ERICSSON MOBILITY REPORT (REVIEW. PART II)

Peter Jonsson
Ericsson, Stockholm, Sweden
www.ericsson.com

DOI: 10.36724/2664-066X-2022-8-4-33-41

ABSTRACT

While 5G roll-outs are by no means complete, they are well under way. For many in the industry, efforts to utilize 5G to go beyond simply providing "fast connectivity" are already in focus. Our articles explore how the industry is already looking forward to what comes next, asking how to: make gains in sustainability; utilize technologies like IoT and edge to maximize efficiencies and push exciting use cases; use 5G as a springboard for innovation; truly capture the opportunities for all consumers and enterprises in all regions; and, with all this growing potential, how to keep 5G safe and secure. 5G technologies play a key role in modernization, providing multiples of capacity while becoming more energy efficient. Innovative network technologies enable service providers to introduce new services that in turn support societies and enterprises to reduce their carbon emission footprint. In this edition, we share some examples of how 4G and 5G technologies make it possible to unleash the power of IoT connectivity to enhance both enterprises' business performance and sustainability. The transition to cellular LPWA and 4G/5G technologies makes it possible to unleash the power of IoT connectivity. We explore the positive impact of these technologies in areas such as business efficiency and sustainability. Deploying edge computing is key to enabling latency-critical and bandwidth-hungry 5G use cases, and can cost less than on-premise IT resource for an enterprise. This capability represents huge untapped growth potential for service providers. As 5G grows in prominence due to advancing digitalization, networks become a more enticing target for threat actors. We explore the motivators, opportunities and capabilities of threat actors, and how to protect 5G networks.

KEYWORDS: _5G technologies, digitalization, mobile subscriptions._

This article was written in cooperation with Telia Company, a market-leading service provider in Sweden, providing innovative services for more digitalized and sustainable societies across the Nordics and the Baltics and with MTN, Africa's largest mobile network operator.

Information about authors:
Executive Editor of Ericsson Mobility Report: Peter Jonsson
Project Manager: Anette Lundvall
Collaborators: Katja Kalliorinne (Telia), Staffan Thorsell (Telia), Amith Maharaj (MTN Group), Emmanuel Lartey (MTN Group), Farhan Khan (MTN Group)
Contributors: Harald Baur, Greger Blennerud, Fredrik Burstedt, Warren Chaisatien, Mikko Karikyto, Anna-Maria Kastedt, Per Lindberg, Michael Martinsson, Rhys Hemi Mataira, Leena Mattila, Amardeep Mehta, Frank Muller, Ravi Shekhar Pandey, Lars Sandstrom

INTRODUCTION

5G technologies play a key role in modernization, providing multiples of capacity while becoming more energy efficient. Innovative network technologies enable service providers to introduce new services that in turn support societies and enterprises to reduce their carbon emission footprint. In this edition, we share some examples of how 4G and 5G technologies make it possible to unleash the power of IoT connectivity to enhance both enterprises' business performance and sustainability.

The transition to cellular LPWA and 4G/5G technologies makes it possible to unleash the power of IoT connectivity. With Telia, we explore the positive impact of these technologies in areas such as business efficiency and sustainability.

MTN considers 5G to be an innovation platform that could completely transform society and businesses. Here's how new ways of working will allow service providers to fully capture the 5G opportunity in Sub-Saharan Africa.

Deploying edge computing is key to enabling latency-critical and bandwidth-hungry 5G use cases, and can cost less than on-premise IT resource for an enterprise. This capability represents huge untapped growth potential for service providers.

UNLEASHING THE POWER OF IOT CONNECTIVITY

Telia Company's purpose is to reinvent better, connected living, and it strives to improve business efficiency. The transition to cellular LPWA and 4G/5G technologies makes it possible to unleash the power of IoT connectivity to enhance enterprises' business performance and sustainability.

At the crossroads of change

In recent years, Telia has seen a continuous rise in the number of cellular-connected IoT devices on its networks across the Nordic and Baltic countries. 2021 saw an increase of 44 %, more than double the growth compared to 2020.

The growth is primarily fueled by large-scale smart meter deployments, based on the low-power wide-area (LPWA) IoT technologies, NB-IoT and Cat-M.

In addition, the adoption of embedded universal integrated circuit cards (eUICC)1 has simplified the global deployment of connected devices by allowing remote SIM provisioning of multiple network profiles.

NB-IoT and Cat-M technologies are ideal for connecting massive volumes of low-cost, low-complexity IoT devices with long battery life and limited data throughput demand.

These technologies, which form part of the 5G standard, are the successors to 2G and 3G networks that are being replaced as the industry moves to adopt broadband and critical IoT, powered by 4G and 5G.

IoT devices migrating to modernized networks

2G and 3G networks are being phased out globally to enable the reuse of valuable radio spectrum for 4G and 5G deployments. By modernizing the networks with the latest technology and replacing old equipment, it is possible to realize new business opportunities and create significant energy savings at the same time.

