Optimizing Diameter Signaling Networks | Heavy Reading white paper

Optimizing Diameter Signaling Networks | Heavy Reading white paper

White Paper

Introduction:The Signaling Plane Renaissance

This white paper examines the challenges and assesses the architectural alternatives for deploying next-generation, Diameter-based signaling.

Signaling systems have always been a vital component of telecom networks. However, with the formal separation of signaling and network bearer through the standardization of "out of band" Signaling System #7 (SS7) protocols in 1980s, a quantum leap was achieved. Not surprisingly, some 30 years later SS7 still remains an industry stalwart for TDM-based fixed and 2G mobile networks supporting Intelligent Network (IN) services.

But as IP networks become commonplace and TDM declines, new open standards based protocols capable of supporting IP services are required. As a result, we now see a renewed focus - a renaissance of sorts on the signaling plane-driven by this shift from TDM services to IP links for IP and Session Initiation Protocol (SIP) based services.

Diameter Next-Gen Network Configurations

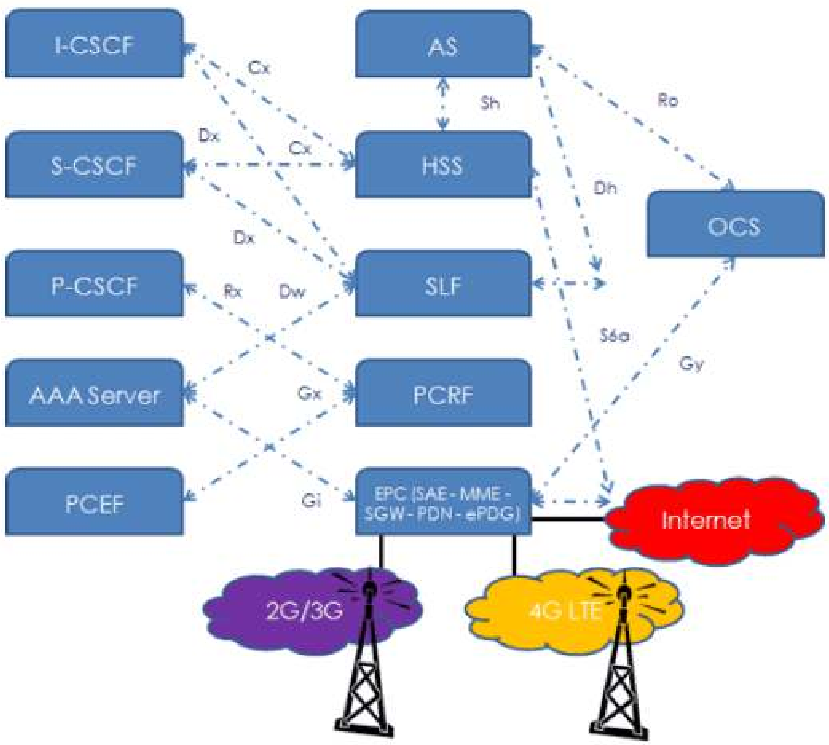
In this section of the white paper, we discuss in greater detail how Diameter nodes are evolving to fulfill next-gen networks signaling requirements and support the massive signaling volume triggered by today's smartphones and applications.

The Next-Gen Signaling Plane

Given the unique attributes of IP networks, activity driving development of next-gen signaling systems started more than a decade ago and ultimately resulted in the completion of the Diameter specification: Internet Engineering Task Force (IETF) RFC 3588.

Since then, Diameter has steadily gained industry-wide acceptance in standards most notably in Release 7 and 8 of the 3rd Generation Partnership Project (3GPP) IP Multimedia Subsystem (IMS) specification. And given the extensive number of interfaces required in core and access IP networks, as per **Figure 1**, this means Diameter will only increase in relevance as IP networks deployments continue to ramp. Currently, there are approximately 50 3GPP and 3GPP2 defined interfaces that utilize Diameter signaling.

