Building trustworthiness into future mobile networks
The widespread dependence of commerce, finance, and other sectors of society on digital networks has placed the telecommunications industry under scrutiny with respect to security, performance, reliability, and compliance with standards and requirements. As many sectors — such as utilities, transportation, and industrial manufacturing — continue to adopt automation and connectivity to improve business efficiency, greater attention to reliability, safety, and security in dimensions that go well beyond conventional telecommunication will be needed [1].

There is a general awareness that as mobile networks are integrated into critical infrastructure, special care will be needed to address the risks associated with the effects of faults and disturbances. These may be due to intentional or unintentional events and can lead to breaches in privacy and confidentiality, denial of service, and other undesirable network incidents. There is a clear demand that users of digital services should be able to trust the networks delivering those services. This is clearly articulated within charters of such initiatives as the EU toolbox for 5G security [2], the NCSC security analysis for the telecom sector, the NIST efforts on trustworthy information systems, the IIC work on trustworthiness for IoT, and the work on security assurance in 3GPP and GSMA [3]. These initiatives draw focus to the fundamental question facing the trustworthiness of mobile networks: How do we know that we can trust networks to perform as expected even in the face of new use cases, new requirements, and new types of attacks? In general, a network’s trustworthiness is composed of complex and holistic views about the integrity of the system. These views cover technical aspects as well as non-technical ones such as business processes, secure supply-chain processes, various actors’ roles and objectives, and regulatory requirements, to mention a few. The technical basis for trust in a system consists of those aspects that are derived from measurable or attestable evidence of compliance to minimum acceptable, service-specific configurable requirements on security, reliability, privacy, availability, safety, and resilience.

In this paper, we focus on the technical aspects for trustworthiness and then the answer to the earlier question lies in the collective span of inputs from multiple proof points, each contributing evidence that networks perform according to specifications and expectations for the configured service while operating in highly dynamic environments. These proof points cover aspects within the lifecycle of standardization, development, deployment, and network operation to assure security, reliability, resilience, safety, and privacy. In addition, networks must operate within the bounds of the rule of the law and, as applicable, incorporate societal values and ethical principles derived from universally understood principles. While there is a growing awareness of the basic problems, this paper explores approaches to the important representative steps needed to answer the most fundamental questions in the area of trustworthy systems design.
Figure 1 illustrates a systemization for trustworthiness in a network that is organized as a process, starting with standardization and system development before extending into a deployment phase. In the system definition phase, standards, activities, and development processes are organized around the proof points mentioned above that realize the operational system capabilities and requirements that can address the elements or operational objectives of trustworthiness. Each of these operational objectives is characterized in terms of one or more proof points of compliance, either by assertion or by way of design and implementation principles. These objectives are monitored in the dynamic system operation environment through system management capabilities supporting measurement, analysis, and reporting, interacting with the deployment characteristics to improve system configuration either by reconfiguration or through the continuous integration and continuous development (CI/CD) cycle. The relationship of design and development of a network, its operational capabilities, and the monitoring and response mechanisms allowing reconfiguration of a deployed system can then be used to mitigate and counter the risks associated with threats, vulnerabilities, or changing conditions. In this sense, the response mechanisms include protection and recovery.
An important aspect of building trustworthiness is the use of open and globally agreed standards to ensure interoperability and adherence to certain protocol and interface specifications, as outlined in [4]. In addition to this, there are various choices — both in terms of technology and methodology — that come into play within the development, deployment, and operation of mobile networks.

As shown in the figure above, this paper addresses three technology areas of key relevance to build trustworthiness in future mobile networks:

1. means for networks to achieve the level of reliability, availability, and resilience needed for use cases beyond enhanced mobile broadband (eMBB), including mission-critical use
2. confidential computing and secure identities to give provable security in cloud and edge compute as well as for connected things and industries
3. artificial intelligence (AI) to automate security assurance and network defense while still providing traceability and explainability of decision-making
Network reliability, availability, and resilience (NRAR)

Mobile networks have always been designed with reliability, availability, and resilience (NRAR) as core properties contributing to maintaining a high level of performance for telephony services. In general terms, network reliability refers to the capacity of a network to perform required actions under stated conditions within a given time interval, network availability means that the network can perform a required function at a given time, and network resilience speaks to the ability to provide and maintain an acceptable service level in cases of faults, disruptions, and challenges affecting normal system operation.

