Independent market research and competitive analysis of next-generation business and technology solutions for service providers and vendors

HEAVY READING WHITE PAPER

Automation in the 4G/5G Mobile Core

A Heavy Reading white paper produced for Cisco Systems Inc.

CISCO

AUTHOR: GABRIEL BROWN, PRINCIPAL ANALYST, HEAVY READING

AUTOMATION, AGILITY, INNOVATION

Automation is viewed through the prism of efficiency and a lower cost of production – and rightly so: Mobile network economics demand a radically reduced cost model. Advanced technologies, such as LTE-Advanced Pro and 5G, offer enormous capacity gains and superior end-user performance. This is driving a much lower cost per bit on transport and is positive for the industry, but must also be matched by lower cost operations.

Platform economics are fundamental to enhancing the feature-set an operator can deploy across its network footprint and its ability to extend into new service domains and customer types. It is not only platform efficiency, however, that is driving operators toward automation. Equally as important is the role automation can play in service creation, management and scaling. Better operational processes translate directly to more agile service delivery and can help make networking as easy to buy and consume as cloud services.

The packet core controls connectivity, session management, mobility, quality of service, policy and more in a mobile network, and is therefore critical to how operators should approach automation of network resources *and* network services. In cloud-native core networks, where functionality is distributed across microservice instances, the need (and opportunity) for automation is even greater. This white paper argues that automation in the 4G/5G mobile core will generate not only large gains in operating efficiency, but will also fundamentally change how customers engage with and consume mobile network services. This will expand operators' addressable markets and underpin their economic future.

Static Networks, Static Services (The Problem Statement)

The problem statement for today's networks is relatively simple: The current operating model is overly dependent on manual configuration and manual oversight and intervention, thus making it rigid in the face of change. Because mobile networks are nationally critical, infrastructure operators are reluctant to reconfigure networks without extensive modeling and testing, which raises the business case hurdle for new services, reducing the propensity to experiment and pursue new opportunities. This leads to "network ossification" with operators finding themselves unwilling, and unwittingly, entrenched against innovation and less able to respond to market disruption. The result is that the addressable market is artificially limited and operators are not able to fully exploit opportunities in the fast-moving Internet economy.

In the core network, virtualization, NFV and the cloud are one of the responses to this problem. The intent is to make it less expensive, easier, and faster to deploy network functions and compose them into end-to-end services. The problem is that NFV has, so far, been characterized by manual processes developed through ad hoc experimentation. For example, on-boarding a VNF – in principle, the first and simplest task in NFV – requires an operator to have a detailed knowledge of the application requirements, the middleware (VM or container – which variety?) and server configuration (CPU type, firmware version, NIC cards, etc.) to achieve acceptable performance. This requires specialist knowledge and varies across vendors, software versions and hardware suppliers. In short, NFV as currently constructed is restricted by too much craft and not enough automation.

Service-Centric Automation (The Opportunity)

The opportunity for mobile operators is to use cloud principles to abstract services from infrastructure. The mobile core, which controls mobile network services, is well-suited to

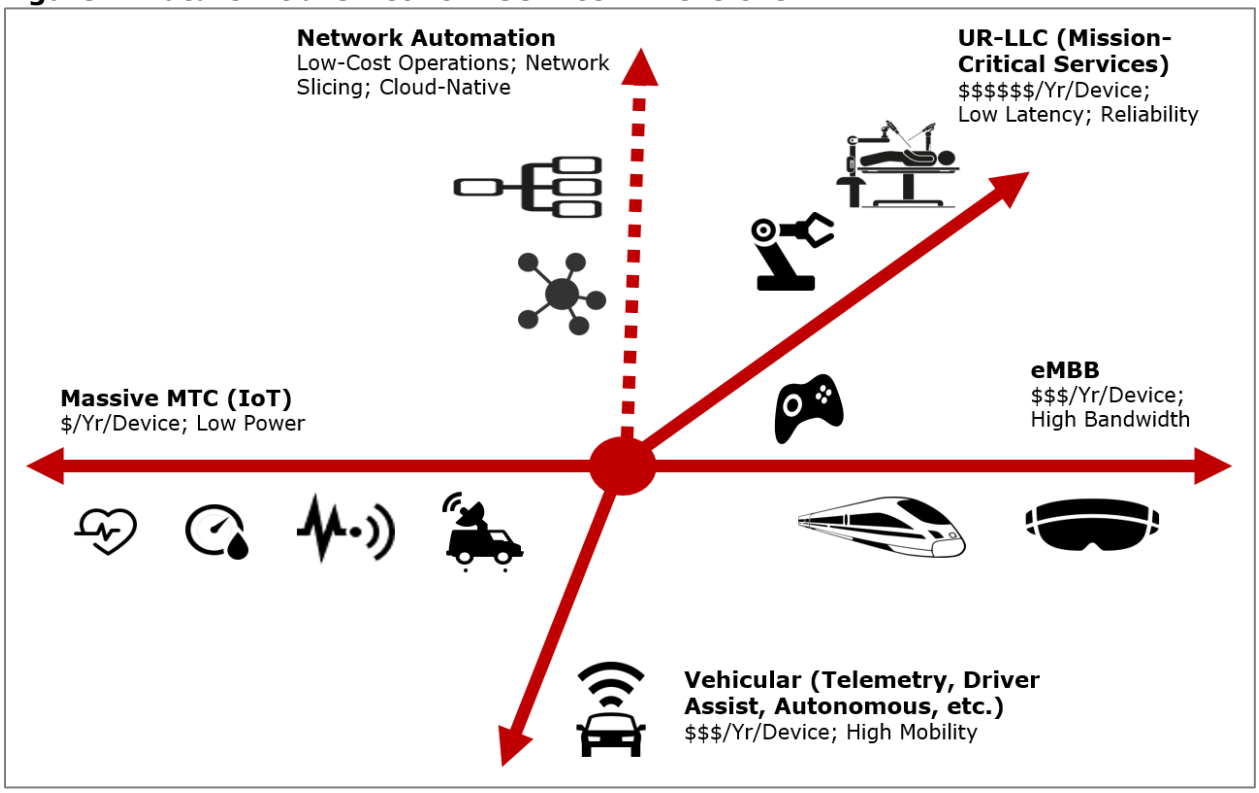
cloud infrastructure because its component functions are already software-centric and operate above the IP transport layer. By decoupling the creation and management of services, operators gain the ability to more quickly address customer demand.

