5G RuralDorset WP6 Neutral Host Task 2:

Security and Privacy

Authors: Dave Happy, Greig Paul, Ross McPherson
Activity lead: Telint Ltd.
Supporting partners: Neutral Networks, Strathclyde University, Satellite Applications Catapult
Delivery date: 31/03/2022

Executive Summary

Security concerns have previously been identified as one of the key barriers to adoption and uptake of rural neutral host (NH) solutions. Operators are concerned that their network may be more exposed to compromise as a result of introducing a complex new architecture and additional equipment to their network, and there are potential aggregations of risk through the interconnection of mobile operators together, closer to the edge. Similarly, mobile operators have contractual obligations around the confidentiality of their customers' information and may also wish to protect information beyond purely traffic data, from a commercial reputation perspective.

One inherent challenge with a neutral host with active equipment sharing is that it involves the sharing of radio access network equipment. This equipment is distributed throughout the mobile network, at less physically secure locations towards the network edge. This means that a multi-layered approach to securing RAN and NH equipment should be taken to protect against the assumed compromise of a site in the network. This is good practice in any case and should be done in a mobile network as a matter of course.

In this report, we explore the requirements and expectations for security in mobile networks, how this is delivered through standard deployment architectures, and the threats faced by mobile operators today. We then consider the emerging and future threats which will be, or are likely to be, faced by mobile operators in the medium to longer term, as a result of wider industry trends. By then exploring a working model for neutral hosting, through the JOTS specification for in-building neutral host, we consider the security implications of rural neutral hosting.

Throughout the document, where threats are identified, potential technical or organisational solutions or enablers are identified. These would be applicable to any mobile operator, local or national, and are summarised below:

- Adoption and uptake of "zero trust" architectures, to reduce the "blast radius" of compromise of any individual component of the network, such that an attacker gains little or no benefit from compromising distributed components of the network.
- Use of robust edge inter-connections between operator networks, with strict traffic filtering. There is a valid concern that a neutral host would "bridge" multiple mobile operators' networks. To some extent, this can be mitigated through physical isolation and use of hypervisors to provide relatively high assurance of isolation of per-operator and per-vendor functions. Operator networks are designed for cross-network peering to take place in the core, and suitable near-edge filtering would need to be in place to enable neutral hosting to interconnect operators' radio access networks (RANs).
- Strong visibility and management of inter-radio interoperability protocols between NH and operator RANs, with verified and robustly tested interoperability protocols (such as X2 and Xn), in order to enable functional and performant handovers between operator and NH radios, without exposing vulnerable implementations to an external network provider.

- Robust isolation of management planes of all protocols across NH and operator RANs, preventing lateral movement and managing "upstream" – if Open RAN and other architectures gain adoption, this will be required to secure an operator RAN. In a NH context, proper isolation of each operator's management and OAM (operation and maintenance) interfaces will be required, alongside robust firewalling and monitoring of traffic, to prevent one compromised radio from being used to attempt to manage or control another radio, or cross into an adjacent network (such as an MNO or NH).
- Technical work should explore the real-world interoperability between radios on the X2 and Xn interfaces, to support performant handovers in the real world. Operators deploy their radios in a vendor-zoned manner in "RAN zones" and a NH operator is almost certain to be deploying a different radio vendor's equipment. In a rural NH context, the ability to deliver seamless handovers into each operator's RAN would be desirable from a user experience perspective, but the security and functionality of interoperability on these interfaces will need to be explored. In particular, strategic reverse engineering (to enable functional interoperability) of wire protocols may be required, in order to support neighbour discovery, neighbour advertisement, and cell-edge interference coordination functions. Without these, the RAN-to-NH interface is likely to have poorer performance, and perhaps lack seamless mobility.
- Further technical work to enable wider use of VoLTE/VoNR voice roaming in the UK, and enhance the ability for UK users to switch operator while preserving VoLTE calling. To enable NH networks to be built without 2G/3G legacy voice telephony integrations and allow existing 2G/3G networks to be sunset. VoLTE/VoNR support must be available for UK handsets, and inbound roamers. This additionally needs work to explore lawful intercept capability impact in a NH environment, including for data traffic if a NH operator provides internet egress capabilities. This could be explored through regulatory options to ensure all UK operators use a single IMS/VoLTE carrier bundle profile, as this would enable users to switch mobile network without losing IP voice capabilities.

After noting the upcoming security challenges national mobile operators will face, similarities between these challenges, and the challenges of supporting a neutral host architecture, were detailed. This suggests that implementing many of the security protections, primarily to adhere to upcoming regulations, and secondly if they want to take advantage of the latest technological advancements such as Open RAN, would actually have a third benefit of positioning them well for supporting a neutral host environment, provided there was a suitable business incentive, as discussed in the previous business study report in this series[1].

[1] https://5gruraldorset.org/app/uploads/2021/12/5G_Rural_Dorset_WP6_Task5_Business_Study.pdf

Contents

Introduction

This report explores the security implications of neutral host and related technical implementations. When considering security in a telecoms network context, it is important to be clear in scope around what is meant by security, and the perspectives through which it is considered. Generally, in information and systems security, the principle of **security** is broken down into the so-called "CIA triad" - **confidentiality**, **integrity,** and **availability**.

Confidentiality refers to the principle of communication between two entities in a system being protected against another entity seeing the contents of it. This can be considered where entities in the system are people or their devices (i.e. user equipment), or where entities in the system are network functions or components (I.e. traffic between a radio and the core network).

Integrity refers to the principle of communication between two entities in a system being protected against undetected modification or tampering with the contents of the message. This can again be considered from the perspective of an end user or the network operator.

Availability refers to the more general concept of the telecoms system being available for use – delivering the expected service reliably, so that it can be used when required. A successful denial of service attack can cause a loss of availability, as could a significant reduction in performance, resulting in an inability to deliver the expected level of service.

From a network operator's perspective, user traffic over external backhaul links (i.e. services like BT's MEAS offering[2]) requires **confidentiality**, to prevent attackers from seeing users' data. European operators tend to implement IPSec based connections between base stations and a Security Gateway (SecGW) to provide this level of **confidentiality. Integrity** protection is also achieved through authentication algorithms in IPSec. **Availability** is provided through sufficient network capacity and redundancy. This would commonly be ensured through service level agreements (SLAs) than a defined technical solution.

This division between purely technical security and practical business decisions is required for all useable systems. Professor Gene Spafford, a leading computer security expert once said "The only truly secure system is one that is powered off, cast in a block of concrete and sealed in a lead-lined room with armed guards - and even then, I have my doubts." Of course, such a computer would not be practical to use, therefore, there is a level of compromise needed between security and practicality for a business to run effectively.

The latest generation of 3GPP mobile standards, commonly known as 5G, presents marketing headlines of higher throughput, greater reliability, lower latency, and increased vendor diversification; but each of these improvements requires architectural and design changes.

One such functional change is the transition from bespoke hardware implementations to softwarization of components running on commercial off-the-shelf (COTS) hardware. This transition began in 4G, but the switch to the Service Based Architecture (SBA) employed in the 5G core, enables the use of virtual network functions (VNFs), and technological advancements such as cloud architecture, virtualisation, orchestration and micro-services.

[2] https://www.btwholesale.com/assets/documents/why_bt_wholesale/Infographic_Why_BT_Wholesale.pdf

This switch to the softwarization of components dramatically reduces the barrier to entry for new vendors entering the market, but it also elevates security concerns of software supply chains and interoperability. Similarly, low latency advancements require operators to deploy additional hardware increasing the threat surface of their network. In fact, many of the advantages of 5G networks come hand-in-hand with new security concerns. Incumbent mobile operators will seek to minimise these security concerns while continuing to deploy new features with the aim of attracting new users.

Many of these security concerns, are reminiscent to the security concerns of incumbent mobile operators sharing infrastructure with a Neutral Host provider, therefore, the authors will detail the similarities between the likely steps which will be taken by incumbents and the security requirements which new operators will likely have to follow.

Current Threats in the Mobile industry

Threat model of Mobile Network

In the hierarchy of sensitivity and significance of secrets in a mobile network, some of the most commercially expensive parameters to change would be the long-term SIM card security keys (called Ki and OP/Opc). These keys are symmetric "shared secret" keys, where a copy of the key is present on the SIM card, as well as in the ARPF (Authentication Credential Repository and Processing Function) network function, for use by the AuSF (Authentication Service Function)[3]. This permits the core network to authenticate a handset, ensure it is legitimately matched to a subscriber, and route and bill calls, SMS and data appropriately. If these shared keys were compromised, a malicious attacker could decrypt all the `over the air' messages for that mobile network, and accordingly an operator would have to re-issue SIM cards to each of the affected users. Changing these keys is possible in a remotely provisionable SIM card; however, once they are compromised, issuance of a new SIM through out-of-band channels would be necessary, as the keys used to secure such an update would no longer be secure. This may require posting a new physical SIM card to users or issuing an updated eSIM for users with eSIM-enabled handsets.

To reduce the possibly of compromise, in a 3GPP network, these long-term symmetric keys are not shared with other network functions – the ARPF holds long-term keys, and computes authentication vectors derived from the long-term symmetric keys. This process of key derivation is carried out according to 3GPP standards[4], and takes into account factors such as the Public Land Mobile Network (PLMN) that a user is connecting to. So, for example, in a roaming scenario, authentication vectors are unique to the PLMN the user is roamed to. In a 4G network, keys given to eNodeBs are derived from their Physical Cell ID (PCI), and frequency/configuration (DL-EARFCN), and K_{ASME} is derived from the identity of the serving network (i.e. the PLMN ID, or MCC/MNC pair, of the core network serving the user)[5]. This means that the symmetric cryptographic keys which are exposed to network functions are per-user, scoped and constrained to the cell in question, and cannot be used in other ways. For example, the K_{ASME} key issued to a roaming cell in another PLMN would not be usable to create a "fake" home radio, as the key would be incorrect, since it was derived for a different MCC/MNC pair.

A mobile operator will inherently be concerned by any business arrangement which results in them exposing home network keys to a third party, as this would allow their customers to attach to cells operated by the third party, unless these are carefully managed and controlled.

[3] https://www.gsma.com/security/wp-content/uploads/2021/11/T-ISAC-5G-Security-Nov21-Jyrki-GSMA.pdf
[4] https://www.etsi.org/deliver/etsi_ts/133400_133499/133401/16.03.00_60/ts_133401v160300p.pdf Section 6.2
[5] https://www.sstic.org/media/SSTIC2016/SSTIC-actes/how_to_not_break_lte_crypto/SSTIC2016-Article-how_to_not_break_lte_crypto-michau_devine.pdf Section 1.2

Privacy and Tracking

One area of concern within the increasingly connected world is the tracking of users. This has been done for a number of years through tracking the International Mobile Subscriber Identity (IMSI), which is a unique and permanent per SIM card-based identifier, broadcast in plaintext when communicating over the mobile network. Therefore, within a localised environment, custom IMSI catching hardware, commonly known as stingrays, can detect which IMSIs are in range and track that user of the phone.

