Ericsson Review
284 23-3176 | Uen

Diameter Signaling Controller in next-generation signaling networks

At the heart of the evolved mobile data network almost everything uses the Diameter protocol to communicate.

JÖRG EWERT, LENNART NORELL AND SONER YAMEN

The Diameter protocol is used widely in all-IP networks. Nearly everything that is connected to or is part of a network uses the protocol in some way. The evolved mobile-data nodes use it to communicate with each other and applications use it to get the data they need. The number of signaling messages being sent is rising rapidly, which is putting pressure on all parts of the network – and particularly on gateways, charging systems, policy servers and user-data repositories. Deploying a Diameter Signaling Controller (DSC) – a key network component – can relieve some of this pressure while boosting operational efficiency and increasing the reliability of the internal signaling network.

Introduction

The evolution of IMS and mobile-broadband network architectures is closely linked with Diameter. The protocol was first introduced for communication over the interfaces between the IMS core and application servers, charging systems and HSS databases. In the EPC, the protocol is used for policy control in addition to accessing user and charging information. The Diameter-signaling architecture for IMS and EPC is illustrated in Figure 1.

For mobile-broadband networks, Diameter performs the same functions as SS7 in roaming interfaces and non-call-related signaling. Indeed the issues related to the increase in Diameter signaling have been likened to the problems that arose when SS7 was introduced in the first mobile networks. Diameter is one of the main pillars in the transformation of network signaling to native IP-based protocols, complementing SIP – which is used for call-control signaling in IMS.

Diameter

The specification of this protocol began in 1998. The objective was to create a framework for authentication, authorization and accounting (AAA) that would overcome the limitations of the RADIUS protocol in terms of reliability, security and roaming support.

The framework is a base protocol[1] that defines the minimum mandatory set of AAA operations. The base protocol defines the message format, which comprises a header and a set of data elements expressed as attribute-value pairs (AVPs). By expressing data as AVPs, applications using the protocol can be extended in the future with-

BOX A Terms and abbreviations

3GPP	3rd Generation Partnership Project
AAA	authentication, authorization and accounting
AVP	attribute-value pair
DA	Diameter Agent
DCCA	Diameter Credit Control Application
DEA	Diameter Edge Agent
DRA	Diameter Routing Agent
DSC	Diameter Signaling Controller
EIR	Equipment Identity Register
EPC	Evolved Packet Core
GSMA	GSM Association
GUI	graphical user interface
hPCRF	home PCRF
HPLMN	Home Public Land Mobile Network
HSS	Home Subscriber Server
IANA	Internet Assigned Numbers Authority
IETF	Internet Engineering Task Force
IMS	IP Multimedia System
IMSI	International Mobile Subscriber Identity
IP	Internet Protocol
IP-CAN	IP connectivity access network
IPsec	IP Security
IPX	IP Packet Exchange
LDAP	Lightweight Directory Access Protocol
LTE	Long-Term Evolution
MAP	Mobile Application Part
MME	Mobility Management Entity
O&M	operations and maintenance
OTT	over-the-top
PCC	policy and charging control
PCRF	policy and charging rules function
P-CSCF	proxy call session control function
PGW	packet data network gateway
PRD	Permanent Reference Document
RADIUS	Remote Authentication Dial-In User Services
SCTP	Stream Control Transmission Protocol
SGSN	Serving GPRS Support Node
SIP	Session Initiation Protocol
SLF	Server Locating Function
SS7	signaling system 7
TCP	Transmission Control Protocol
TLS	Transport Layer Security
TPS	transactions per second
VoLTE	voice over LTE
vPCRF	visited PCRF
VPLMN	Visited Public Land Mobile Network

