The copyrights of this paper are with the IEEE.

Before use, please, consult the IEEE Copyright Policies.

The typeset version of this paper is available from IEEE Xplore.

Enjoy reading!
5G-Enabled Tactile Internet

Meryem Simsek, Adnan Aijaz*, Member, IEEE, Mischa Dohler, Fellow, IEEE, Joachim Sachs, and Gerhard Fettweis Fellow, IEEE

Abstract—The long-term ambition of the Tactile Internet is to enable a democratization of skill, and how it is being delivered globally. An integral part of this is to be able to transmit touch in perceived real-time, which is enabled by suitable robotics and haptics equipment at the edges, along with an unprecedented communications network. The fifth generation (5G) mobile communications systems will underpin this emerging Internet at the wireless edge. This paper presents the most important technology concepts which lay at the intersection of the larger Tactile Internet and the emerging 5G systems. The paper outlines the key technical requirements and architectural approaches for the Tactile Internet, pertaining to wireless access protocols, radio resource management aspects, next generation core networking capabilities, edge-cloud and edge-AI capabilities. The paper also highlights the economic impact of the Tactile Internet as well as a major shift in business models for the traditional telecommunications ecosystem.

Index Terms—Tactile Internet; haptic communications; real-time communication; edge intelligence; ultra-low latency; ultra-high reliability, 5G, massive connectivity, OFDM

I. INTRODUCTION

Mobile communications continues to play an important role in modern economy, including consumer, health, education, logistics, and other major industries. Mobile communications networks of today have successfully connected a vast majority of global population. After creating the Mobile Internet, connecting billions of smart phones and laptop, the focus of mobile communications is moving towards providing ubiquitous connectivity for machines and devices, thereby creating the Internet-of-Things (IoT) [1].

With the technological advancements of today, stage is being set for the emergence of the Tactile Internet in which ultra-reliable and ultra-responsive network connectivity will enable it to deliver real-time control and physical tactile experiences remotely. The Tactile Internet will provide a true paradigm shift from content-delivery to skill-set delivery networks, and thereby revolutionizing almost every segment of the society.

As per ITU [2], the Tactile Internet will add a new dimension to human-machine interaction by delivering a low latency enough to build real-time interactive systems. Further, the Tactile Internet has been described as a communication infrastructure combining low latency, very short transit time, high availability and high reliability with a high level of security [3], [4]. Associated with cloud computing proximity through e.g. mobile edge-clouds and combined with the virtual or augmented reality for sensory and haptic controls, the Tactile Internet addresses areas with reaction times in the order of a millisecond. Example areas are real-time gaming, industrial automation, transportation systems, health and education.

Because the Tactile Internet will be servicing really critical aspects of society, it will need to be ultra-reliable, with a second of outage per year, support very low latencies, and have sufficient capacity to allow large numbers of devices to communicate with each other simultaneously and autonomously. It will be able to interconnect with the traditional wired Internet, the mobile Internet and the IoT – thereby forming an Internet of entirely new dimensions and capabilities. State-of-the-art fourth generation (4G) mobile communications systems do not largely fulfill the technical requirements for the Tactile Internet. Therefore, fifth generation (5G) mobile communications systems are expected to underpin the Tactile Internet at the wireless edge.

5G wireless access is the wireless access solution to fulfill the wireless communication requirements for 2020 and beyond [5]. At ITU, ITU-R working party 5G has the responsibility for the terrestrial radio system aspects of international mobile telecommunications (IMT) systems, which today comprise IMT-2000 (i.e. 3G) and IMT-Advanced (i.e. 4G). 5G is treated under the term “IMT-2020” of which the scope is currently being developed in a new ITU-R recommendation typically referred to as IMT Vision. An early assessment of 5G scenarios and requirements has been developed in the METIS research project [6], [7], and recently also by the telecommunications industry alliance NGMN [8]. Overall there is a common understanding that 5G should not only support an evolution of traditional mobile communication services, such as personal mobile multimedia communication or personal mobile broadband services; 5G should in addition address novel use cases including for example machine type communication (in several fields like e.g. smart energy networks or smart grids, vehicular communication and intelligent transport systems, or sensor networking) or novel ways of media distribution. This range of 5G use cases pushes the requirements on 5G in several dimensions, like latency, data rate, device and network energy efficiency, mobility, reliability, traffic volume density, connection density, etc. The broad range of targeted 5G capabilities will make it an important enabler for the Tactile Internet.

To facilitate the expected massive increase of traffic to be handled in a 5G system, additional spectrum has to be allocated to the 5G wireless access. The spectrum range up to a few GHz is of special importance in order to provide wide-area coverage. However, in order to enable very high capacity
of rapidly moving devices' control circuits is, for example, significantly below 1 ms per sensor [12]–[14]. Hence, the automation in industry is a key application field in the Tactile Internet. Today, control processes are realized by fast wired connection, e.g., the industrial Ethernet. In the future, these wired systems are aimed to be fully or partially replaced by wireless systems in order to enable high flexibility in production, i.e., the industrial revolution [15]. This requires a guaranteed reliability and minimum end-to-end latency and can be enabled by Tactile Internet solutions.

B. Autonomous Driving

Fully automated driving and platooning of vehicles is discussed as a new step in mobility within the context of 5G. A considerable and sustainable reduction of road accidents and traffic jams can be realized by autonomous driving, i.e., vehicle-to-vehicle or vehicle-to-infrastructure communication and coordination. The time needed for collision avoidance in today's applications for vehicle safety is below 10 ms. If a bi-directional data exchange for automatic driving manoeuvres is considered, a latency in the order of a millisecond will likely be needed. This can technically be realized by the Tactile Internet and its 1 ms end-to-end latency.

Fully autonomous driving is expected to change the traffic behavior entirely. Especially small distances between automated vehicles, in particular in platoons, potentially safety-critical situations need to be detected earlier than with human drivers. This requires ultra-high reliable and proactive/predictive behavior in future wireless communication systems.

C. Robotics

In recent years, the technical potential of robotics has increased in various fields. Its demonstrated potential comes with increased complexity, and various challenges, so that autonomous robotics will find their application only in a limited range of rather specific areas, e.g., autonomous driving, in the near future. Remotely controlled robots with real-time, synchronous and visual-haptic feedback, however, seem to be a promising alternative to autonomous robots. Controlling robots must happen at latency reaction times that are fast enough for the robot and its object. If a real-time controlling and communication is not guaranteed, robots will move in fitful manner which may lead to an oscillatory behavior. For many robotics scenarios in manufacturing this has led to a maximum latency target of a communication link of 100 µs, and round-trip reaction times of 1 ms, the target as discussed for the Tactile Internet.

