Ericsson Review

The communications technology journal since 1924

2014 • 7

Communications as a cloud service: a new take on telecoms

July 22, 2014

ERICSSON

Communications as a cloud service: a new take on telecoms

Modern mobile networks are complex systems built with an increasingly broad variety of technologies to serve a wide base of devices that provide an ever-greater range of services. These developments create interesting business opportunities for operators. But they also bring challenges, as new technologies and new expectations need to be managed with the same staff and budget.

BART JELLEMA AND MARC VORWERK

Software as a service (SaaS) is a promising solution for overcoming the challenges of implementing and managing new network technologies. The SaaS approach can provide substantial savings in terms of cost and lead time, and create a new source of revenue for those adopting the role of service provider.

This article shares some of the technical and economical insights and know-how gained from a proof of concept study conducted at Ericsson to explore the implementation of VoLTE as a service.

Why a new take on telecoms?

Today's networks support several technology generations, from 2G to 4G, and as research for 5G is well underway, the next generation is on the commercial horizon. The types of devices connected to networks vary from feature phones to smartphones and tablets to the billions of new connected devices that are emerging to support applications like smart homes and connected vehicles. In short, this is a complex ecosystem based on constant development, which can be difficult to predict and consequently challenging to plan for and budget.

The introduction of 4G LTE networks, for example, brought with it a major overhaul of voice services in core networks – in the move from circuit-switched to IMS. For many, especially niche operators, this type of technology upgrade threatens to stretch organizational capabilities to the limit, even to the point where business profitability is at stake.

To counter this challenge, many operators have turned to Network Functions Virtualization (NFV). By placing core networks in large concentrated data centers, NFV is a way to rationalize and simplify operations as well as speeding up innovation cycles[1]. The addition of multi-tenancy capabilities to NFV makes this approach particularly interesting for global operators, who have a presence in several countries and manage a range of networks through various operating companies.

Apart from addressing the strain on internal resources, NFV opens up the opportunity for operators to provide services, like VoLTE, to other communication service providers. By deploying the necessary IMS network functions for services in a central virtualized data center, and by adopting a SaaS model, operators can unlock the potential of their infrastructure beyond their own portfolios. Virtualized services can then be offered to smaller second and third tier affiliates or MVNOs at a lower cost, with reduced risk, and within a shorter time frame than is normally associated with the introduction of new services using traditional telecom business models.

The SaaS business model allows an operator's partners to circumvent lengthy hardware procurement cycles. This way, the burden of costs and complexities associated with owning a completely new and technologically advanced communications system can be removed. Simply by signing up as a tenant to the existing facilities of a host operator's data center, partners will be able to provide services quickly and cost-efficiently.

Once in place, NFV provides a flexible telecom-grade platform on which a variety of communication services can be offered to people and organizations, in a low-cost, low-impact fashion. Services can be quickly and easily trialed, launched, scaled up or down and decommissioned in line with market demand,

BOX A Terms and abbreviations

ARPU	average revenue per user
CRM	customer relationship management
CSCF	Call Session Control Function
HSS	Home Subscriber Server
IMS	IP Multimedia Subystem
LI	Lawful Interception
MRFP	Media Resource Function Processor
MSC	mobile switching center
MTAS	Multimedia Telephony Application Server
MVNO	mobile virtual network operator
NFV	Network Functions Virtualization
NPV	net present value
O&M	operations and maintenance
OSS	operations support systems
OVF	Open Virtualization Format
P-CSCF	proxy call session control function
SaaS	software as a service
SBG	Session Border Gateway
SLA	Service Level Agreement
SRVCC	single radio voice call continuity
TCO	total cost of ownership
VLAN	virtual local area network
VM	virtual machine
VoLTE	voice over LTE

presenting an operator-branded and guaranteed alternative to the many third-party over-the-top solutions that operate in both the consumer and enterprise communication space.

Concept – heading for the clouds

Today, the purchasing process for a new IMS system can take several months from order placement to an operational system. Once an order is placed, the network system vendor initiates the production process for the node. On completion, the node is then integrated and packaged together with the necessary software elements, tested, shipped, installed at the designated central office site, integrated into the network, tested again, accepted and finally put into operation. Once the system is functional the operator is responsible for operations and maintenance (O&M), often with the support of the vendor.

With a SaaS deployment, operators can purchase a virtualized IMS network slice that is custom-initialized for them in a large data center. Network slices can be tied into existing radio and packet core networks over a remote link – as Figure 1 illustrates.

Working in this way, operators will no longer need to purchase, install or own any hardware, or invest in training staff on a new system. The SaaS approach removes the need to manage software licenses, and reduces system integration from a complete IMS solution to just the points of interconnect with the access network. Ownership and operational details are instead taken care of by the service provider and operators will pay as they go using simple, predictable price models, such as a flat service fee per subscriber. The benefits: no large upfront investments, limited technical and business risks, and much shorter time to revenue.

