5G: Personal Mobile Internet beyond What Cellular Did to Telephony

Gerhard Fettweis, TU Dresden
Siavash Alamouti

Abstract

Cellular technology has dramatically changed our society and the way we communicate. First it impacted voice telephony, and then has been making inroads into data access, applications, and services. However, today potential capabilities of the Internet have not yet been fully exploited by cellular systems. With the advent of 5G we will have the opportunity to leapfrog beyond current Internet capabilities.

Motivation

Life without modern wireless communications, as it existed before 1990, is now difficult to imagine. It empowers our modern life, enables modern societies to operate efficiently, and has had a major impact on modern politics, economy, education, health, entertainment, logistics, travel, and all industries.

In order to have a productive life, we need to obtain and share information from various sources. Thankfully, most of the information and data known to mankind has been digitized in one form or another, and is available for consumption and sharing. Hence, we require reliable and fast data communications for real-time access and sharing of this information, which is stored and consumed in various formats and structures (text, image, audio, and video).

First and second generation mobile communications networks were dominated by analog and then digital audio signals and text messaging. The third generation was more about scaling the number of users on the network for voice communications and text messaging, but was overwhelmed by an unpredictable tsunami of image and video content. This trend is sure to continue. Video, audio, and image formats are going to become richer and will require even more data, most probably beyond improvements in codec technologies. Therefore, the thirst for data communications is going to continue, and our transmission networks will most probably remain the bottleneck. Hence, we need to provide as much capacity as we can, and ensure that we build an efficient and smart architecture that can accommodate future demands for data communications.

Moreover, the landscape of information has expanded greatly in the area of machines. There are many new gadgets and devices (hereby referred to as machines) that are potential sources of valuable information. Thanks to advances in communications technology, machines can be connected and reached cost effectively and will fast become an integral part of the global information network. This will definitely drive an ever increasing demand for monitoring and sensing data and applications. The nature of most of this data will be different than conventional human generated content, and will mostly be short, bursty, and asynchronous. Nevertheless, the number of these devices will be orders of magnitude larger than the conventional communications devices of today. Hence, we need to architect our networks to handle tens of billions of devices growing rapidly, limited only by the capabilities and capacities of our networks.

Beyond our need for communication and sharing is the need to steer/control elements of our surroundings and environment, as gadgets, sensors, and machines help us carry out our day-to-day life more efficiently. Once machines become connected, the next natural leap is to have them controlled remotely. This will generate a completely new paradigm for control communications.

Today we enjoy the power of telephony and data communications. Our fourth generation (4G) networks enable real-time access to richer content and enable early application of machine type communication, while control communications is in its infancy. This article provides a recollection of our current state to motivate and sketch a vision of our future.

The mobile industry has had a chronology of revolutionary applications and technologies that have shaped the daily lives of their customers. First and foremost, the need for untethered telephony, and therefore real-time wireless communications, dominated the success of cordless phones, followed by cellular communications. Soon thereafter, two-way paging implemented by text messaging became another killer application. With the success of wireless LAN technology (WiFi based on the IEEE 802.11 standard), Internet browsing, and the widespread market adoption of laptop computers, untethered Internet data con-

 0163-6804/14/$25.00 © 2014 IEEE

nectivity became a reality and ultimately a necessity for everyone. This phenomenon opened the market for cellular broadband wireless data connectivity. The logical next step was to invent a better user experience for a subset of laptop functions for mobile use and merge it with the cellular telephone, which evolved into today's smartphone. We now enjoy access to the world's information at our fingertips, anytime, anywhere. But, is this the end game? Is everything else going to be evolutionary? As difficult as it is to predict, history has shown that the future is ripe for transformations and inventions, especially since we are far away from an ideally connected world.

When the cellular concept was first introduced, it was not adopted immediately and took more than 20 years to be deployed. There were commercial concerns about the business viability of a consumer mobile communication service. One major concern with the cellular concept was that telephony was already ubiquitous and available everywhere: in people's homes, offices, hotels, public venues, on every block on every street, and even in automobiles if you were willing to pay for it (and in fact very few people did, strengthening commercial concerns about the cellular concept). In other words, there initially was no perceived business case or "killer application" for cellular communications and hence reluctance to make the larger investments in infrastructure required to make it happen. Another reason for the long delay in recognizing the cellular opportunity early on was the underestimation of the prospects of semiconductor scaling. This enabled mobile phones at a size, price, and cost that make them portable and affordable for everyone. As a result, the industry could not predict that cellular communications would create a paradigm shift that would forever change all our lives. The paradigm shift was the personalization of communication and the ability to call people instead of places. Nobody had predicted the scale of impact of cellular communications and that it would be one of the largest businesses in the history of mankind with an immense impact on all economies [1].

