Oracle Communications Diameter Signaling Router

Release 7.0.1 Planning Guide

ORACLE WHITE PAPER | NOVEMBER 2015

Table of Contents

List of Figures

List of Tables

.

Acronyms

Acronym	Description
cSBR	Charging SBR
DA	Diameter Agent
DIH	Diameter Intelligence Hub
DP	Database Processor
DSR	Diameter Signaling Routing
IDIH	Integrated Diameter Intelligence Hub
IMR	Ingress Message Rate
IPFE	IP Front End
MP	Message Processor
MPS	Messages Per Second
N-IDIH	Network Integrated Diameter Intelligence Hub
NOAM	Network Operations, Alarms, Measurements
NE	Network Element
NFV	Network Function Virtualization
PM&C	Platform, Management, and Control
SBR	Session Binding Repository
SBR(b)	SBR – subscriber binding database
SBR(s)	SBR – session database
RMS	Rack Mount Server
SBR	Session Binding Repository
SOAM	System (nodal) Operations, Alarms, Measurements
SS7	Signaling System 7
TOR	Top-Of-Rack (Switches)
VM	Virtual Machine

Terminology

Term	Description
Co-Located	Two or more products in the same cabinet.
Co-Mingled	Two or more products in the same enclosure.
Site	A specific geographic location where DSR equipment is installed.
Geo-Diverse	Refers to DSR equipment located at geographically separated sites

Geo-Redundant	A node at a geo-diverse location which can assume the processing load for another node(s)
1+1 Redundancy	For every 1, an additional 1 is needed to support redundant capacity. The specific redundancy scheme is not inferred (e.g. active-active, active-standby).
N+K Redundancy	For every N, an additional K is needed to support redundant capacity. The specific redundancy scheme is not inferred (e.g. active-active, active-standby).
Node	A DSR node is either a DSR signaling node, an NOAM node or an SDS node.
Ingress Message Rate	A measure of the total Diameter messages per second ingressing the DSR. For this measure, a message is defined as any Diameter message that DSR reads from a Diameter peer connection independent of how the message is processed by the DSR.
Messages Per Second	A measure of the DSR Diameter message processing volume in messages per second. For this measure, a message is defined as: 1. DSR processing of an ingress Diameter message and either transmitting a single outgoing Diameter message or discarding the ingress message. The outgoing message may be a variant of, or a response to, the ingress message. 2. DSR transmission of any Diameter message, as required by DSR configuration, that is associated with incremental actions/events associated with #1 above. For example, the re-routing of a Request upon connection failure or the copying of a Request. Messages excluded from this measure are: • Diameter peer-to-peer messages: CER/CEA, DWR/DWA, and DPR/DPA • Ingress Diameter messages discarded by the DSR due to Overload controls • Answers received in response to Message Copy

References

[1] DSR Alarms, KPIs, and Measurements – Available at Oracle.com on the Oracle Technology Network (OTN)

[2] Platform Feature Guide – Available upon request

[3] DSR Feature Guide – Available at Oracle.com on the Oracle Technology Network (OTN)

[4] DSR 7.0.1 Cloud Benchmarking Guide – Available on MyOracleSupport

Introduction

The Diameter Signaling Router (DSR) creates a Diameter signaling core that relieves LTE and IMS endpoints of routing, traffic management and load balancing tasks and provides a single interconnect point to other networks. The resulting architecture is below. Refer to the DSR Feature Guide for currently supported functionality.

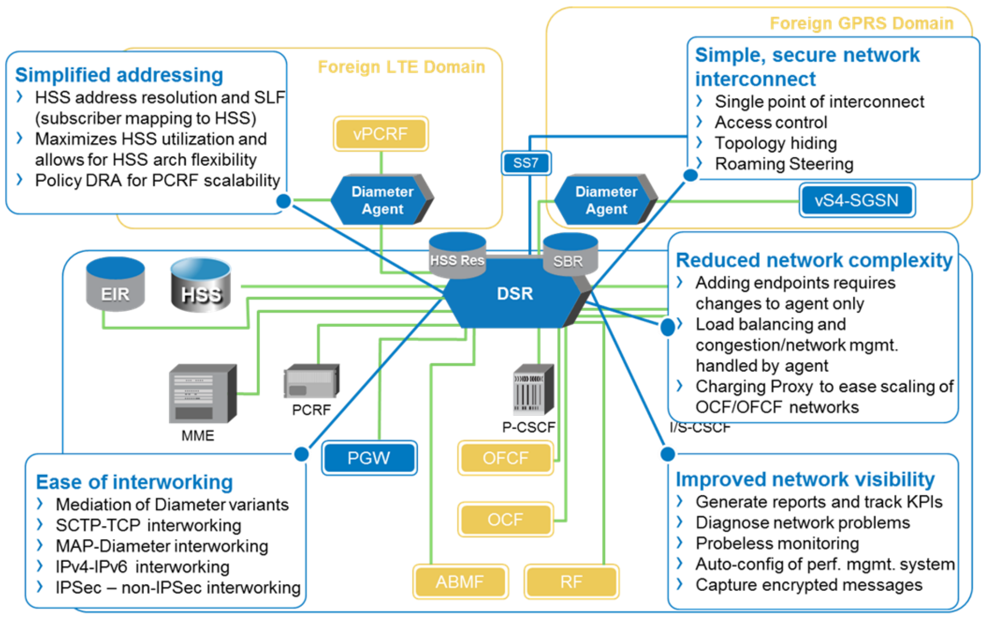

Figure 1. DSR in the Diameter Network

Cloud Deployable DSR

Historically, the DSR is sold and supported as an engineered system. That is, the software, exact hardware, precise configuration of both the hardware and software are specified and tuned for optimal performance, capacity and reliability. The telecommunications market is moving rapidly and many operators are deploying telecom applications on a common and shared infrastructure with many other telecom applications. This effort, along with the automation of application lifecycle is referred to as Network Functions Virtualization (NFV).

Oracle Communications Diameter Signaling Router (OCDSR or DSR) is deployed on a number of platforms. The DSR has a number of deployment scenarios:

» Bare-metal and hybrid (mixture of bare metal and virtual machines) - is the original deployment configuration of the DSR. It scales to very high performance and is widely deployed.

» Fully virtualized - was introduced shortly after bare-metal. It provides virtualization of the DSR, but does not use a cloud manager, and does not co-reside with other applications. Provides compact, cost-effective footprint and is widely deployed.

» Cloud deployable – is recently introduced. It provides full virtualization, assumes the DSR resources are managed by a COTS infrastructure manager, and that the DSR can be one of many applications in the cloud.

» Mix and match – DSR is a network of DSR signaling sites. The deployment infrastructure at each site can vary. E.g. bare-metal at one site, and then cloud deployed at another location.

The above deployments can include HP c-Class, rack-mount servers (HP or Oracle) or cloud deployable. A cloud deployable DSR is a DSR that is ready and able to be deployed into a number of different cloud environments, including but not limited to:

» A customer provided shared resource infrastructure. The DSR is simply one of many applications.
» A dedicated private cloud. The DSR may be the only application, or one of a small set of applications. Services and infrastructure may also be provided by Oracle and deployed at customer's sites. Often (but not necessarily) this is a deployment tuned specifically for the DSR.
» A hosted cloud. The DSR is deployed in an Oracle or operator hosting cloud, and end-customers rent or lease the DSR application from the hosting provider.

The DSR is capable of running on a huge variety of infrastructures, but not all infrastructures are the same and performance, capacity, and latency can vary dramatically based on the chosen infrastructure and how it is deployed. In general, the DSR works best in a high bandwidth, low-latency, high processing power environment (carrier grade cloud). The DSR 7.0.1 Cloud Benchmarking Guide – Available provides details of the infrastructure used for benchmark testing, including the hardware and software. It also describes key settings and attributes, and some recommendations on configuration.

The remainder of this document focuses on bare metal, hybrid, and virtualized RMS but does not reflect requirements for a cloud deployable DSR.

Key Components

DSR network elements are deployed in geographically diverse mated pairs with each NE servicing signaling traffic to/from a collection of Diameter clients, servers and agents. The DSR Message Processor (MP) provides the Diameter message handling function and each DSR MP supports connections to all Diameter peers (defined as an element to which the DSR has a direct transport connection). DSR supports a three tiered OAM architecture with a centralized Network OAM as shown in Figure 2. The Network OAM can be virtualized onto the System OAM blades at one of the signaling sites (for small systems) or deployed as standalone rack mount servers either co-located or separate from the signaling sites. A two tiered OAM architecture (without Network OAM) is no longer supported.

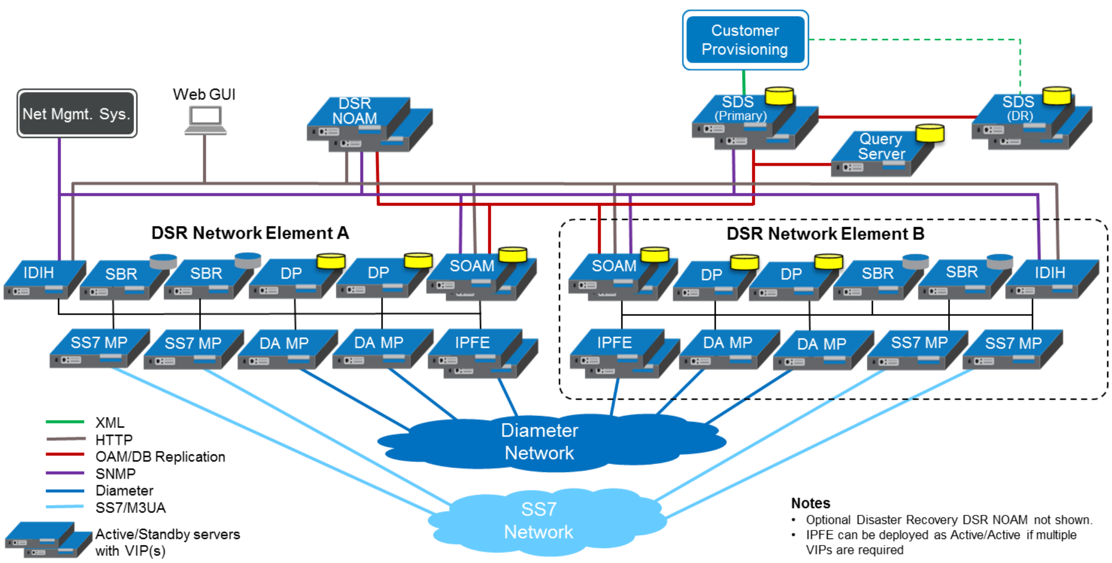

Figure 2. DSR System Architecture

A DSR consists of:

- A Network OAM network element that can be co-located with the database provisioning node or signaling node, and may optionally include a disaster recovery node
- At least one signaling node, and may optionally include up to 31 additional nodes
- A single network may contain up to 32 DSR signaling nodes
- An optional database provisioning node (SDS), and may optionally a include disaster recovery node
- One database provisioning system (SDS nodes) normally serves all the DSRs in a network.

The key components of the solution are:

- Operations, Alarms and Maintenance (OAM) - Required
    - System OAM per signaling node
    - Network OAM
- Diameter Agent Message Processor (DA-MP) - Required
- Management Server (PM&C) - Required
- IP Front End (IPFE) - Optional
- Integrated Intelligence Hub (IDIH) – Optional
- Diameter Intelligence Hub (DIH) – Existing deployments only
- Subscriber Database Server (SDS) - Optional
- Query Server (QS) - Optional
- Database Processor (DP) - Optional
- Session Binding Repository (SBR) – Optional
- SS7 Message Processor (SS7 MP) - Optional

Operations, Alarms and Maintenance

The Operations, Administration and Maintenance components of the DSR include the System OAM located at each signaling node and the Network OAM (NOAM).

Key characteristics of the System OAM at each signaling node are as follows:

- centralized OAM interface for the node
- maintains local copy of the configuration database
- supports SNMP northbound interface to operations support systems for fault management
- maintains event and security logs
- centralizes collection and access to measurements and reports

Key characteristics of the Network OAM are as follows:

- centralized OAM for the DSR network
- supports SNMP northbound interface to operations support systems for fault management
- runs on a pair of servers in active/standby configuration or can be virtualized on the System OAM blades at one signaling site (for small systems with two DSR signaling nodes only)
- optionally supports Disaster Recovery site for geographic redundancy
- provides configuration and management of topology data
- provides mechanism to create user groups with various access levels
- maintains event and security logs

» centralizes collection and access to measurements and reports

Diameter Agent Message Processor (DA-MP)

The DA-MP provides the core routing and processing functionality for the DSR and scales by adding blades.

Key characteristics of a DA-MP are as follows:

» provides application specific handling of real-time Diameter messages
» accesses DPs for real-time version of the subscriber DB, as needed
» interfaces with System OAM

Management Server (PM&C)

Provides the management of the overall platform.

Key characteristics of the management server are as follows:

» not a message processing critical component of the DSR
» deployed as a single server per site but can be supported as a redundant server

IP Front End (IPFE)

The DSR IP Front End provides TCP/SCTP connection based load balancing that makes a large DSR accessible to incoming connections through a minimal number of IP addresses.

Key characteristics of an IPFE are as follows:

» optional but highly recommended component of the DSR
» supports up to two active / standby pairs each supporting 3.2 Gbps bandwidth
» supports TCP, uni-homed SCTP and multi-homed SCTP

Integrated Diameter Intelligence Hub (IDIH)

The IDIH supports advanced troubleshooting for Diameter traffic handled by the DSR.

Key characteristics of the IDIH are as follows:

» optional but highly recommended component of the DSR
» integrates the configuration of trace filters directly into the DSR
» trace filters can be applied against all messages processed by DSR
» transaction data records containing both Diameter Requests and Answers are captured for additional analysis and filtering
» ladder diagrams facilitate troubleshooting concerning message sequence and peer connection issues
» supports detailed decoding of AVPs associated with Diameter protocols

Diameter Intelligence Hub (DIH)

The DIH is supported for existing deployments only. The DIH supports advanced troubleshooting for Diameter traffic handled by the DSR.

Subscriber Database Server (SDS)

The SDS provides a centralized provisioning system for distributed subscriber data repository. The SDS is a highly-scalable database with flexible schema.

Key characteristics of the SDS are as follows:

» interfaces with provisioning systems to provision subscriber related data
» interfaces with OAMs at each DSR network element
» replicates data to multiple sites
» stores and maintains the master copy of the subscriber database
» supports bulk import of subscriber data
» provides web based GUI for provisioning, configuration and administration of the data
» supports SNMP northbound interface to operations support systems for fault management
» provides mechanism to create user groups with various access levels
» provides continuous automated audit to maintain integrity of the database
» supports backup and restore of the subscriber database
» runs on a pair of servers in active / hot standby
» optionally supports Disaster Recovery site for geographic redundancy

Query Server (QS)

The Query Server contains a replicated copy of the local SDS database and supports a northbound SQL interface for free-form verification queries of the SDS Provisioning Database. The Query Server's northbound SQL interface is accessible via its local server IP.

Key characteristics of the QS are as follows:

» optional component that contains a real-time, replicated instance of the subscriber DB
» provides SQL access for offline queries to the SDS Provisioning Database

Database Processor (DP)

The DP is the repository of subscriber data on the individual DSR node elements. The DP hosts the full address resolution database and scales by adding blades.

Key characteristics of a DP are as follows:

» provides high capacity real-time database query capability to DA-MPs
» interfaces with DP SOAM (application hosted on the same blade as the DSR SOAM) for provisioning of subscriber data and for measurements reporting across all DPs

Session Binding Repository (SBR)

The SBR stores diameter sessions and subscriber bindings for stateful applications. Two stateful applications are supported:

1. Charging Proxy Application (CPA) for off-line charging
2. Policy Charging Application (PCA) for Policy Diameter Relay Agent (Policy DRA) and Online Charging Diameter Relay Agent (OC-DRA).

The off-line charging application uses charging SBRs (cSBR), OC-DRA uses session database SBRs (SBR(s)) and Policy DRA uses both session database SBRs (SBR(s)) and subscriber binding database SBR's (SBR(b)). Throughout this document the SBRs are referred to individually when there are significant differences discussed, and referred as SBR, without distinguishing the application, when the attribute applies to all types. The SBR scales by adding blades.

Key characteristics of an SBR are as follows:

» optional component of the DSR
» provides repository for subscriber and session state data
» provides DSRs with network-wide access to bindings

SS7 Message Processor (SS7 MP)

The SS7 MP provides the MAP Diameter interworking function and scales by adding blades.

Key characteristics of an SS7 MP are as follows:

» performs message content conversion between MAP and Diameter
» performs address mapping between SS7 (SCCP/MTP) and Diameter
» supports 3G<->LTE authentication interworking as needed
» interfaces with System OAM

DSR Configuration

The DSR Network consists of a Network OAM, DSR Signaling nodes and optionally a Subscriber Database node. Each of these elements and the configuration attributes are described in the following sections. This document focuses on requirements for bare metal, hybrid, and virtualized RMS. Please refer to DSR 7.0.1 Cloud Benchmarking Guide – Available on MyOracleSupportfor all information regarding cloud deployable DSR.

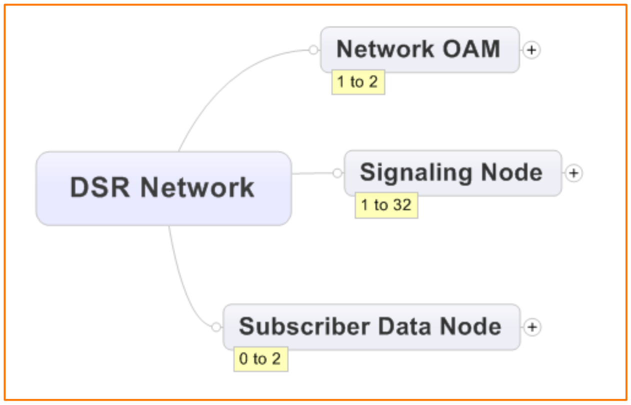

Figure 3 : DSR Configuration

Network OAM Network Element

DSR supports three tiered OAM which includes a Network OAM network element, System OAM and OAM functions on the Diameter Agent Message Processor. Refer to the Feature Guide for more information on three tiered OAM functionality. Two tiered OAM is no longer supported in this release.

Each DSR network requires at least one Network OAM and optionally a disaster recovery Network OAM.

There are several attributes that define a DSR Network OAM network element and these are defined in the below diagram. Items in green are the default selection and white items are optional. This section describes each of these attributes.