Existing NRAR mechanisms designed and deployed for eMBB include aspects such as network node pooling, geographical redundancy, and solutions with multiple frequency bands and multiple cells that provide an abundance of resources. Multiple radio access technologies (multi-RAT) can provide redundancy as well, with different complementing access technologies. These mechanisms are becoming increasingly important for new use cases where requirements on performance and service quality must be verified (for example, when robots are used in conjunction with wireless automation, functional safety requirements call for redundant radio connectivity options for diversity and link reliability). Among other issues, jitter and latency can affect critical path deadlines in a manufacturing or assembly line, allowing, for example, race conditions to interfere with the integrity of a process control system. While higher reliability and lower latency requirements typically come at a cost to capacity and throughput, the 5G system is designed to allow flexibility in configuring radio bearers to meet controllable performance requirements while continuing to evolve to address new requirements.
Security threats to networks that affect availability can result in a diverse attack surface due to (for example) design, deployment, or external vulnerabilities caused by information exposure and leakage among network and service functions. A network can be compromised by lone targeted attacks or through denial of service; the availability of mobile networks is specifically vulnerable to distributed denial-of-service attacks that can occur through a core network or air interface. Ensuring availability in the face of such varied threats will require mitigation and recovery methods aided by automated intelligence in the network.

The availability of the air interface can be improved through careful planning for local deployments, for example, by deploying multi-user multi-input multi-output (MU-MIMO) antenna technologies for additional capacity. There are also other dynamic means by which reliability and availability may be improved. An example is spectrum management to improve connectivity by the associating devices with the best choice of radio resources and with reassignment of spectrum to meet changing requirements. Radio resource management can thus capitalize on spectrum diversity to improve the availability and reliability of an air-interface. Performance management counters that are defined by 3GPP’s Service and System Aspects Working Group 2 (SA2) for radio links to provide event logging towards network management can also be used for more real-time quality assurance and metrics, such as data throughput, latency, and outage probability. Such compliance to quality of service (QoS) requirements on the radio bearer side constitutes an element of assurance delivery for network slices or applications.

Reliability and resilience depend on improved overall observability of a network. While performance management counters within radio access networks (RAN) are currently used to allow visibility of radio-level performance for network domain managers, the process required to initiate measurements and analyses of performance can be made more flexible in future programmable networks, such as by using anomaly detection events as automated triggers to invoke specific measurements. Additionally, the monitoring of core network performance can improve a network’s ability to ensure well-defined objectives for reliability, enabling service providers to react to potential threats and high-probability events that are flagged by data analysis workflows to satisfy compliance requirements. In this scenario, AI tools can help improve monitoring, adaptation, and orchestration in a highly dynamic network environment and thereby address NRAR requirements that are important to stakeholders in the mobile industry.

Other important aspects related to reliability include trusted supply chains, better quality of raw materials, stress-testing of products, provable software techniques to reduce failures and detect problems, and root cause analyses of events coupled with rapid responses to network disturbances. The increased virtualization of network functions and cloud computation will require strong reliance on a chain of trust that can be anchored to trusted hardware. In extreme cases, it may be necessary to reconfigure network routing paths or redundant implementations of network functions to reconcile faults, disturbances, or threats. Cloud implementations of network functions must also handle elasticity of demand in the face of varying loads.

Moving forward, it is essential to raise awareness of NRAR from individual components and to extend it to include all parts of the network. This will help in understanding the impact of network mechanisms in radio, core networks, and transport networks on the end-to-end, application-level experience. New types of NRAR functionalities will also be needed for the virtualized and service-based solutions deployed in 5G networks, although with similar baseline and evolved principles as those existing today.
Building trustworthiness into future mobile networks
Trust foundations: Confidential computing and secure identities
July 2020

Trust foundations: Confidential computing and secure identities

Confidential computing

The service-based architecture (SBA) for 5G core networks uses virtual network functions built from microservices in containers. Through service orchestration, networked functions are automatically placed on state-of-the-art cloud computing platforms that utilize security functions integrated within processor hardware. Hardware-based roots of trust (RoT) form the basis for security assurance in cloud systems and allow retention of data ownership as well as protection of secrets and digital identities.

The realization of confidential computing, such as software guard extensions (SGX) and secure encrypted virtualization (SEV), relies on the use of secure enclaves within processor hardware, generally known as “trusted execution environments” (TEE). The TEE executes codes and stores sensitive data in attestable, dedicated, isolated, and protected areas of the processing system.