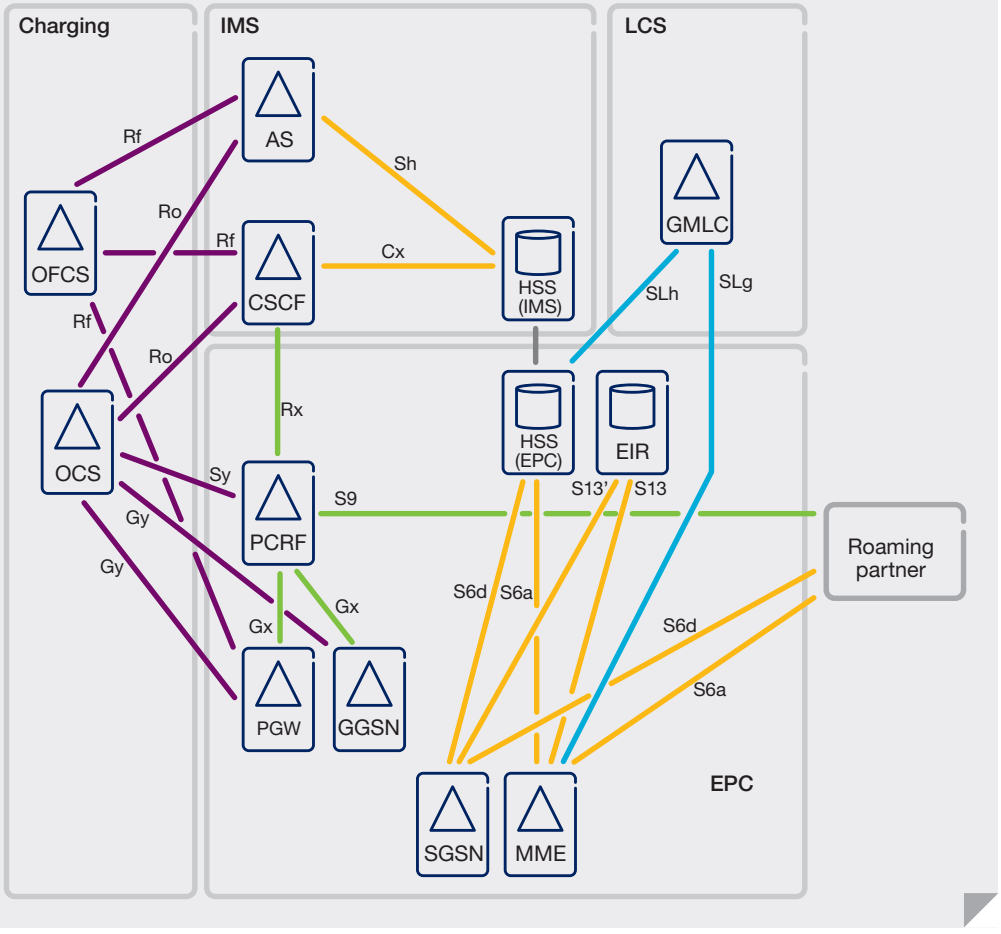
out having to modify existing source code or data inputs; new information is simply added as a new AVP. The base protocol defines a set of commands and AVPs that can deliver the minimum set of signaling functions, such as peer discovery, capability exchange, proxying, loop detection and error handling. The base set of operations can be extended by Diameter applications through the addition of new commands and AVPs. As illustrated in **Figure 2**, Diameter supports both SCTP and TCP transport protocols, and transport security is provided by the TLS or IPsec protocol.

Diameter is a peer-to-peer protocol that uses a request-answer transaction format. A Diameter peer can be a client, a server or an agent – a Diameter Agent (DA). Agents are positioned between clients and servers, and forward client requests to the appropriate server.

There are four kinds of agents: relay, proxy, redirect and translation. A relay agent uses the header information and routing-related AVPs of a message to choose the destination peer. A proxy agent can modify AVPs in the message, it forwards messages, and may use the AVPs to determine the destination or apply a policy – for example, to reject a request. Redirect agents return requests to the originating client, providing information on the appropriate next hop that can service the request. Translation agents translate messages from one protocol into another. This type of agent was originally defined to translate messages from AAA protocols, such as RADIUS, to Diameter. However, translation from LDAP and MAP to Diameter is also of interest.

The Diameter framework can be developed by extending existing applications or by creating new ones. An existing Diameter application can be extended through the addition of optional AVPs. To implement additional functionality new Diameter applications need to be defined which may imply new command codes and new sets of mandatory AVPs. The 3GPP Ro protocol is an example of how an existing Diameter application – the IETF-specified Diameter Credit Control Application (DCCA) – has been extended with additional AVPs to support the exchange of charging information.

FIGURE 1 **Diameter-signaling architecture for IMS and EPC**



However most of the interfaces in 3GPP (S6a, Cx, Rx, Sh and so on) have their own specific Diameter application.

Each application requires a specific ID, which is assigned by IANA; the S6a interface, for example, has the application ID 16777251.