Is the world ripe for another paradigm shift? It sure feels like it. The Internet and mobile communications have become such as integral part of our lives. However, our experience with it is fragmented and far from ideal, very much like in the early days of telephony and before mobile communications.

Starting with connectivity, the experience is poor and unpredictable. We need to be aware of the networks (3G, 4G, WiFi, Bluetooth, etc.), and need to sign in and pair in order to get access, and at times we cannot get connectivity required for the simplest applications, even browsing. 4G cellular will go a long way to address this, but the thirst for data will continue, and we need to provide more and more capacity as time goes by far beyond two to three times the spectral efficiency and an order of magnitude capacity improvement from 4G.

Another area of opportunity is the form and context of communications. We have many ways of reaching others, which can become very confusing. We need to reach a state where we will seamlessly connect to the right individuals at the right time and through the right means. As a consequence, an individual driving should never be enticed to physically pick up a phone to answer a phone call or text. The context should be known, and the best method of communication should be chosen based on the context of the users. Our devices and gadgets need to evolve to be able to present information in the most convenient form possible using wearables (watches, glasses, etc.), and many physical objects should be capable of sending, receiving, and rendering information (smart tables, whiteboards, screens, etc.)

Content discovery is another area ripe for fundamental change. Because of the colossal amount of information on the Internet and intranets, it is becoming more difficult to find relevant information and content. The Internet is an improvement over the old broadcast model of content distribution as it has increased our choice and reach, but it is still far from ideal. Even the vast improvements in speed and quality of search engines have not been able to keep up with the demand for relevant information. The business model for search and discovery has further impacted the quality of content discovery. The more you pay, the more "relevant" your content will become for users and will appear on the top of their list. But more important, you have to search for your relevant content repeatedly and inefficiently.

We believe this is where the next paradigm shift in mobile communications may lie. Instead of the consumers going to the Internet, the Internet will come to them, and in fact we will become nodes on the Internet. We will become both the source of valuable information and the sink for highly personalized information and content. For this to happen, eventually, all people and the information on the context of their environment need to be continuously available to one another. Also, the context needs to be understood and communicated across multiple nodes, and our gadgets and networks need to become smart enough to understand what content is relevant to whom. Clearly, this requires a new level of security, integrity, and safety to be implemented to achieve the necessary privacy for the technology to be accepted and launched successfully.

A well recognized trend is connecting humans with machines and with other machines, referred to as machine-to-machine (M2M) or machine type communications (MTC). Over recent years we have seen a multitude of wireless M2M applications being deployed (e.g., in public transport systems and vending machines). However, the commercial success has been somewhat limited. Why? The previous revolutionary applications have clearly been driven by addressing human needs for basic communications, and our networks have been designed for voice and text applications, whereas M2M, by nature, has a very different set of requirements. The application of cellular communications to M2M has been an afterthought. Undoubtedly, we will see major breakthroughs in this area, and we should strive to understand these requirements on their merits. For M2M to reach its full potential, it needs a network optimized for it. The big question that our industry needs to answer is whether we will have the same network designed for both human and machine communications, a new dedicated network for machines, or a hybrid.

We believe this is where the next paradigm shift in mobile communications may lie. Instead of the consumers going to the Internet, the Internet will come to them, and in fact we will become nodes on the Internet. We will become both the source of valuable information and the sink for highly personalized information and content.

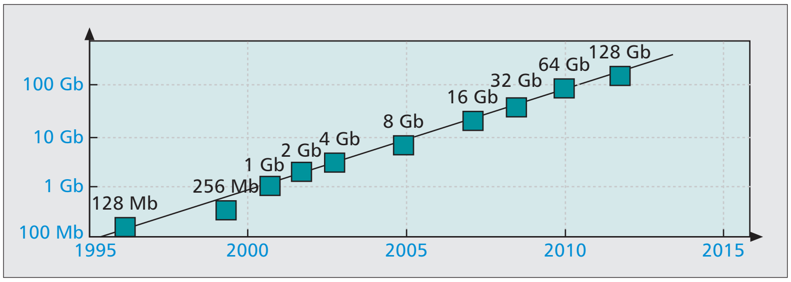

Figure 1. *The market introduction of single-chip flash memory capacity showing a doubling in capacity every 18 months along Moore's Law, leading to a 10× increase every five years.*

In the next section, we discuss the history of wireless and attempt to draw some conclusion to help plan for the future roadmap.

