RYSAVY RESEARCH

How “Title II” Net Neutrality Undermines 5G

April 19, 2017

Table of Contents

Notice

Rysavy Research provides this document and the information contained herein to you for informational purposes only. Rysavy Research provides this information solely on the basis that you will take responsibility for making your own assessments of the information. Although Rysavy Research has exercised reasonable care in providing this information to you, Rysavy Research does not warrant that the information is error-free. Rysavy Research disclaims and in no event shall be liable for any losses or damages of any kind, whether direct, indirect, incidental, consequential, or punitive arising out of or in any way related to the use of the information.

Introduction

Engineers and scientists are applying their ingenuity to make 5G a reality in the United States. 5G, the next generation of wireless technology, will be tested in a series of trials this year and will launch widely in the 2020 timeframe. Preliminary field testing and research show that 5G will further unleash the explosive potential of wireless networking by dramatically lifting data throughput rates, expanding capacity, and reducing network delays. 5G network platforms will accommodate the same variety of applications and uses as many wireline networks. The network and service capabilities that 5G makes possible will enable a vast number of new use cases for mobile and fixed wireless technologies.

Unfortunately, the FCC's current approach to ensuring an open internet is premised on Title II of the Communications Act. When applied to 5G network deployment and operation, the regulations will likely have the perverse effect of thwarting many of the most consumer-friendly 5G use cases. The purpose of this paper is to make clear how some of the Title II-based regulations related to net neutrality will sacrifice the very 5G characteristics that hold the most promise for consumers, innovation, and economic growth across multiple sectors of the U.S. economy. There must be a better way to accomplish the important policy objectives of maintaining an open internet in the United States while also allowing 5G to deliver on its promise.

The Collision of Title II-Based Open Internet Rules and 5G

The millions of mobile applications already transforming the world are just the dawn of the next frontier in mobile broadband—humanity has barely begun exploiting the full potential of wireless technology. The Internet of Things, which will interconnect objects to increase their utility and efficiency, will account for tens of billions of new connections by next decade. IoT's potential is limited only by imagination; use cases include self-driving cars with pre-crash sensing and mitigation, health biometric sensing and response, telemedicine, and proactive monitoring of critical physical infrastructure such as transmission lines.

What many of these new applications have in common are stringent data communication requirements, such as high reliability or minimal delay. This is true even for use cases without particularly onerous bandwidth demands. For example, a self-driving car or autonomous robot may need only a small amount of data, but it might have to receive that data within a few thousandths of a second. In contrast, a user watching a movie is not negatively affected if the video stream leaves the server a second earlier, with no interruption to the viewing experience.

The federal government's decision to classify mobile broadband as a Title II common carrier service, intended to subject mobile broadband to net neutrality rules, happened at the same time that standards-based mobile technologies, such as LTE, were beginning to provide quality-of-service (QoS) management capabilities to network operators. Such QoS capabilities improve mobile traffic flows and can enhance user experiences. The timing of Title II-based net neutrality for mobile could not have been worse. Now should be the time for operators to start using QoS parameters to serve different use cases and to experiment with various business models that could support them. However, with Title II as the baseline regulation, it's unclear that QoS capabilities can be used as intended. The requirement that a heart monitor transmission to a hospital emergency room cannot be treated as any more special than a cat video is absurd.

The FCC's Title II-based restrictions on handling different kinds of traffic based on what the bits require could slam the door on a vast number of new applications that are actually pro-consumer and pro-innovation. Worst case, the full potential of 5G may never be fully realized.