Figure 3 shows a typical Diameter message, comprising a Diameter header and a series of AVPs. The example shown is an Update-Location-Request message sent over the S6a interface from the MME to the HSS.

Main use cases

The purpose of a DSC is to facilitate the ever-increasing use of Diameter signaling in mobile networks, and **Figure 4** shows a reference architecture for the use cases described here. This key

BOX B **Diameter interfaces**

Cx, Sh	subscription and authentication data – IMS	**S6a,S6d**	subscription and authentication data – EPC
Gx	QoS/policy – EPC	**S9**	QoS/policy – EPC
Gy	online charging – EPC	**S13, S13'**	EIR query – EPC
Rf	postpaid charging – EPC/IMS	**SLh, SLg**	location-based services – EPC
Ro	online charging – EPC/IMS	**Sy**	online charging – EPC
Rx	QoS/policy – EPC		

FIGURE 2 **Structure of the Diameter stack**

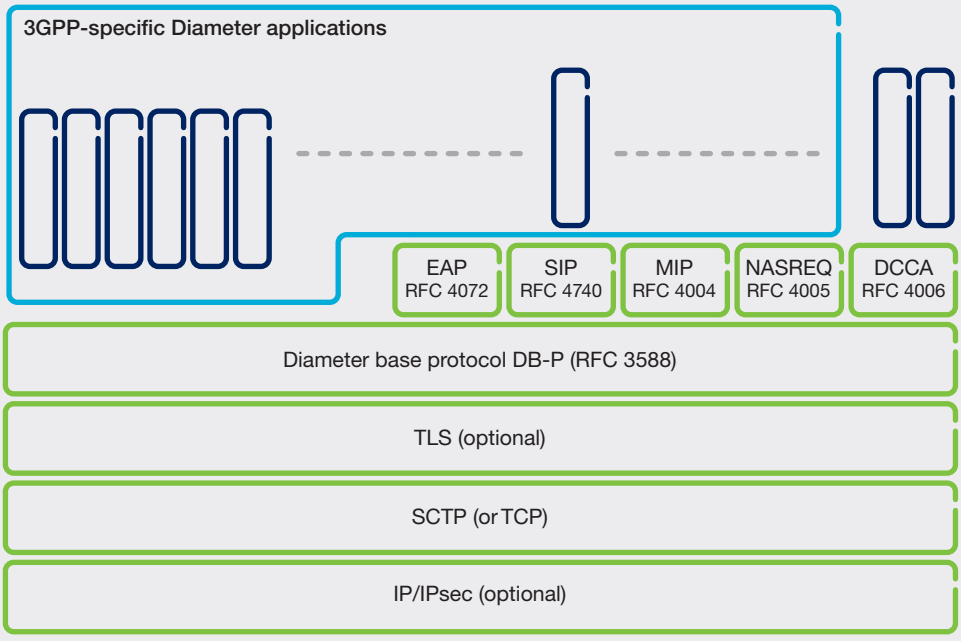
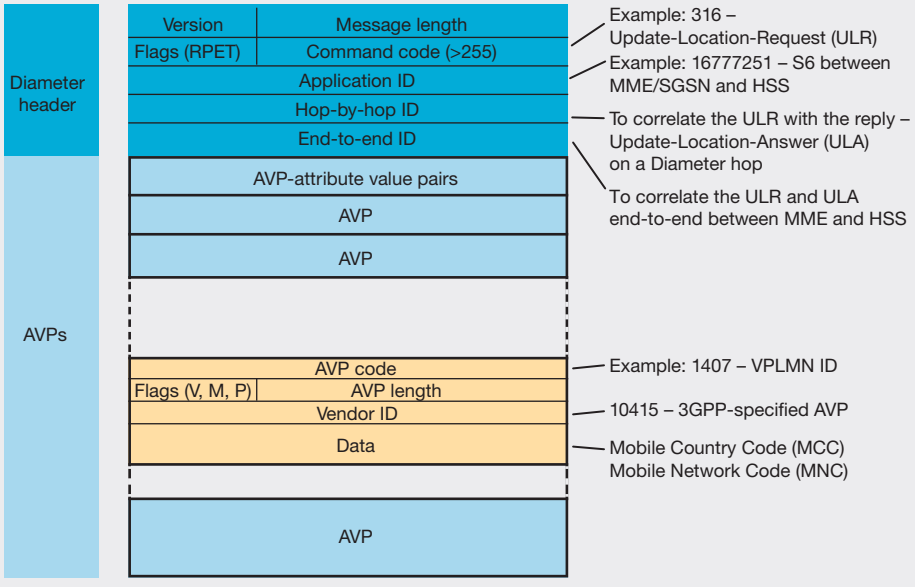