Figure 1: Diameter Interface Summary



Source: Heavy Reading / 3GPP

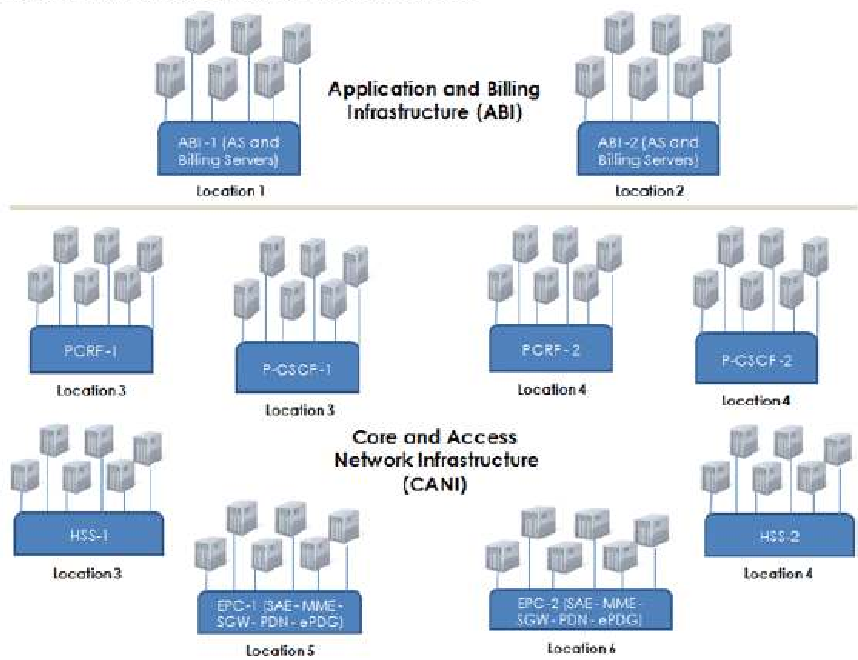
The Impact of 4G Network Distribution

Long Term Evolution (LTE), in many respects, reflects a risk and reward scenario-substantial new revenue potential-but since services will require the implementation of the reference architecture points noted above (e.g., PCRF), consideration must be given to network design to ensure access to services, scaling sites and handling node failure.

Specifically, there are three distinct set of challenges that must be considered.

First, in order to scale Diameter endpoints, the typical approach is to simply add server computing resources on a site level (as per **Figure 2**). The downside of this approach is that each server requires its own link and address scheme for routing purposes.

Figure 2: Distributed Server & Distributed Network Architecture



Source: Heavy Reading

Secondly, the highly distributed nature of next-generation networks must be also factored. In order to document the specific challenges this introduces we have defined two generic types of network sites in this white paper. They are:

- Application and Billing Infrastructure (ABI)
- Core and Access Network Infrastructure (CANI)

The ABI group includes application servers and billing applic a tions. By nature, these tend to be major sites, heavily data-centric but fewer in number. Still, these sites must be geographically distributed to support IT redundancy requirements. As a result, a significant amount of signal exchange with potentially long distances may result.

In addition, the impact of CANI sites must be considered. As the name suggests, most core and access infrastructure - including EPC (SAE, MME, SGW, PDN and ePDG), PCRF, HSS and CSCFs-fall into this grouping.

Since these sites are even more numerous in nature to support redundancy and match subscriber penetration levels by market, an even greater potential exists to overcome signaling networks.

Still, regardless of whether a billing server or a PCRF failure occurs, in all cases both ABI and CANI sites must be able to recover from these failures in real time by rerouting signaling quests to alternate nodes that can be problematic using a **peer-to-peer** connection model.

A final area of apprehension is server resource optimization. Given the cost of both ABI and CANI server infrastructure, network operators must continue to ensure all servers are optimally utilized to reduce opex. This most common approach is to implement a load balancer to route signaling to underutilized servers, leveraging the same base software intelligence used for failure rerouting.