About 30 % of all cellular IoT devices are still connecting through 2G/3G networks. However, enterprises are migrating their IoT devices and services to Cat-M and NB-IoT networks, which are more energy efficient, reliable and have higher capacities.

Across Europe, the sunsetting of 3G networks is happening before 2G, but the order and the schedule varies from country to country and between service providers. Telia will decommission its 3G networks before 2G, with the 3G sunset already in motion across Telia's markets in the Nordics and Baltics.

Globally, the number of IoT devices connected via 2G and 3G has been in slow decline since 2019 (Figure 1). The combined segment of cellular LPWA, broadband and critical IoT (4G/5G) overtook 2G/3G in terms of IoT connection numbers for the first time in 2020. LPWA IoT technologies are expected to make up about 50 % of all cellular IoT connections in 2027.

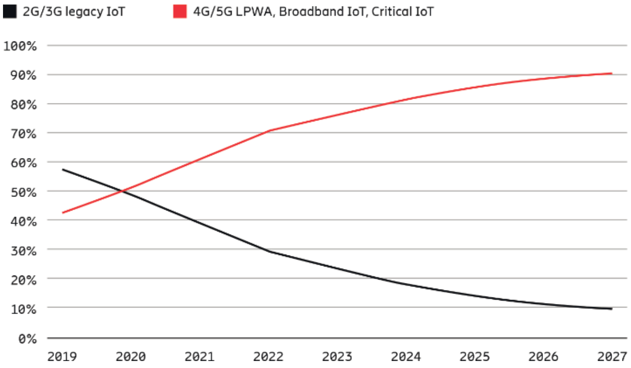

Figure 1. % age share of 2G/3G vs 4G/5G connections for cellular IoT

Source: Ericsson Mobility Visualizer.
Note: NB-IoT and Cat-M access technologies are also referred to as LPWA technologies

Extending IoT connectivity reach with cellular LPWA

LPWA IoT technology supports solutions requiring low total cost, long battery life and the ability to operate in remote locations. Its energy efficiency comes from sending smaller amounts of data at defined time intervals and then quickly powering down the transmitter in between. The two different cellular LPWA IoT network technologies, NB-IoT and Cat-M – both under the 5G standard – are inherently more secure and have longer reach than previous generations. For example, Cat-M can have a reach of up to 100km and NB-IoT up to 120km from a radio base station.

The extended reach and high-penetration capabilities make it possible to cost-efficiently connect sensors in cities, remote rural, coastal, and maritime areas, and even deep inside buildings or underground. In several tests throughout its development, Telia has shown that NB-IoT can connect devices placed as deep as 80m underground.

The transformational power of enterprise digitalization Organizations that embrace this new era of digitalization enjoy increased efficiency and cost reductions, thanks to better predictability and greater control.

Digitalization also means companies are becoming software businesses, generating proprietary data. They are no longer an isolated part in a vertical market, but a data-driven, interconnected element of a wider, digital ecosystem of services.

For example, when an agriculture machinery manufacturer equips a tractor with more than 300 IoT sensors and the ability to process more than 150,000 measurements per second, the business and its value creation changes. The tractor is now a data-generating unit, part of an ecosystem of related services such as weather forecasting, commodity pricing and crop yield predictions.

There are many more examples: A car manufacturer that harnesses IoT connectivity is no longer just selling cars, they are also enabling carpooling services and shared ownership alternatives, while gathering and handling information about the driver, roads, traveling habits and even the weather. Providers of consumer IoT services improve the health and lifestyle of consumers thanks to health monitoring, lifestyle optimization and entertainment apps.

Enterprises can be transformed and their new capabilities turned into new customer values and chargeable services. Internal processes and cost control become more effective too, as every decision can be based on real-time data.

Monitoring enables less repeating and reactive maintenance, and there is no longer a need for so many trips or manual efforts, leading to clear sustainability gains such as reduced CO2 and pollution from fossil-fueled vehicles. Smarter energy systems, smarter grids and better monitoring allow for a more efficient use of resources.

A truly data-driven, or rather data-native, company makes data the basis for all decision making, regardless of whether it relates to technology, business decisions or sustainability.

IoT connectivity goes underground for pest control

As cities expand and urbanization grows, there is typically an increase in common underground pests, such as rats.

Poison traps have traditionally been used for pest control. However, this method allows poison to enter the food chain whereby birds, foxes and domestic pets eat the poisoned rats above ground.

A pest control company in Denmark developed a new digital trap that enables an ethical, non-toxic approach. At first, the solution utilized 2G (GSM/GPRS) for connectivity, but due to the heavy steel covers below the surface of the drains, 25% of the traps could never connect to the network. By migrating from legacy connectivity to NB-IoT technology, the connection success rate rose from 75% to 100%. NB-IoT technology fulfilled the performance requirements to connect the traps deep underground, enabling performance monitoring and information gathering about the number of triggered traps, maintenance needs and sewer flooding in hard-to-reach places. This gave the company a competitive advantage.

