Reliability Implications of Emerging Technologies/Services

Alliance for Telecommunications Industry Solutions
March 2018

ATIS-I-0000065

Abstract

This report provides a qualitative assessment of key emerging technologies to identify potential implications for end-to-end service availability.

Foreword

As a leading technology and solutions development organization, the Alliance for Telecommunications Industry Solutions (ATIS) brings together the top global ICT companies to advance the industry's most pressing business priorities. ATIS' nearly 200 member companies are currently working to address the All- Internet Protocol (IP) transition, 5G, network functions virtualization, big data analytics, cloud services, device solutions, emergency services, M2M, cyber security, network evolution, quality of service, billing support, operations, and much more. These priorities follow a fast-track development lifecycle — from design and innovation through standards, specifications, requirements, business use cases, software toolkits, open source solutions, and interoperability testing.

ATIS is accredited by the American National Standards Institute (ANSI). The organization is the North American Organizational Partner for the 3rd Generation Partnership Project (3GPP), a founding Partner of the oneM2M global initiative, a member of and major U.S. contributor to the International Telecommunication Union (ITU), as well as a member of the Inter-American Telecommunication Commission (CITEL). For more information, visit www.atis.org.

Notice of Disclaimer and Limitation of Liability

The information provided in this document is directed solely to professionals who have the appropriate degree of experience to understand and interpret its contents in accordance with generally accepted engineering or other professional standards and applicable regulations. No recommendation as to products or vendors is made or should be implied.

NO REPRESENTATION OR WARRANTY IS MADE THAT THE INFORMATION IS TECHNICALLY ACCURATE OR SUFFICIENT OR CONFORMS TO ANY STATUTE, GOVERNMENTAL RULE OR REGULATION, AND FURTHER, NO REPRESENTATION OR WARRANTY IS MADE OFMERCHANTABILITY OR FITNESS FOR ANY PARTICULAR PURPOSE OR AGAINST INFRINGEMENT OF INTELLECTUAL PROPERTY RIGHTS. ATIS SHALL NOT BE LIABLE, BEYOND THE AMOUNT OF ANY SUM RECEIVED IN PAYMENT BY ATIS FOR THIS DOCUMENT, AND IN NO EVENT SHALL ATIS BE LIABLE FOR LOST PROFITS OR OTHER INCIDENTAL OR CONSEQUENTIAL DAMAGES. ATIS EXPRESSLY ADVISES THAT ANY AND ALL USE OF OR RELIANCE UPON THE INFORMATION PROVIDED IN THIS DOCUMENT IS AT THE RISK OF THE USER.

NOTE - The user's attention is called to the possibility that compliance with this standard may require use of an invention covered by patent rights. By publication of this standard, no position is taken with respect to whether use of an invention covered by patent rights will be required, and if any such use is required no position is taken regarding the validity of this claim or any patent rights in connection therewith. Please refer to [http://www.atis.org/legal/patentinfo.asp] to determine if any statement has been filed by a patent holder indicating a willingness to grant a license either without compensation or on reasonable and non-discriminatory terms and conditions to applicants desiring to obtain a license.

Copyright Information

Alliance for Telecommunications Industry Solutions
1200 G Street, NW, Suite 500
Washington, DC 20005

Contents

1. Executive Summary

Goal

This report identifies emerging technologies that have the potential to significantly impact service availability or the way that availability is measured or modeled. It does not attempt to quantify the impact of the technologies but will instead provide a qualitative assessment to understand the broad impact (positive/negative/neutral) of a technology on service availability.

Context

Network planning techniques for estimating end-to-end service availability are based on models and deployment experience for existing technologies and network architectures. In some cases, new technologies follow the existing architecture with a simple one-for-one substitution where a new technology completely replaces an existing technology. In these cases it is easy to understand the impact the new technology will have on service availability because the existing network planning methodology can continue to be used with a simple substitution of the new technology. But in other cases, a new technology can dramatically alter the network architecture and failure mechanisms. This may require rethinking approaches for assessing service availability.

Problem Statement

Emerging technologies have the potential to change service availability typically experienced in the network as well as the impact of component failures when they do occur. As a technology is developed and deployed, the service availability implications may not be considered from a holistic end-to-end perspective.

This document identifies key emerging technologies that are being currently developed or deployed and assesses the potential impact each technology could have on end-to-end service availability and resiliency, as well as how the technology could impact recovery from massive failures such as natural disasters. The analysis is limited to a qualitative assessment, simply identifying if a technology is likely to improve or degrade service availability.

Emerging Technologies and Their Reliability Assessments

This document identifies five technologies for analysis and conducted preliminary assessments of the impact each technology could have on service availability.

Network Functions Virtualization:

Current communications networks are powered by physical devices that tend to have short lifecycles as technology changes and evolves. Service availability is determined by device hardware and physical location within the network architecture. Network Functions Virtualization (NFV) leverages standard IT virtualization technology to consolidate many network equipment types onto high-volume servers, switches, and storage that can be located in centralized Datacenters. This removes many of the existing limitations of hardware and network architecture.