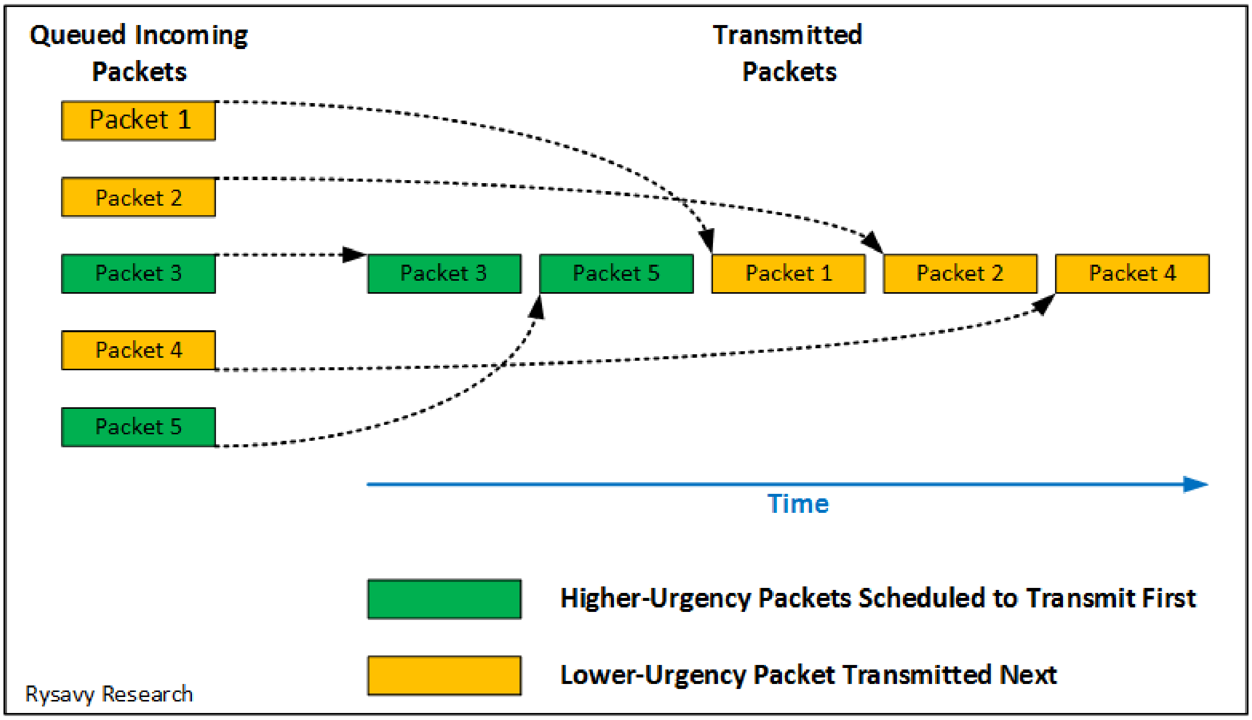
How Quality-of-Service Works

Engineers have designed controls for how packets flow between base stations and users over the radio interface. Traffic-flow parameters include whether bit rates are guaranteed, their priority relative to other traffic flows, the maximum amount of packet delay that can be tolerated by the traffic in question, and the extent of permissible packet loss. LTE specifications define thirteen quality-class identifiers, each with unique parameters.[1] Voice over LTE (VoLTE), which is based on voice-over-IP protocols, uses these QoS mechanisms to provide carrier-grade voice service. Without this control, an LTE voice call would disintegrate if surrounding users were consuming large amounts of data—the network prioritizes voice as higher priority than data. The same prioritization of voice over data also happens in 2G and 3G networks. VoLTE is an operator specialized service, not a broadband Internet service, so Title II net neutrality allows this specific form of prioritization. However, the rules restrict many other potential Internet-based applications from using QoS capabilities.

Figure 1 shows how, in a QoS-enabled network, the network may schedule higher-urgency packets to transmit first, ahead of those with lower urgency.

[1] For details about LTE QoS, refer to 3GPP TS 23.203. *Technical Specification Group Services and System Aspects; Policy and charging control architecture*, available at http://www.3gpp.org/DynaReport/23203.htm. Specifically, see Table 6.1.7, "Standardized QCI characteristics."

Figure 1: Transmission of Packets According to Their Urgency in a QoS-Enabled Network



5G will employ similar, yet more sophisticated, mechanisms to handle different kinds of traffic flows. This is critical because engineers are designing 5G for a wider range of use cases than prior technology generations, such as 3G and 4G. As described below, 5G will employ a "network slicing" architecture that will depend heavily on QoS management. Many of the applications envisioned for 5G are of a control nature, which means they need minimal delay and high reliability.

Table 1 lists some typical applications and their QoS requirements.

Table 1: Examples of Applications and QoS Requirements

Application	Requirements
Speech	Guaranteed bit rate, low delay, but can tolerate some packet loss.
Internet of Things	Varying requirements depending on use case, but mission-critical applications will require low error rate and low delay.
Streaming (music, video)	High throughput, but can tolerate delay and some packet loss.
Health and medicine	Throughput-rate requirements vary. High priority for critical health applications.

