Greenfield Mobile Network Considerations

Converged Networks and Mobility

A Technical Paper prepared for SCTE by

Ravi Guntupalli
Director of Technology, Mass-Scale Infrastructure Group
Cisco Systems, Inc.
raguntup@cisco.com

Ibrahim Ayad
Engineering Architect, Mass-Scale Infrastructure Group
Cisco Systems, Inc.
iayad@cisco.com

Irfan Ali
Principal Engineer, Mass-Scale Infrastructure Group
Cisco Systems, Inc.
irfaali@cisco.com

Table of Contents

Title **Page Number**

List of Figures

Title **Page Number**

List of Tables

Title **Page Number**

1. Introduction

Cable operators continue to expand the service offerings towards the end-users with wireless and cellular capabilities. Some of the operators already leverage Mobile Virtual Network Operators (MVNO) relationships with existing nationwide Mobile Network Operators (MNO) partners to offer mobile services on the macro cellular networks. This enables the Cable Operator to be able to offer a unified experience for the end users and able to tap into additional revenue by offering differentiated services.

However, depending on the markets and needs, the availability of spectrum or regulatory requirements Cable operators are considering deployment of new 4G and 5G networks from the ground up, which are a Greenfield network. Depending on the timing of the market in the region and the device ecosystem availability within the region, it may be possible that the Cable Operator would have to deploy a 4G radio but given the timing of the industry, we anticipate more 5G network deployments to ensure investments are directed towards the network of the future. In some cases, the operator may choose to leverage a 4G+5G Radio access network and terminating on a 4G Core. This architecture – widely known as 5G NSA would allow for the cable operators to leverage existing device ecosystem but also partially build towards the network of the future. However, more and more MSOs seem to be evaluating a 5G Standalone deployments and this paper tries to focus primarily on the 5G SA deployment models but also addresses ability to handle specific scenarios of 4G+5G deployments. In this paper, we consider any 3GPP standards based wireless network that is being built up for the first time by a provider without having the dependencies of legacy 3G network interworking or device support related to 2G/3G.

Additionally, we use the terminology Multiple Service /System Operator (MSO) to refer to cable operators who have decided to deploy Greenfield mobile networks. By contrast, Mobile Network Operators are the traditional mobile operators like AT&T, Verizon, T-Mobile, Vodafone, etc.

While architectures for deployment based on 3rd Generation Partnership Project (3GPP), Cable Labs and other industry standard development organizations (SDOs) are widely available and considered before finalizing deployment, it is also critical to consider some of the lessons learned by wireless network providers over the years ramping up the network capabilities.

Based on some lessons learnt in deploying 4G networks globally as well proven best practices in the industry for macro and microcellular networks, this document addresses and captures multiple key challenges to anticipate and plan for both the architecture and operational models. Given that 5G SA architecture leverages new capabilities like Service Based Architecture and some of the challenges related to 4G networks based on point-to-point interfaces using Diameter and GTP-C interfaces may not always be applicable, the authors believe that initial set of deployments will have some form of point-to-point capabilities – though leveraging new protocols defined in 3GPP. Hence, some of the lessons learnt in 4G networks would absolutely be applicable.

2. 5G – Architecture and Capabilties

There is significant amount of collateral and information related to 5G and the new capabilities that 5G network architectures bring to the table. Rather than reiterating the capabilities of 5G and cover the architecture options in detail, this section provides a very high-level view of 5G, and the nomenclatures used in the rest of the document.

Beyond just being an advancement in mobile generation – 5G was expected to lay a path for new capabilities and services that are yet to come. The underlying premise for the architecture enhancements was to enable capabilities that would allow Service Providers leveraging 5G to break through some of the constraints of legacy internet connectivity paradigms and enable a new set of differentiated services.

Within the scope of a 5G Network, we believe that the network must be flexible enough to deliver a wide range of new services but also be able to address existing legacy applications. Given the complexity of owning a brand new 5G network, the network needs to account for ease of deployment but also for ease of day-to-day operations and management.

The figure below takes a high-level representation of the various domains of an end-to-end 5G network.

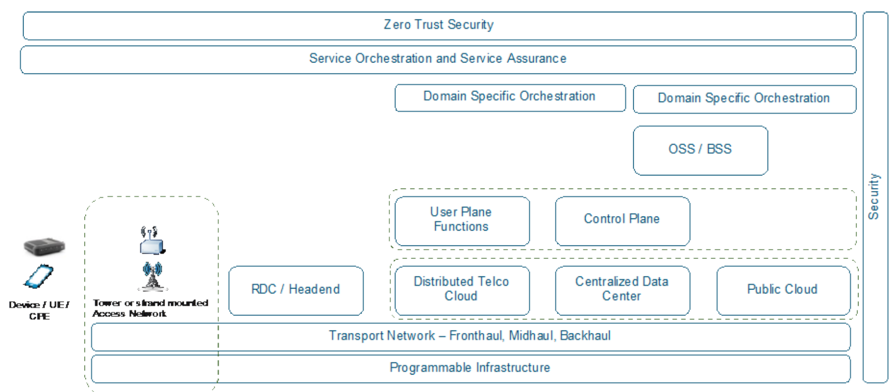

Figure 1: Different Domains of a typical 5G network

At a minimum, the following domains must be individually understood and designed for a successful deployment of 5G network.

- Radio Access Network
- Transport Network
- Packet Core
- Automation and Assurance
- Compute and Infrastructure
- Security

While this document does not go into each of these individual domains, care must be taken to ensure each area is tuned and enhanced to account for 5G network deployments for the specific MSO environment. For subsequent sections, capabilities and unique learnings and architecture options around packet core network are detailed.

A simplified representation of a 5G network architecture is shown below with the key capabilities described.

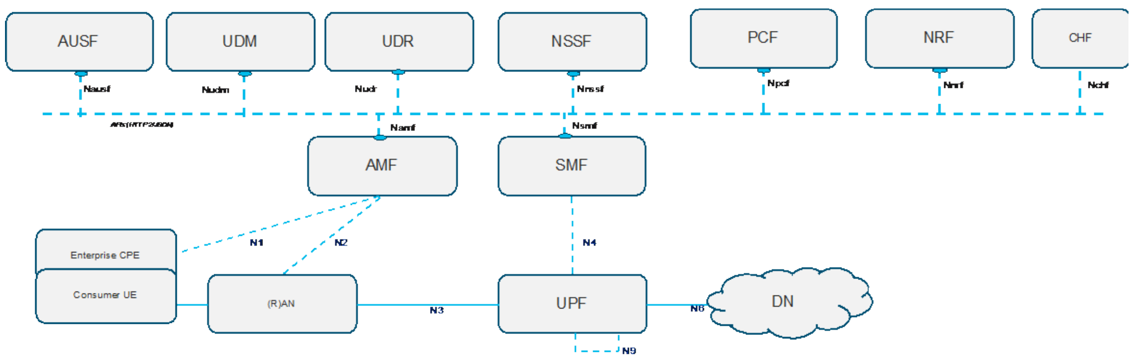

Figure 2: Simplified 5G network architecture with packet core focus

- **Consumer UE / Enterprise CPE**: The end user device that is connecting via the radio access network or in case of MSO environment, via the existing cable network back into the 5G Core network.
- **Radio Access Network (RAN)**: In the case of 5G wireless access technology, the Access Network typically consists of a 5G gNodeB which performs Radio resource management, radio bearer control, scheduling of packets over the air interface among other access management functionality. In MSO deployment scenario, based on 3GPP Release 16 or later specifications, Access Network could be existing access capabilities like Cable plant or WiFi technologies terminating via appropriate trusted or untrusted access.
- **Access and mobility management function (AMF)**: AMF is responsible for Registration management, access control and mobility management function for all accesses (incl. WLAN), Terminates NAS signaling for all accesses (single AMF per UE). AMF receives mobility related policies from PCF (e.g., mobility restrictions) and forwards mobility related policies to the UE (via N1)
- **Session management function (SMF)**: With 5G SA architecture, SMF provides a mechanism to deploy a common session management for all accesses (incl. WLAN) and wireline capabilities. SMF handles all session management signaling with UE (relayed by AMF), controls the user plane functions. SMF interfaces directly with UDM to receive subscription information (no need to go via AMF) along with policy information from PCF.
- **User plane function (UPF)**: UPF supports set of operations (forwarding to other functions, encapsulation/decapsulation, bitrate enforcement, application detection etc.) and acts as a primary anchor point of IP packets in the network. Within 5G SA architecture, SMF dynamically configures UPFs (activates/configures subset of the operations defined above) to provide the traffic handling functionality needed for a session. In addition, one or

multiple chained UP functions can be activated and configured by SMF per session as needed for a scenario. Within the MSO network architecture, UPF can also enable convergence capabilities by unifying access gateway capabilities when leveraging 3GPP Release 16 and additional SDO specifications.

