

## Developing a Framework for Future Mobile Data Pricing

DOI: 10.12776/QIP.V21I2.759

Michael Paetsch, Peter Dorčák, František Pollák, Ľubomír Štrba,  
Branislav Kršák

Received: 03 March 2017

Accepted: 12 March 2017

Published: 31 July 2017

### ABSTRACT

**Purpose:** The revenues for mobile data transmission overtook the revenue of voice calls for the first time in 2014 in the USA. It can be observed that demand for mobile data – largely driven by video and cloud - is increasing exponentially, while overall data revenue is rising only moderately. This will lead to insufficient revenues stream to increase investments into mobile networks and ensure quality service. Consequently, hereof network performance will deteriorate sharply. At the heart of the problem is the current global pricing regime of fixed multiple MB/GB bundles, irrespective of time of the day, intensity of usage (e.g. video vs. email) and underlying economic value of the data. A new framework is proposed as to optimize and align network capacity and implicit data value/utility, which is crucial to ensure customer satisfaction and access justice.

**Methodology/Approach:** The fundamental differences in pricing voice and data in voice and/or data centric networks are analysed in detail. Information has been synthesized as to develop insights into the impact of different devices and type of digital traffic for the overall performance of mobile networks. Based hereupon, a new framework for mobile data has been proposed to address the increasing misalignment between network capacity, usage and underlying data value/utility. Initial solutions have been proposed and discussed.

**Findings:** While voice calls are easily quantifiable and are largely predictable in its occurrence and network load implications, mobile data traffic shows very large variations depending on type of traffic. While social media messaging by many customers consumes very little capacity, consumption of video streaming by relatively few customers can lead already to network saturation.

**Research Limitation/implication:** Carriers set prices for a fixed amount of data – irrespective of intensity and time of data traffic - which leads to sharp spiky type of traffic patterns essentially signalling sharp overuse during busy hours coexist with large period of underused times.

**Originality/Value of paper:** A new framework for proposition building and particularly pricing of mobile data services is provided.

**Category:** Research paper

**Keywords:** mobile cellular networks; spectrum; pricing; utilization of mobile networks; economic value

## 1 INTRODUCTION

Since 2014 US mobile carriers will earn more revenues with mobile data than with mobile voice services. Since voice has been the key revenue driver for mobile carriers during the last 15 years, little price sophistication can be observed as of today, in the data space. In fact, most carriers use data as an add-on to voice. However, given the fact that the industry will eventually evolve towards 100% mobile data, a new framework for mobile data is required.

There are numerous complexities involved in developing such a new framework. There is a huge discrepancy between cost of providing mobile data capacity and revenue generation potential, if compared to voice services. As will be shown this leads to vast variations in pricing data.

The paper will systematically explain the differences between the voice and data centric world and systematically develops a new framework for pricing data services in order to increase data revenues to mobile carriers. This paper starts out by providing a basic understanding of the cost of building and operating networks. Based hereupon, differences between pricing mobile voice and mobile data as well as pricing fixed data and mobile data services are explained in detail. Following, mobile data services are segmented and the challenges between network capex, quality of service and value to the customer are explored. Finally, a new framework for proposition building and particular pricing of mobile data services is provided.

## 2 FUNDAMENTAL CONCEPT OF CELLULAR NETWORKS

### 2.1 The Cellular Radio Concept

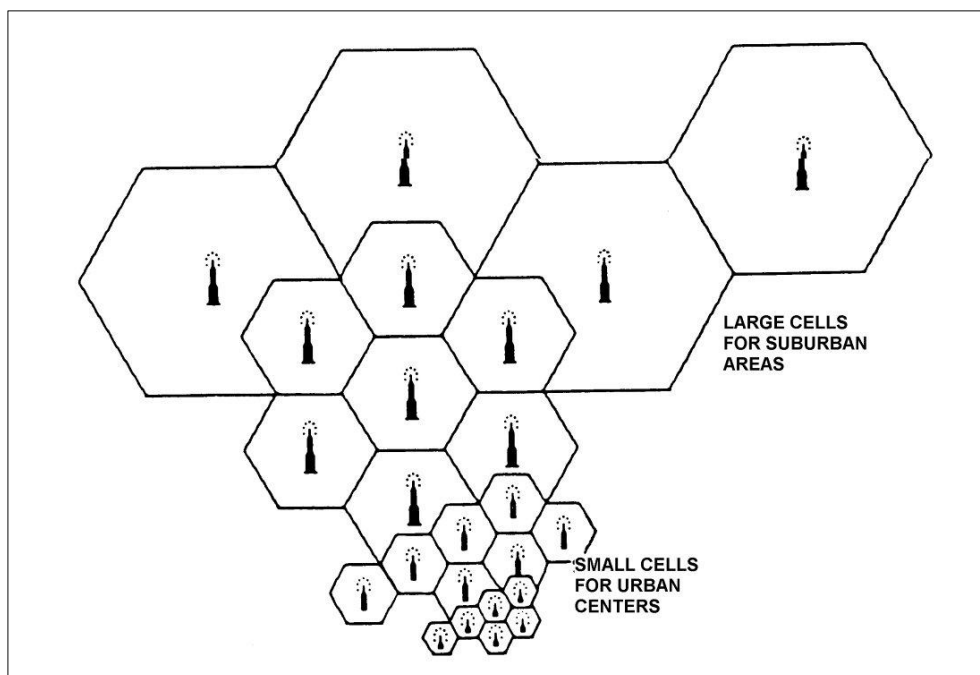
The concept of cellular mobile telephony essentially evolved around the need to make more efficient use of the radio spectrum in order to accommodate a rapidly growing number of users for mobile services. In contrast to mobile radio systems, mobile telephone services differ as far as they are operating in full duplex- mode and are always connected to the public switched network.

The first commercial mobile telephone service in the United States was established in 1946 in St. Louis, Missouri (Steinbock, 2005). These so called pre-cellular mobile telephone systems used a wide-area architecture: one very powerful transmitter provided coverage of 40-50 miles around the base station,

which was normally located in the centre of a metropolitan area. However, this system design had great limitation in terms of capacity, since only very few channels were available to serve the mobile communications needs of an entire city. In 1976, for example, Bell Mobile provided 12 channels for the entire metropolitan area of New York (more than 20 million people) to serve 543 customers, with an additional 3700 on the waiting list (Lee, 1976). In 1947, engineers at Bell Laboratories devised the concept of cellular radio (Calhoun, 1988). Rather than setting up a high-power transmitter to cover a large area (40-50 miles in diameter) the cellular concept proposed to divide the coverage zone into smaller sections, or “cells,” each served by a low-power transmitter. By reducing the coverage areas and establishing a large number of small cells, it became possible to reuse the frequency and thus substantially enhance the capacity of the system.