Navigating treacherous waters with IoT Hundreds of thousands of islands making up the archipelagos of Finland and Sweden are battered by brutal storms every winter. Navigation marks are extensively deployed to support marine traffic safety, but these often break free from their anchors and float across long distances. In the past, local maritime authorities had to go out on resource-demanding and fuel-consuming runs each spring to find the marks and return them to their correct locations.

A Finland-based global provider of advanced tracking and sensor solutions took on the challenge, developing a tracker using NB-IoT and aiming to deploy them in over 20,000 navigation marks in the Finnish archipelago. NB-IoT is the ideal connectivity solution for the hard-to-reach offshore navigation marks thanks to its extensive reach and the ability to operate for up to ten years on a single battery charge. Sea routes will be digitalized by remotely tracking navigation marks, which create savings in cost and resources, reduce CO2 emissions and make the waters safer.

Remove roaming limits

Global multinational enterprises (MNEs) need to connect IoT devices across different countries and regions. For an MNE to procure local solutions from a local service provider in each market would be very challenging to implement and to operate, both technically and commercially.

Using cellular network capabilities, they can change the connectivity profile of devices through eUICC. The SIM profile is changeable over-the-air (OTA) and can be set to become a local network device to fulfill the legal requirements that exist in each market, or to have a roaming profile when allowed.

A Finland-based manufacturer of industrial and marine gearboxes, as well as drives for process industries, needed to set up easy-to-use and cost-effective mobile connections for some of its 200,000 gearboxes across 40 countries. In many critical segments of the process industry, optimized gearboxes that allow for uninterrupted operations and cost-effective maintenance are vital. Unplanned maintenance leads to production loss.

Installed sensors measure data such as oil quality, relative humidity, temperature, gearbox vibrations, pressure, and cleanliness of the equipment. Pre-installed IoT devices transmitted the relevant information to different stakeholders, such as the process control system, the operations and maintenance personnel and the equipment manufacturer. Through eUICC SIM cards, health monitoring the equipment and anticipating subscription costs became transparent and easier to manage.

Transforming tomorrow with IoT

4G and 5G networks will continue to evolve, further enhancing IoT connectivity capabilities with higher data speeds, lower latency, improved security, and extreme reliability. Supported by 4G networks, businesses can achieve better efficiencies and performance with cellular IoT technologies, and Telia's 5G network presently supports use cases such as remotely controlled high-lift wheel loaders, autonomous field robots for mechanical weed control and automated port operations.

Service providers are uniquely positioned to support the digital transformation of a wide range of industries with evolving cellular IoT technologies, as they enable industries to become truly data driven, efficient and sustainable to further contribute to a better society. As 5G and IoT transform connectivity and unlock new intelligence, the possibilities are only limited to what enterprises and service providers can imagine.

THE EVOLUTION OF MTN'S CONNECTIVITY PLATFORM

Continued investment in 4G – and the expansion of 5G – technologies are expected to play a crucial role in realizing MTN's ambitions, and will enable it to meet evolving market demands and monetize new use cases across markets in the Sub-Saharan Africa region.

MTN Group, South Africa, has defined its strategic "Ambition 2025" plan. It is built on MTN's current market position, where connectivity is the foundation, while platforms are gradually expanded to capture new growth opportunities and deliver value. In this context, 5G network deployment and evolution across markets plays an important role in enabling new services for consumers, enterprises, industries and society. For MTN, 5G is an innovation platform with the ability to transform various aspects of business and livelihoods beyond pure connectivity.

Data connectivity and usage – drivers for revenue growth

In the Sub-Saharan Africa region, connectivity is still dominated by 3G and 2G technologies, with 4G only making up around 20 % of mobile subscriptions by the end of 2021.1 However, demand for data connectivity and digital services is increasing across markets. Operating in 18 markets across the Middle East and Africa, MTN is pursuing these new growth opportunities.

Continuous network modernization and coverage build-out, supported by MTN's Rapid Rural Rollout (R3) program, has enabled it to capture strong new subscriber growth and stimulate increased data usage. This has resulted in increased data service revenues, despite price pressure in the markets. In South Africa, MTN networks experienced strong data growth as the number of customers actively using the internet grew by 12.5%, leading to a mobile data traffic growth of almost 60% in 2021. The average mobile data traffic per pre-paid subscriber was 2.3GB and 10.3GB for post-paid subscribers.

MTN considers data as a main driver of revenue growth over the medium term. Initiatives to stimulate further data adoption include data service bundling, segmented value propositions and the development and launch of freemium data propositions, supported by strategic over-the-top partnerships.

MTN's strategic priorities up to 2025

MTN continues to invest in 4G technologies and has expansive plans for 5G to realize the opportunities it has identified to evolve and expand its service offerings for the consumer, enterprise and industry segments. MTN's strategic priorities are articulated in its Ambition 2025 strategic framework, which is underpinned by 10 key technology strategic pillars intended to enable growth in connectivity and platforms businesses.