Application	Requirements
Autonomous vehicles	High throughput and low delay, with low packet loss.
Video conferencing and telepresence	High average throughput, low delay, can tolerate some packet loss on video but less on voice.
Operating system or application update	Can run in the background over an extended period, so QoS requirements are minimal.
Web browsing	High average throughput, low error rate, can tolerate slight delay.

Current wireless networks assign equal priority to all third-party application traffic, regardless of the application type. An analogy is a freeway on which fast-moving cars and slow-moving trucks use all lanes equally. The Information Technology & Innovation Foundation (ITIF) states in a report, “To date, we have been able to muddle through with this ‘best-effort’ system, but many of the exciting innovations around the corner will increasingly require reliable low-latency connections. And while some applications affirmatively need prioritization or some kind of differentiation, other applications can easily tolerate delay or jitter.”[2]

The goal of intelligent traffic prioritization is to maximize the quality of experience across the largest number of users and application types possible, allocating higher priority for those applications that need it while not adversely affecting those that do not.

As ITIF states, “Traffic differentiation simply is not a zero-sum game.” Because applications have varying quality requirements, selective application of QoS results in higher average quality of experience across the subscriber base. The Broadband Internet Technical Advisory Group agrees, stating, “For example, some differentiation techniques improve the Quality of Service (QoS) or Quality of Experience (QoE) for particular applications or classes of applications without negatively impacting the QoE for other applications or classes of applications.”[3]

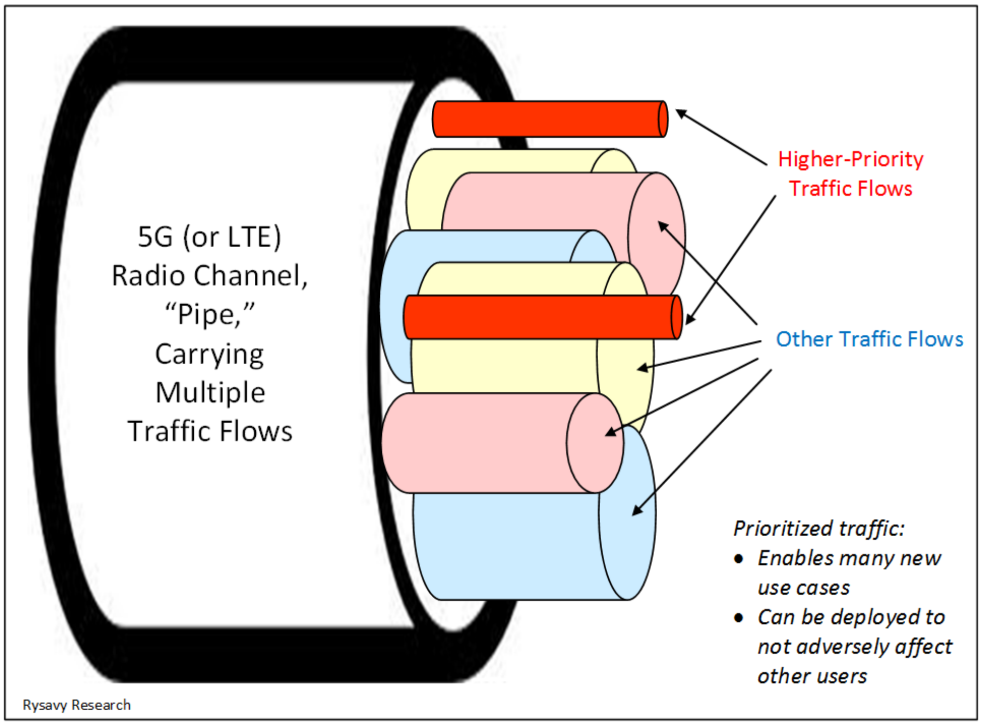
[2] Information Technology & Innovation Foundation, *Crafting a Grand Bargain Alternative to Title II: Net Neutrality with Net Adoption*, October 2015. Available at http://www2.itif.org/2015-alternative-title-ii.pdf.