The ability to automate mobile core processes using a "Lego-block" style composition of services is dependent on tools that map network services to infrastructure resources and ensure that services and users are aligned and optimally located. This will minimize the "craft" that characterizes today's NFV deployments and ultimately extend to how customers interact with networks to configure and consume services. In this sense, the opportunity is to harness end-user demand to drive automation into service providers.

LTE-A Pro & 5G Service Dimensions

One reason why automation is critical in the near-term is the imminent deployment of 5G and advanced 4G networks with capabilities that can radically extend the operator's reach into new industries and new service types. These new markets require a new cost model and new level of flexibility that, if addressed, can open up massive opportunities for mobile operators. **Figure 1** shows service dimensions identified for 5G with example service types.

Figure 1: Future Mobile Network Service Dimensions



Source: Heavy Reading, based on 3GPP SMARTER

New service types identified by the 3GPP and other industry players include cellular V2X, industrial verticals, massive-scale IoT, private networks, production-critical robotics, remote experts and virtual reality. The 3GPP work now also includes a dedicated focus network operations, which includes management of network slices using automation. On automation, however, the industry will need to contribute above and beyond the base standards.

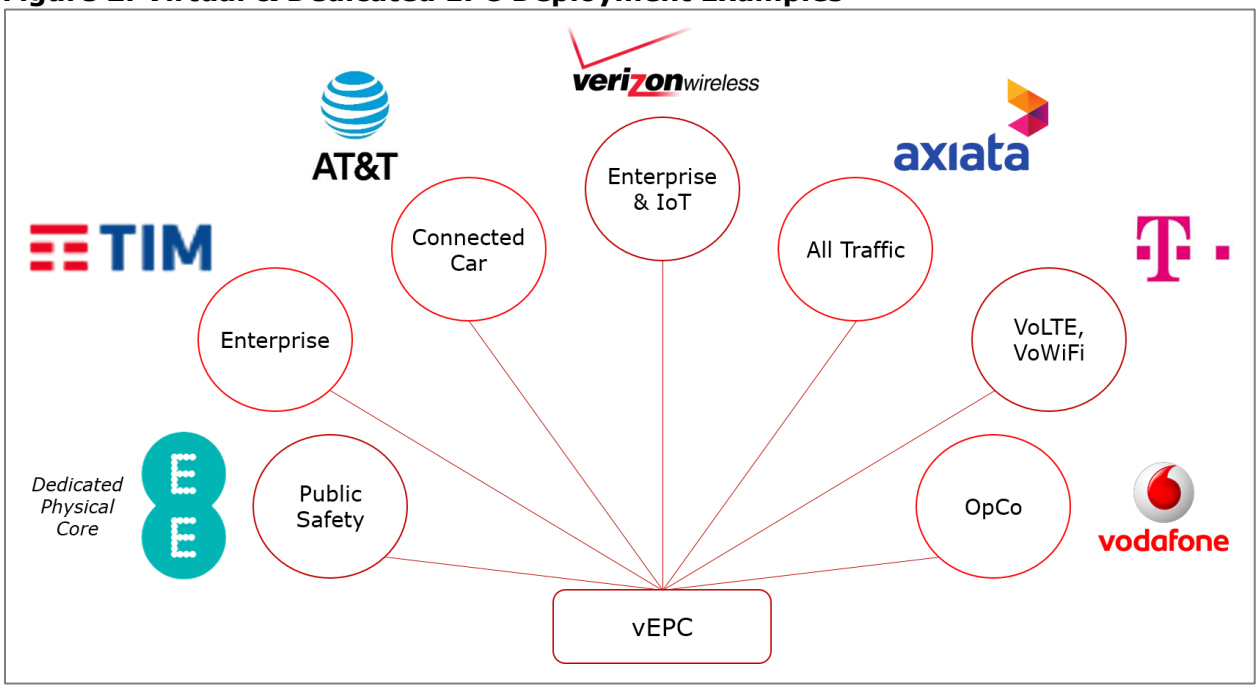
ORCHESTRATION & AUTOMATION: STATE OF THE ART

Mobile operators are making headway in process automation. The sector's day-to-day operations are vastly different, and more efficient, now than a decade ago. Even physical tasks, such as inspecting cell towers, are benefiting from new tools (in this case, remote inspection by drone). Elsewhere, advanced simulation tools, self-organizing networks, automated drive test and predictive analytics are examples of how mobile service providers are using automation and related software tools to improve operations. In the 4G and future 5G Core (5GC) the challenge is to transition to a cloud native model that is automated by design.