Some of the most important pillars are ensuring best-in-class, ubiquitous access across mobile and fixed networks, maintaining network leadership and efficiencies, and the monetization of infrastructure. Other priorities include investment in sustainable technologies and zero-touch, service-aware networks. 5G networks will play an essential role in delivering on the technology pillars to realize the Ambition 2025 plan.

Monetization of network infrastructure includes a network-as-a-service (NaaS) strategy, where network sharing (national roaming, MOCN and MORAN) is the starting point, followed by 5G network slicing which enables exposure of network functionality via APIs to build new enterprise services. An additional step will be the monetization of data exposed via online third parties.

Building 5G for timely monetization

MTN's 5G network build-out strategy is based on meeting evolving market demands with the timely deployment of the relevant technology enablers, in order to optimize the potential for monetizing new use cases. So far, 5G subscriber uptake has been driven by a combination of increased 5G device penetration and fixed wireless access (FWA) subscriber uptake. The average 5G subscriber mobile data consumption is approximately twice that of 4G subscribers.

Mobile broadband and FWA are currently the main 5G services marketed by MTN. It stresses better user experience as the main value, in a manner that relates to consumer needs, rather than bandwidth and latency which are not relative to the consumer. Interest in high-speed, good-quality broadband increased as working from home practices spread during the pandemic. 5G FWA will compete with fiber-to-the-home as an alternative, cost-efficient home broadband solution.

The deployment of 5G SA architecture, enabling network slicing, will be driven by consumer and enterprise use case evolution over time. In the 2023-2024 timeframe, the initial target will be consumers (enhanced mobile broadband/FWA).

This will be followed by deployments for enterprises, as ultra-reliable low-latency communications (URLLC) for critical services – which are crucial for high-end industrial applications – and 5G-era massive machine-type communications (IoT) use cases start to emerge. Over-the-top services will also be an important offering to create stickiness.

The challenge of migrating to SA architecture is not related to the technology as such, but rather how to monetize these new types of services, while also adhering to local market regulations related to net neutrality.

The enterprise opportunities

5G will enable a range of new services across different sectors, such as mining, manufacturing, utilities and agriculture. MTN is sharing information with enterprise customers and industry verticals about the value of 5G connectivity and low latency for optimizing its operations, as well as the introduction of new services. Dedicated private networks are already being deployed in proof-of-concept trials to validate the value of new services.

An AI-based face recognition system at mining sites is one example of a service being evaluated – this is currently 4G based, but will evolve to 5G. According to MTN, the main new opportunities in the African market that can be addressed with 5G technologies are related to areas such as virtual education, industrial automation, telemedicine, remote health care and smart cities.

MTN's 5G deployment strategy

5G is still in its infancy in South Africa. Within the country, MTN is a leading service provider, with around 35 million mobile subscribers. Of these, about 50 % are active mobile data users. At around USD 6.30, it has the highest blended average revenue per user (ARPU)2 of all service providers in South Africa.

MTN launched its 5G commercial services in June 2020 and reached 200,000 5G subscribers by the end of 2021. Continued 5G subscriber uptake will be strongly impacted by the availability of a wider range of low-cost 5G smartphones. In the recent spectrum auction, MTN acquired 100MHz of spectrum across three frequency bands: 40MHz in the 3.5GHz band, 40MHz in the 2.6GHz band and 2x10MHz in the 800MHz band.

MTN's initial 5G network deployment strategy focuses on high-value urban areas and hot spots, where they will deploy high-quality 5G New Radio (NR) equipment on the mid-band 3.5GHz frequency (40MHz bandwidth) as a capacity layer. Initially, hot spots being targeted include key markets, university locations, institutions and residential areas serving consumers with high data usage potential.

Long term, a coverage layer on the 700MHz band will ensure that regulatory requirements for 5G coverage are fulfilled. Deployments in high-band spectrum (mmWave) will be carried out on a more limited basis in areas with high-capacity traffic demand and in areas for deployment of FWA services. 5G is also available in some areas through dynamic spectrum sharing (DSS), a technology which allows both 4G and 5G to be deployed in the same band and on the same radio. MTN has deployed about 1,000 5G mobile sites and aims to reach 25 % 5G population coverage by the end of 2022, and 60 % by 2025.

MTN will begin decommissioning its 3G network in 2025/2026, with 4G and 5G becoming the principal technologies used to deliver telecoms services to its customers. In Sub-Saharan Africa, 5G subscriptions will represent around 10 % of all mobile subscriptions by 2027,3 with South Africa expected to lead the adoption rate in the region. Local market research forecasts that the number of 5G subscribers in South Africa is expected to reach 11 million by 2025.

Strategy execution – addressing the new opportunities

MTN knows that the traditional business models and "ways of doing things" will not be sufficient to enable it to make the most of the emerging 5G opportunities. To really benefit from 5G's capabilities, MTN will need to tie its 5G vision and roadmap closely to its digital transformation strategies. It will need to introduce network slicing if it expects to see revolutionary business models and service pricing.

