Deep Science Publishing
https://doi.org/10.70593/978-81-988918-0-8_12

DeepScience
Open Access Books

Chapter 12: Preparing for what's next: Future-proofing networks for artificial intelligence, IoT, and smart infrastructure

12.1. Introduction

The confluence of artificial intelligence, Internet of Things, and technology-enhanced physical environments is leading to an era of pace-setting transformation within communities, organizations, and entire economies. Responsive, resource-aware, and risk-adjusted networks will be the cornerstone of exponentiating returns blended with remediating governance gaps and collaboration challenges. While dynamically adapting to changing requirements, responsive networks will apply AI machine learning at the edge to optimize themselves. By establishing trustless network slices for billion-device deployments of supervised, semi-supervised, and unsupervised learning, resource usage and risk management for IoT move into the kilobit range. Secure dynamic automation of business processes across smart infrastructure will challenge conventional app-driven non-IoT experiences delivered by smartphones (Arora & Bhardwaj, 2021; Khan et al., 2021; Atakishiyev et al., 2024). As AI, IoT devices, and digital twins of the smart environment enable user-centric digital experiences, they will drive a rupture of the existing limits on physical space and sense-making for intelligence augmentation. Leading network operators will enhance their value mediation role and embrace a change from internalization of user value to externalization as confidence intermediary, while cooperating with other trusted, trusted-less, and non-trusted intermediaries in the converging markets for IoT and public safety. A key methodological contribution lies in the linkage of a macroeconomic monetary growth model with microeconomic service flow allocation rules under resource flow bottlenecks and finite planning horizons. Together with the applied research questions it enables intelligent network factorizations. Thereby, existing high-level derivations of technological readiness for an all-optical future are brought to verification by establishing specific physical layer conditions for arrival at key technology readiness milestones. Potential use cases,

including unified networking for auxiliary sensing and time transfer in emerging wireless connectivity markets, are detailed (Rahman & Thill, 2023; Sadaf et al., 2023).

12.2. Understanding AI, IoT, and Smart Infrastructure

Today's enterprise networks are fundamentally different from their predecessors. The focus has shifted from IT to OT. Enterprise technology implementations over the last few decades have been driven primarily by business needs. New technologies have enabled companies to increase productivity, reduce costs, become more agile, and ensure delivery of higher-quality products and services. Demand for increased productivity drove the Industrial Revolution with its use of mechanical devices to perform tasks previously done by hand. The introduction of new digital technologies in the twenty-first century has propelled this quest for higher productivity to new heights.

The current network evolution is characterized primarily by three technology trends: AI, IoT, and Smart Infrastructure. Business visionaries and technology evangelists are touting the transformative implications of these changes. Most companies have, to one degree or another, already adopted these technologies. The business impact of these operational technology transformations is profound. A survey of over 300 executives at companies that had engaged in a major business model transformation found that 61% of the executives were moving toward using AI compared with 45% who were focusing on data analytics, or 32% using cloud computing. Nearly 44% of the executives said that AI-enabled businesses would have more potential to disrupt their companies compared with only 29% who were focused on cloud computing and 26% who were concentrating on data analytics. This paper examines these technology trends.

12.2.1. Defining AI and Its Implications

A more general term that encompasses narrow AI and that enjoys wide recognition today is machine learning. Here the term is used to cover all of the dense field of AI techniques, from machine learning via artificial neural networks or deep learning, intelligent expert systems, natural language processing, machine vision, speech recognition, planning and scheduling, and the like, to the type of abstraction and general intelligence that humans are capable of exhibiting but that AI systems, however advanced, cannot. As is clear from the long list of AI technologies, there are many forms of AI, formulating solutions and methods to exhibit AI techniques to bring AI to bear to solve certain problems, exhibiting "narrow AI" that specialize in certain areas, some at human level competence, some even at human superhuman competence.

AI is increasingly being successfully employed for tasks previously the reserved province of cognitive human intelligence working in collaboration with the physical actions of the human body. As the decades have progressed, such solutions have become ever cheaper to implement based on advancements in hardware in the form of general purpose microprocessors or graphics processing units, large clouds supplied by vendors or large local services, large databases, particularly if transcultural, and extensive partially or fully supervised training of the AI by humans. AI technology may exhibit high levels of human capability in areas such as machine vision, speech recognition, human interaction, and critical decision-making. However, they are individual tasks, ingeniously automated at high levels of capability, yet employment of the solutions to issues in those verticals does not provide the generalized human capability to transfer knowledge, competencies, and experience across different verticals solving different problems, nor to identify and ask any question about the known universe.