Tracking can also be achieved network wide when coupled with specific mobile operator information, for example, using the ID number of the cell the user is currently connected to, and knowledge of the location of the cells, a user could be roughly tracked throughout the country. This information can be extracted unknowingly from the operator through compromising individuals, hacking computer systems, or crowdsourcing information. The information can also be knowingly sold to create an additional revenue stream for the mobile operator, as demonstrated in the United States, where mobile operators even used their privileged position to sell their consumers' real-time location data to third party companies[6].

Legacy signalling technology employed in 2G/3G networks, known as Signalling System 7 (SS7), additionally offers the ability for malicious attackers to extract user location information without the operator's knowledge. SS7 is the internally standardized telecommunication protocol for communication of data and control signals within a public switched telephone network (PSTN). However, limitations in the design have allowed for individuals to be tracked while roaming aboard [7].

Although the 3GPP were aware of these issues, they persisted through the mobile generations to preserve backwards compatibility. One of the advancements in 5G SA, which aims to improve this situation, is the use of the Subscription Permanent Identifier (SUPI) to replace the IMSI as the globally unique subscriber identity. The main benefit of SUPI over IMSI, is that it is never transmitted over the air in plaintext. Instead, a SUCI (Subscription Concealed Identifier) is transmitted using one of three transmission schemes. Protection Scheme Identifier 1 which does not conceal the identity of the user, but Protection Scheme Identifiers 2 & 3 which employ an Elliptic Curve Integrated Encryption Scheme (ECIES), encrypting the SUCI, which is then mapped to the SUPI to know which user is connecting.

Legacy Systems

The UK Government and MNOs have announced their intention to sunset all 2G and 3G networks by 2033 and transition to modern interoperable networks, such as 4G or 5G[8]. However, a key requirement in deploying new or replacement sites is to delivery of voice service. In fact, voice calling is regulatory requirement from Ofcom. To provide this service UK

[6] https://www.nytimes.com/2020/02/27/technology/fcc-location-data.html
[7] https://www.thebureauinvestigates.com/stories/2020-12-16/spy-companies-using-channel-islands-to-track-phones-around-the-world
[8] https://www.gov.uk/government/news/a-joint-statement-on-the-sunsetting-of-2g-and-3g-networks-and-public-ambition-for-open-ran-rollout-as-part-of-the-telecoms-supply-chain-diversificatio

operators could deliver IMS IP-based calling, but importantly this would require accompanying handset support, for every handset on their network, including in-bound roaming users. This represents a significant challenge considering the numerous models and variations of handsets which would all have to be supported.

Therefore, an operator wishing to deploy modern 4G/5G Open RAN equipment would have to additionally provide 2G/3G legacy network coverage. This challenge was identified specifically by the Diversification Taskforce in their early 2021 report to DCMS, highlighting the challenge for a new-entrant vendor (even one at-scale, like Samsung, as detailed in their evidence to parliamentary committee[9]) which lacked a 2G or 3G RAN offering[10]. European mobile operators have even switched RAN equipment provider based on their (in)-ability to provide legacy equipment[11]. As such, there has been a rush to acquire 2G/3G capabilities in the Open RAN industry, which underlines the ongoing importance of CSFB service provision.

The challenge behind delivering an additional 2G/3G network is the security model is quite different from that in a 4G or 5G network. A large range of legacy protocols and components would require connectivity, and therefore, security protections. These components include the 2G Base Station Controller, the SGSN (Serving Gateway Support Node) and the GGSN (GSM Gateway Support Node). There would also be a need to support connectivity for service delivery, such as SMS, as well as connectivity to a legacy HLR and MSC for voice functionality. This would add significant extra cost and introduce a considerable level of complexity in the architecture since, as the name suggests, these legacy protocols were originally designed for operation in a circuit-switched, rather than packet-switched, network. This will increase the cost of routers and switches required to support carrying legacy protocols as well.

However, as 2G and 3G networks, still carry a significant quantity of voice calls in the UK. This is in-part due to rural areas often not having 4G/5G coverage. Even in urban areas, handset compatibility of IMS (required to carry out calling over the 4G or 5G network) can be limited, and lost in the event of switching network, due to firmware-level incompatibilities.

Despite this, it is important to note that legacy carrier-provided voice and SMS services are increasingly becoming irrelevant for users, as they are largely using over-the-top applications, such as WhatsApp, iMessage/Facetime, Facebook Messenger and similar, rather than their carrier-provided native voice and messaging functionality. This means that there will be commercial challenges in making a viable business proposition that can deliver CSFB. It is a regulatory requirement (at present) for the devices to support it, yet there is already a timeline in place for its removal, and those implementing networks are unlikely to be able to make a commercial case for developing new functionality that already has an end-of-life date in the UK market.

[9] https://committees.parliament.uk/writtenevidence/5863/pdf/

[10] https://www.gov.uk/government/publications/telecoms-diversification-taskforce-findings-and-report/telecoms-diversification-taskforce-findings-and-report

[11] https://www.mobileeurope.co.uk/press-wire/16209-telecom-italia-says-2g-skills-are-minimum-requirement-now-for-o-ran-suppliers

Separation of Operators

In traditional network architectures, the security-critical functions of the mobile network sit almost entirely within the core network[12]. The inter-connections between mobile networks are carried out in the core, in the inter-network signalling components. These are high-risk attack surfaces between mobile networks, since they present points where one compromised network can interact with another network and permit an attacker who has successfully compromised aspects of one network to attempt to enter into another network.

This inter-network signalling enables things like roaming and calling users on different mobile networks, and therefore are also exposed to international operators.

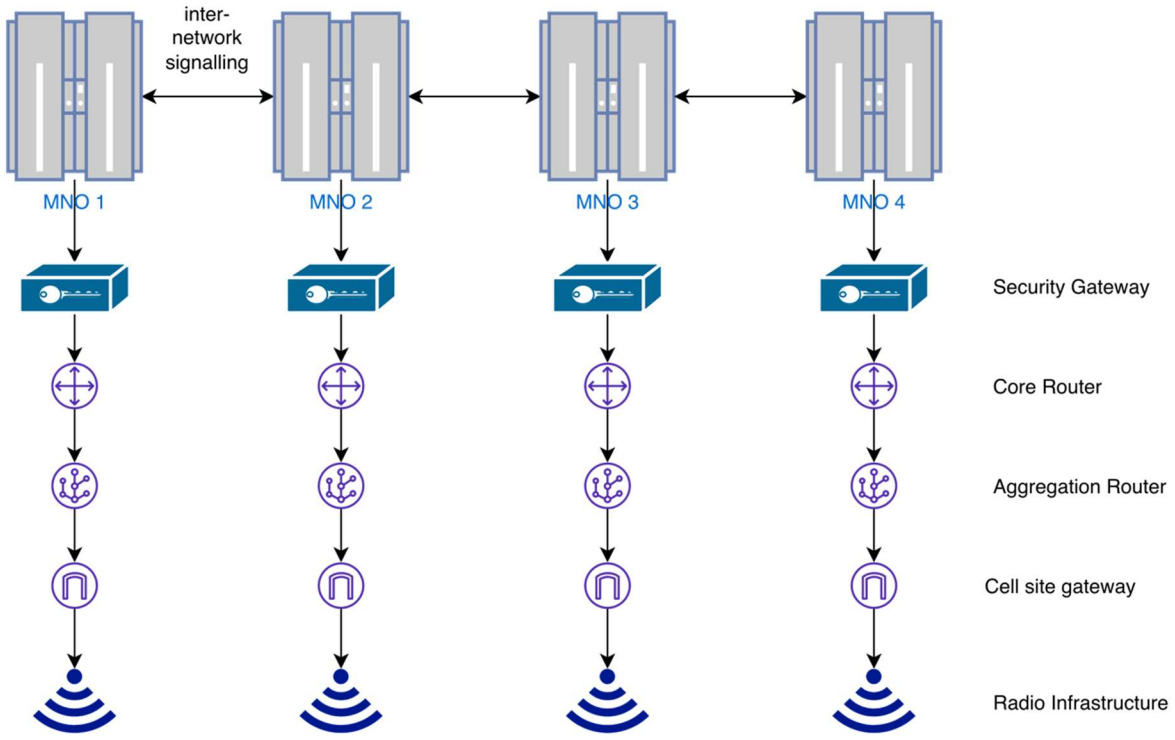

Figure 1 - Illustration of traditional radio network infrastructure and interconnection between operator networks

Traditionally, mobile network operators each run their own RAN, Core and infrastructure. Components downstream from the security gateway can be aggregated and combined across operators, to realise reduced capital and operating costs within the RAN. However, from a security perspective this also introduces an inherent level of inter-connectivity between mobile networks, at a new level of connectivity (i.e. downstream in the RAN), in addition to the traditional inter-network signalling connectivity that is present in the core network. To manage the security implications of this, there is likely to be a technical requirement to introduce greater network visibility closer to the edge of the network.

12 https://www.ncsc.gov.uk/files/Summary%20of%20the%20NCSCs%20security%20analysis%20for%20the%20UK%20telecoms%20sector.pdf

Where networks are inter-connected towards the edge, the costs of monitoring and securing these inter-connections will be inherently greater. This is due to the greater number of such components deployed in the network, and therefore a greater number of cross-network exposed interfaces. Each of these interfaces will require firewall and monitoring infrastructure, as well as skilled human resourcing to monitor them for incidents, potential attacks, or signs of compromise. It is also likely that both operators would want to deploy their own infrastructure in a "back-to-back" configuration, such that both operators are able to control their own monitoring and firewall infrastructure.

Emerging Threats for Mobile Operators

Advancements in architecture and design in 5G, has meant that mobile operators are able to provider lower latency, higher resilience, and increased vendor diversification. However, to achieve these benefits, operators will have to adapt to new technologies, and their associated security risks.

Virtualisation, Containerisation & Sharing of IT Platforms

Virtualisation allows commodity server hardware to run a number of functions inside isolated virtual machines sitting atop a hypervisor. The hypervisor allocates resources to virtual machines and controls their access to the physical hardware of the system. This model can have manageability and security advantages, compared with running all software on a single server, as all the different 'machines' running on the same hardware are isolated from one another. Virtual machines can also be imaged, backed up, and restored more easily than physical computers. With appropriate hypervisors and licences, they can even be transferred seamlessly between hypervisor hosts, allowing for maintenance, upgrades, replacements, or handling equipment failures.

Recently, containerisation has emerged as an alternative sharing platform, with lower processing overheads than virtual machines, as there is no need for a full operating system to be hosted for each container. This means that a container can be more lightweight, and be setup more quickly in response to an increase in demand.

The benefits of virtualisation and containerisation have been exploited extensively by tech companies, such as Google or Microsoft, as this standardised approach abstracts away hardware-specific properties of the underlying host system. As long as the underlying hardware is capable of running the requisite workloads, the virtualisation platform provides a predictable and tested environment for network functions to run and enables rapid system scalability to cope with dynamic demand.