Network slices will be created on demand and will be independently controlled and managed with a degree of customization that could previously only be achieved with dedicated physical networks. Network slices allow partners to integrate into network platforms in a similar way to a dedicated private network, but with far less effort. They will also enable MTN to expand its role from connectivity to other areas of the value chain – such as cloud and edge services, orchestration and applications.

ENABLING DEMANDING USE CASES WITH CSP EDGE COMPUTING

Edge computing is key to enabling latency-critical and bandwidth-hungry 5G use cases, representing significant growth potential for communications service providers (CSPs).

Demand for immersive use cases has been held back by factors in the development of a new ecosystem, including networks, devices and applications. As this ecosystem matures, we expect the value brought about by edge computing will overcome the cost advantages held by large-scale data centers. Our analysis indicates that it is clearly possible for a CSP to build-out edge computing with an annual cost base not materially higher than a data center.

Historically, enterprises could either run their application workloads on-premise, based on the company's own IT infrastructure, or hosted in centralized data centers. There are several fundamental differences between these deployment options, including cost, control, security and regulatory compliance.

With the rollout of 5G, CSP mobile networks present an attractive proposition for running demanding enterprise applications close to target customers.

A cost analysis of deployment shows that the cost to CSPs to deliver edge compute resources to enterprise customers is nearly half of what it would cost for an enterprise to build its own on-premise infrastructure with similar performance, reliability and data security.

Enter edge

Edge refers to the distribution of compute resource and applications to geographically distributed sites on the premises of an enterprise or in a CSP network. It provides compute resource and storage closer to where the data is generated and consumed. It offers significant advantages by enabling advanced data processing capabilities located close to where they are needed, reducing the latency inherent in centralized data centers. Deploying software at the edge comes with an increased cost compared with centralized deployment, but also enables a range of enhanced capabilities, including increased performance, reliability, data security and privacy, as well as reduced cost/bandwidth for the transport network.

Since data does not need to travel to remote locations for processing, analysis and rendering, enterprises can save precious milliseconds on round-trip times (RTT) while benefiting from more reliable data throughput. Enterprise on-premise edge computing can help insulate their networks from cyberattacks and distributed denial-of-service (DDOS) attacks on more centralized locations.

There is also reduced risk of data being intercepted in transit, further adding to the security and privacy features of edge computing.

Edge computing can help organizations to fully comply with jurisdictional data regulations and sovereignty laws by allowing data to be processed close to its source.

CSPs can leverage the proximity of their existing sites to end users to set up edge compute, providing low-latency and high-performance IT capabilities for enterprise workloads as a service. For example, one way enterprises can reduce on-premise IT infrastructure is by deploying “infrastructure-less” branch offices; all IT on-premise applications, from communication, image processing and analytics to specialized enterprise services, can be hosted on the network edge.

A number of considerations must be addressed while rolling out compute capabilities alongside connectivity. There can be limitations to adding resources to some sites due to constraints on space, power and/or network capacity. Another challenge could arise from low fault tolerance of the commercial off-the-shelf (COTS) hardware used at the edge sites. CSPs may also require new sites to provide both continuous coverage and compute capabilities at critical locations to enable particularly demanding use cases.

The cost of the edge

To compare the cost of deploying compute resources at different scales, we convert capital expenditure into depreciation by dividing each asset category by the number of years it will be written down, and then add the resulting depreciation to the annual opex, providing a snapshot of the yearly cost structure. For example, power and cooling systems are written down over 14 years, whereas COTS servers are typically written down over 3 years.

Capex includes:

- Server capex is mainly the cost of COTS servers and virtualization software.
- Other capex consists of the cost of components such as power distribution and cooling systems.

Opex includes:

- The electrical power required to run and cool the servers.
- Other opex, mainly the cost of operations and maintenance (O&M).

As an example, we estimate the cost of compute resources for a CSP in Sweden. Initially, edge compute rollout is expected to be on aggregation sites having power capacity installed up to 10kW, hosting an average of 8 server units, each with 4 cores. With approximately 8,000 access sites and 1 aggregation site per 10 access sites, there is a virtual processor (vCPUs) capacity of 25,600 (800 sites x 8 servers per site x 4 cores per server) for enterprise applications at CSP-owned edge sites.

Capex depends on the required capacity plus redundancy in the edge hardware components to meet the reliability requirements for edge services or applications. The geographic distribution can also be leveraged to improve the system availability by avoiding a single point of failure. We categorize the capex into server capex and other capex due to the faster cycle of server performance improvement compared to others. Servers are typically depreciated over 3 years while investments in power and cooling systems are depreciated over 14 years.

Upgrading aggregation sites with edge compute capability, with an average of 8 units of servers, can draw up to 1.6MW (800 sites x 8 servers per site x 250W per server) for running the servers. With an assumed power efficiency factor of 2, 3.2MW power is needed on average to power all the aggregation sites. The cost of compute resource at each aggregation site is estimated be around USD 20,000. Hence the USD per critical watt for an edge site is USD 20,000/(8 servers x 250W/server) = USD 10/W. This cost is very similar to USD per critical watt for building a large-scale data center.