The fact that the cell radius is determined by the transmitted power, and is therefore variable, enables the systems engineer to continuously decrease the cell size, and consequently, increase the system capacity. Specifically, a reduction in cell radius by 50% leads to a quadrupling of the number of circuits per MHz.

Hence, there are two fundamental ways to increase the overall capacity of a cellular system: first splitting cells into ever smaller sub cells as depicted in Fig. 1 and second adding channels, which in turn requires more spectrum.



*Figure 1 – Cell Splitting (Calhoun, 1988)*

It is important to note that cell splitting is associated with high capital cost. Each cell site requires the purchase of a base station and requires costly connection to a

mobile switch by means of fixed leased line or microwaves. In addition, cell sites require maintenance and above all monthly rent should be paid for the site. Hence, spectrum allocated to operate mobile networks play a central role in the overall economics of a mobile system, as will be further outlined below.

## **2.2 Spectrum and Spectrum Management**

All radio communication systems make use of the electromagnetic spectrum. This spectrum is described as “a form of oscillating electrical and magnetic energy capable of traversing space without the benefits of physical interconnection” (Levin, 1971). Although it appears in theory that the radio spectrum is infinite, there are technical and natural reasons that greatly limit the amount of electromagnetic spectrum suitable for mobile communications systems (Seybold, 2005). Hence, radio spectrum is a very precious, highly contested “limited natural resource” and must therefore be utilized as efficiently as possible. The history of spectrum management bears some resemblance to the development of the automobile: At the beginning of the 20<sup>th</sup> century, only a few cars used the then-available streets, which in turn meant that a driver could practically use the entire road.

With the growing popularity of cars, it soon became obvious that traffic rules would have to be introduced to avoid accidents, as well as to increase traffic flow (serious impairment of movement). In a similar way, everybody could make use of the US spectrum, prior to 1927, by simply writing a letter to the State Department of his or her attention to utilize a certain part of the radio spectrum. Over time, however, it became clear that the radio spectrum can suffer from congestion (i.e. interference and service degradation) if too many systems transmit and receive information on it in an uncoordinated way. Therefore, radio spectrum had to be systematically organized on national level, and – since electromagnetic waves in the low frequency bands do not stop at national levels – also internationally.

Developments in mobile communications systems are largely driven by the availability and/or scarcity of spectrum. Since radio waves with different frequencies have very dissimilar propagation characteristics, certain parts of the radio spectrum are far more suitable for land-mobile-communications systems than others (Seybold, 2005). With large parts of the highly usable radio spectrum virtually completely allocated, the growing demand of both existing and new mobile communications services can *ceteris paribus* only be accommodated by replacing current systems with more spectrum efficient ones, or by the development and introduction of technologies that will allow certain mobile communications systems to economically operate in higher frequency bands that are less crowded.

### 2.3 Advances in Cellular Technology

It is important to understand that cellular mobile communications were initially seen as an add-on, and very elite service to the existing fixed switched network. The first US analog commercial networks based on the AMPS standard (so called 1G), grew from less than 100.000 subscribers in 1983 to more than 10 million subscribers in 1992. As depicted in Fig. 2, the growth rate has been indeed exponentially. It became obvious that an analog system could not cope with this kind of market development.

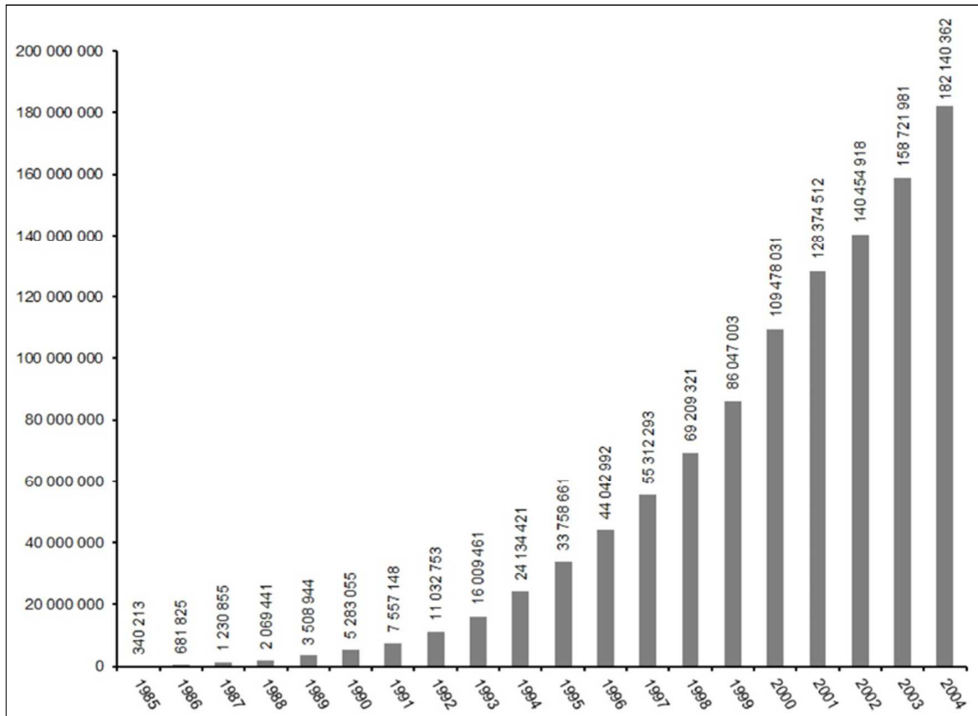


Figure 2 – Year End 2004 Estimated Wireless Subscribers (CTIA, 2005)

This marked the advent of two important developments: first, it became clear that the network technology has to evolve from analog to digital. The work for the first digital network (GSM or 2G) started in the mid-80s and was launched commercially in the 900 MHz band in 1991 (Pelkmans, 2001). The second important impulse stemmed from the insight that instead of a luxury product for a few, mobile communication will become a mass product eventually superseding fixed telephony. This idea was captured in the term “Personal Communication System (PCS)” or “Personal Communication Network (PCN)” (Paetsch, 1993). In the mid-90s, it became apparent that the data rates offered, were too low for fledging data market. Work on the so called UMTS or 3G network started, and the first 3G network was launched in 2001. This network is designed to accommodate data and voice traffic equally.