Remote attestation is a key tool in technically assessing the trustworthiness of a service and can contribute to the assurance of the code (including version and function) that is executing and whether it is executing on the intended hardware. Through attestation, a TEE user can remotely verify properties of an environment that subsequently can be used to qualify the trustworthiness of the TEE service. Attestation can securely bootstrap containers and virtual machines for virtual network functions in infrastructure and have them securely populated with network key material and privacy-sensitive data only after verification. It is furthermore important to automate the process and integrate it with OSS/BSS and monitoring systems. Such secure, trust-bootstrapping techniques are currently being explored for 5G [5]. Examination
of whether the codes themselves are secure and free of malware must, however, be addressed through other means, such as secure software design, including analysis and formal proofs.

Another application of confidential computing is for AI, addressing proprietary concerns surrounding the protection of algorithms and underlying models along with privacy needs in multi-tenant environments. Confidential computing can be used to employ data for AI applications without exposing information by various means, including feeding encrypted data into an enclave, training an AI model without revealing cleartext, and showing only the final model. Another interesting use is in combining data from multiple parties to improve an AI model while restricting or preventing information leakage between participants [6].

Adoption of confidential computing is not without its challenges, though, which can range from choice of TEE framework, impact on code base, hardware-rooted vulnerabilities, and risks of performance degradation. Attestation requires not only a public key infrastructure (PKI), but also creation, distribution, and maintenance of whitelist information of acceptable TEE properties. The flexibility of supporting multiple confidential computing technologies within one framework would be useful in a highly dynamic environment with changing and emerging threats [7]. This is a new area, where efficient practices for mobile networks need to be defined.

Secure digital identities

Secure digital identities play a fundamental role in building trust. Identities are essential for secure communication in several layers and among several entities in mobile networks, including the connectivity layer and application layer. This includes the interaction between a mobile device and a network, intra-network communication between network nodes, and internet communication between network nodes and external servers.

Trust chains are a powerful tool — an identity used in one layer that may serve as an ROT for deriving identities in other layers. One example of this is the specification of the authentication and key management for applications (AKMA) feature [9] in 5G, where the identity in a SIM card — which was developed for subscriber identification and network access control — is also used to provide security for the application layer. The AKMA feature is primarily of interest for IoT applications but may be applied to conventional eMBB use as well.

AKMA and the improved privacy [4] that comes with 5G are features that can widen the use of SIMs as additional hardware ROTs anchored to a subscription. Interest is now increasing in using embedded SIM (eSIM) cards, which rely on embedded (non-removable) universal integrated circuit cards (UICC) along with remotely downloaded subscriber credentials. With eSIMs, IoT service providers can provision devices with a common profile of a global network service provider, enabling them to bootstrap and assume a localized profile after reporting their location. eSIM technologies offer several advantages for IoT, ranging from not requiring the use of card readers, allowing for smaller device sizes, and avoiding the impracticality and expense of manual (physical) card replacement. The industry is currently working on an integrated SIM (iSIM) solution, where the SIM functionality and associated remote provisioning support is implemented in a trusted environment within a system-on-chip (SoC) solution, known as the integrated UICC (iUICC). There
is now ongoing standardization to provide open and globally agreed eSIM and iSIM solutions suitable for all IoT devices [8].

While SIMs are important identities for devices network access, there is a need for identities in various other parts of networks. Service providers have an interest in anchoring trust for certain identities to specific functionalities in core and RAN. On the other hand, service providers and regulators have a direct interest in anchoring trust in the overall network from trusted hardware roots towards the entire chain of hardware, network software, and service-specific functions. Work is ongoing at ETSI and 3GPP to realize this and will continue in the coming years as confidential computing is adopted for use in mobile networks [9].
Assurance and defense: AI as a powerful tool

While AI is widely explored for its potential in addressing security concerns in networks, it is also important to consider security and transparency for the very use of AI itself.

AI for security assurance and compliance

Requirements and auditing for security assurance being addressed through efforts such as NESAS also contribute to increased transparency regarding the security posture of networks [3]. Meanwhile, additional effort is needed on improved mechanisms for monitoring and compliance verification of mobile networks. The monitoring of virtualized network functions with highly dynamic behavior calls for new agile security measures that are at least partially automated and designed to detect and respond to developing threat situations.

