Globberry | Version 1.0 - March 2017

Cloud DRA

DIAMETER ROUTING AGENT

Introduction

This document describes the Cloud Diameter Routing Agent (DRA) software suite developed by Globberry.

Conceptually, DRA is a 3GPP-compatible system for Diameter message distribution and routing used in various signaling flows of the new generation IMS networks.

The word Cloud in the product name reflects the fact that the product is based on micro-service architecture and can be deployed on an arbitrary configuration of physical or virtual servers, including servers in the cloud.

Table Of Contents

The key functions of Cloud DRA are:

- **Message Relay.** Distributing Diameter messages transmitted over various signaling interfaces to the target network components, with flexible and configurable rules for routing, load balancing and failover

- **Subscriber-Dependent Routing.** Distributing Diameter messages to the target components responsible for servicing a specific subscriber group, when subscribers are split into logical groups based on operator's policies

- **Session Binding.** Routing of all related messages for the same data transfer session to the same set of service components, including messages from different Diameter interfaces, such as Gx/Rx in VoLTE

- **Host/Realm Anonymizing.** Deleting the host/realm data from the Diameter responses, thereby preventing external systems from discovering the equipment type and host addresses of any components inside the operator's network, except for the DRA itself

- **Mediation.** Modification of certain attributes inside the routed Diameter messages, with the purpose of maintaining compatibility between various equipment types and/or enriching the messages with session and subscriber information

- **Interworking.** Conversion of RADIUS messages to the Diameter format, and vice versa

- **Throttling.** Limiting the load on internal network components during unexpected peaks/bursts of incoming Diameter messages, in order to protect them from failure and maintain the overall health of the operator's network.

The DRA standard was introduced by 3GPP to address the increased Diameter signaling traffic and growing complexity of 4G LTE networks

With the introduction of Cloud DRA, mobile operators can:

- Facilitate and accelerate modifications to the Diameter routing topology when adding new network components and subsystems, by relocating all routing configuration functionality to the Cloud DRA elements

- Achieve uniform and consistent management of the routing rules for different implementations of the Diameter protocol used in the telecommunications industry, such as Gx, Gy, Sy, Rx, Rf, Ro, S6a/S6d, S9 and S13.

- Regardless of the Diameter interface being used, forward all messages related to the single data transfer session to the same set of network components, for example, to coordinate data and application sessions in VoLTE and other similar applications

- Optimize network resources usage through dividing the subscriber base by home location or any other criteria, and service each of the subscriber groups using the subset of network components allocated specifically for that group

- Decrease the number of connections between components, by an order of magnitude

- Easily scale the existing network under the Diameter load growth by further subscriber segmentation, horizontal scaling of the components and by centralized reconfiguration of the Diameter signaling topology inside Cloud DRA

- Balance the load between components of the same type, as their configuration is scaled horizontally

- Use Cloud DRA as a Diameter Edge Agent (DEA), providing a single entry point for all Diameter signaling exchanges with the external systems, such as billings of the partnering roaming operators

- Defend from external hacking by fully concealing the internal network topology, through anonymizing the realms and host addresses of all network components receiving and processing Diameter requests

- Defend from DoS-attacks and outages by monitoring the incoming Diameter TPS's and enforcing preset limitations on the number of requests passed on to the internal network components during unexpected peaks/bursts of the incoming Diameter traffic