As a result of these concerns, standards development defined the creation of a standalone Diameter Routing Agent (DRA) in 3GPP to support routing Diameter signaling to several nodes such as CSCFs, PCRFs, HSS, and EPC (including MME).

Originally, this functionality was envisioned to be performed by the PCRF. And while this is still a valid implementation option, definition of the DRA is seen as representing a less complex approach for meeting the challenges. The DRA supports several agent capabilities:

- **Relay Agent:** Supports basic Diameter message forwarding - no message modification or inspection function.
- **Proxy Agent:** Supports the ability to inspect and modify Diameter messages - and therefore can be utilized to invoke policy control rules.
- **Redirect Agent:** Stores generic routing information for DRA nodes to query. This ensures individual node routing tables are minimized.

Intra-Network Signaling & Routing Dynamics

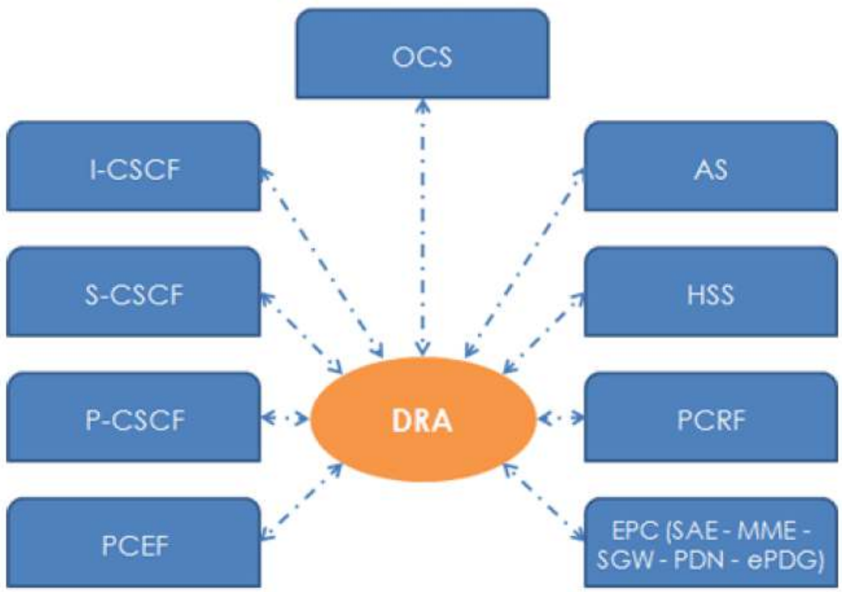
3G has fundamentally changed the service mix for network operators, and 4G will undoubtedly have even greater impact as adoption of complex session-driven broadband services increase.

Therefore, network operators are increasingly concerned that the changes in traffic patterns originating from smart devices could create signaling "bottlenecks" on the Diameter interfaces discussed above, ultimately resulting in network-wide signaling failures.

The most recent example is that of NTT DoCoMo, which suffered a major network outage of approximately four hours on January 25, 2012, that was directly related to an abnormal peak in signaling traffic.

Consequently, interest in DRAs continues to grow. Essentially, by deploying a highly scalable DRA in a centralized architecture vs. peer-to-peer connections, as shown in **Figure 3**, it's possible to load balance signaling, perform session setup, handle failure rerouting and support centralized routing updates.