However, by the turn of the century it became clear that the unbelievable growth in the internet will eventually have to be supported by mobile networks and work on LTE, respectively 4G networks, started (White Paper, 2011). LTE is clearly a network designed to efficiently handle mobile data only (White Paper, 2011). Voice and Data traffic are all IP in a 4G network.

Fig. 3 summarizes the development of mobile communication: almost exactly every 10 years a new cellular technology was introduced into the market, operating on higher frequencies spectrum and featuring ever smaller cells size radiuses. Hence, the amount of base stations required to provide coverage is increasing sharply. This problem gets further aggravated by the fact that 3G and 4G cells are shrinking as the “load” i.e. capacity of a cell gets filled with traffic (Portilla-Figueras, et al., 2009).

1G (1981/82)	2G (1991)	3G (2001)	4G (2011)
<b>Analog Network:</b> Voice of Few subscriber	<b>Digital Network:</b> Voice of Mass Market	<b>Digital Network:</b> Voice & Data Market	<b>Analog Network:</b> Mass Data Market
400-800 MHZ Spectrum	900-1900 MHZ Spectrum	1800-2100 MHZ Spectrum	1900-2800 MHZ Spectrum
Cell Size (radius) up to 15-20 km	Cell Size (radius) up to 8 km	Cell Size (radius) up to 2 km	Cell Size (radius) up to 1.4 km
Data Speed: 0	Data Speed: <0.1 Mbps	Data Speed: <1 Mbps	Data Speed: 3-6 Mbps

*Figure 3 – The Development of Mobile Communication (Paetsch, 1993)*

Especially is these areas substantial differences exist between the voice centric 2G and the data centric 3G/4G network economics. These differences will be discussed further below. The following Fig. 4 shows how the global subscriber base is currently connected towards the different network technologies.

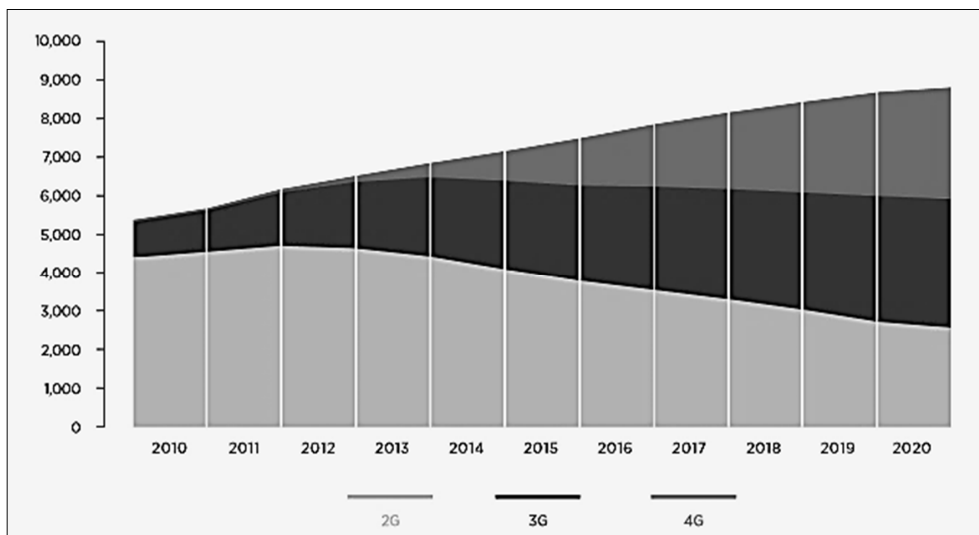


Figure 4 – Global Subscriber Connections in Mio. – Excluding M2M (GSMA, 2016)

### 3 EVOLUTION OF USAGE PATTERNS

As pointed out above, the first twenty years of mobile communications were clearly dominated by providing voice traffic. Fig. 5 shows how mobile data will surpass mobile voice globally. According to GSMA report (GSMA, 2013), the mobile operator data revenues will overtake voice revenues globally by 2018 as we move towards a fully connected world.

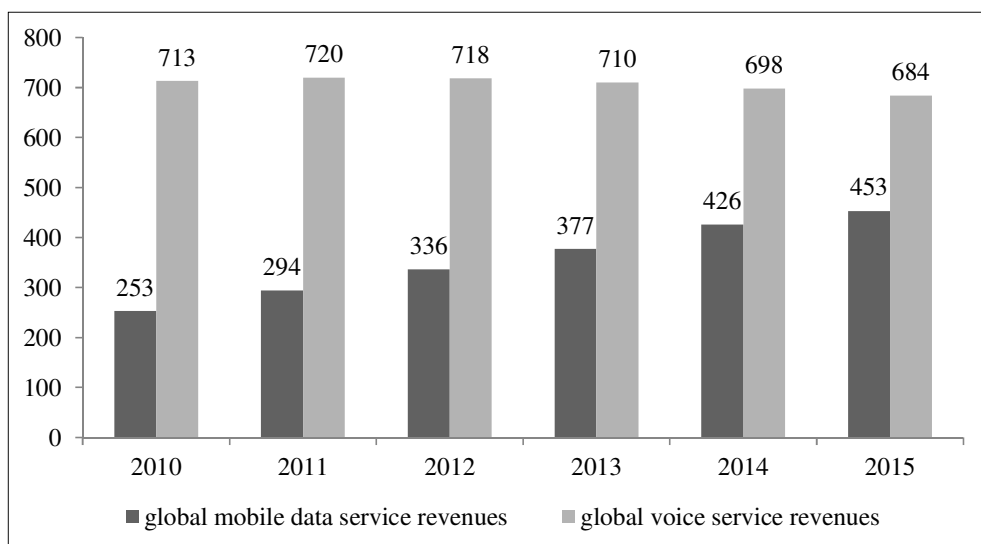
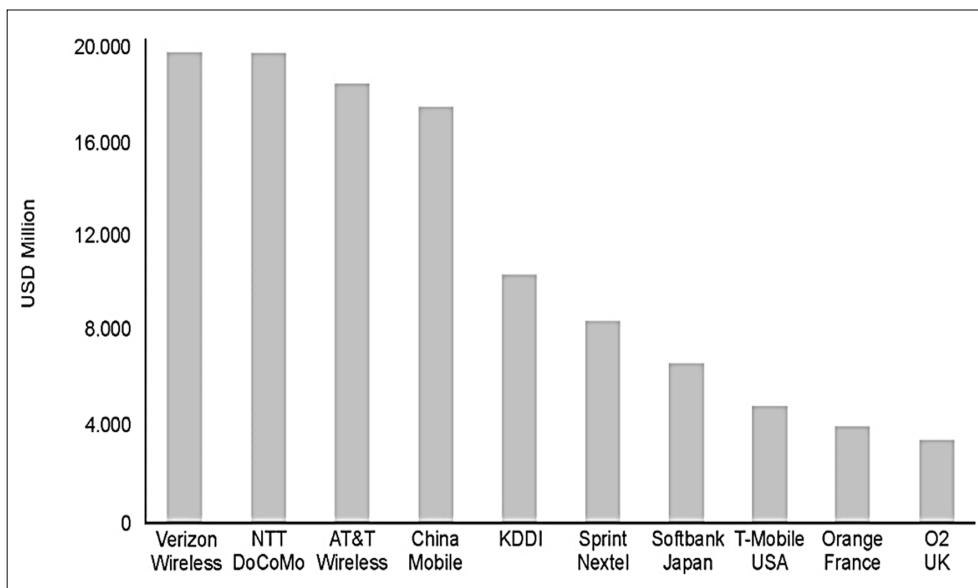