D. Healthcare

Tele-diagnosis, tele-surgery and tele-rehabilitation are just some of the many potential applications of the Tactile Internet in healthcare. Using advanced tele-diagnostic tools, medical expertise could be available anywhere and anytime regardless of the physician's location [16]. Hereby, a tele-robot at the patient's location will be controlled by the physician, so that not only audio and/or visual information but also haptic
feedback is provided. The same technical principle is applied to tele-surgery applications. In tele-rehabilitation techniques can be used for patients to remotely steer and control his/her motions. In all Tactile Internet based healthcare technologies, high fidelity and extreme precision is fundamental to enable the deployment of tele-medical technology.

E. Virtual and Augmented Reality

Existing virtual and augmented reality applications can significantly benefit from the availability of the Tactile Internet. The virtual reality is a shared, haptic virtual environment in which several users are physically coupled via a simulation tool to jointly/collaboratively perform tasks by perceiving the objects not only audio-visually but also via the touch sense. In augmented reality, on the other hand, the combination of real and computer generated content is visualized in the user’s field of view. The major goal of future augmented reality applications, compared to today’s static information augmentation, is the visualization of dynamic content and up-to-date information.

Haptic feedback in virtual reality is a prerequisite for high-fidelity interaction. Especially, the perception of objects in virtual reality via the sense of touch leads to various applications relying on high level of precision. This precision can only be realized if the latency between the users and the virtual reality is a few milliseconds.

The augmentation of additional information into a user’s field of view enables the development of many assistance systems, e.g., maintenance, driver-assistance systems, education. With the Tactile Internet the content in augmented reality can be moved from static to dynamic. This enables a real-time virtual extension of a user’s field of view, so that possible dangerous events can be identified and avoided.

F. Further Tactile Internet Applications

Additional Tactile Internet applications are serious gaming, education, individualized manufacturing, and unmanned autonomous systems. Serious games are real-world simulations designed for the purpose of solving a problem. The end-to-end delay in the interaction between players and games is a key factor influencing the quality of players’ experience and the game’s usability, since the delay influences directly the perceived realism of the game.

Individualized manufacturing unlike the mass production in today’s assembly line based production processes will enable the manufacturing of good in production islands. Hereby, mobile robots will deliver assembly parts on demand. This requires a wireless real-time tactile communication network among the mobile robots.

Unmanned autonomous or remotely controlled systems are increasingly used in a large number of contexts to support humans in dangerous and difficult-to-reach environments, remotely controlled by humans, or for tasks that are too tedious or repetitive for humans. The remote control of an unmanned aircraft, for example, can be realized with high precision and without any reaction delay with a reduced end-to-end latency as a Tactile Internet application.

III. TACTILE INTERNET REQUIREMENTS

The Tactile Internet, wherein humans will wirelessly control real and virtual objects, will not be realized without overcoming the enormous system design challenges. Some of the most stringent design challenges for the Tactile Internet have been recently presented in [17].

Human beings’ interaction with their environment is crucial. Our perceptual processes limit the speed of our interaction with our environment. We experience interaction with a technical system as intuitive and natural only if the feedback of the system is adapted to our human reaction time. Consequently, the requirements for technical systems enabling real-time interactions depend on the participating human senses. Hereby, reaction times of about 100 ms, 10 ms, and 1ms is required for auditory, visual, and manual interaction, respectively. Realizing these reaction times, all human senses can, in principle, interact with machines. Hence, human beings should not only be able to see and hear things far away, but also touch and feel them. Transmitting accurately the equivalent of human touch via data networks is the vision to close the data cycle [18], which is aimed to be realized by the Tactile Internet.

In the following, we highlight the key technical requirements for realizing the Tactile Internet.

A. Ultra-Responsive Connectivity

The Tactile Internet requires ultra-responsive network connectivity i.e., end-to-end latency on the order of 1 ms [19], [20]. For real-time transmission as otherwise the tactile users will experience cyber-sickness, which occurs primarily as a result of conflicts between visual, vestibular, and proprioceptive sensory systems [21]. Thus, if eyes perceive a movement which is slightly delayed compared to what is perceived by the vestibular system while the remainder of the human being’s body remains static, this delay leads to cyber-sickness. This is especially important for technical systems with tactile and haptic interaction or for mission critical communications, e.g. machine-type communication which enable real-time control and automation of dynamic processes in industrial automation, manufacturing, traffic management, etc.

The end-to-end latency (round-trip delay) in technical systems includes the time spent in the transmission of the information from a sensor (or human in case of haptic interaction) via the communication infrastructure to a control server; the processing of the information and the eventual retransmission via the communication infrastructure back to the actuator (human). Considering an end-to-end latency of 1 ms the latency budget for wireless transmission is even lower than 1 ms (see Fig. 3).

B. Ultra-Reliable Connectivity

The phenomenal success of cellular networks has been based on providing ubiquitous and reliable wide area coverage for voice and text communications. With 4G and the success of mobile computing devices the industry is targeting to bring the same reliability and ubiquity of access to mobile internet applications such as web browsing and audio and video
streaming. The Long Term Evolution (LTE) today already is providing effective data rates of around 50 Mb/s. However, considering technology and market forces in 10 years, we must be able to address cellular speeds of 10 Gb/s or more and introduce new applications [5], [6], [22]. While high data rates will be a key feature of 5G networks, another key challenge is to be able to provide carrier grade access reliability. Beyond single digit end-to-end latency, ultra-reliable network connectivity is an important requirement for the Tactile Internet. Reliability refers here to the probability to guarantee a required function/performance under stated conditions for a given time interval [23]. The specific reliability requirements differ for various types of services and applications.

Demands for the highest possible reliability are associated with requirements for real-time response. This becomes clear, with applications addressed in Section II requiring a reliable reception of rapidly transmitted data. A failure rate even below $10^{-7}$ might be necessary in some 5G applications [24], [25]. This corresponds to merely 3.17 seconds of outage per year. Wireless systems of today are built around the perception that a link with 3% outage is a good link. However, when two links with uncorrelated channels are combined, 3% outage per link generates a combined outage of approximately $10^{-3}$. And five uncorrelated links can already achieve an outage of less than $10^{-7}$ [4]!