Content: The Wireless Roadmap

Up until now, telecommunications has been used mainly for transmission of various kinds of content (voice, video, data, email, etc.). The majority of this content is generated or stored in devices. In the past decades, advances in mobile electronic devices have been staggering. Within mobile devices, content is mainly stored in flash memory. The storage capacity of flash memory has kept growing according to Moore's Law: doubling every 18 months. This is equivalent to an order of magnitude increase (10×) every five years, as depicted in Fig. 1. As embedded memories grow, so does the need to transmit more and more data [2].

In the past two decades, wireless communications has advanced greatly. The simplest measure of performance increase has been advances in achieving higher data rates. This trend for increased bandwidth has been following the same exponential increase as Moore's Law for semiconductors, as shown in Fig. 2.

WLAN systems are intended for local area broadband communication services in homes, offices, and hot spots with peak data rates reaching hundreds of megabits per second. With WiGig on the horizon, yet another order of magnitude improvement in speed and latency of data transfer is possible.

The phenomenal success of cellular 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 aiming to bring the same reliability and ubiquity of access to mobile Internet applications such as web browsing and audio and video streaming. Long Term Evolution (LTE) already provides effective data rates of around 50 Mb/s.

The gap between the data rate speed of cellular and WLAN has been 100× (Fig. 2). However, the use cases for mobile data on devices have converged. Consumers only care about applications and services, which demand that the underlying network become completely transparent. In other words, the distinction between applications for local area and wide area communications will disappear, and we will need a truly heterogeneous network, leading to a new inflection point of technology [3].

Another significant driver is the vast amount of spectrum available in the millimeter range beyond 100 GHz for unlicensed WLAN systems with reasonable propagation properties [4], whereas licensed cellular systems might be restricted to sub-100 GHz frequencies only. Recent results in the sub-100 GHz band show promising results for cellular hotspot coverage [5]. The physical layer design for cellular systems is constrained by availability of licensed spectrum, whereas WLAN technologies will have an abundance of frequencies in the unlicensed band. What is very clear, however, is that with the next generation of wireless technologies, the various access technologies will have to be highly coordinated and integrated into a seamless experience, unlike the fragmented experience of today.

Setting the Stage for LTE and Beyond

Considering technology and market forces in 10 years, we must be able to address cellular speeds of 10 Gb/s or more. Current wide area access technologies do not meet this requirement. Using orthogonal frequency-division multiple access (OFDMA) in wireless systems requires analog/digital conversion with higher bit resolution due to its Gaussian amplitude distribution. Hence, for upcoming high data rates, analog/digital conversion with 10-bit resolution alone would most likely represent a power consumption challenge that is difficult to resolve with currently available or projected technologies [6]. This suggests that a new physical layer approach may be needed for 5G cellular communications.

Monitoring

Collecting information by monitoring our environment for understanding the current status or predicting the future is of great value to our future progress in almost all industries. Many of today's smartphone apps fulfill some of our needs (e.g. providing information on location, speed, prices of goods, and weather status). Many new devices such as connected wristbands collect sensory information on activity and vital signs. Obviously, we are at the very early stages of this trend. There are applications that can have significant impact on our daily lives. For example, every individual plant could be monitored and classified. We could ensure proper light and moisture for optimum growth and health, and increase productivity in agriculture and food production. This is at the top of the food chain, which seemingly is furthest from the technology. You can extend this to many other applications.

For new M2M applications to become technically and financially feasible, adding connectivity to devices must be very cheap, and central management of devices needs to be reliable and inexpensive. Given that many of these devices have mobile use cases (e.g., body sensors, or sensors inside cars), most of these devices need wide area connectivity. There is a collective (or social) network aspect to these devices as well, which could further enhance use cases. For instance, for winter sports, if every ski or snowboard could measure and report the status of the run, the information could be aggregated to provide real-time information on the status of every run so that skiers could select their runs based on their desired profile (powder, ice, moguls, etc.) Expanding this vision to our daily lives, you could have hundreds of billions of connected devices assisting human beings in their daily lives, chores, and work [7].