VoLTE as a service

In 2013, Ericsson's R&D and IT divisions carried out a joint project to develop a proof of concept implementation for VoLTE as a service. The objective was to gain an understanding of the technical and economic implications of offering a complex communications solution like VoLTE as a service. For telecom applications, SaaS is a relatively new business model that needs to »

FIGURE 1 The SaaS concept

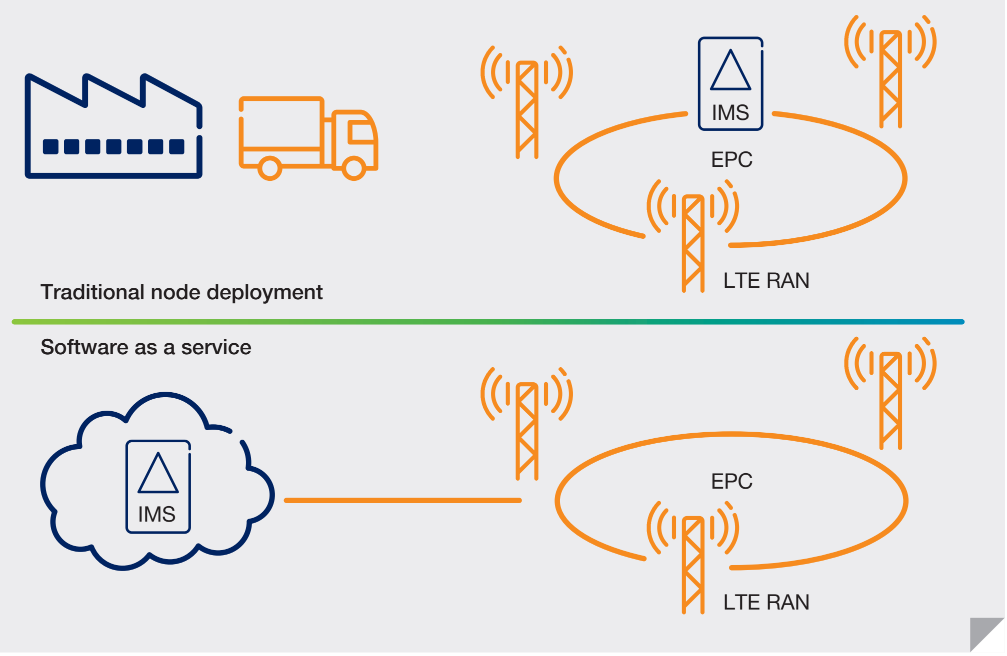


FIGURE 2 VoLTE as a service – architecture

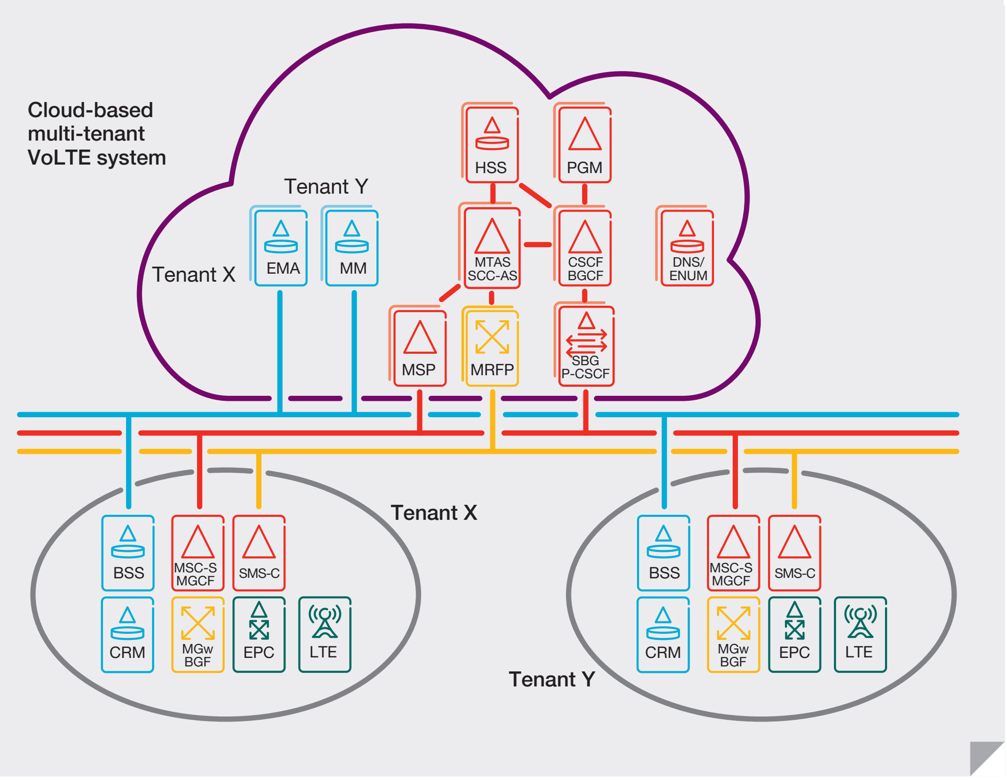


FIGURE 3 Network Functions Virtualization – portfolio migration

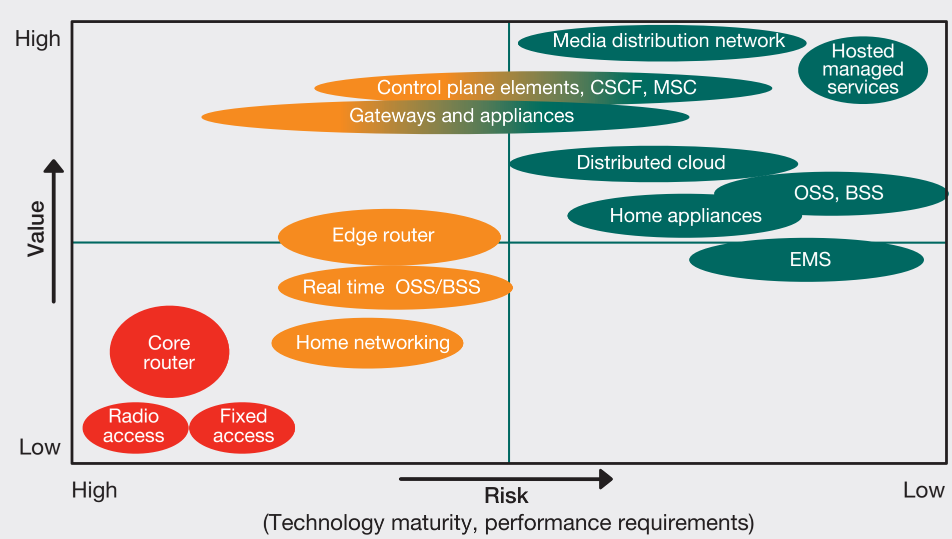
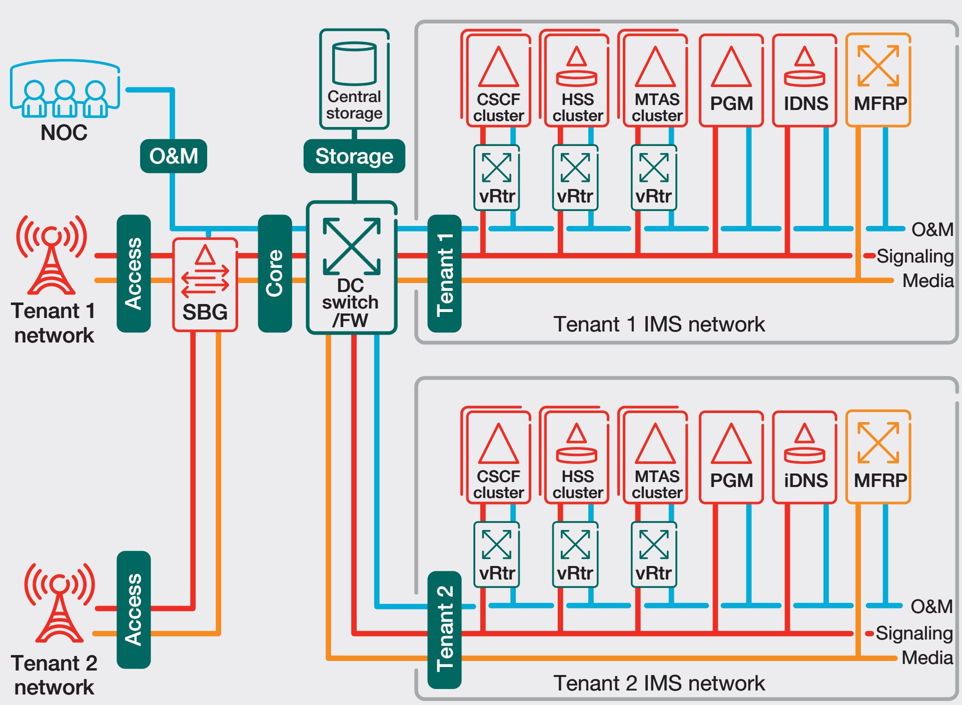


FIGURE 4 IP design



take into consideration the tough requirements of the underlying cloud infrastructure.

From the start of the project, it was clear that turning VoLTE as a service into a viable business proposition, with competitive price levels and sound margins, would require the onboarding and serving of new tenants to be simple, efficient and easily repeated.

Through virtualization techniques, the hosting service provider can deploy multiple VoLTE systems on the same shared data center hardware, while still guaranteeing each tenant their own dedicated, logically separated virtual network. Such a multi-tenant cloud infrastructure makes it possible for service providers not only to share hardware among tenants, but also O&M and engineering staff. The resulting economy of scale is much more significant than any individual small-scale installation could achieve.

To improve repeatability, a high degree of business process automation (auto-deployment and auto-scaling) reduces the time and effort needed to operate services, which in turn reduces costs. And to ensure that customers get what they pay for, the provision of relevant network statistics is essential for billing and to provide proof of Service Level Agreement (SLA) conformance.

A blueprint for the architecture

So how is this done? As shown in Figure 2, the operator's radio and packet core networks as well as their legacy circuit-switched network are connected to a remote virtualized IMS network within a cloud data center over standardized interfaces for signaling, O&M and media.

As illustrated in Figure 3, next generation systems will normally be fully implemented as software without any strong hardware dependencies. Consequently, IMS server-type network functions like CSCF and MTAS are natural candidates for cloud placement.