Figure 5 – Voice vs. Data Revenue 2010-2015 in Billion U.S. Dollars (Statista, 2017a; 2017b)

There are two three major factors driving the consumption of data usage, leading to a situation in which data ARPU (average revenue per customer) exceeds voice ARPU: reduction in voice related pricing, the strong rise of “on-the go computing” devices and migration of a largely fixed internet business model to an “anytime-anywhere” internet model



*Figure 6 – Data Revenues in Billion U.S. Dollars as of Q4/11 (Sharma, 2012a)  
(Note: Verizon & ATT are US Telcos, NTT and KDDI and Softbank are Japanese Carriers)*

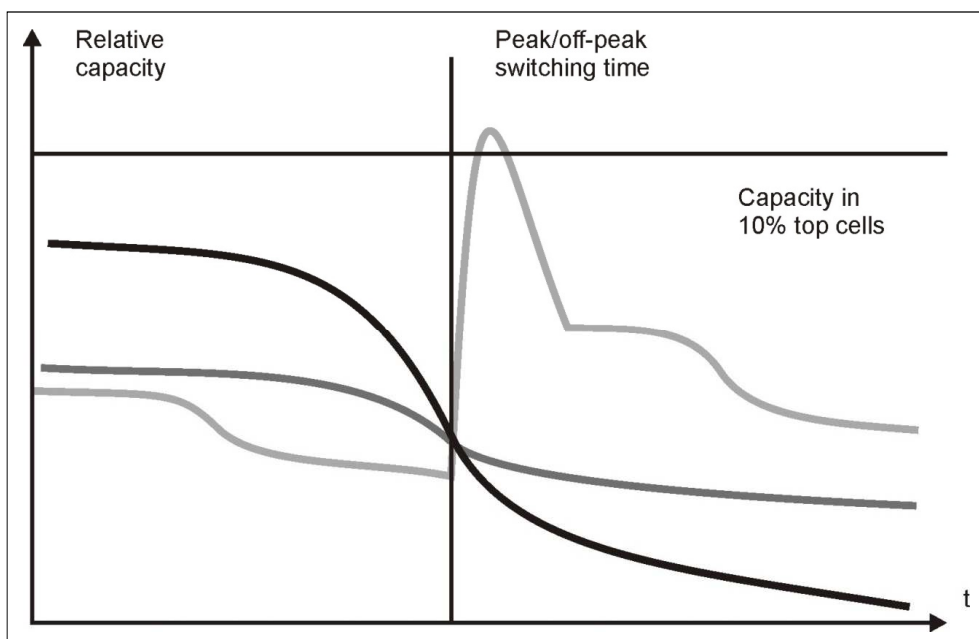
In the US, this inflection point – i.e. the crossover of mobile data exceeding mobile voice revenues – occurred in 2013. Aside from Japan, the US is the leading nation with regard to overall voice revenues as shown in Fig. 6.

### 3.1 The Pricing of Mobile Voice

During the early stages of mobile cellular networks, pricing of voice services was straight forward but expensive. Most mobile networks adopted a peak/off-peak pricing scheme. This provided an adequate pricing mechanism, given the fact that the networks especially in the metropolitan areas had only a limited number of channels available. As early adopters were largely professionals with very little price sensitivity, pricing during the day was extremely high, respectively low during the off-peak time. Switching between off-peak/and peak occurred around 6 pm. The peak pricing would reflect the given capacity at the time of network rollout. With the beginning of mobile personal communication, i.e. the evolution of a mass market in mobile communication, the situation changed fundamentally. The handset prices fell dramatically and as a result, diffusion of mobile phones increased substantially. However, these new added customers had a very different usage pattern. Since these early nonprofessional consumers were



much more price sensitive, the traffic pattern changed fundamentally. As described in Fig. 7 indicated by the black line, early network traffic was clearly reaching the highest level during peak hours and slowed down considerable during off-peak hours. As more consumers were added to the network and at the same time capacity of the network was enhanced, the curve developed into a pattern reflected by the dark grey line in Fig. 7. The higher price sensitivity of these consumers coupled with the high difference between peak and off peak traffic (up to factor 5), lead to a traffic abnormality. As described by the grey line in Fig. 7, consumers virtually sharply decreased mobile phone traffic prior to the peak/off-peak border and “postponed” calls to the past peak time. This resulted in an extremely high traffic peak right after the switching time, often exceeding the capacity levels and therefore negatively influencing customer satisfaction (Khanfir and Fitkov-Norris, 2000; Yaipairoj and Harmantzis, 2004).