Approaches such as deep learning and probabilistic networks can rapidly detect or predict compliance failures, accelerate completion of runtime verification tasks, and (thereby) prevent or shorten the time window of non-compliance. Such methods can also be used to facilitate the investigation of incidents, including those related to non-compliance, and may help in locating the root cause and proposing recommendations for recovery. The potential of using AI to improve compliance is addressed in research work [10].

The use of AI makes it possible to analyze and understand large network systems and to determine the level of compliance with different security requirements and policies. AI can also be employed to provide rapid compliance assurance for 5G network slices that serve specific industries with distinct and stringent security requirements. The complex and dynamic nature of multi-tenanted virtualized networks necessitates the use of AI to properly interpret and verify compliance across multiple network slices.
AI to strengthen network security

AI technology can employ threat intelligence [11], attack detection and response, or automated security management for a comprehensive improvement of network security and associated defense mechanisms. In 3GPP, there exist initiatives to support network-based detection of false base stations, making use of existing mechanisms for measurement data collection [12]. With appropriate system observability, the challenge still remains to sift through observations efficiently and identify signs of false base stations. In some cases, it is possible to define rules for deviation from expected behavior from a network. In other cases, especially when an attacker is resourceful and stealthy, such deviations may defy easy detection. The design of identifiable features pertaining to radio signals and subsequent utilization in AI models is an important research topic.

The 5G system has been designed to support large numbers of IoT devices and the adoption of IoT in the automation of various tasks can have a critical impact on the security of networks. Compromised devices, for example, can severely threaten availability. AI, with its ability to automate analyses of monitoring data, is a valuable tool to address these security concerns. Indeed, AI is already being used for network-based monitoring of IoT devices to detect deviations and trigger network-based responses against malicious or faulty behavior [14].

Responsible AI

The digitalization of economic activities affords the exponential scaling of services for consumers and business entities, promising disruptive and transformative changes in many associated sectors. New business solutions can bring about positive changes towards a more sustainable and safe society; however, negative side effects born of intentional as well as unintentional misuse of information and communication technology (ICT) risks erosion of trust in the underlying technology. This has led to an interest in trustworthy AI, responsible AI, and AI ethics. The adoption of ethical principles that emphasize the ability to explain the analyses of automata and their inference mechanisms is turning out to be one of the most important requirements for trustworthy systems.

The European Commission’s ethics guidelines for trustworthy artificial intelligence (AI) offers the best enumeration of the principles of ethical AI development — including human agency and oversight — technical robustness and safety, privacy and data governance, transparency, diversity, non-discrimination and fairness, societal and environmental well-being, and accountability.

These principles are relevant for the development, deployment, and use of AI-enabled ICT services. Work is ongoing to create tools and methodologies for assessment of the underlying requirements outlined for each principle. In parallel, corporations and governments have a responsibility to raise awareness of AI ethics and to create means for an organic adoption of these guidelines as part of an industry-endorsed certification process [15][16].
Conclusion

5G ambitions can only be realized if mobile networks are trustworthy, performing according to expectations even in the face of attacks, faults, or other disturbances and offering evidence of transparency in their ability to meet those expectations. It is additionally important to conform to well-understood principles that define ethical behavior and transparency and can be explained in human terms.

Trustworthiness related to these expectations is often built through the accumulated input from multiple proof points or key performance indicators that span across network standardization, development, deployment, and operations. Thus, the creation of evidence for the trustworthiness of mobile networks involves many actors, including vendors, service providers, and regulators.

In building trustworthiness for future mobile networks, three technical areas of special relevance include: (1) the ability to achieve reliability, availability, and resilience; (2) confidential computing and secure identities to build trust, and (3) the use of AI.

It is important for the mobile industry to prioritize ongoing efforts in these areas to address security and privacy concerns, especially in at a time where there is greater demand for the openness of interfaces, for vendor inter-operability in networks, and for regulatory compliance mechanisms. Engagement is also needed to cater to the stringent and diversified requirements that will come from new use cases. This will shape evolved mechanisms for reliability, availability, and resilience in mobile networks, including many different mechanisms both on the component level and for the entire network. The industry must also collectively work on the next generation of confidential computing and its role in building trustworthiness within the highly virtualized and dynamic environments of 5G and beyond. There will be new and diversified solutions for identity management, building security and trustworthiness on all different layers in public as well as private mobile networks, particularly regarding identities for core and RAN and redesign of SIM technologies for IoT. Lastly, AI will permeate the mobile networks for many different reasons, most notably to build security and defenses, and as a tool in providing real-time monitoring and compliance verification. To secure acceptance and responsible use of AI, it is important to follow commonly agreed rules. Within mobile networks, AI will itself become part of the overall assessment and assurance mechanism of for the mobile networks. Indeed, AI is already helping to address security concerns for IoT devices.