This has meant that the latest generation of software-based 5G mobile cores have been developed with this scalability and virtualisation in mind. Accordingly, traditional web hosting companies have partnered with 5G core providers to enable rapid deployment of 5G networks. Metaswitch has partnered with Microsoft Azure[13], Athonet has demonstrated its core being run on Amazon AWS[14] and Ericsson has partnered with Google Cloud Platform[15]. Each of these have developed tightly integrated solutions where a consumer is available to quickly spin-up a 5G core instance and connect their RAN.

Virtualisation enables more efficient use of resources, but the security of virtualised hosts is different to physical hosts. Firstly, containers isolation is not as robust as virtual machines

[13] https://www.metaswitch.com/blog/metaswitch-fusion-core-now-integrated-with-microsoft-azure-private-multi-access-edge-compute
[14] https://enterpriseiotinsights.com/20200910/channels/news/athonet-releases-open-5g-core-as-a-service-on-aws-to-spur-private-5g-market
[15] https://www.ericsson.com/en/press-releases/2021/6/google-cloud-and-ericsson-partner-to-deliver-5g-and-edge-cloud-solutions-for-telecommunications-companies-and-enterprises

isolation, since containers do not ship with their own kernel. This means they operate in a shared environment alongside other containers, making "container escapes" possible[16]. Palo Alto Networks reported that in recent months they've discovered three kernel vulnerabilities that allowed for container escape, without additional hardening measures such as Seccomp and AppArmor or SELinux[17].

Virtual Machines are more isolated due to the dedicated kernel, however, almost all systems present a highly privileged management plane interface, for administration and control of the hypervisor. Access to this hypervisor may enable access to multiple different virtual applications running on the same hardware, potentially providing the attacker significantly more access than if a single physical host was compromised.

To isolate the hypervisor's management plane, it should be physically separated (i.e. using separate equipment) from untrusted interfaces, such as those which link to provide connectively to consumers. System-level integrity checks, such as UEFI secure boot, should be used to detect and prevent a compromised hypervisor from being booted. Virtual machines themselves should be integrity checked, using root of trust. Hardware-backed cryptographic keys stores; should be made available directly to the virtual machine through SR-IOV or VT-d. A hypervisor's management plane should be protected to the security level of the highest security function which may be run on it.

Virtual machines should be monitored and, where possible, run from stateless (read-only) snapshots which are unable to boot if tampered with. By doing this, it will be significantly more difficult to compromise a virtual machine and gain code execution which could ultimately result in the ability to compromise the hypervisor or wider network. Platform security must also be considered; to ensure that system firmware updates are shipped and installed on devices, since vulnerabilities in the platform firmware may affect the software running on the system.

Standard IT security mitigations should be followed at every level of the process: Hypervisor, Virtual Machines and Software Application. Hypervisors should be promptly patched, as a hypervisor is a highly privileged process; and often has a complex attack surface. A recent VMWare vulnerability was expected to be exploited within minutes of being disclosed[18], which shows the necessity to patch very rapidly. Moreover, the breadth of different types of security vulnerabilities which can be found in this kind of software[19] shows the challenge in managing these systems.

To provide defence in depth, access control mechanisms should be employed on both the hypervisor and virtual machines themselves, and, where possible, passwords should not be used for authentication. Instead, PKI-based authentication should be carried out by smartcard or other security token. Where backup, `break-glass access' passwords must be used, these should be unique to each device; and sufficiently complex to prevent brute force guessing

[16] https://unit42.paloaltonetworks.com/cve-2022-0492-cgroups/
[17] https://www.paloaltonetworks.com/blog/prisma-cloud/linux-kernel-vulnerabilities/
[18] https://blog.malwarebytes.com/exploits-and-vulnerabilities/2021/09/patch-vcenter-server-right-now-vmware-expects-cve-2021-22005-exploitation-within-minutes-of-disclosure/
[19] https://www.vmware.com/security/advisories/VMSA-2021-0020.html

attacks from succeeding. Additionally, these should be single use passwords and changed after use, to prevent from being stored and re-used.

Similarly, standard software protections should be employed for virtualised components, for example, stack protection. Suppliers should additionally be required to build-in security measures, such as signing at compile time to enable validation. It is also important for software functions to not overlook "historical" security issues on IT platforms, which have generally been ignored in a desktop computing context, but which became more relevant in a telecoms network environment where more skilled adversaries will go to greater effort to compromise a network. Attacks such as LD_PRELOAD dynamic linking overrides[20] should be mitigated, and the full execution environment (including system and application libraries) should be part of integrity measurement of the system in use. Where a system changes from the expected state, it should be removed from production use, isolated, and made available for investigation, to look for any indicators of compromise or other forensically useful artefacts.

Software Orchestration

As network functions transition to being purely software implemented, rather than as physical hardware appliances running software, the network operator must also take into account the provision of a secure and responsive platform to assist running all the various software components. This model of operations is increasingly common in the enterprise IT and cloud computing space, especially with the rise of Platform-as-a-Service and Function-as-a-Service. In these models, application functionality runs independently of operating system and platform. Examples of this in the IT sector include Heroku, AWS Lambda, etc.

In such an environment, there is an important distinction from traditional telecoms equipment, since there is an inherent decoupling of the platform and software execution environment from the software's logic itself. While this can help with software supply chain security, by allowing for independent validation and updating of the platform or network function software itself, it introduces a new layer of complexity – each time dependencies are uncoupled, there is a requirement for integration and validation testing to take place, to ensure that an updated platform or application continues to function with the rest of the stack.

In the context of a mobile network, to avoid downtime during an upgrade, multiple instances of each component need to operate, creating redundancy and the ability for connections to "fail over" to other instances of a component. This means that multiple equivalently configured instances of each network function are required, and a managed approach must be taken to scheduling the downtime, upgrades and restarts of these, lest an upgrade be carried out on all nodes at once, causing an outage.

In order to carry out orchestration effectively, it is paired with automation, which allows for scheduling and management of containers or other workloads. This requires that the orchestration software has privileged access, usually across a cluster of nodes, in order to be able to effectively manage workloads in a decentralised way, without a single point of failure.

[20] https://cloudsofhoney.net/2021/06/24/mitre-attck-t1574-006-dynamic-linker-hijacking-with-ld_preload/

This introduces potential attack vectors, as it makes the orchestration platform software highly privileged and a valuable target.

In many containerisation environments, there is a shared control plane, which is used by privileged users to communicate in API requests, as well as other nodes in the cluster to communicate their status and report back on having carried out tasks they were asked to do. This introduces a potential high-risk point of failure in the "controller" nodes which act as orchestrators and expose a shared API to both management plane traffic, as well as virtual network functions themselves. Approaches to isolation and segregation of this kind of communication should be considered, to prevent a compromised node from sending management requests to the cluster controller.

In addition, there are also resilience implications of such an architecture – as orchestration gains adoption, it is often linked with an increase in the design complexity of the overall solution, since automation helps to make it more feasible and practical to run a more complex overall solution. For example, rather than using an internal data store, it is highly likely that a clustered, orchestrated solution will connect to a separate distributed data store, run across nodes (such as redis or ceph). This makes the continued running of the network function eventually dependent on the orchestration and automation platform, since it ceases to be viable to manually manage and schedule these complex dependencies. In addition, the orchestration platforms themselves are also increasingly complex software, requiring regular updates, in order to remain up to date with the rest of the cluster, and ensure they apply relevant security patches.

There is precedent for this becoming a real issue, Monzo, the UK start-up bank, reported an incident[21] where they encountered an issue upgrading their orchestration stack itself (Kubernetes, often abbreviated to K8s). Due to a complex inter-dependency between different components that were updated at different times, there was an incompatibility in an upgrade, which could not be rolled back, which led to a service outage. Due to the complexity of these systems, Monzo had to resolve multiple chained incompatibilities, arising from complex interdependencies and version incompatibilities, which meant that different components could neither work together, nor easily be downgraded. Resulting in loss of service for Monzo consumers.

In this case, Monzo was able to push out an update to one of the components to an even newer version that they had been testing, which resolved the incompatibility. Had this not been possible, it is likely that a more prolonged outage would have been experienced, requiring rebuilds of the automation and orchestration infrastructure on an older, previously tested version, as well as re-provisioning and recovery of data held in any storage services which were running on that platform.

As orchestration becomes prominent in the telecoms industry, suitable experience and ability to control these complex systems, will become a requirement for mobile operators. This is also outlined in the upcoming telecoms security regulations, where an organisation must maintain the capability to manage their own systems without being entirely reliant on outsourced

[21] https://community.monzo.com/t/anatomy-of-a-production-kubernetes-outage-presentation/37331

providers, and must have a plan in place to "operate the[ir] network without reliance on persons, equipment or stored data located outside the UK"[22].

Software Supply Chain

The GSMA recognises the on-going security issue concerning the supply chain of the software used within mobile networks[23]. Current UK market consolidation has meant that three out of the four UK MNOs have elected to deploy an Ericsson core, while one elected to use a Nokia core. This standardisation within the core architecture represents both a benefit as the commonality between code bases reduces the amount of code to be secured, and a detriment because if a major vulnerability were discovered it could quickly be inflicted on a large proportion of the market.

Modern software is constructed from various dependencies and libraries, and 80-90% of all software is derived from Open Source components[24]. However, operators may not even be aware of the software that is being run on their systems, due to the complex nature of dependencies. For example, LOG4j is a java-based logging library, which is used throughout numerous industries and software products. On 9th of December 2021, a `zero-day' Remote Code Execution vulnerability was discovered in this logging library. A `zero-day' vulnerability means that the developers of the software had no prior knowledge of the potential exploit, and do not have an update to correct it. Remote Code Execution refers to the ability for an attacker to run arbitrary code on the victim's system, without the victim's knowledge or permission. This may allow them to extract passwords, data, or change the functionality of the system.

At least 750 companies, including VMware, Microsoft, Apple iCloud, Samsung Knox and IBM are known to have been affected by this vulnerability[25], with an estimated half of all corporate networks targeted, seeing over 100 attempts to exploit every minute[26]. This vulnerability was particularly challenging, as system administrators may not even have been aware of their potential exposure to risk. This is due to downstream software utilising this vulnerable software but not disclosing their dependencies.

This uncertainty as to what software is included within an overall software package can be resolved through a Software-Bill-of-Materials. This name is derived from when a hardware product is constructed, and a bill-of-materials detailing all the component parts is created. This would detail every component and version number which goes into constructing a particular software product, which in turn would provide system administrators the knowledge around what products they need to patch.