Hence, the simultaneous connection to multiple links (multi-connectivity) might be a potential solution for achieving the required hard-bound (i.e. not average!) reliability for tactile applications [26]. The achieved reliability will also have a positive impact onto delay since less re-transmissions will be needed.

C. Security and Privacy

Safety and Privacy are also the key requirements for the Tactile Internet. With stringent latency constraints, security must be embedded in the physical transmission and ideally be of low computational overhead. Novel coding techniques need to be developed for tactile applications that allow only the legitimate receivers to process a secure message. Absolute security will, hereby, be achieved if an illegitimate receiver cannot decode the data even with infinite computational power. This rises a challenge, especially in massive connectivity applications. Identification of legitimate receivers requires novel reliable and low-delay methods. One such method could be the usage of hardware specific attributes such as biometric fingerprints.

D. Tactile Data

The Tactile Internet must handle the tactile information in the same way as the conventional audio/visual information. Hence, tactile encoding mechanisms are needed which facilitate transmission of tactile information over packet-switched networks. Besides, there must be provisioning of audio/visual sensory feedback due to the highly multi-dimensional nature of human tactile perception [27].

E. Edge Intelligence

The Tactile Internet must overcome the fundamental limitation due to finite speed of light. Without this, the range of tactile services and applications would be limited to 100km (assuming most is through fiber). To overcome this, the Tactile Internet must support a hybrid composition of machine and human actuation mixing real tactile actuation with intelligence-based predictive actuation. Such predictive actuation should be in close proximity of the tactile edge. Therefore, the edge of the network (mobile edge cloud) must be equipped with intelligence to facilitate predictive caching as well as interpolation/extrapolation of human actions. This necessitates the development of novel artificial intelligence techniques for edge cloud architectures.

IV. ARCHITECTURE AND TECHNOLOGY COMPONENTS

Unlike the conventional Internet which provides the medium for audio and visual transport, the Tactile Internet will provide the medium for transporting touch and actuation in real-time i.e., ability of haptic and non-haptic control through the Internet. Unlike auditory and visual senses, the sense
of touch occurs bilaterally i.e., it is sensed by imposing a motion on an environment and feeling the environment by a distortion or reaction force [28]. The key distinction between haptic and non-haptic control is that in case of the former, there is actually a haptic feedback (kinesthetic or vibro-tactile) from the system, in addition to audio/visual feedback, thereby closing a global control loop; whereas in case of the latter, the feedback can only be audio/visual and, hence, there is no notion of a control loop. It should be noted that the haptic control is inherent to a majority of tactile applications.

As shown in Fig. 1, the end-to-end architecture for the Tactile Internet can be split into three distinct domains: a master domain, a network domain, and a controlled domain.

A. Master Domain

The master domain usually consists of a human (operator) and a human system interface (HSI). The HSI is actually a haptic device (master robot), which converts the human input to tactile input through various tactile coding techniques. The haptic device allows a user to touch, feel, and manipulate objects in real and virtual environments, and primarily controls the operation of the controlled domain as discussed later. It should be noted that in some applications, multiple operators can collaboratively control the operation of a single controlled domain. The master domain also has the provisioning for auditory and visual feedbacks. In addition to being important requirement for non-haptic control, the auditory and visual feedbacks play a critical role in increasing the perceptual performance as the human brain naturally integrates different sensory modalities [29].

State-of-the-art haptic devices, available from vendors like Geomagic and Sensable are usually designed in the form of a linkage-based system which consists of a robotic arm attached to a stylus. The robotic arm tracks the position of the stylus and is capable of exerting a force on its tip. To truly realize the vision of the Tactile Internet, further developments on haptic devices are needed; particularly in increasing the degrees of freedom (DoF) to meet the demands of envisioned applications and embedding the network interface for direct or indirect communication with the cellular network.

B. Controlled Domain

The controlled domain consists of a teleoperator (controlled robot) and is directly controlled by the master domain through various command signals. The teleoperator interacts with various objects in the remote environment. Typically, no a priori knowledge exists about the environment. Through command and feedback signals, energy is exchanged between the master and controlled domains thereby closing a global control loop.

C. Network Domain

The network domain provides the medium for bilateral communication between the master and controlled domains, and therefore kinesthetically couples the human to the remote environment. Ideally, the operator is completely immersed into the remote environment. The Tactile Internet requires ultra-reliable and ultra-responsive network connectivity that would enable typical reliabilities and latencies for real-time haptic interaction. The underlying 5G-driven communication architecture, composed of the Radio Access Network (RAN) and Core Network CN), is expected to meet the key requirements in realizing the vision of the Tactile Internet.

To this end, the important functions of the 5G RAN in the Tactile Internet ecosystem are as follows: i) efficient support of various Radio Access Technologies (RATs) such as traditional cellular, emerging millimeter-wave, massive MIMO, full-duplex, etc. ii) Tactile QoE/QoS aware scheduling and radio resource management for tactile applications in co-existence of other vertical applications such as machine-to-machine, vehicle-to-vehicle, smart grids, etc. iii) efficient packet delivery through reliable radio protocols and physical (PHY) layer, and iv) optimal resolution of air-interface conflicts through novel medium access control (MAC) techniques. The key functionalities of the 5G Core Network (CN) relevant to the Tactile Internet are as follows: i) dynamic application-aware QoS provisioning, ii) edge-cloud access, and iii) security.

Although a number of research efforts are focusing on 5G systems, there is no unanimous agreement on a 5G network architecture yet. However, both the academic and industrial communities have a general consensus that 5G networks must be designed in a flexible manner such that one network, based on a common physical infrastructure, is efficiently shared among different vertical applications. Such sharing will be possible through greater degree of abstraction of 5G networks wherein different network slices would be allocated to different vertical application sectors. A network slice is defined as a connectivity service based on various customiziable software-defined functions that govern geographical coverage area, availability, robustness, capacity, and security [30]. Such slicing approach provides more of a network on-demand functionality.

The recent trends of network function virtualization (NFV) (providing abstraction [31]) and software defined networking (SDN) (providing flexibility [32]) are critical in shaping such an envisioned architecture. NFV provides the separation of network functions from the hardware infrastructure; the network function can be managed as a software module that can be deployed in any standard cloud computing infrastructure. On the other hand, SDN provides an architectural framework wherein control and data planes are decoupled, and enables direct programmability of network control through software-based controllers [33].