Good(ish) Progress With NFV

Depending on how you look at it, the industry has made reasonable progress on virtualizing parts of the mobile core – including the packet core, IMS core and ancillary functions – and, in a few cases, has done so at scale. Around the world you can find examples of virtual 4G core projects. **Figure 2** shows some examples related to EPC. Leading operators may have multiple virtual core for enterprises, MVNOs, IoT and projects underway simultaneously.

Figure 2: Virtual & Dedicated EPC Deployment Examples



Source: Heavy Reading

These deployment examples show how operators associate virtual core networks with a use-case and market opportunity. The connected car, for example, maps directly to new revenue from a growth market. IoT positions the operator for future growth in small data services, and so on. This is very positive trend.

The issue currently is that many of the deployments are of "craft-style" deployments. They work – often as well as physical deployments – but a lot of technical development has been expended to get to this state (and develop the craft). Moreover, this work is internally focused and is dominated by questions such as: What NFVI to use? How does the vendor provide

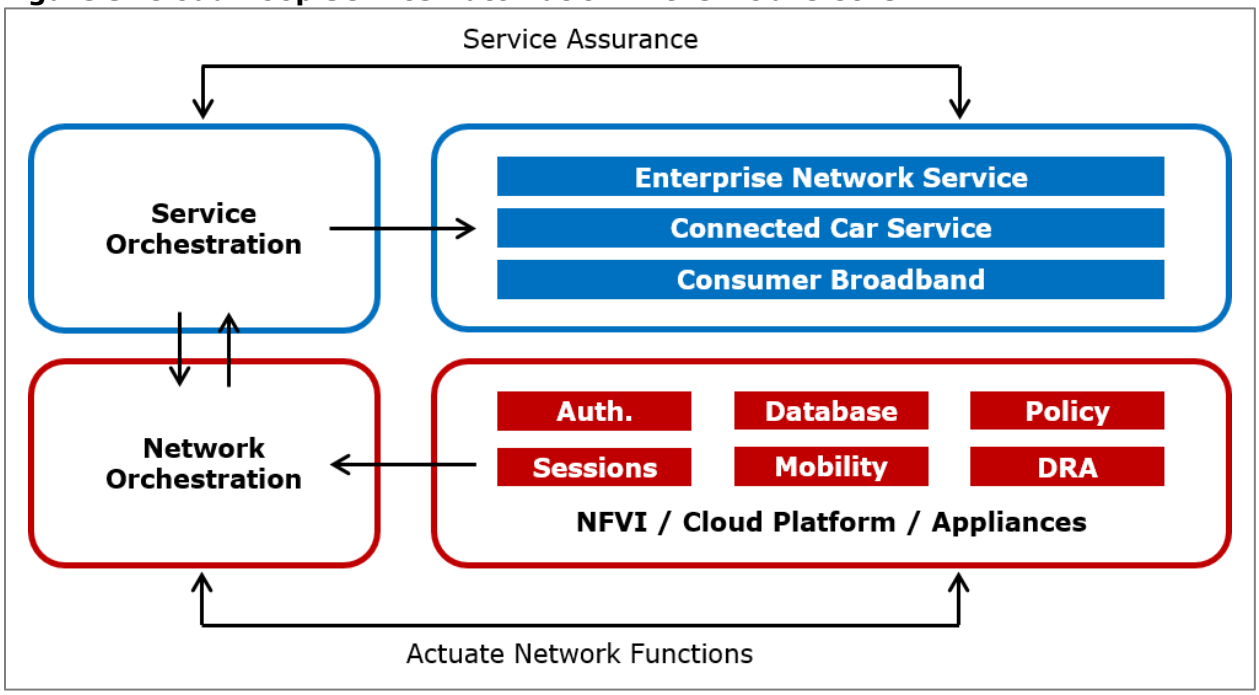
VNF config files? Which VNF-M does it use? How do we troubleshoot? What performance will we achieve with our call model? Have we agreed licensing? Is this on VMware or OpenStack, and so on? These challenge is that while these systems work, the development is internally focused and there is not enough attention on how customers will use and interact with services. There is a need to orientate NFV toward service outcomes.