Figure 3: DRA Simplified Deployment Scenario



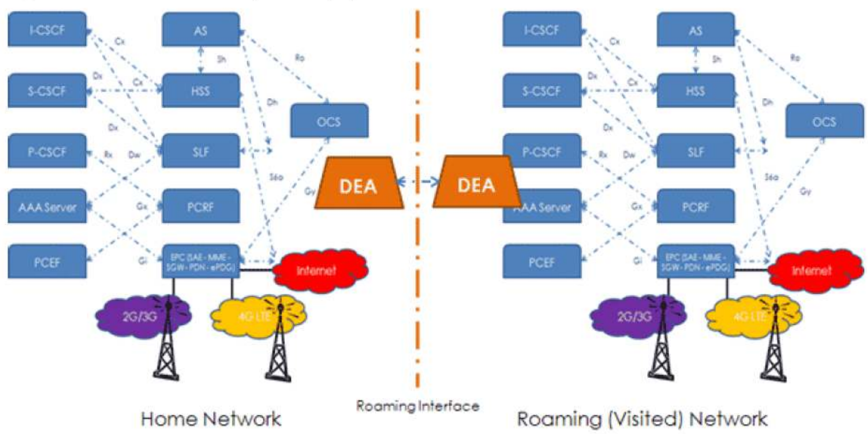
Source: Heavy Reading

Harkening back to the era of SS7, the D-link STP interconnection model was defined to meet the same scalability and routing challenges that A-link connections between peer-to-peer nodes in the same network (intra-network) would encounter.

Conversely, signaling and routing challenges must be considered on not only an intra-network basis, but also an inter-network basis to support roaming.

Therefore, as illustrated in **Figure 4**, the GSM Association (GSMA) defined the Diameter Edge Agent (DEA) functionality based on DRA to support roaming. Like a DRA, the DEA supports the ability to act as a network proxy, or simply a relay. As a result, even though DEA and DRA have unique network topology profiles, since they both support similar functionality, some vendors have developed multipurpose products that support both functions.

Figure 4: DEA Inter-Network Simplified Deployment Scenario



Source: Heavy Reading

Nevertheless, it's also important to note that Diameter products also must support a broad spectrum of 2G legacy protocol interworking to facilitate a graceful evolution path and roaming. For example, the Signaling Delivery Controller (SDC), a DRA and DEA compliant product from F5 Traffix Systems supports interworking with a full range of legacy protocol including Radius, LDAP, SS7 and 2G mobile GPRS Tunneling Protocol (GTP).

Quantifying the DRA Value Proposition

In this section of the white paper, we evaluate and quantify the value proposition of deploying a DRA to support a next-generation service such as voice over LTE (VoLTE).

VoLTE Signaling Implications

Since LTE was designed as end-to-end, IP-based network, one of the main challenges identified early on was how to most effectively support legacy circuit-switched (CS) voice services.

While solutions such as falling back to 2G or 3G networks may be implemented in some conditions, in 2010 the telecom industry reached broad consensus that the GSMA VoLTE implementation that is based on IMS would be adopted to ensure seamless roaming. The implications from a signaling perspective are wide-ranging. This includes handling of the message exchange across several interfaces, including HSS and MMEs, PCRF to enforce policy control and CSCFs to establish and maintain session control.

There are several implementation options for handling VoLTE signaling. These options are:

- 1. Peer-to-Peer Deployment
- 2. DRA–CANI Deployment
- 3. DRA–ABI Deployment
- 4. DRA–CANI and ABI Deployment

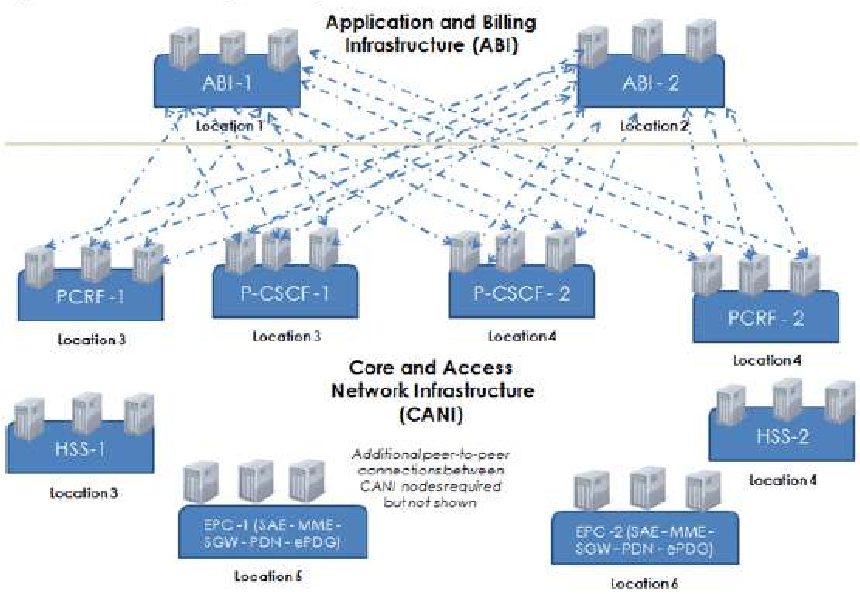
Below, we analyze each of these options in turn.