- Policy control function (PCF): The PCF Provides QoS and charging rules to Session Management Function and interfaces with external Application function (e.g., IMS) when applicable. PCF also provides mobility related policies directly to AMF (e.g., mobility restrictions for stationary devices (FWA) and optionally provide policies to the UE e.g., on network discovery/selection. Additional support for network discovery/selection policies, while defined in 3GPP require UE/CPE capabilities to be deployed in the network.
- Unified Data Management (UDM): UDM uses subscription data (including authentication data) that may be stored in UDR and is responsible for generation of 3GPP AKA Authentication Credentials, identification Handling and Access authorization based on subscription data.
- Authentication Server function (AUSF): AUSF Supports authentication for 3GPP access and untrusted non-3GPP access and supports recover from synchronization failure in certain cases.
- Unified Data Repository (UDR): UDR carries the subscription data and offers services for UDM, PCF etc., to allow for retrieval of appropriate user data to be used for allowing the device onto the network and applying the corresponding policy. In MSO deployments, UDR could be a mechanism to unify the wireless and wireline subscription data in the future.
- Network Slice Selection Function (NSSF): Network Slice allows for a self-contained, logical portion of an E2E network resources within a service provider but to ensure the UE/CPE is allowed access to the appropriate slices defined, NSSF was introduced as a new capability within the 5G SA Architecture and helps selecting the set of Network Slice instances for the specific devices.
- Network Repository Function (NRF): With the introduction of Service Based Architecture, Service Providers have the ability of the steering away from somewhat of a point-to-point network architecture in 4G networks. NRF enables service registration from NFs and acts as a service discovery function – either for enable direct or indirect communication between various functions defined above.
- Charging Function (CHF): The charging function is responsible for generating charging data records (CDRs), based on usage information obtained from the UPF and SMF. The CHF interfaces to the operator's billing system.

Note that while this section captured some of the high-level capabilities of the critical network functions in the 5G architecture, 3GPP TS 23.501, Section 6.1 carries a comprehensive list of capabilities for the complete set of the network functions. Full set of functions as defined by 3GPP are listed in the Appendix A.

3. What could 5G be for MSO's

Cable operators get presented with more opportunities every day in different countries to expand the footprint and offerings to the end users by enhancing the service capabilities and enable new set of customer use cases and experiences. With more spectrum options being made available by regulatory bodies in different countries – example CBRS and C-band in US, Canada's 3.5 GHz auction and upcoming auctions in India as well as other countries - the ability for Cable operators to enhance their network to offer mobility solutions to their customers is more viable than ever.

Especially with 5G architecture capabilities defined in 3GPP specially to address the wireline and wireless convergence (5WWC Work Item in Release 16 of 3GPP specifications[1])and additional capabilities being addressed by Cablelabs[2] Cable operators are certainly closer to achieving the goal of becoming MSOs. Cable operators' interest for foraying into the mobility domain is being driven by a multitude of factors – some country specific and some globally applicable.

Offering a mobile service to the users could:

- Significantly increase revenues – new wireless service to existing customers or new standalone wireless customers
- Enhance customer experience by offering seamless service experience – extending the home experience to wider area
- Offer differentiated services that could not be offered on wireline alone – broadband services including voice

Depending on the country, regulatory requirements, and competitive positioning – Cable operators are typically presented with multiple options.

1. Tight partnership with an existing Mobile network provider
2. MVNO network agreements with one or more existing Mobile network provider
3. Build a standalone network with nationwide coverage
4. Build pockets of wireless network with MVNO or partnership with Mobile network provider for enhanced coverage by leveraging small cells or micro coverage

When it comes to a mobility network built from the ground up, some Cable operators have traditionally relied on unlicensed spectrum to offer mobility for the end users, leveraging capabilities of WiFi which is being enhanced every day with new capabilities as well. WiFi6 will continue to be a critical part of the MSO architecture it is critical to understand the capabilities of 3GPP to deploy 5G networks in the licensed bands as well as shared access and unlicensed access being leverage would be available. In addition to network capabilities, enhanced device capabilities also offer new potential options for MSOs that have traditionally not been present. As an example, being able to leverage eSIM capabilities will enable MSOs to migrate subscribers faster and more seamlessly onto the new MSO networks.

---

[1] https://portal.3gpp.org/desktopmodules/WorkItem/WorkItemDetails.aspx?workitemId=830050
[2] https://www.cablelabs.com/specifications/WR-TR-5WWC-ARCH

While some MSOs could have common requirements and the end use cases being similar, as captured in Cable and Mobile Convergence, A Vision from the Cable Communities Around the World[3] and Section 4, "A Survey of Mobile Deployment Plans by MSOs Around the World", the deployment models and the architecture options being considered vary per country and per MSO.

3GPP network architecture and the spectrum options could be classified into the following broad set of deployment architectures:

- Consumer Fixed Wireless Access
- Consumer Enhanced Mobile Broadband Access
- Enterprise focused wireless access – Private Networks or a Private access via public access

Deployment of RAN assets could vary as well – including but not limited to strand mount access nodes, small cells, dedicated base stations or shared resources access nodes. We believe there are 3 main architecture deployment models as listed below when the MSO chooses to deploy a wireless network based on 3GPP standards.

3.1. MNO Partnership

This is a scenario where the nationwide coverage is provided by the MSO with a tight partnership with an existing MNO. The MSO can build and deploy a selected set of markets or regions in the country to offload data via MSO network – which reduces cost of delivering the service and / or offers a more differentiated service compared to the MNO network. This can be achieved via roaming relationship with a single identity or DSDS capabilities[4].

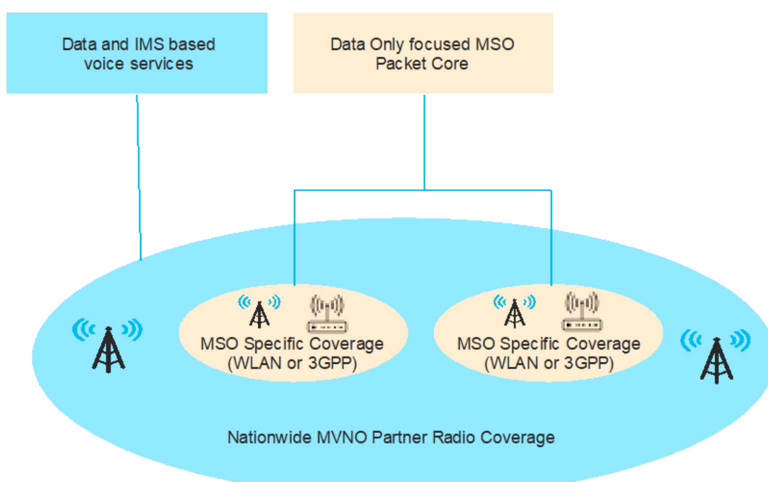

Figure 3: MNO Partnership for nationwide coverage with MSO pockets of coverage

[3] https://www.nctatechnicalpapers.com/Paper/2020/2020-cable-and-mobile-convergence
[4] DSDS details depicted in "Cable and Mobile Convergence, A Vision from the Cable Communities Around the World" section 4.1.5

3.2. Nationwide MSO

Ability to offer a nationwide coverage provided by the MSO on Day 1 leveraging existing WLAN assets and deploying new 3GPP RAN where applicable and available based on regulatory requirements. These scenarios would require the MSO offer full voice capabilities in the network and deliver regulatory compliance like Emergency calling from Day 1. In lieu of a tight MNO partnership, the MSO can leverage roaming partnerships but may be challenging to achieve roaming within the same country without a tight relationship with the MNO. While the network is being expanded and built out with 3GPP, the MSO may have some coverage gaps and impacting the subscriber experience.