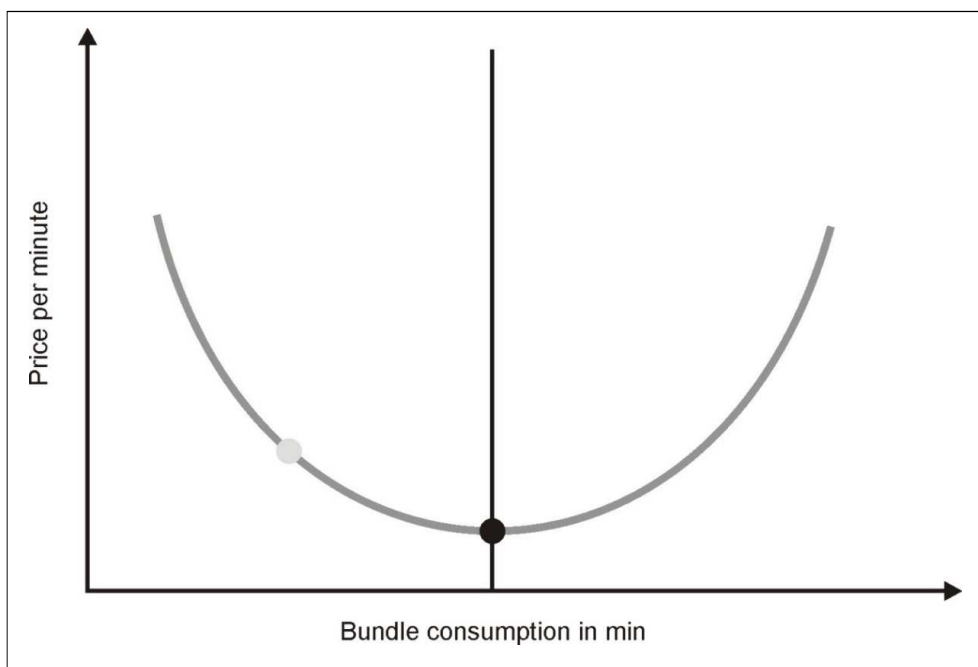


*Figure 7 – Systematic Overview: Network Load in peak/off-peak Pricing Scenario over Time*

This in turn forced the mobile operators increasingly to expand capacity to meet off-peak traffic. As a consequence many ways of pricing network capacity in order to maximize revenues were discussed (Basar and Srikant, 2002; Fitkov-Norris, 2003; Baochun, Nahrstedt and Xue, 2006). In most European countries, however, most mobile operators adopted a non-dynamic pricing scheme which is far easier to communicate to consumers, known as airtime bundles. A typical airtime bundle comprises a given number of minutes for a fixed price. A larger airtime bundle would offer a lower minute price than a lower airtime bundle. The consumers could use these airtime bundles around the clock, the concept of

peak/off-peak disappeared. This tariff structure represented a fundamental shift toward a much more pro-usage scenario (i.e. personal communication) and marked the beginning of the direct attack to substitute rather than amend the fixed network (Vogelsang, 2010). The bundling airtime pricing had furthermore the advantage to appear much more attractive to consumers than it really was. This is described in Fig. 8. If a consumer purchases a certain airtime bundle B (e.g., 50 minutes) for a certain price P (e.g., 20 EUR), he or she will focus on a certain price per minute (PpM).

In fact, the consumer will automatically calculate  $P/B$  thereby inevitable calculating the lowest possible average price per minute possible (0.40 EUR), market in the Fig. 8 a yellow dot (0.40 EUR per one minute). Interestingly, it is practically impossible for the consumer to reach exactly this point in the course of one month. If for example, the consumer was on vacation and used only 50% of his airtime in a given month, the true cost for the average price of a used minute is increasing sharply (e.g., 0.80 EUR). Likewise, prices for minutes exceeding the bundle are substantially higher, thereby also driving up the average cost per used minute substantially. While the consumer liked the easiness and transparency of the tariff, he systematically misjudged the true PpM cost.



*Figure 8 – Systematic Overview: Average Airtime Price per Minute in an Airtime Bundling Pricing Concept*

In light of the ever larger built-up of network capacity of 2G/3G, the pricing schemes eventually changed to the most aggressive form: flat pricing for fixed network and eventually fixed and mobile telephony. This pricing scheme was

adopted first in 2007 in the USA and is experiencing widespread adoption. As Fig. 9 indicates this flat pricing scheme did not only increase the used minutes per user and progressively lowered the cost per consumed minute but also lead to a rising number of consumers that terminated their fixed access completely (Rodini, Ward and Woroch, 2003). This effect is known as fixed-to-mobile substitution and is an accelerating trend in advanced mobile markets with a dense fixed network infrastructure such as Europe and the USA (Banerjee and Ros, 2004; Albon, 2006).

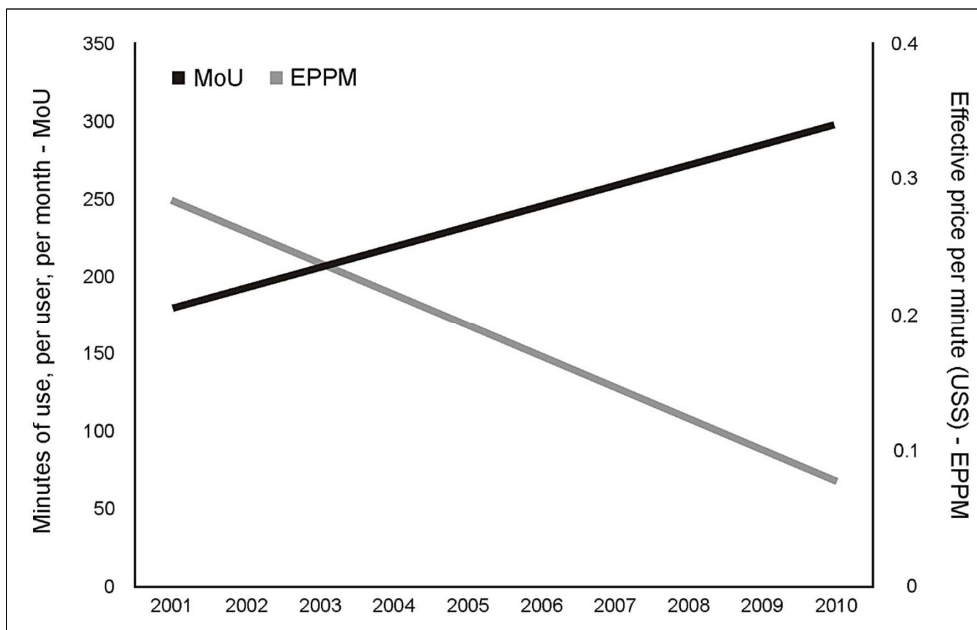


Figure 9 – Development of Use and Effective Pricing of Mobile Switched Minutes 2001-2010 Globally (Note: MoU user per month, EPPM Worldwide effective price per minute) (Wireless Intelligence, 2012)

### 3.2 On the Go Computing

Aside from the sharply falling effective prices for mobile minutes, the appearance of on-the go computing devices such as smartphones and tablets greatly influences the change in usage pattern away from voice towards data (Specifically, the launch of the first iPhone allowed for the first-time mobile web browsing on an acceptable level).