We need to translate the industry vision for M2M into requirements to the extent to which we can envision and attempt to draw some conclusions on technology requirements. An ideal M2M device is one that can be activated only when needed, can run "forever" on a small battery or an energy scavenging unit, and transmits the required amount of information only when needed. The rate at which the data is transmitted and the amount of information communicated depends highly on the particular use case.

An example specification for the use case of a smart meter is to transmit a packet of 25-byte payload every 100 s, to operate on an average 200 μW power budget (a 2 in solar cell), and be able to provide at least a 10 dB better link budget than GSM.

Setting the Stage for LTE and Beyond

We believe it is technically and financially feasible to provide M2M services on a very large scale. However, this is not true with current cellular systems as their protocols require too much communications overhead for synchronization, channel allocation, and mobility/connectivity management. As a result, they are far from meeting the power requirements for a broad-scale M2M cellular sensing system. Hence, a new (5G) standard beyond LTE is needed.

Immersive Steering and Control: The Tactile Internet

In this section we focus on some new use cases for real-time services. Real time is a highly subjective term and depends on the use case. We define a service to be real time when the communication response time is faster than the time constants of the application. We consider four types of physiological real-time constants: muscular, audio, visual, and tactile. For literature on human reaction analysis we refer to [8], and an overview is given in [9].

Humans have the ability to react to sudden environmental changes using our muscles; for example, when reacting to a sudden unforeseen incident by hitting the brakes in a car, or quickly pulling back a hand after touching a hot platter on a stove. There are two distinctly different timescales of reaction, depending on being prepared or unprepared for the situation. If unprepared, the sensing to muscular reaction time is in the range of 500 ms to 1 s. When translating this to comparable situations in technical applications, this sets the targets for specification. An easily understood example is interactive browsing of the web. The page build-up after clicking on a link should be a fraction of this time so that we get a sense of immediacy. Henceforth, real-time browsing interaction is experienced if new web pages can be built up after clicking on a link within a few hundred milliseconds. A shorter latency(i.e., a faster reaction time of the web) is not necessary for creating a real-time experience as the

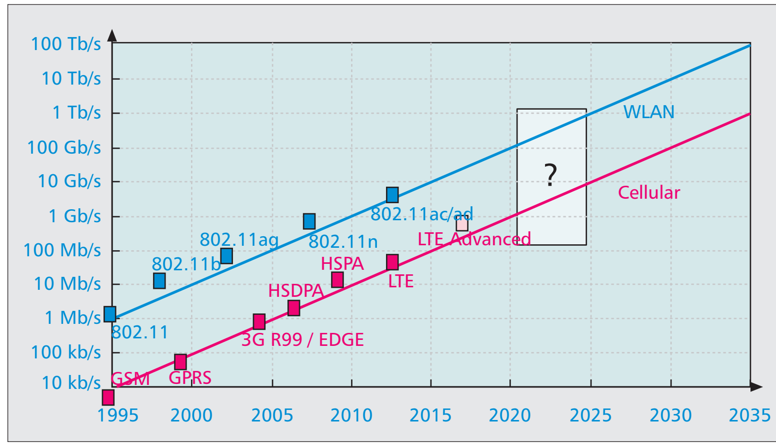

Figure 2. *The Wireless Roadmap, showing the market entry of technologies as tracked by the authors.*

difference is not perceivable by humans. This reaction time has generally been met by initial 802.11b and 3G cellular systems.

Clearly, if humans are prepared for a situation, faster reaction times are needed, such as when driving a formula 1 car in a race or during high-speed typing on a keyboard. In a communications setting this would be the case, for example, if while clicking a link the new web page is known, and the user is preempting where to click next. The International Association of Athletics Federation (IAAF) has come to the conclusion that the fastest "prepared" neural-muscular physiologic reaction time is 100 ms. For instance, in track and field competitions, any sprinter who reacts faster than 100 ms after the shot is considered to have a false start, even though some studies show that 85 ms could be possible [10].