The industry groups defining standards for NFV explicitly recognized the potential availability implications of NFV and included several work items addressing all aspects of service availability, including scalable architectures, resiliency, and active monitoring for fault detection. NFV requirements were then specified to ensure that NFV will support existing reliability and availability requirements.

5G and Network Slicing:

Specifications for 5G technology (3GPP Release 15), targeted for completion mid-2018, promise to support ultra-low latency, high bandwidth, and more efficient connectivity for billions of IoT devices. To address the challenge posed by divergent service requirements, the 5G core network introduces the concept of "network slicing", where a common infrastructure can be configured into distinct "slices" that optimize functionality depending on the service needs. Network slicing leverages NFV to dynamically combine virtual network functions to implement the desired overall functionality.

Network slicing is unlikely to significantly alter service availability, for many of the same reasons that NFV will continue to meet existing requirements. However, network slicing may make it feasible to offer support for specialized services that require significantly enhanced performance, potentially including enhanced availability where necessary for specific services.

LTE Direct:

LTE Direct (LTE-D) is a device-to-device discovery and communication technology that allows devices to discover and communicate directly over licensed spectrum. This capability may be of interest to the public safety community to support functionality such as off-network push-to-talk (PTT) group communications. LTE Direct could enable first responders to communicate even when the cellular network is not available (e.g., during disaster recovery).

LTE-D should be considered as a fallback for public safety networks that must continue to function during catastrophic failures of the cellular network. As such, it may not have an impact on end-to-end service availability, but it could significantly enhance service survivability for critical first responder communications.

Unmanned Aerial Vehicles (UAV):

Service providers are currently identifying potential applications for drones to enhance communications networks by providing a "Cell on Wings" mounted on a low-altitude tethered drone as well as high-altitude drones. UAVs are already proving useful during disaster recovery, accelerating service restoration for first responders as well as for the general public.

UAVs have the potential to improve service availability although their greatest impact will be to enhance network resiliency by dramatically reducing the recovery time after major disasters.

Reliability Implications of Monocultures of Software Solutions

In spite of the incredible diversity of applications available today, key components of the underlying infrastructure often use common software components. In many cases, this software is open source, which can provide an advantage by allowing the software to receive the widest possible review by industry experts. However, in spite of widespread popularity, maintenance of critical components is not always well resourced. When security vulnerabilities or other faults do occur, a wide range of applications can be simultaneously impacted with potentially serious availability implications.

The widespread use of particular software solution components can offer advantages for reliability, but also creates the risk that previously dormant weaknesses may be exposed

creating sudden, widespread, failures. Continued monitoring of the service reliability implications of software monocultures is appropriate.

Conclusion

This report examines key emerging technologies to assess their impact on end-to-end service availability and service survivability. Existing service availability metrics will continue to apply as emerging technologies are deployed, in part because availability is being considered in the development of the new technologies. However, some technologies, such as 5G network slicing, LTE-D, and UAVs, could improve network survivability, particularly for first responders, and offer enhanced availability for specialized services (e.g., remote medicine) that might require extremely high availability.

All of the technologies identified in this report should continue to be monitored, but no specific actions are recommended at this time.

2. Scope & Purpose

Scope

This report:

1. Outlines network reliability and service availability expectations arising from emerging technologies.
2. Examines a representative set of emerging technologies to determine their potential impacts on service availability in terms of their current developments.
3. Provides a qualitative assessment of what can be expected from a reliability perspective, once these technologies are placed into service.

Purpose

The paper educates service providers, policymakers, and others about the challenges posed by emerging technologies from a reliability perspective. It provides an overview as to how the ICT industry is preparing in terms of potential reliability and service availability impacts arising from the introduction of these technologies. Finally, it makes recommendations for each technology whether it should be considered as:

- Watch List – Sufficient industry preparation with potentially minimal reliability impacts.
- Needs Attention – Insufficient industry preparation that may lead to potentially significant reliability impacts.

3. References

[ETSI NFV White Paper] – "Network Functions Virtualisation, An Introduction, Benefits, Enablers, Challenges & Call for Action", ETSI NFV Introductory White Paper, October 2012, http://portal.etsi.org/NFV/NFV_White_Paper.pdf

[ATIS-1000076] – "Reliability and Quality of Service Enablers – PSTN Transition to IP Packet Networks", ATIS Standard ATIS-1000076, February 2017.

[ATIS-0100003] – "User Plane Priority Levels for IP Networks and Services", ATIS Standard ATIS-0100003, November 2004.

[ATIS-0100006] – "Service Restoration Priority Levels for IP Networks", ATIS Standard ATIS-0100006, March 2006.

[ATIS-0100008] – "Defects Per Million (DPM) Metric for Transaction Services such as VoIP", ATIS Standard ATIS-0100008, May 2007.

[ATIS-0100012] – "Standard Outage Classification", ATIS Standard ATIS-0100012, April 2013.