Other domains such as SD-WAN have, arguably, adopted cloud principles faster than mobile networks and provide a model for how to map network technology development to how customer want to consume and interact with services. With the right toolsets and processes mobile operators can achieve the same kind of benefits.

Cloud-Loop Service Automation

Automation should address both services and infrastructure. These domains overlap but are often considered monolithic when, in fact, they are different processes with different intentions and outcomes. **Figure 3** illustrates the relationship. To deliver the desired outcome, service activation and network instantiation work in combination to ensure the path optimization at a flow level to direct the user to the correct network slice.

Figure 3: Cloud-Loop Service Automation in the Mobile Core



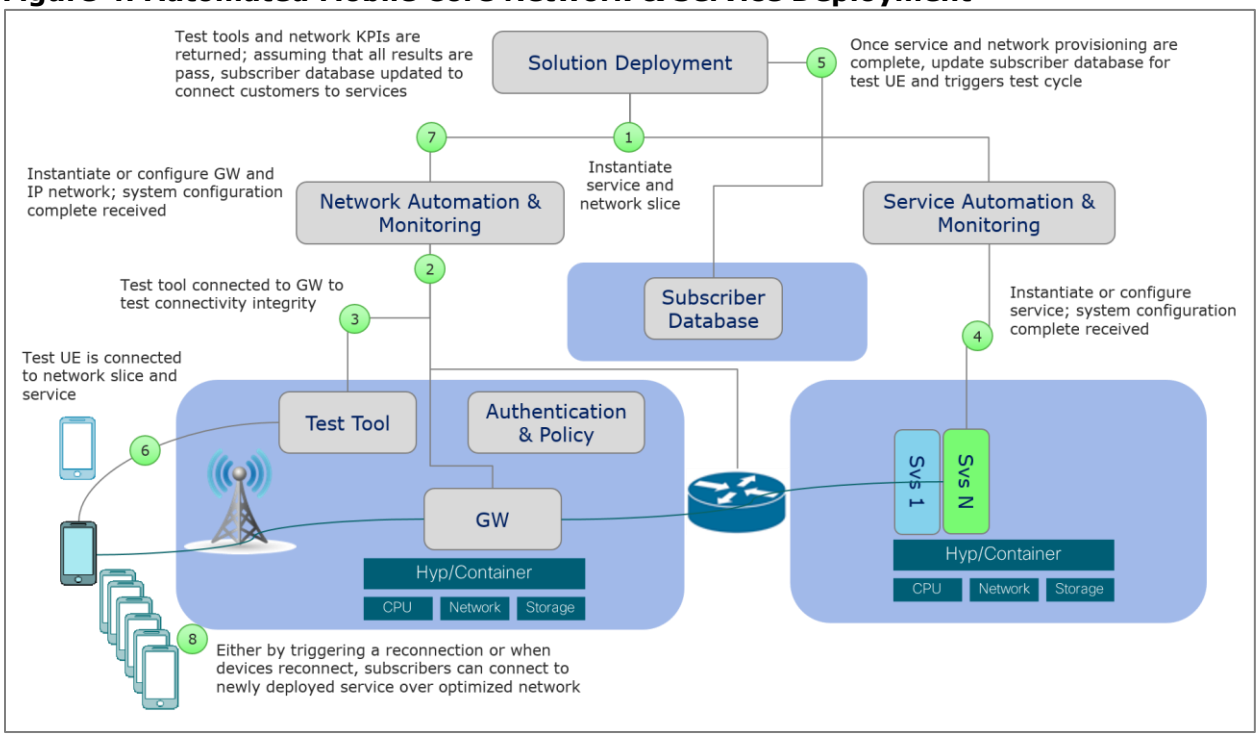
Source: Heavy Reading

In this model, network services (blue) are created using network functions (red) running on cloud infrastructure and physical appliances. In the example, services are "network slices," which in today's 4G network would mean services separated by APN and/or a dedicated packet core instance. Service assurance is used to monitor end-to-end performance, with the service orchestrator and network orchestrator working to reconfigure or scale the network, per the needs to the application. Actuating network functions to change configuration in response to feedback from service assurance serves to "close the loop" and automate both the service lifecycle and the underlying network infrastructure.

Automated Service Introduction: An Example

A more concrete example is shown in **Figure 4**. In this case, a new network slice, with appropriate IP services, is set up in the mobile core. This could be, for instance, a connected car service or a new enterprise network. In 4G terms, this is akin to a virtual dedicated core network gateway provisioned for use by a user group.

Figure 4: Automated Mobile Core Network & Service Deployment



Source: Cisco

The example shows network automation and service automation as distinct but interrelated processes. Monitoring and service assurance runs across both domains to provide a feedback loop. It also shows how it is important to not just provision a network instance and associated service, but also be able to direct the user to the new services, either at network attach stage or by forced reattachment. This process is important to operators that want to dynamically add and change services without having to plan for the manually onerous – and relatively slow and risky – migration of users between gateways. Integrating testing into the automation process removes long test cycles and the potential for human error from the process.

