. Miscellaneous Considerations

The LFAP protocol may be used in a Network Address Translation (NAT) environment, where the FAS is on one side of the NAT and the CCE is on the other. There is however one caveat, if the FAS attempts to contact any of the FAS’s sent in the CR message, reachability is not guaranteed. Therefore an LFAP server must not depend reaching those FAS’s.
11. Author’s Addresses

Paul Calato
Phone: +1 (603) 557-6913
Email: calato@riverstonenet.com

Mike MacFaden
Phone: +1 (408) 878-6525
Email: mrm@riverstonenet.com

Riverstone Networks, Inc. is located at:
5200 Great America Parkway
Santa Clara, CA 95054
USA
12. References


[4] Several documents are used to describe the IPsec protocol suite. The interrelationship and organization of the various documents covering the IPsec protocol are discussed in RFC 2411, November 1998.


Expires December 2001