12.2.2. The Role of IoT in Modern Networks

The IoT can be summarized as a collection of sensors that report to the outside world. The sensors could be put into a vehicle, a power station, or on a person. They can use GPS for positioning, and other sensors to determine traffic conditions, identify whether the vehicle has crashed or the person has fallen, and do more research on the structure, along with the comforts, safety, and modifications of living conditions. This data is transmitted to a cloud system for data analysis. Based on the data, corporations and governments can determine when maintenance is required or whether services such as power or water need to be regulated. By describing what we call the Smart Infrastructure (SI) system, we address the issues of helping the citizens and municipalities protect resources and the local economy and design areas and buildings that are designed for and operated in a manner that meet, preserve, and enhance the natural resources and quality of life.

The data that these IoT Sensors and actuators create can be a big burden on any network but not if they are utilized wisely. The burden of data and transmissions can be up to a billion transactions a day! We've witnessed the move from a model that utilized data once a day while tracking IO or ETL processes to possibly a billion data messages a day as the Digital Twins pull real-time updates from the network. In fact, a company has stated that the future of AI requires that there be teradata and expertise that allow the AI programs to be taught and validated. These burdens change the requirements necessary for the network to transfer the data and place requirements that would have been considered impractical just a few years ago. These sensing and control requirements will become social at their core to the extent that the whole infrastructure for IoT, AI, Data, and Control becomes a cloud system.

12.2.3. Smart Infrastructure: Concepts and Applications

Smart infrastructure denotes a development and management paradigm that leverages advanced digital connectivity and intelligent systems in the creation and operation of engineered physical systems. Smart infrastructure is generally governed by both systems engineering principles and smart systems technologies. By their very nature, engineered physical systems are complex systems composed of interdependent elements that work in synergy to produce and provide some kind of output or service. Smart infrastructure elements are enhanced by their connectivity and control systems intelligence. As an example, the energy value chain offers a traditional manufacturing industry that converts raw material inputs into complex final products, fuels.

Smart infrastructure must be at or near ubiquity within a defined geographic area, a country region for example. Smart infrastructure operates by facilitating the movement from one location to another of people, goods, and services. Smart infrastructure connects people virtually and physically to inform, deliver, and share. Smart infrastructure is concrete and tangible but also intangible and virtual. Smart infrastructure relies on smart systems intelligence in managing the flow of people, goods, and information in time and space. Smart infrastructure is represented by all of the processes associated with the movement of people, goods, and information in time and space. Smart infrastructure includes massive capacity. Information flows through the facility comprise the inner workings of the smart infrastructure but are typically not controlled by the infrastructure.

Not only does the digital economy rely on smart infrastructure, smart infrastructure itself is transforming as industry adapts in response to the digital economy, creating new opportunities for investment. It is the disruption to the economy from digital technologies, such as additive manufacturing, autonomous systems, and the Internet of Things, that is changing how it is defined, built, financed, and operated in ways that might not have been envisioned. Smart infrastructure is fragile, easily toppled or destabilized. In short, resilience is the Achilles heel of smart infrastructure--for the economy it is too little, it is over-utilized.

12.3. Current Network Architectures

The dominant approach to networking, despite recent interest in alternative and virtualized models, is derived from the Open Systems Interconnection model. The model is a well-known networking architecture that provides a simplified view of logical network organization and creates a layered abstraction of networking functionality. The main purpose of the model was to promote interoperability between telecommunications systems. To accomplish this, the model established a layered framework that set out the

major functions required at each layer and defined the interfaces between these layers. According to the model, there are seven layers in a complete telecommunications architecture. Starting at the top, the layers are as follows: Application, Presentation, Session, Transport, Network, Data Link, Physical. The Transport Layer provides for transparent transfer of data between end systems. The Network Layer can be thought of as an operating system for the network. The main functions requested of the Network Layer are to determine on what basis the connections are to be made, to provide the various services required by the upper layers, and to control the network resources associated with those upper layers.