Peer-to-Peer Deployment

This approach is unique from the other options in that it does not leverage a DRA in any way. Rather, as per **Figure 5**, it utilizes peer-to-peer Diameter signaling interfaces between CANI and ABI sites. Key implementation attributes and characteristics of this approach include:

- **Failover:** Relies on dual homing between nodes on a standalone basis. Node failure is not broadcast to other nodes. Failover is done on a server to server level vs. utilizing pooled resources.
- **Scalability:** Relies on scalability of site nodes vs. pooling of resources.
- **Security and Authentication:** ABI and CANI nodes are visible to all other nodes and therefore susceptible to hacking.
- **Administration and Routing:** Changes to routing table is labor-intensive since routing tables of individual nodes typically must be updated to reflect any change to sites with which it connects. Trouble shooting is extremely complex and time consuming due to several connections

Figure 5: Peer-to-Peer Message Exchange



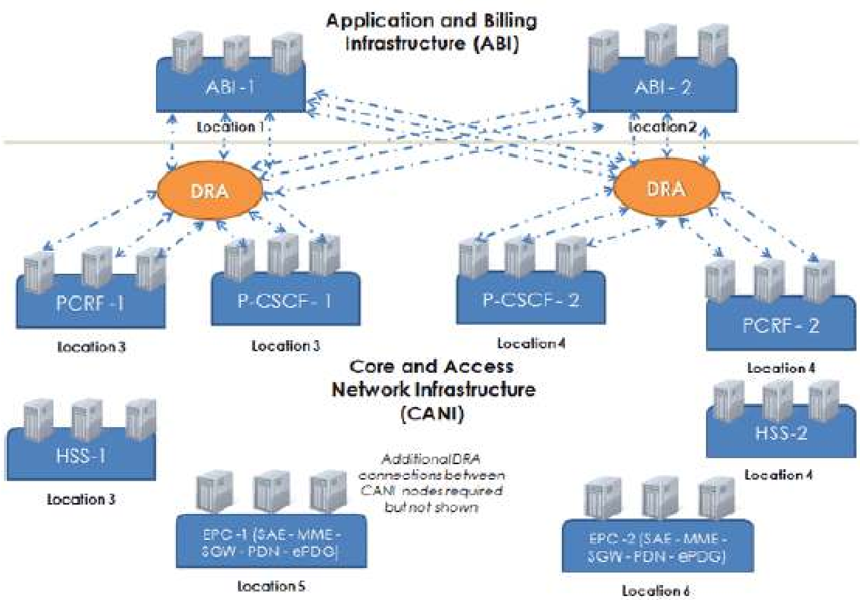
Source: Heavy Reading

DRA– CANI Deployment

In this scenario, as per **Figure 6**, DRAs are deployed to optimize CANI node signaling routing. ABI servers are not optimized. Key implementation attributes and characteristics of this approach include:

- **Failover:** Since all signaling traffic is routed via the DRA, and it receives health messages from each server, if a failure is encountered, additional DRA-based servers, regardless of location, can carry the load utilizing a predefined server failover order.
- **Scalability:** Utilizing DRA to pool server resources and as a load balancer also means that server capacity is optimized vs. the peer-to-peer model in which load balancing is not possible.
- **Security and Authentication:** Unlike the peer-to-peer model, the topology of CANI servers behind the DRA can be hidden to ABI servers and other networks/users.
- **Administration and Routing:** Changes to route management of CANI sites are transparent to ABI. CANI troubleshooting is simplified over peer-to-peer since fewer connections exist. In addition, since the DRA is a central focus for Diameter messaging processing and standards based, upgrading to a new 3GPP Diameter release is simplified vs. having to coordinate on a peer-to-peer level. In turn, this also permits network operators to deploy or overlay " best of breed " vendor solutions for components such as PCRF and HSS.

Figure 6: DRA – CANI Message Exchange



Source: Heavy Reading

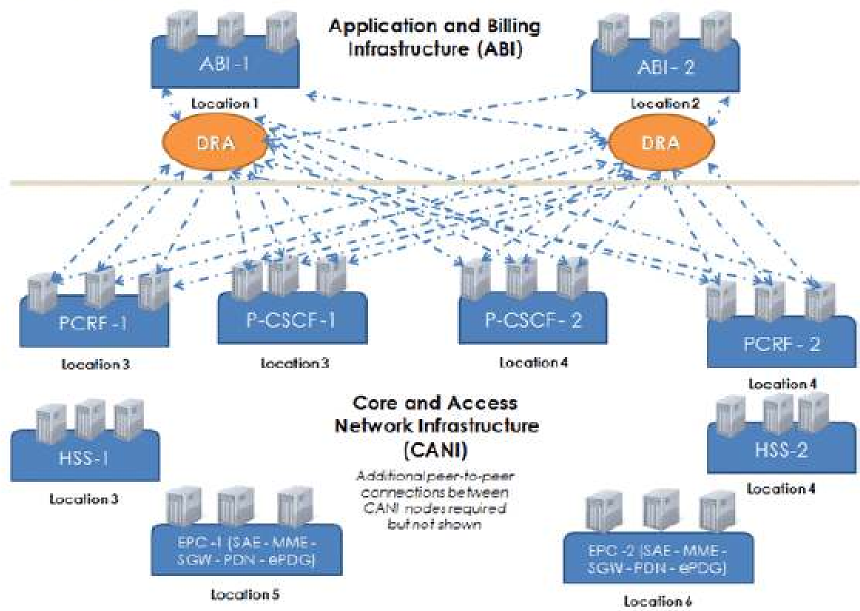
DRA–ABI Deployment

In this scenario, as per **Figure 7**, DRAs are deployed to optimize ABI node signaling routing. CAMI servers are not optimized.

Key implementation attributes and characteristics of this approach include:

- **Failover:** Since all signaling traffic is routed through the DRA, and it receives health messages from each server, if a failure is encountered, additional DRA-based servers regardless of location can carry the load employing a predefined server failover order.
- **Scalability:** Utilizing DRA to pool server resources and as a load balancer also means that server capacity is optimized vs. the peer-to-peer model in which load balancing is not possible.
- **Security and Authentication:** Unlike the peer-to-peer model, the topology of ABI servers behind the DRA can be hidden to CANI servers and other networks/users.
- **Administration and Routing:** Changes to route management of ABI sites are transparent to CANI. ABI trouble shooting is simplified and less time-consuming over peer-to-peer, since fewer connections exist.