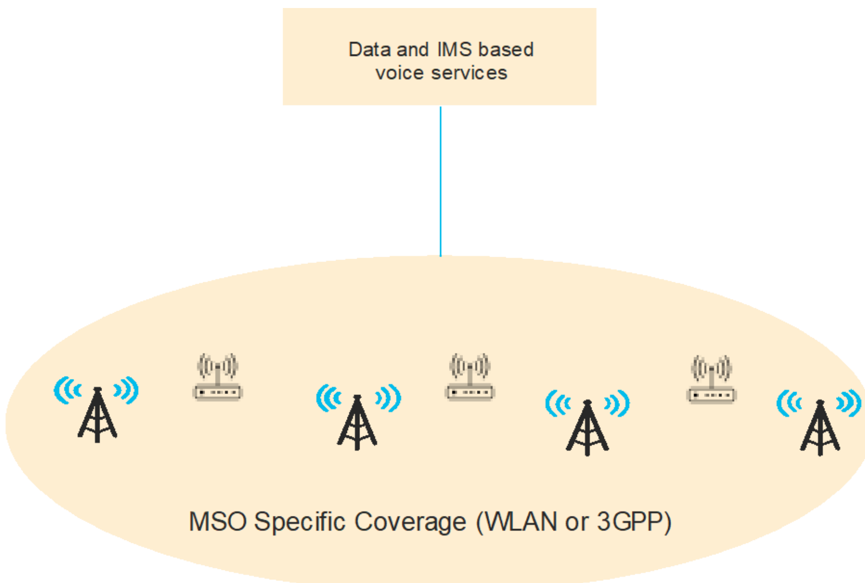

Figure 4: Nationwide MSO network without MNO partnership (except roaming)

3.3. Private Networks

A scenario where the MSO could choose to focus primarily on new opportunities only using the licensed spectrum and target the Enterprise Private Network markets. The MSO may choose to offer a consumer service only via a pure MVNO partnership as an option – however, the spectrum assets are leveraged only for the enterprise deployments to begin with. With this model, the MSO could always expand to nationwide MSO or MNO partnership models in the later phases. MSO may choose to deploy a dedicated or a multi-tenant core for enterprises

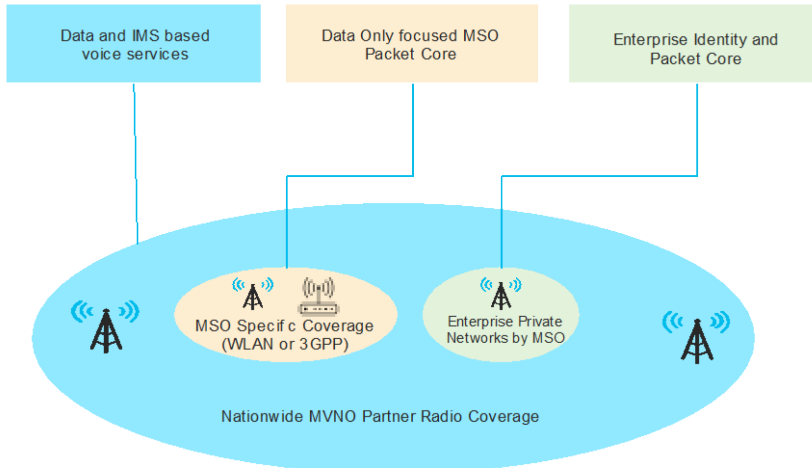

Figure 5: Focus on Private Networks and Limited MSO consumer coverage with nationwide MNO partnership

All these network options could exist by themselves as standalone or fully integrated. Introduction of additional capabilities could be in phases – as an example,

- 5G Fixed Wireless Access Network Deployment – Data offload only
- Voice Network introduction – migrate from MVNO Voice
- WiFi integration
- 5G Differentiator Features Deployment: Slicing; MEC; External APIs and Roaming supported
- 5G Core Advanced Features Deployment: Network Analytics (Artificial Intelligence (AI)/Machine Learning (ML)); Common Data Layer deployment

The sequence of the phases does not always have to be as depicted above but rather based on operator priorities.

While the 3GPP architecture seems simple on paper, what drives the deployment complexity is the service differentiation the operator chooses to provide towards the end user. As an example, a SMF/UPF combination depicted in the 3GPP architecture is supposed to deliver at a high level, the following base functionality for simple packet routing.

- Session Management and UE IP address allocation
- Configures traffic steering at UPF to route traffic
- Interfaces towards Policy control functions
- Lawful Interception and other regulatory requirements
- Policy rule enforcement, (e.g., Gating, Redirection, Traffic steering) and QoS handling
- DPI and application detection

In addition to the functionality defined above, due to operational requirements an operator will have:

- Local Redundancy in case of card or a process failure
- Geo-Redundancy in case of site failure or connectivity challenges
- Local Policy in case of temporary glitches in connectivity
- Local storage or accounting records and NAT Binding records

In certain scenarios, building out as a greenfield mobile network operator, the MSOs will be presented with choices to build a simplified network. However, it could end up being a competitive differentiation or a service parity with the MVNO network scenario which may end up forcing the MSOs to deploy an equivalent capability in the network.

As 5G network architectures evolve and deployment plans are formalized, the industry is driven not just by competitive pressures, but also based on evolving use cases which are heavily dependent on the availability of the device ecosystem that supports various 3GPP capabilities. The current expectation of the market trend as seen by the authors is depicted below.

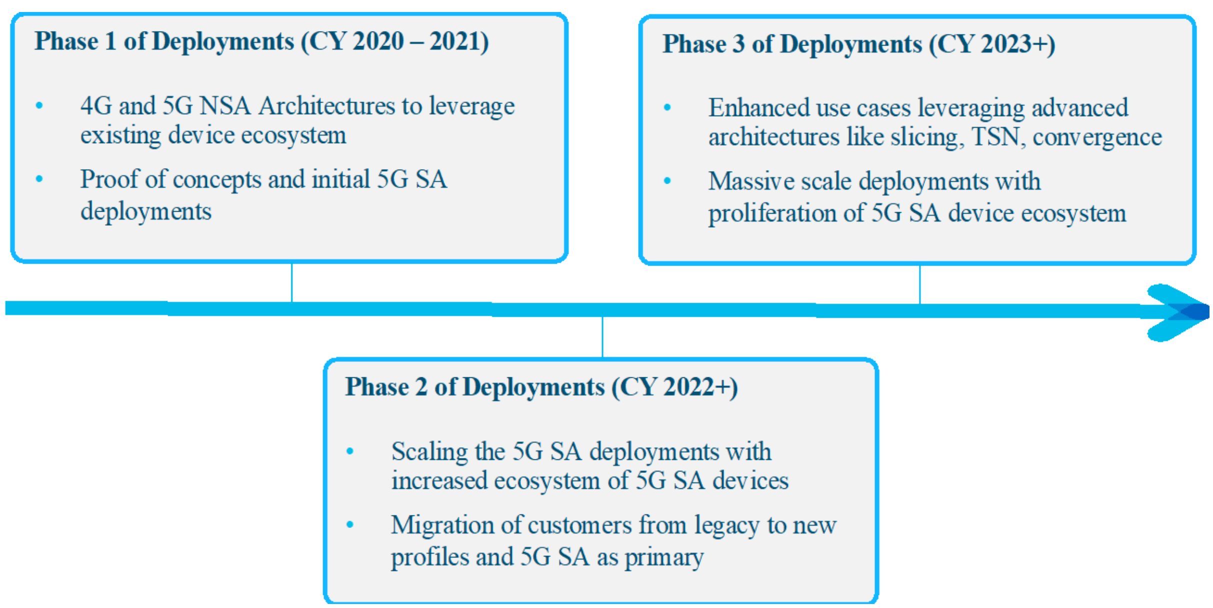

Figure 6: Likely timeline of deployments based on industry momentum

While there have been enough publications around what 3GPP network architectures in 5G offers for MSOs, technical capabilities, deployment models, this paper focuses on highlighting some of the key lessons learnt in existing Mobile network deployments – during 4G deployments and some initial 5G deployments. Intent of this paper is to offer solutions to common pit falls in mobile networks as it stands today – to account for these from the get-go and avoid any potential network issues and service interruptions. The authors believe that the lessons learnt during the initial 4G deployments, and the maturity achieved over the period of 10+ years in 4G networks should be leveraged as 5G networks are being built out.