Whilst SDN/NFV was initially proposed for the Internet infrastructure, the approach is now being considered/used closer to the edge in the cellular CN and even Cloud-RAN. That opens the interesting possibility to provide an end-to-end flexible/abstracted architecture based on radio-aware SDN/NFV slicing in the networking domain and network-aware Radio Resource Management (RRM) & scheduler approaches in the wireless domain. Through such a coupling, it is possible to design one network in a flexible manner offering different end-to-end network slices to different vertical applications. For example, the logical architectural approach, illustrated in Fig. 2, builds on a common programmable physical infrastructure and an NFV-enabled network cloud that
Late goals: Actuator, Sensor, discussed in more details in subsequent sections. These are now underactuated by cloud and content-centric designs. These are now underactuated by cloud and content-centric designs.

The VNF management and orchestration framework is used to monitor, manage, and troubleshoot VNFs software. Such an architectural approach enables flexible and dynamic slicing of end-to-end network and service resources, which is particularly attractive to cater for the requirements of tactile as well as other vertical applications.

As highlighted in Section III.A, for achieving a round-trip latency of 1ms, the communication delay due to the speed of light needs to be considered as well. Within 1 ms light travels (assuming mostly fiber) some 200 km. This means that the maximum distance for a steering and control server (see Fig. 3) to be placed from the point of tactile interaction by the users is 100 km away. This distance assumes no processing nor network congestion delays anywhere along the communication. Considering a more realistic scenario by taking the additional signal processing, protocol handling, and switching and network delays into account, this requires the control/steering server, to be in the range of a few kilometers from the tactile point of interaction. Hence, control and steering servers need to be built as close to the base station (access point) as possible.

The concept of putting servers at the edge of the mobile radio access network has been coined as the (mobile) edge cloud and the possibly best way to enable this proximity is to combine servers into the same box as base stations and access points. As a solution to this, the vision of a highly adaptive and energy-efficient computing (HAEC) box has been established [34], [35]. The HAEC box is a novel concept on how computing design can be built by utilizing optical and wireless chip-to-chip communication to achieve significant increase in performance compared to today’s servers. Hence, the HAEC box is a very performant, energy-adaptive computing platform which can replace a base station and a control/steering server.

The integration of wireless and optical transceiver functionality is integrated into one 3D-stacked chip, where individual dies are thinned and vertically stacked on each other. This allows very compact and powerful systems, with the additional benefit of mixed technologies, e.g. a digital processor could be realized in a certain process (e.g. complementary metal-oxide semiconductor (CMOS)), while the analog front-end could be realized in another process, allowing higher transmit power. Such a technology offers short connections between front-end. With the availability of 3D chip stack integration, antenna designs will not be limited to chip based materials, i.e. semiconductors and insulator materials. Different materials can be used in one chip stack. This enables, for example, antenna designers to consider materials with low losses at high frequencies and low permittivity for broadband antenna design.

The main target of the HAEC box is to allow enhanced run-time adaptivity together with adaptive high performance computing in an energy efficient way. Hereby, energy efficiency is achieved by providing direct links, which reduce the amount of switches between nodes, and by possibly completely turning off links depending on the required data rates. In addition, HAEC box like servers must have real-time operating systems which can guarantee an extremely small response time, so that they can cope with the latency requirements of Tactile Internet applications.

In addition, the HAEC box is expected to have $10^4$ times the computing performance per unit volume of today’s servers [34]. These 3D chip stacks will not only contain processors but also memory yielding an extreme powerful server. Equipped with a very performant energy-adaptive computing platform and being in proximity, i.e. at the network edge, a system of unprecedented local compute power can be realized. This will enable many new Tactile Internet applications.
B. PHY Design for Ultra-Low Latency

In order to obtain a 1 ms end-to-end latency for the Tactile Internet, it is important to understand the chain between sensors and actuators. Fig. 3 shows an exemplary of latency objectives of a mobile-wireless communication system for the Tactile Internet. The sensor measures, pre-processes and provides its data to the embedded system controlling the air interface. The air interface then passes the data through all protocol layers to the physical (PHY) layer. The same happens at the receiving side, for example a base-station with a connected (mobile) edge-cloud, with the data provided to a control server.

Most of today’s broadband communication systems are based on Orthogonal Frequency Division Multiplexing (OFDM) mainly due to its robustness against multi-path channels [36], [37]. However, to achieve an end-to-end latency of 1 ms, the physical transmission must have very small packets which requires a one-way PHY layer transmission of 100 µs as shown in Fig. 3. Since the packet error correction encoding at the transmitter and the error correction decoding and detection at the receiver limit the packet size to less than the target latency, a packet must be smaller than 100µs packet duration. In current Long Term Evolution (LTE) cellular systems, however, the sub-carrier spacing is 15 kHz and the duration of one OFDM symbol is on the order of 70 µs. This numerology together with the reference symbols design, channel estimation, and channel coding requires significant revision of the cellular PHY for the Tactile Internet, which might become reality with the 5G system.

One way to overcome these limitations for achieving 1 ms delay in OFDM based systems is to change the OFDM numerology, i.e. symbol duration, sub-carrier spacing etc., and enable high levels of diversity, and fast channel estimation together with fast channel decoding (e.g. with convolutional codes). In [24], [38], it has been shown that an OFDM based system with changed OFDM numerology can achieve reliable transmission with 1 ms delay.

In general, in OFDM, a cyclic prefix (CP) is added to each symbol to avoid inter-symbol interference. The low latency required for Tactile Internet applications, however, demands for short bursts of data, meaning that OFDM signals with one CP per symbol may present a prohibitive low spectral efficiency. Additionally, OFDM square pulse shaping leads to high out-of-band (OOB) emission which poses a challenge for opportunistic [37] and dynamic spectrum access [39], [40]. These challenges motivate an investigation into alternative waveforms to OFDM for the next generation networks. Hence, alternative multicarrier schemes are currently being evaluated as candidates for the PHY layer of the fifth generation of mobile communication systems.

For suitably low latency in Tactile Internet applications high efficiency must be achieved with short burst transmissions. An efficient way to achieve this is to filter a group of subcarriers to reduce the OOB emission. In this case, since the bandwidth of the filter covers several subcarriers, its impulse response can be short, so that high spectral efficiency is reached in short burst transmissions. These multiplexing schemes mainly do not consider any CP, so that they are more sensitive to small time misalignment.