[ATIS-0100020] – "Quantifying the Impact on IP Service Availability from Network Element Outages", ATIS Standard ATIS-0100020, October 2008.

[ATIS-0100021] – "Analysis of FCC-Reportable Service Outage Data Version 2", ATIS Standard ATIS-0100021, December 2012.

[ATIS-0100025] – "A Methodology For Estimating the Availability of Access IP routers in Terms of Customer Facing Line Card Availability", ATIS Standard ATIS-0100025, June 2009.

[REL001] – "Network Functions Virtualisation (NFV); Resiliency Requirements", ETSI GS NFV-REL 001, v1.1.1, January 2015.

[REL002] – "Network Functions Virtualisation (NFV); Reliability; Report on Scalable Architectures for Reliability Management", ETSI GS NFV-REL 002, September 2015.

[REL004] – "Network Functions Virtualisation (NFV); Assurance; Report on Active Monitoring and Failure Detection" ETSI GS NFV-REL 004, v1.1.1, April 2016.

[REL005] – "Network Functions Virtualisation (NFV); Accountability; Report on Quality Accountability Framework", ETSI GS NFV-REL 005, v1.1.1, January 2016.

[REL006] – "Network Function Virtualisation (NFV); Reliability; Specification on Software Modification Process", Draft ETSI GS NFV-REL 006, v0.0.8, work in progress.

[REL009] – "Network Function Virtualization (NFV); Reliability; Specification of Requirements to Support NFV Reliability and Availability", Draft ETSI GS NFV-REL 009, v0.0.1, work in progress.

4. Definitions & Abbreviations

For a list of common communications terms and definitions, please visit the *ATIS Telecom Glossary*, which is located at < http://www.atis.org/glossary >.

For purposes of this report, the following terms are defined:

- **Availability:**
  1. The degree to which a system, subsystem, or equipment is operable and in a committable state at the start of a mission, when the mission is called for at an unknown time (i.e., a random, time).
     Note 1: The conditions determining operability and committability must be specified.
     Note 2: Expressed mathematically, availability is 1 minus the unavailability.
  2. The ratio of (a) the total time a functional unit is capable of being used during a given interval to (b) the length of the interval.
     Note 1: An example of availability is 100/168 if the unit is capable of being used for 100 hours in a week.
     Note 2: Typical availability objectives are specified in decimal fractions, such as 0.9998.
  3. Timely, reliable access to data and information services for authorized users [INFOSEC-99].
  4. The prevention of denial of service [CESG].
  5. The property of being accessible and useable upon demand by an authorized entity [7498-2].

6. The prevention of the unauthorized withholding of information or resources [ITSEC].
7. The property of an object being accessible and usable upon demand by an authorized subject [POSIX.6].

- **Network Availability:** The probability a network can perform its required functions [T1.Rpt24-1993].

- **Reliability:**
  1. The ability of an item to perform a required function under stated conditions for a specified period of time.
  2. The probability that a functional unit will perform its required function for a specified interval under stated conditions.
  3. The continuous availability of communication services to the general public, and emergency response activities in particular, during normal operating conditions and under emergency circumstances with minimal disruption.

- **Resiliency:** The ability to provide and maintain an acceptable level of service in the face of faults and challenges to normal operation, ranging from simple misconfiguration or equipment failures to large scale natural disasters.

5. Service Availability

This report focuses on end-to-end service availability from the end user's perspective. However, it is also recognized that many factors contribute to the service availability, and that it is therefore appropriate to also consider reliability (component and network) and resiliency (the ability to maintain availability in spite of faults). These terms are distinct, but they are all related, as discussed below. The analysis in this report will assess the impact of emerging technologies on all aspects of service availability.

Reliability is the probability that a piece of equipment performs satisfactorily for a given period of time. Reliability can be stated mathematically as "Reliability = (1 − Probability of Failure)". In communications networks, equipment can be repaired or replaced, so while reliability is an important cost factor, it is not the sole determinant of the end user's

experience of a service. The end user is often more concerned with the probability that the service will be working at any given time – i.e., the service availability.

Service Availability is determined by the reliability (mean time between failures, or "MTBF") and by how quickly service can be restored (mean time to repair, or "MTTR"). Again, mathematically, service unavailability (probability the service will not be working) is equal to (MTTR / MTBF), and the service availability is equal to (1 – unavailability). This is often measured in minutes of downtime per year. Service availability can be improved by either making the time between failures longer, or by reducing the time required for repair. It is worth noting that while service availability is an important metric, it is a simple ratio that cannot distinguish between an hour-long outage every year and a full day outage every 24 years. However, these two scenarios could have a dramatically different human impact. To address this, the concept of resiliency is introduced.

Resilience is the ability to maintain an acceptable level of service in the face of faults. This is a somewhat objective metric, but in general, longer outages increase the problems experienced by end users, even if these outages occur less frequently, so resiliency is essentially a measure of the ability of the network to restore service within an "acceptable timeframe". In practice, resiliency focuses on the ability of a network to restore service after major, often catastrophic, outages. Note that resilience and survivability are often used interchangeably depending on context.