---

[22] https://www.gov.uk/government/consultations/proposal-for-new-telecoms-security-regulations-and-code-of-practice/telecoms-security-proposal-for-new-regulations-and-code-of-practice

[23]https://www.gsma.com/futurenetworks/resources/open-networking-the-security-of-open-source-software-deployment/

[24] https://www.linuxfoundation.org/press-release/spdx-becomes-internationally-recognized-standard-for-software-bill-of-materials

[25] https://gist.github.com/SwitHak/b66db3a06c2955a9cb71a8718970c592

[26] https://www.zdnet.com/article/log4j-flaw-nearly-half-of-corporate-networks-have-been-targeted-by-attackers-trying-to-use-this-vulnerability/

Another consideration for telecoms operators are the misaligned incentives between them and their potential downstream software vendors. Namely, the main priority of many open-source software developers is functionality[27], rather than security, and developers may not be security minded or may not have designed the software with the same threat model in mind.

Commercial open-source software, is similar to traditional open source software, and is often produced by the same development community. Where it differs is there is a commercial agreement is available for support, expert advice, and dedicated patches/features for the buyers' specific use case. The code itself commonly remains open source and freely available, as the agreement is based around a service rather than a product. The most well-known implementation of this model is perhaps RedHat Enterprise Linux; although it uses entirely open-source code, businesses pay for continued support and updates to maintain service to their customers. Commercial open-source helps to address some of the funding gaps experienced by open-source projects. However, this still creates an external dependency and trust on ongoing support and maintenance from an organisation which may not continue as long as the mobile operator requires.

Proprietary software packages which contain open-source code, should use automated scanning tools within continuous integration and continuous delivery (CI/CD) development environments. These tools are able to identify obsolete and vulnerable products, alerting the developers to the need to update or replace out of date components within the stack. Such tools should consider the operating system platform, system libraries, container base (if applicable), and application code (including dynamically and statically linked external code), when doing so.

The Huawei Cyber Security Evaluation Centre (HCSEC), a joint initiative between the UK Government and Huawei to evaluate the security of Huawei products for use within the UK, discovered that an older version LTE eNodeB product contained "70 full copies of 4 different OpenSSL versions, ranging from 0.9.8 to 1.0.2k (including one from a vendor SDK) with partial copies of 14 versions, ranging from 0.9.7d to 1.0.2k, those partial copies numbering 304. Fragments of 10 versions, ranging from 0.9.6 to 1.0.2k [...]"[28]. This significant diversity spread throughout the code base would be unmaintainable from a developer perspective. The report went on to detail "There were also a large number of files, again spread across the codebase, that had started life in the OpenSSL library and had been modified by Huawei". OpenSSL is a software library used for the exchange of cryptographic keys that forms the basis of a secure communication channel.

OpenSSL versions 1.0.1 through 1.0.1f, in particular, experienced a critical security flaw in 2014 known as Heartbleed, where anyone on the internet was able to retrieve private memory of the application running vulnerable versions of the OpenSSL. Publicly available code to exploit this vulnerability has been available since 2014, but the splintered nature of the Huawei code made

[27] https://www.linuxfoundation.org/wp-content/uploads/2020FOSSContributorSurveyReport_121020.pdf

[28] https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/790270/HCSEC_OversightBoardReport-2019.pdf

it challenging to patch all the various versions and therefore vulnerabilities. In 2019, this is understood to be resolved through gradual consolidation of versions.

As part of Cisco's response to the upcoming NCSC Telecom Security Requirements[29], they have detailed the requirement to employ a single, well maintained, in-house version of critical software, such as OpenSSL. This benefits the security and on-going maintenance of all products relying on this OpenSSL library, as only a single copy needs to be maintained.

On-going Support

Existing telecoms operators often outsource the management and maintenance of their equipment to a managed service provider (MSP). This would commonly be the equipment manufacturer. This does provide a number of advantages for operators. For example, updates and testing can all be standardized across different markets reducing the effort undertaken and the associated costs. The downsides, are that it creates an external dependency on the on the management company.

The Competition and Markets Authority[30], raised this concern with the Home Office dual dependency on Motorola Solutions Inc (Motorola). Motorola are involved in both the existing Airwave network used by Police and Ambulance for critical communications within the UK, and the upcoming LTE-based replacement Emergency Services Network (ESN). In the long term, ESN is expected to save the Home Office £200 million a year. But extending Airwave's lifespan costs ~£500 million annually. The initial switch off date was due to be 2019, but this has repeatedly been pushed back until the ESN system is fully operational with current estimates being Q4 2026[31]. The increasing maintenance costs result from simple supply and demand, as a technology becomes older, and falls out of fashion, the ability pool of available engineers to maintain it becomes diminished. Therefore, the costs to acquire this skill-set increases.

The dependency on an MSP, can also present a security risk, as detailed in the widespread Kaseya ransomware attack[32]. DCMS have additionally called for views on the security of managed service providers[33]. Finally, as operators deploy multiple different vendors' kit, it may be unclear which vendor is responsible for the interoperability between different pieces of equipment.

RAN Diversification

Advances in Open RAN increase interoperability of different RAN components, which in turn reduces the barrier for new entrants in the market. Rather than having to construct an entire eco-system to compete with established players, new entrants can focus on one particular area,

[29] https://trustportal.cisco.com/c/r/ctp/trust-portal.html?search_keyword=security%20declaration#/pdfViewer/c%2Fdam%2Fr%2Fctp%2Fdocs%2Fcorporate-policy%2Fcisco-security-declaration.pdf
[30] https://www.gov.uk/government/news/cma-opens-investigation-into-motorola-s-airwave-network
[31] https://www.ukauthority.com/articles/airwave-network-deal-extended-until-end-of-2026/
[32] https://www.globalsign.com/en/blog/kaseya-attack-2021-are-ransomware-attacks-inevitable
[33] https://www.gov.uk/government/publications/call-for-views-on-supply-chain-cyber-security/call-for-views-on-cyber-security-in-supply-chains-and-managed-service-providers

increasing the potential pool of vendors and removing the recently emerged duopoly within the UK market.

However, from a security perspective, there are additional considerations which should be made when reducing the barrier to entry. Firstly, this may also result in a race for the bottom with vendors concentrating less on security at the expense of reduced prices and faster time to market. Secondly, the disaggregation (splitting) of components within an Open RAN environment also increases the number of interfaces and therefore the number of potential attack vectors. In addition, the introduction of more vendors means a greater number of third-party and open-source software libraries which require risk-management and patching. There will also be increased complexity in building appropriately isolated management plane interfaces to keep different vendors' equipment isolated, and a heightened need for awareness of potential container and hypervisor escape scenarios, where multiple Open RAN functions are run on the same physical server node.

Open RAN features five additional interfaces:

- A1 - Connection between Service Management and Orchestration and the RAN Intelligent Controller
- E2 – Connection between RAN Intelligent Controller and CU/DU
- F1 – Connection between CU and DU
- Open FrontHaul - Connection between DU and Radio Unit
- O2 – Connection between Orchestrator and Cloud Platform

These introduce increased connectivity between RAN functions from potentially different vendors, which increases the risk of security concerns arising from a heightened number of potentially privileged interfaces exposed to an increased number of virtual or physical network functions, from an increased number of vendors.

Telecoms operators have historically been used to peering only with trusted partners, as opposed to working with a diverse range of suppliers and operators. The GSMA and 3GPP have jointly attempted to increase the number of suppliers while retaining a focus on security through their Network Equipment Security Assurance Scheme (NESAS)[34], which aims to provide a centralized validation and security audit of vendors to simply the purchasing process for operators. NESAS defines a set of security requirements and provides a consistent framework for the assessment of pre-defined test cases. However, it should be noted that currently only limited subset of vendors had applied and undertaken these tests, and the draft Telecommunications Security Code of Practice stipulates the full findings would have to be evidenced as the current evaluation criteria is focused on brevity and ease of reading rather than detailed technical descriptions,

Another issue that MNOs are likely to face is integrating between and with existing equipment. To deliver reliable performance in a modern wireless network, there is a range of performance optimisations which should be carried out cross-radio. The X2 and/or Xn interfaces are used in 4G and 5G, respectively, to facilitate communications between radios. For example, a radio vendor can offer improved performance at the cell edge (and thus reduce inefficient utilisation

[34] https://www.gsma.com/security/network-equipment-security-assurance-scheme/

of radio resource blocks) by carrying out coordination with the relevant neighbouring cell, to ensure that the resource blocks allocated to a user near the cell edge are left free in the adjacent cell, reducing the likelihood of interference, and thus improving performance. This requires responsive, near real-time inter-radio traffic over the X2/Xn interface. These protocols are not interoperable beyond the basic functionality included in the standards, which does not offer these kinds of performance features[35].

Traditionally, before the rise of Open RAN and a push for interoperability of interfaces used on RAN equipment, network operators have deployed their networks using mono-vendor "RAN zones". In each zone, radios from a single vendor are used. This creates an economy of scale for the vendor in question. Therefore, to gain the full benefits of an increased vendor pool and interoperability, inter vendor X2/Xn performance additional work will have to be undertaken to integrate both systems. Smaller RAN manufacturers, such as Mavenir, have attempted to undertake this work but have been meet with prohibitively expensive licensing costs from Tier 1 vendors to improve this interoperability[36].

Localised Cell Sites

As the frequency of mid and high band 5G services is greater than previous generation services, the cell size will decrease, requiring a larger number of cells to cover the same area. This will also be coupled with the need for specific high-capacity areas, such as density populated areas. The Small Cell Forum (SCF) has explored this consideration extensively and summarised that as the number of remote elements will increase, the current physical protection of assets will no longer be viable. Instead cryptographically secure, zero-trust architecture between network components should be implemented.

This approach is in line with Ofcom's proposals for mmWave-based densification of mobile networks in urban environments, rather than increasing availability of spectrum to meet increasing user demand[37].

In a rural environment, where equipment is further from police or private security responses, there will be an increased need for platform assurance and security. Where a NH operator were to seek to install a UPF or other equipment at the network edge, this would also be a concern from a security perspective. The threat of physical compromise of a site is one of the key driving concerns behind the MNOs' approach to security of assets and network equipment, and is what gives rise to the Joint Operators Technical Specifications (JOTS) architecture for neutral host solutions.

[35] https://www.gov.uk/government/publications/5g-supply-chain-diversification-strategy/annex-5g-networks-overview--2

[36] https://www.lightreading.com/open-ran/ericsson-nokia-accused-of-charging-exorbitant-fees-to-open-networks/d/d-id/775017

[37] https://www.ofcom.org.uk/__data/assets/pdf_file/0017/232082/mobile-spectrum-demand-discussion-paper.pdf

Operators' Perspectives on Security of Neutral Host Architectures

After establishing likely future security issues for MNOs as a whole, we will now examine the existing mobile operator's security architectures for enabling Neutral Hosting, and importantly how these may align with the wider security objectives of the industry. The Joint Operators Technical Specifications (JOTS)[38], published by Mobile UK, presents the UK operators' aligned and coordinated technical specifications for the deployment of shared wireless solutions. As stated in the introduction to JOTS, "The aim of the JOTS forum is to specify the performance, coverage and reliability of wireless systems that are shared by mobile operators."