To optimize use of bandwidth, most media handling will most likely continue to take place in the tenant network, with the possible exception of the MRFP. Certain network functions, such as HSS, can be placed either in the cloud or in the tenant network, depending on operator preference or to comply with local regulatory requirements with respect to user databases.

To integrate with the operator's various business support, customer care and other IT systems, the virtualized IMS network will provide billing and provisioning capabilities.

When another operator becomes a tenant, a copy of the virtualized IMS network can be instantiated in the data center and the whole onboarding process simply repeated.

For commercial deployment, at least two data center locations are needed to provide geo-redundancy. Alternatively, the tenant could operate a single non-redundant system in their own network and rely on a secondary virtualized system as an overflow and failover mechanism – geo-redundancy as a service.

As an additional offering, service providers can include smaller regional satellite sites that host the IMS media plane nodes. In such topologies, the satellite centers can be used to house not just media gateways but also network functions like Lawful Interception for IMS (LI-IMS), an anchor MSC for SRVCC and /or an SBG/P-CSCF. Providing media-plane nodes in this way reduces the impact of introducing IMS to an operator's existing core network to practically nothing. Taking a coverage area the size of North America as an example, approximately 24 regional sites would be required to provide this service.

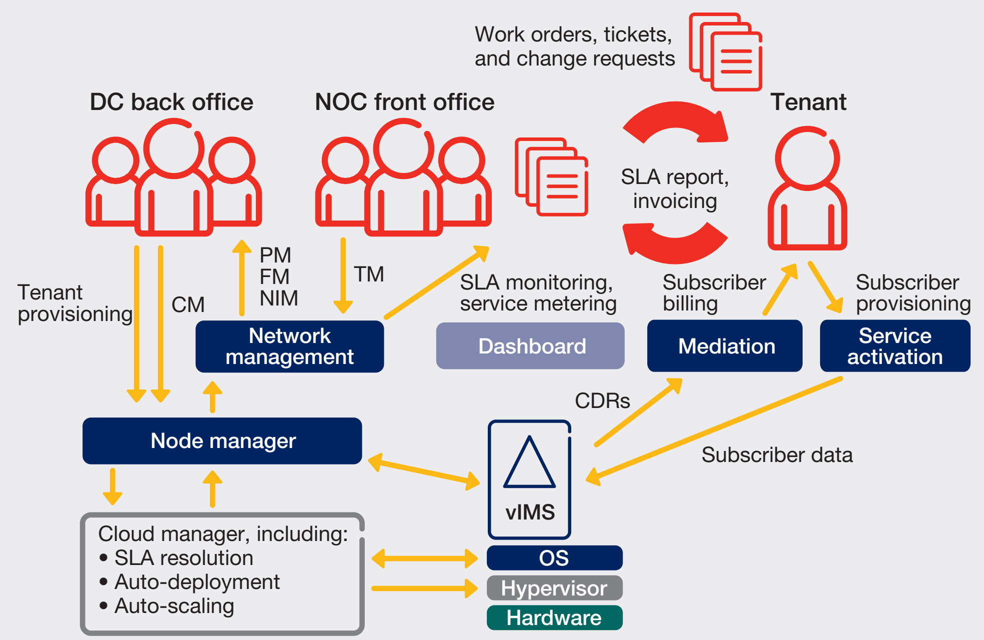
A significant change

By the end of December 2015, roaming fees within the European Union will no longer exist; rates for voice calls and data transmission will be the same as in the subscriber's home market[2]. This drastic change for consumers is likely to stimulate traffic and motivate operators across Europe to centralize their core network infrastructures – as physical location will no longer influence billing rates.

Hardware

From a hardware perspective, data centers will need to be equipped with enough servers to host virtualized versions of the number of tenant IMS networks anticipated. In addition, high capacity physical IP switches and central storage will be needed. As hardware is completely decoupled from software through virtualization middleware, service providers have the freedom to select the x86-based hardware of their choice, as long as it meets the set target specifications in terms of performance, bandwidth and memory of the virtualized network functions – including some virtualization overhead.

FIGURE 5 Operations and maintenance view



BOX B

Legend for Figure 5

CM – configuration management
DC – data center
FM – fault management
NIM – network inventory management
NOC – network operations center
PM – performance management
TM – ticket management

Operations and maintenance

As shown in Figure 4, the IP plan needs to be designed so that each tenant has their own set of dedicated VLANs – at least for O&M, signaling and media – that are separated from all the other tenants to avoid interference and maintain security.

For O&M, the service provider's back office can perform tasks such as configuration management, performance management, fault management and network inventory management through a managed services portal. This is similar to the way network management works in the service provider's own IMS network. The front office can process work orders and change requests received from tenants, tickets from field engineers, and take care of invoicing and SLA reporting.

As shown in Figure 5, tenants will be provided with O&M access rights to their specific network for provisioning subscribers and retrieving detail records for charging. This access will connect to the tenant's back-end IT systems like CRM and billing systems, and a dashboard function will allow the tenant to view key performance statistics for their network.

Northbound interfaces from the different virtualized network functions are generally not affected by virtualization.

