Network Working Group
INTERNET-DRAFT

Brian Bidulock
OpenSS7 Corporation

Expires in six months

January 2, 2003

Load Grouping Extension
for
Signalling User Adaptation Layers
<draft-bidulock-sigtran-loadgrp-01.ps>

Status of this Memo

This document is an Internet-Draft and is in full conformance with all provisions of Section 10 or RFC 2026. 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/ietf/1id-abstracts.txt

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

To learn the current status of any Internet-Draft, please check the Directories on ftp.is.co.za (Africa), nic.nordu.net (Europe), munnari.oz.au (Pacific Rim), ftp.ietf.org (US East Coast), or ftp.isi.edu (US West Coast).

Abstract

This Internet-Draft describes an extension parameter and procedure to the Signalling User Adaptation Protocols [IUA...TUA], based on the Stream Transmission Control Protocol (SCTP) [RFC 2960] which permits an Application Server Processes (ASP) to indicate its placement within a Load Group and permits an Signalling Gateway (SG) to distribute traffic over Load Groups under Application Server Process (ASP) control.

The described procedure provides for Override, Load-share and Broadcast traffic mode operation within a Load Group consisting of one or more Application Server Processes (ASPs) resulting in a mixture of traffic modes within an Application Server (AS). The parameters and procedures described here supplement the *UA Load Selection* [LOADSEL] extension parameters and procedures for improved distribution of traffic over ASPs and Signalling Gateway Processes (SGPs).

1. Introduction

1.1. Scope

This Internet-Draft provides the *Load Distribution* parameter and associated procedures in extension to the parameters and procedures of the Signalling User Adaptation Layers (UAs) [IUA...TUA], and the *UA Load Selection* [LOADSEL] extension, for the purpose of permitting Application Server Process control over grouping of ASPs into Application Servers as part of the procedures of these UA protocols.

This *Load Grouping* extension is intended to be compatible with UA implementations not supporting this extension.

1.2. Change History

Changes from Version 0.0 to Version 0.1:

- added this section,
- updated release version and dates,
- updated references,
- updated postscript diagrams,
- minor formatting changes,
- reworked the procedures section,

- added interworking rules,
- change *Load Selection* to *Load Selector* to match *LOADSEL* [LOADSEL],
- added examples,
- updated author's address.

1.3. Terminology

LOADGRP supplements the terminology used in the UA documents [IUA...TUA] and the *UA Load Selection* [LOADSEL] extension by adding the following terms:

Load Distribution (LD) – a *Traffic Mode Type* which is applicable within a Load Group.

Load Group (LG) – a group of ASPs that share the same traffic distribution within an Application Server. An ASP is permitted to belong to more than one Load Group for a given AS.

Load Selector (LS) – an identifier that uniquely identifies a Load Group within an Application Server. This identifier is only guaranteed unique within the scope of an Application Server and must be combined with a *Routing Context* or (set of) *Interface Identifier(s)* to uniquely define a Load Group at a Signalling Gateway.

Signalling User Adaptation Layer (UA) – one or more of the Stream Control Transmission Protocol (SCTP) [RFC 2960] Signalling User Adaptation Layers [IUA...TUA].

1.4. Overview

Existing UA procedures with regard to traffic distribution and ASP traffic management provide a mechanism to select the algorithm for coordinating state and distributing traffic over a number of Application Server Processes (ASPs) serving an Application Server. These existing procedures provide only simplified distribution approaches which are not amenable to large scale systems that need to adapt to dynamic traffic loading or need live reconfiguration for maintenance purposes.

LOADGRP extensions to the Signalling User Adaptation Layers [IUA...TUA] permits close control over the grouping of Application Server Processes serving an Application Server that provides the capability to group ASPs into load groups not existing in the current scheme.

Under the existing approach, each Application Server Process acts independently with respect to an Application Server. This approach does not provide for the grouping of ASPs into work groups, not does it provide coordinated control of the role of groups of ASPs serving an ASP. As a result, the existing scheme does not scale well in this regard.

LOADGRP provides for the grouping of ASPs into load groups with a traffic distribution mode (Load Distribution) within the load group that is independent of the Application Sever traffic mode.

1.4.1. Existing Traffic Distribution

Figure 1 illustrates the existing traffic distribution algorithm that is used across the Signalling User Adaptation Layers.

When an SGP distributes a Signalling User Adaptation Layer message toward the Application Server based on the *Routing (Link) Key*, it selects an ASP that is active for the AS according to a *Traffic Mode Type* that is associated with the AS. The *Traffic Mode Type* describes three general distribution algorithms: Override, Load-share and Broadcast.

The detailed actions taken for these distribution algorithms are described in Section 4 of the Signalling User Adaptation Layer specifications [IUA...TUA]; however, they can be summarized as follows:

Traffic Mode Type Override:– When distributing messages to an Override Application Server, the SGP selects the ASP which is active for the Application Server. In an Override Application Server, at most one ASP can be active for the AS at any given point in time. The active ASP for the AS is selected.

Traffic Mode Type Load-share:– When distributing messages to a Load-share Application Server, the SGP selects one of the ASPs that are active for the Application Server using an implementation dependent load-sharing algorithm based on some unspecified aspect of the traffic or static configuration data.

Traffic Mode Type Broadcast:– When distributing messages to a Broadcast Application Server, the SGP sends a copy of the message to each of the ASPs that are active for the Application Server. (The ASPs themselves decide which, if any, of the ASPs will process the message.)

In general, for the Signalling User Adaptation Layers, the *Traffic Mode Type* is a characteristic of the Application Server, and an Application Server can only have associated with it only one *Traffic Mode Type*, and thus, only one traffic distribution algorithm can be used across the ASPs that are serving a given Application Server.

LOADGRP enhances the traffic distribution algorithms of the existing Signalling User Adaptation Layers by introducing an additional level of distribution.

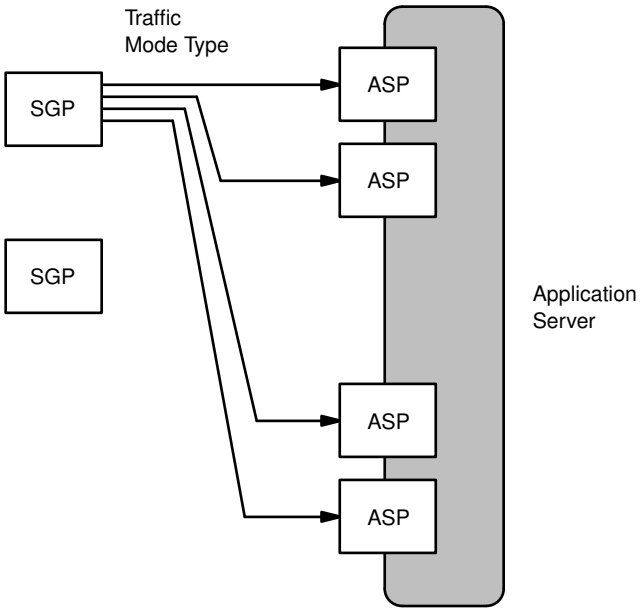


Figure 1. Existing Traffic Distribution

1.4.2. Extended Traffic Distribution

Figure 2 illustrates the extended traffic distribution algorithms that are used across the Signalling User Adaptation Layers as a result of the messages and procedures in *LOADGRP*.

LOADGRP introduces the concept of a Load Group. A Load Group is a logical grouping of Application Server Processes (ASPs) into traffic groups serving an Application Server. Signalling Gateway Processes (SGPs) distribute traffic first over *Load Groups* and then distribute traffic within the Load Group. Each Load Group describes and is identified by a *Load Selector* [LOADSEL] within the Application Server. The *Load Selector* identifies the traffic flows that will be distributed to an associated Load Group within an Application Server.

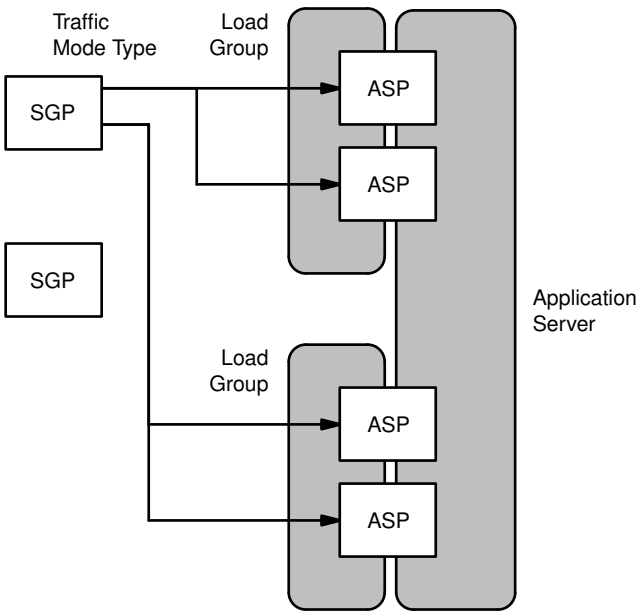


Figure 2. Load Group Traffic Distribution

When an SGP distributes a Signalling User Adaptation Layer message toward an Application Server based on the *Routing (Link) Key*, it first selects an Load Group that is active for the Application Server according a traffic distribution algorithm determined by the *Load Distribution* that is associated with the Application Server and the *Load Selector* position of the Load Group within the AS.