JOTS originated as a technical forum in 1999, creating a joint technical requirement document for third-party indoor GSM coverage solutions, and has evolved since then to keep up the pace with subsequent generations of mobile networks. JOTS was originally defined for indoor Distributed Antenna System (DAS)-based solutions.

Our earlier reports[39] in this series compared and explored the differences between DAS and active neutral hosting. As a summary, a DAS system is generally based around a small number of operator base stations, with the RF outputs from these base stations routed throughout a large building. A DAS system may have multiple antennas throughout a building, all served from the one base station (and thus sharing the combined capacity of the operator's spectrum across the whole building). More advanced DAS systems may have multiple operator base stations, to give increased capacity by serving different areas of a building from different operator base stations.

An extension to the JOTS specification was developed between 2018 and 2020, to create a series of 5 annexes to the JOTS specification, for Neutral Host In-Building (NHIB) scenarios. It is important to note that the JOTS NHIB work is clearly predicated on indoor usage (as the name suggests), and language used in the documents makes clear that the authors envisage it being used for "venues". This means that NHIB in itself is not necessarily a suitable model for the deployment of a rural neutral hosted network. Nonetheless, the JOTS approach is highly relevant, as it presents a security architecture which has been agreed by mobile operators, and which can be used to integrate with their core networks.

An outdoor rural NH network would have to integrate with existing RAN equipment to support handovers and other functional and performance requirements, but otherwise could be adapted to adhere to the JOTS architecture and security properties.

The subsequent sections present a critical evaluation of the JOTS specification annexes, and potential implications for rural neutral host operators. Where there are areas of potential improvement, such as in opportunities for enhanced security, these are also highlighted. Note that we do not conduct a detailed analysis of the DAS-based JOTS specification, since it is inherently driven by the design principle of an operator's base station being co-located on the DAS site, and under the control of the operator. In evaluating the security requirements for an outdoor rural NH operator, the annexes for NHIB are the relevant specifications.

[38] https://www.mobileuk.org/jots

[39] https://5gruraldorset.org/app/uploads/2021/12/5G_Rural_Dorset_WP6_Task5_Business_Study.pdf

Introductory Terminology

In order to keep the use of terminology consistent with JOTS, readers should note that JOTS NHIB proposes three separate operational domains – Operator Domain, Neutral Host Domain, and Retailer Domain.

The Operator Domain is run by a mobile network operator (MNO) and implements a Tier-2 security gateway or another IPSec termination device, with Public Key Infrastructure (PKI) to authenticate the Tier-1b security gateway connection, and appropriate connectivity to the mobile core network.

Neutral Host Domain refers to a secure data-centre location where "Tier-1f" security gateway components (for termination of IPSec tunnels from neutral hosted cell sites) are located, alongside connectivity from the Neutral Host Domain to the Operator Domain, to connect from the "Tier-1b" Security Gateway in the Neutral Host domain to the Tier-2 Security Gateway in the Operator network. The Neutral Host Domain also needs to implement base station aggregation, and RBAC-controlled management for multiple tenants (i.e. Operators), as well as the provision for in or out-of-band management connectivity to each operator. Neutral Hosts further need to implement specific technical PKI and follow security governance measures, such as ISO27001, and previously CAS-T, which is in the process of being updated to the forthcoming Telecoms Security Requirements, under the Telecoms (Security) Act 2021.

Retailer Domain refers to a commercial entity, whose business is built around the provision of in-building coverage solutions to venues. It is not the venue in question. Retailers are responsible in the JOTS model for the procurement, provision, configuration, security and ongoing maintenance of site equipment and infrastructure, as well as management of the relationship with the venue or end customer.

The Neutral Host and Retailer could both be the same commercial entity in this model, but the domains have inherently different physical security requirements.

JOTS Architecture Requirements

The first JOTS annexe outlines the functional and security architecture for NHIB installations. Paragraph 10 indicates that dot1q[40] VLAN encapsulation is required to be used, so that 802.1p[41] MAC-layer packet priority marking can be implemented, to prioritise NH traffic on a shared LAN or other infrastructure. This is relevant to rural NH environments, especially where a local ISP or other alternative backhaul infrastructure provider may be used, as it would introduce technical expectations on the implementation of backhaul. It is also permitted for a L2 VLAN from the Neutral Host Domain to be extended to the Retailer domain. While either an L2 or L3 network can be deployed at the Retailer domain, paragraph 11 indicates the need for an appropriate IP addressing scheme to be used. It is possible for NAT to be used in this process, which may assist smaller Retailer operators in managing their requirements for publicly routable IP addressing. The LAN addressing can be re-used in a NAT'ed environment, or

[40] https://www.ieee802.org/1/pages/802.1Q-2014.html
[41] https://docs.microsoft.com/en-us/previous-versions/windows/desktop/qos/802-1p

dedicated private addressing can be used in a Metro Ethernet (i.e. VLAN'd) environment to reach the Neutral Host Domain. The latter may be an attractive option where a rural internet provider is acting as both Retailer and Neutral Host, since they already operate and manage the intermediate network.

Paragraph 14 indicates that DNS resolution must be available to the small cell service, either provided locally, through the Neutral Host Domain, or through public DNS with appropriate security measures in place. This presents an opportunity for a NH provider to add security measures. For example, DNS logging and active monitoring, in addition to filtering, could be used to provide a level of alerting for basic attacks or anomalous behaviours; if there is an attempt made by a piece of software in the base station infrastructure to reach an unexpected hostname, this may provide useful intelligence about an upstream supply-chain compromise of a relevant component. The high-profile Solarwinds supply chain compromise had a number of unique DNS lookups[42], which would be detected in a suitably monitored and filtered environment, providing early indication of attempts being made to connect to new and unexpected hostnames.

Paragraph 16 sets out an outline security architecture for how a Base Transceiver Station (BTS) should be configured to communicate with small cells and the aggregation node (Controller). The controller is expected to be discovered via a DHCP option 43 request or similar, and IP address, subnet, and gateway via regular DHCP request. As an alternative option, tunnel endpoint addresses could be determined via FQDN resolution. This presents an opportunity for additional security measures, by restricting the available FQDNs, and in hardening and minimising the complexity of DHCP implementations on the network, to validate and respond to only well-understood and expected messages, in order to mitigate advanced attacks that may attempt to gain code execution through a potential future DHCP exploit. Privilege separation of DHCP client services on devices would be another worthwhile implementation-level security feature, to avoid DHCP clients running with elevated privileges.

Paragraph 16 goes on to also outline how small cell nodes will initiate a connection to their Controller or Tier-1f SecGW, depending on the aggregation architecture deployed. This will be an IPSec tunnel, with IKE-SA handshaking and PKI-based certificate exchange.

Paragraph 16 further discusses that Operation and Maintenance (OAM) connectivity should be achieved using the same tunnel-inner address as used in S1-AP or S1-U connectivity, but that other models can be considered where a separate tunnel is used for OAM connectivity. This presents a significant opportunity for improvement of the JOTS design – robust isolation of the management plane and control plane would be beneficial, by isolating privileged management functions from regular control plane traffic, which may give access to platform functionality through protocols like SSH, especially in modern, interoperable software-based or virtualised RAN infrastructure. By isolating management traffic from user and control plane traffic, this prevents malformed packets from being treated incorrectly as management traffic, and exposing users to management plane functionality.

[42] https://truefort.com/wp-content/uploads/2021/01/TrueFort-SolarWinds-IoC.pdf

If the same network interface is desired to be used, contrary to the best-practice isolated management plane approach proposed above, management plane protocols should be implemented through a VPN or similar secure, tunnelled protocol, rather than exposing plain OAM traffic over the interface. Otherwise, there is a risk that misconfigured equipment could allow users to target traffic towards the OAM interfaces with their user plane traffic.

Paragraph 16 requires per-operator management interface segregation where different endpoints are deployed for different MNOs. This is an important security consideration when implementing a NH architecture and proper isolation architectures should be considered, beyond merely the management plane endpoint interface.

Paragraph 18 illustrates an internet-based NH **retail** connectivity concept, through which a BTS is connected through a local network, 1:1 NAT to permit pass-through on a regular CPE, then is transmitted across the internet to the BTS controller in the **Neutral Host domain**.

Figure 2 - NH retail architecture with controller in the NH domain[43]

In this scenario, as the controller is located in the **Neutral Host domain**, there is a requirement for good backhaul performance over the internet, specifically with low latency and jitter. This would be a suitable model for deployment across a fibre altnet provider-run network, for example. One advantage in this model is that by locating the controller in the **Neutral Host domain**, this reduces the costs of equipment in the **retail domain** (marked venue in the above figure), since in essence, only a BTS, local network, and CPE are required.

Where suitably low latency and jitter cannot be attained through the connectivity available, the below architecture is proposed, with the controller located on the local network at the BTS site. This will increase the per-site cost.

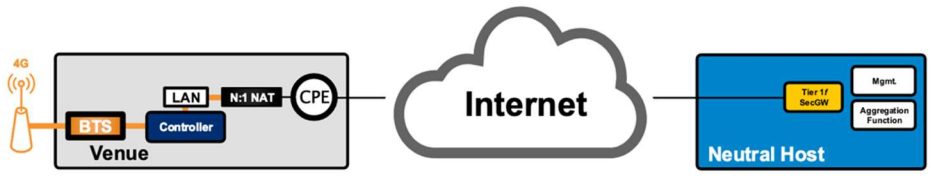

Figure 3 - NH retail architecture with controller in retailer domain [44]

[43] https://uploads-ssl.webflow.com/5b7ab54b285dec5c113ee24d/5fabc97b674e5182cf149531_JOTS-NHIB-Specification-Annex-1-Architecture.pdf
[44] https://uploads-ssl.webflow.com/5b7ab54b285dec5c113ee24d/5fabc97b674e5182cf149531_JOTS-NHIB-Specification-Annex-1-Architecture.pdf

More generally, it is worth noting that the JOTS specifications have been written from the perspective of traditional RAN equipment. In the 5G RuralDorset project, the feasibility of deploying coverage using tier-2 vendor (i.e. non-traditional) RAN equipment has been trialled and validated. It is likely that new JOTS-like architectures for Open RAN deployments would need to emerge, were O-RAN standards to gain traction in the UK market. In the meantime, it is likely that there may be an intermediate step, where solutions not fully compatible with the full O-RAN standard-set could be used in a NH network which adheres to the JOTS specifications, while not necessarily implementing some of the RAN-specific aspects as envisaged under JOTS.

For example, the concept of a traditional BTS-and-controller architecture may be able to be made more flexible with software defined radio base-stations, where a Radio Unit (RU) implements the 5G PHY in an FPGA at the top of a mast, and a commodity off-the-shelf (COTS) server is situated near the base of the mast, or at a nearby aggregation site within a ~20km fibre run. This was demonstrated within the 5G RuralDorset program where 5G Baseband Units (BBUs) where co-located in a base station hotel arrangement and dedicated CPRI fibre was run to the RU on the mast.