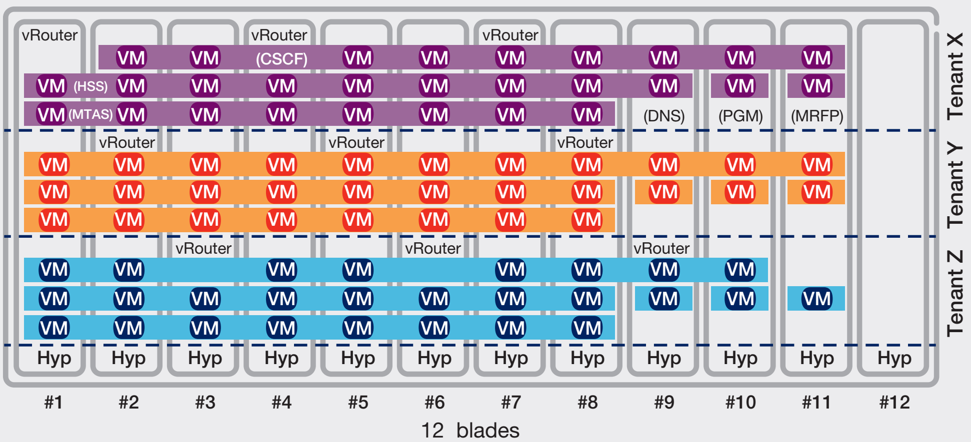
The exact implementation of the IMS network – its internal structure, what software and which release is used – is entirely at the discretion of the service provider. In other words, the implementation is transparent and of no real concern to the tenant. Their only concern lies with the behavior of the service, the agreed service level and the interfaces exposed at the points of interconnection.

Features – under the hood

Multi-tenancy

Modern server blades house multiple processor cores on which virtual »

FIGURE 6 Multi-tenancy



machines (VMs) can be placed. Virtualized IMS network functions like CSCF or HSS use a number of virtual machines for traffic processing, as shown in Figure 6, which act much like a physical node with several blades as part of a cluster. As illustrated, these virtual machines should be spread horizontally over multiple blades, so that the failure of one blade will never bring down an entire node.

The remaining cores can then be used for other network functions or even other tenants.

Auto-deployment

Onboarding a new tenant sets a deployment function into motion. As shown in Figure 7, this function executes an IMS network deployment sequence using a cloud orchestration tool in combination with scripts that parse the customer-specific environment settings. Any necessary adaptations are executed inside the deployed VMs.

To save time during the onboarding process, tenant VLANs can and should be prepared ahead of time. Software images for each virtualized network function are built and uploaded (in advance) to the cloud manager in, for example, Open Virtualization Format (OVF)[3], and are kept in storage. From there, the deployment function can instantly clone network functions for new tenants.

To connect them to their pre-assigned VLANs, the virtual machines are linked to the appropriate port groups and powered on. The deployment function loads a data transcript onto the VMs to create an operational virtualized network function and configures the application interfaces, so that they form an integrated IMS network. All of this post-configuration work can be scripted; and any data transcript common to all tenants can be included in the software image.

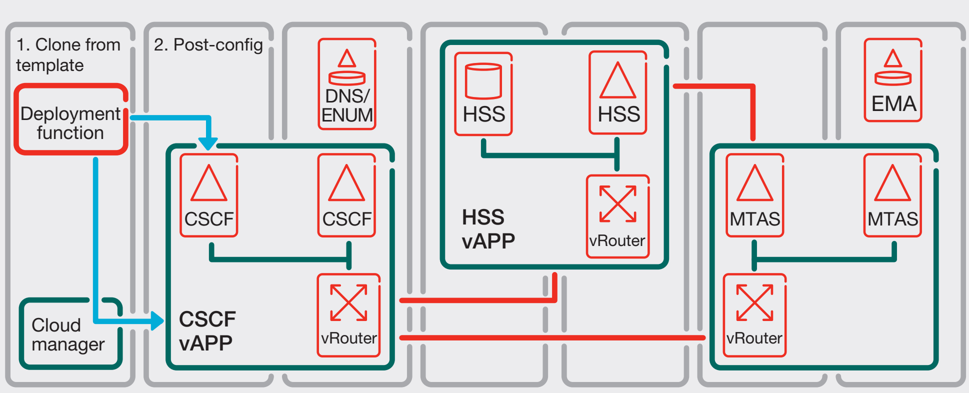
Once all the network functions and connections between them are established, the next step is to connect the virtual IMS network to the tenant's access network and IT systems before provisioning the first users. The high degree of preparation and process automation, together with the use of hardware capacity available in the data center, and prestorage of software images, results in drastically reduced installation times. The complete software installation for an IMS network can be fulfilled in just a few hours, compared with the several days it would normally take to set up a traditional central office environment with physical nodes. Time to revenue from contract signing to commercial launch could be reduced to a matter of weeks, rather than months.

Auto-scaling

The ability to scale networks is a key business enabler. In the proof of concept project, the Ericsson team developed a controller function that worked in conjunction with the cloud manager to determine when and where networks need to be scaled. As shown in Figure 8, the controller continuously monitors the average processor load on each of the virtualized network functions, by reading the load figures from the guest operating system. This approach has proven to be more accurate than using the measurements provided by the hypervisor, as the hypervisor cannot, for example, determine the priority and necessity of currently executed tasks from the outside.