The *Traffic Mode Type* continues to describe three general distribution algorithms: Override, Load-share and Broadcast. The change in the behavior of the SGP when selecting an ASP for traffic distribution introduced by *LOADGRP* is that the SGP also takes into account the concept of a Load Group as identified within an AS by its *Load Selector*.

The extended procedures can be summarized as follows:

Traffic Mode Type Override:– When distributing messages to an Override Application Server, the SGP first selects the Load Group that is active for the Application Server. In an Override Application Server, at most one Load Group can be active for the AS at any given point in time. The active Load Group for the AS is selected.

Traffic Mode Type Load-share:– When distributing messages to a Load-share Application Server, the SGP selects one of the *Load Groups* that are active for the Application Server using an implementation dependent load-sharing algorithm based on the *Load Selector* [LOADSEL] associated with the Load Group.

Traffic Mode Type Broadcast:– When distributing messages to a Broadcast Application Server, the SGP sends a copy of the message to each of the *Load Groups* that are active for the Application Server. (The *Load Groups* themselves decide which, if any, of the *Load Groups* will process the message.)

After selecting an Load Group according to the *Traffic Mode Type* for the Application Server, the SGP then selects an ASP within the Load Group based on the *Load Distribution* that is associated with the Load Group. The *Load Distribution* describes the same three general distribution algorithms that are provided in the *Traffic Mode Type:* Override, Load-share and Broadcast. When selecting an active ASP within a Load Group, the procedures that the SGP will follow can be summarized as follows:

Load Distribution Override:– When distributing messages within an Override Load Group, the SGP selects the ASP which is active for the Load Group. In an Override Load Group, at most one ASP can be active for the Load Group at any given point in time. The active ASP for the Load Group is selected.

Load Distribution Load-share:– When distributing messages within a Load-share Load Group, the SGP selects one of the ASPs that are active for the Load Group using an implementation dependent load-sharing algorithm based on some unspecified aspect of the traffic or static configuration data.

Load Distribution Broadcast:– When distributing messages within a Broadcast Load Group, the SGP sends a copy of the message to each of the ASPs that are active for the Load Group. (The ASPs themselves decide which, if any, of the ASPs will process the message.)

The result of *LOADGRP* is that two levels of traffic distribution are provided for, permitting more flexible membership of ASPs serving Application Servers, and provides the Application Server Process with more control over the traffic patterns for which it is active.

1.5. Sample Configurations

Although *LOADGRP* does not restrict either *Traffic Mode Type* or *Load Distribution* to a static assignment, there are, for example, six (6) combinations of static *Traffic Mode Type* and *Load Distribution* assignment under this scheme. Not all of these combinations provide for interesting or useful configurations as follows:

Table 1. Sample Configurations

Mode	Traffic Mode Type	Load Distribution	Description
1	Override	Load-share	This mode permits a load-share group of ASPs to override another load-share group of ASPs.
2		Broadcast	This mode allows "mirrored" ASPs to override each other.
3	Load-share	Override	This mode allows ASPs to override each other within a traffic slot of a load-share group.
4		Broadcast	This mode permits "striping" and "mirroring" with load-sharing under ASP control.
5	Broadcast	Override	This mode permits a joining ASP to knock another ASP out of a Broadcast slot for an Application Server.
6		Load-share	This mode permits "mirroring" and "striping" including automatic load-sharing within a mirror image.

2. Conventions

The keywords **MUST**, **MUST NOT**, **REQUIRED**, **SHALL**, **SHALL NOT, SHOULD**, **SHOULD NOT**, **RECOMMENDED**, **NOT RECOMMENDED**, **MAY**, and **OPTIONAL**, when they appear in this document, are to be interpreted as described in [RFC 2119].

3. Protocol Elements

The following subsections describe the parameters which are added by *LOADGRP*, their format and the message in which they are used.

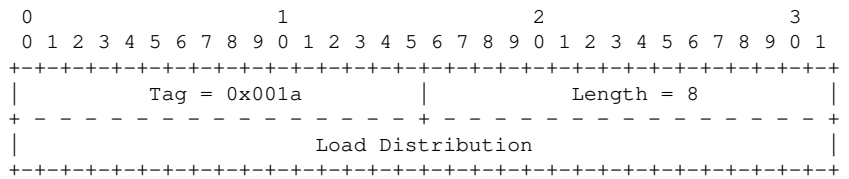
3.1. Parameters

LOADGRP adds one new parameter: the *Load Distribution* parameter.

3.1.1. Load Distribution

The *Load Distribution* parameter is used in the **REQ REQ**, **REQ RSP**, **ASPAC** and **ASPAC ACK** messages. It is used in conjunction with the *Traffic Mode Type* parameter [M2UA...TUA] and *Load Selector* parameter [LOADSEL] to further refine the traffic distribution algorithm for an ASP in a Load Group serving an Application Server. The *Load Selector* parameter identifies the Load Group for which the ASP is registering, activating or deactivating and the *Load Distribution* parameter identifies the traffic distribution that is to be used within the Load Group.

The *Load Distribution* parameter is formatted as follows:



EDITOR'S NOTE:– The parameter tag values shown as **`0x001a`** above will be assigned by IANA within the common parameter range of the SIGTRAN UAs and may change its value in further versions of this document.

The *Load Distribution* parameter contains the following fields:

Load Distribution field: n x 32-bits (unsigned integer)

The *Load Distribution* has the same purpose for the Load Group that the *Traffic Mode Type* has for the Application Server: it identifies the traffic distribution algorithm to be used within the Load Group. Valid values for *Load Distribution* are as follows:

1 Override

2 Load-share
3 Broadcast

3.2. Messages

LOADGRP adds the *Load Distribution* parameter as an **OPTIONAL** parameter to be used in conjunction with the common *Traffic Mode Type* in the following messages: [1]

REG REQ	Registration Request message
ASPAC	ASP Active message
ASPAC ACK	ASP Active Ack message

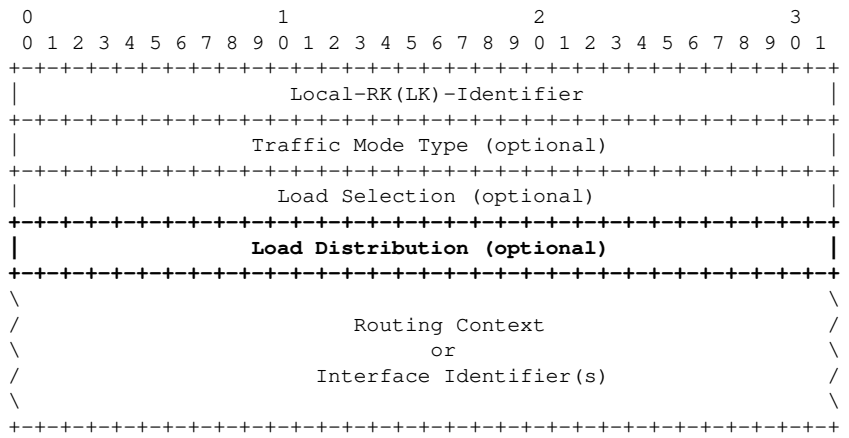
3.2.1. Registration Request (REG REQ)

LOADGRP supplements the **Registration Request (REG REQ)** message by permitting the following optional parameters to be included in the *Routing (Link) Key* [M2UA...TUA] parameter or *Link Key* [IUA...M2UA] parameter within the message:

Extension Sub-parameters	
Load Distribution	Optional

The *Load Distribution* parameter is used in the Routing (Link) Key to further refine the traffic distribution to be received by the registering ASP.

The format of the resulting *Routing Key* or *Link Key* parameter is as follows:



When an ASP wishes to register within a Load Group associated with an Application Server, it includes the *Load Selection* parameter and *Load Distribution* in the *Routing (Link) Key* for that Application Server in the **REG REQ** message.

The *Load Distribution* parameter indicates the traffic distribution method to be used within the Load Group as identified by the Load Selection.

No other changes to the **REG REQ** message, *Routing Key* or *Link Key* parameters format are provided by *LOADGRP*[1].