FIGURE 3 **Diameter message structure and an example message**



network component supports the scaling of Diameter-signaling networks, which will be significantly beneficial when the growth of mobile data and VoLTE traffic requires the deployment of multiple HSS, MME, PCRF, P-CSCF and PGW network elements.

It is expected that the growth in signaling traffic (accumulated transactions per second) will reflect the growth in mobiled ata traffic – a tenfold increase between 2011 and 2016[2]. In the near term, this growth will be driven primarily by LTE rollouts. From 2014, growth will be accentuated by the introduction of VoLTE. The amount of signaling traffic will be further increased as the installed base of GSM/WCDMA packet-switching networks starts to transition from SS7 to Diameter.

Centralized routing

As illustrated in **Figure 5**, positioning a DSC centrally in the core network dramatically reduces the number of connections required. As the number of peer relations to configure in a full mesh is directly proportional to the square of the number of interconnected devices, centralized positioning of a DSC can reduce configuration time considerably, and the time and effort required to add and configure another Diameter signaling peer – such as an MME – can be controlled.

With a well-designed centralized DSC, signaling flow can be monitored, network faults can be isolated, and signaling traffic can be rerouted for maintenance purposes – making network expansion a much simpler process.

When the DSC routes a request, two or more peers can usually serve it. The DSC uses a load-balancing algorithm to distribute load over the available servers. The complexity of this algorithm can vary, ranging from a simple round robin to a more evolved solution that takes current server loads into consideration, for example.

Overload protection

A DSC can enhance the robustness of the network by supporting intelligent context-aware throttling of signaling load for servers suffering from unbalanced load, and for clients that misbehave.

Overload and signal flooding can be caused, for example, by:

- the mass registration of mobile phones and inter-node signaling directly following network recovery;
- the handover of huge numbers of sessions following a radio-network failure; or
- reregistration of smartphone applications following the failure of an OTT server.

The DSC can protect itself and other connected server peers from overload by throttling messages. However, any message discarded by the DSC implies a lost processing effort (blind load) in the originating network elements and increases the time until the network settles to a normal state again. Therefore, it is important to perform load regulation as close as possible to the source of the overload.

LTE roaming support

In the EPC, there are direct Diameter interfaces that connect the network elements of the visited network – MME, SGSN and vPCRF – to the equivalent elements of the home network: HSS and hPCRF. To simplify the roaming interface, an additional functional entity – the Diameter Edge Agent (DEA) – has been defined in the LTE Roaming Guidelines (GSMA PRD IR.88).

The DEA provides an entry point that hides the topology of the network behind it and advertises itself to roaming partners as a Diameter relay serving all Diameter applications in the network. As such, the DSC should be considered as a signaling firewall that protects the internal network from malformed messages and unauthorized senders. This firewall functionality is implemented by filtering messages through a set of customizable Diameter message screening rules and message normalization can be done by rebuilding all messages using a specified layout.

Session binding

3GPP has defined a new functional element for policy and charging control (PCC) architecture. This element – the Diameter Routing Agent (DRA) – ensures that all Diameter sessions established over the Gx, Rx and S9 reference points for a specific IP-CAN session use the same PCRF instance when multiple PCRFs have been deployed.

For example, sessions created over Gx for LTE users may be routed to a specific PCRF based on the subscriber identifier – IMSI. When another session is created over Rx – for a VoLTE voice call, for example – this session is identified by the assigned IP address. To ensure that session requests are routed to the correct PCRF, the DRA must maintain the relationship between the subscriber identifier, the assigned IP address and the chosen PCRF instance for the duration of the IP-CAN session.

Like the DEA, the DRA is a functional element of the DSC. Consequently, maintaining all these relationships is a challenge because DSCs are typically deployed as redundant pairs. The stored state for all IP-CAN sessions must be shared between the redundant DSC pair so that a DSC outage does not result in major loss of IP-CAN sessions.