[3] Broadband Internet Technical Advisory Group, *Differentiated Treatment of Internet Traffic*, October 2015. Available at http://www.bitag.org/documents/BITAG_-_Differentiated_Treatment_of_Internet_Traffic.pdf.

Differentiation is not a zero-sum game. Selective application of QoS increases the quality of experience across the subscriber base.

Figure 2 shows how a 5G wireless network could use QoS management to allocate different priorities to different traffic flows based on their urgency.

Figure 2: Radio Resource Management in a Wireless Network



How Title II Undermines 5G

Within any specific coverage area, a cellular network has a limited amount of aggregate capacity available to users. This capacity is determined by the amount of spectrum deployed and the spectral efficiency of the technology. When the amount of demand is less than the available capacity, applications function well for users in that coverage area. But when demand exceeds capacity, congestion results and applications suffer. The effect is analogous to too many cars traveling on a highway. Initially, cars simply slow down in the face of traffic ahead. If the traffic continues unabated, cars come to a halt. Similarly, in the face of network congestion, applications will initially operate more slowly; for example, a file download takes longer to complete. But as congestion gets worse, packet delays or dropped packets increase to the extent that network transactions time out and fail entirely.

Operators mitigate the worst effects of congestion by deploying more spectrum when possible, installing more cell sites, using more efficient technology, and offloading some traffic onto other networks such as Wi-Fi. But eliminating congestion entirely is impossible. Even a small number of users in the same geographic area simultaneously using high-bandwidth applications, such as video, can consume the entire capacity of a cell. Operators cannot predict how many mobile users will be present at any moment in any location, nor can they know what those users will be doing. Network investment can ensure a high quality of experience on average but cannot guarantee it for all users at all times.

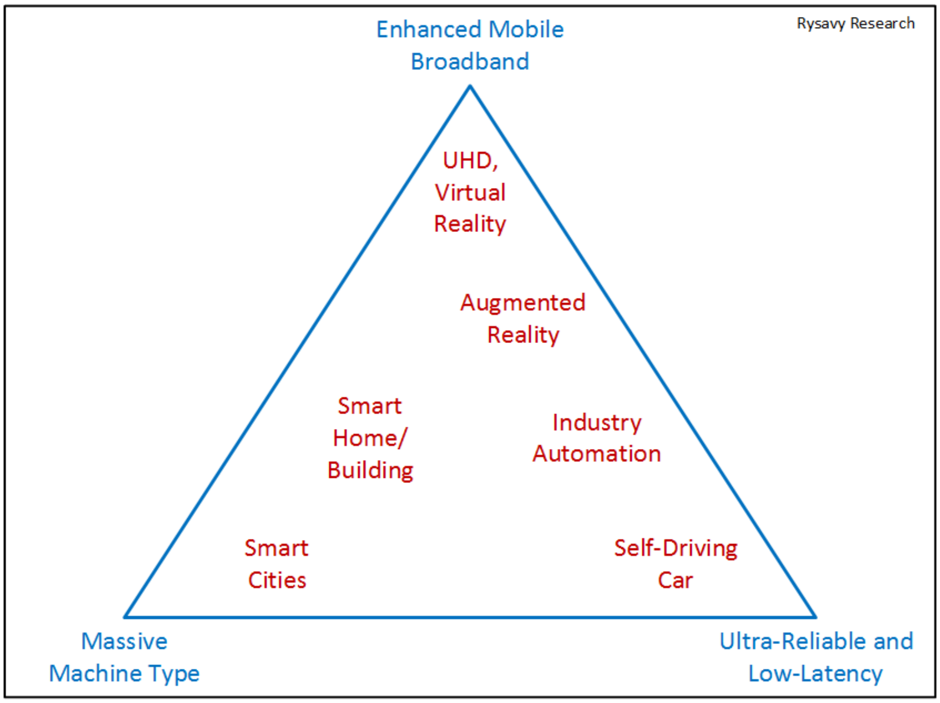
The effect of congestion is neither uniform nor consistent, meaning the degree of congestion can change moment by moment. Therefore, absent network tools to manage around congestion, no user can completely depend on the network for a critical operation. This was an acceptable solution in years past, but consider the applications being developed for 5G: A medical device that has detected a possibly life-threatening event and is trying to send that data to a server for immediate evaluation could fail. A self-driving car may detect debris on the road but not be able to swerve in time. A stock trade could be executed moments too late. A 5G network environment can ensure these critical connections are protected from congestion effects—but only if QoS and other tools being built into the 5G standards are allowed to work.