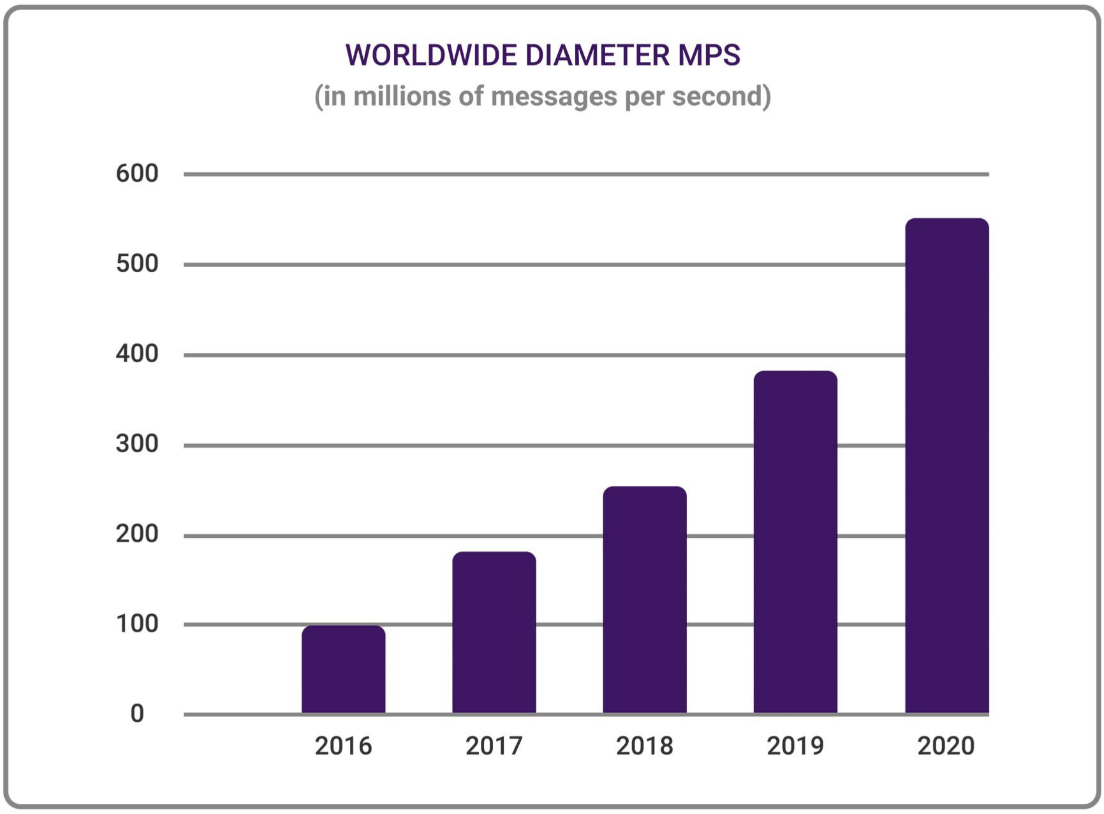
By 2020, LTE networks will generate over 500 million Diameter-messages per second (500,000,000 MPS)

- Seamlessly integrate network components from different vendors using different Diameter "dialects" – including "dialects" with proprietary AVPs – by adjusting the contents of the Diameter messages at the time of their routing

- Enrich Diameter messages with subscriber and/or session attributes known to the telecommunications operator, but absent from the incoming messages, for various reasons

- Introduce network components with RADIUS interfaces into the networks with Diameter-based signaling

- Collect all statistical and performance data (KPIs) for all Diameter signaling interfaces within the operator's network from a single location.



Expected growth in the number of worldwide Diameter message exchanges
Source: Oracle Corp.*

DRA in Next Generation Networks

The role of Diameter signaling within next generation networks matches the role of the SS7-signalling in traditional switched networks. Using the common Diameter protocol as a basis, 3GPP, IETF and GSMA have developed a set of detailed specifications – including Gx, Gy, Sy, Rx, Rf, Ro, S6a/S6d, S9, S13 – regulating all aspects of Diameter signaling in the new IP-based networks. It is impossible to imagine modern networks without Diameter interfaces; at the same time, the need for backward compatibility with the existing mobile networks may require interaction between Diameter and the traditional signaling interfaces, such as SS7.

Mobile data usage grows exponentially in all parts of the world, and with it grows the number of Diameter messages generated every day. By many estimates, by 2020, LTE networks worldwide will generate over 500 million Diameter messages per second (500,000,000 MPS). The year-to-year increase in the overall number of generated Diameter messages is approximated at 50%, and in some markets – such as in Eastern Europe – at 100%. Besides the traditional use of Diameter for charging and policy enforcement (PCC), the growth of mobile and signaling traffic in the upcoming years will also be driven by the new mobile technologies and applications, such as VoLTE, LTE Broadcast, Car Connectivity and Internet of Things/M2M.