FIGURE 4 **Reference architecture for use cases**

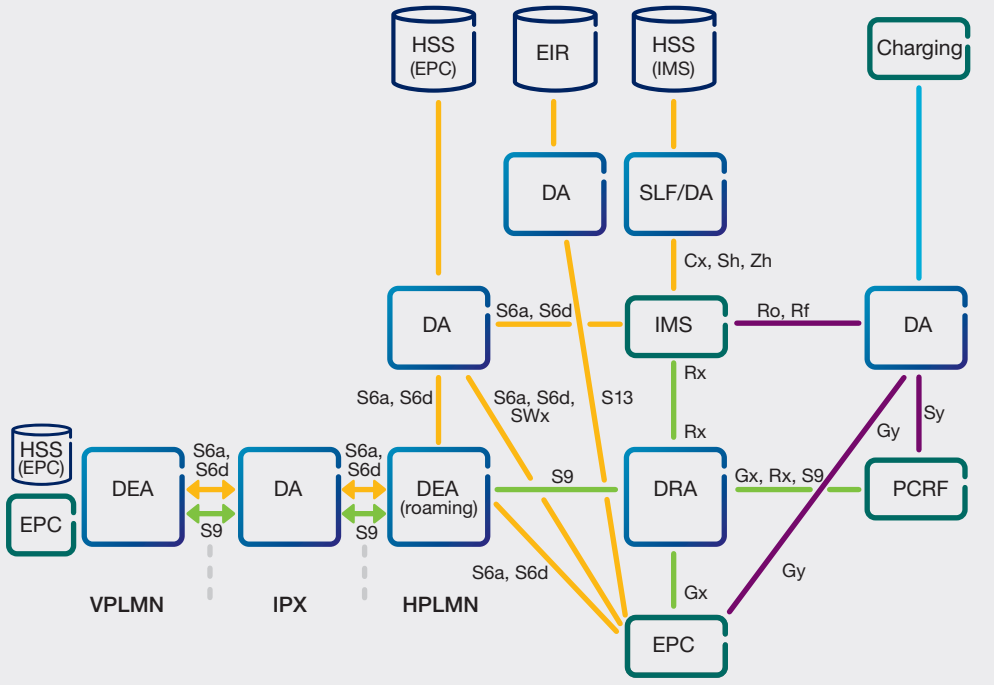


FIGURE 5 **Centralized versus distributed Diameter-signaling network architecture**

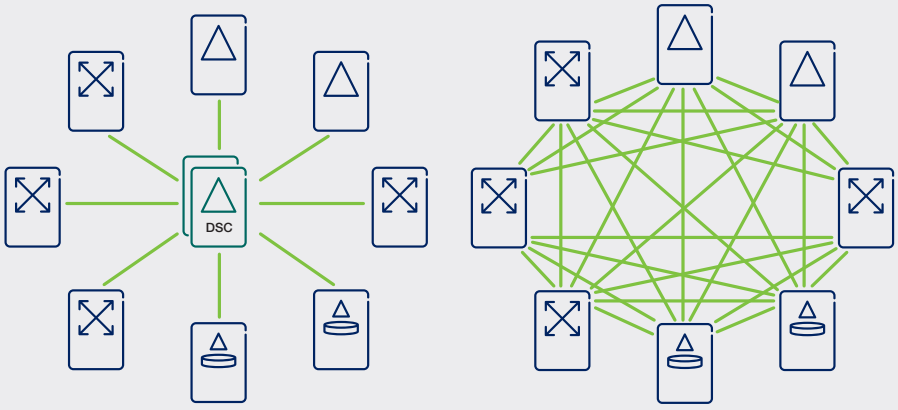
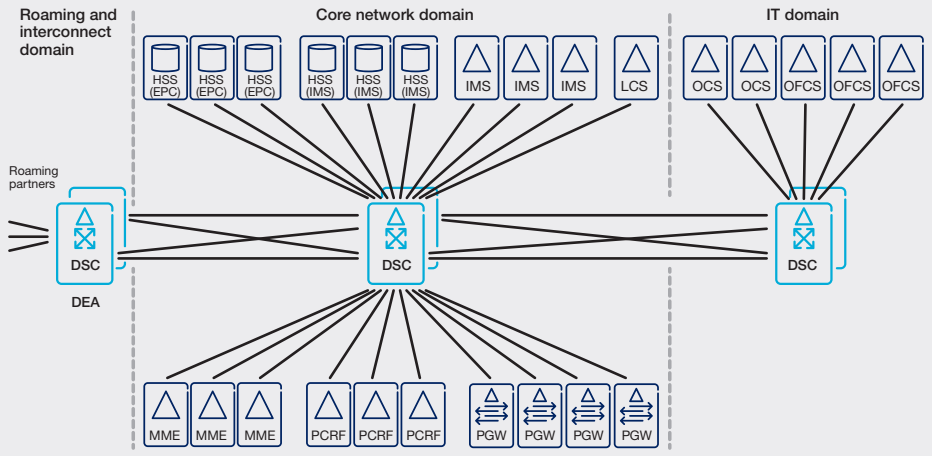