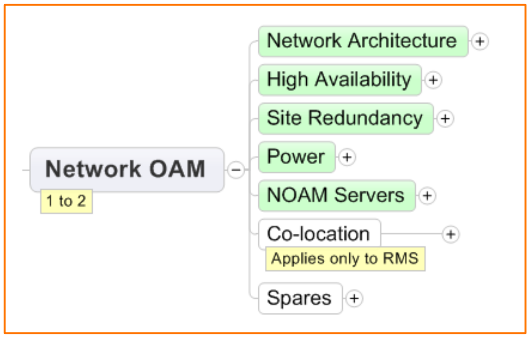

Figure 4 : Network OAM Configuration Attributes

Network Architecture

The network architecture shown below applies to the cases where the NOAM NE is deployed on rack mount servers. In the cases where the deployment of the Network OAM is co-resident on C-Class blades, the networking for the DSR Signaling node is applicable.

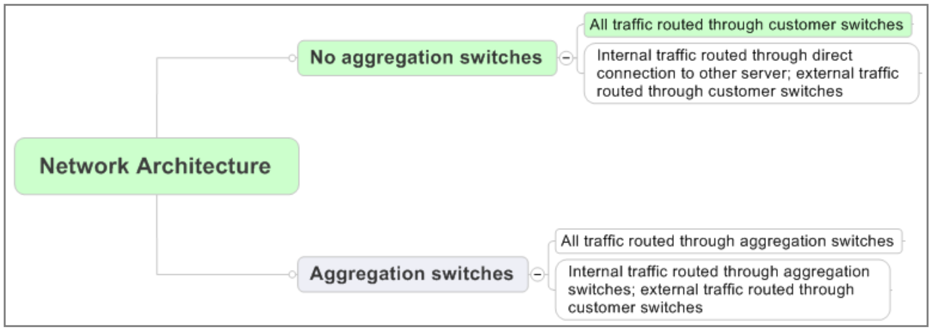

Figure 5 : Network OAM Network Architecture

High Availability

The NOAM provides a network OAM and is not considered to be a real-time signaling critical component of a DSR network deployment. However, customers requiring an NOAM, generally require that the NOAM is highly available.

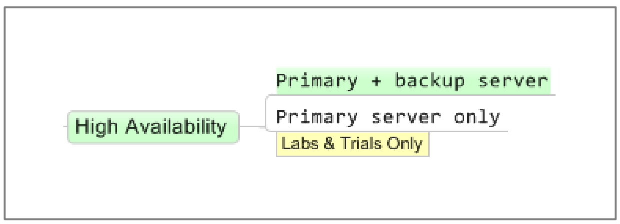

Figure 6: NOAM High Availability

Site Redundancy

The NOAM is not considered to be a real-time critical component of a DSR network deployment. However, a disaster recovery node is optionally available to provide site redundancy in cases of a complete loss of a site.

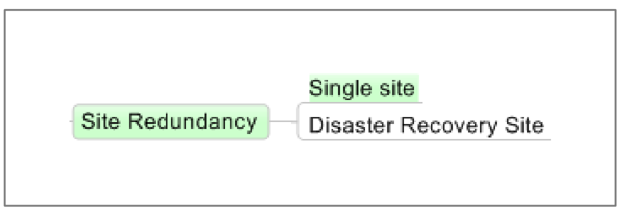

Figure 7: NOAM Site Redundancy

Power

The NOAM can be deployed with either AC or DC power. DC power is supported on HP hardware only.

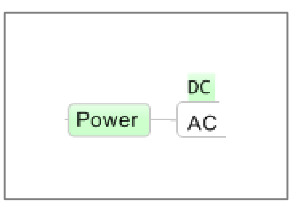

Figure 8: NOAM Power

NOAM Servers

The NOAM can be deployed on dedicated rack-mount servers either at a standalone site or co-located at one of the DSR signaling sites. If co-located at one of the DSR c-Class signaling sites, the active or standby NOAM can be co-resident on the management server. The NOAM can also be deployed on the System OAM blades at one of the DSR c-Class signaling sites. This option is supported for small systems with no more than two signaling nodes. If co-located at one of the DSR RMS signaling sites, the NOAM is co-resident on the core RMS servers.

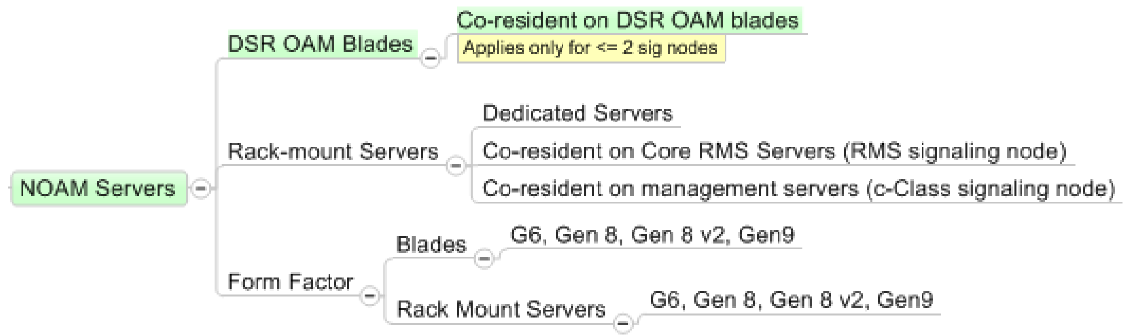

Figure 9: NOAM Servers

Co-Location

The rack mount server form factor of the NOAM may be co-located with one or more products. Co-location is the ability for the NOAM to be physically located in the same site footprint (cabinets) as other products.

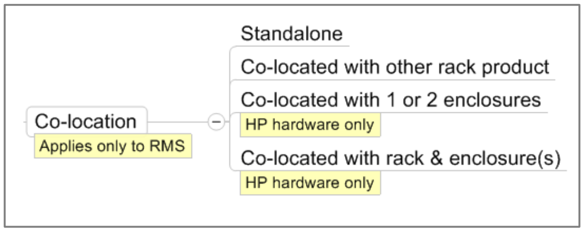

Figure 10: NOAM Co-location

Co-located products share resources such as cabinet and power. Examples of co-located products with NOAM are:

» NOAM rack mount + DSR signaling node in an enclosure
» NOAM rack mount + SDS rack mount
» NOAM rack mount + Policy rack mount

Spares

Spares are optional for the NOAM since it is not a real-time critical signaling component.

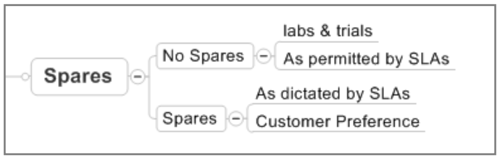

Figure 11: NOAM Spares

c-Class Signaling Node

There are several attributes that define a DSR signaling node and these are identified in the node attributes diagram below. Items in green are the default selection and white items are optional. This section describes each of these attributes.

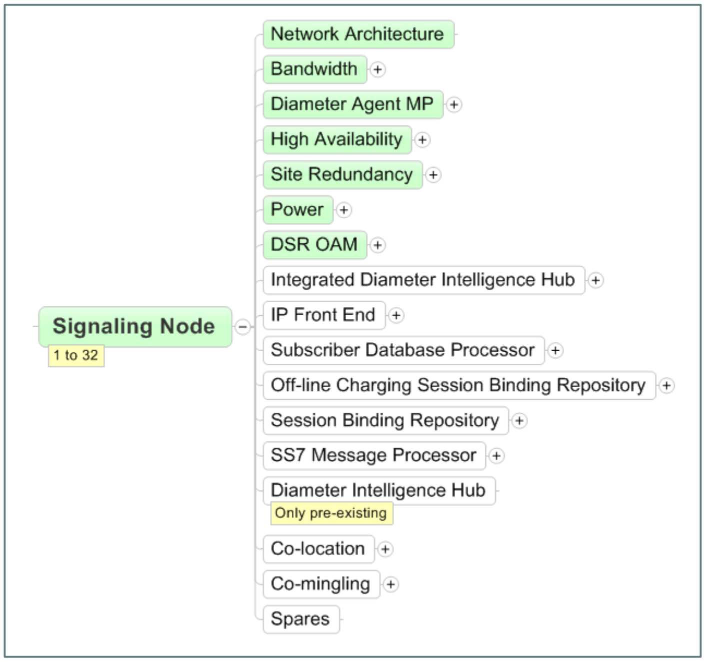

Figure 12 : DSR Signaling Node Attributes

Network Architecture

Customers have differing requirements for how to implement DSR signaling nodes into their networks. These requirements include support for both layer 2 and layer 3 demarcations and support for physical separation of the different types of traffic. The following diagram shows the network architecture options for DSR signaling nodes. Green indicates the default shipping baseline.

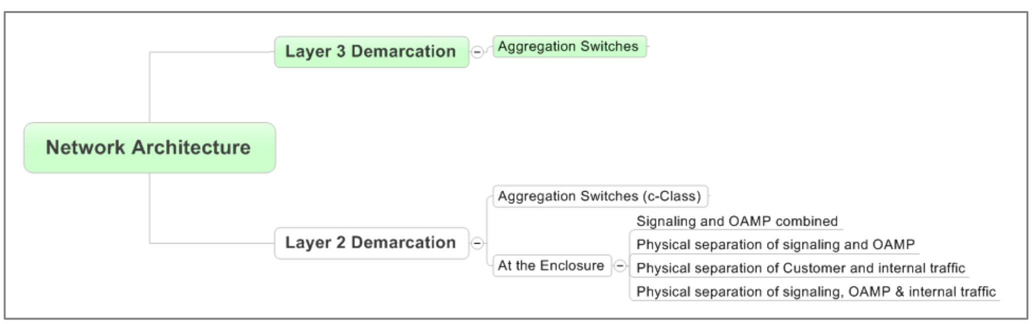

Figure 13: Network Architecture

Each DSR signaling node will use one network architecture:

» Layer 3 Demarcation with aggregation switches
» Layer 2 Demarcation with aggregation switches
» Layer 2 Demarcation at the enclosure with:

» Signaling & OAM traffic combined, OR
» Physical separation of signaling and OAM traffic, OR
» Physical separation of customer and internal traffic, OR
» Physical separation of signaling, OAM and internal traffic

These options result in multiple network topologies which can be supported based on customer needs. Topologies differ from each other in number of physical networks supported and whether aggregation switches are present. Each Topology can have multiple variations based on whether 1G or 10G connectivity is required and whether backup network is needed.

TABLE 1: DSR SIGNALING NODE NETWORK TOPOLOGIES

Topology	Equipment Needed	Purpose
1	One pair of enclosure switches per enclosure. One pair of aggregation switches	Provide aggregated uplinks to the customer network with Layer-3 demarcation.
2	One pair of enclosure switches per enclosure. One pair of aggregation switches	Provide aggregated uplinks to the customer network with Layer-2 demarcation.
3	One pair of enclosure switches per enclosure.	Provide integration with Customer L2 domain where aggregation switches are not used.
4	Two pairs of enclosure switches per enclosure.	Provide physical separation of traffic between OAM and Signaling.
5	Three pairs of enclosure switches per enclosure.	Provide physical separation of traffic between OAM and two separate Signaling networks.
6	Three pairs of enclosure switches per enclosure.	Provide physical separation of inter-enclosure traffic from customer's OAM and Signaling networks.

Bandwidth

Enclosure bandwidth is driven by the traffic sent to and from the blades within a given enclosure. The amount of bandwidth needed in certain situations drives the size and number of switches required.

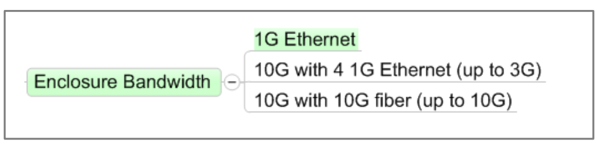

Figure 14: Enclosure Bandwidth

» DSR supports 1G enclosure switches. This implies a 1G connection from the blade to the enclosure switch.
» DSR supports 10 G enclosure switches. This implies a 10G connection from the blade to the enclosure switch.
» DSR supports 10 G fiber connection from the enclosure to the customer's switches or to another enclosure switch when 10G enclosure switches are used.
» DSR supports 4 x 1 G connections from the enclosure to the aggregation switches, customer's switches or to another enclosure switch when 10G or 1G enclosure switches are used.

When multiple physical networks are used, and therefore multiple sets of enclosure switches are used, then each physical network may use 1G or 10G enclosure switches depending on the bandwidth requirements of the specific deployment.

Diameter Agent Message Processors

The DSR Diameter Agent message processors (DA-MPs) provide the core routing and processing functionality for the DSR.

The DSR c-Class form factor supports DA-MPs in an active-active configuration for scaling up to sixteen total DA-MPs per DSR signaling node. DSR also supports DA-MPs in an active-standby configuration for up to two DA-MPs per DSR signaling node.

DSR supports simplex configurations for labs and trials. Refer to the Platform Feature Guide [2] for more details on the supported hardware.

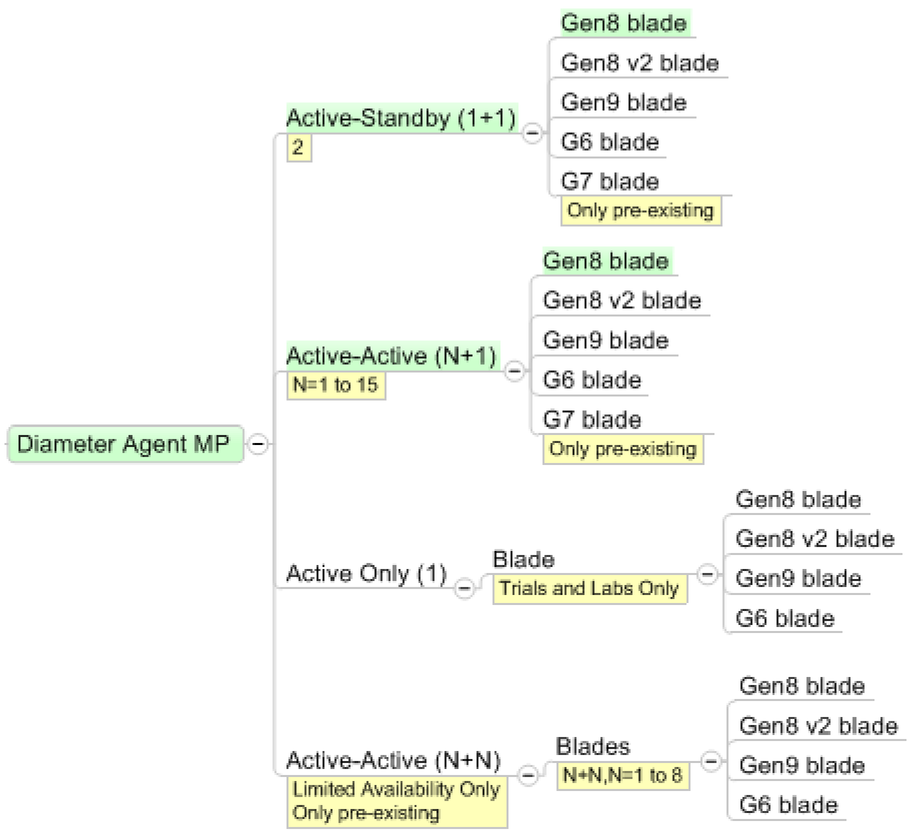

Figure 15: Diameter Agent MP Configurations

High Availability

High availability is an expected attribute of any core signaling product. DSR is no exception. A typical DSR node is engineered with high levels of redundancy to ensure a very high availability of the system at all times.

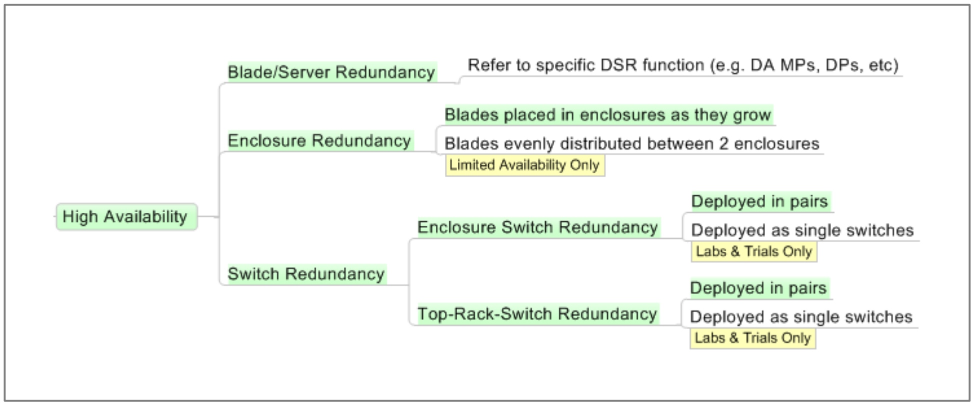

Figure 16 : DSR High Availability

A DSR site supports redundant:

- Power
- Enclosure switches
- Aggregation (top-rack) switches and pass-thru modules (when used)
- Application blades or servers where applicable (DA MP, DP, SBR, IPFE)
- System OAM blades

Some DSR customers require very high nodal availability, and for these customers, backplane enclosure redundancy is also supported. When required, a DSR node will support:

- Redundant enclosures
  - Note: this is different than a normal two enclosure configuration where more slots are required. In this case, there may be enough capacity to house all the blades in one enclosure, but the customer’s redundancy requirements dictate the need for the additional enclosure.
- When redundant enclosures are used, SOAM, DA-MP, and DP blades are split evenly between enclosures.

The management server is not a message processing critical component of the DSR, and is deployed as a single server per site. In cases where DSR is co-mingled with another product that uses a redundant management server, the DSR supports the redundant management server.

The Diameter Intelligent Hub is not a message processing critical component of the DSR, and is deployed as a single server or blade (existing configurations only) per node.

The SS7 MP is not HA.

For cost, footprint and power efficiencies in a lab environment, a DSR lab can be deployed with a single:

- Power
- Enclosure switch
- Aggregation (top-rack) switch and pass-thru module (when used)
- Application blades or servers where applicable (DA-MP, DP, SBR, IPFE)
- System OAM

Site Redundancy

The DSR can be deployed in mission critical core routing of networks, or in less critical areas of the network such as off-line billing. Since there are a variety of needs for high system availability, the DSR can be deployed with or without geo-redundancy.

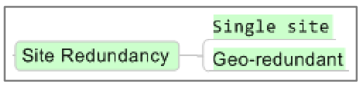

Figure 17 : Site Redundancy

The DSR can be deployed as a single node at a single site or with geo-redundancy. Normally, this is a mated pair. DSR also supports a geo-redundant model that allows for a mated-triple. In this model, any one of the three nodes forming the mated triple could take all the traffic of the three nodes. Although mated triplets are supported, they are not normally recommended due to the cost, added complexity of configuration and operational maintenance of an extra signaling node. The mated pair configuration supports 99.999% availability.

Power

The DSR signaling node supports DC and AC International and AC North America power supplies.