6. Emerging Technologies and Their Reliability Assessments

This section provides a representative list of emerging technologies, describes current reliability assessment work being done on each technology, and makes recommendations as to how the industry needs to react to these developments.

Network Functions Virtualization

Introduction

Currently, telecommunications networks are powered by physical devices that are often proprietary in nature with vendor-driven "bells and whistles" that drive up cost and essentially lock the network operator into one or two vendors. These physical devices tend to have short lifecycles as technology changes and evolves.

The concept of Network Functions Virtualization (NFV) "aims to address these problems by leveraging standard IT virtualization technology to consolidate many network equipment types onto industry standard high-volume servers, switches and storage, which could be located in Datacenters, Network Nodes and in the end-user premises." [ETSI NFV White Paper].

The European Telecommunications Standards Institute (ETSI) started an Industry Specifications Group (ISG) to study and standardize concepts related to NFV. This ISG has published over 60 documents, including detailed architectural specifications, technical reports, and proof of concepts (PoC) all promoting a shared NFV experience of implementation and testing. Other SDOs are also investigating specific aspects of NFV.

Current Industry Assessments on Reliability Impacts

Reliability Basics – ATIS Initiatives

Reliability basics have been developed in many SDOs. ATIS Committees, notably the former Performance, Reliability and Quality of Service (PRQC) Committee and the Network Reliability Steering Committee (NRSC), were at the forefront in the development of key industry standards related to reliability and priority aspects as summarized in [ATIS-1000076].

Key ATIS standards that apply to NFV reliability include the following:

- Admission Control Priority Levels [ATIS-0100003]: This document initiated the concept of prioritizing admission control based on the criticality of the incoming traffic. Three broad classification levels are specified: Emergency traffic (e.g., ETS, E-911) get the highest priority followed by "normal" traffic (e.g., Enterprise and business customers), and "best effort" traffic (e.g., ISP OTT). Within each level, multiple sub-levels are permitted depending on country/regional requirements or the network type (wireless, broadband) of traffic. This document is the basis for the ITU-T Specification Y.2171 for admission control priorities.
- Restoration Priority Levels [ATIS-0100006]: This document extends the priority levels for admission control to the restoration of traffic in case of network failures. The same three broad classification levels are specified. This document is the basis for the ITU-T Specification Y.2172 for restoration priorities.

- Availability Criteria: Many ATIS documents specify availability classification schemes and measurement methods:
    - Defects per Million [ATIS-0100008].
    - Standard Outage Classification [ATIS-0100012].
    - Availability Metrics for IP-Based Networks [ATIS-0100020].
    - Analysis of FCC-Reportable Service Outage Data Version 2 [ATIS-0100021].
    - Methodology for Estimating Availability of Access IP Routers [ATIS-0100025].
    - Network Resiliency Planning for Enterprise Customers [ATIS-0100028].

Many of these concepts have been included in the development of NFV Reliability in the ETSI NFV ISG work. In particular, the ATIS initiated ITU-T priority classification schemes have been specifically referenced in the ETSI documents.

Reliability Work in ETSI NFV

The goal of reliability standards development in the ETSI ISG Reliability Working Group (REL WG) is to ensure all aspects of Reliability are addressed in an NFV environment with emphasis on:

- Service Availability
- Network Resiliency

The outputs from this Working Group include:

- Reliability Guidelines/Requirements for NFV Architectures
- Methods for achieving Desired Availability

To date, five reports have been published while two specifications and two reports are in progress.

A representative set of completed and ongoing specifications work is as follows:

1. Resiliency Requirements – Technical Report [REL001]: This report was published in January 2015. It provides an exhaustive description of a wide-range of reliability aspects that could be impacted by NFV. It includes service

availability levels for admission control and restoration priorities (work initiated by ATIS and standardized in the ITU-T). It also includes guidelines for maintaining service availability and network resiliency including:

   a. Failure Detection and Remediation
   b. Fault Management
   c. Service Deployment

2. Scalable Architectures for Maintaining Reliability – Technical Report [REL002]: This report was published in September 2015. It describes a Cloud Based method for scaling out new virtual functions to meet increased traffic loads as well as the maintenance of required availability levels when virtual functions fail. The key to this methodology is the storage and maintenance of required state when scaling out new VNFs as well as instantiating new VNFs when existing virtual functions fail.

3. Active Monitoring and Failure Detection – Technical Report [REL004]: This report was published in April 2016. This report describes methods for periodic testing of Virtual Network Functions (VNF) in live environments to determine the status of virtual functions. It provides active monitoring methods for virtual functions to indicate timely detection of failures. It also provides fault notification Quality Indicators and methods of measurement.

4. Quality Accountability Framework – Technical Report [REL005]: This report was a joint interactive work between the ETSI NFV ISG and the TL9000 SDO. The main intent was to clarify the roles of these two bodies in the development of Quality Accountability in an NFV environment. The report describes key Service Quality Measurements and demarcation points.