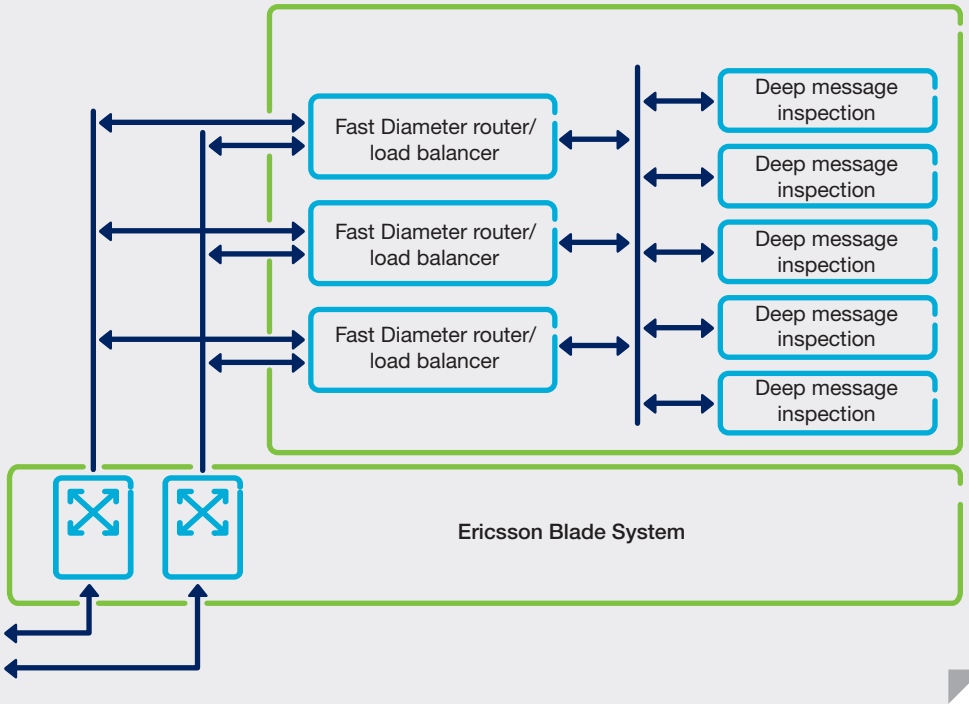
FIGURE 6 **DSC including logical DA, DEA and DRA elements**



Address resolution of nodes

When user data is distributed over multiple instances, the DSC must be able to locate the correct instances for a particular user. 3GPP specifies this in the reference points from EPC and IMS to the HSS that stores the user data. For Diameter applications using reference points Cx, Dx, S6a, S6d, Sh and Dh, the DSC uses the subscriber identity stored in an AVP in the message to select the appropriate server instances (typically two or more), and it then applies load-balancing. To do this, the DSC must be aware of the Diameter application so that the request can be routed to the appropriate server instance.

FIGURE 7 **Sample DSC topology**



The Diameter interfaces to other nodes that store user data, such as the PCRF or online charging systems, may also need to utilize this feature in the DSC.

For centralized user database deployments or when an SLF is used to identify the location of user data, the DSC only needs to identify the set of front ends or SLFs that route the request further. In this case, the DSC does not need the subscriber identity to route the message.

Deployment topologies

In small to medium-sized networks (serving up to 10 million subscribers), the Diameter signaling network can be collapsed into a single mated pair of DSCs.