3.2.2. Registration Response (REG RSP)

LOADGRP adds the following *Registration Status* values to the *)Registration Status* field in the **REG RSP** message:

16 Error – Unsupported/Invalid Load Distribution

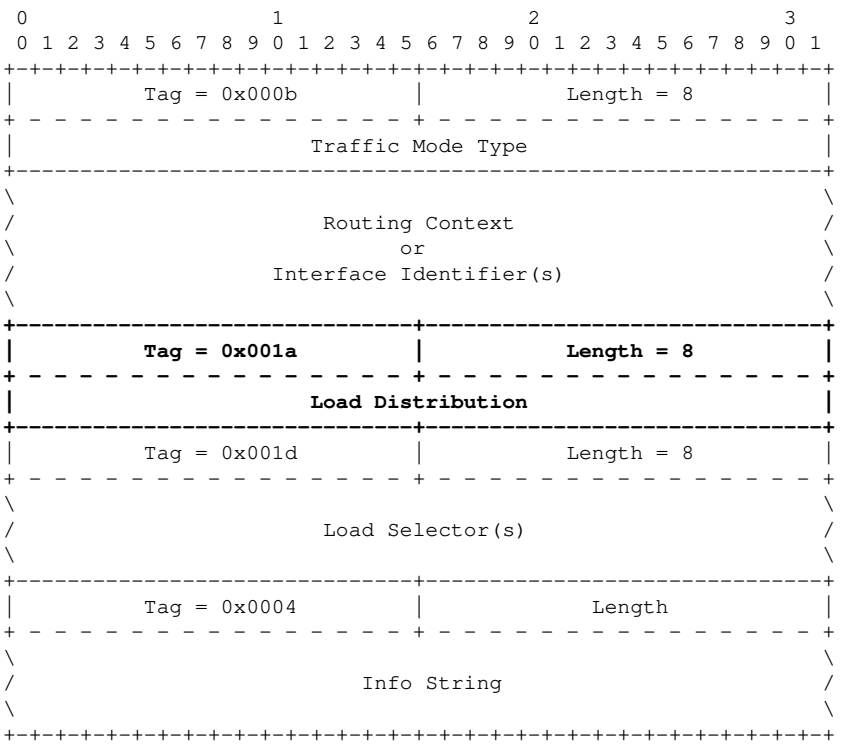
EDITOR'S NOTE:- The *Registration Status* value shown as `16` above will be assigned by IANA as a value of the UA-specific *Registration Status* parameter for each SIGTRAN UA and may change its value in further versions of this document.

3.2.3. ASP Active (ASPAC)

LOADGRP supplements the **ASPAC** message by permitting the following optional parameters to be included in the message:

Extension Parameters	
Load Distribution	Optional

The format of the resulting **ASPAC** message is as follows:



EDITOR'S NOTE:– The parameter tag values shown as **0x001a** above will be assigned by IANA within the common parameter range of the SIGTRAN UAs and may change its value in further versions of this document.

When an ASP wishes to activate only within a Load Group associated with an Application Server, it includes the *Load Selector* and *Load Distribution* parameters in the **ASPAC** message.

The *Load Distribution* parameter indicates the traffic distribution method to be used within the Load Group as identified by the Load Selector.

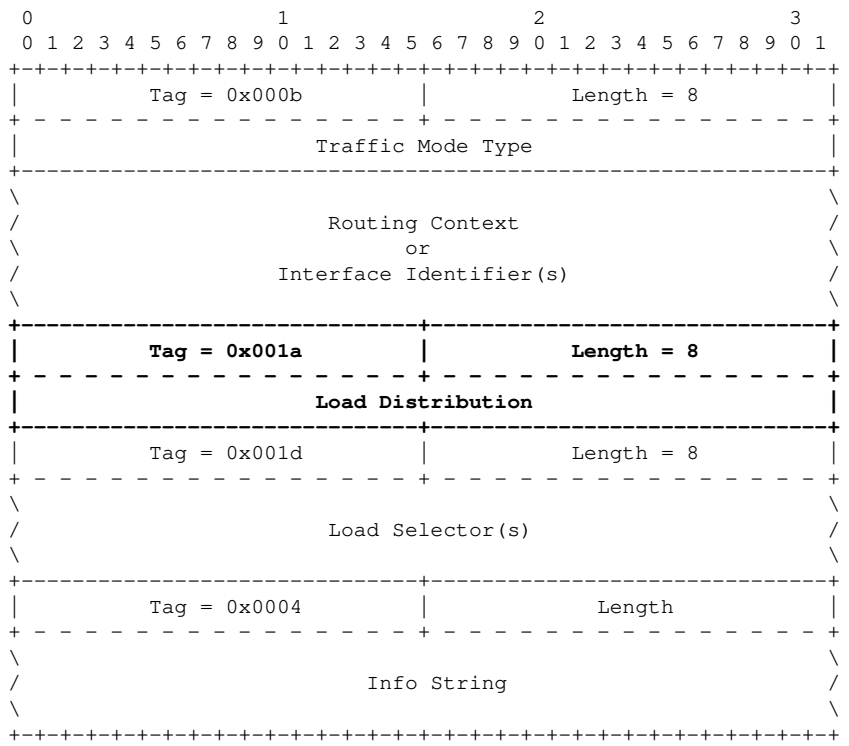
No other changes to the **ASPAC** message format are provided by *LOADGRP*[1].

3.2.4. ASP Active Ack (ASPAC ACK)

LOADGRP supplements the **ASPAC ACK** message by permitting the following optional parameters to be included in the message:

Extension Parameters	
Load Distribution	Optional

The format of the resulting **ASPAC ACK** message is as follows:



EDITOR'S NOTE:– The parameter tag values shown as **0x001a** above will be assigned by IANA within the common parameter range of the SIGTRAN UAs and may change its value in further versions of this document.

When an ASP has requested activation within a Load Group or the SG otherwise activates the ASP within a Load Group the SG responds with an **ASPAC ACK** message including the *Load Selector* that identifies the Load Group and optionally includes the *Load Distribution* for the Load Group in which the ASP has been activated. If the ASP included the *Load Distribution* parameter in the **ASPAC** message, the SG **MUST** include the *Load Distribution* parameter in the response **ASPAC ACK** message.

The *Load Distribution* parameter indicates the traffic distribution method to be used within the Load Group as identified by the Load Selector.

No other changes to the **ASPAC ACK** message format are provided by *LOADGRP*[1].

3.2.5. Error (ERR)

LOADGRP supplements the **Error (ERR)** message by adding the following values to the common mandatory *Error Code* parameter in the **ERR** message:

28 Unsupported Load Distribution

EDITOR'S NOTE:- The *Error Code* value shown throughout this document as 28 will be assigned by IANA as a value of the common *Error Code* parameter for SIGTRAN UAs and may change its value in further versions of this document.

This new error code is interpreted as follows:

The *"Invalid/Unsupported Load Distribution"* error would be sent by an SGP if an ASP sends an ASP Active with an invalid or unsupported Load Distribution. An example would be a case in which the SGP did not support override within a Load Group.

No other changes to the **ERR** message or *Error Code* parameter values are provided by *LOADGRP*. See Section 4 for extension procedures associated with the **ERR** message.

4. Procedures

LOADGRP provides for an additional level of control over the traffic distribution patterns within an Application Server. *LOADGRP* provides the *Load Distribution* parameter which may be optionally included in the **REG REQ**, **ASPAC** and **ASPAC ACK** messages. In addition, it supplements the ASP State Maintenance and ASP Traffic Maintenance procedures.

4.1. ASP State Maintenance

In addition to the SGP maintaining the state of each remote ASP in each Application Server that the ASP is configured to receive traffic, under *LOADGRP*, the SGP **MAY** also maintain the state of each remote ASP in each Load Group within an Application server that the ASP is configured to receive traffic. The Load Group state is maintained separate from the ASP and AS states.

4.1.1. Load Group State

The state of the Load Group is maintained in the Signalling User Adaptation Layer on the SGPs. The state of the Load Group changes due ASP state transitions. The possible states of a Load Group are:

LG-DOWN: The Load Group is unavailable. This state implies that all related ASPs are in the **ASP-DOWN** state for this Load Group. Initially, the Load Group will be in this state.

LG-INACTIVE: The Load Group is available but no application traffic is active (i.e, one or more related ASPs are in the **ASP-INACTIVE** state, but none are in the **ASP-ACTIVE** state within the Load Group).

LG-ACTIVE: The Load Group is available and application traffic is active. This state implies that at least one ASP is in the **ASP-ACTIVE** state within the Load Group.

4.1.2. Application Server State

Where ASPs are configured for operation within *Load Groups* under *LOADGRP*, the Application Server state is interpreted as provided in the Signalling User Adaptation Layer specifications [M2UA...TUA].

4.2. ASP Traffic Maintenance

4.2.1. ASP Up Procedures

LOADGRP extends the ASP Up procedures to include the concept of the Load Group.

Whenever an SGP moves an ASP to the ASP-UP state, it also moves the ASP to the ASP-INACTIVE state in all Load Groups in all Application Servers for which the ASP is configured. This may have the effect of changing the Load Group state, requiring the generation of **NTFY** messages as described under ""Notify Procedures"" below.

4.2.2. ASP Down Procedures

LOADGRP extends the ASP Down procedures to include the concept of the Load Group.

Whenever an SGP moves an ASP to the ASP-DOWN state, it also moves the ASP to the ASP-DOWN state in all Load Groups in all Application Servers to which the ASP belongs. This may have the effect of changing the Load Group state, requiring the generation of **NTFY** messages as described under ""Notify Procedures"" below.

4.2.3. ASP Active Procedures

When activating, *LOADGRP* extends the UA activation procedures by permitting an optional *Load Distribution* parameter to be included in the ASP Active (**ASPAC**) and ASP Active Ack (**ASPAC ACK**) messages. The *Load Selector* parameter is used to indicate for which Load Group the concerned Application Server is becoming active [LOADSEL] and the *Load Distribution* parameter is used to indicate the traffic mode of the Load Group.

LOADGRP supplements the ASP Active Procedures as follows:

When an ASP sends an **ASPAC** message to activate the ASP within an Application Server, the ASP **MAY** choose to activate within a Load Group for the specified AS by including the *Load Selector* parameter in the **ASPAC** message.