When the load for a particular network function like CSCF exceeds its set upper limit, which can happen for example during traffic peaks, the controller requests the cloud manager to scale out. The cloud manager powers up another CSCF virtual machine,

FIGURE 7 Auto-deployment



which then joins the existing cluster and rebalances the traffic. Similarly, a node can be scaled in during periods of low traffic.

The user interface for the controller allows engineering staff in the data center to set upper and lower processor-load thresholds for scaling in and out of network functions. Additional parameters, such as the minimum and maximum number of traffic processors, can also be set so that a node has a guaranteed minimum redundancy without monopolizing more than its fair share of available resources.

Figure 9 shows an example time series taken from a test session carried out during the proof of concept project. The scaling mechanism for the CSCF kicked in just before the 12:58 time stamp, as the processor load exceeded the set maximum (indicated by the red line). Three minutes later, at approximately 13:01, the CSCF was running on a cluster of five instead of the original four traffic processors.

Depending on the existing traffic load, it takes between five and 10 minutes to add capacity automatically (by scaling a virtualized network function out by one traffic processor) to a live node in a virtualized data center. In contrast, adding a physical hardware board to a live physical node on such a time scale is unimaginable.

Service-level monitoring

SLAs are highly varied in nature, covering different aspects of a service such as customer ticket turn-around times and other logistical matters. As far as technical content is concerned, SLAs between service providers and tenants are best kept simple and transparent. Many network statistics can be made available for information purposes, which is fine, but the list of contracted KPIs that carry financial implications are best kept as simple as possible (see **Table 1**).

In its simplest form, billing tenants for the use of VoLTE as a service can be based on the actual number of active users during a given time period – assuming a certain maximum traffic volume. The volume can be defined in terms of the maximum number of simultaneous sessions (the current licensing model) or by average voice minutes per subscriber. As voice »

FIGURE 8 Auto-scaling

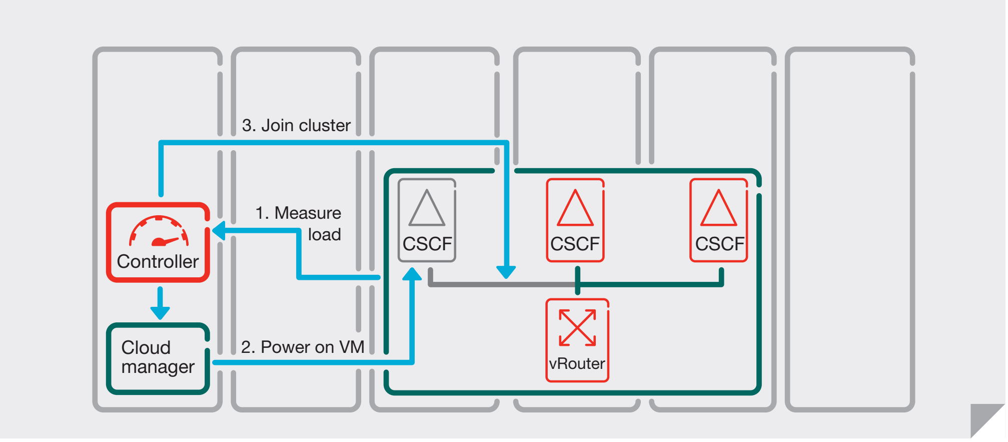


FIGURE 9 Example time series

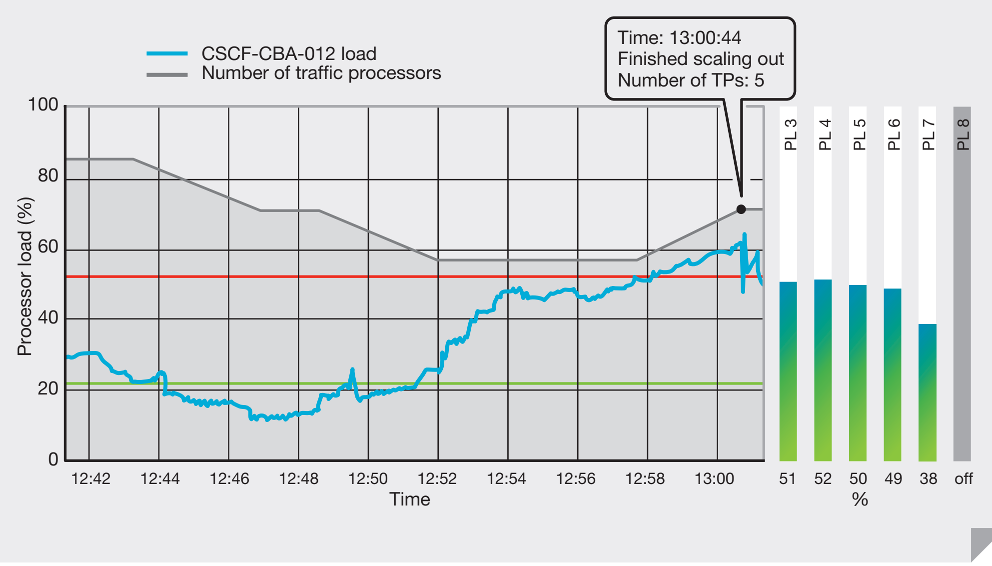
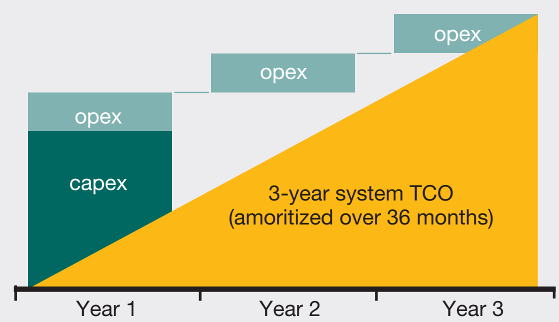