Opex is the sum of electricity cost and O&M. For the current study, we assume it to vary in the range of USD 0.10-0.15/kWh. For O&M, the cost of full-time employees required to manage and maintain the distributed edge servers is projected.

We constructed four different scenarios to estimate and compare the compute resource cost, based on USD per vCPU-hour.

- Scenario 1 is a base case with costs assumed for a small- or medium-sized enterprise handling its compute needs with its own IT infrastructure.
- Scenario 2 is an estimation of cost for a large-scale data center to provision the same capacity as the first case.

- Scenario 3 is built around provisioning the capacity used in the first two cases by deploying edge computing on the CSP network.
- Scenario 4 is an extension of the third case, with the addition of the cost to implement a set of measures to reduce power consumption. These include using renewable energy, dynamic usage of battery/power storage at peak times and advanced cooling technologies, including a heat exchanger for the server cabinets.

Server capex is the most significant parameter for all the scenarios except the base case where O&M (other opex) dominates due to the lack of scale.

Electricity cost is the second largest factor in USD/CPU-hour for scenario 3. This leads to the significance of additional power efficiency elements in scenario 4.

With an estimate of expenditure in use cases suitable for edge deployment, the cost of edge compute resources can be just 10 % more than that of a large-scale centralized one. Capacity utilization is the most important parameter for increasing the cost efficiency of the edge resources (Fig. 2).

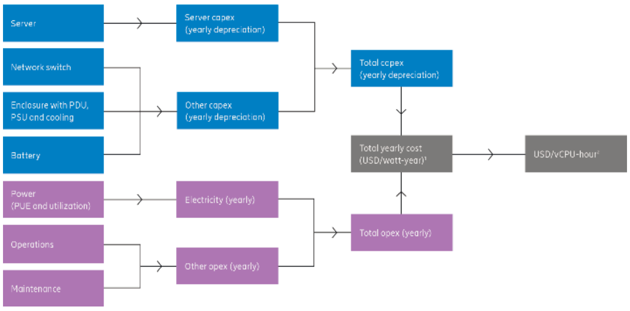
Server
Server capex (yearly depreciation)
Network switch
Enclosure with PDU, PSU and cooling
Other capex (yearly depreciation)
Battery
Total capex (yearly depreciation)
Total yearly cost (USD/watt-year)¹
USD/vCPU-hour²
Power (PUE and utilization)
Electricity (yearly)
Operations
Other opex (yearly)
Maintenance
Total opex (yearly)

Figure 2. Annual cost estimation framework for compute resources

CSP edge infrastructure resources are marginally more expensive than those at a large-scale data center but much less than those at an enterprise on-premise compute solution. CSP edge infrastructure also provides better latency and proximity to enterprise applications.

When comparing the costs of a large-scale traditional data center and a CSP network edge, we need to consider that those alternatives enable different use cases. Positive features for a CSP-operated edge include high location sensitivity, reduced latency (in the millisecond range), and guaranteed connectivity. However, edge compute infrastructure will have limited scalability compared to large data centers.

The short- to mid-term edge opportunity for CSPs should be seen in the wider context of the enterprise opportunity, where edge computing will be an enabler for a broad range of use cases, for example offerings such as private 5G networks, IoT platforms, cloud gaming and immersive experiences with XR. In the long term, when compute is deeply integrated in mobile networks, the most demanding use cases, including closed-loop industrial control systems, industrial robotics, extended reality with real-time synchronous haptic feedback (the Internet of Senses) and negotiated automatic cooperative driving for autonomous vehicles, will open up an expanding set of opportunities.

SECURING 5G NETWORKS IN AN EVOLVING THREAT LANDSCAPE

5G is, by design, more secure than previous generations, but it is being deployed and operated in an evolving and complex threat landscape. New, demanding use cases served by telecom networks can increase attack motivations and attack vectors are multiplying. These factors are exponentially increasing the need to protect networks.

The evolving 5G threat landscape

With the introduction of 5G and billions of new devices, the threat landscape in which telecom networks operate is evolving significantly. Networks provide vital infrastructure for business-, mission- and society-critical applications, and as a result, threat actors are motivated to constantly evolve to seek out weaknesses.

Safeguarding 5G networks

As the value and volume of personal, business sensitive and public service information increases with continued digitization, security and privacy laws and regulations have been expanding. This is a reaction to decreasing risk tolerance and the deteriorating cyber security environment.

Regulators know the importance of 5G and see safeguarding these networks as vital. The threat landscape for 5G is more complex than with previous generations due to the convergence with traditional IT, enabling IT threat actors to attack telecom networks in a similar way. In addition, networks often have new functionalities, such as network slicing for service separation and isolation, along with an increased use of AI/ML for automation. While AI is widely explored for its potential in addressing security concerns in networks, it is also important to consider the security and transparency of AI. Edge computing places cloud resources closer to the access, bringing new challenges whilst enabling mission-critical, low-latency applications.