Various modulation schemes are discussed for 5G, e.g. Filter Bank Multicarrier (FBMC), Universal Filtered Multicarrier (UFMC), Bi-orthogonal OFDM (BFDM), etc. [41–43]. Another modulation scheme discussed for 5G, is the Generalized Frequency Division Multiplexing (GFDM), one promising solution for the 5G PHY layer [44], [45]. GFDM is a flexible multicarrier modulation scheme allowing to cover CP OFDM as a special case. In addition, GFDM is based on the modulation of independent blocks consisting of K subcarriers carrying M subsymbols. The subcarriers are filtered by circularly shifting a prototype pulse in time and frequency domain. This makes GFDM a waveform capable to achieve low OOB emission, which is a major feature for 5G networks. For low-latency real-time applications, the signal length must be reduced [46]. Because GFDM is confined in a block structure of $M \times K$ samples, it is possible to design the time-frequency structure in a way that the time constraints of Tactile Internet applications can be achieved. The increased complexity of GFDM will be manageable with the evolution of electronics. A flexible, customizable Field Programmable Gate Array (FPGA) platform [47] has been used to develop a GFDM proof-of-concept and testbed for experimental research.

In addition to the waveform design, further enhanced techniques are needed at PHY layer to cope with Tactile Internet requirements. Hereby, high reliability can be realized by efficiently using channel coding techniques to exploit diversity levels [24], [38].

To enable low latency in Tactile Internet applications, fast decoding techniques are desirable. Compared to channel codes with iterative decoding, convolutional codes can start decoding as the data arrives, so that they offer faster receiver processing. In [48], [49], for example, low-density parity-check convolutional (LDPC) codes with stringent latency constraints which allow to combine operation close to the channel capacity with a low structural latency by using windowed decoding has been analyzed. Based on optimized decoding schedules for LDPC codes, a decoding delay of only 100 ns was observed.

In addition, for achieving very high reliability convolutional codes do not have error floors (like turbo codes) and they perform similarly well for short message sizes, so that they are often applied to control applications which are also discussed with the Tactile Internet applications.

Finally, the frame structure for Tactile Internet shall be designed in a way that it supports fast decoding. This can be realized by short transmission time intervals and by placing reference symbols and control information at the beginning of the frame, so that the channel estimation can lead to quick decoding and data can be decoded with smallest decoding delays when it is received.

VI. Wireless Access

Significant changes to the way the wireless access is handled are also needed. Notably, radio access protocols as well as radio resource management require a fresh approach. To this end, we outline some recent developments in both domains.
A. Radio Access Protocols

Radio access protocols need to support the range of relevant frequencies for 5G wireless access. This spectrum flexibility comprises that different duplexing schemes can be used, like frequency division duplex (FDD) and time division duplex (TDD) where the uplink-downlink slot can be configured very dynamically. Furthermore, the access protocols need to support operation in licensed spectrum, as well as in shared-licensed or license-exempt spectrum use.

For Tactile Internet applications, a substantial role of radio access protocols is to provide very low latencies and at the same time provide very high reliability and availability [50]. A high level of diversity in space and frequency is needed to provide high reliability levels [24], [38]. Multi-connectivity is a way to provide this diversity, where the dimensions of multi-connectivity are frequency and space. Connectivity for the device (potentially simultaneously) can be established via multiple frequency layers and / or via multiple sites. On different frequency layers different configurations of the 5G radio interface can be used, e.g. according to the spectrum properties. It is desirable that LTE carriers can be tightly integrated into the multi-connectivity with 5G due to their wide deployment and high availability, in particular during the early 5G deployments [51]. Multi-connectivity via multiple sites shall be flexible to work independent of the type of backhaul connection that exists between the sites (e.g. fibre cable, wireless links or copper cables), and should thus be able to handle various backhaul latencies. A common radio resource control protocol layer is well suited to integrate different frequency layers and spatial transmission paths. Besides joint radio resource management over multiple layers, it provides control plane diversity and fast control plane switching for robust signaling and radio link management procedures like mobility. For the use plane it enables use plane aggregation (for high peak rates) and fast user plane switching (for reliability). By coordinating connectivity states on multiple frequency layers (and RATs) via a common control plane, devices can have optimized sleep modes with very low energy consumption and fast connectivity activation [52]. A common control also allows to separate control signaling from user plane transmission, so that for example system information does not need to be transmitted per frequency layer or at all sites. The overall benefits of an integrated common protocol layer are increased reliability and diversity, increased user plane data rates, high spectral efficiency, and better energy efficiency for the device and the network.

The latency of the radio transmission depends on how quickly radio resource can be allocated for a device when a data packet arrives at the radio interface. Network based scheduling has proven to be an efficient solution, but it comes at the cost for uplink transmission of a scheduling request and scheduling grant phase prior to the actual data transmission. Contention-based transmissions can allow quicker uplink access to the radio channel. However, if collision probabilities are high large delays can occur due to backoff and retransmission schemes. An instant uplink access method is desirable where certain radio resources can be instantly used by devices while collision probabilities are controlled to remain sufficiently low. For periodic traffic types persistent scheduling is desirable. Network based scheduling remains an efficient resource allocation scheme for traffic that does not have prohibitively low delay requirements.

B. Radio Resource Management

Once the access protocols and mechanisms are determined, resources need to be allocated which is typically the role of the RRM protocols. The RRM challenge w.r.t. the Tactile Internet is that the outlined use cases will require round-trip latencies of as little as 1 ms as well as high reliability and capacity (data rates). In some cases, wired access networks are partly meeting the requirements, but wireless access networks are not yet designed to match these needs. Scaling-up research in this area will be essential.

As established above, the 5G access network needs to cope with one-way latencies of only about 100 µs (Fig. 3). The resource allocation of the available physical blocks needs to be done up to 10-times faster than in LTE. The transmission errors inherent to wireless systems will need to be ironed-out through careful design. Applications in industrial environments for example, where potentially huge numbers of robots and machines will work in close proximity, will create challenging interference conditions not satisfied by current wireless systems. Classical approaches for medium-access control to be reconsidered in such an environment, and new techniques may be needed to drive latency down to a bare minimum.