5. Software Modification Requirements – Normative Specification [REL006]: This is a normative specification. The goal is to ensure that NFV software upgrades and updates are enabled with no negative impacts on service availability. The requirements include the modification (upgrades/updates) of VNF, Management and Orchestration Components (Orchestrator, Virtualized Infrastructure Manager, and the VNF Manager), as well as NFVI Components (Controllers, Hypervisor, Compute, Network, and Storage Nodes).

6. Specification of Requirements to Support NFV Reliability and Availability – Normative Specification [REL009]: This is a normative specification and is currently under development. Target date for completion is 1Q18. The goal of this document is to create a set of formal requirements for service availability in an NFV environment that can be utilized by network operators and equipment suppliers. The topics covered include the following:
    a. General Requirements
    b. Fault Management, Failure Prevention, & Detection
    c. Resiliency Flows
    d. Management & Orchestration Component Requirements to Support VNF Availability
    e. NFVI Component Requirements to Support VNF Availability

These documents are written with the specific purpose of providing reliability and availability guidelines and requirements for vendors and operators as they develop and implement VNF into their networks. Various SDOs (e.g., TL9000) engaged in specialized technologies that will utilize NFV have referenced the work of the REL WG from reliability perspectives.

Conclusion

Network reliability and service availability questions are being covered in the ETSI NFV ISG Reliability Working Group and is starting to receive attention from other SDOs. As ongoing Normative Specifications are published, it is expected that the telecommunications industry will have a broad range of reliability requirements that will guide the development and implementation of VNFs without significant impacts on service availability.

5G and Network Slicing

Introduction

3GPP is currently in the midst of working on 5G technology (Release 15) requirements which is targeted to complete in mid-2018. It is a wireless technology with a promise to deliver ultra-low latency requirement of 1ms and a high-bandwidth requirement of 10Gbps throughput per user, with exponential number of connections of billions of devices. Target use cases for 5G technology cover a wide range that includes Enhanced

Mobile Broadband (eMBB), Massive Machine Type Communication (eMTC), and Ultra Reliable and Low Latency Communications (URLLC).

On the RAN side, New Radio (NR) technology is introduced as part of 5G release 15. 5G NR aims to make wireless broadband the same as wireline with the fiber-like performance at a significantly lower cost. In addition, 5G NR will scale to efficiently connect the massive IoT and will offer new types of mission-critical services; and finally, another 5G NR requirement will be able to reach down to 0.5ms one-way for 5G-RAN latency. 5G NR is made possible by the new flexible OFDM algorithm and configurations, complemented by a shortened slot structure and reduced processing times with improved reliability.

On the Core network side, Service-based Architecture is introduced as part of the 5G core architecture, in which the Network Functions (NF) components now support a set of NF Services and Operations that are defined in the architecture. NF Services, NF Operations, and their interfaces are currently being defined in 3GPP SA2 and CT working groups. NF services must be self-contained and scale independently and must allow independent software upgrades without affecting other NFs or NF Services. The underlying technology for the Network Functions are based on SDN/NFV technologies. NF Services must be discoverable via a common registrar and only authorized consumers can access the NF services.

One of the challenges for 5G networks is simultaneously supporting a diverse set of services (eMBB, eMTC, and URLLC). Network Slicing is a concept that helps realize this in an economically feasible manner by leveraging one common infrastructure. The standards for network slicing is far from completed. Various architectural definitions of network slicing are being defined with subtle differences in NGMN, 3GPP, and IETF. At a high level, it is an end-to-end architecture that is logically separated (sliced) to support the specific needs of the different industry segment or services categories. It allows for a collection of network and application functions to be grouped together logically to support different categories of services (e.g., eMBB, URLLC). Each network slice draws from a common pool of physical and logical resources yet is independent and can support customized service behavior or service SLAs. Different network slices can be constructed from some common NFs and some slice-specific NFs. In the extreme case, a network slice could be constructed completely from slice specific NFs. A highly secure slice that requires complete isolation from other slices would be an example of the latter.

Current Industry Assessments on Reliability Impacts

The underlying technology for the 5G Core Network Functions are based on the virtualized technology. ETSI NFV defines the interfaces between various NFV components including the VNF/EM to the VNFM (Ve-Vnfm interface), VNFM to NFVO (Or-Vnfm interface), and NFVO and OSS/BSS (Os-Ma). The normative stage 3 specs are being defined in the ETSI NFV SOL documents and are expected to complete later this year. 3GPP SA5 working group is responsible for the Telecom Management aspects of the standards. Several 3GPP SA5 specifications related to management of virtualized network functions have made reference to the ETSI NFV documents.

The 3GPP SA5 specifies that the VNF application-specific metrics would be fed directly into the NM (from EM) using the existing 3GPP ltf-N interface, while the VNF virtualized resources data would be fed into NM from NFVO, with data coming from VNFM, VIM, and NFVI as specified by ETSI NFV. In addition, the 5G Core architecture supports features for reliability such as pooling of certain Network Functions and procedures for Network Function failover and reselection.