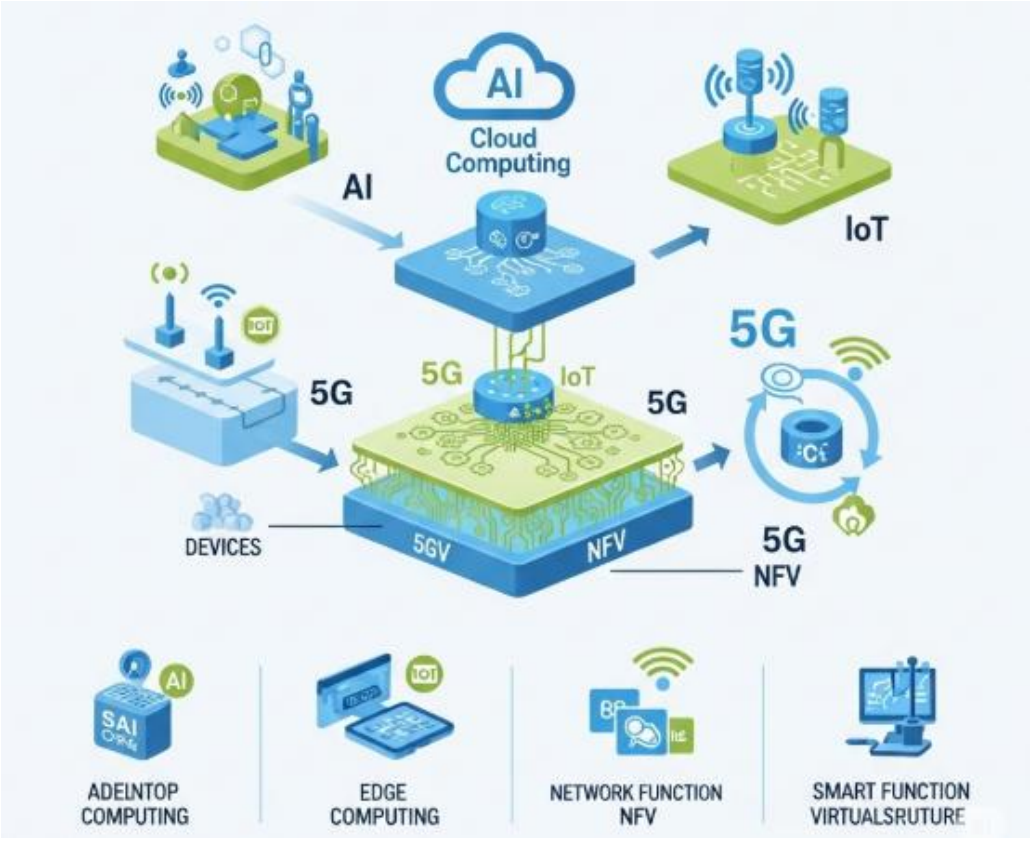

Fig 12.2: Current Network Architectures

These functions involve establishing, maintaining, and terminating connections; forwarding data packets through the network; automatically detecting network failures, selecting alternative paths, and dynamically reconfiguring the network to continue data flows; and multiplexing various upper layer services on a single physical channel. Existing networks are further organized based on dedicated physical and logical functional structures. For example, the core of the Internet comprises a set of routers that contain hierarchical destination-based forwarding tables and that exchange with one

another reachability and cost information. The goal of this self-configuring network is to push traffic along the lowest cost path via the routers controlling the largest flows and that have the most storage capacity.

12.3.1. Overview of Traditional Network Models

A network is a system of interconnected IT (information technology) nodes and links via any means of communication. Interconnectivity is facilitated via established, standardized protocols to ensure security, privacy, and integrity. As a design blueprint, these protocols govern how packets of information are formatted, structured, transmitted, routed, and received within or between networks, services, and applications, over shared or dedicated communication links. All physical and abstract relationships among all constituents of a network can be modeled, enabling a holistic approach to monitor, manage, control, direct, and supervise all physical and logical flows at many levels of abstraction.

Computer network architectures can be categorized into traditional horizontal, layered models, or newer, non-hierarchical modern design, with few or no underlying abstractions. The most common layered model is the OSI reference model, which consists of seven distinct layers, each defining a well-specified functional interface to adjacent levels in the model, abstracted to limit external hardware and software dependencies. Layers 1 through 4 are most closely associated with computer networks. Layers 5 through 7 encompass Application Logic Interfaces at the user level and other lower layers, and primarily involve configuration, control, and monitoring functions and services.

The TCP/IP Internet suite is organized horizontally into four representation layers, along with defined application interface functions for external service access, while the Ethernet architecture defines lower-layer hardware and operation functions for packet routing and transmission services. All models share common abstractions. Network routers and switches execute packet transmission and routing functions for packets classified into several priority classes, providing users with varying quality-of-service guarantees, while various classes of servers provide control, configuration, monitoring, and library functions for external entities, while also managing overall network operation, optimizing various performance criteria via operational feedback from equipment and user connections.