Figure 7: DRA – ABI Message Exchange



Source: Heavy Reading

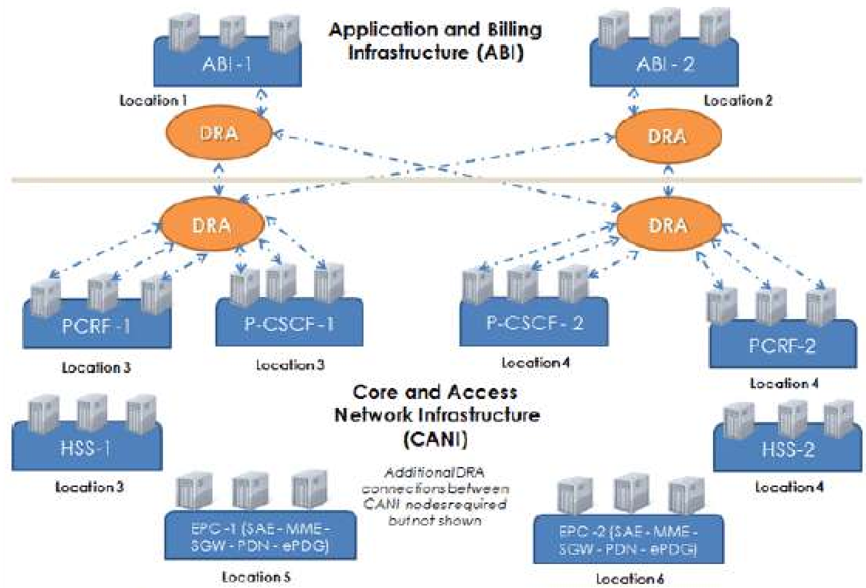
DRA-CANI & ABI Deployment

In this scenario, as per **Figure 8**, DRAs are deployed both in ABI and CANI sites to optimize signaling routing.

Key implementation attributes and characteristics of this approach include:

- **Failover:** Since all signaling traffic is routed through DRAs, failures in either domain are transparent to each other.
- **Scalability:** Utilizes a DRA to pool server resources in both domains. This not only reduces overall connections, but it also introduces several other routing options, including the ability to use load balancing between CANI and ABI nodes vs. simply within the individual domains.
- **Security and Authentication:** Topology of CANI and ABI sites are hidden.
- **Administration and Routing:** Any changes to routing are transparent to both ABI and CANI sites. Troubleshooting of ABI and CANI sites is simplified and less time-consuming over peer-to-peer, as well as the other two DRA deployment options, since a fewer number of connections exist.

Figure 8: DRA – CANI & ABI Message Exchange



Source: Heavy Reading

Evaluation Criteria	Peer-To-Peer	DRA–CANI	DRA–ABI	DRA–CANI & ABI
Failover	**Low** Server-to-server methodology	**Medium** CANI failures transparent to ABI	**Medium** ABI failures transparent to CANI	**High** CANI and ABI failures both transparent
Signaling Transport Scalability	**Low** Difficult	**Medium** CANI optimized; ABI not	**Medium** ABI optimized; CANI not	**High** CANI and ABI both optimized
Security & Authentication	**Low** All topologies visible	**Medium** CANI topology hidden; ABI visible	**Medium** ABI topology hidden; CANI visible	**High** CANI and ABI topologies both hidden
Administration & Routing	**Low** Upgrade all node approach	**Medium** CANI simplified; ABI servers unchanged	**Medium** ABI simplified; CANI servers unchanged	**High** CANI and ABI both simplified
Overall Value Proposition	**Low**	**Medium**	**Medium**	**High**

Figure 9: Implementation Options Side-by-Side Comparison Source: Heavy Reading

Conclusion & Summary

In many respects, the impacts of 4G all-IP-based services on next-gen signaling networks are only now starting to be understood. However, early experience has shown that these networks can be overcome by the amount of signaling resulting from smart devices and advanced services, even before the impact of roaming traffic is factored.

Furthermore, since 4G networks will need to be carefully engineered on an end-to-end basis to minimize investment, next-gen signaling networks by default will have to be highly scalable, reliable and cost-efficient. For that reason, although DRAs are not specifically required, given the scope of challenges currently identified, we believe DRAs have several advantages over a peer-to-peer approach.