4. Greenfield Network Considerations

4.1. Converged 4G+5G core Deployment

One of the important decisions an MSO faces would be to either deploy a 5G only RAN (also called as 5G Stand-alone (5G SA) or Option 2 RAN) or 4G 4G-and-5G non-standalone (5G NSA) or Option 3 RAN. The high-level architecture for the two options is shown in the Figure below.

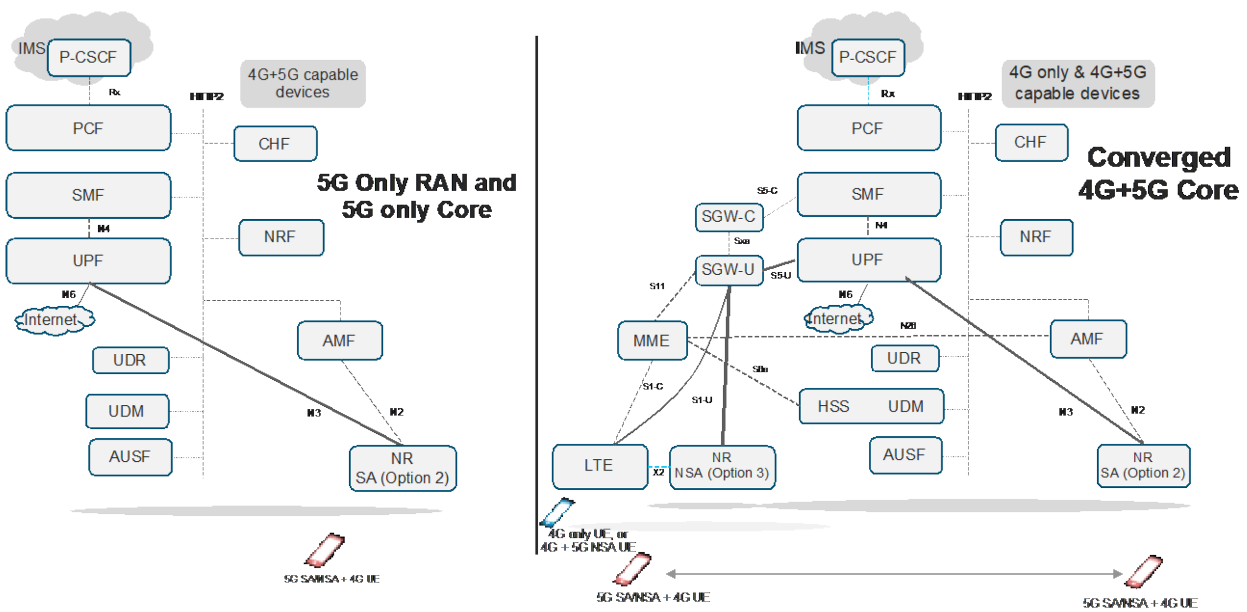

Figure 7: Converged 4G+5G Core architecture vs. 5G SA architecture

The left side of the Figure depicts the key network functions of 5GS system that only serves 5G RAN (NR) in a stand-alone mode. This would imply that in the gNB in NR band for the specific country is deployed and there is no 4G LTE deployment. Also, the mobile supports 5G SA mode of operations.

On the right side is a converged 4G and 5G core that also support 4G LTE radio and core network. The 4G related elements includes the MME and Serving Gateway (SGW). Also interworking interface between the MME and AMF (N26) is used to enable seamless mobility between NR and LTE. With this deployment the MSO can offer both LTE and NR in different frequency bands in the 150 MHz CBRS frequency spectrum, as an example. In addition, the mode where NR radio is added for additional downlink and uplink data rates with signaling going via the LTE cell. This configuration is called 5G NSA or Option 3. This requires the deployment of both LTE and NR i.e., eNB and gNB. The same NR radio can operate both in Option 3 and in Option 2 mode. With this deployment mode an operator can support legacy 4G UEs and the new 5G NSA and 5G SA UEs.

The main driving factor in such a decision is the availability of 5G SA capable devices. Though the number of devices that support 5G (NSA and/or SA) is increasing it is still very small compared to the 4G UEs (May 2021: approximately 800 device types that support 5G NSA or 5G SA compared to about 25,000 device types that support 4G technology[5]). Furthermore, the number of 5G devices that support 5G SA is a small fraction of the devices that support 5G (NSA or SA). Though the number of devices (phones and other form factors) that support 5G will increase, this number will still be smaller than those that support 4G. Depending on the timeline of deployment, it is possible for an MSO to deploy the converged 5G and 4G core. The operator should deploy this converged core also in comparison to deploying a separate 4G only EPC core (with PGW and PCRF instead of SMF+UPF and PCF), since EPC core cannot support 5G NR SA mobiles and EPC also is based on legacy telecom specific protocols like Diameter and GTP-C whereas the converged 5G and 4G core is based on protocols that have much wider deployment in the market, e.g., HTTP.

If the MSO chooses not to deploy 4G radio and no support for 4G only devices is required, the MSO can choose to deploy a SA core. However, given that MSO subscribers could potentially roam into a country that does not yet have 5G deployments or a partnership with a roaming provider with 4G only capabilities, it may be required to terminate the legacy GTP-C interfaces from the roaming partner. This would mean that the converged 4G+5G core will primarily be leveraged as a 5G SA Core when the device is on the MSO deployed RAN but when the device is connecting from a 4G only partner RAN, the converged core capabilities of exposing legacy GTP-C interfaces i.e., S8 interface could be leveraged. Note that in this case, N26 interface is not required unless a tight roaming relationship with handovers is planned.

4.2. MNO/MSO demarcation points

As stated in the previous sections, the most common deployment model mobile network deployment for an MSO could be one with partnership with an MNO, where the MNO provides nationwide coverage, and the MSO provides coverage in pockets (e.g., in urban dense areas). The network architecture diagram for such a deployment scenario is shown in the figure below where the MNO is providing 4G including 5G NSA coverage and the MSO has deployed 5G SA in NR bands in its pockets.