When an SGP receives an **ASPAC** message with a *Load Selector* parameter in the message, or where the SGP is otherwise configured to activate the ASP in a configured Load Group, when the SGP moves the ASP to the ASP-ACTIVE state for the AS, it also moves the ASP to the ASP-ACTIVE state for the Load Group identified by the *Load Selector* parameter. In either case, if the activation is successful, the SGP includes the *Load Selector* parameter in the **ASPAC ACK** message.

If the *Load Selector* in the **ASPAC** message is invalid, the SGP responds with **ERR**("Invalid Load Selector"). If the *Load Selector* parameter is not present in the **ASPAC** message and the SGP is configured to require one, or the *Load Selector* parameter is not valid or not configured for the Application Server, the SGP responds with **ERR**("Missing Parameter"). If the *Load Selector* parameter contains an invalid or unsupported *Load Distribution* value, or the *Load Distribution* parameter is missing but the SGP cannot determine the distribution applicable to the Load Group without one, the SGP responds with **ERR**("Unsupported Load Distribution").

There are three modes of Application Server traffic handling in the SGP at the Application Server level and three modes of AS traffic handling at the Load Group level: Override, Load-share and Broadcast. When included, the *Traffic Mode Type* parameter in the **ASPAC** message indicates the traffic handling mode to be used in a particular Application Server between Load Groups. When included, the *Load Distribution* parameter indicates the traffic handling mode to be used between ASPs within the Load Group indicated by the *Load Selector* parameter.

In the case of an Override mode AS, reception of an ASP Active message at an SGP may move a Load Group to the LG-ACTIVE state. When an LG moves to the LG-ACTIVE state in an Override mode AS, this causes the (re)direction of all traffic for the AS to the active Load Group. Distribution of traffic within the Load Group is determine on the basis of the Load Distribution mode of the Load Group. Any previously active Load Group is now considered to be in state LG-INACTIVE and **SHOULD** no longer receive traffic from the SGP within the AS. The SGP then **MUST** send a Notify message **NTFY**("Alternate ASP Active") and include the *Load Selector* parameter in the Notify message indicating the Load Group that has become active.

In the case of a Load-share mode AS, the reception of an ASP Active message at an SGP that moves a Load Group to the LG-ACTIVE state causes the direction of specific traffic flows associated with the *Load Selector* to the Load Group. The assignment of traffic flows to *Load Selector* values is implementation dependent, but could be based on specific information in the protocol data message.

In the case of a Broadcast mode AS, reception of an ASP Active message at an SGP that results in moving a Load Group to the LG-ACTIVE state within the AS causes the direction of traffic to the newly active Load Group in addition to all the other LGs that are currently active for the AS. The algorithm at the SGP for broadcasting traffic within an AS to all the active ASPs is a simple broadcast algorithm, where every message is sent to each of the active Load Groups.

4.2.4. ASP Inactive Procedures

When deactivating, *LOADGRP* extends the UA activation procedures by permitting an optional *Load Selector* parameter to be included in the ASP Inactive (**ASPIA**) and ASP Inactive Ack (**ASPIA ACK**) message. The *Load Selector* parameter is used to indicate for which Load Group the concerned Application Server is becoming inactive.

LOADGRP supplements the ASP Active procedures of the UAs as follows:

When an ASP wishes to withdraw from a specific Load Group within an Application Server, the ASP sends an ASP Inactive message to the SGP with a *Load Selector* parameter included in the message. In the case where the ASP does not include the *Load Selector* parameter in the ASP Inactive message, the SGP must know via configuration data which Load Groups the ASP is a member. Upon receiving an ASP Inactive message with included or implied *Load Selector*, the SGP moves the ASP to the ASP-INACTIVE state in each of the Load Groups indicated.

4.2.5. Notify Procedures

When the SGP or IPSP notifies its UA peer with a **NTFY** messages which concerns an ASP, it **MUST** include the Load Group (if available) along with the *ASP Identifier* in the message. The **NTFY** messages to which this applies are:

NTFY("AS-ACTIVE"), *NTFY*("AS-PENDING"), *NTFY*("AS-INACTIVE") –

When the SGP or IPSP notifies the active and inactive ASPs in an AS that it has detected the transition of the Application Server to the AS-ACTIVE, AS-PENDING or AS-INACTIVE state, or that it has detected the transition of a Load Group to the LG-ACTIVE state for the AS, it *MUST* include the *Load Selector*, if available, indicating the Load Group which has changed state along with the *ASP Identifier*, if any, in the **NTFY** message.

When the *Routing Context* (*Interface Identifier*) associated with the Application Server cannot be implied at the ASP from static configuration data, the *Routing Context* (*Interface Identifier*) **MUST** also be placed in the **NTFY** message.

NTFY("ASP Failure") –

When the SGP or IPSP notifies the active and inactive ASPs in an AS that it has detected the failure of an ASP or the failure of an association to an ASP (i.e. SCTP Communication Lost Indication), it **MUST** include the *Load Selector* (if available) with the *ASP Identifier* in the message.

When the *Routing Context* (*Interface Identifier*) associated with the Application Servers cannot be implied at the ASP from static configuration data, the *Routing Context* (*Interface Identifier*) **MUST** also be placed in the **NTFY**("ASP Failure") message.

The notifying SGP or IPSP **MAY** also place the *Load Selector* parameter into the **NTFY**("ASP Failure") message to indicate the traffic which was applicable to the load selection at the time of the failure.

The purpose of this procedure is to inform the active and inactive ASPs, not only of the ASP failure, but of the identity of the ASP and the load selection for which the failed ASP was responsible.

NTFY("Alternate ASP Active in AS") –

When the SGP or IPSP notifies the previously active ASP in a override AS that an alternate ASP has become active, it **MUST** include the *Load Selector* indicating the Load Group, if any, with the *ASP Identifier*, if available, in the **NTFY** message.

The purpose of this procedure is to inform the previously active ASP, not only of the that another ASP has taken over the traffic for which the notified ASP was previously responsible, but to also indicate the identify of the overriding ASP and the load selection that was overridden.

When an ASP becomes ASP-INACTIVE for a Load Group or Application Server for the first time, the SGP **MAY** send **NTFY** messages indicating the state of the Application Server and the Load Groups in the application server to the newly inactive ASP. Application Server state is notified by sending the appropriate **NTFY** message without a *Load Selector* parameter present in the message. Load Group state is notified by sending the appropriate **NTFY** message with a *Load Selector* parameter indicating the Load Group in the message.

4.3. Registration

UAs permit Application Server Processes to register for the *Routing Context* (*Interface Identifier*) associated with a *Routing Key* (*Link Key*). *LOADGRP* extends the registration procedure to also permit the Application Server Process to register a *)Load Distribution* against a *Load Selector* identifying a Load Group in addition to the *Routing Context* (*Interface Id*).

This is performed by extending the registration procedure of *Load Selection* [LOADSEL] by adding the *Load Distribution* parameter to the *REG REQ* message during regsitration. The *Load Distribution* parameter is analogous to the *Traffic Mode Type* parameter.

In addition to the normal registration procedures of the UAs, the following additional error conditions can occur:

"Error – Unsupported/Invalid Load Distribution"

This error **MUST** be sent in an **Registration Response (REG RSP)** message whenever the SG determines that the *Load Distribution* associated with a Load Group is invalid, is not supported by the SG, or is required to determine the Load Distribution algorithm of the Load Group but is missing from the *Load Selection* in the **REG REQ** message.

4.4. Interworking Procedures

5. Examples

Figure 3 illustrates the example configuration that is used for all the examples in this section. The example configuration consist of:

- Two SGs (SG1 and SG2) acting as STPs in the SS7 network and consisting (for example) of a single SGP. Each SG is connected to each of the ASPs in the example configuration.

- Four ASPs (ASP1, ASP2, ASP3 and ASP4). Each ASP is connected to both of the SGs in the example configuration.
- Two Load Selectors (LS1 and LS2) are associated with the Application Server. The traffic that corresponds to each Load Selectors and the Load Distribution within each Load Selector is different in each example.

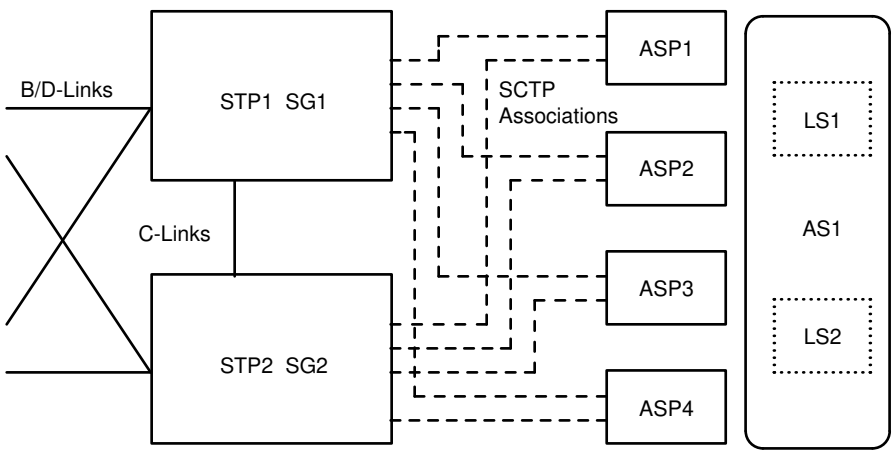


Figure 3. Example Configuration

5.1.1. Initialization

Figure 4 illustrates the common initialization procedure use for all of the examples. Each ASP establishes an SCTP Association with SG1 and sends and ASP Up message to which it receives and ASP Up response. The ASPs are not statically configured to serve specific AS or LS within the AS, so no Notify messages are received. The same sequence of messages are also exchanged with SG2.