New ideas and concepts to boost access networks’ inherent redundancy and diversity need to be researched to address the stringent reliability requirements of Tactile Internet applications. Multi-connectivity, as well as tight integration with LTE, is one approach to increase reliability, in particular if some of the access layers have limited availability due to challenging propagation conditions as in high frequencies. Simplified and fast resource-access schemes as well as efficient signaling protocols need to be designed to optimize the use of the underlying physical resources. Further, radio resource management has to control the interference levels, so that low latency communication can also be provided with high reliability and availability.

Radio resource allocation is a key component of radio resource management. It has direct impact on throughput, latency, reliability, and QoS for various services. With the introduction of tactile applications into the 5G ecosystem, resource allocation becomes particularly challenging as the available resources would be shared between tactile applications and other human-to-human (H2H) or machine-to-machine (M2M) applications, having different and often conflicting service requirements.

In state-of-the-art LTE networks, packet scheduling is a key resource allocation technique. Packet scheduling takes into account the QoS requirements, buffer status reports, and channel quality of users to maximize the spectral efficiency. Since the nature of tactile applications is different than H2H and M2M applications, the scheduling requirements are different as well. Therefore, using one scheduler for different applications may not result in optimal resource allocation decision.
Due to stringent service requirements, radio resources must be provided on priority for tactile applications. It is particularly desired that there is no external competition on radio resources for tactile applications. A robust approach to guarantee this is to allocate a separate end-to-end “slice” of available resources to tactile applications, which remains dedicated for any ongoing operation. Since the resource requirements will change over time, such an end-to-end slicing must be allocated dynamically.

Against this background, a dynamic and flexible resource allocation scheme is desirable in 5G that maximizes the utility of various applications by ensuring an efficient utilization of radio resources. Such dynamic and flexible resource allocation ought to be achieved through end-to-end “virtualization” of resources, i.e., using the virtualization methods in the SDN domain and seamlessly connect it to the advanced RRM 5G schedulers. Such end-to-end “virtualization” enables flexible slicing, isolation, and customization of resources across different vertical applications and user devices. It has the following key benefits in general and specifically in context of the Tactile Internet.

- It comes as a natural solution for cloud-RAN approach, which is increasingly gaining popularity for 5G.
- It ensures resources are allocated to tactile applications on priority with no external competition.
- It facilitates application of tactile-specific scheduling algorithms for maximizing the utility of tactile applications.
- It enables secure resource allocation owing to isolation of allocated slice from rest of the radio resources.

Based on such virtualized approach and illustrated in Fig. 4, we propose a resource allocation scheme that consists of the following key steps:

- **Dynamic Resource Slicing.** In this step, resources are allocated to different “slices” as in traditional RRM. For efficient utilization of scarce radio resources, a combination of bandwidth-based and resource-based provisioning can be used. In bandwidth-based approach, resource allocation is defined in terms of aggregate throughput that will be obtained by the flows. On the other hand in resource-based approach, a fraction of base station’s resources are allocated to each “slice”. Depending on the number of active devices and exploiting the channel state information, the number of radio resources allocated can be determined statistically.

- **Resource Isolation.** We propose an end-to-end isolation of radio resources between tactile applications/users and other applications/users, and connect it to the slicing in the SDN. For other applications/users, resources can be managed in two distinct ways: (a) isolation of resources across user devices but not across applications, and (b) isolation of resources across applications but not across user devices. Such isolation scheme not only guarantees end-to-end availability of resources for haptic applications, but also allows for customization of resources for other applications.

- **Resource Customization.** In this step, resources are customized according to the service requirements of different applications. For example, tactile applications typically generate a very high packet load and therefore, dynamic scheduling schemes may not be feasible due to disproportionally large signaling compared to tactile data. Hence persistent scheduling schemes are needed to tactile applications wherein radio resources are allocated for a given set of sub-frames.

**VII. Network and Cloud Designs**

Subsequently, we discuss non access stratum infrastructure requirements. In the context of the Tactile Internet, the most important are core networking as well as cloud designs.

**A. Core Network**

In the Tactile Internet ecosystem, the core network must provide adequate QoS as well security for tactile applications. Overall, a thin core network with substantial decrease in the protocol overhead is desirable. The thinning of the core network can be achieved by its functional decomposition and moving some of core functionalities to the access network. This will reduce the number of nodes in the data path and hence reduce the end-to-end latency.

The SDN paradigm is particularly attractive for the mobile core network. The SDN-enabled core network will introduce the programmability and hence the flexibility to tailor the data flow inside the core network with respect to the requirements of tactile applications. This will result in improved service quality and user experience.

Regarding security, the current IP Security (IPSec) protocol functionalities are sufficient for providing the required security in the Tactile Internet. However, there are two major challenges: First, whilst it is possible to handle encryption at the eNB, the current trend is that operators handle security in the P-GW. Since it might be placed hundreds of kilometres from the eNB, very large delays are incurred which are not acceptable with the Tactile Internet. Second, security operations (encryption, decryption, authentication, hashes, etc) are generally very time-consuming which calls for novel and
innovative security solutions which are ultra-quick (and possibly piggybacked into communications operations, such as the channel encoder). Whilst being beyond this paper, innovative approaches are clearly needed to provide adequate security for tactile applications with minimal delays (see Section III-C).

An important issue from the protocol stack perspective is the small packet header to payload ratio. For example, the payload of one packet for a 3-DoF haptic stream is only 6 bytes. With the transition towards IPv6, the issue becomes particularly challenging as the header size doubles. Current header compression techniques [53], [54] have been designed for specific Internet protocols like RTP and might suffer significantly from unreliable wireless links. Besides, there is an inherent trade-off between header compression and the inherent delay this process might entail. Hence, this issue needs to be revisited.

The core Internet latency must be reduced which is variable and largely dictated by queuing delays and geographical routing policies. Faster forwarding techniques improving the packet forwarding efficiency by using high-speed cache and data-flow-based technology can be applied to reduce the latency of the core Internet. In addition, anycast addressing (as discussed in IPv6) can be considered for proximity issues in Tactile Internet. This can be for highly reliable transmissions, especially when the closest interface fails, or to reduce latency through proximity.