[5] https://gsacom.com/webinar/the-5g-story-so-far-5g-spectrum-networks-and-devices-in-1h-2021/

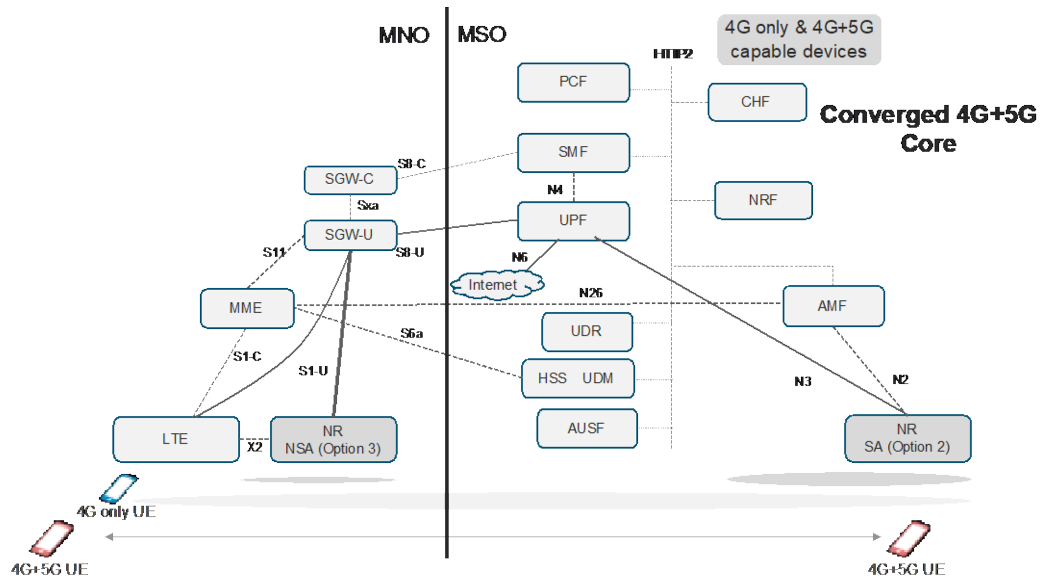

Figure 8: Demarcation of MNO and MSO networks in partnership when single-SIM is used.

The key interworking interfaces between the MNO and MSO are those that cross the thick black demarcation line between the MNO and MSO. The S6a interface between the MNO's MME and the UDM+HSS of the MSO is used to authenticate the MSO's UE when that UE is in MNO's coverage area. The S8-C and S8-U interfaces between the SGW-C and SMF and SGW-U and UPF respectively, are used to anchor the IP connection of the mobile in the MSO's SMF and UPF (this is also called home-routed scenario in 3GPP specifications). The reason for home-routing is that primary service policy and charging are provided by the SMF and PCF in the home-network of the UE, i.e., the MSO. The N26 interface between the MME and AMF is used to transfer contexts between the MNO and MSO network to enable seamless mobility.

Though 3GPP specifications exists that provides all the details of the call-flows to support the above architecture, the main challenge is in deployments of these inter-operator interfaces. The vendors providing the network functions on the two sides of the demarcation lines may be different and may require extensive inter-operability testing. Also, the interactions between the radio networks to handovers from the MNO network to the MSO network when a mobile is in the coverage of the MSO's radio network and vice-versa require quite a lot of radio planning (e.g., configuring neighboring cell information and updating these as the radio network of the MNO or MSO changes).

To mitigate some of these challenges, a DSDS based approach could be leveraged as described very well in Cable and Mobile Convergence, A Vision from the Cable Communities Around the World[6] and Section 4.1.5, “Dual-SIM”.

[6] https://www.nctatechnicalpapers.com/Paper/2020/2020-cable-and-mobile-convergence

4.3. Single vs. Multivendor strategy – Open Interfaces

While 3GPP provides a very well debated and industry vetted architecture with the operator, vendor and entire ecosystem contributing, it sometimes is unable to accommodate every single deployment scenario and use case. Especially, when it comes to requirements driven by MSOs, since the use case requirements and deployment models are not always aligned with established MNOs in the market, there could be a need to extend of enhance the capabilities for MSO deployments. Given that 3GPP and other SDOs have a set schedule for completing the documentation of specifications and priorities, it is likely that timelines for deployment of the operator may not align

This typically drives the MSOs to consider custom implementations. Take a scenario where an MSO would like to identify a customer not just via mobility identity but based on wireline identity or a cable identity. If leveraging 3GPP Release 15 specifications, this is not possible today without custom adaptation as 5WWC work item in Release 16 would add some of these capabilities. In these scenarios, an MSO may consider adding a Cable Line ID as a custom attribute across the nUDM service and on nChf service to identify the mobile user and tie into the Cable Service.

A similar capability was leveraged in 4G LTE networks extensively – especially on diameter interfaces. This was typically achieved via vendor defined AVPs or vendor specific AVPs. Additional details are captured in the Diameter RFC 3588, 5.3.3. Vendor-Id AVP. While this was a well understood mechanism, due to competitive scenarios, it is not always viable and possible to take this custom vendor defined AVPs to SDOs. In such scenarios, it may be required to continue supporting custom implementations which would require ensuring interoperability across vendor products.

Since 5G Service Based Interfaces support extensions or information elements that could added beyond the specifications, this could be leveraged by the MSO to achieve the use case and faster go to market option without having to wait for 3GPP or other SDO. Invariably, such custom behavior creates a challenge during software upgrades or interoperability across different vendor solutions.

We recommend that operators need to ensure that all deployment interfaces are open and compliant to 3GPP specifications. This is typically achieved via vendor-to-vendor interoperability testing either from a lab to lab or within the MSO labs. Any extensions or custom attributes implemented as described above by any vendor specific to the MSO network are recommended to be fully documented and more importantly, should be protected by a feature capability exchange and only utilized when the peer node indicates support.

- Option 1: Implement feature capability exchange: The recommended way is to ensure the initiator of the message that includes the custom information does so only once it has been established that the receiver can gracefully handle the custom information provided. This could be achieved in multiple ways in 5G network deployments.
  - In case a client / server relationship is established – leverage feature capability during initial connection establishment

    - Leverage NRF capability exchange to communicate and register feature versions which include the custom attribute supports and only provide the producer information when the client specifically requests for this capability

- Option 2: Implement configuration options: This option provides an easy implementation choice by only including the custom attributes if specifically configured to do so. This removes the complexity of feature capability exchange or relying on NRF but poses an operational challenge of making sure every NF is appropriately configured and upgrades are sequenced and managed in a controlled manner
- Option 3: Ignoring uninterpreted attributes and default behavior: Alternate to a feature capability exchange or configuration is ensuring that the information passed through the custom attributes is ignored when a receiver is not able to interpret it and ignoring such information does not impact service. This is critical even if one of the additional options is implemented to avoid any service disruptions to minor changes in software releases or attribute parameters. The receiver should log an error so the issue could be debugged and addressed appropriately but while this is underway, network service is maintained without disruption, albeit with likely not all capabilities.

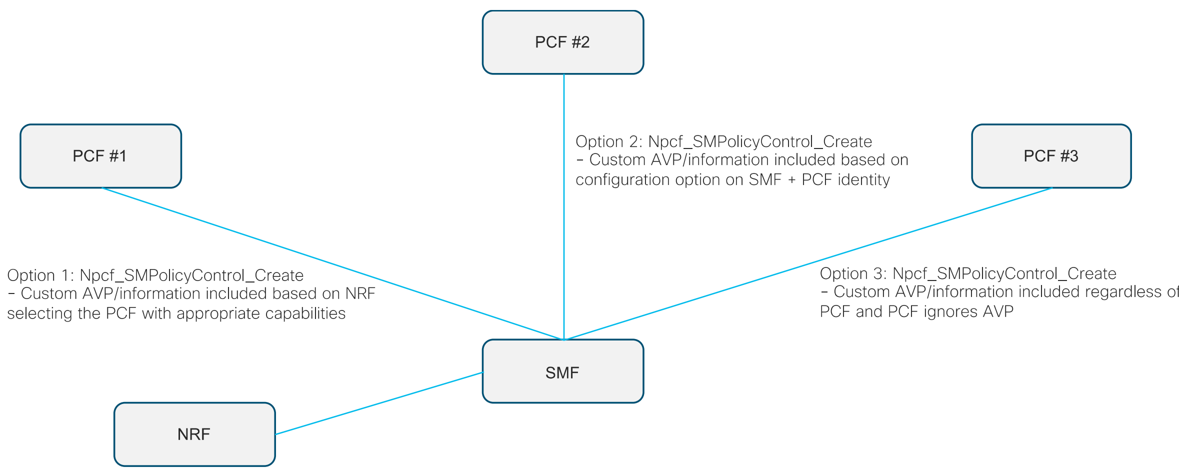

Figure 9: Options to achieve interoperability with custom implementation

We believe that at the end of the day – 3GPP compliant interfaces are the best way to deploy networks. The reason for recommending this is to ensure potential introduction of additional vendors in the mix. As an example, to achieve certain tracking area mapping capabilities, it was required in initial days of 4G LTE RAN ecosystem to implement custom behavior on the S1-AP interface between eNodeB and MME. However, it had eventually gone to a point where it was almost impossible to interwork an eNodeB from the RAN vendor with any other vendor's MME without impacts to service or capabilities which created a "lock in" problem for the operator and not being able to deploy best of breed network functions or a dependency on the operator introducing new services as the radio provider feature acceleration may not be at par with the operator requirements or a market specific mapping of RAN and MME i.e., tight dependencies cause operator to define a clear boundary of operation for radio and packet core

As an example, a radio vendor that has the radio infrastructure deployed in one region was able to only utilize MMEs from that vendor. During a system outage or migration or other similar scenarios, the operator is not able to leverage the additional capacity that is available from a peer market or region which is provided by another Radio vendor. Especially, the use of ASN.1 as a protocol between Radio and MME decreases the ability to interwork once custom attributes are introduced.