Due to the sharp increase in Diameter signaling, the next generation network architecture must place a greater emphasis on the flexibility of the component scaling. In the traditional peer-to-peer configurations, all interacting components are directly connected to each other:

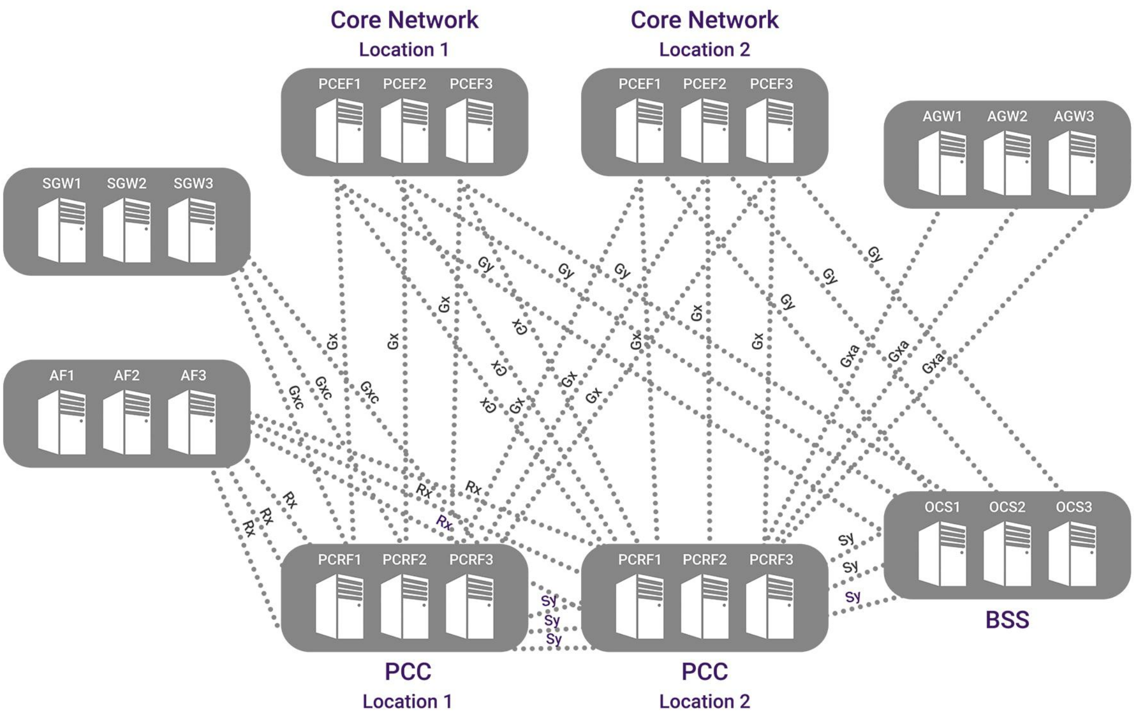


Figure 1. Peer-to-peer network topology

The number of interconnected Diameter interfaces represents an obvious problem during network upgrades, because installation of the new hardware not only requires the reconfiguration of the components being added, but also necessitates changes to many other components within the network. Load rebalancing among the new components and their inclusion in the existing failover algorithms may also become a challenge. The need to have direct network connectivity between all of the components increases their vulnerability to hacking attacks. Finally, the Diameter interfaces already implemented within the network tie the operator to specific hardware vendors; adding components from a new vendor usually requires substantial effort adapting existing interfaces to the Diameter "dialects" used by that vendor.