The process outlined in the diagram above is as follows:

1. The operator instantiates a new service and network slice via service orchestrator.
2. Network automation tool uses service commands to configure gateway and IP network. The toolset returns "system configuration complete" message to service orchestrator (a.k.a. solution deployment).
3. Test tool connected to gateway to test connectivity integrity.
4. Service automation tool used to instantiate and configure service. "System configuration complete" message turned to service orchestrator (a.k.a. solution deployment).

5. Once service and network provisioning are complete, the service orchestrator (a.k.a. solution deployment) updates the subscriber database to trigger a test cycle for the test UE.
6. The test UE is connected to the newly-created network slice and service.
7. Test tool returns network key performance indicators (KPIs). Assuming the results "pass," the subscriber profile is updated in the database to connect customers to the new service on the next attach request.
8. Either by triggering a reconnection, or when devices reconnect, subscribers are connected to the newly-deployed service over an optimized network path.

CLOUD & NEXT-GEN AUTOMATION

Cloud is the major enabler of next-generation mobile core automation. The intent is to abstract services from infrastructure such that a service can be deployed on demand using a console. This is dependent on automation *internally* to the cloud platform in terms of network orchestration and workload placement and *externally* in terms of enabling product managers to design and deploy services without having to request engineering resources, test cycles, and so on.

Cloud-Native Mobile Core

Virtualization was an early step to creating a software-based mobile core; while this has proven effective in some ways, the operational aspects of NFV have tended to reflect the hardware world. This experience shows that operators need a "Lego-block" way to create end-to-end network services using modular, interoperable software components. This is often described as "cloud native" and involves VNFs being designed or refactored to run in a cloud environment. The principles of cloud native mobile core networks are:

- VNF software disaggregation and refactoring. This is the transition to the use of microservices to create mobile core network functions.
- Scale-out/scale-in. Networks services should scale with demand and in response to resource availability or constraints.
- Automated lifecycle management. The network should not require human configuration of infrastructure or service logic during "normal" operations.
- Distributed N+K redundancy. Classic mobile core networks typically rely on 1+1 and N+1 redundancy at the chassis and line-card level. Cloud core networks are distributed and on demand.
- DevOps-like operations. Rather than the waterfall processes, operators are seeking operational agility, which in time may also extend to in-house software development and deployment.
- Self-service and on demand. Ultimately, network services should be as easy to buy, consume and quit as cloud services.

These are generic principles for cloud-centric telecom networks. When applied to the mobile packet core, there are some specific implications:

Stateless & State-Efficient Operation

In mobile core networks today, many different nodes maintain subscriber state information in databases integrated within the appliance – for example, an MME, S-GW, P-GW, PCRF, OCS, etc., each retain state information for each user for the duration of a session. This reflects a monolithic view of network function design and makes it hard to keep the different functions synchronized. If state can be removed from VNFs into either a separate data store or to a microservice, this brings the inherently stateful 3GPP architecture more into line with modern, stateless cloud applications.

Generally speaking, it is simpler to make user-plane functions stateless, or at least to remove state while a user session is in progress and retain it in an external common data store. Control-plane functions are more complex in that they need state information to operate correctly; in this case, it may be better to think in terms of being "state-efficient" by moving state to a dedicated microservice rather than to an external data store.

By removing state from the runtime part of the application, it is easier to terminate a VNF by migrating users/session to a new VNF instance (which re-acquires state information, as needed, from the separate data store or microservice). This changes how network functions are deployed, used and maintained. No longer does the operator need to upgrade an in-service VNF; instead it replaces it with a newer, improved version. The ability to rapidly add or remove VNFs from a service path, at low risk of service disruption, will bring much greater to agility to operating procedures.

API-Driven Operations

A cloud-based core network uses APIs to control VNF instantiation, to combine VNFs into service paths and to scale and terminate VNFs. This is again linked to the idea of Lego-block-like service creation. In this model, services are created using a console, which selects from a library of VNFs to create sub-network instances and, in turn, end-to-end network services (or "network slices"). This should include integrated testing, as shown in **Figure 3** above. From console to instantiation and operation in the network requires well-defined APIs. There are many important and candidate APIs involved in NFV process automation and in hybrid NFV-Physical services – for example, API specification is underway in several fora, including ETSI NFV, ONAP, MEF, IETF, TM Forum and so on (see also **Figure 6** for API types).

API-driven services are even more central to 5G. An important part of 5G is creating a network platform that is open to multiple industries for diverse services. This means these customers should be able to consume, and interact with, the network. This is addressed in the 5G core by the Network Exposure Function (NEF), which acts as an API gateway and will allow external users, such as enterprises or partner operators, the ability to monitor, provision and enforce application-level policy for users inside the operator network. This could be useful for federated services, such as international or cross-operator network slicing, or for hybrid enterprise/operator networks. The NEF (and associated Nnef interface) is being specified as part of the 5G system architecture in TS 23.501 in Release 15. It is likely to be some time, however, before this capability is commercially used.