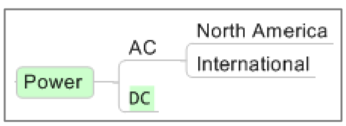

Figure 18 : Site Power

DSR OAM

The DSR requires that System OAM is present in each DSR node. The platform requires a management server at each site. These two components form the basis of the overall management of the DSR signaling node.

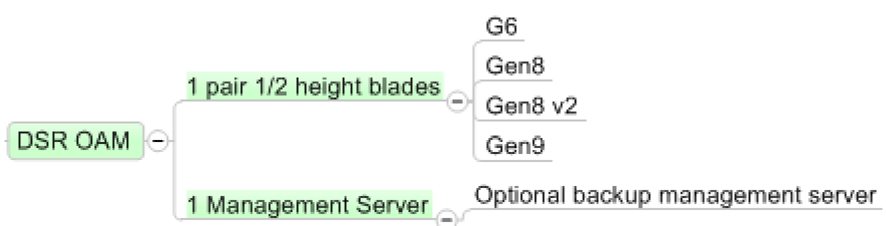

Figure 19 : System OAM&P

» System OAM blades are deployed in active-standby configuration with two blades per node.
» DSR uses a non-redundant management rack-mount server. This management server may be shared with other products using c-Class at the same site.

Enclosures

DSR is supported in up to three HP C-Class enclosures per node. Additional enclosures are supported for future expansion and for co-mingling applications.

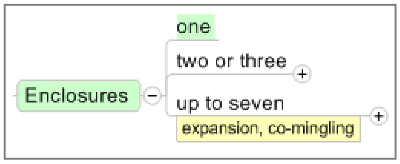

Figure 20 : Enclosures

The DSR blades may be placed in any position within an enclosure except as noted below. There are no restrictions (beyond physical) on where blades can be placed in an enclosure. That is, DA-MPs, DPs, DIH, and System OAMs, can go in any position within an enclosure. The following are the current known physical limitations:

- The DIH storage sidecar must go to the immediate right of the DIH blade (existing deployments only).
- Dividers to support ½ height blades span 2 slots (odd + even slot to the right), so full height blades may not be configured into slots where the ½ height dividers exist

In a multi-enclosure DSR configuration, the DSR blades may be placed in any enclosure. There are no restrictions (beyond physical) on which types of blades can be placed in which enclosures. In a multi-enclosure DSR configuration, blades with the same function may be split into different enclosures. For example: one System OAM, DA-MP, and DP in one enclosure with the redundant mates in a second enclosure.

Integrated Diameter Intelligence Hub (IDIH)

The Integrated Diameter Intelligence Hub is an optional but normal component of most DSR deployments. IDIH provides troubleshooting capabilities for the DSR and supersedes the Diameter Intelligence Hub described below. All new deployments requiring DSR troubleshooting will use IDIH.

The IDIH function is provided on a single server per signaling node. IDIH is supported on the rack mount server form factor for all new deployments. IDIH can be implemented on a c-Class blade for existing DIH deployments only by repurposing the existing DIH blade. DIH rack mount server hardware at existing deployments can also be repurposed for IDIH.

DSR supports one IDIH per signaling node. DIH and IDIH can run simultaneously on the same node.

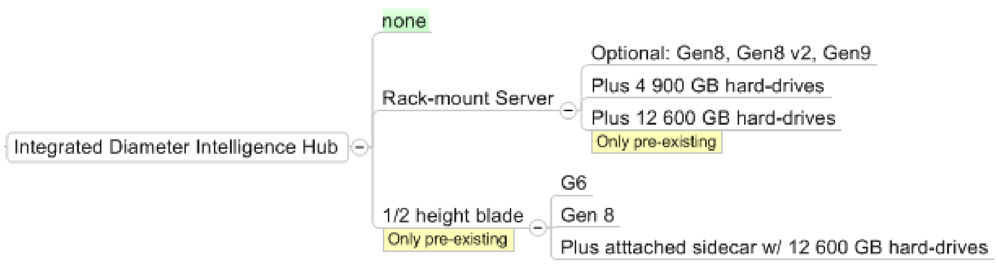

Figure 21 : Integrated Diameter Intelligence Hub

New deployments of IDIH servers allow for co-residence with the management server function. In other words, one server can be used for IDIH plus the management server function. IDIH cannot be co-resident with the NOAM function.

IP Front End

The DSR IP Front End provides TCP/SCTP connection based load balancing to hide the internal hardware architecture and IP addresses from the customer network. This is an optional component in a DSR signaling node.

IPFE blades can be deployed with an active-standby or active-active configuration. In either configuration, the maximum number of supported IPFE blades is four per DSR signaling node.

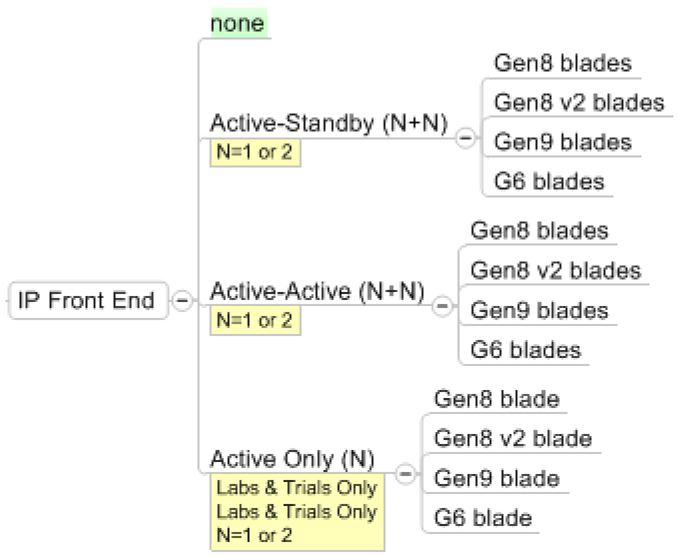

Figure 22 : IP Front End

Subscriber Database Processor

The DSR subscriber database processor (DP) provides real-time access to a subscriber database. This is an optional component in a DSR signaling node. DPs are deployed in conjunction with the Subscriber Database Servers (SDS) to provide full subscriber address resolution capability. DP blades are deployed in an active-active configuration up to a maximum of ten blades per DSR signaling node.

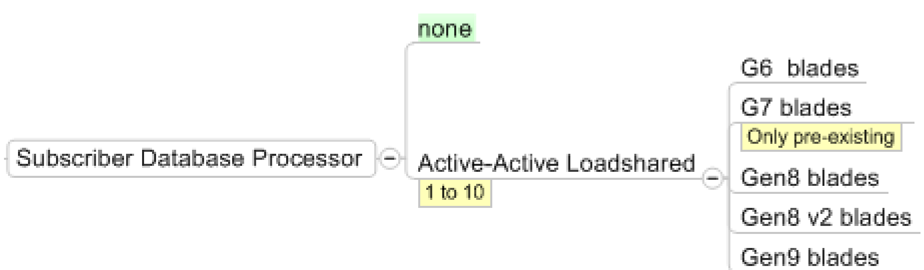

Figure 23 : Subscriber Database Processor

Off-Line Charging Session Binding Repository

The DSR Charging SBR provides a repository of session data required by the Charging Proxy Application (CPA). This is an optional component in a DSR node. The Off-line Charging Session Binding Repository is available as a non-geographically redundant configuration only.

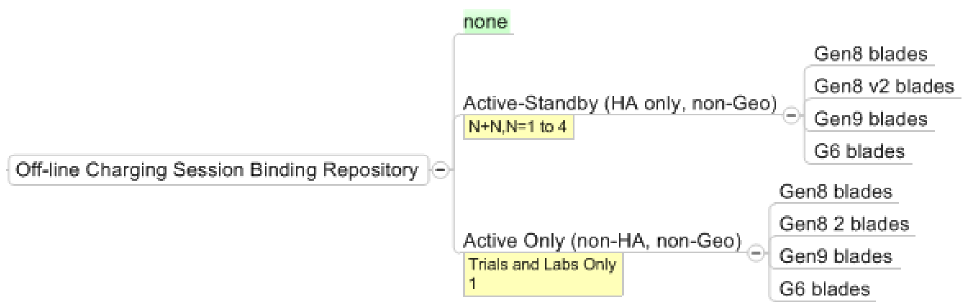

Figure 24: Off-Line Charging SBR Configurations

Session Binding Repository

The Session Binding Repository (SBR) provides a repository of session data and subscriber binding data required by the Policy and Charging Application (PCA). The PCA supports both the Policy Diameter Relay Agent (Policy DRA) and Online Charging Diameter Relay Agent (OC-DRA). The SBR provides two primary functions in the architecture:

- SBR(b) - A subscriber binding database – which maps subscribers to specific PCRFs (or other element)
- SBR(s) - A session database – which maintains the state of policy and/or charging related Diameter sessions

In this release of DSR, the OC-DRA uses the session database but does not currently use the subscriber binding database. Policy DRA uses both the session database and the subscriber binding database.

Throughout this document subscriber binding and session databases are referred to individually when there are significant differences discussed, and referred as SBR when the databases can be treated as a generic resource.

DSR supports an optional Gateway Location Application (GLA) on nodes that have access to the SBR databases. This application provides a Diameter signaling mechanism to access a subset of the state in the SBR databases. If GLA is implemented, one additional blade is equipped per network to allow for the query capability.

The SBRs are pooled and distributed resources for maintaining state information. The state information is scoped differently based on the database function. For example, in the Policy DRA application:

- Subscriber binding state is scoped for the entire Policy DRA network. The pool of SBR(b)'s maintaining subscriber binding are used by the entire set of DSR signaling nodes running the Policy DRA application in a given network.
  - Example: in an eight node network, all eight nodes communicate and use the subscriber binding database.
- Session state is scoped for a DSR mated pair. The pool of SBR(s)'s maintaining the session state is shared between a signaling node and its mate. Note: a mate is not always used by all operators.
  - Example: continuing the eight node example above – this network is comprised of four mated pairs. Each mated pair has its own session database. So there are four session databases.

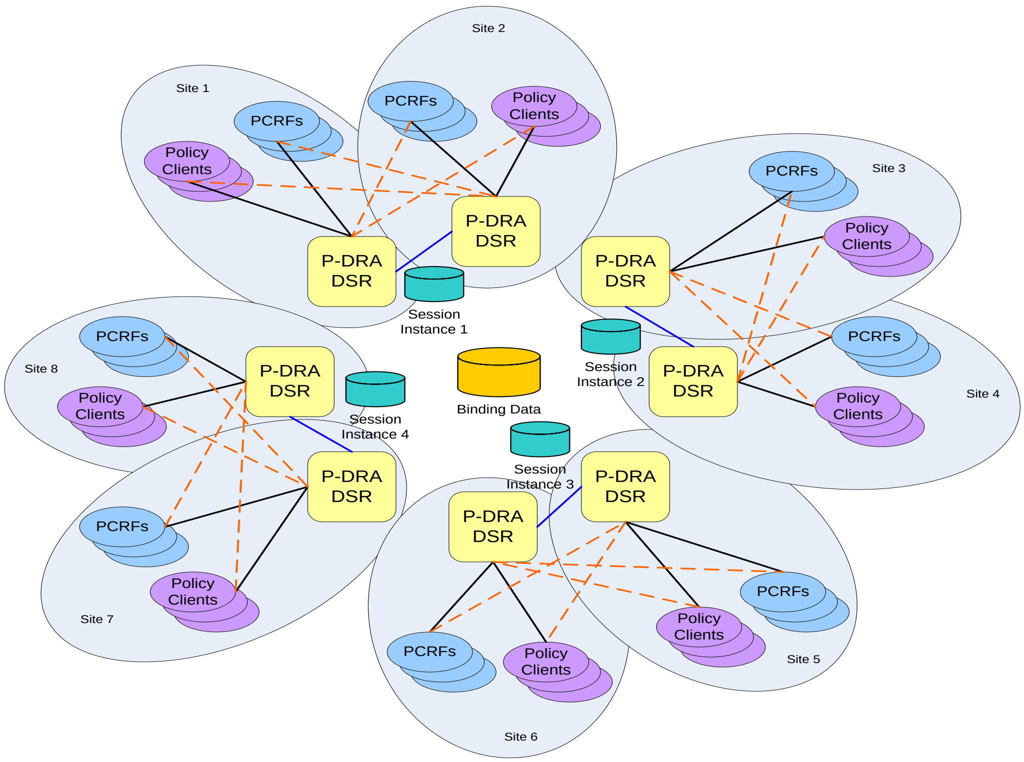

Figure 25: Policy DRA Database Architecture

In summary, the session database is active in every DSR signaling node running the Policy DRA application, but the subscriber binding database is normally only active in a subset of the DSRs. The following figure shows the SBR relationship with the two databases and the deployment architecture (form factor).

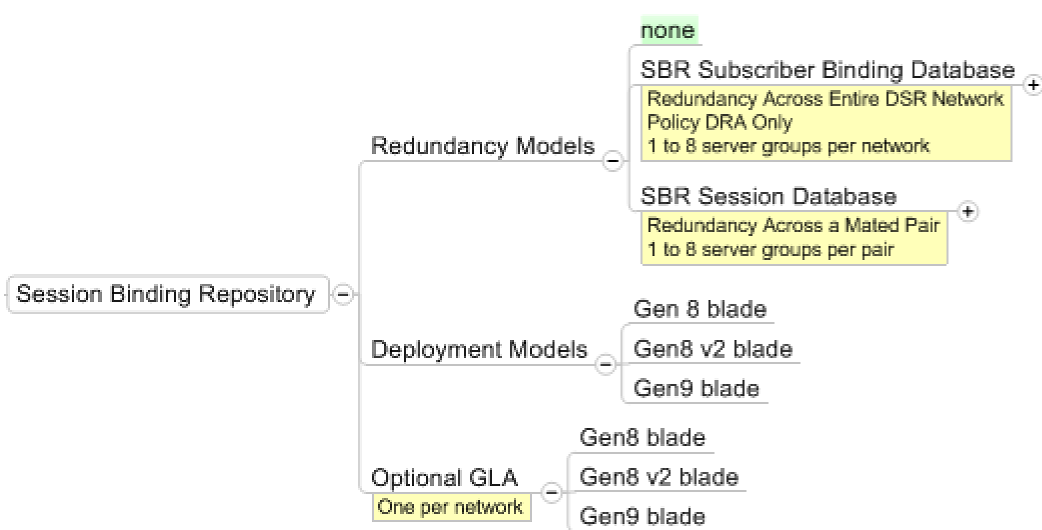

Figure 26: DSR SBR

To describe the configuration of the SBRs, the concept of a server group is introduced. A server group is a collection of servers that work together to provide a specified functional behavior. In this case, a server group provides SBR functionality. Each SBR Server Group consists of one, two, three or four servers, depending on the redundancy model desired. The PCA implementation requires that each mated pair in a network has the same number of session SBR(s) server groups. The following figure shows the supported redundancy configurations for SBR Server Groups:

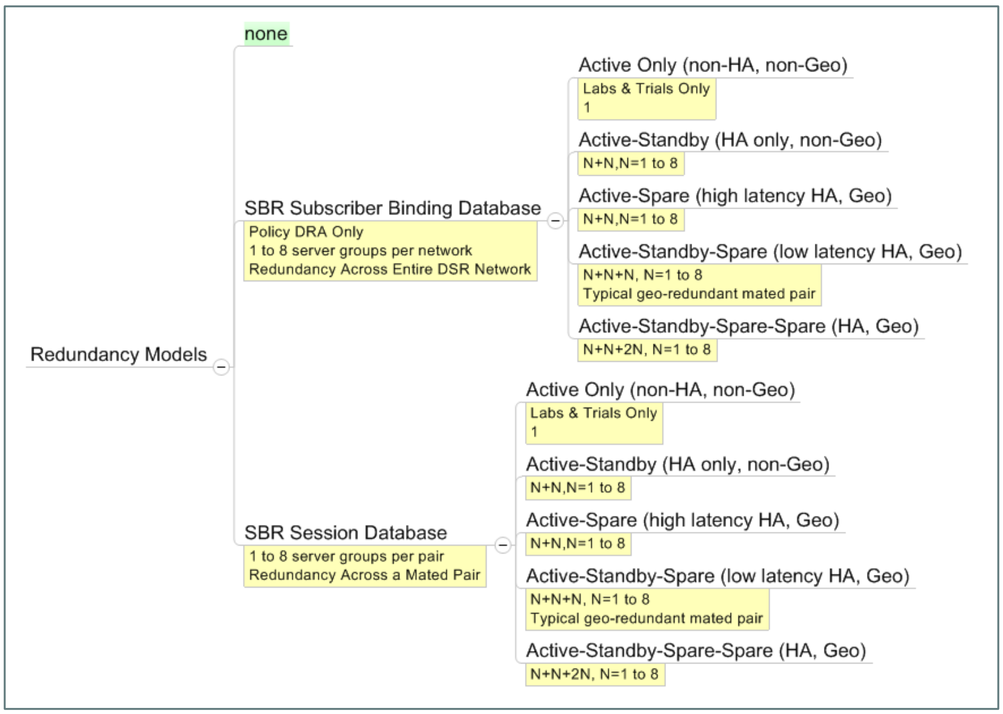

Figure 27: SBR Redundancy Model

SS7 Message Processor (SS7 MP)

The SS7 Message Processor is an optional component of the DSR which is required to support the MAP Diameter interworking function (IWF). MAP Diameter IWF is supported for both c-Class and rack mount server signaling node configurations. The SS7 MP is supported on Gen8 or Gen9 hardware configurations but can be deployed in signaling nodes with G6, G7,Gen8, and Gen9 hardware. However, the DA-MPs associated with the MAP Diameter IWF must be either G6,Gen8 or Gen9 hardware.

The following figure shows the redundancy models and server counts supported for the SS7 MP.

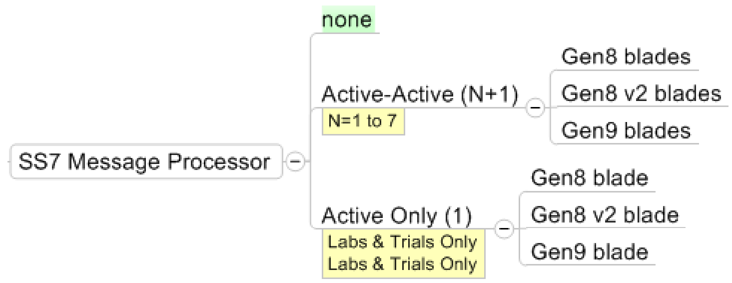

Figure 28: SS7 MP Configurations

Diameter Intelligence Hub (DIH)

The Diameter Intelligence Hub is an optional component of DSR deployments and is superseded by the Integrated Diameter Intelligence Hub described above. The DIH is supported for existing deployments only. The DIH provides troubleshooting capabilities for the DSR.