B. Ultra-Reliable Edge-Cloud

As we will see, the Tactile Internet profits from cloud technology being natively embedded into the networking technologies. The advantage of cloud computing and storage is that resources are not only shared by multiple users but also dynamically reallocated per demand [55]. The move away from heavy dedicated infrastructure to a shared one maximizes computing capabilities whilst minimizing power consumption, rack space, software licenses, among others. This, in turn, allows one to get applications up and running faster, with improved manageability and less maintenance, and the ability to adjust to fluctuating and unpredictable data flow demands during run-time. Technical enablers for cloud capabilities are low-cost computers/servers; low-cost storage; high-capacity networks; and secure software capabilities in form of virtualization and service-oriented architectures [56], [57]. This is really important to ensure a scalable and trusted up-take of Tactile Internet technologies [58].

As discussed before, the distance from the cloud to the wireless edge can maximum be in the order of a few kilometers so that the 1ms challenge can be met. This, however, requires the cloud application server to be hosted at the edge of the operator’s core network (at best). This offers some important advantages: notably, the applications are stored and executed in a very trusted and secure environment which is particularly paramount to critical Tactile Internet applications. On the downside, scalability is an issue and so is delay if the server is placed deep into the CN.

To circumvent this issue, the notion of Cloudlets has recently been proposed [59]. Here fairly distributed cloud storage and computing capabilities are placed at the very edge of the network. In the case of the cellular network, this would be equivalent to putting it into the RAN or at edge between RAN and CN. That has interesting implications since content (Tactile Internet AI engines, see below) and network management (Cloud-RAN) can be merged into a single infrastructure [60]–[62]. Separation in functionality is ensured via different virtualization approaches [63]. This approach bodes well for the Tactile Internet since infrastructure investments to upgrade to Cloud-RAN capabilities are already underway. Again, the downside is the dependence on the operator since the networking infrastructure will need to be used.

The cloud community, however, has been pushing the edge-cloud concept even further through the Fog Computing and Nebula paradigms [64]–[67]. The latter refers to a dispersed cloud infrastructure that uses (voluntary) edge resources for both computation and data storage. Early trials were able to validate an architecture that enables distributed data-intensive computing through a number of optimizations including location-aware data and computation placement, replication, and recovery. The system was shown to be robust to a wide array of failures and substantially outperform other cloud approaches. If security and trust are solved to the same degree as cellular networks, then the Nebula approach could power Tactile Internet applications at the haptic user equipment (UE).

The different cloud instantiations are shown in Fig. 5. In general, it has become obvious that there is an increasing demand for edge-cloud capabilities, partly because of scalability and partly because of end-to-end delay reasons. For the cellular 5G networking architecture to scale and support that trend, 3GPP will need to further relax some requirements about injecting non-3GPP traffic into the core. First steps had been done with femtocell technologies, i.e. Home eNB (HeNB), HeNB-Gateway, etc; that work would need to be refined to facilitate the trusted and authenticated nebula paradigm whilst not being bound to the current embodiment of the CN. Further recent work has concentrated to incorporating very heterogeneous trusted/untrusted real-time/non-real-time traffic into 3GPP mobile systems [68]–[73].
VIII. ARTIFICIAL INTELLIGENT CAPABILITIES

Content and skillset data will be transmitted over a significantly more powerful 5G core network as well as the next generation Internet. The finite speed of light, however, is the biggest adversary in facilitating the required real-time experience advocated by the Tactile Internet. Whilst the advances on hardware, protocols and architecture are paramount in diminishing end-to-end delays, the ultimate limit is set by the finite speed of light.

Breaking the laws of physics not being an option, other – more sophisticated – techniques need to be invoked to facilitate the required paradigm shift. In essence, so we argue, this will be provided by unprecedented edge artificial intelligence (AI) engines which are cached and then executed in real-time close to the tactile experience. The respective components are discussed subsequently.

A. Advanced Content-Caching

With a proper cloud technology in place, the Tactile Internet application content needs to be loaded or ported. A typical example would be an AI algorithm (see below) which is tailored to work in the context of a remote dentistry operation, or remote car servicing.

These advanced caching techniques and user-oriented traffic management approaches at the edge of the network improve network performance by de-congestion of the core network and reduction of end-to-end latency – the latter is particularly important to the Tactile Internet. Important here is to understand that the caching for a specific Tactile Internet application will not help the actual application at hand but other Tactile Internet applications which run in parallel over the same network; and vice-versa.

Significant work has been conducted on optimum edge-cloud caching policies [74]–[79]. Worth highlighting is the recent work [74] which advocates for proactive caching by using advanced predictive process capabilities as well as the massively improved cloud technologies discussed before. With this approach, peak traffic demands are substantially reduced by intelligently serving predictable user demands via caching at base stations and users’ devices. Whilst the advocated approach pertains to rather long-term windows and file structures, it forms the foundation for predictive Tactile Internet caching, as will become apparent in the subsequent section.

B. Tactile Artificial Intelligence Engines

Arguably the most important content to be stored are AI engines which predict the haptic/tactile experience, i.e. acceleration of movement on one end and the force feedback on the other. That allows to spatially decouple the active and reactive end(s) of the Tactile Internet since the tactile experience is virtually emulated on either end; this, in turn, allows a much wider geographic separation between the tactile ends, beyond the 1ms-at-speed-of-light-limit.

The algorithmic framework is currently based on simple linear regression algorithms which are able to predict movement and reaction over the next tens of milliseconds, mainly because our skillset driven actions are fairly repetitive and exhibit strong patterns across the six degrees of freedom. When the predicted action/reaction deviates from the real one by a certain amount $\epsilon$, then the coefficients are updated and transmitted to the other end allowing for corrections to be put in place before damage is done at e.g. a deviation of $\delta$. This has been illustrated in Fig. 6.

More sophisticated algorithms have become available, ranging from [80]–[86]. For instance, [82] employed a prediction method for three-dimensional position and force data by means of an advanced first-order autoregressive (AR) model. After an initialization and training process, the adaptive coefficients of the model are computed for the predicted values to be produced. The algorithm then decides if the training values need to be updated either from the predicted data or the current real data.

The AI algorithms, ported through intelligent edge-caching and stored on appropriate cloud technologies, aid in enabling the perception of real-time interaction, stabilizing the tactile system, and consequently enhancing the QoE of the Tactile Internet user.

IX. ECONOMIC IMPACT

The Tactile Internet is characterized by extremely low latency, ultra-high reliability, availability and security. Prior Internets are only supporting content delivery, whereby content can be multimedia content or static content, e.g. video streams, data files, voice, or email. With the Tactile Internet real and virtual objects can be steered directly at real-time interaction speeds. Hence, not only content will be transmitted, but control information. The paradigm shift of the Tactile Internet is therefore that communication is built for enabling steering and control. This is a big difference to content delivery in today’s technologies.