5G Use-Case Models Depend on Ability to Provide QoS

The International Telecommunication Union is the organization charged with setting 5G objectives and approving final, technical standards for how 5G networks interface with one another and enabled devices. The ITU's recommendation M.2083-0[4] defines use cases using the following model.

[4] Available at http://www.itu.int/rec/R-REC-M.2083-0-201509-I

Figure 3: ITU 5G Use-Case Model



Enhanced mobile broadband is faster Internet, a turbo-charged version of today's LTE-Advanced networks. "Massive machine type" refers to millions of sensors and controls placed throughout cities, homes, and businesses to improve energy efficiency, transportation, and other logistics. But it is the new ultra-reliable and low-latency category, also referred to as mission critical, that opens cellular networks to capabilities never before possible, such as advanced industry automation and autonomous vehicles. This category of 5G application will depend on the ability to deploy traffic prioritization.

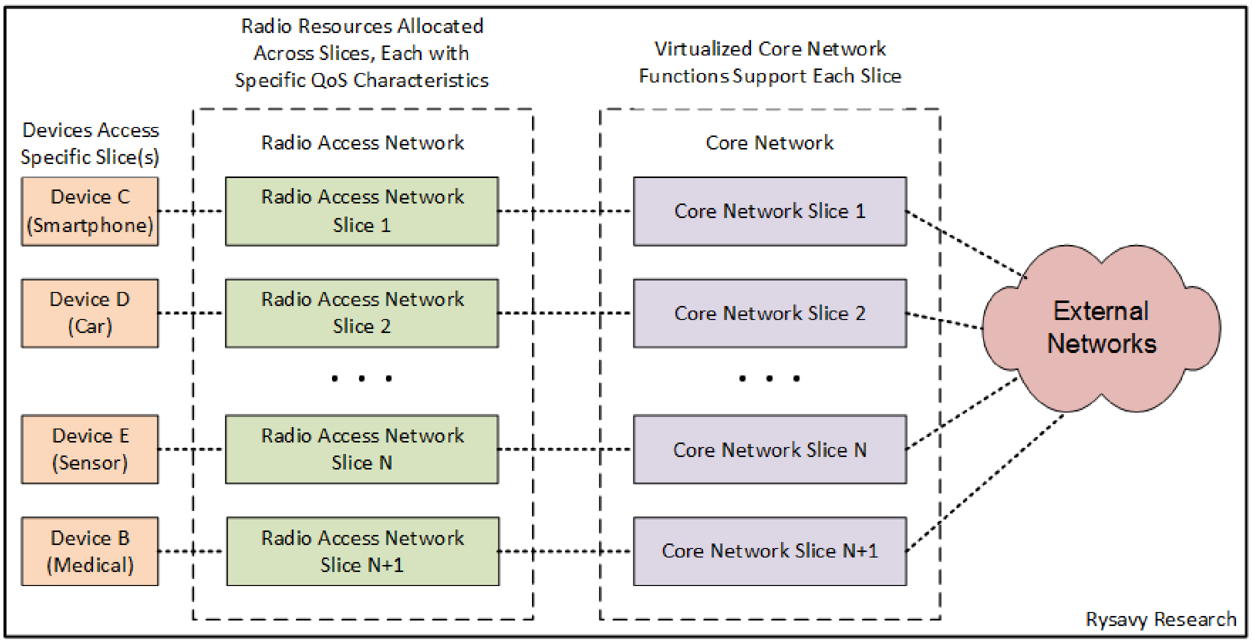
Developers expect response times of less than a millisecond with 5G, ten times lower than with LTE, in which 10 msec latencies are typical. But unprioritized and competing with other traffic, the latency (round-trip time in the network) can be ten times higher, for example, 100 msec. At 60 miles per hour, a car travels nine feet in 100 msec versus only one inch in 1 msec. In a scenario of an intelligent highway warning a car of a pedestrian on the road at a blind curve, that could be the difference between life and death.