The DIH function is provided on a single server per signaling node. DIH deployment requires use of external switches to provide port mirroring for traffic capture.

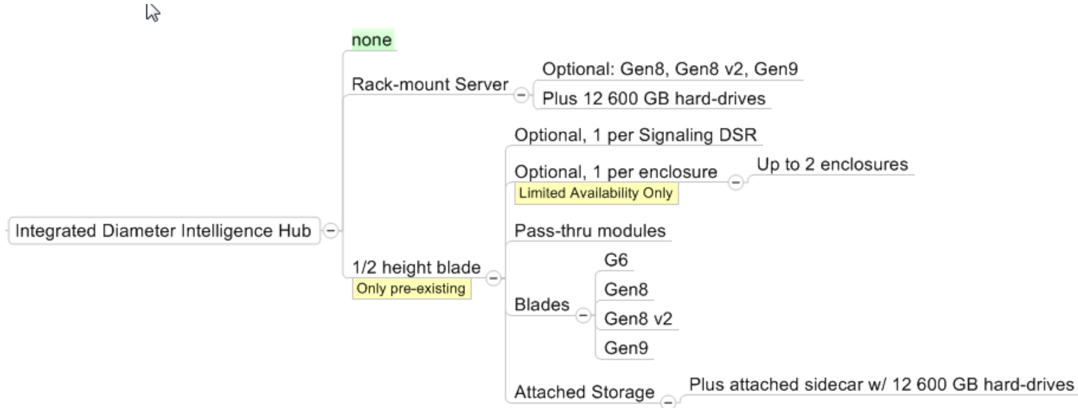

Figure 29 : Diameter Intelligence Hub

A DIH in c-Class form factor is supported for existing deployments only and consists of several physical components as listed below. DSR supports one DIH per enclosure up to two enclosures in limited availability only.

- A DIH processing blade
- A storage (side-car) blade
- 12 X 600 GB hard drives
- 2 X 1 GB pass-thru modules

Co-location

The DSR may be co-located with one or more products. Co-location is the ability for the DSR to be physically located in the same site footprint (cabinets) as other products.

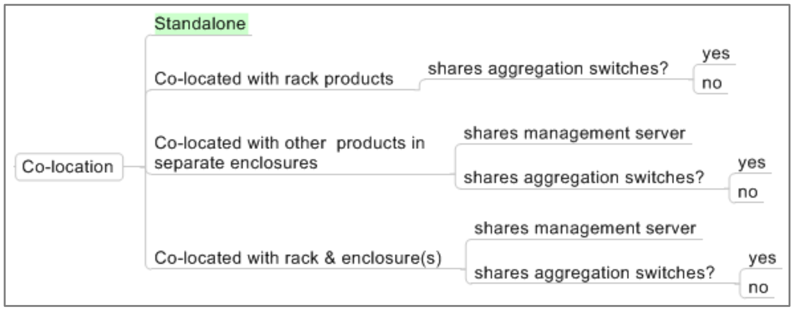

Figure 30 : DSR Co-location

Co-location allows for different products at the same site to share certain resources:

» Aggregation switches – when in use
» Management server – when applicable
» Physical cabinet and power

Here are some examples of co-located products with a DSR:

» DSR in an enclosure + rack mount SDM
» DSR in an enclosure + rack mount SDS (DSR subscriber database provisioning system)
» DSR in an enclosure + 2nd DSR in a second enclosure
» DSR in an enclosure + Policy Management in a second enclosure
» DSR in an enclosure + rack mount PIC PMF

Provided there is physical space, the DSR and one or more additional products are allowed to co-locate within the same physical cabinet. Co-locating products:

» share the same power supplies provided the power supplies are appropriately sized for the full co-located solution
» share the same aggregation switches provided the co-locating solutions are using aggregation switches
» share the same management server provided the co-locating solutions are using a management server

Co-mingling

The DSR may be co-mingled with one or more products. Co-mingling is the ability for the DSR to be physically located in the same enclosure(s) as other products.

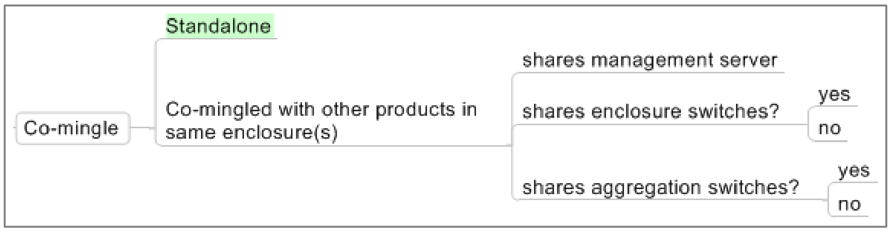

Figure 31 : DSR Co-mingling

Co-mingling allows for different products in the same enclosure(s) to share certain resources:

» Aggregation switches – when in use
» Management server
» Enclosures
» Enclosure switches

Here are some examples of co-mingled products with a DSR:

» DSR + SDM + Policy Management (sometimes called LTE-in-a-box)
» DSR + DSR

Provided there is physical space, the DSR and one or more additional products are allowed to co-mingle within the same physical enclosure.

» Co-mingling and co-location are allowed at the same site deployment.
» Co-mingling products are allowed to share the same power supplies provided the power supplies are appropriately sized for the full co-mingled solution.
» Co-mingling products are allowed to share the same aggregation switches provided the co-mingling solutions are using aggregation switches.
» Co-mingling products are allowed to share the same enclosure switches.
» Co-mingling products can use the same management server.

Spares

To maintain a high availability system, spare parts for key components are required.

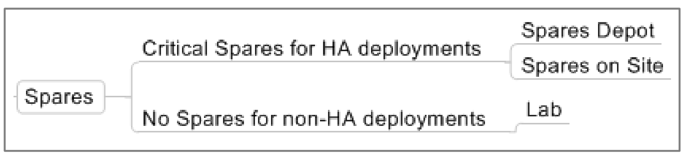

Figure 32 : Spares

Rack Mount Server (RMS) Signaling node

DSR supports a rack mount server (RMS) form factor for DSR signaling nodes. The rack mount server configuration makes use of virtualization to host multiple functions per server. The core RMS servers host the DA-MP, NOAM, SOAM, management server and IPFE functions as shown in the figure below. DP and SBR functions are not supported in the RMS form factor. Optional expansion servers (up to four) support three virtual machines (VM) which can be a mix of DA MPs and SS7 MPs. The IDIH is a series of virtual machines provided on a separate physical server in the rack mount server signaling node. DIH servers are supported for existing configurations only.

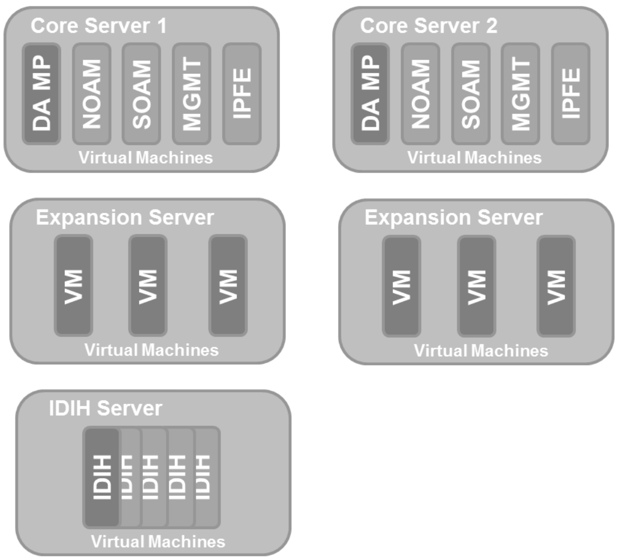

Figure 33: Rack Mount Server Signaling Node

There are several attributes that define a DSR rack mount signaling node and these are identified in the node attributes diagram below. Items in green are the default selection and white items are optional. This section describes each of these attributes.

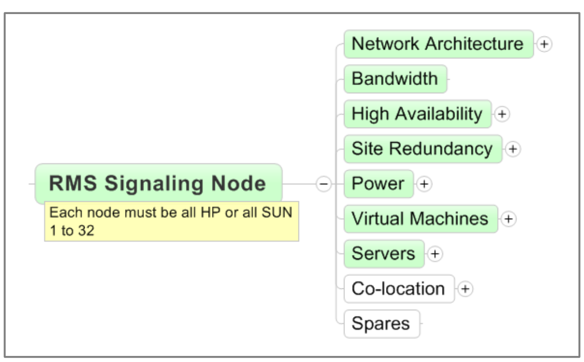

Figure 34 : RMS DSR Signaling Node Attributes

Network Architecture

Customers have differing requirements for how to implement DSR signaling nodes into their networks. These requirements include support for both layer 2 and layer 3 demarcations and support for physical separation of the different types of traffic. The following diagram shows the network architecture options for DSR RMS signaling nodes. Green indicates the default shipping baseline.

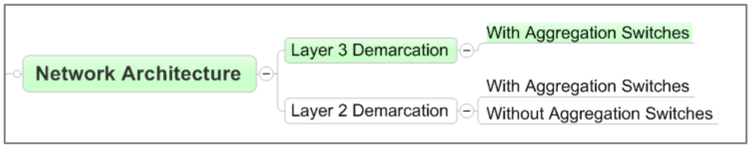

Figure 35: RMS Network Architecture

Each rack mount DSR signaling node will use one network architecture:

- Layer 3 Demarcation with aggregation switches
- Layer 2 Demarcation with aggregation switches
  - Physical separation of inter-RMS traffic from signaling and OAM traffic
- Layer 2 Demarcation for Rack Mount Server deployments without aggregation switches
  - Direct connection for inter-RMS traffic, separate signaling and OAM

These options result in multiple network topologies which can be supported based on customer needs. Topologies differ from each other in number of physical networks supported and whether aggregation switches are present. Each Topology can have multiple variations.

TABLE 2: DSR RMS SIGNALING NODE NETWORK TOPOLOGIES

Topology	Equipment Needed	Purpose
1	One pair of aggregation switches Rack Mount Server	Provide aggregated uplinks to the customer network with Layer-3 demarcation.
2	One pair of aggregation switches Rack Mount Server	Provide aggregated uplinks to the customer network with Layer-2 demarcation.
3	Rack Mount Server	Provide integration with Customer L2 domain where aggregation switches are not used.
4	Rack Mount Server	Provide physical separation of traffic between OAM and Signaling.
5	Rack Mount Server	Provide physical separation of traffic between OAM and two separate Signaling networks.
7	Special topology for two rack mount server deployments only. Does not require any Oracle supplied switches.	Provide direct connection from RMS to customer network(s) and provide ability to prevent inter-RMS traffic from impacting customer switches
8	One pair of aggregation switches Rack Mount Server	Provide physical separation of inter-RMS traffic from customer's OAM and Signaling networks in an RMS-only network.

Bandwidth

The DSR rack mount servers support eight 1G connections from the server to the aggregation switches and/or customer switches.

High Availability

High availability is an expected attribute of any core signaling product. DSR is no exception. A typical DSR node is engineered with high levels of redundancy to ensure a very high availability of the system at all times.

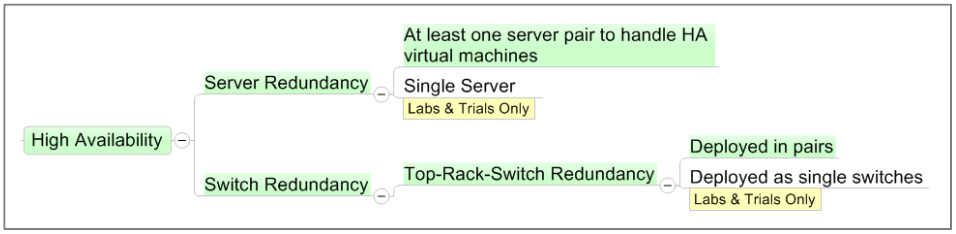

Figure 36: RMS High Availability

An RMS DSR site supports redundant:

» Power
» Aggregation (top-rack) switches, pass-thru modules (when used)
» System OAM, DA MP, IPFE (VMs)
» Network OAM if applicable (VMs)

The management server is not a message processing critical component of the DSR, and is deployed as a single virtual machine per site. In cases where DSR is co-mingled with another product that uses a redundant management server, the DSR supports the redundant management server.

The Diameter Intelligence Hub and Integrated Diameter Intelligence Hub are not message processing critical components of the DSR, and are each deployed on single servers per node.

The SS7 MP (VM) is not HA.

Single servers are supported for labs and trials.

Site Redundancy

The DSR can be deployed in mission critical core routing of networks, or in less critical areas of the network such as off-line billing. Since there are a variety of needs for high system availability, the DSR can be deployed with or without geo-redundancy.

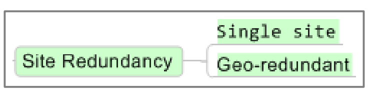

Figure 37 : RMS Site Redundancy

The DSR can be deployed as a single node at a single site. The DSR can be deployed with geo-redundancy. Normally, this is a mated pair.

Power

The DSR signaling node supports DC and AC power.. DC power is supported on HP hardware only.

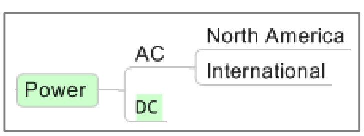

Figure 38 : RMS Site Power

Virtual Machines

As described above, the rack mount server configuration makes use of virtualization to host multiple functions per server. Each of these functions is described below.

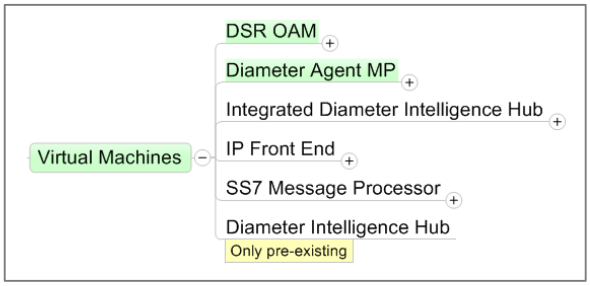

Figure 39 : RMS Virtual Machines

DSR OAM

The DSR requires that a System OAM is present in each DSR node. The platform requires a management server at each site. These two components form the basis of the overall management of the DSR signaling node. In the rack mount server form factor, the System OAM and management server functions are virtualized onto the core servers with the DA-MP, NOAM (if applicable) and IPFE (when used) functions. Refer to Figure 33: Rack Mount Server Signaling Node. Core servers are deployed as a server pair for high availability. DSR also supports simplex configurations for labs and trials.

Diameter Agent Message Processors

The DSR Diameter Agent message processors (DA-MPs) provide the core routing and processing functionality for the DSR. The rack mount server form factor supports a minimum of two DA MPs per production node in an active-active configuration. Additional DA-MP virtual machines can be added through the addition of expansion servers. Each DA MP supports up to 10k MPS but in this release the total nodal capacity for an RMS signaling node is 50k MPS.. Refer to Figure 33: Rack Mount Server Signaling Node. DA MP VMs are deployed with an N+K redundancy model to provide high availability. DSR also supports simplex configurations for labs and trials.

Integrated Diameter Intelligence Hub (IDIH)

The Integrated Diameter Intelligence Hub is an optional but normal component of most DSR deployments. IDIH provides troubleshooting capabilities for the DSR and supersedes the Diameter Intelligence Hub described below. All new deployments requiring DSR troubleshooting will use IDIH. Refer to Figure 33: Rack Mount Server Signaling Node. The IDIH function is provided on a single server per RMS signaling node. DIH rack mount server hardware at existing deployments can be repurposed for IDIH. DIH and IDIH can run simultaneously on the same node.

IP Front End

The DSR IP Front End provides TCP/SCTP connection based load balancing to hide the internal hardware architecture and IP addresses from the customer network. An RMS DSR always includes IPFE and it is a deployment decision whether or not to use it. In the rack mount server form factor, the IPFE function is virtualized onto the core servers with the DA-MP, NOAM, System OAM and management server functions. Refer to Figure 33: Rack Mount Server Signaling Node. Core servers are deployed as a server pair for high availability. DSR also supports simplex configurations for labs and trials.

SS7 Message Processor (SS7 MP)

The SS7 Message Processor is an optional component of the DSR which is required to support the MAP Diameter interworking function (IWF). The SS7 MP virtual machines are deployed on expansion servers. Refer to Figure 33: Rack Mount Server Signaling Node. Each SS7 MP supports up to 10K MAP MPS but in this release the total nodal capacity for an RMS signaling node is 40k MAP MPS. Two SS7 MP VMs are supported on each expansion server up to a maximum of four SS7 MP VMs.

Diameter Intelligence Hub (DIH)

The Diameter Intelligence Hub is an optional component of DSR deployments and is superseded by the Integrated Diameter Intelligence Hub described above. The DIH is supported for existing deployments only. The DIH provides troubleshooting capabilities for the DSR. The DIH function is provided on a single server per signaling node. DIH deployment requires use of external switches to provide port mirroring for traffic capture.

Servers

The RMS signaling node supports HP hardware configurations. Where applicable, HP hardware versions can be mixed in a single node.

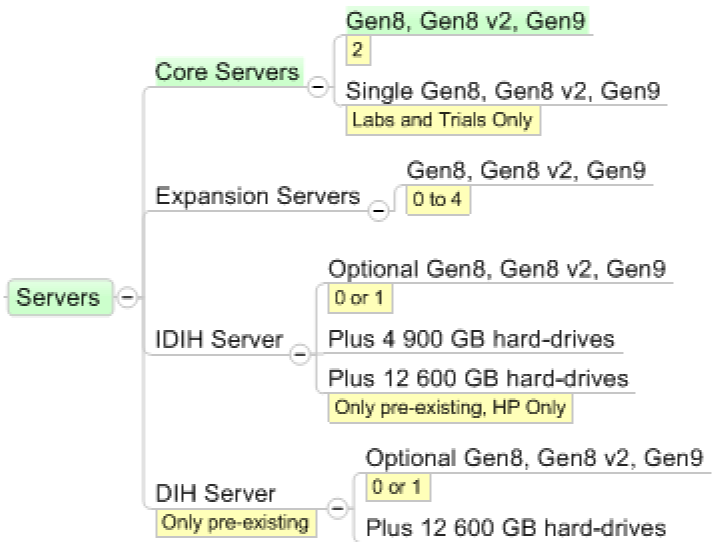

Figure 40: RMS Server Configurations

Co-Location

The rack mount server form factor of the DSR may be co-located with one or more products. Co-location is the ability for the RMS DSR to be physically located in the same site footprint (cabinets) as other products.