This opens up completely new opportunities for existing and new applications in many fields, of which some are listed in Section II. Hence, the Tactile Internet will have a marked impact on business and society, introducing numerous new opportunities for emerging technology markets and the delivery of essential public services.

Being an enabler of skillset delivery, the Tactile Internet is a very timely technology for service and skillset driven...
economies like the ones predominantly found in Europe. Especially, in the business-to-business ecosystem the Tactile Internet will be a strong enabler to drive markets for autonomous cars, remote medical care, and many other industries. For consumers it will revolutionize the way they interact with their environment and surroundings. A preliminary market analysis as presented in [4], shows the potential of the Tactile Internet in different markets such as in mobility, manufacturing, event organization, education with entertainment (edutainment), (health) care, smart grid and further emerging markets like agriculture, drones, constructions, etc. It has been shown, that a preliminary market analysis revealed that the potential market could extend to 20 trillion US dollar worldwide. This is around 20\% of today’s worldwide gross domestic product (GDP) [87].

The Tactile Internet will also have a repercussion on the telecommunications ecosystem. Notably, as illustrated in Fig. 5, the content-bearing cloud infrastructure needs to be brought really close to the edge so as to enable the 1ms latency requirement. That, in turn, allows the telecommunications operators to charge the content providers and thereby opening an additional stream of revenues. Whilst this happens to a limited degree with companies like video-streaming service Netflix today, the Tactile Internet will see an explosion of edge-content caching through cloud and thereby an exponentially increasing business-to-business (B2B) market opportunity for operators.

Furthermore, telecommunications vendors may also play an increasingly direct role in this ecosystem. The reason is because (at least early) Tactile Internet applications will be less consumer but more B2B driven. B2B customer acquisition however is more in-line with vendors, and indeed currently vendors push hard for alliances with various B2B companies, like oil/gas, transport, etc. Once locked into these markets, the vendors will be able to procure the best operator deals which in a sense flips the business model of cellular communications.

Overall, the Tactile Internet opens up massive business opportunities for the operators, vendors, over-the-top-content providers; and society at large.

X. CONCLUSIONS AND FUTURE WORK

The different embodiments of the Internet will be dwarfed by the emergence of the Tactile Internet that will be able to deliver real time control and physical tactile experiences remotely. It will revolutionize almost every segment of the society. It is expected that the next generation 5G mobile communications systems will underpin the Tactile Internet at the wireless edge.

It was thus the aim of this article to investigate the interesting area of 5G and Tactile Internet intersection. After discussing the exciting Tactile Internet applications, key technical requirements have been identified for the Tactile Internet. The paper covered several technical issues and challenges pertaining to 5G networks that must be addressed to enable the Tactile Internet including innovations in end-to-end architecture, revolutions in hardware and PHY layer transmission novel approaches to radio access and core protocols as well as to resource management and unprecedented edge-cloud and AI capabilities. The biggest challenge — given the unprecedented requirements on delay and reliability — will be to ensure tight whilst at the same time scalable integration of the various technology components into a single, seamless end-to-end networking experience. Once achieved, the Tactile Internet will have a massive impact onto business and society. It will create new opportunities for vendors, operators, content providers and other members of the service chain.

To sum up, research on the Tactile Internet is still in its infancy. Besides generating interest in the research community, the paper aims to open several areas for future research including ultra reliable network connectivity in the wireless domain, reducing end-to-end latency through innovations in air interface, backhaul, core networking, and the Internet, efficient ways of coding tactile information, and overcoming the physical limit due to finite speed of light.

ACKNOWLEDGMENT

The authors thank the researches at the Collaborative Research Center 912 "Highly Adaptive Energy-Efficient Computing" supported by the German Research Foundation (DFG) for their supports [35]. That work was also partially funded by the European Commission H2020 5GPPP projects NORMA and VirtuWind. Finally, we would like to thank Stefan Parkvall, Johan Torsner and Icaro Da Silva for their valuable comments.

REFERENCES


E. Bastug, M. Bennis, and M. Debbah, “Living on the edge: The role of proactive caching in 5g wireless networks,” *Communications Magazine, IEEE*, vol. 52, no. 8, pp. 82–89, Aug 2014.


Joachim Sachs is a principal researcher at Ericsson Research working on future wireless communication systems. After studies in Germany, France, Norway and Scotland, he received diploma and doctorate degrees from Aachen University and the Technical University of Berlin, Germany, respectively. In 2009 he was a visiting scholar at Stanford University. Since 1995, Joachim has been active in the IEEE and the German VDE Information Technology Society (ITG), where he currently co-chairs the technical committee on communication networks and systems.

Gerhard Fettweis Prof. Gerhard Fettweis earned his Ph.D. under H. Meyr’s supervision from RWTH Aachen in 1990. After one year at IBM Research in San Jose, CA, he moved to TCSI Inc., Berkeley, CA. Since 1994 he is Vodafone Chair Professor at TU Dresden, Germany, with 20 companies from Asia/Europe/US sponsoring his research on wireless transmission and chip design. He coordinates 2 DFG centers at TU Dresden, namely cfaed and HAEC. Gerhard is IEEE Fellow, member of the German academy acatech, and his most recent award is the Stuart Meyer Memorial Award from IEEE VTS. In Dresden he has spun-out eleven start-ups, and setup funded projects in volume of close to EUR 1/2 billion. He has helped organizing IEEE conferences, most notably as TPC Chair of ICC 2009 and of TTM 2012; and as General Chair of VTC Spring 2013 and DATE 2014.

Mischa Dohler (S’01–M’03–SM’07–F’14) is full Professor in Wireless Communications at King’s College London, Head of the Centre for Telecommunications Research, co-founder and member of the Board of Directors of the smart city pioneer World-sensing, Fellow and Distinguished Lecturer of the IEEE, and Editor-in-Chief of the Wiley Transactions on Emerging Telecommunications Technologies and the EAI Transactions on the Internet of Things. He is a frequent keynote, panel and tutorial speaker. He has pioneered several research fields, contributed to numerous wireless broadband, IoT/M2M and cyber security standards, holds a dozen patents, organized and chaired numerous conferences, has more than 200 publications, and authored several books. He acts as policy, technology and entrepreneurship adviser. He has talked at TEDx and had coverage TV & radio.