Technically, this could be expanded to run over existing public fibre, reducing deployment costs, but testing to ensure data is not exposed to untrusted third-parties would be required. Additionally, the implementation security of the CPRI interface would need to be evaluated, to ensure that it has reasonable robustness measures in place to handle malformed or attacker-generated inputs.

In a scenario where a local ISP or alternative network provider wishes to act as the NH provider and also as the retailer, a metro ethernet (MEN) based architecture may be appropriate, as shown in the figure below:

Figure 4 - A metro-ethernet example of a NH deployment, where a local ISP or alt-net acts as NH and retailer[45]

In this architecture, a L2 VLAN can be deployed across the provider's own network eliminating the need for a controller and making an in-venue controller optional. This architecture is similar to that demonstrated in the 5G RuralDorset project, where a "base-station hotel" setup was deployed by an ISP, and CPRI traffic carried over their ISP full-fibre network.

Paragraph 32 makes provision for synchronisation in the network and allows for it to be carried out through a range of protocols, including networked options like NTP, PTPv2 or SyncE, or alternative options like GPS or OTA. The JOTS specification permits network synchronisation to take place outside the IPSec tunnel, with appropriate firewall rules in place. In such a situation, a

[45] Ibid.

neutral host operator will need to ensure that the protocol implementations for any exposed time sync protocols are suitably hardened – as an internet-exposed synchronisation service would be a significant attack surface for an attacker seeking to compromise the system.

Overall, the presence of the JOTS neutral-host-in-building approach shows that there is a technical ability for an appropriate architecture for NH service delivery to be carried out, in a manner which is compliant with operators' security requirements. We do note, however, that the current approach to this is focused entirely on "in-building" scenarios, and that for rural deployment of NH, there would need to be greater consideration around inter-RAN connectivity for handovers between the NH radio network and the operator's wider area RAN. While this may be less of a concern for in-building use, where the benefit to the operator of delivery of coverage indoors is greater for high-dwell-time indoor users. In an outdoor environment, with greater user mobility to be expected, a lack of performant handovers would be a concern.

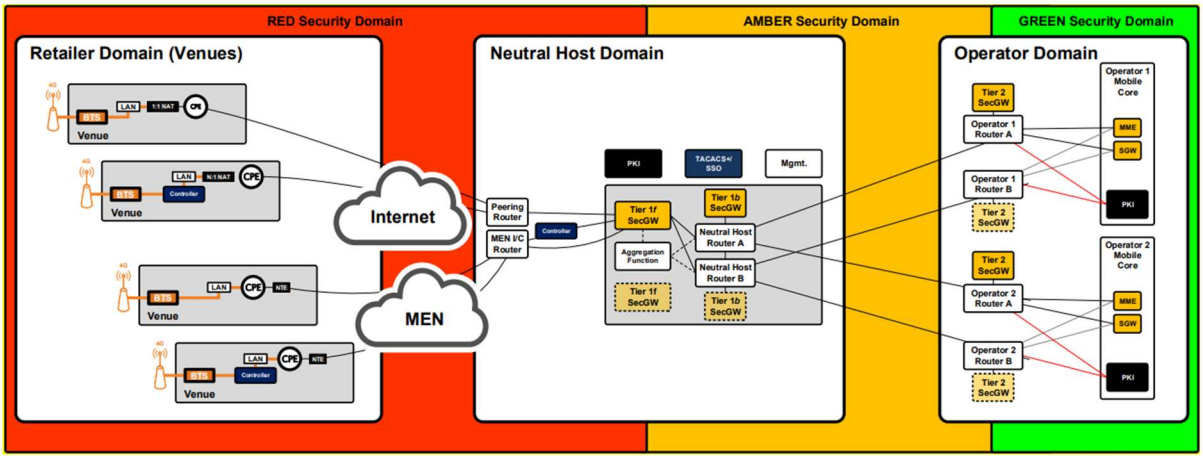

Figure 5 - Illustration of the JOTS colour-coded security zoning model for a NHIB scenario [46]

In addition, as discussed in previous reports[47] on the feasibility and practicality of NH, there are a number of other considerations around spectrum access. It is likely that a NH operator will require access to spectrum that is shared with traditional operator, which means that there will be performance and user experience benefits where both traditional operator and NH have sufficient functional interoperability to permit cell edge interference coordination to take place. While this is, as discussed earlier in this report, not a cost-effective option due to the demands of incumbent RAN vendors, it is likely that this may change with time, as a result of Government pressures for increased interoperability, openness and interchangeability of equipment in the RAN. This is likely to enable and increase the potential for rural NH service provision.

[46] Ibid.

[47] https://5gruraldorset.org/app/uploads/2021/07/5G-RuralDorset-WP6-Neutral-Host-Task-4-Spectrum.pdf

Existing MNO Shared Architectures

There are two widespread network sharing paradigms in use in nationwide telecoms networks – Multi Operator Radio Access Network (MORAN) and Multi Operator Core Network (MOCN). Both modes are standardised in 3GPP standards and widely supported in 2G, 3G, 4G and beyond[48]. In a MORAN setup, all physical RAN equipment (i.e. antennas, towers, radio units, baseband units, power supply, etc.) is shared between two or more operators. Separate radio carriers are used per-operator, meaning that each operator needs access to their own spectrum.

In a MOCN setup, in addition to physical RAN equipment being shared, the radio carriers are also shared – one RF carrier announces multiple PLMNs, allowing users from both operators to attach to the same cell carrier. MOCN therefore permits carriers to pool their spectrum if they were to choose to do so.

It is worth noting that, commercially speaking, operators are unlikely to do this, given the general scarcity-driven approach taken to spectrum in mobile networks; spectrum is finite in quantity, and mobile operators pay large amounts of money to Governments to access it. This means that operators are generally territorial around spectrum, since they use it to attempt to deliver a better service to their users. Thus, an operator with more users in a given area will generally require more spectrum to deliver the same quality of service and experience than an operator with fewer customers in that same area.

In a MOCN network, a level of aggregation and decision-making needs to take place in the RAN, to determine which mobile network the traffic should be routed to, so that it can be sent to the correct operator's core. Since users from different operators will be attached to the same cell, this decision needs to be taken in the software implementation of the MOCN infrastructure. By way of example, Parallel Wireless (an Open RAN vendor) offers a MOCN product, which illustrates that their Open RAN controller can act as an intelligent router for different subscribers, based on routing criteria like PLMN, SIM ID, or specific user identity via IMSI etc, as well as support local breakout for private network traffic[49].

Some use-cases of MOCN can result in "roaming" approaches being taken to delivery of service, involving the caching of subscriber information in other networks. This was seen in a case study involving Parallel Wireless that enabled both RAN and core sharing between Mayutel and Telefonica[50]. In this scenario, the two networks effectively interoperated through Parallel Wireless' technology, so that Mayutel had access to share a single RAN with Telefonica, while also sharing core infrastructure ("authentication was performed via inter-core network communication for a user who is on a roaming core network"). By having both operator cores coordinate through an Open RAN controller, there was a greater level of coordination than would generally be seen between operators. For example, the partners core network would be able to authenticate users from the home users' network, through authentication credential caching, and Mayutel base stations could connect to Telefonica's core network. This case study

[48] https://www.parallelwireless.com/products/multi-tenant-and-sharing/
[49] https://www.parallelwireless.com/products/mocn/
[50] https://www.parallelwireless.com/wp-content/uploads/Parallel-Wireless-Mayutel_Case-Study.pdf

is helpful as it illustrates that many of the traditional perceived barriers to neutral hosting can be worked around where suitable commercial incentives are in place. Namely, Telefonica is a household name in the mobile telecommunications space, and this carries a level of weight, which is notable.

Rural Neutral Hosting Security Considerations

A traditional mobile network's resources can be split into discrete sections, core functions, transport network, gNB and radio resource blocks, as detailed in Figure 6 below. Sharing these resources may result in security concerns from operators from an availability point of view. This is commonly known as Quality of Service (QoS), and represents an important consideration for mobile operators.

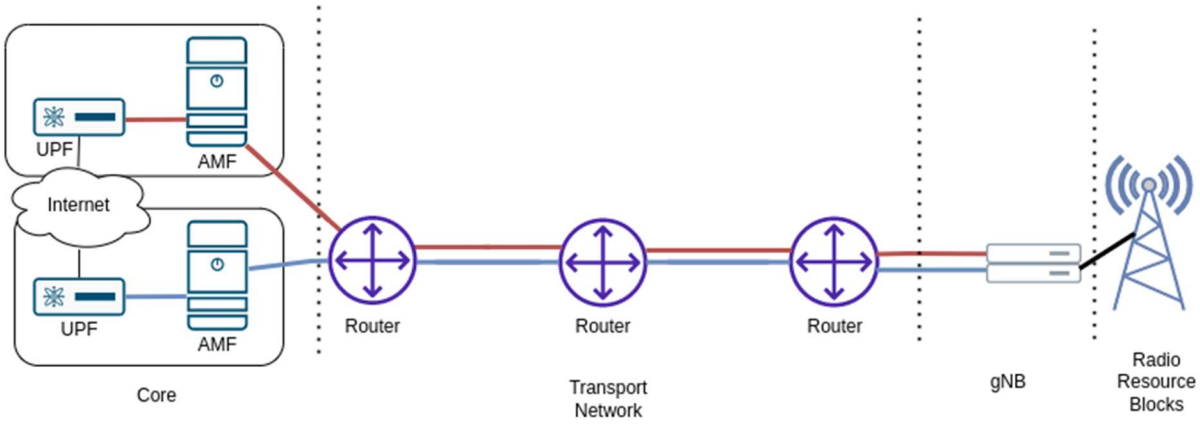

Figure 6 – Illustration of RAN network sharing opportunities in a rural NH context

Starting from the right of Figure 6, radio resource blocks have traditionally been shared by allocating specific radio resource blocks in line with each operator's spectrum holdings. However, a recent study conducted by Orange Labs in France[51] demonstrated sharing the total number of radio resource blocks between networks/users results an overall better quality of service than traditional frequency allocations. A dynamic radio resource block allocation scheme was employed based on user demand. A minimum bound was employed to ensure any potential service level agreements (SLAs) could be continued to be met. The results signified this method of sharing would result in higher bandwidth and lower latency for each set of consumers without compromising SLAs, thereby not risking the service availability.

The transport network provides the connection between the gNB/base station and the core network. In traditional infrastructure this would be routed via either leased line or a Multiprotocol Label Switching (MPLS) network. This process pre-programs defined routes into a series of routers connecting the base station to the core network. Once a data packet destinated for the 5G core network enters the network, an MPLS label is attached to the packet defining the next router in the chain to the core, and the packet is forwarded accordingly. This process is repeated until the packet reaches the final destination. The advantage of such a deployment is that the pre-determined route can additionally be configured with automatic fall-back routes in the event of a failure. MPLS networks also have an established track record of providing multiple different users simultaneous service, and over-the-top encryption and

[51] M. Kassis, S. Costanzo and M. Yassin, "Flexible Multi-Operator RAN Sharing: Experimentation and Validation Using Open Source 4G/5G Prototype," 2021 Joint European Conference on Networks and Communications & 6G Summit (EuCNC/6G Summit), 2021, pp. 205-210, doi: 10.1109/EuCNC/6GSummit51104.2021.9482466.

authentication can be added to packets. Therefore, the sharing of transport network infrastructure would not present an availability or confidentially security concern.