12.3.2. Limitations of Existing Architectures

In an overview of traditional network paradigms for IoT, one of the first mentions could be TCP/IP that was designed for wired point-to-point communication systems, with an inherent absence of mobility, and became a de facto reference for any services and applications dealing with long-distance communication. Originally, a reference framework proposed a single communication paradigm for all applications over the Internet. The Socket Interface was basically defined as an abstraction for portable network communication. However, some applications, due to their own intrinsic properties, can only adopt a specific protocol, without incurring the overhead imposed by a common paradigm.

At the same time, some inconsistencies and gaps in the TCP/IP specification have implemented the need for more specific transport protocols for different applicability areas. A first problem with IP is that it is hierarchical for network address and flat for host address, making routing scalable by IP prefixes but consuming too many bits in the address encoding. Internet addressing does not specify connectivity and uniqueness, which might be predominating features in some new categories of networking services like sensor networks and networks of networks. This latter deficiency is both a fault of the routing algorithm and of the Multicast Extension of IP, that also only allows host-to-host connectivity. Moreover, Mobility is only partially handled by IP and personalization and renumbering of a communication can only be implemented by extensions of the current Internet architecture. There are no facilities for Quality of Service for multimedia services and a flat translation between protocols also translates the well-established notion of Quality Levels.

12.4. Emerging Technologies and Trends

The recent emergence of 5G wireless communication networks and associated technologies points to some key trends for future-proofing networks. First, communications are expected to support the three main pillars of digitized societies: AI services as a service, IoT implementations-as-a-service, and smart solutions-as-a-service. In this context, future-proofing communications-related enterprise services may precondition the deployment of telecommunications networks that create new types of capacities and lower costs to increase the levels of service excellence for support of differentiated quality as perceived by users.

5G and Beyond

5G is the technologic platform due to the support of diverse service requirements with very different levels of performance. 5G is also expected to support a scale of service demands in terms of the number of networked devices that will interconnect our

infrastructures and services. 5G has the concept of digital twins, but they are strictly related to the necessity of providing intelligence to systems, processes, entities, and services involved in the operation of connected infrastructures, such as for industrial IoT. We must go one step further. Rather than in just one direction, which is the provision of intelligence mainly provided by Artificial Intelligence and Machine Learning and transmitted by digital networks, layers of intelligence must be distributed in every step of the processing, from local one-hop communications among near or adjacent devices, to longer-distance transmissions involving the service-aware coordination of many hierarchically interconnected or interrelated fragments of these digital twins.

Edge Computing

Edge computing is perhaps the most visible technology that will allow for this distribution of intelligence along the data processing flow although it is not alone, Fog Computing, hierarchical Systems of Systems, and Cloud Computing also play similar roles. Neural Processing Units are processors that exist as accessible silicon circuitry on edge devices and also on distributed clouds that are closer to the immersed devices of the IoT application. NPUs open new ways of embedding AI services within IoT devices. Innovation-rich companies are already hybridizing decisions and/or control loops in digitized factories based on AI capabilities in terms of Neural Networks distributed across the Industry 4.0 reference architecture.

12.4.1. 5G and Beyond

5G has been defined with the goal of enabling new use cases and market segments for mobile communications, which is largely focused on consumer services. These use cases are based on major advances in three diverse technology pillars. The first use case has to do with enhanced mobile broadband, which is focused on the augmented and virtual reality-driven demands for extreme capacity and low latency in wireless access. The second use case, referred to as ultra-reliable low latency communications, is driven primarily by demands for low latency and high reliability in support of industrial applications. The third use case, called massive machine-type communications, is driven by the Internet of Things market and calls for efficient support of a massive number of low-cost devices with reduced energy consumption for battery life extension.

In addition, 5G mobile wireless communications is considered as a new infrastructure for enabling technical and social innovations that leverage the ongoing digitization of business and non-business applications across most, if not all, sectors of the economy. Therefore, advanced 5G mobile communication capabilities and associated digital economic innovations will also directly benefit the transport, utilities, logistics, construction, and taxation sectors. 5G has adopted a new service-centric architecture that

supports end-to-end slices tailored for specific consumer and enterprise service requirements. However, given the wide range of features and requirements across the targeted verticals and sectors of the economy, 5G is the first mobile system to incorporate network slicing, which allows multiple logical networks to run on top of a common physical network infrastructure.

12.4.2. Edge Computing

With the enormous growth in the amount of data generated at the far edge of the Internet due to the large-scale deployment of Artificial Intelligence, Internet of Things, and smart infrastructure, and for which considerable reliance is placed on latency-sensitive, computation-intensive, and bandwidth-hungry state-of-the-art data analytics and machine learning algorithms, storing and processing this data centrally at cloud data centers, or even at intermediate edge data centers closer to the cloud may not be efficient or cost-effective. This has led to the strategic and architectural shift in networking as well as computing, to place storage and computing resources where the data is generated or consumed, leading to the large-scale deployment of traditional edge servers at the physical locations where the data is generated, or consumed, and edge computing and edge analytics frameworks that use a combination of edge servers, cloudlets, fog, and multi-access edge compute servers.