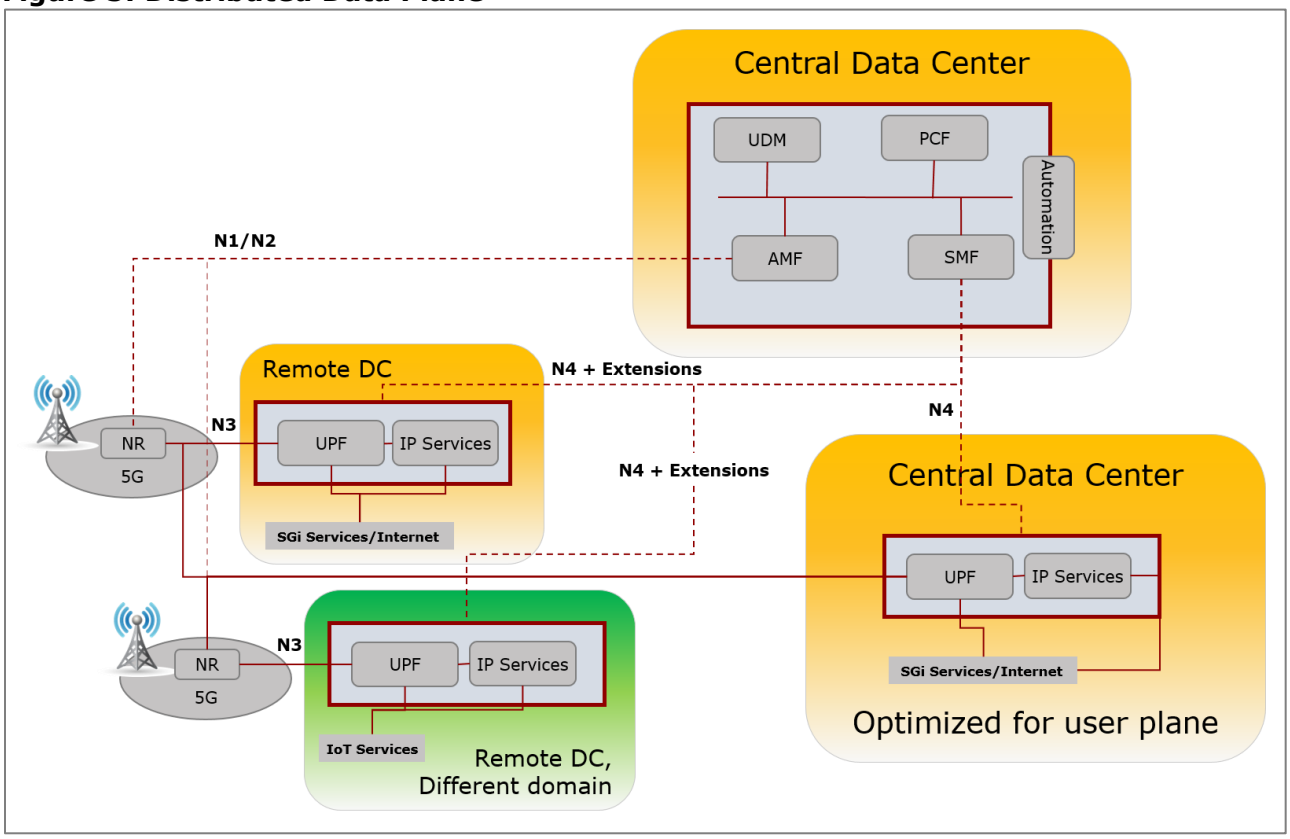
Distributed User Plane

Centralized control and distributed processing is a defining characteristic of the cloud. The same concept is used in software-defined networks and can be applied to 4G and 5G mobile core networks. In the 4G Evolved Packet Core, control- and user-plane separation (CUPS)

has been specified to decouple handling of user traffic from control of it. This enables independent scaling, which could mean, for example, that more traffic for the same number of users is easily accommodated, and offers the potential for distributed user-plane nodes, which could mean, for example, nodes placed in distributed data centers (a.k.a. MEC with full core EPC capabilities) or enterprise premises.

Whereas CUPS was developed for 4G after the initial standardization, in 5G the concept is inherent to the system architecture, which formally distinguishes the user-plane function (UPF) from access management function (AMF), session management function (SMF) and the rest of the control plane. This architecture is shown in **Figure 5**, with the control plane deployed centrally and user-plane nodes distributed across multiple data centers, alongside IP services needed to process the flow. User groups can be directed to the correct user-plane instance according to subscriber profile and policy.

Figure 5: Distributed Data Plane



Source: Cisco

A cloud model is inherently more flexible in terms of where functions (a.k.a. workloads) can be placed, making it appropriate for a distributed network architecture. For example, an operator or enterprise may wish to place user-plane nodes on premises to optimize performance or reduce dependencies on a WAN link or in an edge data center, such as a rearchitected central office. Once the edge cloud infrastructure is in place, it is important to be able to program these edge devices and functions using toolsets to translate intent into actions, automatically. To ensure users terminate at the location hosting the services, other functions, such as DNS, CDNs and IP pools, also must be configured. In a dynamic network, this quickly becomes complex and can only be achieved through automation.