However, it may still be beneficial to divide the signaling network across several interconnected DSCs, each serving a separate and independently administered domain. In the example shown in **Figure 6**, one DSC pair handles the international Diameter traffic as a DEA, a second DSC pair handles the national core Diameter signaling, and a third DSC pair handles the signaling to charging systems. The DSC pairs act as an interface between domains, and consequently each domain can evolve without impacting any other.

Very large networks (serving more than 10 million subscribers) are often subdivided into geographical regions. For such networks, DSC pairs may be deployed for each region, where all DSCs are linked to each other providing the signaling interconnect.

The Ericsson approach

As a central node in the network, the DSC must be carrier-grade complemented with a robust network design – which requires both node- and network-level redundancy. At node level, N+1 redundancy is applied for traffic-carrying blades, and other hardware components are 1+1 protected. Network-level redundancy is achieved via mated DSC pairs, which can be situated at different geographical sites.

Figure 7 illustrates the DSC, which from a software-architecture perspective is divided into two: a front-end router that handles most message routing through a peer table and an extended diameter-routing table; and a back-end part handling deep message inspection.

The front end can handle frequently occurring proxy functions and is also responsible for the load balancing and throttling functions. This part is optimized for high-capacity message processing with minimum delay.

The back-end part is invoked when a message requires screening and when – based on highly customized business logic – modification of the message content is required. The back end supports a comprehensive GUI for easy customization of message handling – an essential function for configuring the interfaces with roaming partners.

The front- and back-end parts are administrated by a common operations and maintenance (O&M) concept. However, two different O&M roles are defined to fit with the main use cases: basic diameter signaling control and diameter message manipulation.

Each blade in the DSC runs both a front-end and a back-end software process. Consequently, all blades are equal in terms of software configuration, which eases expansion and supports a cost- and time-effective configuration process. The load balancer distributes incoming messages to the front end. And requests that require invocation of back-end processes are distributed so that load is balanced.

The DSC can be scaled up from minimum configuration of two traffic-serving blades to any size simply by adding or removing blades. The configuration of cooperating nodes is unaffected. In one magazine, the DSC can be expanded with up to 10 traffic blades handling up to 300k TPS, and for very high capacity needs, the DSC can comprise several magazines.

Because individual DSC blades are not visible to neighboring network nodes, they can be added, taken out of service or removed without any impact on the network. In this way, software and hardware upgrades and maintenance can be performed during normal operation without creating disturbances.

Conclusion

The emergence of the Diameter protocol as a fundamental part of network signaling in the EPC and IMS has created the need for a signaling-controller network element that facilitates configuration and increases the robustness of the network. This network element, the DSC, is mission-critical and requires telecom-grade software and hardware. A high degree of flexibility will also be necessary to manage the evolution of Diameter signaling during the coming decade.

Acknowledgements

Sven Steinacker, Martin Johansson, Volker Kleinfeld and Antonio Alonso

References

1. IETF, 2003, RFC 3588, Diameter Base Protocol, Available at: http://www.ietf.org/rfc/rfc3588.txt
2. Ericsson, 2012, Traffic and market data report, Available at: http://www.ericsson.com/res/docs/2012/traffic_and_market_report_june_2012.pdf

Jörg Ewert

joined Ericsson in 1999 as a system manager for mobile softswitch O&M. He later became the leader of the O&M system design team. In 2005, he joined the product management team at Business Unit Networks, where he was responsible for network management and security. He is now responsible for network evolution in product line switching at BU Networks. Ewert studied business administration at the University of Hagen, Germany, and holds an M.Sc. and a Ph.D. in physics from the University of Göttingen, Germany.

Lennart Norell

joined Ellemtel in 1977 to work on the development of the AXE system. He moved to Ericsson in 1982 and has since held various management and technical expert positions within system and product management in the telecom and datacom product areas. He has been active in the design of IMS, where he previously led the strategic systems management unit. He is currently an expert in IMS and core network architectures at Business Unit Networks, focusing on voice and video over LTE and network signaling. He has an M.Sc. in electrical engineering from Chalmers University of Technology, Sweden.

Soner Yamen

joined Ericsson in 1995 as a support engineer. In 1999, after working as a solution and project manager in Turkey, he transferred to system management at Ericsson Eurolab, Germany. In 2000, he joined the product management team, where he is currently responsible for Network Architecture in product line switching at Business Unit Networks. He holds an M.Sc. in electrical engineering from the University of Wisconsin-Madison in the US.