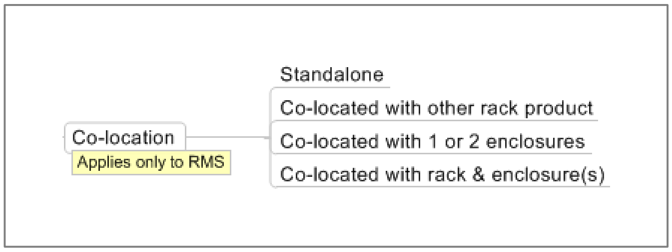

Figure 41: RMS Co-location

Co-location allows for different products at the same site to share certain resources:

» Physical cabinet and power

Some examples of co-located products with an RMS DSR are:

» RMS DSR signaling node + SDS rack mount
» RMS DSR signaling node + Policy rack mount

Spares

To maintain a high availability system, spare parts for key components are required.

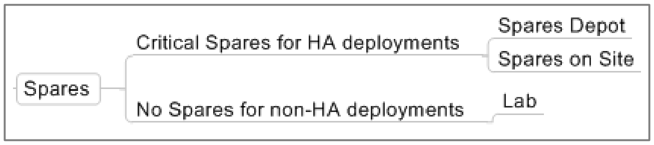

Figure 42 : RMS Spares

Subscriber Database Server (SDS) Node

There are several attributes that define a Subscriber Database Server (SDS) and these are defined in the diagram below. The typical configuration is shown in blue and common supported alternatives are white.

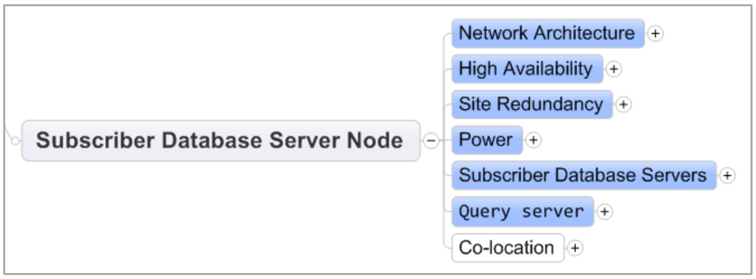

Figure 43 : Subscriber Database Server Node Attributes

There is only one SDS form factor supported for DSR, and that is HP Rack-mounted servers Gen8,Gen8 v2, or Gen9.

SDS uses the standard platform supported rack mount cabinets and allows for the use of cabinets specified by the customer provided those cabinets meet the minimum requirements as specified by Oracle cabinet requirements.

Network Architecture

The SDS is only deployed with DSR when a provisioning feed to a DSR subscriber database is required. In conjunction with the subscriber database processors (DPs) at the DSR signaling nodes, the SDS supports full subscriber address resolution capabilities.

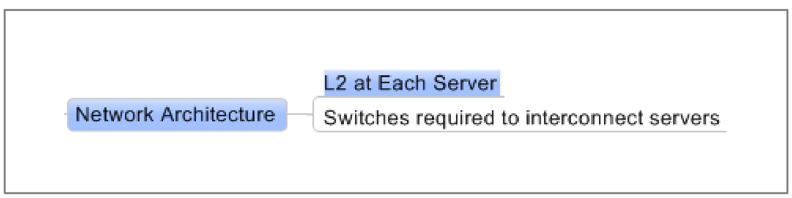

Figure 44 : SDS Network Architecture

» SDS supports layer 2 demarcation to the customer network at the server level.

High Availability

The SDS provides a provisioning feed to the subscriber database and is not considered to be a real-time critical component of a DSR network deployment. However, customers requiring a subscriber database are generally hosting large network solutions and still require that the SDS is highly available.

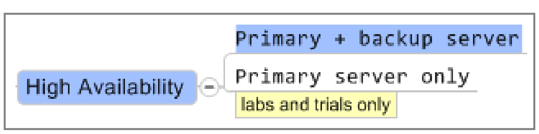

Figure 45 : SDS High Availability

» SDS is deployable in a highly available configuration where primary and backup servers are deployed as a pair.

» In configurations where a highly available SDS is not required (e.g. lab), then the SDS is deployed using a primary server only.

Site Redundancy

The SDS provides a provisioning feed to the subscriber database and is not considered to be a real-time critical component of a DSR network deployment. However, DSR optionally supports an SDS Disaster Recovery site for geographic redundancy

Figure 46 : SDS Site Redundancy

» The SDS can be deployed as a single site.
» The SDS can be deployed as a pair: the primary site, and a disaster recovery site.
» The disaster recovery site, when used, has a functionally duplicated configuration to the primary site.

Power

The SDS supports both AC and DC power options.

Subscriber Database Servers

DSR supports one or two rack mount servers per site. Two servers are supported for production sites and one server is supported for lab configurations.

Figure 47 : SDS Servers

Query Server

The query server allows non-real-time applications to query the provisioned SDS database.

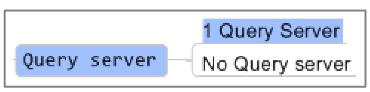

Figure 48 : SDS Query Server

» SDS allows for an optional query server.

Co-location

The SDS may be co-located with one or more products. Co-location is the ability for the SDS to be physically located in the same site footprint (cabinets) as other products.

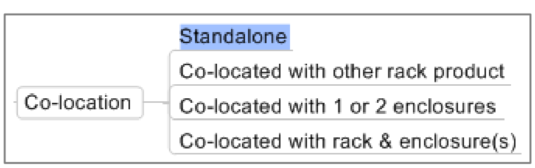

Figure 49 : SDS Co-Location

Co-location allows for different products at the same site to share certain resources such as the physical cabinet and power. Here are some examples of co-located products with an SDS:

» SDS rack mount + DSR in an enclosure
» SDS rack mount + policy in an enclosure
» SDS rack mount + NOAM rack mount

DSR Planning

Planning for DSR blades in the c-Class Form Factor

System OAM

The DSR nodal OAM is handled in DSR by a pair of active-standby blades referred to as System OAM. The System OAM manages the OAM for entire DSR node independent of how many enclosures the DSR spans or the physical location of those enclosures. The System OAM is sized to accommodate the OAM requirements for DSR and for co-resident OAM applications as described below. There is no expansion or scaling of the System OAM blades.

The System OAM blades support the following primary applications:

» System OAM which is the primary DSR OAM application supporting the System OAM, Diameter Agent, IP Front End, and the Session Binding Repository

» DP OAM provides the nodal management and distribution of the DPs and subscriber database. This is an optional application and is only present when DPs are used.

» Optionally co-resident Network OAM provides the network level OAM function. This application is only present when the NOAM function is co-resident at the DSR signaling node and is applicable for small systems with no more than two DSR signaling nodes.

System OAM applications have certain times when bursts of data are transferred to or from the DSR. For example: transferring measurements data off the System OAM, or transferring the subscriber database to the DP OAM from the SDS during initialization of the system. The applications are designed to handle this type of bursty transfer.

The System OAM, when it is also running the DP OAM, is potentially a RAM intensive function. The DP OAM is designed not to exceed the available memory; however RAM is the most likely resource constraint.

Measuring System OAM Utilization

In this section, one of the key metrics for managing the performance of the System OAM is shown. There are many more measurements available, and these can be found in [1] DSR Alarms, KPIs, and Measurements.

Table 3: Key Metric for System OAM Blade

Meas ID	Name	Group	Scope	Desc	Recommended Usage	
					Condition	**Action**
31056	RAM_UtilPct_Average	System	Blade	The average committed RAM usage as a percentage of the total physical RAM.	If the average Ram utilization exceeds 80% utilization	Contact Oracle

System OAM System Engineering

TABLE 4: DSR SYSTEM OAM SYSTEM CAPACITY

	System OAM
Number of System OAMs	2 = 1+1
System OAM supported applications	DSR SOAM, DP OAM, DSR NOAM, Spare SOAM
Form factor	HP G6, Gen8, Gen8 v2, Gen9

Diameter Agent MP

MPS Calculations for DA-MP

The Diameter Agent message processor (DA-MP) provides the primary Diameter processing and routing capabilities of the DSR. DA-MPs are a critical component of any DSR signaling node. DA-MP processor capabilities are primarily bound by the processing speed of the blade, read/write speeds on I/O ports of the blade and with increased message size as described below, RAM utilization. .

DSR supports Diameter message sizes up to 60K bytes. Normal message size is defined as messages up to 16K bytes and large messages are defined as messages between 16K and 60K bytes. Typically, less than 1% of the messages traversing the DSR are large messages. The percentage of large message sizes is likely to grow in some networks and varies from network to network. For these reasons, engineering the system to at least 5% mix of large messages provides support for growth of message sizes in the networks for the foreseeable future.

As the average message size grows in the network, so do the bandwidth requirements. While a given Diameter Agent supports an average 2K message size, the bandwidth available to the Diameter Agent may limit the actual average message size that can be supported in a given configuration.

» For example: a DA running at 50K MPS with average message size of 2K bytes would use about .75 Gb of bandwidth, but if the signaling network is engineered with 1Gb enclosure switches, then the maximum throughput available to a given DA is 1 Gb.

» A second example: 10Gb enclosure switches are used so the per DA bandwidth is not a concern, however the aggregate of the signaling traffic would be limited to 10 Gb which equates to a little less than 14 DAs at 50K MPS & 2K bytes message size. A second enclosure could be used and the DAs distributed between the enclosures, in which case each enclosure would be able to support 10 Gb.

DSR supports multiple HP blade hardware types for DA-MPs:

» G6 half-height

» G7 full-height (existing implementations only)

» Gen8 half-height

» Gen8 v2 half-height

» Gen9 half-height

Multiple DA-MP blades are supported in an active-active configuration up to a maximum of sixteen DA-MPs per DSR signaling node. DSR also supports existing active-standby configurations for up to two DA-MPs per DSR signaling node. The different DA-MPs have differing capacity limits, however, the methodology for dimensioning and determining how many DA-MPs are needed is exactly the same. The dimensioning of the DA-MPs needs to consider two primary factors:

» The set of capabilities and applications running (e.g. policy proxy, core routing, address resolution), and
» The desired processing throughput and its impact on critical resources (i.e. MPS, number of connections, IPSEC MPS).

To keep the management of the DA-MP performance simple, performance is divided into three groups:

» **Core** – this group consists of all routing, mediation, screening features and anything that doesn't fit into the next two groups.
» **Database** – this group consists of all database features including HSS/HLR address resolution features both range based and subscriber based address resolution (RBAR and FABR), and the inter-working function (IWF) required for MAP Diameter interworking.
» **Stateful** – this group consists of all subscriber/session stateful features including policy DRA, online charging DRA and offline charging proxy and includes Gateway Location Application (GLA).

If multiple applications are running on the same DA-MP, then the application with the lowest rating applies to the entire DA-MP.

In addition to the above, the use of IPSEC has a significant impact on the performance of the DA-MP and has its own capacity rating. The IPSEC de-rating only applies to the portion of the traffic running IPSEC.

Another factor that impacts performance is the support of application chaining. Application chaining refers to the case where a single DSR message encounters more than one application. This release supports application chaining for IWF with RBAR or FBAR, Policy DRA with RBAR or FABR and Online Charging with RBAR or FABR. The application chaining de-rating applies when more than 10% of the traffic on a given DA-MP is subject to application chaining.

Message size and number of connections also have an impact on the overall performance of the blade; however Oracle has factored these items into the capacity ratings of the blade. The following formulas can be used to determine the required number of DA-MPs:

Number of DA-MPs = greater of (MPS required/MPS per blade, connections required/connections per blade) + 1

Please refer to section 4) DA-MP System Engineering for the specific values for MPS per blade and connections per blade. If IPSEC is in use, then adjust the MPS per blade value by 25% to 40% de-rating for the % of traffic using IPSEC.

Examples:

» Customer A uses HSS address resolution and is running 20K MPS S6a traffic. In this case, a pair of ½ height DA-MPs easily supports the traffic.
» Customer B is running the policy proxy and needs to run 100K MPS policy proxy traffic using ½ height blades. Then 4+1= 5 blades are required.

Measuring DA-MP Utilization

In this section, only the key recommended metrics for planning expansions of the DA-MP are discussed. There are many more measurements available on the DA-MP, and these can be found in [1] DSR Alarms, KPIs, and Measurements.

Table 5: Key Metrics for DA-MP

Meas ID	Name	Group	Scope	Desc	Recommended Usage: Condition	Recommended Usage: Action
10202	RxMsgRateAvgMp	MP Performance	Server Group	Average Ingress Message Rate (messages per second) utilization on a MP server. The Ingress Message Rate is the number of ingress Diameter messages that are targeted for Relay Agent routing (non-zero Application ID).	When running in normal operation with a mate in normal operation, and this measurement exceeds 30% of the rated maximum, OR exceeds 60% of the rated capacity when running without an active mate.	The operator should determine if Additional growth in diameter traffic is continuing. If so, an estimate of the additional rate of MPS should be calculated and additional DA-MP blades should be planned for. This condition could also be an anomalous spike in traffic, and the operator may choose to ignore the occurrence. Above 40% in normal operation indicates an immediate need for additional DA-MPs.
10204	EvDiameterProcessAvg	MP Performance	Server Group	Average percent Diameter Process CPU utilization (0-100%) on a MP server.	When running in normal operation with a mate in normal operation, and this measurement exceeds 30% of the rated maximum capacity, OR exceeds 60% of the rated capacity when running without an active mate.	If additional growth in the system is anticipated, then consider adding an additional DA-MP. It's possible that the traffic mix is different than originally dimensioned (e.g. 40% IPSEC instead of the originally dimensioning 5%). In these cases, re-assess the dimensioning with the actual traffic/application mix and add additional DA-MPs blades as needed.
10205	TmMpCongestion	MP Performance	Server Group	Total time (in milliseconds) spent in local MP congestion state	Any number greater than 0.	After eliminating any configuration, or major failure conditions, then is a late indication that additional DA MPs are needed. It is highly desirable that

						planning for additional DA-MPs happens before this condition occurs.
10133	RxMsgSizeAvg	Diameter Performa nce	Server Group	The average ingress message size in Diameter payload octets.	Average message size > 2000 bytes	DA-MP dimensioning assumes 2K average message size. This information is used to dimension IPFEs and DIH/IDIH. No action required if there are no alarms associated with the PDU message pool (available memory for messages). If PDU message pool is exhausting, contact Oracle.
31056	RAM_UtilPct_Average	System (blade)	System	The average committed RAM usage as a percentage of the total physical RAM.	If the average Ram utilization exceeds 80% utilization	Contact Oracle
31052	CPU_UtilPct_Average	System (blade)	System	The average CPU usage from 0 to 100% (100% indicates that all cores are completely busy).	When running in normal operation with a mate in normal operation, and this measurements exceeds 30% of the rated maximum capacity, OR exceeds 60% of the rated capacity when running without an active mate.	If additional growth in the system is anticipated, then consider adding an additional DA MP. It's possible that the traffic mix is different than originally dimensioned (e.g. 40% IPSEC instead of the originally dimensioning 5%). In these cases, re-assess the dimensioning with the actual traffic and application mix and add additional DA-MPs blades as needed.

Measuring DA-MP Connection Utilization

In this section, only the key recommended metrics for planning expansions of the DA-MP connections are discussed. There are many more measurements available on the DA-MP connections, and these can be found in [1] DSR Alarms, KPIs, and Measurements.

Table 6: Key Metrics for DA-MP Connections

Meas ID	Name	Group	Scope	Desc	Recommended Usage	
					Condition	Action
10500	RxConnAvgMPS	Connection Performance	connection	Average Ingress Message Rate (messages per second) utilization on a connection.	Minor alarm is set by default at 50%, major at 80%. Ingress message rate per connection is customer configurable with a max per connection of 10,000	Configure additional connections

DA-MP System Engineering

Table 7: DSR DIAMETER Agent MP Capacity

		Performance Grouping		
		Core Routing	**Database** **IWF** **IWF with RBAR Chaining**	**Stateful Applications (incl GLA)** **IWF with FABR Chaining**
1/2 height Gen8, Gen8 v2, Gen9 blade	**MPS**	45,000	40,000	30,000
	Connections	2000	2000	2000
	IPSEC	40%[1] degradation	40%[1] degradation	40%[1] degradation
1/2 height G6 blade	**MPS**	40,000	35,000	30,000
	Connections	1000	1000	1000
	IPSEC	25%[1] degradation	25%[1] degradation	25%[1] degradation
Full height G7 blade[5]	**MPS**	50,000 (80K[2])	40,000 (70K[2])	n/a
	Connections	2000	2000	n/a
	IPSEC	25%[1] degradation	25%[1] degradation	n/a

1. Degradation applies only to the portion of traffic running IPSEC.

2. Capacities advertised are commercial capacities: 1+1 non-scaled DA-MP performs significantly better as shown in ().

3. Entire blade is rated at the capacity of the slowest application running on the blade. For example: a blade running core routing and policy proxy (a stateful application) is rated at 30K MPS on the Gen8 blade.

4. When application chaining is used and makes up more than 10% of the overall traffic, then a 15% derating of the blade is applied.

5. IWF traffic is not supported on the G7 blade.

Extensive use of diameter mediation has an impact on the diameter agent performance. The DA-MP performance capacities in the table above assume no more than five mediation templates.

Table 8: DSR DIAMETER Agent System Capacity (C-Class)

	Performance Grouping		
	Core Routing	Database	Stateful Applications
Max Number of DA-MPs	16	16	16
Total MPS	720,000	640,000	480,000 - policy and online charging DRA 240,000 – offline charging proxy
Total Connections	16,000	16,000	16,000
Mix G6, G7, Gen8& Gen9 DA-MPs	Yes	Yes	Yes
Peer Route Tables (PRTs)	100	100	100
Routing Rules per PRT	1,000	1,000	1,000
Routing Rules per System	10,000	10,000	10,000
Peer Route Lists	2,000	2,000	2,000
Peer Routing Groups	6,000	6,000	6,000
Range based routing rules	1,000,000	1,000,000	1,000,000

The DA-MP is the main Diameter processing engine of the DSR. The DSR's main capacity attribute is messages per second (MPS) and having a precise definition of exactly what constitutes an MPS is important.

MPS is a measure of the DSR Diameter message processing volume in messages per second. For this measure, a message is defined as:

» DSR processing of an ingress Diameter message and either transmitting a single outgoing Diameter message or discarding the ingress message. The outgoing message may be a variant of, or a response to, the ingress message.

» DSR transmission of any Diameter message, as required by DSR configuration, that is associated with incremental actions/events associated with #1 above. For example, the re-routing of a Request upon connection failure or the copying of a Request.

Messages excluded from this measure are:

» Diameter peer-to-peer messages: CER/CEA, DWR/DWA, and DPR/DPA

» Ingress Diameter messages discarded by the DSR due to Overload controls

» Answers received in response to Message Copy

The following table further clarifies and illustrates what is included in Messages per Second.