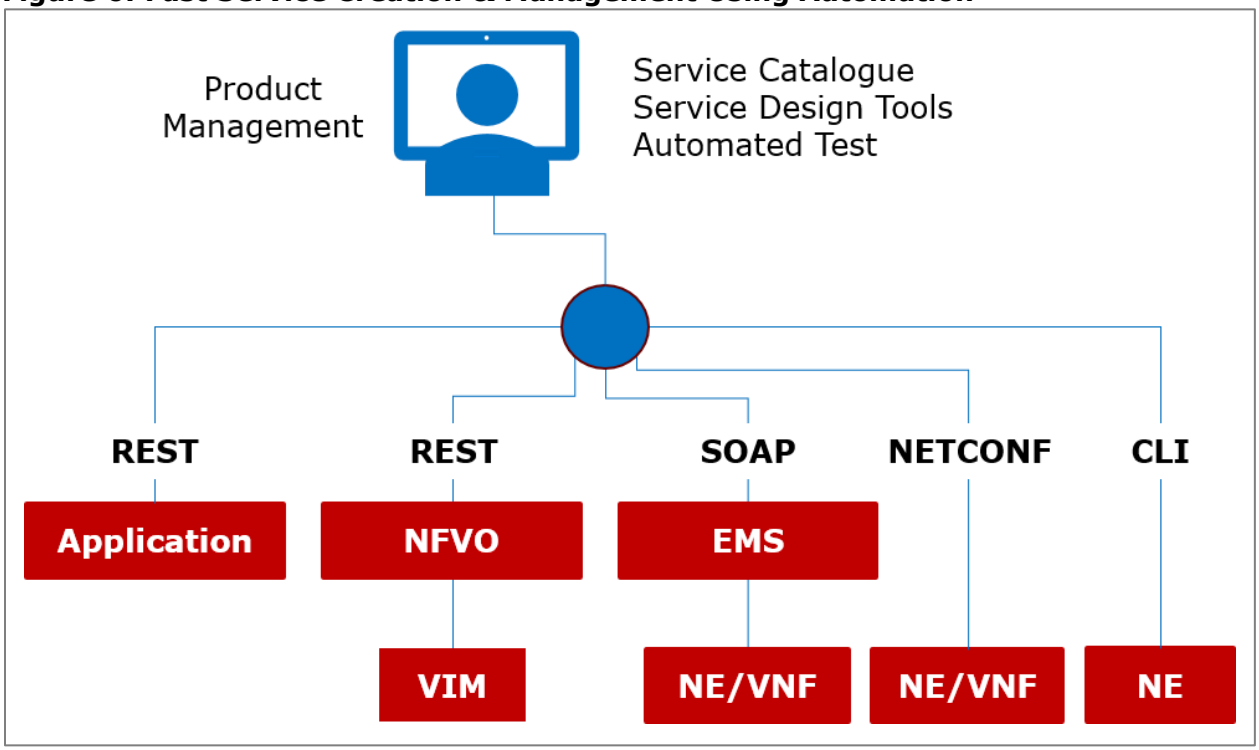
SERVICE CREATION & DEVOPS

Internal network operation is important, but to be transformative to operators, automation must extend out of the network and impact the customer directly. Automation of service creation and service management has the potential to make operators more agile and responsive and, therefore, more able to participate in the dynamic, internet economy.

Console-Based Service Management

The service design and implementation process often involves multiple teams from marketing, product management, finance and engineering. This is appropriate for major lines of business and critical new services. The same structure, however, is not suited to scenarios where the service template is already designed and in operation, but the customer requests a change or product managers identify an add-on opportunity or want to test a new feature. Adapting or amending a network slice may fall into this category. In this case, a model where product management can work with subject matter experts to configure network services without having to request engineering resources and long test cycles is more appropriate. As shown in **Figure 6**, this means console-based service configuration with automated actuation in the network.