5G Networking Slicing and QoS Management

5G needs QoS management, not only for traffic prioritization to support mission-critical applications, but also to enable a fundamental capability in its architecture: network slicing. Network slicing, implemented through virtualization, will allow an operator to provide different services with different performance characteristics to address specific use cases. Each network slice operates as an independent, virtualized version of the network. For an application, the network slice is the only network it sees. The other slices, to which the customer is not subscribed, are invisible and inaccessible. The advantage of this architecture is that the operator can create slices that are fine-tuned for specific use cases. One slice could target autonomous vehicles, another enhanced mobile broadband, another low-throughput IoT sensors, and so on.

Figure 4 shows the network slicing architecture, with devices having access to only the slice(s) for which they have a subscription. Each slice has radio resources allocated, with specific QoS characteristics. Within the core network, virtualized core network functions support each slice and provide connections to external networks.

Figure 4: 5G Network Slicing Architecture



A recent report on network slicing from 5G Americas lists the following examples of slices: serving a utility company, servicing remote control for a factory, serving a virtual operator, and optimizing for streaming video.[5] Operators will be able to provision devices through account configuration so the devices can access specific slices. For consumers, one slice might be for best-effort, unprioritized Web browsing while another slice could support prioritized telepresence that needs low latency and high bandwidth.

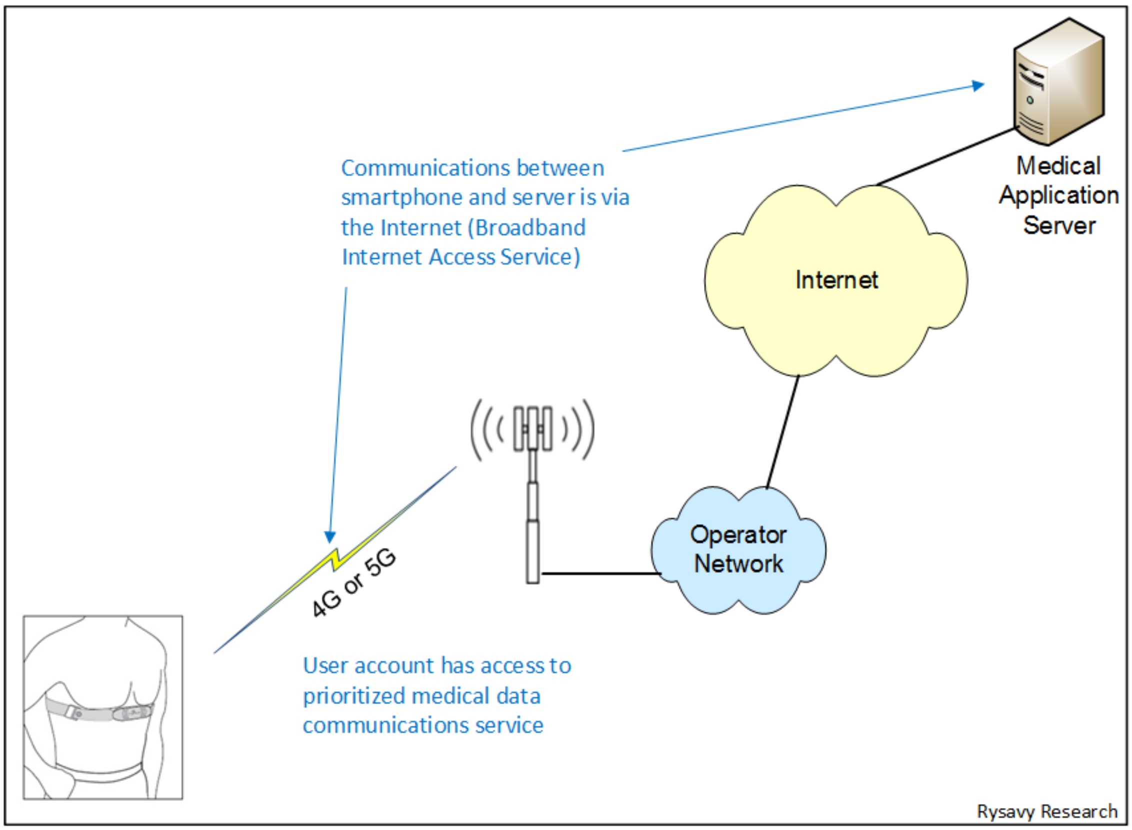
An example of an application using network slicing is a patient wearing body sensors that monitor a heart condition. The sensors continually report vital stats and GPS location to a medical application server, which in turn analyzes the data in real time, taking into account the patient's medications and medical