The sharing of Virtual Network Functions (VNFs), is one of the changes within a 5G architecture. These functions are used for routing, session management, and data egress. Modern 5G core functions are commonly structured using a series of virtual network functions. The advantage of this design, from an operator perspective, is that commodity server infrastructure can run core functions, rather than requiring decided hardware. This model has been extremely popular within the cloud services model of deploying applications where resources can be scaled according to demand. Common platforms and examples include Microsoft's Azure, Google's Cloud Platform, and Amazon's Web Services.

Securing the sharing of these resources from a confidentiality perspective, is dependent on a number of different aspects. First, securing the hypervisor. The move to virtualized/ containerised applications running on COTS hardware creates a prioritized management plane, to manage system resources and enable the more efficient use of resources. If this elevated control software is compromised, the applications running on to of this infrastructure cannot be assumed to be secure. This complexity would be additionally increased with the increasing number of components, and in order to create a responsive, consistent and repeatable environment, a software orchestration platform would likely be employed. Given this software's privileged position additional security checks would be required to ensure this orchestration platform is not and cannot be malicious.

Moreover, as the softwarization of components continue, it would also be required for a Neutral Hosting provider and any partnering incumbent operator to validate the supply chain of any software components employed. This applies to both the initial supply of software and the on-going support and maintenance, ensuring that any vulnerabilities are patched.

Secondly, within a rural Neutral Hosted environment, the financial model for local provider is unlikely to be able to support the higher expense of Tier 1 vendor equipment, whereas a national operator with significant buying power would be able to negotiate a preferential deal. This is likely to result in a national and local operator each deploying different vendors' equipment and additional interfaces, and each of the interactions between vendors presents a potential security vulnerability. Therefore, securing connections between different vendors would be required within a Neutral Host environment.

Thirdly, existing national operator commonly has physical protection around their remote infrastructure. Within a rural NH environment, the national operator may have concerns around the physical security of a local operator's infrastructure, as an intruder may be able to use the network connection as a route into the national operator's core network.

Hosted SMFs/UPFs

Sharing a data egress point, such as a UPF, presents a potential point of data compromise/inspection from existing operators' perspective. This presents a legal and technical challenge for existing operators, particularly when employing MEC elements.

In a 5G network, the UPF acts as the router/gateway to the wider internet or an external network. In the case of Multi-Access Edge Compute (commonly known as MEC), a UPF can act as a selective router, akin to a content delivery network (CDN) edge cache or facility, routing certain IP address subnets to the MEC nodes.

In traditional pre-5G mobile networks, a strict vertical hierarchy was used in building mobile networks – traffic from the radio access network was all passed up to the core network, where it would be served by one of a limited number of core network "points of presence" to the wider internet.

In the 5G world, with increasing quantities of data, and looking towards latency and QoS becoming product/service differentiators, this approach is changing. Mobile network operators are increasingly decentralising their network and introducing extra core sites, in order to have traffic egress their networks closer to the edge of the network (and thus the user)[52]. In a NH network, one of the main costs for the NH operator would be delivering sufficient backhaul to each national MNO which has NH service delivered through their infrastructure, to trunk user traffic back to the core network. Given each operator's core is separately connected to the NH, to ensure isolation, this introduces extra cost, if all user traffic is transmitted back to the core network.

One opportunity to improve this would be for a NH operator to operate their own UPF function and use MEC capabilities to egress as much user traffic as possible through their own single internet connection. This would have implications for mobility in the network, since a user entering or leaving the NH zone would receive a new client IPv4 address. Nonetheless, this is unlikely to have any major impact on user privacy, since in a 3GPP network architecture, traffic is only secured to the gNodeB (with the exception of, for example, NAS authentication traffic). This is done/achieved through shared symmetric keys which are transmitted to the gNodeB by the core network. This means that user traffic content is visible to an NH operator. From this perspective, as long as a NH is able to gain sufficient contractual guarantees from their upstream internet service providers, there is likely to be no meaningful impact on user traffic privacy in a NH architecture.

Additional security could be provided through the use of a UDP-based VPN connection to tunnel their traffic to a static endpoint. This would undo the efficiency gains of distributed UPFs but would render user traffic opaque against analysis. With the rise of use of encrypted protocols, the traffic visible to a NH operator is likely to be at packet header inspection level only – source and destination IP, port, and sequence numbers. As encryption of DNS widens in use through DNS over TLS and DNS over HTTPS, and technologies like Apple's Private Relay service grow in use, there will be even less risk from giving a NH access to user traffic. With the likely forthcoming introduction of Encrypted Client Hello in TLS connection negotiation[53], a backhaul provider is likely to only see what is conveyed in the TCP/IP headers of user traffic. Where IMS is used with an IPsec tunnel, and the tunnel is terminated in the home operators' core, this will protect content of user voice and SMS calls which are carried over IMS.

[52] https://www.ispreview.co.uk/index.php/2020/11/three-uk-add-20-data-centres-to-boost-mobile-performance.html
[53] https://blog.cloudflare.com/encrypted-client-hello/

A NH provider hosting SMF/UPF functionality would need to implement support for law enforcement retention capabilities, and there may be challenges in suitably identifying targeted users and isolating their traffic, since the UPF and similar functions would not have visibility of subscriber identifiers. There would also be security problems with placing retention capabilities at the edge of networks, outside of physically secured environments. In addition, were an operator to implement MEC, there may be obligations on an NH operator to deliver law enforcement retention capabilities on such MEC traffic. Routing targeted users' traffic via the existing operator core network may provide a detectable difference in quality of service or latency, which may tip off a law enforcement target. Further work would be required with practical trials, in order to establish the impact that local traffic routing and egress would have, and the ways of meeting law enforcement obligations by NH providers.

Inter-Radio Firewalling

The first point of compromise which must be considered in a NH environment is protecting inter-radio interfaces. A traditional radio in a RAN has X2 or Xn connectivity to proximate peer nodes, in order to facilitate accelerated handovers directly between radios, which requires direct connectivity to be possible between peer radios.

In a NH network, proper security must be put in place between radios, since it is possible that these radios could be operated by a different party – perhaps a mobile network operator, or perhaps another NH operator. In particular, it must not be possible for X2/Xn peers to be exposed to each others' management planes, as this would enable significant potential for lateral movement and management of another radio's management plane, were a radio (or its network interface) to be compromised.

In a vendor-disaggregated RAN, such as Open RAN, where commodity servers may be used in the RAN, it is particularly important to ensure that privileged protocols and management protocols, such as SSH, are not exposed to other RAN components. These services and protocols should be bound only to the dedicated management plane, and on-server (as well as adjacent-to-server) firewalls should all be configured to block and alert to such traffic and ports. The presence of attempts to initiate an SSH connection between radios would be a strong indicator of potential compromise of another node in the RAN, and would be an incident that the relevant operator should respond to.

In a NH environment, where radios from different responsible operators may need to peer together over X2/Xn to facilitate handovers or similar, it is important for each operator to implement their own "hardened edge" facing other radios in their RAN. This may require re-architecture of the security model of the RAN, to ensure that there is sufficient logging and monitoring in the RAN to detect attempts made to reach privileged services on a given node.

Isolation of RAN Management Planes

Another consideration for a Neutral Host provider is management plane protocols should only be exposed on dedicated management planes, which should be isolated from control and data plane traffic. The management plane should be specific to the operator of the radio (i.e. if a

different entity operates a given radio in a NH environment, a different management plane should be used). This helps to ensure that an attacker is unable to cross organisational boundaries in the event of a compromise of equipment.

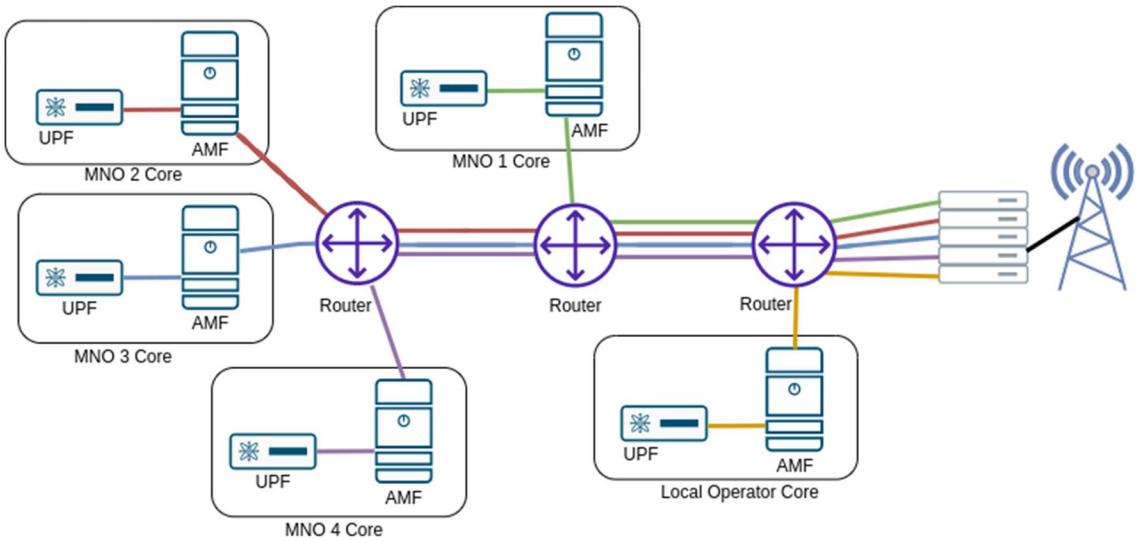

Figure 7 - Multiple mobile operators each using a private management plane to control gNodeBs

It must be assumed that if a RAN node is compromised, an attacker will be able to cause arbitrary traffic to egress on each interface, including the management plane. This means that it is also important for suitable firewalling to be in place, as well as appropriate measures to prevent lateral movement between radios.

At present, the JOTS approach to neutral hosting appears to focus on the installation of separate per-operator radio equipment. This yields separate management planes. For a NH scenario with shared radio and baseband unit equipment, this will require careful design and implementation of the shared, neutral equipment.

In addition, any NH-operator management plane access is likely to give control over the hardware or platform itself and would need to be protected to a high level of assurance. This should at least commensurate to that of an operator, but noting the potential for that access to give management access (at least to that one radio) for equipment serving each operator.

From a security perspective, the management plane of mobile communication networks presents a significant threat attack surface. Security of the management plane is important in any telecoms network, but is particularly important in a neutral hosted environment, where there is hardware sharing. As discussed earlier, in a neutral host network architecture, RAN equipment acts to bridge together multiple core networks – there is an aggregation of risk in the equipment in the RAN.