Edge computing pushes storage and processing resources close to endpoints. Push edge processing close to where data is generated, and act on data locally while data still resides in edge devices. It helps gain efficiency and reduce costs for the applications that are latency-sensitive or bandwidth-constrained. The architecture provides significant improvement for AI, IoT, and smart infrastructure, including the key performance metrics of latency, bandwidth, reliability, efficiency, cost, privacy, and security for most applications in these areas, and mitigates the risks and drawbacks associated with dependence on the cloud or centralized data centers. The optimal use of edge computing, especially in mission-critical AI, robotics, and domain applications have considered the need for fault-tolerance, dependability, and reliability associated with the distribution and decentralization of processing and decision-making.

12.4.3. Network Slicing

For the past few decades, we have been talking about QoS networks that can perform some level of prioritization of the resources being provided for the different types of flows present in the network. In that sense, we have made some progress through complex queuing and scheduling mechanisms as well as Layer 2 technologies. However, these have only gone a short way toward fulfilling the high bandwidth and throughput

demands being provided to the different services being transported. Apart from visible services such as video, each user or machine might be utilizing multiple flows for invisible service types such as data collection or admin and sync services that also demand a constant level of throughput.

Network slicing is a virtual networking architecture where a single physical network can be divided into several virtual networks, each offering its own network protocol optimized for niche operations or services. This allows for a low-cost solution to build a physical network that offers flexible, dedicated networks for customers. A given physical infrastructure is divided into a certain number of slices, where each slice designed for a particular service type optimally utilizes the physical network resources according to the guarantees for that particular service level. Each slice is isolated from other slices and can customize its services. Network slicing reduces network service provisioning time to minutes with no service interaction.

Beyond flexibility and lower capital costs, the advantages of network slicing also include lower operational costs per slice and improved agility, as new network services can be provisioned and decommissioned on demand with minimal operational activity. Network slicing provides immediate value incentives for business-to-business network services too. As networks become increasingly attuned to the services they carry, any particular service will require fewer resources since the network optimization will be transparent to the end-user.

12.5. Designing Future-Ready Networks

Future-ready networks provide the capability to not only support previously unimagined demands from customers today but also to customize and evolve to meet unimagined needs for services and offers from users tomorrow. This quality of services is made possible through an excellent combination of connectivity and compute functions. While virtualization technologies for both connectivity and compute domains have progressed over the recent years, for a variety of reasons, they have not reached sufficient maturity to fully enable the promise of custom-designed services with agility. This inertia in the evolution of flexibility and adaptability in functions has delayed the processes of de facto enabling a service ecosystem that creates favorable economics for offering niche services based on specific user and application requirements.

At the same time, the rapidly widening gap between the ever-growing magnitude of demand for connectivity and the ability to scale the underlying infrastructure has turned into a strategic imperative for the CIOs around the world. Addressing this gap requires radical changes to the underlying infrastructure technologies. In particular, the upgrade cycles must become more frequent, the cost of operations must come down, the

processes of planning upgrades or reallocating resources must become more and more agile, and the traffic patterns experienced today cannot be the only design criteria for building and designing future-ready networks.

Aiming to address the above challenges and imperatives, the increasing level of collaboration and investment that is dedicated to exploring this consonance is exemplified by initiatives such as various open-source networking initiatives. With much of the current foundation surveys focusing on the transport connectivity aspect, in this section, we briefly cover the implications of these efforts from several topical perspectives: scalability, flexibility, adaptability, and security, compared to both the existing operational networks and the requirements for satisfying the service ecosystems in the future timeframes.