Similar rules apply to the non-radio deployment scenarios but due to the ability to extend diameter or GTP protocols without causing interoperability issues based on proposals above has resulted in somewhat of a manageable model within the packet core. Any deviations from common practices and not documenting could result in an interoperability issues and delay in service launch or disruption to existing services. Relying on open interfaces for scalable network architecture and debugging across network functions is critical for long term success of the solution.

4.4. Policy and charging architecture

The sheer number of different data-plans that are offered by MNOs is an evidence of the complicated policy and charging rules and traffic measurements that are performed in MNO networks. Specialized application charging rules, such as zero-rating of Netflix or Spotify traffic, require significant DPI and application detection capability in the network.

3GPP PCC architecture example for 5G System is depicted in the figure below and with all the options supported, the deployment could be daunting with ability to create policies across multiple service enablers in the network and ensuring policies do not contradict each other in this scenario. A well-defined PCC rule can help with detection of a service data flow and providing parameters for policy control and/or charging control and, for PGW enhanced with ADC, for application detection and control.

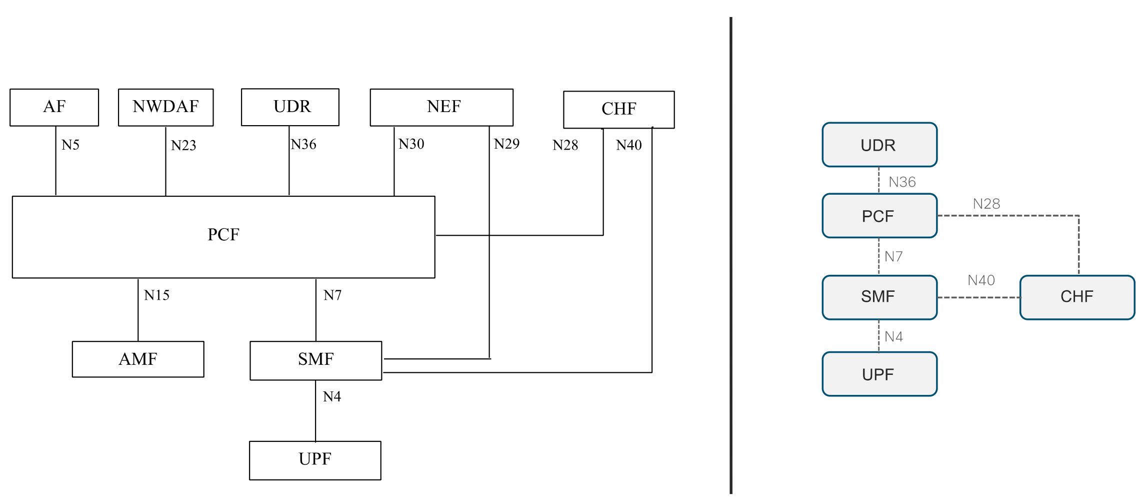

Figure 10: 3GPP PCC Architecture for 5G (left). Simplified PCC architecture for MSO (right).

There are two main drivers for the complex policy and charging infrastructure in MNO networks:

1. Requirement to support dynamic QoS, e.g., for providing periodic scheduling with low latency for voice traffic. (NOTE: This is not typically for charging perspective, since most MNOs are not monetizing voice traffic), and
2. Data caps and rate-enforcement required either for on-line charging (pay-as-you-go) or for offline charging (monthly payment) data-caps per month.

MSOs can simplify these by not supporting voice traffic natively. Not having to support voice or other traffic that requires dynamic signaling of filters and QoS rules at least does not require to support the N5 interface between the PCF and the application function (AF).

MSOs should look at only providing offline charging and not have to implement the more complicated on-line charging model. Typical MSO customers have at the minimum monthly subscriptions with them. Trying to support the more complicated on-line charging (pay-as-you-go) model will simplify the real-time requirements for policy and charging.

With the above simplifications, the only rules that an MSO needs to look at is enforcing are data-cap rules that are simpler to define and maintain. This design would align with similar billing and accounting capabilities offered within the Cable access for end users today in most networks. However, if the MSO chooses to offer more aligned capabilities as MNO, this will require similar PCC architecture and complex rule definition as done on the MNO network today.

Number of policies deployed could have impact on the overall performance of the system. When PCEF needs to perform Deep Packet Inspection or process policy while handling small packets, the overall system capacity could be impacted. Any DPI needs to be accounted for when deploying the static and predefined rules being installed on the gateways and only service / revenue impacting rules should be considered. Changes in policy based on specific locations require the PCF to be aware of the current location of the subscriber and in a mobility intensive environment, this could generate signaling all the way from the RAN, MME/AMF, SGW, PGW/SMF and eventually to the PCF causing massive signaling. Smart Phones tend to create Service Request in the network at an average of 2 mins and any location changes triggered during service request could generate unnecessary signaling.

Location based policy should be higher level granularity e.g., Presence Reporting Area or regional boundaries instead of individual RAN locations. Any rules installed by the PCF should be considered for the depth of the packet inspection required and impact to the system performance.

Designing for a simplified but robust policy, charging architecture and rule definition would ensure a seamless service experience for the end users and enabling new capabilities in the network with limited planning.

4.5. Device Behavior

When a network function or node is not able to handle the messages or is in maintenance mode, it is important that the downstream peer functions are aware of this. It is essential that the operators define a clear expected behavior within a client for error codes received from peer nodes and act in a graceful manner. As an example, if the PCF is unable to handle the message due to a session that is not found, the error code provided towards the SMF should provide enough information for the SMF to either initiate the session recovery procedures or gracefully disconnect the user and reestablish the session. While some of the network error behaviors could be gracefully handled or at least addressed with software updates, during the initial days of LTE launch, a critical lesson learnt was on the unexpected UE behavior.

As LTE networks were deployed, it became apparent that the error codes were not interpreted the same way, despite some documentation in 3GPP. This resulted in unpredictable behavior during error scenarios, but also resulted in non-uniform behavior in different markets or geographic regions of the same operator where different vendor solutions were deployed.

This was eventually addressed in LTE with much more prescriptive behavior expectations on error codes at the device (Reference: TS 24.301[13]) but also a massive effort to map error codes from every diameter interface to eventual error code mapping at the device. While some of these are based on lessons learnt from real deployments, some were mere clarifications of predefined behaviors in 3GPP.