The problems can be solved by introducing DRA into the operator's network:

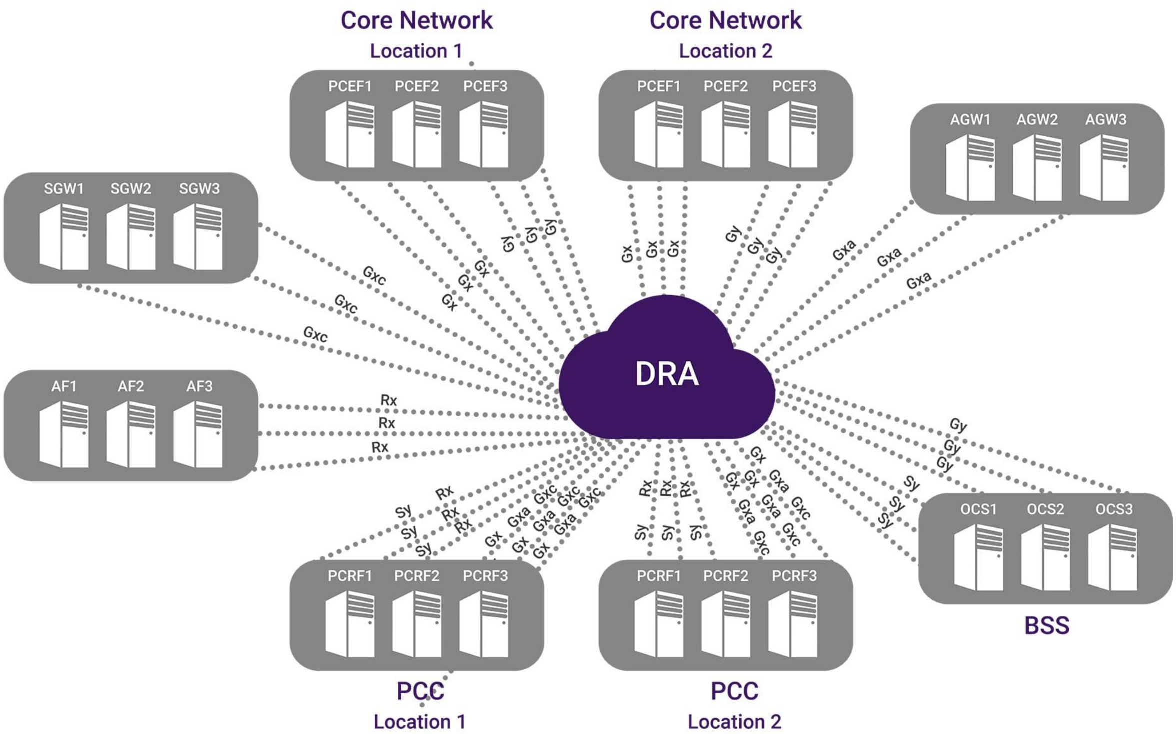


Figure 2. Network topology with DRA

- Instead of a large number of interconnected Diameter interfaces, all Diameter connections are consistently routed via DRA
- Addition of a new component to the operator's network requires reconfiguration of only the DRA and the component itself
- Flexibility of the DRA configuration and routing enables an easy and straightforward upgrade of the load balancing and failover algorithms
- Ability to adjust the content of the Diameter messages as they are routed through DRA facilitates the addition of components from alternative vendors that may use different "dialects" of the Diameter protocols.

For operators who have Diameter interfaces accessible from outside – for example, open to their roaming partners – Cloud DRA can also be used as a **Diameter Edge Agent (DEA)**:

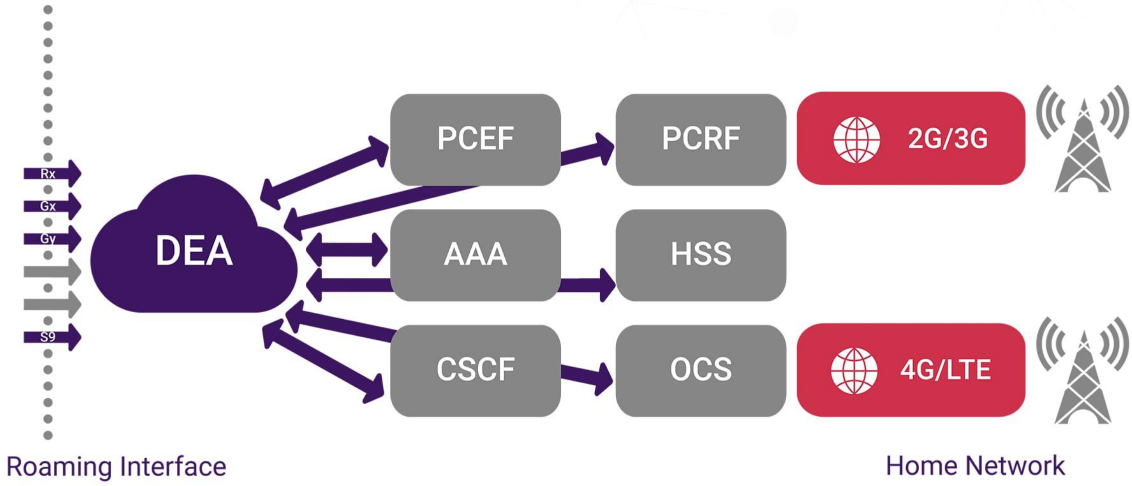


Figure3. Using DRA as a Diameter Edge Agent

- Ability to adjust the content of the Diameter messages as they are routed through DEA (DRA) enables integration with roaming partners using different Diameter “dialects” in their own networks
- The anonymizing feature will enable DEA (DRA) to fully conceal specific topology of the home network (hosts and realms of the internal components), thereby substantially decreasing the vulnerability to outside hacking
- The throttling feature of DEA (DRA) will guarantee that any Diameter request bursts, from either roaming network accidents or from intentional DoS-attacks, will not lead to an uncontrollable increase in the Diameter load on the internal components, thus eliminating the risk of home network failures.

Cloud DRA: Functionality

Message Relay

Within CSP architecture, several components with the same functionality receiving Diameter messages can be combined into a group, and Cloud DRA can distribute incoming Diameter requests to a random member of the group. If one or more components become unavailable, those components are excluded from the delivery list, and the requests are distributed evenly among the remaining components. Cloud DRA monitors availability of the components, and if the failed components become available again, request distribution to them will resume after a certain time.

The selection of the target component group that should receive and process the incoming Diameter request is configurable in Cloud DRA, and, in general, depends on the attributes of the Diameter request, its type, and on the client system submitting the request.

Cloud DRA can store in its memory the correspondence between the subscriber's identity (MSISDN/IMSI), session identifier and the target component for request delivery, and guarantee distribution of all requests for the same subscriber session to the same target component selected at the beginning of the session.

For each component group, Cloud DRA can store the list of IP-addresses / hosts that can submit Diameter messages to that group, and suppress messages from any other sources.

Subscriber-Dependent Routing

Cloud DRA can select the target component for the Diameter request routing of a Diameter request based on the subscriber's attributes fetched from the Subscription Profile Repository (SPR). The operator may choose to divide his or her subscribers into groups using a number of criteria from the subscriber profile: PostPaid vs. PrePaid, by geographical location (home region), etc.

Such a division can be helpful, for example, for configuring the number of parallel Diameter client components with identical functionality but a different set of served subscribers, when performance of a single component is insufficient for handling all the existing load on the Diameter interface. In many cases, such subscriber segmentation can save hardware resources and memory, as each of the target Diameter servers must only store the information on its own group of subscribers. The need for data replication between the parallel components is either eliminated or substantially decreased.

Cloud DRA can analyze the values of the subscriber's attributes, and, in certain cases, form and return the default response itself, instead of delivering the request to the target component. For example, if the target Diameter component performs online charging via Diameter Gy, Cloud DRA itself can respond to requests for the offline subscribers who do not need to be charged via Diameter Gy online.