Figure 6: Fast Service Creation & Management Using Automation



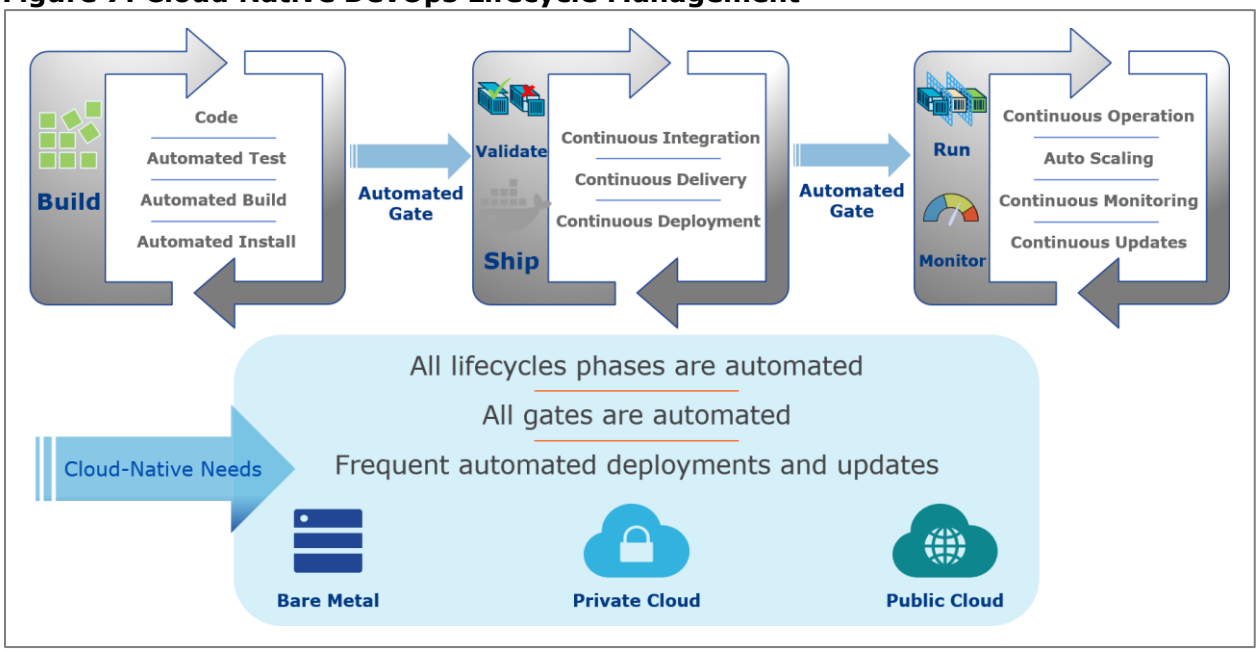
Source: Heavy Reading

The ultimate extension of this concept is self-service networks, such that an enterprise, or perhaps even individual users, can configure and use network services as easily as they consume cloud services today. Converting customer requests made via a console into network actions, without manual intervention, is thus a key objective of automation. This will be gradual process as toolsets, network programmability and service assurance mature.

DevOps-Like Operations

Another defining feature of cloud applications is DevOps. This refers to a culture that empowers smaller teams to act independently, in an agile manner, without being encumbered by traditional "waterfall" processes. The concept is borrowed from software development and startups and is being promoted by operator technology executives as way to make their organizations move faster. **Figure 7** shows how a DevOps-style service lifecycle management is intrinsically linked to automation, with each lifecycle phase – build, ship and monitor – incorporated within an automation loop.

Figure 7: Cloud Native DevOps Lifecycle Management



Source: Cisco

DevOps is an attractive idea, but requires translation, not to mention a cultural shift, to be useful to mobile network operators. In 2006, Werner Vogels, the CTO of Amazon Web Services, said "you build it, you run it" in reference to the idea that a developer of a software module (a.k.a. a microservice) would also be responsible for operations, ongoing maintenance and improvement. The same idea is inherent to perhaps the canonical example of an application built from microservices, Netflix, where individual developer teams could create and deploy new code quasi-independently. In a telecom operator environment, this is obviously not done – no major operator would allow individual developers to commit code to VNFs running in live networks – and, in any case, this is probably not a good idea for nationally critical telecom infrastructure.

Nevertheless, the idea that product managers, as discussed above, could create custom services using a library of modular functions, deploy them and change them as needed is attractive and meaningful. This could be thought of (to borrow a phrase) as DevOps at the service level. An example, still aspirational today, might be a mobile virtual private network (VPN) service in which policy and security can be customized and changed for a customer on demand, without having to perform lengthy regression tests on many other services. This is closer to what the telecom industry means when it talks about DevOps, showing how automation directly impacts the customer – and therefore, the operator's bottom line.

CONCLUSION & SUMMARY

Automation is about efficiency and a lower cost of production, which are essential features of LTE-Advanced Pro and 5G networks. This white paper has discussed how operators can generate large gains in operating efficiency in the 4G/5G mobile core by automating deployment of network services. By developing repeatable processes and automating them, operators can radically recue the risk to service disruption from human error and be more agile in how they design, deploy and test services. In a dynamic networking environment, the paper has argued that operators can only achieve their business objectives through automation

Equally as important, efficiency is the role automation can play in service creation, management and scaling. Better operational processes translate directly to more agile service delivery that will ultimately make networking as easy to buy and consume as cloud services.

This will fundamentally change how customers engage with mobile network services and will expand operators' addressable markets and underpin their economic future. For enterprises that want simpler, faster, more efficient services, mobile core automation enables operators to meet their needs. Over time, as the tools and technology mature, for those enterprises that wants more direct control of services, automation will allow operators to convert customer demand directly into network configurations instantiated as services. As 5G and network slicing come to the fore, this capability will be essential.