Various 5G use cases drive different network slices with various KPI requirements. For example, a network slice to support the healthcare segment (e.g., robotic remote surgery) or autonomous vehicle segment (e.g., driverless V2V communication) would require an ultra-low latency of less than 5ms with bandwidth throughput of 10 Mbps (or more with video support) and must be ultra-reliable with end-to-end local, geographic, and network redundancy; while the entertainment slice (e.g., virtual reality gaming) would require a similar low latency with a much higher bandwidth throughput and redundancy requirement that is not as stringent as healthcare or autonomous vehicle service. There would be no room for error, jitter, or service downtime in the healthcare or autonomous vehicle network slice. Potential key issues for network slicing management that are currently in discussion in 3GPP SA5 include how to manage a network slice that supports multiple services, and how to manage orchestration of slice across multiple administrative domains.

Exponential connections and high throughput requirements will drive network cell site densification which further adds complexity to the network because it increases the number of cell borders, where interference becomes a problem and handoffs introduce the possibility of dropped connections.

Recommendation

The stringent performance and availability requirements of 5G technology drives new work items and standardization in the network management requirements for 5G core. The high-level requirements related to reliability include:

- Policy-based process automation and orchestration to enable flexibility and elasticity in the scaling of network capacity.
- Visibility and aggregation of performance and fault metrics from all layers of the network and ability for the system to automatically scale or self-heal accordingly.
- The architecture should allow for resolution of network problems in real-time, including pooling of certain Network Functions and procedures for Network Function failover and reselection dependent on operator configuration and deployment choices.

LTE Direct

Introduction

LTE Direct (LTE-D) is a device-to-device discovery and communication technology. It allows devices to discover and communicate directly with each over licensed spectrum. While LTE-D devices do not communicate through a cell tower, LTE-D can require the authorization of a cell tower to authorize direct peer-to-peer communication.

Under the LTE-D protocol, devices both broadcast and listen for 128-bit packages of data, called "expressions", which contain basic information about their user's interests. LTE-D devices transmit expressions which indicate their interests and desires and filters the expressions they receive to search for expressions of interest. When an application detects an expression that's relevant to what it does, that application can then go into action, providing something to the user. For example, if two friends have devices that are sending out expressions, then a social-networking app that both of them use might pop up notifications for each saying the other friend is nearby.

Expressions can be private and discreet (targeted securely for certain audiences only) or public (transmitted so that any application can receive them). Public expressions are available to all devices and any application. Private expressions may be limited to specific users and or applications.

Current Industry Assessments on Reliability Impacts

Figure 1 shows the characteristics of LTE-D.

Figure 1 - LTE-D Characteristics

Of particular importance to the public safety community is the addition of LTE-D broadcast capability that can enable off-network push-to-talk (PTT) group communications in a manner equivalent to legacy public safety communication systems (e.g., P25 and TETRA). Off-network communication enables first responders to communicate even when the cellular network is down, for example, in a major catastrophic event.

LTE-D is being developed and specified in 3GPP:

- TS 22.803, Feasibility study for Proximity Services (ProSe)
- TS 23.303, Proximity-based services (ProSe); Stage 2
- TS 33.303, Proximity-based Services (ProSe); Security aspects
- TS 24.333, Proximity-services Management Object (MO)
- TS 24.334, Proximity-services (ProSe) User Equipment (UE) to ProSe function protocol aspects; Stage 3

Figure 2 shows the evolution of LTE-D in 3GPP.