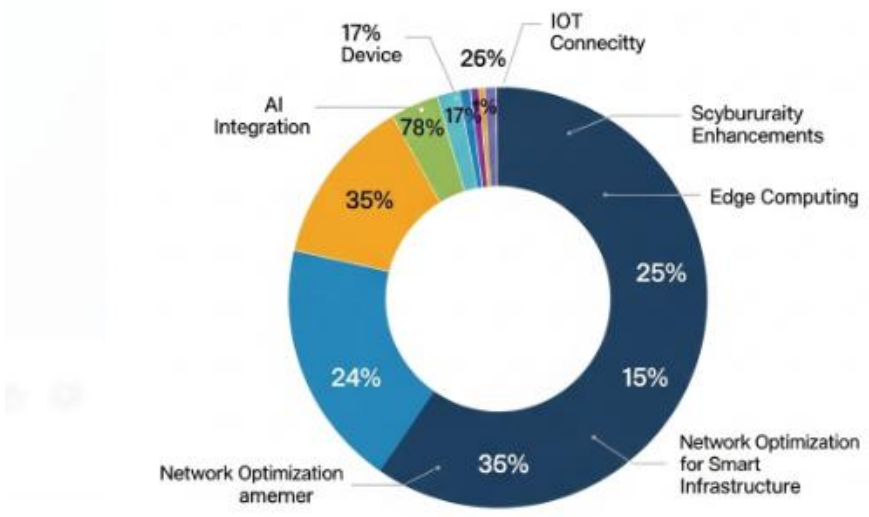

Fig : Future-Proofing Networks for AI, IoT, and Smart Infrastructure

12.5.1. Scalability Considerations

A key challenge for current and future networks is to provide sufficient capacity and performance to support the ever-increasing need to transfer data from one host to another. Although the rapid advances in technology, such as chip speeds and storage capacities, have enabled a continued increase in computing and communication capabilities, the augmentation of network carrying capacity does not come for free. With the growing amounts of data to be transported, influenced and fueled by the growth of the number of devices connected to the Internet, the ability of traditional service providers to offer services with predictable performance at a reasonable price is being severely challenged. Future computer networks must provide considerable capacity to

cope with the demand and transport additional information without creating undue congestion and delays. Several design principles influence the scalability of services. First, services ought to be designed so that as user demand increases, only a portion of the service has to be augmented to satisfy the demand. Second, an increasing number of users should not result in disproportionate increases in the resources required to provide a service. Third, wherever possible, the incremental cost of providing the service should fall steadily as the number of users grows larger. Suitable design principles are, for example, to provide a hierarchy in the network clearly separating edge and core computers, allowing for appropriate local caching and for quality-of-service control functions. Moreover, while current networks are designed to support a few high-bandwidth network applications, future networks must provide an increasing number of user classes and support a variety of bandwidth levels for large amounts of different data sources in a dynamically changing environment. In addition, the ability to sustain predefined quality levels, such as latency, jitter, and reliability on critical data paths, is extremely important.

12.5.2. Flexibility and Adaptability

Enterprises, agencies, and institutions should review their decisions to lock specific technology infrastructures and services, whether by contractual provisions or technology adoption processes. Partnerships focused on operational efficiency and co-development have often begun unique network services defined on proprietary technology and specifications. These multilayered, multidimensional networks may serve a system of specific institutional needs well, but they risk higher degrees of alienation from broader industry or public needs, leading to lock-in scenarios, reliance on narrow vendor ecosystems, and erratic technology upgrade paths. This section considers some options policy- and network operators can pursue to create network service environments that are flexible and adaptable in the present and in the future.

Flexibility and adaptability in network technologies can take many forms. For example, smart infrastructure services like smart cities may contemplate physical infrastructures like lighting or energy metering managed remotely by large institutional owners but allow third parties to develop crowdsourced secondary application services. These applications might enable route optimization for potential autonomous vehicles seeking charging stations or warnings of polluted air spaces, which can lead to travel disruptions for humans and AI services alike. Similarly, adaptive controls can manage facility access and use across public, commercial, and industrial domains. Typical services deployed on edge/networking infrastructure should allow periodic local optimization and adjustment to efficiently handle unexpected congestion, varying patterns in device traffic, and other forms of variation, especially with the range of potential device types

and activities. Control at the access layer can coordinate edge response and anticipate needs on shorter timescales for auxiliary services relying on warehousing or time guaranteed by the device lifecycle.

12.5.3. Security Challenges and Solutions

Security is historically the most serious challenge for any information sharing and communication system because it is always on the weakest link that information gets hacked. The integrated 5G, AI, IoT, and smart infrastructures envisioned will produce the largest data sharing and information connectivity public infrastructures in human history. The objective of data sharing will be to provide other organizations whether governmental, commercial or for individual profit with knowledge actionable through the lens of their own private data and specific analytical objectives. Every step we implement this data sharing objective obviously requires our attention to be on: security of data both in transfer and calculations; privacy protection at the received data level and for the individuals involved in original data; personal data anonymization so that only generalized, non-revertible information on individuals is shared; ethical use of data e.g. for implementing prediction and machine learning based AI solutions. These considerations remain essentially unchanged with respect to those of existing systems.