TABLE 9: MPS DEFINITION

Scenario	IMR	MPS	Comment
Ingress Request processing resulting in the Request being routed upstream (with or without local DSR application processing of the Request)	1	1	Request-in, Request-out
Ingress Answer processing resulting in forwarding of Answer downstream (with or without local DSR application processing of the Answer)	1	1	Answer-in, Answer-out
Ingress Request processing resulting in Answer message sent by DSR to originator (with or without local DSR application processing of the Request)	1	1	Request-in, Answer-out

Ingress Request discarded due to validation error or overload	1	1	Request-in, nothing-out
Ingress Answer discarded due to validation error	1	1	Answer-in, nothing-out
Initial copy and transmit of a Request to a DAS	1	1	Incremental action/event associated with receipt and processing of initial ingress Request message
Ingress Answer triggering reroute of the pending Request message	1	1	Answer-in, Request-out
Request reroute due to connection failure or Answer response timeout (including reroute of copied Requests to DAS for same reasons)	0	1	Incremental action/event associated with receipt and processing of initial ingress Request message
Ingress Answer from a DAS terminated by DSR due to Request copy completion or termination	1	0	Answer-in, nothing-out

Database Processor

Subscriber and Processing Calculations for DP

The database processor provides an in-RAM database for extremely fast and reliable access to subscriber information. DPs are an optional component of a DSR signaling node, and are needed when an application using a subscriber database is utilized. There are two primary resources being utilized on the DP blade:

» Memory – bounds the size of the in-memory database on each DP

» I/O & processor – bounds the number of database accesses that can occur

Using the baseline hardware, a single DP can hold up to 500M entities. Entities may be either text based strings (text based IMPU/IMPIs) or numeric strings (e.g. IMSI, MSISDN). Text based string entities are limited to a system maximum of 10 million. Each subscriber may have one or more entities associated with it. Differing applications and specific needs of a specific application implementation will dictate how many entities are needed for each subscriber. So, the maximum number of subscribers varies based on the number of entities required for each subscriber. For example:

» Customer A uses HSS address resolution with IMSI's in the form of numeric strings. In this case, 500 million subscribers can be supported.

» Customer B uses both IMSI's and MSISDN numeric strings for HSS address resolution. In this case, 250 million subscribers can be supported.

The following formula can be used to determine the required number of available subscribers:

Number of subscribers = 500 M Entities / number of entities per subscriber (with a max of 10M text based entities)

The full 500M entities can be supported by a single DP blade. An additional blade is used for redundancy. The DP blades are active load shared. Scaling beyond 500M entities is planned in a future release.

Each DP blade is projected to support up to 140K database accesses per second. Up to 10 DP blades can be deployed in a DSR node (9+1). The number of database accesses varies on the amount of traffic, type of traffic, and application implementations needing subscriber database access. For example:

» Customer B has a DSR node which is sized to process 200K MPS. The customer is running HSS address resolution on S6a traffic. Approximately 30% (60K MPS) of the traffic is S6a, of which 30K MPS is S6a Request messages and 30K MPS is S6a Answer Messages. Since only S6a Request messages need to access the database => 30K database access per second.

The following formulas can be used to determine the number of database accesses, and the number of DP blades needed:

Number of database accesses/s = MPS per node × %traffic on applicable interfaces × %msgs needing DB access

Number of DPs = upper bound (database accesses/s / 140K access/s) + 1

Measuring DP Utilization

In this section, only the key recommended metrics for managing the performance of the DP are discussed. There are many more measurements available on the DP, and these can be found in [1] DSR Alarms, KPIs, and Measurements.

There are two key components of the subscriber database within a DSR Signaling node: the database processors (DPs), and OAM component which runs on the System OAM blades. The key metrics for managing the DP blades are:

Table 10: DP Performance Metrics

Meas ID	Name	Group	Scope	Desc	Recommended Usage	
					Condition	**Action**
4170	DpQueriesReceived	DP	System (blade)	The total number of queries received per second.	When running in normal operation with a mate in normal operation, and this measurement exceeds 30% of the rated maximum capacity, OR exceeds 60% of the rated capacity when running without an active mate.	The operator should determine if additional growth in the number traffic requiring subscriber database look-ups is continuing to grow. If so, an estimate of the additional rate of database lookups should be calculated and additional DP blades should be planned for.
31056	RAM_UtilPct_Average	System	System (blade)	The average committed RAM usage as a percentage of the total physical RAM.	If the average Ram utilization exceeds 80% utilization	Contact Oracle
31052	CPU_UtilPct_Average	System	System (blade)	The average CPU usage from 0 to 100% (100% indicates that all cores are completely busy).	When running in normal operation with a mate in normal operation, and this measurements exceeds 30% of the rated maximum capacity, OR exceeds 60% of the rated capacity when running without an active mate.	Oracle considers this measurement of lesser importance to the DpQueriesReceived. However, this measurement in conjunction with DpQueriesReceived can be used to indicate the need to add additional DPs.

While memory is a consideration for the DPs, the SDS provides the centralized provisioning of the entire DSR network. It is recommended that the metrics from the SDS are used for managing the number of available subscribers instead of the metrics of individual DPs.

The OAM application related to the DPs runs as a virtualized application on the System OAM blades.

DP System Engineering

Table 11: DSR Database Processor Capacity

	Performance Grouping
	Database
Number of Entities	500,000,000
Number of Text Entities	10,000,000
Mix G6, G7, Gen8 & Gen9 DA-MPs	Yes
DB lookups per second	140,000

Table 12: DSR Database Processor System Capacity

	Performance Grouping
	Database
Max Number of DPs	10 (up to 9+1)
Max Number of Subscribers	500,000,000
DB Lookups per second	900,000

IP Front End

Bandwidth Calculations for IPFE

The IPFE is managed by bandwidth. The bandwidth of an IPFE pair is 3.2 Gbps. For the IPFE bandwidth and capacity planning, customers should consider the impact of IPFE blade failure resiliency and DSR node redundancy in their network. Examples:

» For an IPFE blade pair at a given DSR site, when the aggregate traffic across the IPFE blade pair exceeds 3.2 Gbps then a single IPFE blade failure at the site would cause the traffic on the remaining IPFE blade to exceed its rated capacity.

» If you ARE engineering for a 'local site IPFE blade failure AND a mate site DSR node failure' simultaneously, then, the aggregate traffic across the redundant IPFE blade pairs (i.e. all 4 IPFEs across the 2 sites) must not exceed 3.2 Gbps, since under such a scenario, all of the Client traffic would traverse the single remaining IPFE blade in this failure mode.

» If you are NOT engineering for a 'local site IPFE failure AND a mate site DSR node failure' simultaneously, then, the normal volume of traffic across each IPFE blade and the TSA redundancy scheme across the mated DSR pair dictates the capacity planning threshold.

» If you have a normal distribution of half the Client traffic to each IPFE blade pair, in the worst case all traffic ends up being carried by a single IPFE blade in the remaining DSR, the capacity planning threshold for IPFE bandwidth could be a percentage below 1.6 Gbps for the aggregate traffic across each IPFE blade pair.

In summary, customers should consider following network and deployment attributes for IPFE bandwidth and capacity planning:

» The traffic on a given IPFE blade
» The aggregate traffic across an IPFE blade pair at a given DSR site
» The aggregate traffic across redundant IPFE blade pairs in a mated DSR pair (i.e. all 4 IPFEs across the 2 sites)

To determine how much bandwidth is needed for an IPFE deployment, use the average Diameter message size (from measurement RxMsgSizeAvg or predicted message size) and the MPS to lookup the required bandwidth in the table below. The table indicates the resultant DSR MPS for various average Diameter Request message sizes assuming the following:

» All network traffic sent from clients to the DSR will traverse the IPFE, while all network traffic sent from the DSR to clients will bypass the IPFE.
» Under normal operating load, clients will perform standard aggregation of Diameter PDUs into TCP packets, up to the path MTU. The path MTU between clients and the IPFE is assumed to be at least 1500 bytes.

Table 13: IPFE Bandwidth and Resultant DSR MPS

IPFE Bandwidth	Average Diameter Message Size (bytes)	Resultant DSR MPS
0.5 Gbps	500	125,000
0.5 Gbps	1,000	62,500
0.5 Gbps	1,500	41,666
0.5 Gbps	2,000	31,250
1.0 Gbps	500	250,000
1.0 Gbps	1,000	125,000
1.0 Gbps	1,500	83,333
1.0 Gbps	2,000	62,500
1.5 Gbps	500	375,000
1.5 Gbps	1,000	187,500
1.5 Gbps	1,500	125,000
1.5 Gbps	2,000	93,750
2.0 Gbps	500	500,000
2.0 Gbps	1,000	250,000
2.0 Gbps	1,500	166,667
2.0 Gbps	2,000	125,000
2.5 Gbps	500	625,000
2.5 Gbps	1,000	312,500
2.5 Gbps	1,500	208,333
2.5 Gbps	2,000	156,250
3.2 Gbps	500	800,000
3.2 Gbps	1,000	400,000

3.2 Gbps	1,500	266,666
3.2 Gbps	2,000	200,000

Measuring IPFE Utilization

In this section, only the key recommended metrics for managing the performance of the IPFE are discussed. There are many more measurements available on the IPFE, and these can be found in [1] DSR Alarms, KPIs, and Measurements.

Table 14: IPFE Blade Performance Metrics

Meas ID	Name	Group	Scope	Desc	Recommended Usage	
					Condition	**Action**
5203	RxIpfeBytes	IPFE Performance	Server Group	Bytes received by the IPFE	If the number of (bytes * 8 bits/byte)/(time interval in s) is > 3.0 Gbps	If the traffic is expected to grow then, consider adding an additional IPFE pair
31052	CPU_UtilPct_Average	System	System (blade)	The average CPU usage from 0 to 100% (100% indicates that all cores are completely busy).	When running in normal operation with a mate in normal operation, and this measurements exceeds 30% of the rated maximum capacity, OR exceeds 60% of the rated capacity when running without an active mate.	Contact Oracle
31056	RAM_UtilPct_Average	System	System (blade)	The average committed RAM usage as a percentage of the total physical RAM.	If the average Ram utilization exceeds 80% utilization	Contact Oracle

IPFE SYSTEM Engineering

Table 15: DSR IPFE MP Capacity

Bandwidth Supported	3.2 Gbps
Connections	16000
Number of active DA-MPs load-balanced per IPFE pair	16
Supported Protocols	TCP, SCTP (uni-homed or multi-homed)
Form Factor	HP G6, Gen8, Gen8 v2, Gen9 half-height blade

Table 16: DSR IPFE MP System Capacity

Max Number of IPFE blades	4 = 2+ 2
Total Bandwidth	6.4 Gbps
Total Connections	16000

OFCS Session Binding Repository

Sessions and MPS Calculations for Charging SBR

The OFCS SBR is managed by the number of concurrent sessions and the amount of charging MPS.

The following formulas can be used to determine the number of SBRs per node based on the required number of concurrent sessions and the amount of charging MPS:

Number of SBRs = Greater of:

*1. (Required Concurrent Sessions / number of concurrent sessions per SBR pair) * 2, and*

*2. (Charging MPS / maximum charging MPS per SBR pair) * 2*

Measuring SBR Utilization

In this section, only the key metrics for managing the performance of the SBR are shown. There are many more measurements available on the SBR, and these can be found in [1] DSR Alarms, KPIs, and Measurements.

Table 17: SBR Blade Performance Metrics

Meas ID	Name	Group	Scope	Desc	Recommended Usage	
					Condition	**Action**
31052	CPU_UtilPct_Average	System	System (blade)	The average CPU usage from 0 to 100% (100% indicates that all cores are completely busy).	When this measurement exceeds 60% of the rated capacity.	If the traffic is expected to grow, consider adding an additional SBR pair
31056	RAM_UtilPct_Average	System	Blade	The average committed RAM usage as a percentage of the total physical RAM.	When the average Ram utilization exceeds 80% utilization	If the traffic is expected to grow, consider adding an additional SBR pair
12109	Sbr.RxReqRatePeak	SBR Performance	Blade	The maximum number of transactions/second processed by the SBR during the reporting interval.	When the number of transactions/second exceeds 80% of maximum or 37,000 TPS	If the traffic is expected to grow, consider adding an additional SBR pair

Charging SBR System Engineering

Table 18: OFCS SBR MP Capacity

	Performance Grouping Stateful Applications OFCS Charging Proxy
Concurrent Sessions per SBR pair	25M
MPS (of CPA traffic) per SBR pair	60,000

Table 19: OFCS SBR System Capacity

	Performance Grouping Stateful Applications OFCS Charging Proxy
Number of SBR MPs	4+4
Total Concurrent Sessions	100M
MPS (of CPA traffic)	240,000

Session Binding Repository

Sessions, Subscribers and MPS Calculations for SBR

The SBRs are managed by the number of concurrent sessions, the number of active subscribers (Policy DRA only) and the amount of Policy and/or Online Charging DRA MPS. Other factors in determining the SBR architecture include the DSR deployment architecture (small, large, lab, trial, number of nodes etc.) and the SBR server group and redundancy model desired.

The following information provides guidance on the decision making to determine which redundancy model is appropriate for a particular configuration.

Table 20: Redundancy Model Decision Making

Redundancy	Highly Available	Geo-redundant	Min No of Rqd Nodes	Pros	Cons	Typically Used…
Active Only	No	No	1	Small footprint/power. Lease Cost Option. Functionality with no replication.	Low availability/reliability	Not allowed in production. Labs/trials only. Used for demonstrating behavior when HA/geo/capacity testing is not needed.

Redundancy	Highly Available	Geo-redundant	Min No of Rqd Nodes	Pros	Cons	Typically Used…
Active-Standby	Yes	No	1	Small footprint/power. Lower Cost Option. Single site high availability.	Not geo-redundant. Moderate availability	Labs/trials where HA testing is needed. Singleton production DSRs where lower availability is acceptable.
Active-Standby-Spare	Yes	Yes	2	Geo-redundant, highly available. Fast switchover to standby on loss of active.	Higher cost, footprint, power. Asymmetrical deployment model for mated pairs (i.e. one node can end up with more servers than the other).	Typical production model when HA & geo are needed.
Active-Spare	Yes	Yes	2	Geo-redundant, highly available. Symmetrical deployment model for mated pairs. Most cost effective for HA/Geo.	Slow switchover to spare on loss of active. Moderate availability due to longer switchover time.	Acceptable production model when HA & geo are needed. Consider for lower cost solutions where moderate availability is acceptable
Active-Standby-Spare-Spare	Yes	Yes	3	Geo-redundant, highly available. Fast switchover to standby on loss of active.	Highest cost, footprint, power. Asymmetrical deployment model for mated triples (i.e. one node can end up with more servers than the others).	For use with mated triples. This is not a common redundancy model, and should only be used with prior approval

To deploy each of the redundancy models, varying numbers of servers are needed:

Table 21: Required Server by Redundancy Model

Redundancy	Number of Servers per Server Group			Redundancy Model
	Servers in Node A	Servers in Node B	Total Servers in Group	
Active Only	1	0	1	non-HA, non-Geo
Active-Standby	2	0	2	HA only, non-Geo
Active-Standby-Spare	2	1	3	HA, Geo
Active -Spare	1	1	2	HA, Geo
Active - Standby -Spare - Spare	See note	See note	4	HA, Geo

Note: This redundancy model is for use with mated triples and should only be used with prior approval.

Here are two examples:

1. One server group is needed to support a small lab using Active Only. In this case, 1 server group = 1 server, so only a single SBR is needed.
2. Four server groups are needed to support a network consisting of a DSR mated pair using Active-Standby-Spare. In this case 4X3=12 servers are needed. In the table above, the Node A and Node B designations are arbitrary. That is, for any server group, the node A and node B designations can be reversed. This allows for relatively uniform distribution of the servers among the DSRs. In this example, there are four Active-Standby servers or 4X2= 8 servers in node A, plus four standby servers 4X1=4 servers in node B. This is a valid deployment scenario, but to balance the server deployments, the first node would be assigned the node A role for two of the server groups and then would be assigned the node B role for the remaining 2 servers groups. The mate would have the node roles reversed. The result would be that each node would have 6 SBR servers deployed as shown below.

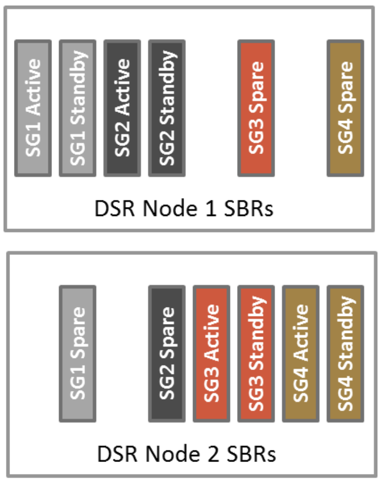

Figure 50: SBR Distribution Example (SG=Server Group)

The subscriber binding and session databases have dedicated server groups. The system scales by adding server groups, however in the short term server groups cannot be dynamically added to the system, so the network should be engineered to configure enough server groups for at least 18 months of growth.

Examples:

1. Very small DSR network with one mated pair of DSR nodes and Active-Spare server group. The number of SBRs is dependent upon the redundancy model selected. In this case, there will be two SBRs that will provide subscriber binding and two SBRs that will provide session binding.
2. Small DSR network with one mated pair of DSR nodes, Active-Standby-Spare server groups with dedicated session and subscriber binding databases. The number of SBRs required can be determined using the formulas below.

Number of session SBR(s) servers = Greater of:

*1. (Required Concurrent Sessions / maximum number of concurrent sessions per SBR server group) * 3 (for Active-Standby-Spare redundancy model)*

*2. (P-DRA MPS / maximum P-DRA MPS per SBR server group) * 3 (for Active-Standby-Spare redundancy model)*

Number of binding SBR(b) servers = Greater of:

*1. (Active subscribers / maximum number of subscribers per combined SBR server group) * 3 (for Active-Standby-Spare redundancy model)*

*2. (P-DRA MPS / maximum P-DRA MPS per SBR server group) * 3 (for Active-Standby-Spare redundancy model)*

If the required number of concurrent sessions is 60M, active subscriber entities is 40M and the amount of P-DRA MPS is 60,000 then the resultant number of SBRs required for the Active-Standby-Spare redundancy model is:

» Session SBR(s) = greater of (60M concurrent sessions / 60M) *3 and (60k MPS / 200k MPS) *3 = 3 session SBR(s) servers

» Binding SBR(b) = greater of (40M entities / 50 M) *3 and (60K MPS / 250k MPS) *3 = 3 binding SBR(b) servers

Measuring SBR Utilization

In this section, only the key metrics for managing the performance of the SBR are shown. There are many more measurements available on the SBR, and these can be found in [1] DSR Alarms, KPIs, and Measurements.