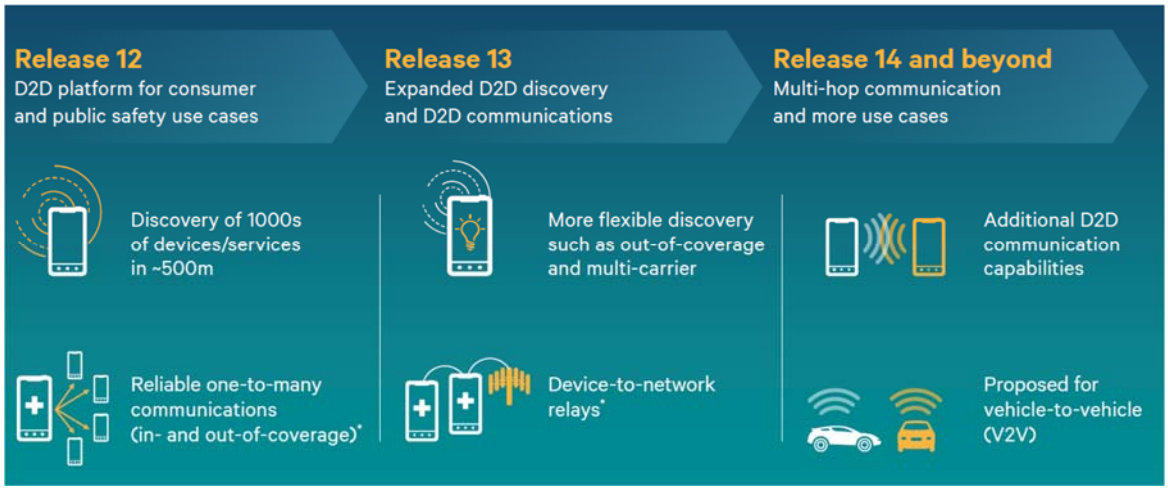

Figure 2 - The Evolution of LTE-D in 3GPP

Recommendation

LTE-D should be considered as a fallback option for public safety networks that must function when cellular networks are not available or there is a catastrophic failure of the network. It is a tool to extend the reliability of critical communication where other reliability schemes have failed.

Unmanned Aerial Vehicles

Introduction

The Unmanned Aerial Vehicles (UAV) market is currently experiencing dramatic growth, both in terms of the capability of the technology and the number of users. This is especially true for low-altitude (<400 ft.), lightweight (<50 lbs.) drones that are used for personal use (e.g., hobbyists) and for emerging commercial applications. Other classes of drones, such as high-altitude, long flight-duration drones, are also experiencing rapid technological progress, but these generally have a longer planning horizon and are less likely to be directly applicable to service availability; therefore, this analysis only considers low-altitude lightweight drones. The context of this assessment specifically considers drones that are deployed to provide a radio base station for network coverage.

There are two types of low-altitude drones: those with an internal power source such as batteries, and drones that are "tethered" to a ground station that can provide power and

communications. Both types will be considered in this assessment, beginning with free-flying drones, since this is the category that most associate with drones.

Although the technology for free-flying drones has dramatically improved recently, there are still fundamental limitations, both in terms of technology and regulation. These limitations include:

- **Payload**: The net payload capacity of current drones is limited, which means that mounting a mobile phone base station on a drone with a reasonable flight time is problematic.
- **Flight duration**: Current low-altitude free-flying drones have limited flight. Increasing the flight time reduces the net payload capacity, while increasing the payload reduces the flight time. Today, this severely limits the applications of low-altitude.
- **Backhaul**: Connecting a base station mounted in a drone into the broader network (i.e., backhaul) introduces additional weight and power limitations.
- **Regulatory restrictions**: These currently impose additional restrictions on the use of drones, including:
    - **Line of sight**: Currently drones are generally limited to operation within direct line of sight to an operator.
    - **Restricted airspace**: Drones are currently prohibited from operating in many areas (e.g., close to airports).
    - **Overflight restrictions**: Drones are currently not allowed to fly over people without their permission, which further limits the areas where they can operate.

The above limitations restrict the range of viable communication applications for drones today, but these limitations are systematically being addressed.

- **Technology**: Ongoing improvements in battery technology are enhancing flight and payload capabilities, while new generations of mobile base station technology are reducing size and weight. This is continually improving the communications capabilities of drones.
- **Regulatory:** The current regulatory restrictions on drones in many cases are a result of limited functionality of the existing generation of the

technology. As capabilities are enhanced, in particular identification and location capabilities, regulatory restrictions may be eased.

Tethered drones: The use of tethered drones can mitigate many of the limitations of free-flying drones. The tether can be used to supply power from the ground, enabling long-duration flights. A backhaul link can also be provided via the tether. In addition, many regulatory concerns and restrictions are not an issue when the drone is tethered and unable to fly beyond the limits of the tether (e.g., line of sight, overflight, and restricted airspace). However, the use of a tether also introduces additional complexity in some ways, including:

- **Power**: A long-term local power source must be provided. During disaster recovery operations, this can be challenging, although solutions such as portable generators are available.
- **Backhaul**: The tether provides a backhaul link to the ground, but connectivity to the network is still necessary. Local network failures would further complicate backhaul, especially for wired connectivity.

Although tethered drones still present operational challenges, they are significantly more feasible than untethered operation, and it is expected this will remain the case for some time. Therefore, tethered drones are expected to be predominantly used for mid-term applications.

Current Industry Assessments on Reliability Impacts

Service providers are currently identifying potential applications for drones to enhance communications networks by providing "Cell on Wings" mounted on a tethered drone. Tethered drones can partially address the limitations of free-flying drones by providing a continuous power source as well as a path for backhaul communications. Although future developments will improve free-flying drones, communications services will likely be provided only by tethered drones for the foreseeable future.

Given these limitations, even when tethered, near term use of low-altitude drones is likely to be limited to applications such as the following:

- **Events**: For well defined, short-duration events, drones may be able to provide enhanced communications. The fact that such events are typically planned well in advance in environments with good infrastructure support

means there are opportunities to address current limitations and to provide an effective tether location.

- **First responder support**: The next likely application will be support for first responders dealing with disasters that impact communications infrastructure. In this scenario, drones could provide valuable functionality even without backhaul connections to the network.
- **Disaster recovery**: Tethered drones can provide temporary base stations (i.e., a "cell-on-wings") when existing infrastructure has been destroyed. This application is already emerging and is expected to increase over time.