Attacks on telecom networks are rising

Threat actors are increasingly skilled and pervasive, and attacks are becoming more frequent. Research from CrowdStrike, a US cyber security company, shows which industry verticals are most frequently impacted by targeted intrusions.1 The data showed that, between July 2020 and June 2021, the telecom industry was the most targeted, attracting 40 % of attacks compared to 10 % for the next-highest industry vertical. It should be noted that the data does not distinguish between the telecom enterprise and the telecom network intrusions for the industry.

What motivates threat actors?

The main motivations to target telecom networks are surveillance/espionage, financial gain and disruption/sabotage. In recent years, the most common type of attack in the cybersecurity landscape has been the deployment of financial gain ransomware.

To achieve bigger payoffs, ransomware operators have shifted their targeting to high-profile organizations in industries such as manufacturing. Threat actors know this industry sector has a low tolerance towards downtime and is more inclined to pay out as a result.

With increased use of 5G within different industry verticals' networks, the motivation to attack 5G networks should be looked at from the perspective of the related industry sector.

Personal data is also always of high interest. One objective of espionage is to obtain call metadata, especially call detail records (CDRs). This means customer billing and customer care systems are primary targets. LightBasin was observed targeting business support systems to obtain CDRs.

Disruption is the least typical of these motivations for targeting telecom networks. These attacks often have their roots in ideology, driven by personal, group or nation-state agendas. During the first quarter of 2022, a number of these attacks occurred on European networks, including targeted attacks to prevent local gamers from participating in a tournament and network-wide disruptive cyberattacks, putting critical services at risk.

Due to a shift in the tactics used by cybercrime and nation-state threat actors, and the increasing use of common IT platforms in telecoms, the likelihood of attacks has increased.

The opportunities for threat actors

New features within 5G networks bring many advantages, enabling new use cases. However, the technical complexities can create new opportunities for threat actors.

The ongoing transformation to cloud native introduces new concepts, new deployment methods and more complex partnership structures. With this trend, deployments are becoming more complex. This requires new types of competence and skill sets, from both vendors and service providers. Consequently, the risk for misconfigurations, which expose weaknesses, is increased.

Vulnerabilities in virtualization, cloud services, or network slicing can have a considerable impact, as they may enable access to unauthorized resources (Fig. 3).

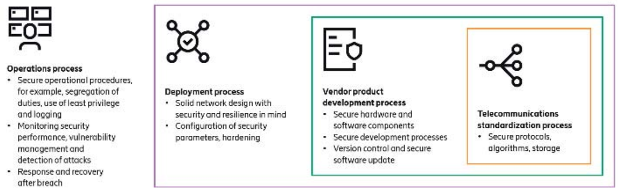
Operations process
• Secure operational procedures, for example, segregation of duties, use of least privilege and logging
• Monitoring security performance, vulnerability management and detection of attacks
• Response and recovery after breach
Deployment process
• Solid network design with security and resilience in mind
• Configuration of security parameters, hardening
Vendor product development process
• Secure hardware and software components
• Secure development processes
• Version control and secure software update
Telecommunications standardization process
• Secure protocols, algorithms, storage

Figure 3. Protecting 5G end-users requires a holistic approach including the four key layers

5G will connect billions of devices, and not all these devices have sufficient security protection. Devices used for Industrial IoT are often optimized for a specific task, with design driven by cost efficiency.

Vulnerabilities in these devices can be used to target the 5G network, or the industry vertical. This requires protection of devices to be provided from the network side. In general, any exposed interface provides an initial entry point for a threat actor. LightBasin accessed target networks via incorrectly exposed interfaces on the GPRS roaming exchange (GRX), a closed inter-service provider network.

Threat actors are increasingly using valid credentials for accessing targets. In addition to the traditional social engineering techniques for obtaining human identities, threat actors are looking for weaknesses presented by the surge of machine identities that are needed in cloud-native deployments. Strong multi-factor authentication, with management and monitoring of privileged accounts, is essential to prevent and detect account misuse. It will also limit the impact of credential theft and the exploitation of vulnerabilities.

What are the capabilities of threat actors?

Threat actors have shown the capability to build targeted and context-specific malware. Nation state threat actors routinely exhibit good operational security and use various defense evasion techniques to hide their activities, making it possible for them to move laterally in the target organization before being noticed. For instance, LightBasin carefully deleted traces in log files after their activities.

Threat actors try to blend their communication into normal traffic and use legitimate protocols, such as ICMP and HTTP. In addition to these, LightBasin used telecom-specific protocols to bypass firewalls and stay under the radar.

As the industry moves away from proprietary protocols and dedicated infrastructure, intrusion of telecom networks does not necessarily depend on extensive knowledge of these networks and their protocols. Threat actors targeting telecommunications networks will increasingly resort to routine vulnerability exploitation, supported by public availability of exploit code.

Even though 5G interconnects are more secure, older network generations will be used for several years, and attacks via interconnected interfaces will continue and will be more complex and difficult to detect as threat actors increasingly focus on defense evasion.