Table 22: Key Metrics for Session SBR Blades

Meas ID	Name	Group	Scope	Desc	Recommended Usage	
					Condition	**Action**
31052	CPU_UtilPct_Average	System	System (blade)	The average CPU usage from 0 to 100% (100% indicates that all cores are completely busy).	When this measurement exceeds 60% of the rated capacity.	Contact Oracle
31056	RAM_UtilPct_Average	System	Blade	The average committed RAM usage as a percentage of the total physical RAM.	If the average Ram utilization exceeds 80% utilization	Contact Oracle
11372	SbrPolicySessionRecsAvg	SBR Session Performance	Server Group	The number of policy sessions in progress	If P-DRA function is enabled and OC-DRA is not enabled and average exceeds 48M. If both P-DRA and OC-DRA are enabled this average must be combined with the SbrOcSessionRecsAvg and the combined average exceeds 48M	Contact Oracle
11441	SbrOcSessionRecsAvg	SBR Session Performance	Server Group	The number of online Charging sessions in progress	If OC-DRA function is enabled and P-DRA is not enabled and average exceeds 48M.	Contact Oracle

					If both P-DRA and OC-DRA are enabled this average must be combined with the SbrPolicySessionRecsAvg and the combined average exceeds 48M	

Key metrics for managing the Binding SBR(b) blades are:

Table 23: Key Metrics for Managing the Binding SBR(b) Blades

Meas ID	Name	Group	Scope	Desc	Recommended Usage	
					Condition	**Action**
31052	CPU_UtilPct_Average	System	System (blade)	The average CPU usage from 0 to 100% (100% indicates that all cores are completely busy).	When this measurement exceeds 60% of the rated capacity.	Contact Oracle
31056	RAM_UtilPct_Average	System	Blade	The average committed RAM usage as a percentage of the total physical RAM.	If the average Ram utilization exceeds 80% utilization	Contact Oracle
11374	SbrPolicyBindingRecsAvg	SBR Binding Performance	Server Group	Average number of active SBR Policy bindings.	When this average exceeds 40M.	Contact Oracle

SBR System Engineering

Table 24: Subscriber SBR(b) Server Group Capacity

	Performance Grouping Stateful Applications	
	Policy DRA	**Online Charging DRA**
Subscriber Binds	50M	N/A
MPS (of Policy DRA traffic)	250,000*	N/A

** assumes up to 17% CCR-I + CCR-T + CCA-I + CCA-T traffic mix with up to 2 alternate keys, assumes average session size of 100 bytes. If GLA application is running, each GLA application MPS counts as 3 MPS for the purposes of dimensioning the SBRs.*

Table 25: Session SBR(b) Server Group Capacity

	Performance Grouping **Stateful Applications** **Policy DRA and/or Online Charging DRA**
Concurrent Sessions	60M
MPS (of Policy DRA and/or Online Charging DRA traffic)	200,000*
Topology hiding (Policy DRA only)	125K with 100% Rx TH** 135K with 100% Gx TH, 100K with 100% Rx & Gx TH

** assumes up to 17% CCR-I + CCR-T + CCA-I + CCA-T traffic mix with up to 2 alternate keys, assumes average session size of 100 bytes. If GLA application is running, each GLA application MPS counts as 3 MPS for the purposes of dimensioning the pSBRs.*

*** assumes up to 32% Rx traffic and up to 25% Gx traffic*

Table 26: SBR Nodal / Mated Pair Capacity

	Performance Grouping **Stateful Applications** **Policy DRA and/or Online Charging DRA**
Max Session Server groups per node/mated pair	8*
Max Concurrent Sessions per node/mated pair	240M
Max DRA MPS per node/mated pair	450K
Max Subscriber Binding Server groups per node/mated pair (Policy DRA only)	8*

** Actual blade count depends on the selected redundancy model*

Table 27: SBR Network Capacity

	Performance Grouping Stateful Applications	
	Policy DRA	Online Charging DRA
Max P-DRA nodes per network	32 (16 pairs)	N/A
Max Subscriber Binding Server groups per network	8	N/A
Max subscriber binds per network	200M	N/A
Max concurrent session binds per network	480M	

SS7 Message Processor

The SS7 message processor supports MAP Diameter Interworking Function (IWF) in conjunction with the DA-MPs. The performance ratings for the SS7 MPs are based specifically on MAP messages with assumptions of a 170 byte average message size with a 4K byte maximum message size and 20 second average transaction hold time.

The SS7 MP requires HP Gen8, Gen8 v2, or Gen9 blades but the associated DA-MPs can be G6, Gen8, Gen8 v2, or Gen9. HP G7 is not supported for DA-MPs providing the Interworking Function.

Measuring SS7 MP Utilization

In this section, only the key recommended metrics for planning expansions of the SS7 MP are discussed. There are many more measurements available on the SS7 MP, and these can be found in [1] DSR Alarms, KPIs, and Measurements.

Table 28: Key Metrics for SS7 MP and Associated IWF DA-MP

Meas ID	Name	Group	Scope	Desc	Recommended Usage	
					Condition	Action
11054	MAP Ingress Message Rate	MAP Diameter Interworking function (MD-IWF)		Average number of MAP messages (both requests and responses) received per second from SS7 network	When running in normal operation with a mate in normal operation, and this measurement exceeds 30% of rated maximum, OR exceeds 60% of the rated capacity when running without an active mate.	If additional growth in MAP traffic is expected, an estimate of the traffic should be calculated and additional SS7 MP blades should be planned for. This condition could also be an anomalous spike in traffic, and the operator may choose to ignore the occurrence.

						Above 40% in normal operation indicates an immediate need for additional SS7 MPs
31056	RAM_UtilPct_Average	System (blade)	System	The average committed RAM usage as a percentage of the total physical RAM.	If the average Ram utilization exceeds 80% utilization	Contact Oracle
31052	CPU_UtilPct_Average	System (blade)	System	The average CPU usage from 0 to 100% (100% indicates that all cores are completely busy).	When running in normal operation with a mate in normal operation, and this measurements exceeds 30% of the rated maximum capacity, OR exceeds 60% of the rated capacity when running without an active mate.	If additional growth in the system is anticipated, then consider adding an additional SS7 MP.

SS7 MP System Engineering

Table 29: SS7 MP Capacity

		Performance
		MAP Diameter IWF
1/2 height Gen8, Gen8 v2, Gen9 blade	**MAP MPS**	30,000

Table 30: SS7 MP System Capacity

	Performance
	MAP Diameter IWF
Max Number of SS7 MPs	8
Maximum Nodal MAP MPS	210,000

Planning for DIH and IDIH

Integrated Diameter Intelligence Hub (IDIH)

The IDIH is an optional component of the DSR and provides trouble shooting capabilities for the DSR. Only one IDIH is supported per DSR node.

Table 31: Integrated Diameter Intelligence Hub Capacity

	Performance Rating
IDIHs supported	up to 1 per DSR signaling node
Supported Capacity	1K TTRs/s with 2.5KB average TTR size, 700 TTRs/s with 5.0KB average TTR size
Supported Interfaces	Diameter - Sh, Cx, Gq, Cc, S6a, S6d, Gx, Rx, Gy, SLg, SLh, Gxa, SWm, SWx, STa, S6b, S9, Sd, Sy, S13
Form Factor	HP Gen8, Gen8 v2, Gen9 half-height blade or HP Gen8, Gen8 v2, Gen9 rack-mount server

Diameter Intelligence Hub (DIH)

The DIH is an optional component of the DSR and is supported for existing deployments only. DIH provides trouble shooting capabilities for acquiring up to 20K MPS of port mirrored Diameter Agent signaling traffic and analyzing and creating xDRs for up to 10K MPS assuming average message size of 500 bytes. The DIH is supported on HP G6,Gen8 and Gen9 blades or on HP Gen8 or Gen9 rack mount servers.

Planning for Network OAM

DSR supports three tiered OAM topology which includes the Network OAM network element. Two tiered OAM topology is no longer supported in this release. Each DSR network requires one Network OAM and optionally a disaster recovery Network OAM.

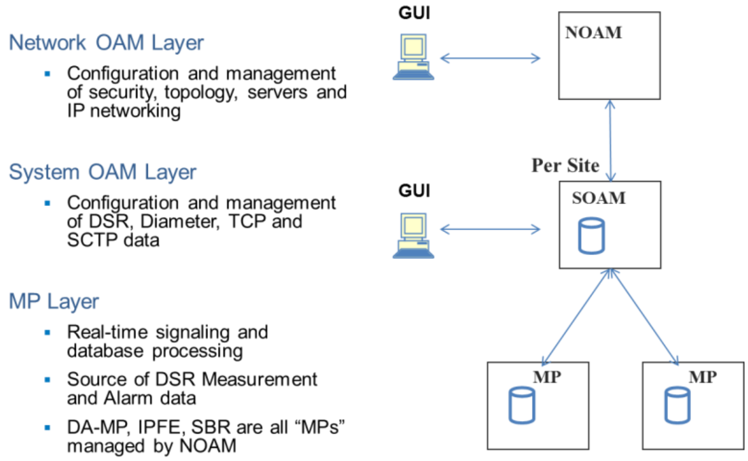

Figure 51 : Three Tier OAM Topology

Deployment Alternatives

The Network OAM consists of a pair of active-standby servers/blades. The NOAM can be deployed on dedicated rack-mount servers either at a stand-alone site or co-located at one of the DSR signaling sites. If co-located at one of the DSR signaling sites, the active or standby NOAM can be co-resident on the management server located at the DSR signaling site. The NOAM can also be deployed on the System OAM blades at one of the DSR signaling sites. This option is supported for small systems with no more than two signaling nodes.

Network OAM deployment alternatives:

» Active-standby rack mount servers at a stand-alone site (management server function co-resident on one of the servers)

» Active-standby rack mount servers co-located at a DSR signaling node (management server function co-resident on one of the servers)

» NOAM application co-resident on the active-standby System OAM blades at a DSR signaling node

The NOAM is appropriately sized to handle the DSR network for the foreseeable future. When deployed on rack mount servers, either at a stand-alone site or co-located at a DSR signaling node, the Network OAM can manage up to 32 DSR signaling nodes. Each signaling node in the DSR network can support up to 48 servers. However, the NOAM NE supports a maximum of 768 subtended servers per network. Only some types of servers are included in the NOAM managed server count:

» SOAM, DA MP, IPFE, SBR and SS7 MP servers count towards the NOAM managed server total in a network. If any of these functions is deployed as a virtual machine (VM), each VM counts as a server.

» DP, DIH and iDIH do not count towards to the total number of NOAM managed servers supported in a network.

Examples:

1. Node 1 has 2 SOAMs, 2 IPFEs, 4 DA MPs and 2 DPs. The server count for purposes of NOAM managed servers is 8 (DPs are not included)

2. Node 2 has 2 SOAMs, 2 IPFEs, 16 DS MPs and 12 SBRs. The server count for purposes of NOAM managed servers is 32.

To calculate the number of NOAM subtended servers in a network, take the sum of the server counts in each node.

When deployed co-resident on the System OAM blades, the Network OAM can manage up to two DSR signaling nodes.

The DSR Network OAM is potentially a RAM intensive function. The Network OAM is designed not to exceed the available memory; however RAM is the most likely resource constraint.

Measuring Network OAM Utilization

In this section, only the key recommended metrics for managing the performance of the Network OAM are discussed. There are many more measurements available, and these can be found in [1] DSR Alarms, KPIs, and Measurements.

Table 32: Key Metrics for Managing the Network OAM Servers

Meas ID	Name	Group	Scope	Desc	Recommended Usage	
					Condition	**Action**
31056	RAM_UtilPct_Average	System	System (server)	The average committed RAM usage as a percentage of the total physical RAM.	If the average Ram utilization exceeds 80% utilization	Contact Oracle

Network OAM System Engineering

Table 33: DSR Network OAM Capacity

	Capacity
Number of NOAMs	primary/backup, plus optional disaster recovery site
Number of DSR signaling nodes supported	32 2 (when co-resident on System OAM blades)
Form factor	G6, Gen8, Gen8 v2, Gen9 blades or Gen8, Gen8 v2, Gen9 rack mount servers

Planning for SDS

The SDS consists of a primary and optional backup Subscriber Database Server and an optional Query Server. The SDS is only used in conjunction with DSR Signaling sites when full address resolution features are needed.

Subscriber Database Servers

Considerations for the SDS

The SDS is appropriately sized to handle the database and database expansion for the foreseeable future. There are two scaling factors that should be considered for the SDS:

- The provisioning rate. This is the rate at which new subscribers and associated information can be written to the SDS database. The current provisioning rate is set at 200 write/updates per second.
- Number of DSR Signaling nodes synchronized. The SDS can synchronize up to 32 DSR signaling nodes.

Measuring SDS Utilization

In this section, only the key recommended metrics for managing the performance of the SDSs are discussed. There are many more measurements available on the SDS, and these can be found in [1] DSR Alarms, KPIs, and Measurements.

Table 34: Key Metrics for SDS Servers

Meas ID	Name	Group	Scope	Desc	Recommended Usage	
					Condition	**Action**
4110	ProvMsgsReceived	Prov	Server Group	The total number of provisioning messages that have been received.	If the rate exceeds 150 message/s	Contact Oracle
31052	CPU_UtilPct_Average	System	System (server)	The average CPU usage from 0 to 100% (100% indicates that all cores are completely busy).	When this measurements exceeds 65% CPU utilization	Contact Oracle
31056	RAM_UtilPct_Average	System	System (server)	The average committed RAM usage as a percentage of the total physical RAM.	If the average Ram utilization exceeds 80% utilization	Contact Oracle

SDS SYSTEM Engineering

Table 35: DSR Subscriber Database Server (SDS) System Capacity

	Database
Number of servers	1+1
Number of query servers (optional)	1
Max of subscribers	500,000,000*
Number of Entities	500,000,000
Number of Text Entities	10,000,000
Number of writes/updates per second	200

1 subscriber = 1 entity

Table 36: DSR Subscriber Database Server (SDS) Network Capacity

	Database
Number of SDSs (active + optional disaster recovery SDS)	2 = 1+1
Number of DSR signaling nodes synchronized	32

Query Server

Processing Calculations for Query Server

The Query Server has been designed to be a best effort server for free-form variable length and complexity queries. DSR supports 1 query server, however the system has been designed to add additional query servers if needed.

A new query server may be needed if the latency of the average query-response grows to a point that is not acceptable by the querying application.

Measuring Query Server Utilization

In this section, only the key recommended metrics for managing the performance of the Query Servers are discussed. There are many more measurements available on the Query Server, and these can be found in [1] DSR Alarms, KPIs, and Measurements.

Table 37: Key Metrics for the SDS Query Server

Meas ID	Name	Group	Scope	Desc	Recommended Usage	
					Condition	Action
31052	CPU_UtilPct_Average	System	System (server)	The average CPU usage from 0 to 100% (100% indicates that all cores are	When this measurements exceeds 65% CPU utilization	Contact Oracle

				completely busy).		
31056	RAM_UtilPct_Average	System	System (server)	The average committed RAM usage as a percentage of the total physical RAM.	If the average Ram utilization exceeds 80% utilization	Contact Oracle

Planning for DSR Servers in the Rack Mount Server Form Factor

A rack mount server based DSR signaling node consists of a number of virtual machines (VM) which provide specific capabilities and the servers they are hosted on. The RMS signaling node consists of core servers, expansion servers, IDIH servers and DIH servers (existing deployments only). The following table describes the VM to server mapping.

Table 38: VM to Server Mappings

Virtual Machine	Hosting Server	Optional/ Mandatory	VMs Per Server	Max VMs Per Node	When Used?
NOAM	Core Server 1	Optional	0-1	0 or 2	Min 2 per network. 0 if standalone NOAM node provided (see above for standalone NOAM planning)
	Core Server 2	Optional	0-1		
SOAM	Core Server 1	Mandatory	1	2	Mandatory 2 per node
	Core Server 2	Mandatory	1		
Management Server (PM&C)	Core Server 1	Mandatory	1	1-2	Mandatory 1 per node, optional 2nd PM&C
	Core Server 2	Optional	0-1		
DA MPs	Core Server 1	Mandatory	1	2 to 8	Core servers each get 1 DA MP, optional expansion servers allow for capacity expansion.
	Core Server 2	Mandatory	1		
	Expansion Server 1	Optional	0-3		
	Expansion Server 2	Optional	0-3		
	Expansion Server 3	Optional	0-3		
	Expansion Server 4	Optional	0-3		
IPFE	Core Server 1	Mandatory	1	2	DSR RMS will always contain IPFE. Specific deployments may choose to not use IPFE.
	Core Server 2	Mandatory	1		
SS7 MPs	Expansion Server 1	Optional	0-2	0-4	SS7 MPs for MAP Diameter IWF are optionally deployed.
	Expansion Server 2	Optional	0-2		
	Expansion Server 3	Optional	0-2		
	Expansion Server 4	Optional	0-2		
DIH VM group	DIH Server	Optional	0-1	0-1	Optional, pre-existing deployment
IDIH VM group	IDIH Server	Optional	0-1	0-1	Optional, new deployments

Core Rack Mount Server

The core rack mount servers host the NOAM, SOAM, DA MP, Management Server and IPFE virtual machines and are deployed in a pair for production nodes. Core servers can be deployed on HP Gen8,Gen8 v2 or Gen9.rack mount servers.

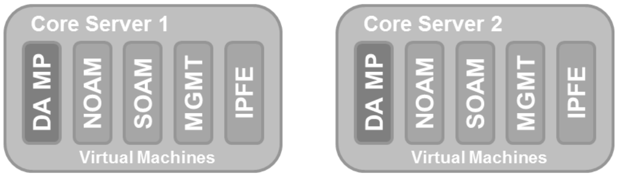

Figure 52 : Core Servers

RMS core servers host multiple VMs. Not only are metrics at the server level required but also key metrics for the DA-MP should be managed. RMS core servers are sized to meet the supported node capacities.

In this section, only the key recommended metrics for managing the performance of the RMS core servers are discussed. There are many more measurements available on the RMS core servers and the hosted VMs, and these can be found in [1] DSR Alarms, KPIs, and Measurements. The key metrics for managing the RMS Core Servers are:

Table 39: Key Metrics for RMS Core Servers

Meas ID	Name	Group	Scope	Desc	Recommended Usage	
					Condition	**Action**
10202	RxMsgRateAvgMp	MP Performance	Server Group (VM)	Average Ingress Message Rate (messages per second) utilization on a MP server. The Ingress Message Rate is the number of ingress Diameter messages that are targeted for Relay Agent routing (non-zero Application ID).	When running in normal operation with a mate in normal operation, and this measurement exceeds 30% of the rated maximum, OR exceeds 60% of the rated capacity when running without an active mate.	If additional Diameter MPS growth is expected, an estimate of the expected traffic should be calculated and additional DA-MP blades should be planned for. This condition could also be an anomalous spike in traffic, and the operator may choose to ignore the occurrence. Above 40% in normal operation indicates an immediate need for additional DA-MPs.
10133	RxMsgSizeAvg	Diameter Performance	Server Group	The average ingress message size in Diameter payload octets.	Average message size > 2000 bytes	DA-MP dimensioning assumes 2K average message size. This information is used to dimension the IPFEs and DIH/IDIH. No action required provided there are no alarms associated with the PDU message pool (available memory

						for messages). If PDU message pool is exhausting, contact Oracle.