For all security issues, increased size of data at transfer, and operated on, increases the size of the susceptible attack space. Yet, it is basically the achievable data mining accessible knowledge via enormous collaborative databases with interconnected personal information that opens the question of who is willing to collaborate, what incentive mechanisms must be designed to ensure that this collaboration takes place. Security threats after that will primarily evolve from who is interested in any act of sabotage such as fraud altering predicted knowledge, creating the maximum chaos possible through announcement of threatening information while only seeking to promote fear rather than support the certainty of protection or publicize distrust of rating system operators based on data query access. Basically these individuals will wish to become notoriety. In this case, severe personal context-based incentive conditions will be needed to counter those individuals.

12.6. Integration of AI in Network Management

Networks are complex entities that operate in a complex and dynamic environment, often with many external influences that change continuously over time. With the advent of new generation networks, ensuring the optimal and secure functioning of these technological ecosystems is becoming an even greater challenge and involves new and multiple dimensions. Intelligent solutions can be of great help in this dynamic and

multifaceted environment. Widespread use of AIOps in network management by Communication Service Providers is starting to make these powerful solutions a commodity, making Artificial Intelligence (AI)-driven solutions available for Operators of Critical Infrastructure, the Internet of Things, industry, smart cities, and other systems that are characterized by critical infrastructure, the need for real time response, and/or the use of complex and large networks. AIOps and Intelligent Network Management solutions that integrate AIOps are serving as the necessary complement for these use cases, given the critical role that communications play in the operation of these infrastructures and business sectors.

Data analytics has provided important tools that let us describe, diagnose and understand networks. AI-driven tools and solutions expand existing capabilities and bring IT relevance into the decision-making process. Three common examples are AI-driven optimization models, predictive maintenance models and Quality of Experience assurance models. Optimization models are used to better operate the network across multiple variables. Optimization models can be used in multi vendor environments, improving their decision-making processes. Predictive maintenance models allow operators to avoid problems by anticipating them and replacing functions, services, or equipment before actual failures. QoE models allow better understanding of how the end customers are experiencing services as well as providing specific stimuli to optimize their experience.

12.6.1. AI-Driven Network Optimization

In traditional network management, data is simply processed and passed on to the networking experts where decision making takes place. In a scenario, where the number of network events is huge and simply thresholding is not enough, making sense of the network data requires expert knowledge at scale. AI and Machine Learning techniques can help automate this expert knowledge. Such systems can learn from the experts and suggest actions, and in some situations also decide on actions based on what they have learnt. Such systems will utilize historical data of the network – which may include data from different sources in the network – events from all network elements, related telemetry data, and use the learnt knowledge to reduce time and effort to manage the networks. Such solutions can help in root cause analysis and diagnosis of problems taking multiple inputs, in determining impact of identified issues, in predicting impact and troubleshooting suggestions of certain alerts, in suggesting resolution for certain issues, and more. As ML and AI technologies evolve, the scope of automation with such systems also helps in real-time decision making and can be extended to taking actions directly to resolve issues without human intervention.

AI enabled automation will result in increased uptime, improved service quality, and less effort from the network experts. However, within the scope of operations, AI has a clear advantage in narrow domain expert systems, and not general domain experts. AI accelerates what human experts do in the operational scope on a day to day basis. Such edges of AI are clearly seen in areas such as machine vision, speech to text translation, and various other domains. Given the multi-dimensional data of the networks, AI techniques have gained wide acceptance and maturity in enabling intelligent decision support systems for network performance management, and it is practically possible to deploy several such systems today.

12.6.2. Predictive Maintenance

One specific AI capability that is particularly relevant to network management is predictive maintenance. Closely integrated into Network Management Systems, predictive analytics forecast and predict hardware and software component failures, service disruptions, and performance degradation. Based on these actionable predictions, appropriate measures can be triggered to avoid performance degradation or outages. With the advent of the 5G technology, communication infrastructure is expected to be more sensitive, flexible, and dynamically programmable. Therefore, the probability of several failure points within the equipment gets increased. Artificial Intelligence adopts predictive maintenance strategies to recognize and minimize the equipment failure frequency by predicting the time lapse between services.

Predictive Maintenance is a smart concept that compares information from sensors and other systems, identifies patterns, and informs when it's the right time to check an asset to avoid an outage. This technology helps industry, factory, and other asset managers understand their networks better. It minimizes the use of technicians for PM-related tasks, improves network reliability, and increases the RoI. Upcoming paradigms of Predictive Network Management include Self-Managed Networks and Cognitive Radio Networks. The latest research direction studies advanced techniques from AI and Machine Learning. Machine Learning and big data analysis become key enablers of the recently emerging concepts of self-managed, self-optimizing, and cognitive networks and play a very important role in predicting the future behavior of network elements for prompt decision-making and action.