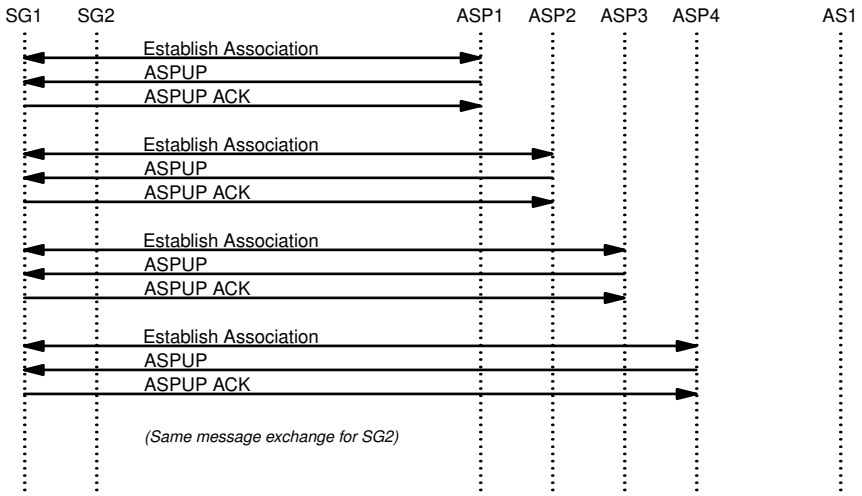


Figure 4. Example – Initialization

5.2. M3UA with Override LG, Load-share AS, based on CIC

This example is for an M3UA [M3UA] configuration with the Application Server (AS1) configured with a Traffic Mode Type of Load-share. The Application Server (AS1) has associated with it a Routing Key (RK1) that consists of a Destination Point Code that corresponds to the AS1 (MGC1) point code (PC1), an Originating Point Code that corresponds to a remote MGC2 point code (PC2), and the SI value for ISUP (SI=5). The Load Selectors (LS1 and LS2) correspond to two sets of CIC values which correspond to two different trunk groups between MGC1 and MGC2 (TG1 and TG2).

5.2.1. Activation

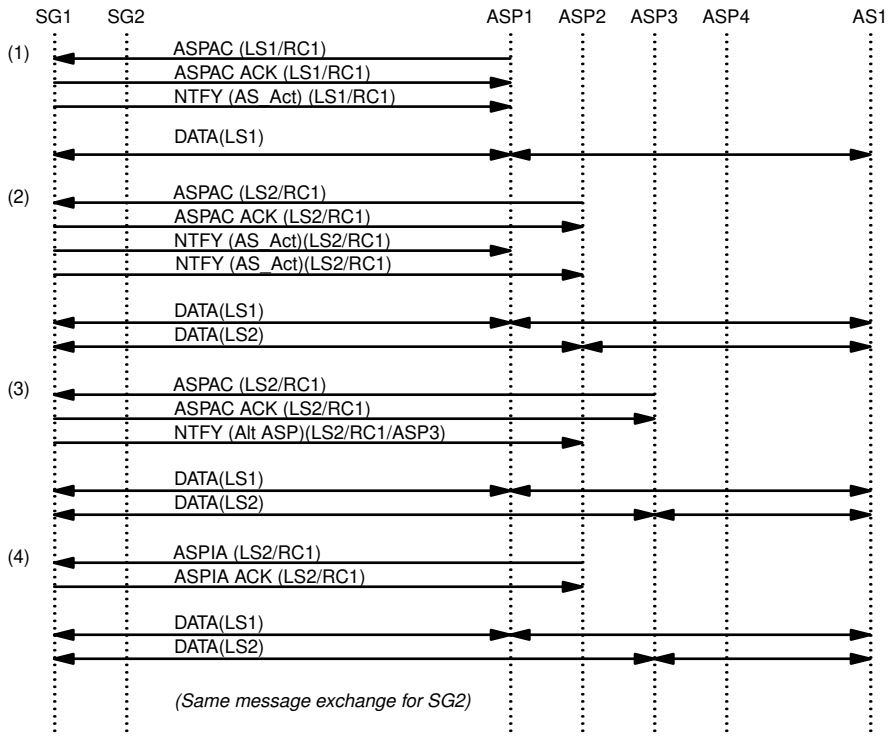


Figure 5. M3UA Example – Activation

Figure 5 illustrates activation of the ASPs for Load-share Load Distribution within the Override Application Server. The sequence of events is as follows:

(1) ASP1 sends an ASP Active message to SG1 identifying *Load Selector* LS1 within Application Server AS1/RC1, optionally including a *Load Distribution* indicating ""Load-share,"" and receives an acknowledgment and a notification. Data is transferred between the SG and ASP1 for Load Group LS1 within AS1.

(2) ASP2 sends an ASP Active message to SG1 identifying *Load Selector* LS2 within Application Server AS1/RC1 and receives an acknowledgment and notification. ASP1 also receives notification that LS2 is active for AS1/RC1. Data is transferred between the SG and ASP2 for AS1/RC1 load-shared on CIC between LS1/ASP1 and LS2/ASP2.

(3) ASP3 sends an ASP Active message to SG1 identifying *Load Selector* LS2 within Application Server AS1/RC1 and receives an acknowledgment. Because LS2 is an Override Load Group, ASP2 gets a notification that ASP3 is now the active ASP for LS2/AS1/RC1. Data is transfer to ASP3 for LS2 and remains transferred to ASP1 for LS1.

(4) ASP2 sends an ASP Inactive message to SG1 identifying *Load Selector* LS1 within Application Server AS1/RC1 and receives an acknowledgment. Because ASP2 is not active in LS2, and LS2 is an Override Load Group, SG1 continues to load-share between LS1 to ASP1 and LS2 to ASP3.

5.2.2. Failure of ASP1

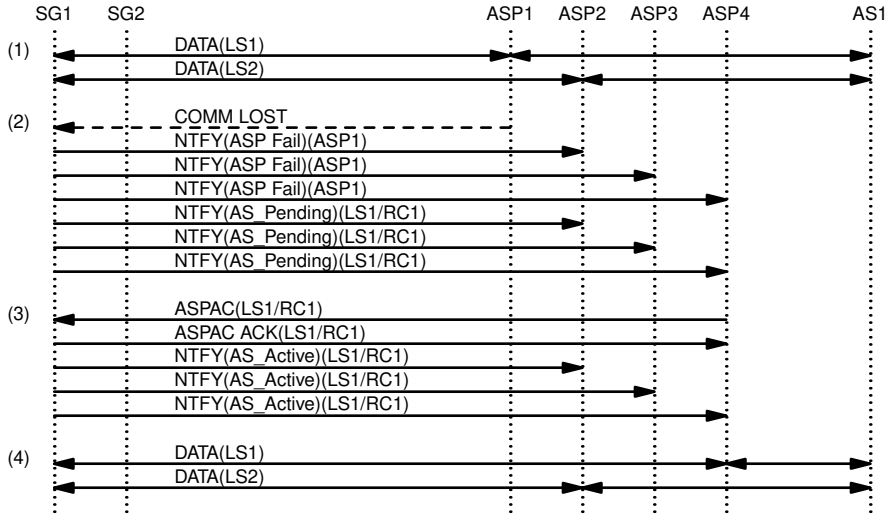


Figure 6. M3UA Example – Failure of ASP1

Figure 6 illustrates the failure of ASP1. The sequence of events is as follows:

(1) Data for LS1 within AS1 is exchanged between SG1 and ASP1. Data for LS2 within AS1 is exchanged between SG1 and ASP2.

(2) Communication is lost between SG1 and ASP1. ASP2, ASP3, and ASP4 are notified of the failure of ASP1 and the change of state of AS1 to **AS-PENDING** for LS1. Data for LS2 in AS1 is unaffected.

(3) ASP4 (spare) responds to the **AS-PENDING** notification and activates for LS1 in AS1/RC1. ASP2, ASP3 and ASP4 receive an **AS-ACTIVE** notification. Data for LS1 in AS1 is now exchanged with ASP4. Data for LS2 in AS1 continues to be exchanged with ASP2.

5.2.3. Sparing

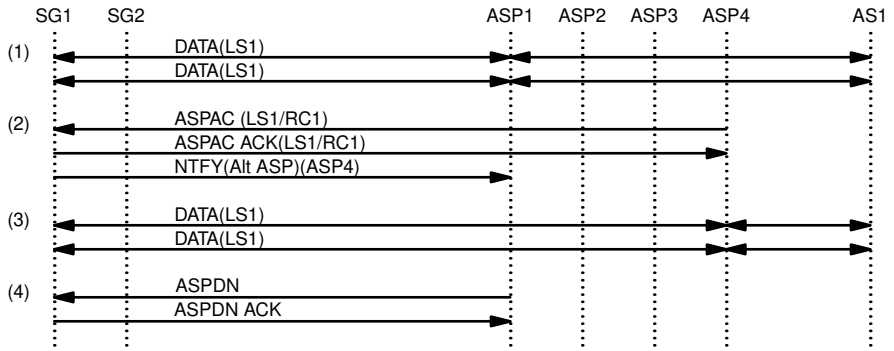


Figure 7. M3UA Example – Sparing

Figure 7 illustrates a sparing situation where one ASP takes over traffic from another so that the original ASP can be taken out of service. The sequence of events is as follows:

(1) Data for LS1 in AS1 is exchange between SG1 and ASP1.