In the longer term, it seems likely that ongoing technology progress will allow drones to be used in a wider range of applications and for longer durations. This could include expanded use of "cell-on-wings" beyond disaster recovery to include spontaneous "events", though still largely in a tethered configuration. In these cases, UAVs have the potential to improve service availability although their greatest impact will be to enhance network resiliency by dramatically reducing the recovery time after major disasters.

Recommendation

The immediate impact of low altitude UAVs on service availability is limited, but modest technology improvements, combined with the use of tethered drones, could quickly expand use. In addition, current regulatory restrictions are being studied, and work is underway to allow these restrictions to be relaxed. Taken together, these developments suggest that UAVs will have a significant impact on service availability, particularly resiliency by reducing the duration of major network outages. It is recommended that UAVs be identified as a technology to watch for potential mid-term reliability implications.

Reliability Implications of Monocultures of Software Solutions

Introduction

Today's ICT world includes an incredible amount of diversity in the applications delivered to end-users. However, in the infrastructure underlying these services there are many areas where a high proportion of the deployment uses the same software or hardware components. Often these components are critical to the correct operation of end-user services. Across parts of the industry this could be regarded as creating a monoculture of

particular software solution components. The table below gives some examples of dominant components in particular domains.

Table 1 – Examples of Dominant Technologies for Critical Software Solution Components

Domain	Example Dominant Components (non-exhaustive)
Server CPU	Intel Xenon and other Intel CPUs
OS Kernel	Linux Kernel, Windows Kernel
Virtualization	VM-Ware
Commercial Public Cloud	Amazon EC2, Google Cloud Platform, Microsoft Azure
Containerization	Docker
SQL Database	MySQL, PostgreSQL
Secure Sockets	OpenSSL
Web Server	Apache
Smartphone OS	Android, iOS

Current Industry Assessments on Reliability Impacts

Reliability Advantages

In some respects, the widespread use of particular solution components offers advantages in terms of reliability. The ability of components, like those shown in the table above, to successfully serve many applications in use today demonstrates that they have achieved excellent operational reliability in practice.

The use of well-known and popular solution components makes it easier to access technical expertise in system design and deployment which can help avoid operational problems and increase reliability. Potential reliability limitations of popular solution components may be well-known in the industry and may be addressed by improvements to those components or by engineering solutions to avoid reliability problems.

One source of unreliability can be incompatibilities between different solution components. The use of popular solution stacks and the use of a relatively small set of alternative solutions can help avoid interworking problems as it limits the number of component combinations that are frequently encountered.

On the other hand, it should be noted that the popularity of particular solution components does not necessarily mean that maintenance of these components is well resourced. For example, it was reported that prior to the discovery of the "Heartbleed" bug the OpenSSL library received very limited funding and was maintained by only one person[1].

Reliability Disadvantages

The main reliability problem with the dominance of particular solution components occurs when a change in circumstances uncovers a previously dormant problem. This can then lead to widespread disruption to all infrastructure that uses the component. For example:

- Unix and Linux represent time as the number of seconds since 00:00:00 UTC on 1 January 1970. In 2038 this value will become too big to represent as a signed 32-bit integer, which will create a "rollover"[2]. This, potentially, could create a variety of unpredictable failures on solution components that rely on Linux time information. Similar problems with time representations may occur on other platforms such as GPS.
- The discovery of security weaknesses in common solution components can lead to a sudden wide-spread malicious attack which may disrupt the operation of services built using those components. With the rise of ransom-ware based attacks it seems likely that attackers will regard the threat of sever service disruption as a potential source of income.

For solution components that are offered "as a service" another source of reliability problems can be the failure of critical components leading to the loss of entire zones of a server cloud. There have been several examples where many services that rely on the same commercial public cloud have lost due to failures in the cloud.

Recommendation

The widespread use of particular software solution components is common in the ICT industry. This can offer advantages for reliability but does also create the risk that

---

[1] https://arstechnica.com/information-technology/2014/04/tech-giants-chastened-by-heartbleed-finally-agree-to-fund-openssl/

[2] https://en.wikipedia.org/wiki/Year_2038_problem

previously dormant weaknesses may be exposed and create sudden, widespread, failures. In today's climate, there may be particular concern about the deliberate exploitation of such weaknesses for attempted extortion or as an ideologically driven cyber-attack.

7. Summary

This report examines key emerging technologies to assess their impact on end-to-end service availability and service survivability. Existing service availability metrics will continue to apply as emerging technologies are deployed, in part because availability is being considered in the development of the new technologies. However, some technologies, such as 5G network slicing, LTE-D, and UAVs, could improve network survivability, particularly for first responders, and offer enhanced availability for specialized services (e.g., remote medicine) that might require extremely high availability.

All of the technologies identified in this report should continue to be monitored, but no specific actions are recommended at this time.