Table 1: SLA reporting

Service Metering	**Service Level Monitoring**
Tenant gets billed per number of users + premium for traffic coverage	Tenant gets credited in case of failure to meet SLA
Key performance indicators	**Key performance indicators**
Number of users	System availability (%)
	IMS registration time (msec)
Traffic volume (average session duration and/or number of concurrent sessions	IMS registration success ratio (%)
	VoLTE setup success ratio (%)

FIGURE 10 Pricing model



minutes readily translate to the payment plans offered by most operators, this model is probably preferable for the majority of tenants. Similar consumption indicators aligned with operator-to-consumer price models can be created for all other services.

While threshold limits are good for SLAs and planning, service providers are not likely to cut off traffic when an agreed maximum for a tenant is reached – as long as continued service does not overload the system or infringe on other tenants. However, a premium may be charged.

To keep service level monitoring relatively straightforward, the proof of concept project created example reports for system availability, registration time, registration success rate and call establishment success rate. If any of these resources underperformed during a billing period, the tenant would receive credit on their next payment.

All of these counters and statistics are already available in today's typical IMS products. By collecting, filtering and combining them into a customized business intelligence report, they can be easily communicated and turned into actionable data.

In a commercial setup, this data would be fed from the OSS into a specialized SLA management tool, in which KPI values are continuously compared against predefined thresholds to detect and record SLA violations. A number of approach warning levels are usually defined below the critical level, so that O&M staff can be alerted and take appropriate actions before any impact on business is felt.

Table 2: TCO comparison – an example in USD thousands

system	capex	opex	as a service	capex	opex
Hardware	2,400	1,000	Setup fee	450	
Software	3,300	2,500	Service fee*		24,098
Systems Integration	1,600				
Project	900				
Staff		4,300			
Facilities		500			
Utilities		200			
Lab costs	5,000				
3 year TCO		21,700	**3 year TCO**		24,548
NPV		20,791	**NPV**		20,791

*36 months x 200,000 subscribers x USD 3.35

Financials – where is the money?

In the traditional system-sales model, total cost of ownership (TCO) is defined as the initial purchase price including related project costs, plus recurring running costs such as support agreements, O&M staff, rent and power. In the SaaS model, this will be replaced by a single line item – service fees – under opex. Unfortunately, estimating a reasonable price level for VoLTE as a service – one that the tenant can afford and that keeps the service provider in business – is not a simple task.

One potential pricing model (shown in **Figure 10**) is based on the traditional total cost of ownership for a three-year period, amortized over 36 equal monthly payments. Payback times of less than three years tend to result in a service that is too expensive for the tenant, and calculating over longer periods tends to make the model unattractive for service providers.

Parameters like operator size and running costs – rents, engineer salaries and electricity – vary greatly from one part of the world to another, and so the economy of scale and benefit to operators in different markets will vary. In conjunction with the proof of concept project, a study aimed to estimate the service price for VoLTE for a typical second or third tier operator with between 100,000 and one million subscribers.

The study estimated and adjusted for net present value (NPV) and the required initial capex and opex over three years to own, deploy and run an IMS system for VoLTE. The resulting estimation set the fee for VoLTE as a service to be somewhere between USD 1 and USD 5 per subscriber per month. An example of the type of calculation used in the study is given in **Table 2** for a mock tenant with 200,000 subscribers.

To match the price points for a service with the average cost per subscriber incurred by operators at different ends of the scale, some sort of tiered price model is needed – a suggested model is shown in **Figure 11**.

If the average revenue for voice services is assumed to be USD 40 per subscriber per month, a fee of USD 1-5 per subscriber per month for VoLTE as a service is between 2.5 and 12.5 percent of the corresponding ARPU it generates, which is a fair business case.

Looking at the addressable market, the number of subscribers connected to second and third tier operators amounts to 22 million in North America alone.

Evolution – beyond the horizon

As illustrated in **Figure 12**, rolling out VoLTE might be the initial motivation for a second or third tier operator to switch to the software as a service model. Doing so would allow such operators to roll out VoLTE in the same time frame (2014-2015) as their larger competitors – and secure their market share.

Subsequently, operators could broaden the scope of their offerings to include customized services for enterprises, the retail industry and many other verticals. The SaaS platform could be further utilized by opening it up to internet-application and web developers to create a whole new range of converged services.