(2) ASP4 (spare) activates for LS1 in AS1 and receives an acknowledgment. ASP4 has overridden ASP1 and a notification is sent to ASP1 that indicates that ASP4 in now the "Alternate ASP Active for AS".

(3) Data for LS1 in AS1 is now being exchanged between SG1 and ASP4.

(4) ASP1 can now be taken down and out of service.

5.3. SUA with Load-share LG, Load-share AS based on GT

This example is for an SUA [SUA] configuration with the Application Server (AS1) configured with a Traffic Mode Type of Load-share. The Application Server (AS1) has associated with it (RK1) that consists of Destination Point Code and Subsystem Number that corresponds to the AS1 (HLR1) point code (PC1). The Load Selectors (LS1 and LS2) correspond to two sets of Global Titles which correspond to Mobile and GSTN numbering.

5.3.1. Activation

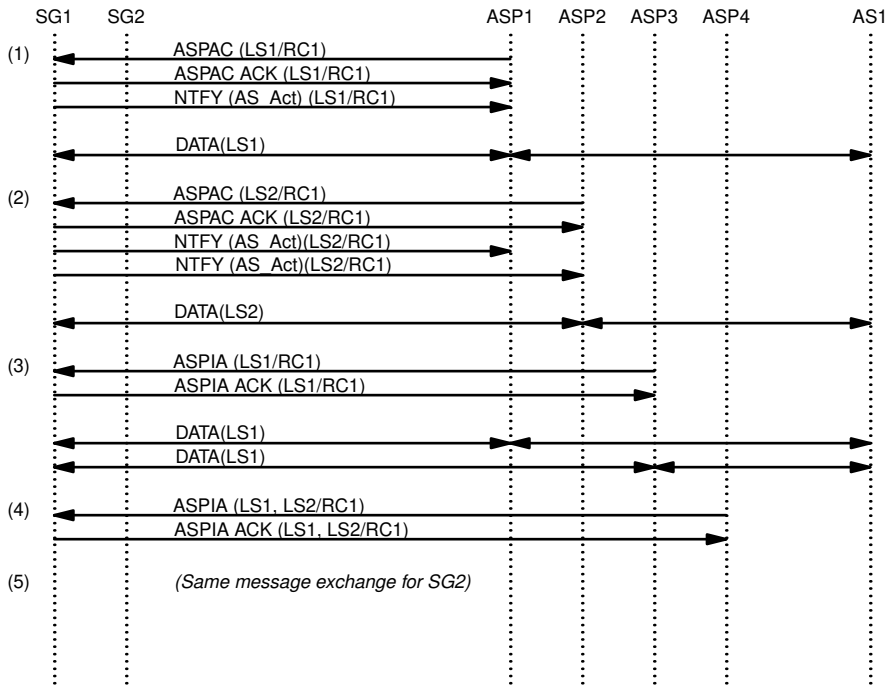


Figure 8. SUA Example – Activation

Figure 8 illustrates activation of the ASPs for Load Selectors within the Load-share Application Server. The sequence of events is as follows:

(1) ASP1 sends an ASP Active message to SG1 identifying Load Selector LS1 within Application Server AS1/RC1 and receives an acknowledgment and a notification. Data is transferred between the SG and ASP1 for Load Selector LS1 within AS1.

(2) ASP2 sends an ASP Active message to SG1 identifying Load Selector LS2 within Application Server AS1/RC1 and receives an acknowledgment and a notification. ASP1 also receives notification that AS1/RC1 is active for LS2. Data is transferred between the SG and ASP2 for Load Selector LS2 within AS1.

(3) ASP3 sends an ASP Active message to SG1 identifying Load Selector LS1 within Application Server AS1/RC1 and receives an acknowledgment. Data is load-shared between the SG and ASP1 and ASP3 for Load Selector LS1 within AS1.

(4) ASP4 sends an ASP Inactive message to SG1 identifying Load Selector LS1 and LS2 within Application Server AS1/RC1 and receives an acknowledgment.

(5) The same exchange is repeated for SG2.

5.3.2. Failure of ASP1 and ASP2

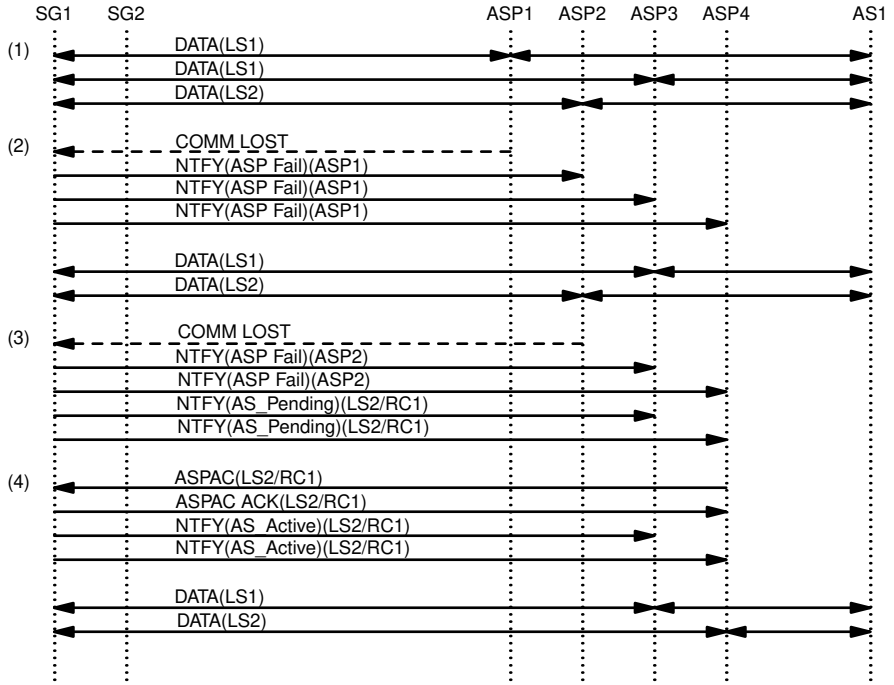


Figure 9. SUA Example – Failure of ASP1 and ASP2

Figure 9 illustrates the failure of ASP1 followed by the failure of ASP2. The sequence of events is as follows:

(1) Data for LS1 within AS1 is load-shared between ASP1 and ASP3. Data for LS2 within AS1 is transferred to ASP2.

(2) Communication is lost between SG1 and ASP1. ASP2, ASP3, and ASP4 are notified of the failure of ASP1. Data for LS1 in AS1 is now transferred to ASP3 only. Data for LS2 in AS1 is unaffected and is still transferred to ASP2.

(3) Communication is lost between SG1 and ASP2. ASP3 and ASP4 are notified of the failure of ASP2 as well of the AS1 state change to **AS-PENDING** for LS2. Data for LS2 is queued at the SG.

(4) ASP4 (spare) responds to the **AS-PENDING** notification and activates for LS2 in AS1/RC1. ASP3 and ASP4 receive an **AS-ACTIVE** notification for LS2. Data for LS2 in AS1 is now transferred to ASP4.

5.3.3. Sparing

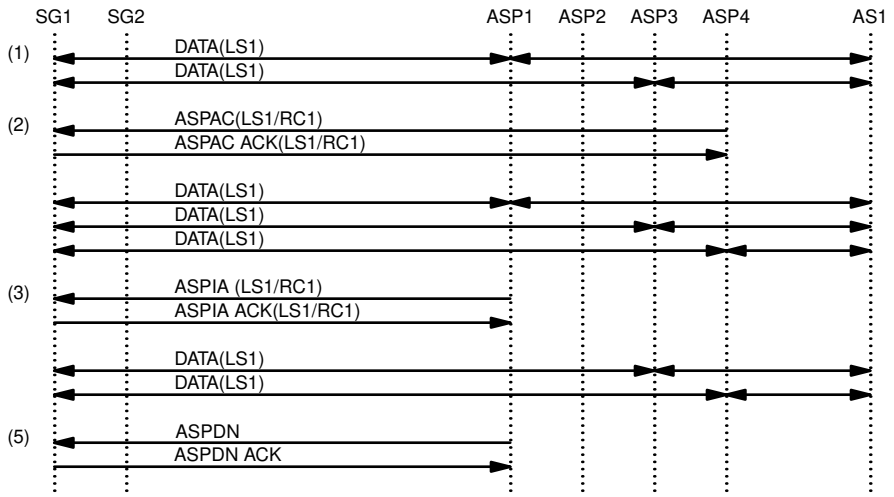


Figure 10. SUA Example – Sparing

Figure 10 illustrates a sparing situation where one ASP takes over traffic from another so that the original ASP can be taken out of service. The sequence of events is as follows:

(1) Data for LS1 in AS1 is load-shared by SG1 between ASP1 and ASP3.

(2) ASP4 (spare) activates for LS1 in AS1 and receives and acknowledgment. Data for LS1 in AS1 is now load-shared by SG1 between ASP1, ASP3 and ASP4.