Degraded Mode

Sometimes, all components in the target Diameter group may become unavailable, because of an incident. In such conditions, Cloud DRA can be temporarily switched to a so-called “Degraded Mode”, forcing the Cloud DRA itself to generate default answers to all messages on the current Diameter interface – rather than pass them along to the target Diameter components, as in the normal mode of operation. The Degraded Mode can be used by operators at times of network problems and outages when, instead of cutting off all network access, they would rather provide service to their subscribers, even on incorrect terms.

Session Binding

The routing rules for different Diameter interfaces, such as Gx and Rx, are configured on Cloud DRA independently. By configuring matching sets of rules for a few related Diameter interfaces, Cloud DRA administrators can ensure that all messages for the same data session will be delivered to the same target component, regardless of the client generating the Diameter message and the interface in question. This will give the target component a chance to accumulate signaling from different incoming Diameter interfaces and process the information holistically. For example, in VoLTE scenarios, when all messages are delivered to the same target PCRF component, a Diameter Rx message can trigger activation of a new voice bearer within a context of the general data session, initiated earlier with a Diameter Gx message.

Host/Realm Anonymizing

Cloud DRA selects the component group and the specific target component based on the Destination-Host, Destination-Realm, Application-Id, as well as Origin-Host, Origin-Realm and some other AVPs in the Diameter request. Before distributing the message, Cloud DRA substitutes the values in the Destination-Host and Destination-Realm AVPs with the respective values for the target component to which the Diameter message is being delivered. When the response is routed, a reverse substitution is made to the Origin-Host and Origin-Realm AVPs.

If the external client systems direct their messages to the Destination-Host and Destination-Realm of Cloud DRA rather that the final target component, and then Cloud DRA determines the real target based on Application-Id and other AVPs in the message, the Origin-Host and Origin-Realm fields in the responses will also contain addresses of the Cloud DRA itself. As a result, all information regarding the internal topology of the operator’s network is hidden from the client systems.

The described behavior pertains not only to the CCR/CCA message exchanges, but also to the RAR/RAA and ASR/ASA exchanges, when the Diameter communication is initiated by an internal component located “behind” Cloud DRA.

Mediation

When a Diameter message is routed through Cloud DRA, the value of any AVPs within the message can be adjusted. Such adjustments can be divided into two major categories:

- Conversion between different Diameter "dialects", when values with the same meaning are described in the source and target systems by different constants, conflicting data types (such as Integer vs. String), or they are in different AVPs.
This category also includes interpretations of values in proprietary AVPs, when meaningful data is transmitted in AVPs unknown to other Diameter vendors

- Enrichment of Diameter messages with data from the SPR, including personalized subscriber data that can affect the conditions of the data service, such as values for QoS or a Static IP address.

Interworking

Besides Diameter, Cloud DRA can also receive and deliver RADIUS protocol messages. RADIUS attribute values are converted to the respective Diameter AVPs and are transmitted to the target system, following all routing rules configured in Cloud DRA in a standard fashion. The responses are converted back to the RADIUS format and delivered to the client system.

During conversion to the Diameter format, RADIUS data can be enriched with attributes from the SPR, as described above. In this case, such enrichment is especially important, as RADIUS messages may lack certain key attributes, due to the limitations of the RADIUS format.

Throttling

Cloud DRA administrators can configure throughput limits for Diameter interfaces, in terms of TPS and bursts, both on the incoming connection and on the outgoing connections towards each of the target components. When the throughput limit is exceeded, the Diameter message is not delivered; instead, the client system receives a preconfigured response from the Cloud DRA itself, with a given return code. This functionality protects the target Diameter components from an arbitrary load increase to levels they cannot handle.

Cloud DRA administrators can specify a few different throughput limits for different Diameter request types, thereby providing a set of subsequent cutoff levels for the Diameter messages of different importance. As an example of possible use, through configuration of message priorities, the operator may guarantee that the existing data sessions are serviced first, and new data sessions are started only when sufficient bandwidth for processing the Diameter requests is available.