These technologies and mechanisms, together with a vast number of other technologies and mechanisms, will help build the foundation for a trusted and trustworthy mobile networks for the future.
References

1. Key 5G use cases and requirements, May 18, 2020. 
   https://www.5g-acia.org/publications/key-5g-use-cases-and-requirements
4. 5G security - enabling a trustworthy 5G system, Ericsson white paper March 2018, re-published January 2020. 
   https://www.ericsson.com/en/reports-and-papers/white-papers/5g-security---enabling-a-trustworthy-5g-system
7. https://github.com/openenclave/openenclave
9. 3GPP TS 33.535, technical specification, authentication and key management for applications; based on 3GPP credential in 5G (AKMA),
12. 3GPP TS 33.581, technical specification, security architecture and procedures for 5G system.
13. European Commission, ethics guidelines for trustworthy AI 
Authors

Kumar Balachandran is an expert in wireless communications networks and has been with Ericsson Research since 1995. He has a BE (Hons) in electronics and communications engineering from the National Institute of Technology, Tiruchirappalli (formerly known as Regional Engineering College, Tiruchirappalli), acquired in 1986 as well as holds an M.S. and PhD in computer and systems engineering from Rensselaer Polytechnic Institute in Troy, NY, awarded in 1988 and 1992 respectively. His recent contributions have been in the area of shared spectrum, and he has been a prominent contributor to the specification of the Citizens Broadband Radio Service (CBRS) in the WINnForum and the CBRS Alliance. He is currently working on research problems pertaining to radio resilience and system reliability. He has served as a panelist and invited speaker at several prominent conferences, is well published, has contributed to several books, and has been named on over 100 issued US patents as inventor.

Ben Smeets is a senior expert in trusted computing at Ericsson Research. He holds a Ph.D. in information theory from Lund University, Sweden, where he also serves as a professor. He joined Ericsson Mobile Communications in 1998 and started out working on security solutions for mobile phone platforms. Smeets Ben is currently working on trusted computing technologies in connection with containers and secure enclaves.
Building trustworthiness into future mobile networks
Authors
July 2020

Michael Liljenstam is a principal researcher in security with Ericsson Research. His research interests include cellular infrastructure security, the intersection of AI/ML and security, network security, and trusted computing. He received his PhD from the Royal Institute of Technology (KTH), Stockholm, Sweden, in 2000 and held positions both as Research Associate at the Institute for Security Technology Studies at Dartmouth College, NH, and Visiting Research Assistant Professor at the University of Illinois, Urbana-Champaign, IL. Following that, he worked for a period as a consultant at Omicron Ceti AB, Stockholm, Sweden, before joining Ericsson Research in 2007. He has more than 30 academic publications across areas of cellular system security, software security, and network malware and has several granted patents. Michael has served on several program committees and reviewed papers for a number of conferences and periodicals.

Jukka Ylitalo is a principal researcher in security with Ericsson Research. Jukka holds M.Sc. and D.Sc. (Tech.) degrees from Helsinki University of Technology, Finland. Since 2001, Jukka has been working at Ericsson Research and different business units in various roles and has headed R&D in a start-up company. He has published scientific articles in the field of security and mobility, contributed to security standardization, and has several granted patents. Jukka is currently working on trustworthy system design for future networks.
As head of the security department at Ericsson Research, Eva Fogelström, together with her team, engages in technologies that will build security for 5G and beyond. Their scope includes areas like trusted computing, identity management, AI, post-quantum crypto, and methods for security assurance. Driving security topics in standardization forums is an important part of her work. Another part is to engage in academic collaborations and external projects, combining top-notch technology competence with know-how of mobile network security know-how. Eva holds a Ph.D. in telecommunications and an M.Sc. in Electrical engineering, both from the Royal Institute of Technology (KTH) in Stockholm, Sweden. Eva has been with Ericsson since 1997, working in the fields of security, mobility, and standardization.
Acknowledgement

The authors would like to thank Mikael Anneroth, Christoph Bauman, Vesa Lehtovirta, Jonas Lundberg, Lina Pålsson, Per Ståhl, Gunnar, Thrysin, Yosr Jarraya, Jari Vikberg, Jorden Whitefield and the reviewers for their contributions to the paper.