In addition to providing a centralized point to support legacy network protocol interworking and roaming, the load balancing and routing capabilities ensure that next-gen signaling networks are cost efficient, scalable, reliable and aligned with the spirit of all IP networks to support extensible service models.

Appendix A: F5 Traffix Systems Solution

This Appendix provides an overview of F5 Traffix Systems ' Diameter based solution, the Signaling Delivery Controller.

Description	Benefits
The F5 Traffix SDC is a third-generation Diameter signaling solution that has unmatched product maturity in its three years as a commercial router and dozens of live deployments. As the market's only full Diameter routing solution combining 3GPP DRA, GSMA DEA and 3GPP IWF, the SDC platform goes far beyond industry standards' requirements. With unbeaten performance and ROI ratios of value/cost and capacity/footprint, it benefits operators' balance sheets as well as operational requirements. When operators deploy the Signaling Delivery Controller, they benefit from an "all-in-one platform" consisting of: Core Router with a DRA (Diameter Routing Agent) for failover management and efficiency, Edge Router with a DEA (Diameter Edge Agent) for roaming and interconnecting with security, Diameter Load Balancer for unlimited scalability enabling cost-effective growth, Diameter Gateway for seamless connectivity between all network elements, protocols, and interfaces to enable multi protocol routing and transformation, WideLens to benefit from network visibility for immediate identification and root cause analysis of network problems, capacity planning and providing KPIs to marketing, Network analytics for context-awareness and subscriber intelligence, Diameter testing tool for continual monitoring of network performance and operation.	Top-down, purpose-built architecture design for network wide Diameter signaling Enhanced signaling congestion and flow control and failover management mechanisms Enhanced signaling admission control, topology hiding and steering mechanisms Advanced context aware routing, based on any combination of AVPs and other dynamic parameters such as network health or time Advanced Session Binding capabilities beyond Gx/Rx binding Comprehensive testing tool suite for Diameter testing automation including stress and stability of all Diameter scenarios Supports all existing Diameter interfaces (50+) and seamlessly supports adding new ones Supports widest range of message oriented protocols for routing, load balancing and transformation (e.g., SS7, SIP, Radius, HTTP/SOAP, LDAP, GTP, JMS and others) Field proven, highly scalable solution with field proven linear scalability achieved via Active/Active deployment mode that exemplifies single node from connected peersperspective Supports SCTP, TCP, TLS, IP-Sec, IPv4, IPv6 Runs on off-the-shelf hardware

Figure A1: F5 Traffix Systems Diameter Solution-Signaling Delivery Controller (SDC) Source: F5 Traffix Systems

Appendix B: Background to This Paper

Original Research

This Heavy Reading white paper was commissioned by F5 Traffix Systems, but is based on independent research. The research and opinions expressed in this report are those of Heavy Reading, with the exception of the information in Appendix A provided by F5 Traffix Systems.

About the Author

Jim Hodges

Senior Analyst, Heavy Reading

Jim Hodges has worked in telecommunications for more than 20 years, with experience in both marketing and technology roles. His primary areas of research coverage at Heavy Reading include softswitch, IMS, and application server architectures, protocols, environmental initiatives, subscriber data management and managed services.

Hodges joined Heavy Reading after nine years at Nortel Networks, where he tracked the VoIP and application server market landscape, most recently as a senior marketing manager. Other activities at Nortel included definition of media gateway network architectures and development of Wireless Intelligent Network (WIN) standards. Additional industry experience was gained with Bell Canada, where Hodges performed IN and SS7 planning, numbering administration, and definition of regulatory-based interconnection models.

Hodges is based in Ottawa and can be reached at hodges@heavyreading.com

F5 Networks, Inc.
401 Elliott Avenue West, Seattle, WA 98119
888-882-4447 www.f5.com

Americas
info@f5.com

Asia-Pacific
apacinfo@f5.com

Europe/Middle-East/Africa
emeainfo@f5.com

Japan
f5j-info@f5.com