[5] 5G Americas, *Network Slicing for 5G Networks & Services*, November 2016. Available at: http://www.5gamericas.org/files/3214/7975/0104/5G_Americas_Network_Slicing_11.21_Final.pdf.

history. Proper analysis depends on receiving accurate data regularly and without interruption—even if the network is congested because a half-dozen people are streaming video while on the same bus as the patient. Because the connection is via the Internet, the communications service is considered Broadband Internet Access Service (BIAS), as defined by the FCC.

To support this application, the operator could provide a network slice customized for medical communications, perhaps one in which throughput rates are modest but the need for reliability and low-latency is high. In other words, the virtualized network that the medical application accesses via the network slice is optimized for the specific needs of health monitoring. In the event of a health event requiring treatment, detected either by the user's monitoring equipment or by the server using more sophisticated data analysis, the system advises the user to see a doctor. If the situation is critical, the monitoring equipment or server could summon an ambulance.

Figure 5: Medical Monitoring Example



Even with access to new spectrum and expected peak throughputs that will exceed 1 Gbps, 5G networks will be required to manage latency, reliability, massive numbers of connections, and a mix of stationary and mobile users. Fundamental to this task will be managing QoS. Different slices will have different QoS requirements, inherently invoking traffic management within each slice. As the 5G Americas paper states, “Each slice is defined to meet different service/application requirements, which are represented in a certain QoS level. A QoS level can be defined by certain performance descriptors such as delay, jitter, packet loss and throughput.” In addition, the amount of radio and network resources to apply to each slice, determined based on demand across all slices, will require QoS management.

Conclusion

The communications requirements of today’s mobile network applications span a huge range. One application may need high throughput but can tolerate significant delay. Another may need to send only a small number of bits, but these must traverse the network with minimal delay. Future Internet of Things innovations, from intelligent highways to smart-grid monitoring, will only increase the rich variety of application diversity. QoS mechanisms in 4G, and those under development for 5G, provide for application developers and operators to specify needs and for the network to dynamically accommodate them.

Developers and operators have a financial stake in enabling a diverse range of services, and empowering them to optimize networks to meet competing needs will result in the highest possible quality of experience for the largest number of users. The business case for massive 5G investment can only be made by being able to support all potential applications. Current simplistic views of network neutrality are blind to the fact that different types of applications have different network requirements.

The U.S. wireless industry is at a critical juncture. The United States has assumed global leadership in 4G and enjoys deep LTE penetration, leading smartphone platforms, and a vibrant application ecosystem. But globally, countries and companies are investing in and concentrating on what will come next with 5G. Constraining 5G with rules that unnecessarily undermine its potential is economic folly.

About Rysavy Research

Peter Rysavy is the president of Rysavy Research LLC, a consulting firm that has specialized in wireless technology since 1993. Projects include analysis of spectrum requirements for mobile broadband, reports on the evolution of wireless technology, evaluation of wireless technology capabilities, strategic consultations, system design, articles, courses and webcasts, network performance measurement, test reports, and acting as an expert in patent-litigation cases. Clients include more than ninety-five organizations.

Peter is a broadly published expert on the capabilities and evolution of wireless technology. He has written more than 160 articles, reports, and white papers, and has taught more than forty public wireless courses and webcasts. He has also performed technical evaluations of many wireless technologies, including cellular-data services, municipal/mesh Wi-Fi networks, Wi-Fi hotspot networks, mobile browser technologies, wireless e-mail systems, and social networking applications.

From 1988 to 1993, Peter was vice-president of engineering and technology at Traveling Software (later renamed LapLink); projects included LapLink, LapLink Wireless, and connectivity solutions for a wide variety of mobile platforms. Prior to Traveling Software, he spent seven years at Fluke Corporation, where he worked on data-acquisition products and touch-screen technology.

From 2000 to 2016, Peter was also the executive director of the Wireless Technology Association, an industry organization that evaluates wireless technologies, investigates mobile communications architectures, and promotes wireless-data interoperability. Peter Rysavy graduated with BSEE and MSEE degrees from Stanford University in 1979. More information is available at http://www.rysavy.com.