The next shorter real-time latency constant is experienced with the hearing system. To humans, conversations appear as real time when we receive the audio signal within 70 to 100 ms. This is why the International Telecommunication Union (ITU) has set this as a minimum latency requirement for telephony, and speech delays on telephone lines have to be on that order of magnitude. This is not only applicable to two-way communications. For instance, lip synchronization between the video stream and the soundtrack needs to be within the same time lag; otherwise, the sound seems disconnected to movements on the video. Assuming the core network is provisioned properly and the wireless network is far from being fully loaded, an over-provisioned LTE network or even later generation 3.x generation networks meet this requirement. As a result, Internet videoconferencing (e.g., Skype) is generally viable over averagely loaded cellular networks today.

When immersed in watching a scene, our vision has a typical temporal resolution of 100 Hz. Modern TV sets therefore have a minimum picture refresh rate of 100 Hz to allow for a seamless video experience, translating into a maximum10 ms inter-picture latency requirement.

The most challenging latency requirement comes from tactile/haptic action of human limbs with visual or audio feedback. Latency requirements for tactile action are on the order of 1 ms.

What we experience today is only the very first glimpse. Obviously, with challenges as pointed out here, 5G cellular communication will be another paradigm shift that will redefine our future, impacting out societies in ways which cannot be foreseen.

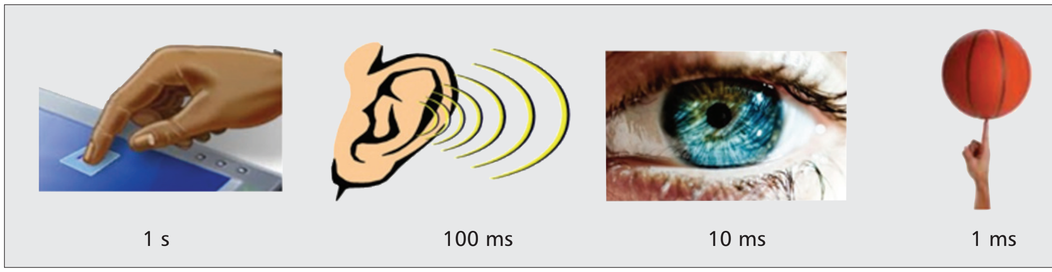

Figure 3. *Coarse categories of physiological real-time constants.*

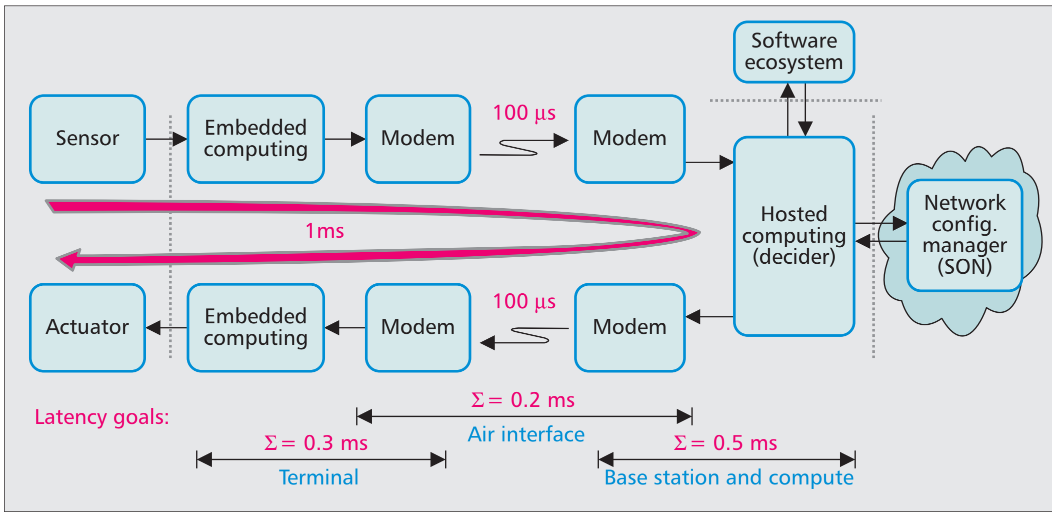

Figure 4. *The impact of breaking down the 1 ms round-trip delay.*

For some results on haptic interaction see [9]. A good example is moving or dragging an object on a touch screen. At a relatively fast speed, our finger moves at 1 m/s over the screen. To create a seamless experience the object must follow the movements of the finger without experiencing a displacement between finger and object. If we assume 1 ms of latency, and 1 m/s speed for the movement of the finger, the resulting displacement is 1 mm, which is acceptable. Much larger latencies would be unacceptable. For instance, a latency of 25 ms would lead to 2.5 cm or 1 in of displacement, which is clearly noticeable and results in a very poor experience.