For years, the existing mobile phone technology and form & design function proved unsuccessful of accelerating mobile internet usage. As Fig. 10 demonstrates, this only changed with launch of the iPhone. The iPhone had a market share of only 11% in 2009 but was responsible for over 65% of all web request in mobile networks. The world market leader (Symbian) on the other hand had 49% installed base in the networks, but only 7% of the total mobile web sites were requested by this phone. The on-the-go computing devices will

transform the IT ecosystem only comparable with the transformation from mainframe to PC. As indicated in Fig. 10, the growth in IT hardware is clearly moving away from traditional desktop and laptop devices towards smartphones and tablets.

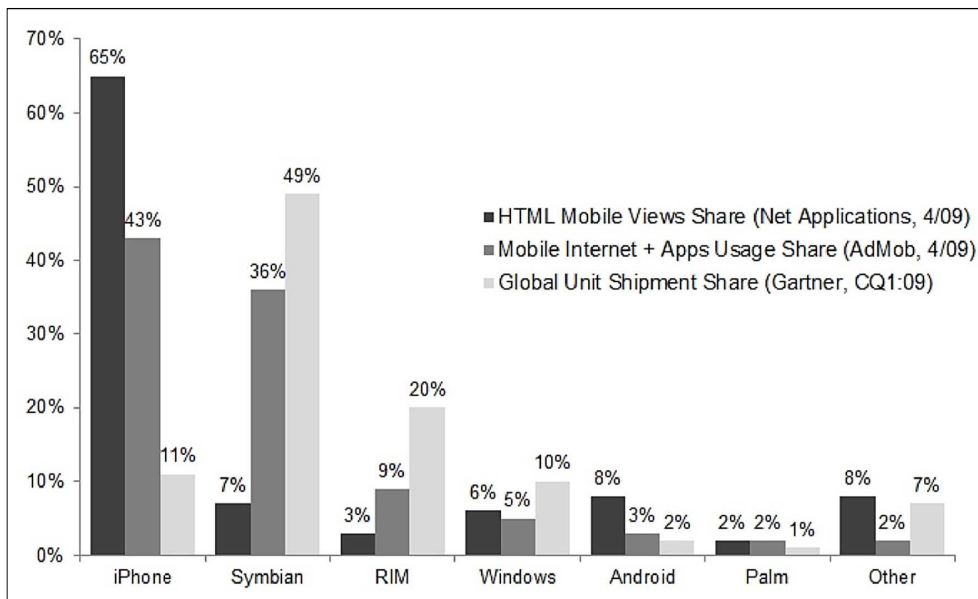


Figure 10 – Web Usage vs. Operating System (Gartner Group, 2009)

In contrast to traditional IT hardware, are on-the-go computer devices not only highly mobile and a have long battery life, but can be switched on or off in seconds.

### 3.3 Mobile Internet

Especially the trend toward on-the-go computing described above, lead to an explosion of network data traffic. In fact, it can be argued that the desire for customers to access the internet has been already building up steadily, but was effectively suppressed by the non-availability of adequate hardware. Hence, a de facto market vacuum was building up, which in turn led the iPhone to become the device with the fastest ever recorded diffusion speed. This becomes evident when studying Fig. 11. While 3G network were launched already in 2001, only a very slow data growth could be observed until 2007 (Phase A). As pointed out in Fig. 12, available equipment in Phase A did not entice consumers to make use of the mobile internet. This changed fundamentally with the introduction of fully screen phone with touch sensitive screens.

Another invention fuelling the use of mobile data were the so-called Apps. Apps are small native applications, which are easy to download and use on a smart phone. Around these apps, a whole new ecosystem comprising millions of applications developed in only a few years. These apps are perfectly suited for operating in a mobile environment: since these small program is residing on the smart phone itself, only data for refreshing information had to be transported, providing consumers with a very good user experience (Kessler, 2012).

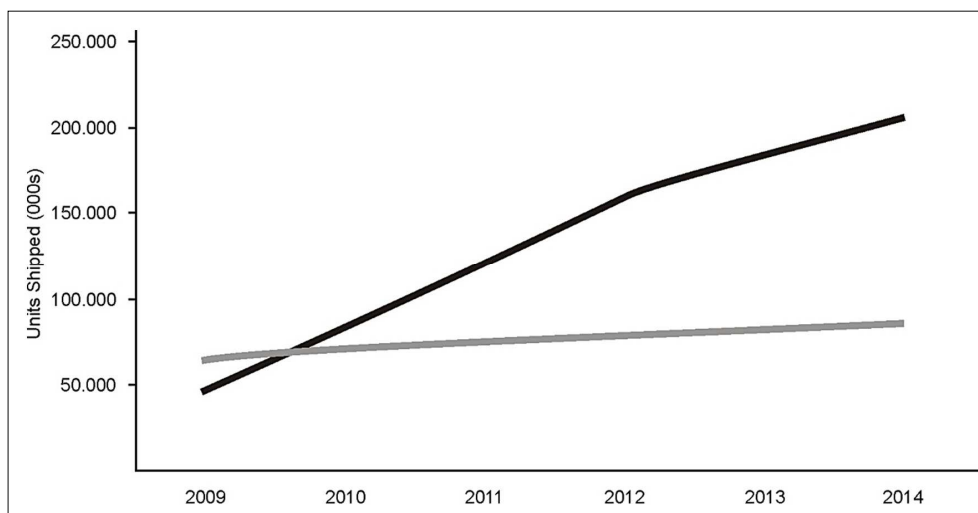


Figure 11 – Development of Desktop/Laptop Shipments (grey line) vs. Smartphones/Tablets (black line) 2009 to 2014 (Sharma, 2012b)