The management plane interfaces of equipment are used to configure and manage it. RAN equipment is not straightforward passive equipment – it needs to be configured based on operator parameters, such as carrier widths, centre frequencies, etc. Manual configuration of neighbouring cells and carrier aggregation pairings may also be required, depending on the level of interoperability that a given RAN device has with other RAN devices. This is particularly

relevant in multi-vendor or vendor-diversified RAN environments, where only baseline 3GPP standards of interoperability will be supported, and optional functionality, like neighbour detection and discovery, may not be supported.

All RAN equipment will have a management plane of some kind. Recent disaggregated RAN solutions, including Open RAN, for example, are likely to have standard IT management plane protocols used, such as SSH or similar, since many of the network functions in an Open RAN network will run on commodity IT servers, using general purpose operating systems, like Linux. To support scalable remote management and support of the devices, the management plane will need to be used to support remote connectivity and administration of the servers, and enable upgrading and configuration of network functions.

In a MOCN environment, for example, a number of key aspects of network operations are not covered in the 3GPP MOCN standards[54], such as segregation of usage information and counters by PLMN, QoS management, as well as fault, configuration, accounting, performance and security management. These introduce new management considerations which operators need to take into account in any kind of neutral host environment.

In a neutral host network where one operator is the "owner and operator", some of this complexity can be managed. The RAN equipment continues to be a point of aggregation between multiple mobile networks, but the equipment "owner and operator" can act as the management operator of the equipment in question. The "owner and operator" would be responsible for provisioning, maintenance and upgrades of the equipment. They would need to implement business processes to interface with each operator sharing the infrastructure, in order to carry out appropriate advertisement of neighbouring cells, and similar operator-specific activities.

Another neutral host model would be for a new-entrant or other independent operator-agnostic provider to act as "owner and operator" of the infrastructure. In this case, they would need the ability to remotely manage the network equipment, to provide an effective monitored service. This means that they would need access to the management plane interface of the RAN equipment in question. If the equipment were managed by an operator, they could use VLAN isolation to connect the management interface to an isolated VLAN on their backhaul network, and deliver remote management connectivity through this interface. Where RAN equipment is managed and operated by an independent third party, it is likely that they would need to either rent capacity from one of the operators connected to the device or provide their own connectivity to the RAN equipment.

Where the equipment "owner and operator" is an independent network provider, the provision of their own backhaul infrastructure may provide a suitable option. The operator of the neutral host environment can provide backhaul services to mobile network operators, while using their own network connectivity to provide management connectivity to the equipment.

[54] https://www.itu.int/ITU-D/tech/events/2011/CrossReg_BWA_Chisinau_October11/Presentations/CrossReg_Broadband_2011_Presentation_P21.pdf

Blocking lateral movement

From a security architecture perspective, operator RAN networks are assumed to be independent and not interconnected at the edge. Peering between operators is carefully managed and carried out in the core, using cross-network signalling interfaces.

In order to protect a NH infrastructure deployment, it is important to prevent lateral movement by an adversary within a NH environment. This means that, for example, if one radio unit or site is compromised, this should not aid the attacker in accessing another radio site.

Practically, this means that management of radios should be carried out across a dedicated network interface, which is not exposed to other radios. Network level access controls should prevent such traffic, in addition to firewalls at each radio preventing management traffic other than from designated permitted hosts upstream in the network. There should also be no static, symmetric secrets in use which are shared between multiple radios – each radio site should have its own unique secrets and use private asymmetric keys, such that full compromise of a single site does not give access to credentials which could assist in accessing or compromising another site.

If X2/Xn handovers are supported, it is important to ensure that this is carried out on a separate network interface from management traffic, and that management protocols are not bound to this network interface. In addition, for higher assurance, consideration should be given to having untrusted interfaces (i.e. those for 3GPP protocols like handover) implemented through a virtualised, hypervisor-based runtime environment, with the handover network interface exposed on a dedicated network interface card (NIC). In such a configuration, it is important to ensure that there is adequate isolation between the east/west facing handover interfaces and the management planes, such that compromise of one radio does not give rise to the ability to compromise another radio.

In a scenario world with interoperable radios, it is possible to envisage a future where there may be a desire to support handovers between NH radios and a traditional operator RAN. This would necessitate the exposure of X2/Xn interfaces between equipment belonging to both the NH and the operator, at the edge in their RAN.

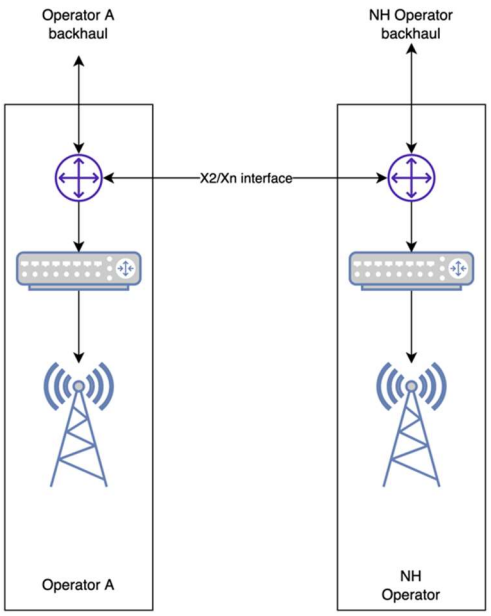

Figure 8 - Illustration of the placement of a near-edge X2/Xn interface in the RAN between a neutral host and operator

Interoperability between Vendors

To support X2 handovers between the NH radios and regular operator radios, significant care in the design of a secure solution for connectivity between each MNO's ran and the NH radios would also be required. This would create a high-risk, difficult-to-monitor interconnect directly into the operator RAN, to support handovers.

To do this securely, the NH would need to implement an isolated gateway for interconnection with each operator RAN. At areas near operator RAN zone boundaries, the NH would need to work with the operator to ascertain which zone to interconnect with. This would significantly raise the complexity of an operator RAN, by supporting this kind of interconnectivity at the "edge" of the RAN, allowing for direct peer-to-peer X2 connectivity.

There will be a tension between security and cost efficiency of a NH operator in this context, and isolation techniques for creation of these gateways should be explored. This could be done through virtualisation and platform-backed isolation of network interfaces, with an independent and discrete NH-operator management plane. This would enable high-assurance isolation of each other operator's RAN, to prevent compromise of one RAN gaining the ability to transmit traffic into another RAN (through the gateway), or compromise of the NH to allow compromise of one or more operator RANs.

Legacy Systems

2G/3G systems would prove a particular challenge where inbound roaming users are present, as it would require NH operators to consider a number of other areas, such as their obligations for Lawful Intercept (LI) support, if they would not route all voice calls through the visiting operator's core network. Due to the relatively limited number of VoLTE roaming arrangements in place, inbound roaming users are likely to require 2G and/or 3G connectivity for voice calling until it is phased out in the UK, making this a requirement in a rural NH environment.

Voice calling is a specific challenge for a rural NH operator, which would not be faced by an urban NH operator. In an urban environment, NH can offer extra infill capacity, without impacting the ability of a user to make an emergency call. Emergency calling procedures are relatively robust, allowing for domestic emergency "roaming" through emergency attaches to be used, as well as attempting (or in many cases preferring) to carry out an emergency call on a 2G or 3G CSFB network by downgrading, using the signalled multi-RAT neighbour cell information announced by a 4G or 5G network.

Cell Site Physical Security

One inherent security challenge in a RAN or NH network is around the relative ease of physical compromise of deployed edge devices, such as network equipment, radios, servers, etc. Traditionally this could be handled through alarmed cabinets with remote monitoring of alarms, and similar physical security techniques, coupled with the relative obscurity of telecoms equipment. In a rural scenario, a police response to a break-in at a cell site is likely to be lengthy, due to the distances that masts may be from local communities (especially for coverage-oriented hill-top sites).

A number of techniques can be used to mitigate this risk. Firstly, all network access from the site should be authenticated by the upstream router, which should be located at a separate location, ideally itself on a physically secured site. This could be carried out using 802.1x network level authentication or similar. This prevents an attacker from trivially connecting their laptop to the network switch, and attempting to explore the network.

In addition, secrets and cryptographic key material stored at network-edge sites should be secured on systems which are backed by hardware-root-of-trust technology (i.e. TPM 2.0), with system secure boot policies validated as part of this, to prevent the extraction of cryptographic key material or 802.1x authentication keys from the deployed devices. As Open RAN and other software-based solutions which interact with the 3GPP stack are increasingly deployed, this will be necessary to protect K_{ASME} per-SIM authentication keys from compromise if they are locally cached. This approach can also be used to protect IPsec tunnel keys and other secrets which are used (i.e. authentication parameters which can remotely authenticate to the system), in order to ensure the general integrity of the system before it comes online.

Conclusions

A rural Neutral Hosting network architecture is similar to a suitably protected 'zero trust architecture' of a national operator. JOTS has been designed with an in-building solution in mind, rather than a rural environment, but the underlaying architecture could be applied.

This demonstrates, from a security perspective, that it would be possible to provide rural NH coverage, assuming that MNOs are suitably incentivised to look for enabling solutions, rather than find commercial barriers. One way through which this could be done would be coverage obligations.

In order to deliver a secure NH solution, it is necessary for the NH provider to be security-aware, and able to architect and implement a secure network of distributed endpoint devices. The provider should also keep these updated, maintained and available, while simultaneously carrying out appropriate proactive monitoring to detect attempts to compromise those endpoint devices. To deliver this viably at scale, a zero-trust architecture would be required, using suitably locked-down endpoint equipment to act as COTS hosts for the virtualised workloads that are likely to be required to enable a NH operator to provide independent and suitably isolated interfaces to each operator.

In addition, NH operators will need the capability to manage the security of their own supply chain, and the radio infrastructure they deploy – similar to how a national MNO needs to manage the security of their RAN. In particular, since a NH radio would become a point of aggregation across multiple MNOs, albeit through independent backhaul networks, it is critical that the risks of compromise to MNO networks through interfaces such as the NAS or S1/NGAP signalling layer are mitigated. It is important to note, however, that this is not a NH-specific threat and that, were there to be a NAS-related or S1/NGAP vulnerability in operator core networks, it is likely that an adversary could compromise the network simply with a SIM card for each operator, connecting over their regular RAN.

In many ways, one could argue that an MNO unwilling to integrate with a NH on security grounds may be attempting to protect a non-zero-trust RAN that would be vulnerable to these kinds of issues already. In other words, an operator that was ready and willing to enable NH interconnections with suitable NH operators would have confidence in the robustness and security of their RAN infrastructure's peering connections over X2/Xn interfaces, the correctness of those implementations, and appropriateness of monitoring in place. The rationale for this is that a suitably hardened zero-trust MNO RAN should be designed such that a compromised node does not have management access to any other node, and that it is not in a position to allow for lateral or vertical movement in the network, thus containing the attack until the node can be removed from service and the issue remediated. This approach would put an MNO in a strong position to be able to securely interconnect with NH operators.