12.7. Conclusion

Modern life relies on the IoT environment where IoT devices take real-time decisions. These devices are connected to the servers where data is processed and receive real-time control signals. Internet connectivity is a must to enable remote computation and control.

Micro-data centers are often employed to handle the heavy computational load and act as a mediator between the cloud and the IoT. In this environment, local AIs can be deployed to do the computation close to the sensors or to do local decision making. If they can make the correct decisions the number of communications interacting with the cloud can dramatically be lower down. Cloud-based AIs can take higher-level high specific decisions and can coordinate the local AIs. Bandwidth is a critical resource in such environments. The need to optimize the energy consumption and minimize the bandwidth of such massively many-to-many communications arises. Sophisticated modulation schemes and AI-based controls can be employed to ensure that the future AI-based services can be delivered with specific quality-of-services that suit the customer requirements. A novel spectrum sharing model in which the entire bandwidth is divided into different channels where power and bandwidth for each channel can be dynamically allocated depending upon the mix of traffic. Traffic demands can be predicted and future futures of a specific network can be planned in advance. Such optimization can result in networks capable of serving a large number of communications within their coverage area with very small investment on infrastructure upgrade or deployment. These various design and implementation guidelines for future-proofing intelligent networks will enable an inflection point in the trajectory of the faster internet which the IoT environment relies on in the next few years along with the deployment of new technologies. The system will reduce latency, conserve power, reduce carbon footprint, ensure the delivery of services with stringent quality-of-performance Guarantee, and drive research around enabling intelligent systems for work forces where people are afraid of machines becoming unemployable.

12.7.1. Future Trends

Smart Infrastructure (SI) refines and extends the Internet of Things (IoT) paradigm by performing the role of an active supervisor or manager of social infrastructure components via Intelligence algorithms and Edge Computing modules, possibly assisted by external services. SI is more than a sum of smart services; it is the service environment for delivering a large variety of promising smart services, with differentiated characteristics and quality levels. Enabling these smart services relies on solid data foundation: an efficient sensor network infrastructure extensively deployed within the vicinity of most infrastructure components requiring smart service provisioning, capable of anticipating smart service requirements. Initializing SI creates an efficient infrastructure cohabitation ecosystem based on an efficient division of labor. Smart Services, embedded into infrastructure capable of observing, learning, predicting and sometimes foreseeing the external conditions that trigger service delivery, are simple and intuitive, and do not require user education or specialist know-how competence for activation or interpersonal decision-making. This implicit service initiation creates a

cohabitation ecosystem based upon humanitarian considerations, with a limited number of security premises that potentially conflict with individualistic expectations.

We explore future trends in the interface between AI, the Internet of Things (IoT), and the Smart Infrastructure (SI) space. Our view is one based upon increased convergence of successful business cases driving demand in the AI user domain and in the vertical SI/IOT/AI "solution" components and enablers, which communicate traffic-services driven by offered services business models. Convergence implies crossover. Indeed, many of the primary infrastructure components were built as vertical business silos. However, each of these business segments has strong and converging expertise in advanced sensing, enabling networking, communications, AI on the boundary/edge, realization at-scale of common software-enabled features of shared utility, and overall successful business case delivery of user-facing services. The environmental and societal needs for rebuilding and upgrading these key active infrastructure pathways to sustainability is increasing. And, hence, their close cooperative inter-relationships also.

References

Arora, P., & Bhardwaj, S. (2021). Methods for Threat and Risk Assessment and Mitigation to Improve Security in the Automotive Sector. Methods, 8(2).

Atakishiyev, S., Salameh, M., Yao, H., & Goebel, R. (2024). Explainable artificial intelligence for autonomous driving: A comprehensive overview and field guide for future research directions. IEEE Access.

Rahman, M. M., & Thill, J. C. (2023). Impacts of connected and autonomous vehicles on urban transportation and environment: A comprehensive review. Sustainable Cities and Society, 96, 104649.

Rahman, M. M., & Thill, J. C. (2023). Impacts of connected and autonomous vehicles on urban transportation and environment: A comprehensive review. Sustainable Cities and Society, 96, 104649.

Sadaf, M., Iqbal, Z., Javed, A. R., Saba, I., Krichen, M., Majeed, S., & Raza, A. (2023). Connected and automated vehicles: Infrastructure, applications, security, critical challenges, and future aspects. Technologies, 11(5), 117.

Khan, S., Sharma, I., Aslam, M., Khan, M. Z., & Khan, S. (2021). Security challenges of location privacy in VANETs and state-of-the-art solutions: A survey. Future Internet, 13(4), 96.