Also, extreme situations where the 1 ms latency requirement can be experienced are when moving a 3D object with a joy stick or in a virtual reality environment. If the time lag between the virtual picture and the movement is above 1 ms, motion sickness can occur, which within this context is referred to as cyber sickness [11]. Hence, a real-time cyber physical experience requires an extreme latency constraint of below 1 ms. Current wireless cellular systems miss this target by nearly two orders of magnitude (Fig. 3).

The Real-Time Challenge at 1 ms: "The Tactile Internet"

When translating the latency restrictions into building electronic systems, we need to consider the speed of light. Light travels 300 km within 1 ms. In order to meet a 1 ms roundtrip latency constraint, we need to consider limitations in hardware and software design. Figure 4 shows one possible latency budget over a communications chain, taking into account the latencies from the sensor through the operating system, the wireless/cellular protocol stack, the physical layer of terminal and base station, the base station's protocol stack, the trunk line to the compute server, the operating system of the server, the network within the server to the processor, the computation, and back through the equivalent chain to the actuator.

Setting the Stage for LTE and Beyond

Each and every element of this communications and control chain must be optimized for latency. Assuming this, the latency budget for the physical layer is at most 100 μs. LTE has an OFDMA symbol duration of 70 μs. Therefore, if we were to meet the latency requirements for tactile communications, we would need to design a different physical layer. Hence, 5G addressing these new interesting application areas of the "Tactile Internet" [12] requires a completely new physical layer for or an overlay on top of OFDMA for these applications.

Conclusions

We have envisioned the high-level requirements for 5G cellular communications with a user-centric approach, considering three dimensions of innovation: higher data rates to address network traffic demands of the future, operations for

M2M sensing devices, as well as 1 ms real-time latency (Fig. 5) to meet requirements for tactile control.

As all three dimensions may not need to be addressed simultaneously for any class of service, it can be assumed that it is feasible to design a new 5G system that can meet these differing requirements, and that it may differ greatly from 4G LTE:

- Cellular communications with data rates of 10 Gb/s can enable ubiquitously available immersive virtual reality at levels not foreseen.
- Context awareness and personalization built into all devices and services can transform the way we communicate as humans.
- It is most likely that we will have various access schemes to fulfill all our communications needs. These schemes need to become highly integrated and coordinated to provide a seamless experience to users and applications.
- Communicating sensors embedded in our environment can enable a host of new services. We can move from cellular/wireless data communications to a new level of wireless monitoring.
- A round-trip latency of 1 ms can potentially move the world from enjoying today's wireless communication systems into the new world of wireless control systems: the Tactile Internet. It will dramatically change our life, impacting all aspects of application areas such as health, safety, traffic, education, sports, games, and energy. For a first overview on application areas and challenges see [12].

* Discussions on setting the stage for LTE and beyond shows that a new physical layer may be needed, possibly parting from OFDMA and embracing new concepts such as generalized frequency-division multiplexing or filter bank multi-carrier [13].

The result may be a revolutionary leap from today's wireless communications to personalized context-based communications and future monitoring and control networks. The future opportunities in mobile communications are larger than anyone can foresee. What we experience today is only the very first glimpse. Obviously, with challenges as pointed out here, 5G cellular communication will be another paradigm shift that will redefine our future, impacting our societies in ways that cannot be foreseen.

Acknowledgments

We would like to acknowledge the contributions of our teams at Vodafone Group R&D and TU Dresden in shaping the vision, and we would also like to thank the reviewers for their valuable comments.