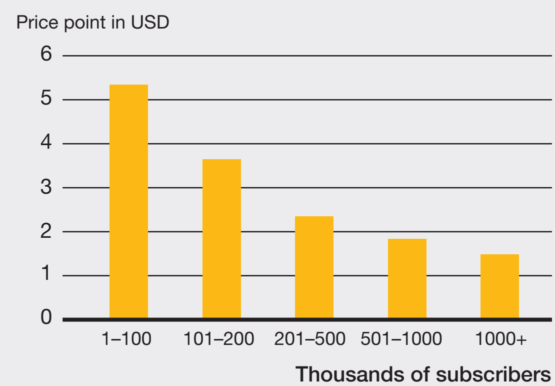
Second and third tier operators are the most obvious first adopters of this type of business model for voice – or rather VoLTE. Once rooted, adoption is likely to rise up the food chain. Many operators, both big and small, have opted for the managed service approach for their voice networks, gaining efficiency and freeing up resources to focus on customers and on improving operator brand value.

Operators already include unlimited voice and unlimited text in their data plans, rendering these services to the level of a commodity, or a fundamental product that cannot really be charged for, but neither can they be taken out of the service offering. And so software as a service – the ultimate form of a managed service – is the most natural evolution path.

Conclusion

By following this route, service providers will be able to offer all managed networks from the same platform and housed under the same roof. For operators, the ability to outsource the responsibility for voice shifts price pressure on to a third party who can provide the right expertise, efficiency and scale.

FIGURE 11 **Tiered price model**

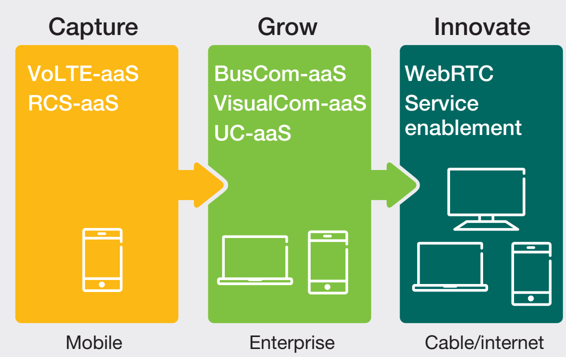


BOX C

Legend for Figure 12

aaS – as a service
BusCom – business communication
RCS – Rich Communication Suite
UC – unified communication
VisualCom – visual communication
WebRTC – Refers to standardization for real-time browser capabi ities.

FIGURE 12 **Service evolution**



References

1. Ericsson, February 2014, White Paper, The real-time cloud – combining cloud, NFV and service provider SDN, available at: http://www.ericsson.com/news/140220-the-real-time-cloud_244099438_c
2. European Commission, Digital agenda for Europe, Roaming, available at: https://ec.europa.eu/digital-agenda/en/roaming
3. DMTF, Open Virtualization Format, available at: http://www.dmtf.org/standards/ovf

Additional reading

ETSI, Network Functions Virtualisation, available at: http://www.etsi.org/technologies-clusters/technologies/nfv

Ericsson Review

To bring you the best of Ericsson's research world, our employees have been writing articles for Ericsson Review – our communications technology journal – since 1924. Today, Ericsson Review articles have a two-to-five year perspective and our objective is to provide you with up-to-date insights on how things are shaping up for the Networked Society.

Address:
Ericsson
SE-164 83 Stockholm, Sweden
Phone: +46 8 719 0000

Publishing:
Ericsson Review articles and additional material are published on: www.ericsson com/review. Use the RSS feed to stay informed of the latest updates.

Ericsson Technology Insights
All Ericsson Review articles are available on the Ericsson Technology Insights app available for Android and iOS devices. The ink for your device is on the Ericsson Review website:www.ericsson.com/review. If you are viewing this digitally, you can:
download from Google Play or
download from the App Store

Publisher: U f Ewaldsson

Editor:
Deirdre P. Doyle
deirdre doyle@jgcommunication.se

Subeditors:
John Ambrose, Nathan Hegedus and Ian Nicholson

Art director and layout:
Carola Pilarz

Illustrations:
Claes-Göran Andersson

ISSN: 0014-0171

Volume: 91, 2014

Marc Vorwerk

joined Ericsson in 2000. Today, he is a senior specialist for cloud computing, and has previously worked on multi-access, IMS and media-plane management research – developing early prototypes and participating in European research projects. He began utilizing virtualization and cloud over six years ago, and has been an evangelist within Ericsson to promote the benefits of these technologies. Today as a senior specialist he is a team leader, an innovation event presenter and provides customer-engagement support. He holds an M.Sc. in electrical engineering from RWTH Aachen University, Germany.

Bart Jellema

joined Ericsson in 1989. He has held several system and product management roles in Canada, Germany and the Netherlands. He currently works with the core networks architecture and technology team in the area of cloud and NFV, and is involved in the establishment of Ericsson's new global ICT centers. He has been active in standardization, holds several patents and is a speaker for Ericsson at innovation events. He holds a B.Sc. in electrical engineering from the University of Applied Sciences, Eindhoven, the Netherlands.

Ericsson
SE-164 83 Stockholm, Sweden
Phone: + 46 10 719 0000

ISSN 0014-0171
284 23-3227 | Uen