CDR Generation

For certain Diameter interfaces, Cloud DRA may produce CDR file flows registering the history of all events and sessions that occur on the given interface. The set of the Diameter fields recorded in the CDR depends on a specific Diameter interface. It may contain some common fields – such as time, session identifier, MSISDN, IMSI,

client and server addresses – as well as fields specific to a given Diameter interface, such as APN in Diameter Gx or traffic volume in Diameter Gy. A single CDR may represent an individual Diameter message or a Diameter “session”: a combination of messages denoting start and end of service delivery to a subscriber.

CDR files document the complete service history for individual subscribers and can be analyzed in case of any disputes or discrepancies. They also provide a good source for network statistics, and can even be used to charge subscribers for certain services or events in the offline mode.

KPI Collection and Metrics

For the Diameter interfaces of all types, Cloud DRA collects the statistics on the number of messages, TPS performance and other similar parameters, saving them to the log files with given frequency. The Cloud DRA analytical scripts can process and transmit these statistics as metrics to external monitoring systems that can then display and control the performance of various Diameter interfaces in real time. Cloud DRA also provides additional metrics for error management, system resource usage and monitoring of various other indicators of the system’s health.

Cloud DRA: Architecture

Cloud DRA is the service provider-level system with microservice based architecture. Cloud DRA's logical modules exist as autonomous components, interacting with other Cloud DRA modules via ReST/http. The system uses a Service Discovery mechanism to detect running modules and build links between the components in real time. Such architecture offers great flexibility for horizontal scaling of the internal Cloud DRA components, as additional modules of known types can be started up in the Cloud DRA environment in real time, automatically taking over some of the load. The architecture also ensures high reliability and accessibility, as separate modules are executed in separate containers, and failure of a single module simply leads to redistribution of its load among all other remaining modules of the same type, through the automatic rebalancing algorithms implemented in Cloud DRA.

Cloud DRA contains modules of the following types:

- Diameter and RADIUS interface adapters, converting messages to and from Cloud DRA's internal format
- Message routers, making and executing decisions about message distribution and delivery
- Message processors, converting and enriching attribute values inside Diameter messages
- Business Rule servers, supplying user-configured rules for routing and data conversion to the other modules
- SPR adapters, fetching the subscriber data from the Subscription Profile Repository
- Message brokers and file generators for saving logs, CDRs and other permanent data
- Monitoring modules for interfacing with external monitoring systems
- Service Discovery clients and servers, automating the integration and interaction of all the modules above.

At the operator's discretion, all Cloud DRA modules can be deployed to either physical or virtual servers, including servers in the cloud. When selecting a deployment option, the only requirement is the operator's ability to guarantee reliable physical http-connectivity between all Cloud DRA modules.

The reliability and high availability of the Cloud DRA can be achieved by either N+1 reservation of all aforementioned modules, or by deployment of several "complete" Cloud DRA instances interacting with each other. In the second scenario, the external Diameter clients can balance their requests between several Cloud DRAs. The reliability and availability of the system can be further enhanced with geographical reservation, by deploying Cloud DRA instances/modules on servers in different locations or different regions.

Cloud DRA: Roadmap

The following functionality is planned for future Cloud DRA releases:

- Addition of an industrial-grade rule engine for greater flexibility in configuration of the routing and balancing rules, as well as for defining the logic for the attribute value conversion and enrichment during the routing
- Automatic detection of network congestion and definition of alternative rules for routing and balancing during congestion
- Priority processing of Diameter signaling for certain high-value applications.

Standards and Specifications

3GPP

Diameter Routing Agent (DRA): TS 29.213

Subscription Locator Function (SLF): TS 29.228

IETF

Diameter Relay/Proxy Agent (DRL): RFC 3588, RFC 6733

Diameter Redirect Agent (DRD): RFC 3588

Diameter Translation Agent (TLA): RFC 3588

GSMA

Diameter Edge Agent (DEA): IR.88

S6a/S6d Proxy: IR.88

Globberry

globberry.com

info@globberry.com