(3) ASP1 deactivates for LS1 in AS1 and receives and acknowledgment.

(4) Data for LS1 in AS1 is now load-shared by SG1 between ASP3 and ASP4.

(5) ASP1 can now be taken down and out of service.

5.4. TUA with Broadcast LG, Load-share AS based on DID

This example is for an TUA [TUA] configuration with the Application Server (AS1) configured with a Traffic Mode Type of Broadcast. The Application Server (AS1) has associated with it (RK1) that consists of Destination Point Code and Subsystem Number that corresponds to the AS1 (HLR1) point code (PC1). The Load Selectors (LS1 and LS2) correspond to two sets of Dialog Ids which correspond to even and odd Dialog Ids.

5.4.1. Activation

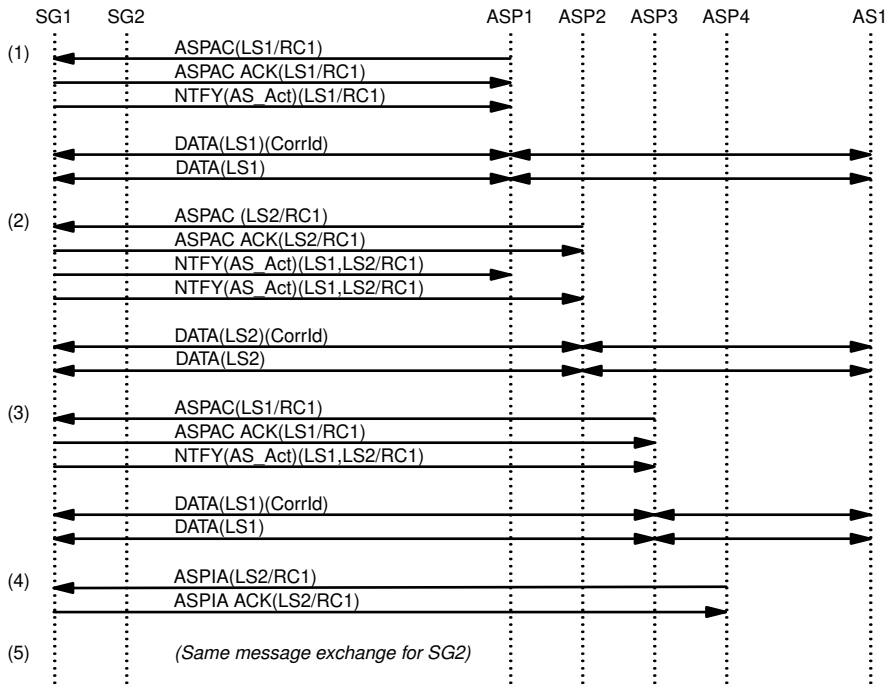


Figure 11. TUA Example – Activation

Figure 11 illustrates activation of the ASPs for Load Selectors within the Broadcast Application Server. The sequence of events is as follows:

(1) ASP1 sends an ASP Active message to SG1 identifying Load Selector LS1 within Application Server AS1/RC1 and receives an acknowledgment and a notification. Data is transferred between the SG and ASP1 for Load Selector LS1 within AS1. The initial Data Messages for LS1 within AS1 are tagged with a *Correlation Id*.

(2) ASP2 sends an ASP Active message to SG1 identifying Load Selector LS2 within Application Server AS1/RC1 and receives an acknowledgment and a notification. ASP1 also receives notification that AS1/RC1 is active for LS2. Data is transferred between the SG and ASP2 for Load Selector LS2 within AS1. The initial Data Messages for LS2 within AS1 are tagged with a *Correlation Id*.

(3) ASP3 sends an ASP Active message to SG1 identifying Load Selector LS1 within Application Server AS1/RC1 and receives an acknowledgment. Data is broadcast the SG and ASP1 and ASP3 for Load Selector LS1 within AS1. The initial Data Messages for LS2 within AS1 are tagged with a *Correlation Id*.

(4) ASP4 sends an ASP Inactive message to SG1 identifying Load Selector LS1 and LS2 within Application Server AS1/RC1 and receives an acknowledgment.

(5) The same exchange is repeated for SG2.

5.4.2. Failure of ASP1 and ASP2

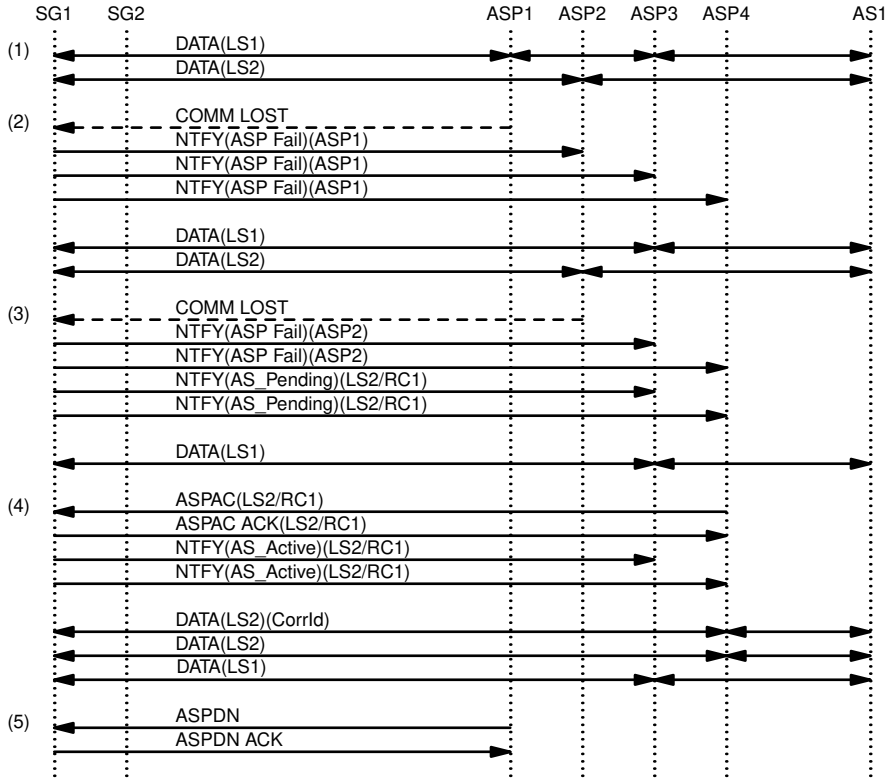


Figure 12. TUA Example – Failure of ASP1

Figure 12 illustrates the failure of ASP1 followed by the failure of ASP2. The sequence of events is as follows:

(1) Data for LS1 within AS1 is broadcast to ASP1 and ASP3. Data for LS2 within AS1 is sent to ASP2.

(2) Communication is lost between SG1 and ASP1. ASP2, ASP3, and ASP4 are notified of the failure of ASP1. Data for LS1 in AS1 is directed toward ASP3 only. Data for LS2 in AS1 is unaffected.

(3) Communication is lost between SG1 and ASP2. ASP3 and ASP4 are notified of the failure of ASP1 as well of the AS1 state change to **AS-PENDING** for LS2. Data for LS2 is queued at the SG.

(4) ASP4 (spare) responds to the **AS-PENDING** notification and activates for LS2 in AS1/RC1. ASP3 and ASP4 receive an **AS-ACTIVE** notification. Data for LS2 in AS1 is now exchanged with ASP4. Initial DATA messages for LS2 in AS1 are tagged with a *Correlation Id*.

5.4.3. Sparing

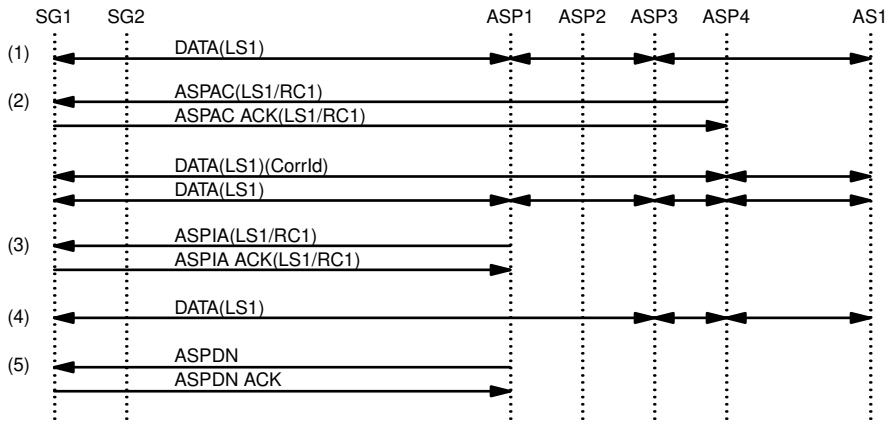


Figure 13. TUA Example – Sparing

Figure 13 illustrates a sparing situation where one ASP takes over traffic from another so that the original ASP can be taken out of service. The sequence of events is as follows:

(1) Data for LS1 in AS1 is broadcast from SG1 to ASP1 and ASP3.

(2) ASP4 (spare) activates for LS1 in AS1 and receives and acknowledgment. Data for LS1 in AS1 is now being broadcast from SG1 to ASP1, ASP3 and ASP4. Initial data for LS1 in AS1 is tagged with a *Correlation Id*.