Expansion Rack Mount Server

RMS expansion servers host up to three VMs which support SS7 MPs and expansion capability for DA MPs.

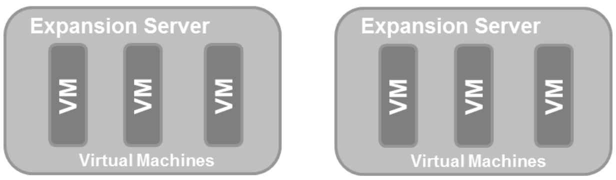

Figure 53 : Expansion Servers

RMS expansion servers host multiple VMs. Not only are metrics at the server level required but also key metrics for the VMs should be managed. SS7 MP VM metrics are included in the table below. DA MP metrics should also be consulted if DA MP VMs are hosted on the expansion server. Refer to the core server section above for DA MP metrics.

In this section, only the key recommended metrics for managing the performance of the RMS expansion servers are discussed. There are many more measurements available on the RMS core servers and the hosted VMs, and these can be found in [1] DSR Alarms, KPIs, and Measurements.

Table 40: Key Metrics for RMS Core Servers

Meas ID	Name	Group	Scope	Desc	Recommended Usage	
					Condition	**Action**
11054	MAP Ingress Message Rate	MAP Diameter Interworking function (MD-IWF)		Average number of MAP messages (both requests and responses) received per second from SS7 network	When running in normal operation with a mate in normal operation, and this measurement exceeds 30% of the rated maximum, OR exceeds 60% of the rated capacity when running without an active mate.	The operator should determine if Additional growth in MAP traffic is continuing. If so, an estimate of the additional rate of MPS should be calculated and additional SS7 MP blades should be planned for. This condition could also be an anomalous spike in traffic, and the operator may choose to ignore the occurrence. Above 40% in normal operation indicates an immediate need for additional SS7 MPs

31052	CPU_UtilPct_ Average	System	System (server)	The average CPU usage from 0 to 100% (100% indicates that all cores are completely busy).	When this measurements exceeds 65% CPU utilization	Contact Oracle
31056	RAM_UtilPct_ Average	System	System (server)	The average committed RAM usage as a percentage of the total physical RAM.	If the average Ram utilization exceeds 80% utilization	Contact Oracle

Rack Mount Server System Engineering

Scaling of the DSR in an RMS configuration is achieved by adding additional VMs running as DA MPs. Each DA MP supports up to 10,000 MPS and the maximum supported per node is 50,000 MPS. In a high availability configuration, the VM redundancy model is described as N+K (or 5+3 for the maximum DA MP configuration). An example RMS configuration is shown below. If any one server fails, there are at least five DA MP VMs to support the maximum of 50,000 MPS. Additional expansion servers provide the SS7 MP VMs in this example. However, SS7 MP VMs can reside on the same servers with DA MP VMs. Each expansion server supports a maximum of three VMs. The maximum number of SS7 MP VMs per server is two.

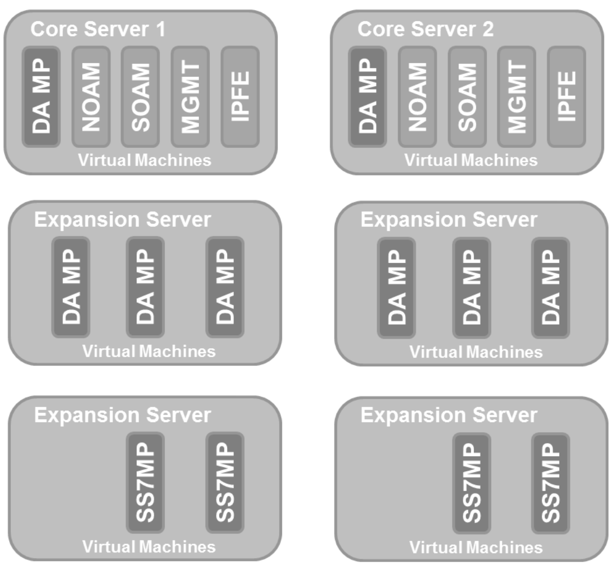

Figure 54 : Example RMS Configuration

Table 41: DA-MP VM Capacity

		Capacity
Gen8, Gen8 v2, Gen9 Server	**MPS**	10,000
	Connections	1000
	IPSEC	40%[1] degradation

[1]Degradation applies only to the portion of traffic running IPSEC.

Table 42: DA-MP RMS System Capacity

	System Capacity
Max Number of DA MPs	8
Maximum Nodal MPS	50,000
Total Connections	8,000
Peer Route Tables (PRTs)	100
Peer Routing Rules per PRT	1,000
Peer Routing Rules per system	10,000
Peer Route Lists	2,000
Peer Routing Groups	6,000
Range based routing rules	1,000,000

Table 43: SS7 MP VM Capacity

		Capacity MAP Diameter IWF
SS7 MP VM on RMS	**MAP MPS**	10,000

Table 44: SS7 MP RMS System Capacity

	Capacity MAP Diameter IWF
Max Number of SS7 MPs	4
Maximum Nodal MAP MPS	40,000

DIH and IDIH Servers

At RMS signaling nodes, DIH and IDIH are supported on rack mount servers only. Refer to the planning section above for DIH and IDIH capacities.

Baseline Reference/Assumptions/Limitations

This document is based on Platform 6.7 and the Platform Feature Guide – Available upon requestshould be referenced for further details on the platform.

Signaling Node Cabinet Layouts

Sample DSR Layout (DC)
Cabinet 1 Front View

Sample DSR Layout (DC)
Cabinet 1 Rear View

Power Requirements (-48VDC)	
Recommended Power Feed	8 x 60A
Maximum Power (W)	4,110
Maximum BTU/hr	14,024
Average Power (W)	3,576
Average BTU/hr	12,202

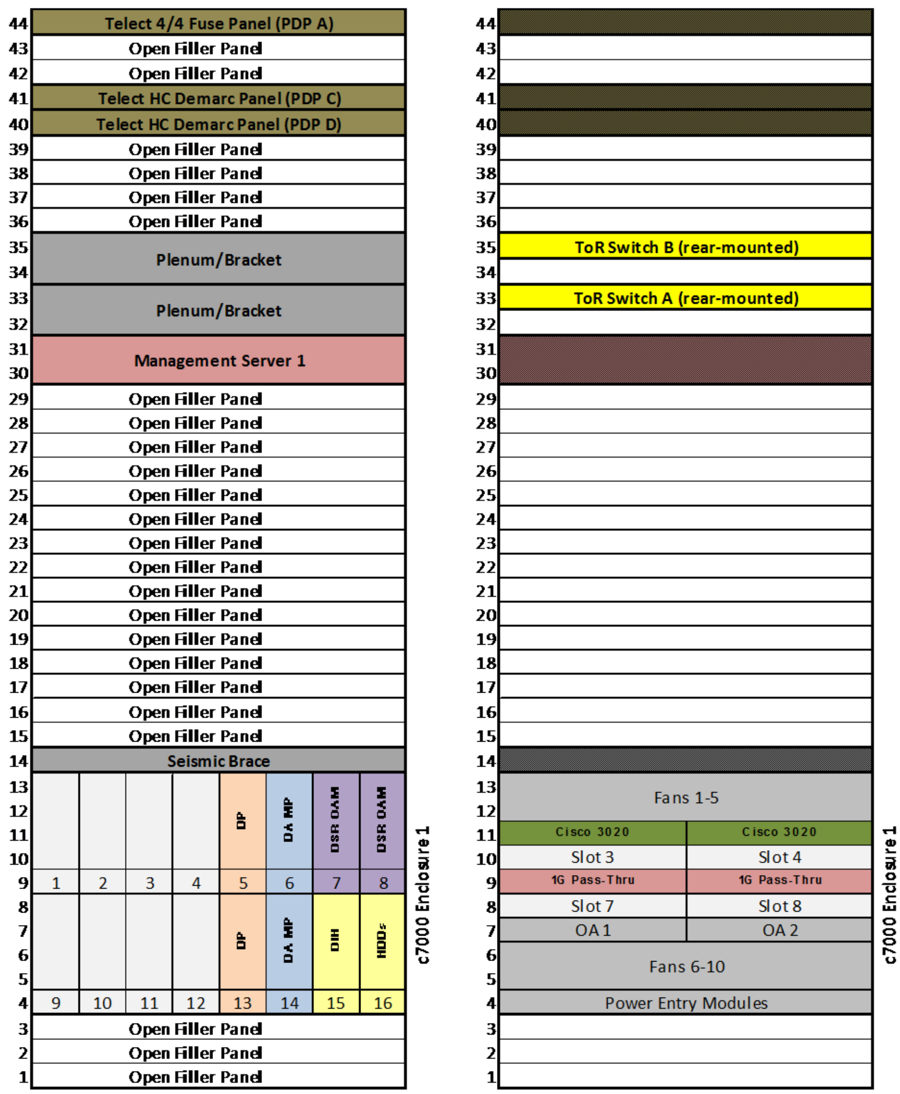

Figure 55 : DC c-Class Example Configuration

Sample DSR Layout (AC)
Cabinet 2 Front View

Sample DSR Layout (AC)
Cabinet 2 Rear View

Power Requirements (208VAC)	
Recommended Power Feed	2 x 50A
Maximum Power (W)	4,110
Maximum BTU/hr	14,024
Average Power (W)	3,576
Average BTU/hr	12,202

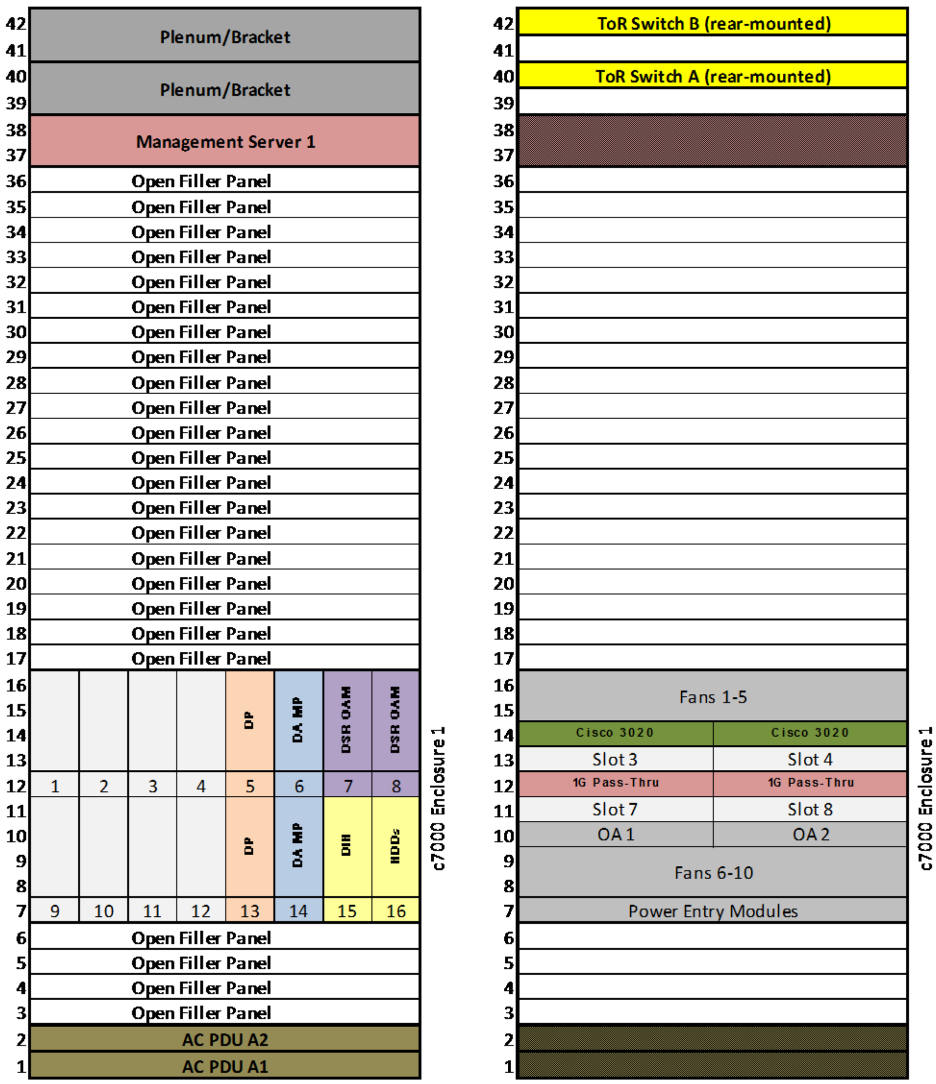

Figure 56 : AC c-Class Example Configuration

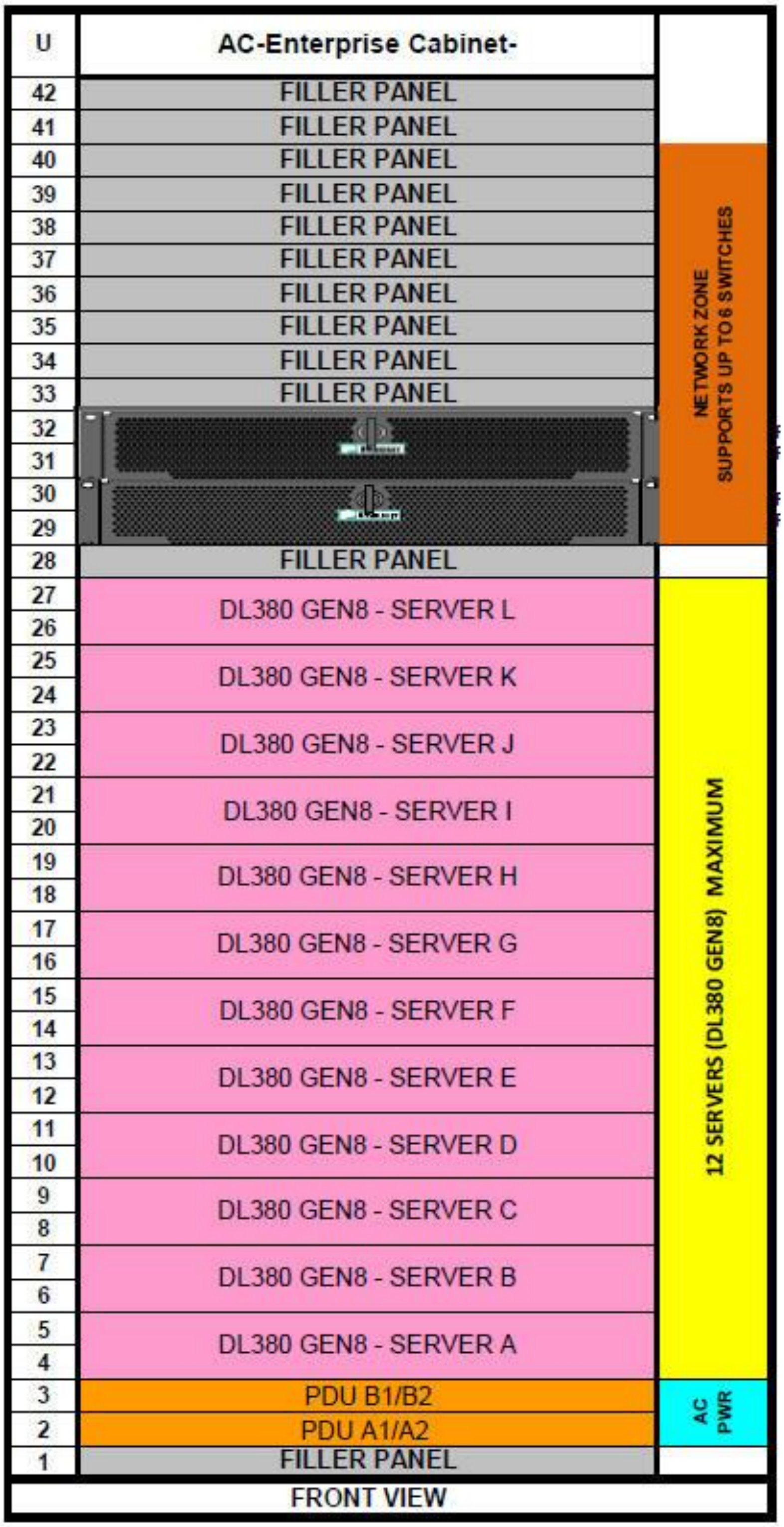

Figure 57: AC RMS Example Configuration

U	DL380 GEN8 GEN8 DC RMS Cabinet
44	PDP A - TELECT 4/4 PANEL
43	PDP B - TELECT 4/4 PANEL
42	FILLER PANEL
41	PDP C - TELECT 4/4 PANEL
40	FILLER PANEL
39	FILLER PANEL
38	FILLER PANEL
37	FILLER PANEL
36	FILLER PANEL
35	FILLER PANEL
34	FILLER PANEL
33	FILLER PANEL
32	FILLER PANEL
31	SWITCH B
30	
29	SWITCH A
28	
27	DL380 GEN8 - SERVER L
26	
25	DL380 GEN8 - SERVER K
24	
23	DL380 GEN8 - SERVER J
22	
21	DL380 GEN8 - SERVER I
20	
19	DL380 GEN8 - SERVER H
18	
17	DL380 GEN8 - SERVER G
16	
15	DL380 GEN8 - SERVER F
14	
13	DL380 GEN8 - SERVER E
12	
11	DL380 GEN8 - SERVER D
10	
9	DL380 GEN8 - SERVER C
8	
7	DL380 GEN8 - SERVER B
6	
5	DL380 GEN8 - SERVER A
4	
3	FILLER PANEL
2	FILLER PANEL
1	FILLER PANEL

Figure 58: DC RMS Example Configuration

Oracle Corporation, World Headquarters
500 Oracle Parkway
Redwood Shores, CA 94065, USA

Worldwide Inquiries
Phone: +1.650.506.7000
Fax: +1.650.506.7200

CONNECT WITH US

blogs.oracle.com/oracle

facebook.com/oracle

twitter.com/oracle

oracle.com

Integrated Cloud Applications & Platform Services

White Paper Oracle Communications Diameter Signaling Router Release 7.0.1 Planning Guide
November 2015