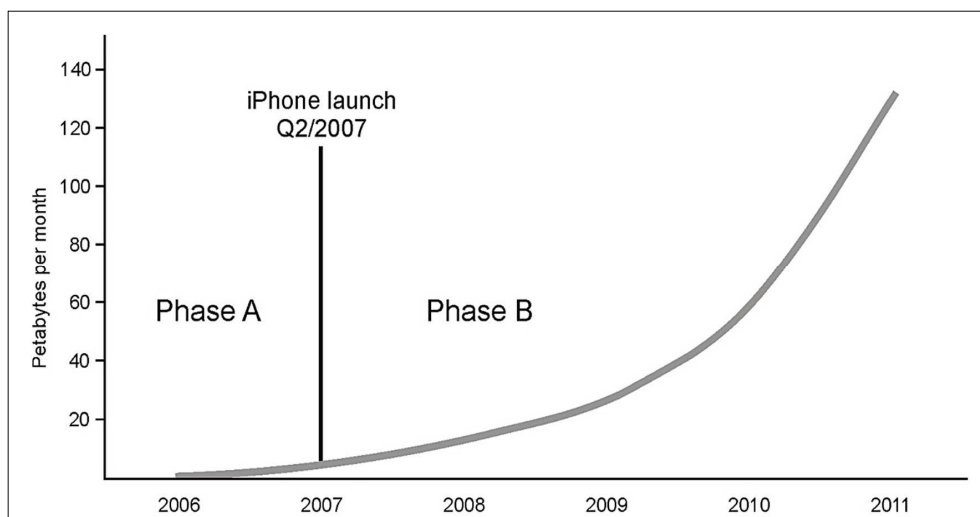


Figure 12 – US Mobile Data Growth 2006-2011 (Sharma, 2012c)

In summary, it becomes clear that the usage pattern of mobile communications services is shifting dramatically towards mobile data away from mobile voice. While mobile data has so far been viewed, and priced as an appendix to voice, this development will require telcos to fundamentally review data pricing on its own. As outlined in the following chapter 4, pricing data is a highly complex undertaking requiring a new approach and a new framework.

### 3.4 Data Usage and Data Pricing and Data Revenue

The above described massive shift from voice centric mobile phones to smartphones is accompanied by a parallel trend towards digital media. While email and chat clients do not consume a lot of network capacity the opposite is true for data intensive video streaming. Large part of the network capacity is used over relatively long periods of time. The strong trend towards usage intensive applications such as video and browsing continuous has reached almost 75% of total data traffic in certain European networks (GSMA, 2014). Consequently, the demand for data is growing rapidly and will reach 22GB per mobile subscriber in the USA and 12 GB in Europe in the USA (GSMA, 2016).

As can be seen from the following Fig. 13, there is fundamental mismatch between the accelerating demand for data on one hand and the monetization potential or revenue on the other hand.

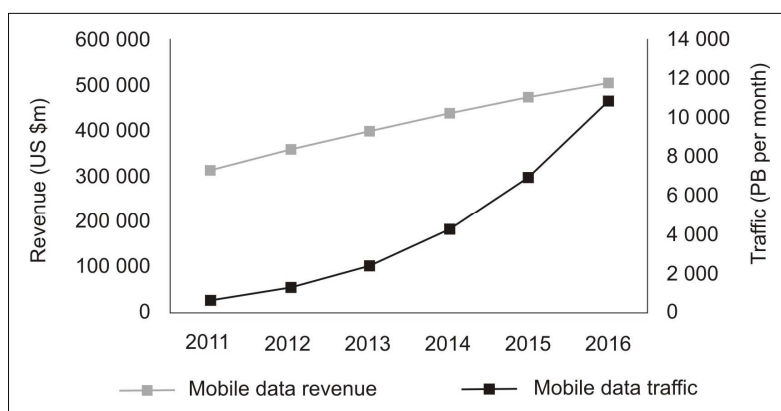


Figure 13 – Development of Mobile Data Revenue and Mobile Data Traffic (Ernst & Young, 2013)

It becomes obvious that there is a fundamental mismatch between data consumption and data revenue, which will lead to a situation in which telco carriers cannot invest sufficiently into their networks to sustain the strongly rising traffic. Under the current data pricing regime and in the absence of any new intelligent data pricing strategy this will lead to highly congested networks very similar to rush hour traffic in metropolitans. Overall, this appears not a viable long term strategy, given the importance of mobile data networks for the economy especially when moving into a cloud centric macro IT architecture.

## 4 A NEW FRAMEWORK FOR DATA PRICING

### 4.1 Key Differences between Voice and Data Pricing

In order to understand the principle differences of voice and data pricing, above outlined characteristics of 2G and 3G/4G networks play a crucial role. Below, the key differences of voice and data centric networks are grouped and subsequently outlined.

*Cell Scale vs. capex in a voice centric vs. data centric network:* The cell size of a 2 G can be more than 20 km. Importantly, the cell size is fixed and therefore the capacity in a given cell accurately predictable. Only if capacity demand is increasing and/or in-house coverage should be improved, cell splitting is performed. This means that capital outflow is - apart from the initial investment – very much in sync with rising number of customers and usage.

In data, centric network, this is fundamentally different. As pointed out, 3G cells do not have a fixed cell size. Rather the radius of the cell is shrinking when the cell gets loaded with traffic, thereby making it necessary to add cells for coverage under load. Another substantial difference is the cell radius itself. While 2G cells can be around 20 Km 3G/4G cells have a maximum cell size of only 1 to 1.5 km. This means, however, that the “capex-characteristic” differs in the sense that much more cells along with back link facilities, rent, maintenance, etc. have to be installed right in the beginning. Moreover, given the much smaller cell radius, the cell does not provide capacity scale to a large area, but rather a very small area. So, it is perfectly possible that one 3G/4G cell is fully loaded leading to blocking, while the other one is practically empty.

*One device provisioning in voice networks vs. multi-device provisioning in data networks:* In the previous decades, mobile communications centred essentially around the provision of voice services. Hence, voice represents an on-off type of service, occupying always exactly the same load/capacity from a cell. This is again fundamentally different in a data centric network for a variety of reasons. First, as outlined below in Fig. 14, data networks are accessed by a large variety of hardware ranging from smartphones, tablets to laptops, etc. The consumption of bandwidth for each device varies greatly if averaged over the month.

This has several important implications: a voice centric network handled only voice calls via one device, allowing very accurate prediction of traffic load. The capacity load of data network can vary greatly, depending on the observed mix of devices in the cell at a given time.