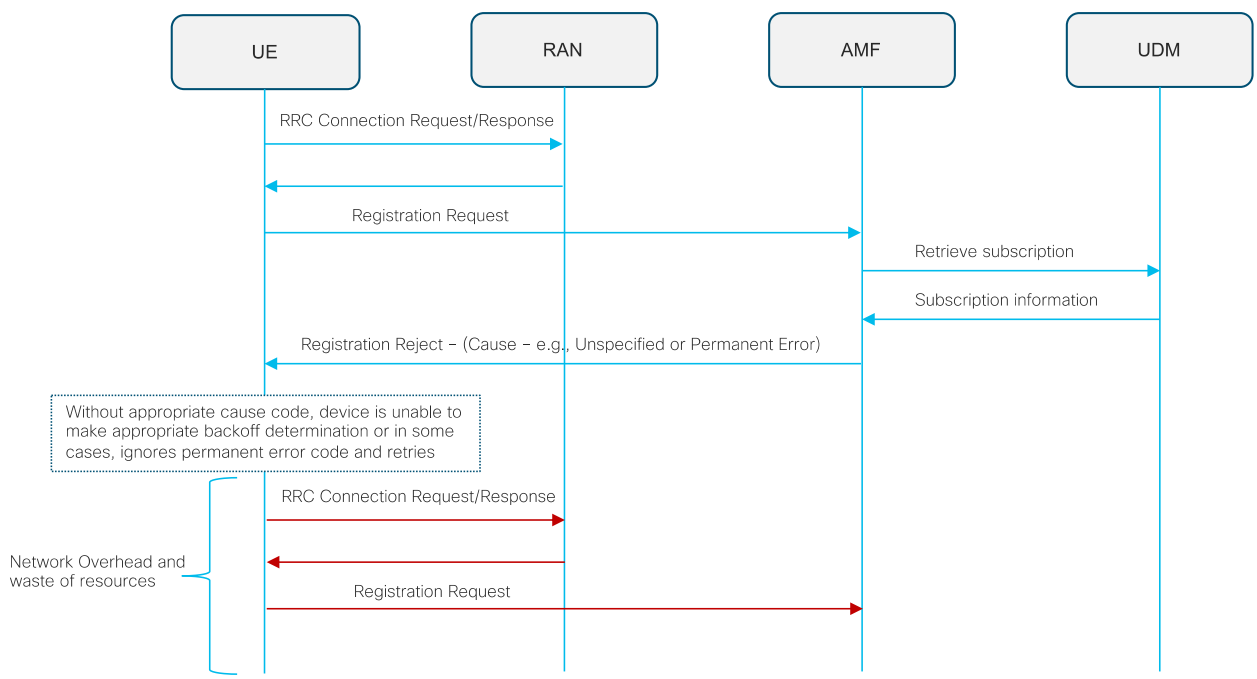

Figure 11: Call flow depicting UE errors and retries

Fortunately, most of this work has been forward ported into the 5G specifications already and similar capabilities as mature 4G networks exists already in 5G networks. However, it is imperative to plan for previously unseen behaviors in the network. Especially, new capabilities and interfaces in 5G defined around Service Based Architecture, slicing capabilities on devices and additional new functionality defined in 5G – it is very likely unforeseen scenarios will occur in the network.

It is critical to also note that the error scenarios are not just to be UE focused but rather a network focused approach as well. Some of these could be avoided by:

- Formal interoperability testing as described in previous sections or relying on statements of compliance to begin with
- Adding intelligence into the network to decommission or remove certain software functions in the network based on error trends
- Configurable error code behaviors

Accounting for unforeseen device behaviors and designing for graceful recovery mechanisms would help ensuring a reliable end use experience.

A couple of good examples of what has been achieved in 4G networks and positive outcomes of lessons learnt could be based on the overload scenarios or race conditions in the network.

Race conditions have typically occurred when the network and device are attempting conflicting procedures which are unfortunately not completely preventable but by adding message priorities and sequenced queues, the impact of these race conditions could be mitigated to a certain extend.

For overload scenarios, based on ecosystem and industry discussions, IETF work on overload control (Reference: https://datatracker.ietf.org/doc/html/rfc8582 [12]) or GTP-C load / overload control was introduced into 3GPP specifications. It should be noted that these are not widely leveraged today, partially due to the timing of availability of these capabilities in the specifications. By the time these specifications were available in 4G, there was an industry momentum within 5G and focus shifted towards 5G software development and implementation in the field. However, most of these capabilities have been made available with 5G specifications from day 1 and so could be leveraged as needed from initial deployments for MSOs.

One additionally way of ensuring device vs. network race conditions or duplicate messaging is avoided is by fine tuning the timers in the network and the message queues with validity timers / expiry tags for messages so that any messages stuck in queues are discarded after a certain time. In some rare scenarios, if a message from the queue is processed after the device internal timers have already expired and a procedure from the network is executed, this could result in out of band processing of the messages resulting in incorrect and unexpected behavior from the devices.

While there may be some reluctance to use some of these advanced capabilities in the specifications to begin with, we believe that having the capabilities planned in the network from the initial get go would be important.

5. Other topics for consideration

While not captured in detail in this document for brevity, it is critical for MSOs to take additional topics into consideration as listed below.

1. **Virtualization Stack**: Every MSO may choose to have a platform of choice – which may or may not be driven directly by the 5G deployment strategy. It is likely that some MSOs have already charted a path for virtualization of Network Functions in the existing deployment and service strategy. In some ways, the same platform could be the entry path for 5G Core cloud-native NF's or implementation of Open vRAN architectures that are critical for success of 5G. Depending on the capabilities of the MSO, operational focus within the company, a CaaS and PaaS capability may be under evaluation or already in deployment which could be leveraged or any existing investments on NFVI could be leveraged. Given that this is a choice for the MSO which could be influenced by other factors in the company – additional details are not discussed in this document.
2. **Convergence architectures**: 3GPP has defined the baseline architecture for convergence of wireline and wireless services as part of the 5WWC work item in Release 16 specifications. However, with additional dependencies being addressed in CableLabs and the detailed analysis and discussing in Cable and Mobile Convergence, A Vision from the Cable Communities Around the World[7], this document does not address convergence but could be a critical decision point for the MSOs before finalizing the 5G deployment architectures.
3. **Support of Voice Services:** Given the stringent performance and regulatory requirements for voice, in a deployment model where MSO partners with an MNO (see previous sections), it could be simpler for the MSO to have the MNO support voice services whereas the MSO provides internet connectivity as native voice support and client capabilities are made available specific to the MSO architecture. One challenge of achieving voice services delivered via the MNO partner with data services from the MSO native network is that either the device will have to support dual SIM capabilities or when the voice network of MNO is leveraged, the data is offloaded to the MNO partner network as well while the device is in a voice call
4. **Support for lawful interception**: Lawful interception is a required functionality by regulatory bodies in most countries where mobile networks are deployed. As cable operators consider their journey from MVNO to MSO, they will need to plan and deploy lawful interception. Most vendors who provide infrastructure equipment support lawful interception on their network function. Also, MSOs will need to create a system for handling lawful interception request from law enforcement agencies in their jurisdiction. Since information about targets of lawful interception needs to be guarded very closely in an operator's network, this involves only enabling a select few operation staff to have access to this information and for configuring and monitoring the system for lawful interception.

[7] https://www.nctatechnicalpapers.com/Paper/2020/2020-cable-and-mobile-convergence

5. **Roaming partnerships**: Roaming architectures are very well defined in 3GPP but most importantly a clear cross operator engagement model defined by GSMA to enable global roaming. It is recommended that MSOs leverage the existing guidelines of GSMA[8] for 5G being developed. One of the important considerations for the MSO as they finalize the deployment model is the any dependencies related to anchoring devices from 4G networks in countries where 5G is not yet available. Using a converged core architecture could simplify the operations and supporting different roaming models as described in Section 4.1
6. **RAN considerations**: This document does not cover RAN architectures in detail as there has been a significant amount of collateral and analysis on the various RAN deployment architectures. Please refer to Security Benefits of Open Virtualized RAN9 or additional standardization details as part of O-RAN alliance10. O-RAN could be discussed as a standalone document alone.
7. **SIM profiles and Device Management**: SIM profile development and device management capabilities could be driven by the operator choice of device partner ecosystem, compliance to OMA-DM specifications and partnerships with specific UICC provider itself. Given that the SIM profiles and UICC management are usually not disclosed publicly due to security concerns, this document does not specifically capture the lessons learnt from 4G deployments with respect to device management. However, this would be a critical area of investment for the MSO to ensure ability to configure new APNs/DNNs within the network and subsequently provisioning the device with the capabilities or being able to manage the MAPCON profiles on the device for WLAN vs. 3GPP selection. Note that with 5G, given the capabilities of ANDSF from 4G networks are now integrated with the PCF, we anticipate that once the devices start supporting the capability, network selection would become simpler to manage in 5G.
8. **Network Failures and Recovery**: Network failure scenarios are inherent in the regardless of the architecture and stability of the network functions and their underlying infrastructure. In the move to cloud infrastructure, this is even more prevalent as the virtualization infrastructure typically does not achieve more than 99.9% redundancy. To achieve a 5x9s service or greater, the redundancy of the network functions must be increased to support more common cloud infrastructure failures. Care needs to be taken to address:
    - Network Function Failures
    - Software Upgrades and Configuration Failures
    - Blast Radius of a Failure