(3) ASP1 deactivates for LS1 in AS1 and receives and acknowledgment.

(4) Data for LS1 in AS1 is now broadcast from SG1 to ASP3 and ASP4.

(5) ASP1 can now be taken down and out of service.

6. Security

LOADGRP does not introduce any new security risks or considerations that are not already inherent in the UA [IUA...TUA] Please see the "Security" sections of the UA [IUA...TUA] for security considerations and recommendations that are applicable to each UA.

7. IANA Considerations

LOADGRP adds the following parameter tag value (described in Section 3) to the Common Parameter numbering space for the UAs [IUA...TUA].

Tag Value	Parameter Name
0x001a	Load Distribution

EDITOR'S NOTE:– The *Load Distribution* parameter tag value shown throughout this document as **`0x001a`** will be assigned by IANA within the common parameter range of the SIGTRAN UAs and may change its value in further versions of this document.

LOADGRP adds the following value to the *Error Code* parameter of the UAs.

28 Unsupported Load Distribution

EDITOR'S NOTE:- The *Error Code* value shown as `28` above will be assigned by IANA as a value of the common *Error Code* parameter for SIGTRAN UAs and may change its value in further versions of this document.

LOADGRP adds the following value to the *Registration Status* field of the *Registration Result* parameter for the UAs [M2UA...TUA].

16 Error – Invalid/Unsupported Load Distribution

EDITOR'S NOTE:- The *Registration Status* value shown as `16` above will be assigned by IANA as a value of each UA-specific *Registration Status* parameter for each SIGTRAN UA and may change its value in further versions of this document.

Acknowledgments

The authors would like to thank Tolga Asveren, Ken Morneault, Barry Nagelberg, Benjamin J. Wilson, Jacques Rajchgod, Greg Sidebottom and Gery Verwimp for their valuable comments and suggestions.

Notes

[1] For a detailed description of these messages, see Section 3 of the specific UA document [M2UA...TUA].

References

IUA. K. Morneault, S. Rengasami, M. Kalla and G. Sidebottom, "ISDN Q.921-User Adaptation Layer," RFC 3057, The Internet Society (November, 2000). [Normative]

DUA. R. Mukundan, N. Mangalpally, K. Morneault and A. Vydyam, "DPNSS/DASS 2 Extensions to the IUA Protocol," <draft-ietf-sigtran-dua-04.txt>, Internet Engineering Task Force - Signalling Transport Working Group (October 2002). Work In Progress. [Informative]

V5UA. E. Weilandt, N. Khanchandani and S. Rao, "V5.2-User Adaption Layer (V5UA)," <draft-ietf-sigtran-v5ua-03.txt>, Internet Engineering Task Force - Signalling Transport Working Group (June 2002). Work In Progress. [Informative]

GR303UA.
R. Mukundan, K. Morneault, "GR-303 extensions to the IUA Protocol," <draft-ranjith-sigtran-gr303ua-00.txt>, Internet Engineering Task Force - Signalling Transport Working Group (October 2002). Work In Progress. [Informative]

M2UA. K. Morneault, R. Dantu, G. Sidebottom, B. Bidulock and J. Heitz, "Signaling System 7 (SS7) Message Transfer Part 2 (MTP2) - User Adaptation Layer," RFC 3331, Internet Engineering Task Force - Signalling Transport Working Group (September, 2002). [Normative]

M3UA. G. Sidebottom, K. Morneault and J. Pastor-Balbas, (eds), "Signaling System 7 (SS7) Message Transfer Part 3 (MTP3) - User Adaptation Layer (M3UA)," RFC 3332, Internet Engineering Task Force - Signalling Transport Working Group (September, 2002). [Normative]

SUA. J. Loughney, G. Sidebottom, L. Coene, G. Verwimp, J. Keller and B. Bidulock, "SS7 SCCP-User Adaptation Layer (SUA)," <draft-ietf-sigtran-sua-14.txt>, Internet Engineering Task Force - Signalling Transport Working Group (June 30, 2002). Work In Progress. [Normative]

ISUA. B. Bidulock, "SS7 ISUP-User Adaptation Layer (ISUA)," <draft-bidulock-sigtran-isua-00.txt>, Internet Engineering Task Force - Signalling Transport Working Group (January 5, 2003). Work In Progress. [Normative]

TUA. B. Bidulock, "SS7 TCAP-User Adaptation Layer (TUA)," <draft-bidulock-sigtran-tua-01.txt>, Internet Engineering Task Force - Signalling Transport Working Group (January 2, 2003). Work In Progress. [Normative]

RFC 2960.
R. Stewart, Q. Xie, K. Morneault, C. Sharp, H. J. Schwarzbauer, T. Taylor, I. Rytina, H. Kalla, L. Zhang and V. Paxson, "Stream Control Transmission Protocol (SCTP)," RFC 2960, The Internet Society (February 2000). [Normative]

LOADSEL.
B. Bidulock, "Load Selection Extension for Signalling User Adaptation Layers (LOADSEL)," <draft-bidulock-sigtran-loadsel-01.txt>, Internet Engineering Task Force - Signalling Transport Working Group (January 2, 2003). Work In Progress. [Informative]

RFC 2119.
S. Bradner, "Key words for use in RFCs to Indicate Requirement Levels," RFC 2119 - BCP 14, Internet Engineering Task Force (March 1997). [Normative]

Author's Addresses

Brian Bidulock
OpenSS7 Corporation
1469 Jeffreys Crescent
Edmonton, AB T6L 6T1
Canada

Phone: +1-780-490-1141
Email: bidulock@openss7.org
URL: http://www.openss7.org/

This draft expires July, 2003.

List of Tables

Table 1 Sample Configurations 5

List of Illustrations

Figure 1 Existing Traffic Distribution 3
Figure 2 Load Group Traffic Distribution 3
Figure 3 Example Configuration 11
Figure 4 Example – Initialization 12
Figure 5 M3UA Example – Activation 13
Figure 6 M3UA Example – Failure of ASP1 14
Figure 7 M3UA Example – Sparing 14
Figure 8 SUA Example – Activation 15
Figure 9 SUA Example – Failure of ASP1 and ASP2 16
Figure 10 SUA Example – Sparing 17
Figure 11 TUA Example – Activation 18
Figure 12 TUA Example – Failure of ASP1 19
Figure 13 TUA Example – Sparing 20

Table of Contents

Status of this Memo 1
Abstract 1
1 Introduction 1
1.1 Scope 1
1.2 Change History 1
1.3 Terminology 2
1.4 Overview 2
1.4.1 Existing Traffic Distribution 2
1.4.2 Extended Traffic Distribution 3
1.5 Sample Configurations 4
2 Conventions 5
3 Protocol Elements 5
3.1 Parameters 5
3.1.1 Load Distribution 5
3.2 Messages 6
3.2.1 Registration Request (REG REQ) 6
3.2.2 Registration Response (REG RSP) 6
3.2.3 ASP Active (ASPAC) 6
3.2.4 ASP Active Ack (ASPAC ACK) 7
3.2.5 Error (ERR) 8
4 Procedures 8
4.1 ASP State Maintenance 9
4.1.1 Load Group State 9
4.1.2 Application Server State 9
4.2 ASP Traffic Maintenance 9
4.2.1 ASP Up Procedures 9
4.2.2 ASP Down Procedures 9
4.2.3 ASP Active Procedures 9

4.2.4 ASP Inactive Procedures 10
4.2.5 Notify Procedures 10
4.3 Registration 11
4.4 Interworking Procedures 11
5 Examples 11
5.1.1 Initialization 12
5.2 M3UA with Override LG, Load-share AS, based on CIC 12
5.2.1 Activation 13
5.2.2 Failure of ASP1 14
5.2.3 Sparing 14
5.3 SUA with Load-share LG, Load-share AS based on GT 15
5.3.1 Activation 15
5.3.2 Failure of ASP1 and ASP2 16
5.3.3 Sparing 17
5.4 TUA with Broadcast LG, Load-share AS based on DID 17
5.4.1 Activation 18
5.4.2 Failure of ASP1 and ASP2 19
5.4.3 Sparing 20
6 Security 20
7 IANA Considerations 20
Acknowledgments 21
Notes 21
References 21
Author's Addresses 22
List of Tables 23
List of Illustrations 23
Table of Contents 23

Copyright Statement

Copyright © The Internet Society (2003). All Rights Reserved.

This document and translations of it may be copied and furnished to others, and derivative works that comment on or otherwise explain it or assist in its implementation may be prepared, copied, published and distributed, in whole or in part, without restriction of any kind, provided that the above copyright notice and this paragraph are included on all such copies and derivative works. However, this document itself may not be modified in any way, such as by removing the copyright notice or references to the Internet Society or other Internet organizations, except as needed for the purpose of developing Internet standards in which case the procedure for copyrights defined in the Internet Standards process must be followed, or as required to translate into languages other than English.

The limited permission granted above are perpetual and will not be revoked by the Internet Society or its successors or assigns.

This document and the information contained herein is provided on an "AS IS" basis and **THE INTERNET SOCIETY AND THE INTERNET ENGINEERING TASK FORCE DISCLAIMS ALL WARRANTIES, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO ANY WARRANTY THAT THE USE OF THE INFORMATION HEREIN WILL NOT INFRINGE ANY RIGHTS OR ANY IMPLIED WARRANTIES OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE.**