Trust in mobile networks is paramount

Trust in mobile networks, especially 5G, is the foundation for digitalization. To enhance trust, the GSMA Network Equipment Security Assurance Scheme (NESAS), jointly defined by 3GPP and GSMA, provides an industry-wide security assurance framework to facilitate improvements in security levels.

NESAS defines security requirements and an assessment framework for secure product development and product lifecycle processes, and uses 3GPP-defined security test cases for the security evaluation of network equipment. NESAS is intended to be used alongside other mechanisms to ensure a network is secure and, in particular, to ensure an appropriate set of security policies covering the entire lifecycle of a network is in place.

3GPP standardization made major improvements in terms of security and privacy compared to 4G. 5G has been designed with new functionality that is intended to make it more resilient towards various existing frauds, subscriber privacy and eavesdropping issues, than earlier generations.

For instance, the industry is putting considerable effort into protecting the interconnect networks between the service providers, encrypting, and otherwise hiding subscriber identifiers, and preventing the modification of the user data sent between user equipment and radio base stations.

5G also provides a standardized and well-defined way to deploy zero-trust functions like authentication and authorization of API usage, and protected communication between and to the 5G network functions.

It's time for the active defense of telecom networks

With networks being used in new contexts, connecting a greater variety of mission-critical processes, it is no longer enough to rely solely on standardized and regulatory-based security controls. Now the active defense of telecom networks is also required.

The entire industry is currently accelerating the journey from passive defense to active defense strategies. The embedded security inside network products is critical but still not enough. The telecom networks of today are built to evolve, and security must do the same.

Securing 5G networks

Telecom networks' availability and performance are more valuable than ever, which makes them attractive targets for malicious actors. Powerful security monitoring and automation, identity management, effective incident response handling and solid business continuity planning are critical to securing networks. Building a secure 5G network requires a holistic approach, rather than a focus on individual technical parts in isolation, to protect end users. Network operations is one of four key layers enabling the holistic approach, alongside standards, product development processes and network deployments.

Methodology

Forecast methodology. The forecast time in the Mobility Report is six years. The subscription and traffic forecast baseline is established using historical data from various sources, validated with Ericsson internal data, including measurements in customer networks. Future developments are estimated based on macroeconomic trends, user trends, market maturity and technological advances. Other sources include industry analyst reports, together with internal assumptions and analyses. Historical data may be revised if the underlying data changes – for example, if service providers report updated subscription figures.

Mobile subscriptions. Mobile subscriptions include all mobile technologies. Subscriptions are defined by the most advanced technology that the mobile phone and network are capable of. LTE (4G) subscriptions, in most cases, also include the possibility for the subscription to access 3G (WCDMA/HSPA) and 2G (GSM or CDMA in some markets) networks. A 5G subscription is counted as such when associated with a device that supports New Radio as specified in 3GPP Release 15, and connected to a 5G-enabled network. Mobile broadband includes radio access technologies HSPA (3G), LTE (4G), 5G, CDMA2000 EV-DO, TD-SCDMA and Mobile WiMAX. WCDMA without HSPA and GPRS/EDGE are not in-cluded. FWA is defined as a connection that provides broadband access through mobile network enabled customer premises equipment (CPE).

Subscribers. There is a large difference between the numbers of subscriptions and subscribers. This is because many subscribers have several subscriptions. Reasons for this could include users lowering traffic costs by using optimized subscriptions for different types of calls, maximizing coverage and having different subscriptions for mobile PCs/tablets and mobile phones. In addition, it takes time before inactive subscriptions are removed from service provider databases. Consequently, subscription penetration can be above 100%, which is the case in many countries today. However, in some developing regions, it is common for several people to share one subscription, for example via a family- or community-shared phone.

Mobile network traffic. Ericsson regularly performs traffic measurements in over 100 live networks covering all major regions of the world. These measurements form a representative base for calculating worldwide total mobile network traffic. Mobile network data traffic also includes traffic generated by FWA services. More detailed measurements are made in a select number of commercial networks with the purpose of understanding how mobile data traffic evolves. No subscriber data is included in these measurements.

Population coverage. Population coverage is estimated using a database of regional population and territory distribution, based on population density. This is then combined with proprietary data on the installed base of radio base stations (RBS), together with estimated coverage per RBS for each of six population density categories (from metro to wilderness). Based on this, the portion of each area that is covered by a certain technology can be estimated, as well as the % age of the population it represents. By aggregating these areas, world population coverage per technology can be calculated.

REFERENCES

[1] https://www.ericsson.com/en/reports-and-papers/mobility-report.

[2] Ericsson ConsumerLab, 5 ways for a better 5G. 2021.

[3] Source for network and device statistics: GSA and GSMA. 2022.

[4] https://www.ericsson.com/en/reports-and-papers/mobility-report/mobility-visualizer.

[5] https://www.ericsson.com/en/internet-of-things/platform.

[6] ericsson.com/mobility-visualizer.