[8] NG.113 5GS Roaming Guidelines v4.0 @ https://www.gsma.com/newsroom/resources/ng-113-5gs-roaming-guidelines-v4-0/
[9] https://www.cisco.com/c/dam/en/us/solutions/service-provider/pdfs/5g-network-architecture/white-paper-sp-open-vran-security-benefits.pdf
[10] https://www.o-ran.org/

6. Conclusion

We acknowledge and understand that every MSO has different reasons for investing in building a new Mobile Network. Whether it is to reduce the cost structure in offering mobile services by offloading as much traffic as possible from the MVNO partners network or to compete with differentiated services offering in the market, MSOs now have a unique opportunity to leverage new spectrum assets to do so. With 5G architecture defined in 3GPP and other SDOs enabling convergence capabilities in the future, investing in 5G to build a standalone network could be ideal in some cases but due to regulatory requirements or device ecosystem dependencies or roaming opportunities, there may be a need to support 4G / LTE networks as well.

As MSOs explore the option of natively building a greenfield network it is critical to ensure some of the lessons learned by MNOs while building the 4G networks or during the launch of VoLTE networks should be leveraged to ensure the new 5G networks by MSOs do not have the same challenges.

As described in the various sections, the following would be critical considerations, though not a comprehensive list.

1. Partnership scope with MNO partner or MVNO partner to ensure ability to influence policy across the network for seamless experience for the end user regardless of which network (MNO vs. MSO greenfield) the user device connects from
2. Augmenting voice and required regulatory compliance capabilities while the native coverage by MSOs is being expanded / built
3. Ensuring open interfaces across the network architecture and functions – to be able to leverage best of breed architecture vs. a single vendor strategy risking potential "lock in"
4. Reducing the complexity of policy and charging architecture but still being able to address inline service requirements to offer similar or same capabilities as the partner network
5. Expect that even with wide interoperability testing and device compliance, there will be a scenario where one device firmware/model may misbehave and with the volume of devices, this could quickly snowball into a massive network challenge. Preparing network with various mitigation capabilities when such device behavior is encountered could reduce subscriber impact and reduce operation costs
6. Design for various potential race conditions and network failure scenarios and fallback mechanisms from day 1 and ensure mobility is accounted for as devices connect from alternate access types or locations during a mass failure or pockets of failures in the access/core network

With the extensive knowledge that the MSOs have gathered over the years with existing networks and the interest to offer differentiated service to the end users, we believe that MSOs are at a cusp of changing the dynamics on wireless/wireline networks as a whole and the entire Service Provider model. Taking into consideration the variety of challenges that mobility and wireless pose and leveraging best practices will make this transition smoother and more successful.

7. Appendix A

A complete set of 3GPP network functions as defined in Release 15 and Release 16 specifications is listed below. The authors believe that while these capabilities are eventually required to be able to offer a wide set of use cases and completive solutions, not every function is required from day 1 of the deployment. It is possible to introduce most of the functions into an existing Greenfield network after launch without disruption of service or a major redesign. Some functions, like BSF or SCP may need some consideration before launch and planning for the future.

Table 1: Complete list of the 3GPP defined network functions

5G-EIR	5G-Equipment Identity Register
AF	Application Function
AMF	Access and Mobility Management Function
AUSF	Authentication Server Function
BSF	Binding Support Function
CAPIF	Common API Framework for 3GPP northbound APIs
CHF	Charging Function
ePDG	evolved Packet Data Gateway
GMLC	Gateway Mobile Location Centre
LMF	Location Management Function
LRF	Location Retrieval Function
N3IWF	Non-3GPP Interworking Function
NEF	Network Exposure Function
NR	New Radio
NRF	Network Repository Function
NSSF	Network Slice Selection Function
NWDAF	Network Data Analytics Function
PCF	Policy Control Function
(R)AN	(Radio) Access Network
SEAF	Security Anchor Functionality
SEPP	Security Edge Protection Proxy
SMF	Session Management Function
SMSF	Short Message Service Function
UDM	Unified Data Management
UDR	Unified Data Repository
UDSF	Unstructured Data Storage Function
UPF	User Plane Function

Abbreviations

3GPP	Third Generation Partnership Project
5GC	5G Core
5GS	5G System
AF	Application Function
AMF	Access and Mobility Management Function
AP	Access Point
APN	Access Point Name
AUSF	Authentication Server Function
CBRS	Citizen broadband radio service
CDR	Charging Detail Records
CHF	Charging Function
CPE	Customer premise equipment
CUPS	Control plane and user plane separation
DN	Data Network
DSDS	Dual-SIM Dual-Standby
eMBB	Enhanced mobile broadband
EPC	Evolved Packet Core
ePDG	Enhance Packet Data Gateway
EPS	Evolved Packet System
eSIM	embedded SIM
FWA	Fixed wireless access
LTE	Long Term Evolution
MCC	Mobile Country Code
MNC	Mobile Network Code
MNO	Mobile Network Operator
MSO	Multiple System Operator
MVNO	Mobile Virtual Network Operator
NEF	Network Exposure Function
NF	Network Function
NFV	Network Function Virtualization
NG-RAN	Next Generation RAN
NR	New Radio
NRF	Network Function Repository Function
NSA	Non-Stand Alone
NSSF	Network Slice Selection Function
NWDAF	Network Data Analytics Function
OCS	On-line Charging System
OfCS	Off-line Charging System
PCF	Policy Control Function
PCRF	Policy and Charging Function

PLMN	Public Land Mobile Network
SA	Stand Alone
SBA	Service Based Architecture
SBI	Service Based Interface
SMF	Session Management Function
SMSF	Short Message Service Function
SEPP	Security Edge Protection Proxy
RAN	Radio Access Network
UE	User Equipment
UPF	User Plane Function
UDM	Unified Data Management
UDR	Unified Data Repository
URLLC	Ultra-Reliable Low Latency Communication
VoLTE	Voice over LTE
VoNR	Voice over NR
vRAN	Virtual RAN
WWC	Wireless-Wireline Convergence

Bibliography & References

[1] 3GPP TS 23.501, System architecture for the 5G System (5GS)
[2] 3GPP TS 23.502, Procedures for the 5G System (5GS)
[3] "5G Implementation Guidelines: SA Option 2", June 2020, GSMA
[4] Wireless and Wireline Convergence for the 5G system architecture, 5WWC, WI # 830050 - 5WWC
[5] 5G Wireless Wireline Converged Core Architecture Technical Report, CableLabs
[6] Jennifer Andréoli-Fang, PhD, CableLabs; John T. Chapman, Cisco et al, Cable and Mobile Convergence: A Vision from the Cable Communities Around the World
[7] The 5G story so far: 5G Spectrum, networks and devices in 1H 2021, Global mobile Suppliers Association
[8] NG.113 5GS Roaming Guidelines
[9] Eric Hanselman, Security Benefits of Open Virtualized RAN
[10] O-RAN Architecture Description, O-RAN Alliance
[11] O-RAN Use Cases Detailed Specification 5.0
[12] Diameter Overload Rate Control, RFC 8582
[13] 3GPP TS 24.301, Non-Access-Stratum (NAS) protocol for Evolved Packet System (EPS); Stage 3
[14] 3GPP TS 32.255, Telecommunication management; Charging management; 5G data connectivity domain charging; Stage 2

------ End of Document ------