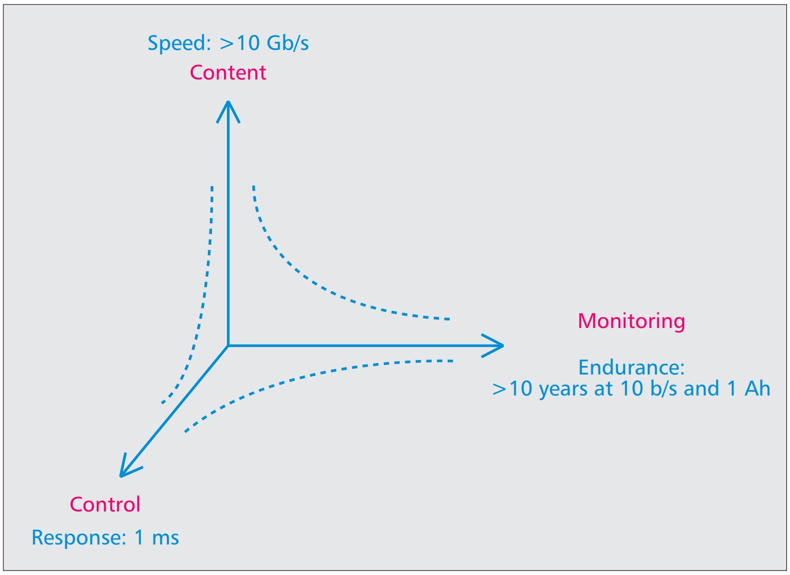

Figure 5. *The wireless network moving from communications of content to monitoring to control. The dashed lines indicate that applications allow for a trade-off between the three depicted criteria of optimization: speed, response time, and operation endurance.*

References

[1] Vodafone Institut, "Mobile Technologies — The Digital Fabric of Our Lives," www.vodafone-institut.de/uploads/media/Studies_komplett_FINAL_von_250913.pdf, Sept. 2013.

[2] G. P. Fettweis, "A 5G Wireless Communications Vision," *Microwave J.*, Dec. 14, 2012.

[3] C. Christensen, "The Innovator's Dilemma," *Harper Business Essentials*, 1997.

[4] J. Well, "Faster Than Fiber: The Future of Multi-Gb/s Wireless," *IEEE Microwave Mag.*, May 2009, pp. 104–12.

[5] T. Rappaport *et al.*, "Millimeter Wave Mobile Communications for 5G Cellular: It Will Work!," *IEEE Access*, vol. 1, Digital Object Identifier 10.1109/ACCESS.2013.2260813, May 2013, pp. 335–49.

[6] S. Krone, "Wireless Communications with Coarse Quantization: Information-Theoretic Analysis and System Design Aspects," *Vogt Verlag*, 2012.

[7] M.A. Uusitalo, "Global Vision for the Future Wireless World from WWRF," *IEEE Vehic. Tech. Mag.*, vol. 1, no. 2, 2006, pp. 4–8.

[8] J. Johnson, *Designing With the Mind in Mind*, Morgan Kaufman, 2010.

[9] E. Steinbach *et al.*, "Haptic Communications," *Proc. IEEE*, vol. 100, no. 4, Apr. 2012, pp. 937–56.

[10] M. T. G. Pain and A. Hibbs, "Sprint Starts and the Minimum Auditory Reaction Time," *J. Sports Sciences*, vol. 25, no. 1, Jan. 2007, pp. 79–86.

[11] T. DeFanti and R. Stevens, "Teleimmersion," Ch. 6, *The Grid: Blueprint for a New Computing Infrastructure, Elsevier Series in Grid Computing*, pp 131–55.

[12] G. P. Fettweis, "The Tactile Internet — Applications & Challenges," accepted for the *IEE Vehic. Tech. Mag.*, to appear 2014.

[13] G. Wunder *et al.*, "5GNOW: Non-Orthogonal Asynchronous Waveforms for Future Mobile Applications," in this issue, *IEEE Commun. Mag.*.

Biographies

Gerhard P. Fettweis (Gerhard.Fettweis@vodafone-chair.com), born March 16, 1962, received his Ph.D. from RWTH Aachen in 1990. Thereafter he joined IBM Research and then TCSI, California. Since 1994 he is Vodafone Chair Professor at TU Dresden, Germany. So far he has co-founded 11 startups. He runs the German Science Foundation's CRC HAEC and COE cfAED. He helps organize IEEE conferences (e.g., TPC Chair ICC 2009 and TTM 2012, General Chair VTC Spring 2013 and DATE 2014).

Siavash Alamouti, born March 16, 1962, received B.A.Sc. and M.A.Sc. degrees in electrical engineering from the University of British Columbia. He is currently an entrepreneur helping build new companies focused on personalization. He was the group R&D director for Vodafone from 2010 to 2013. Before that he was an Intel Fellow and CTO of the Mobile Wireless Group. Prior to joining Intel in 2004, he held positions at various companies including Vivato, Cadence, and AT&T. He is most well known for the invention of the Alamouti Code.