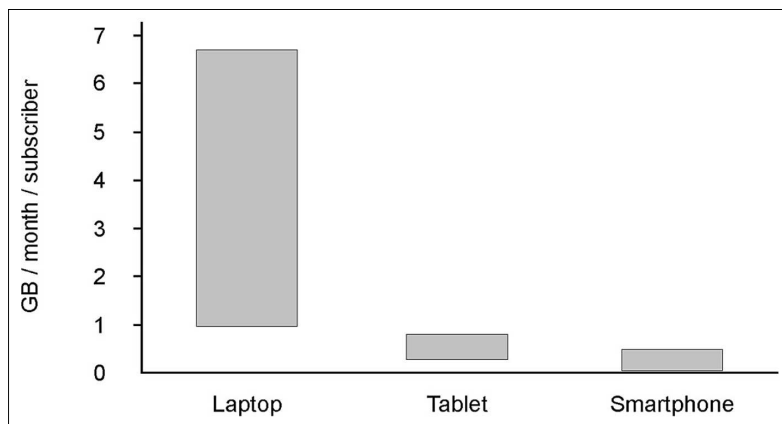


Figure 14 – Monthly Traffic Volumes in 3G by Type of Equipment (Malik, 2012)

One service provisioning in voice networks vs. multi-service provisioning in data networks: aside from the complexity regarding multiple equipment with different capacity demands, services provided in voice and data network differ equally large. Historically, a voice network provides only one service and this is voice.

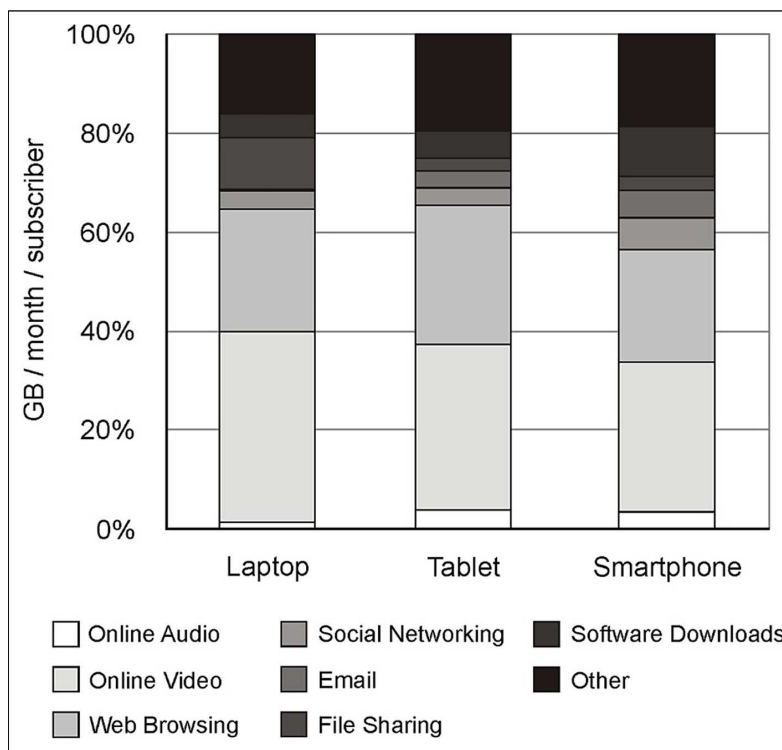


Figure 15 – Overview of Various Services Used by Different Equipment (Malik, 2012)



The capacity of one voice call is always identical and it therefore easy to quantify the capacity. This contrasts sharply in a data centric world. It is very important to cluster the different types of traffic (Fig. 15). Online services such as online audio and only video are called real-time services, for the application cannot tolerate transmission delays without the customer experiencing a sharp and instantly noticeable drop in quality or even usability.

Services like web browsing, social networking, email, file sharing and software downloads are so called non-real time services. Small delays or a certain deterioration in transmission speed increases the waiting time of the consumer but does not render the service useless. The by far fastest growing type of services in data networks are the real time streaming services for music and video (Youtube, Napster, Somify are only a few of well-known streaming platforms). However, it should be mentioned that streaming absorbs high rates of capacity measured in Mbps.

*Quality of service in a voice centric vs. data centric network:* measuring quality in a voice centric network is very easy and transparent. A customer can either receive/make a voice call or not. A situation in which a voice call cannot be completed is referred to as blocking. Quality is therefore easily measurable by the blocking rate of each cell at the busy hour of the day. In a data centric network, quality is much harder to define, measure and ensure. Given the fact that the cell size is much smaller, and the multitude of devices and real-time (RT) non-real time (NRT) traffic can vary widely, traffic throughput and especially service degradation can vary sharply.

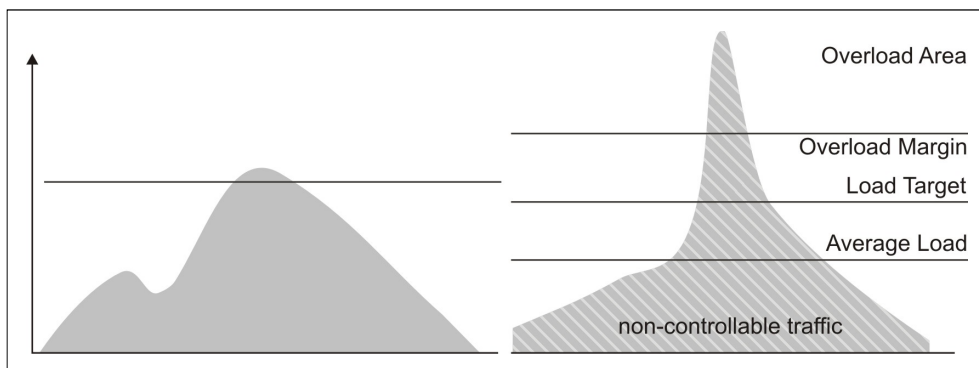


Figure 16 – Peak Shaped Traffic in Voice and Data Centric Networks

In summary, a voice network comprises larger cell sizes that are used by only one device producing one type of service, i.e. voice. The criteria for quality (i.e. blocking rate) is transparent and easy to measure. In sharp contrast, hereto, peaks in data centric networks are much more pronounced and statistically or mathematically far less predictable. First, the cell size is substantially smaller in coverage size and second different devices accessing different type of traffic require a widely fluctuating capacity at a given time. It is important to note that

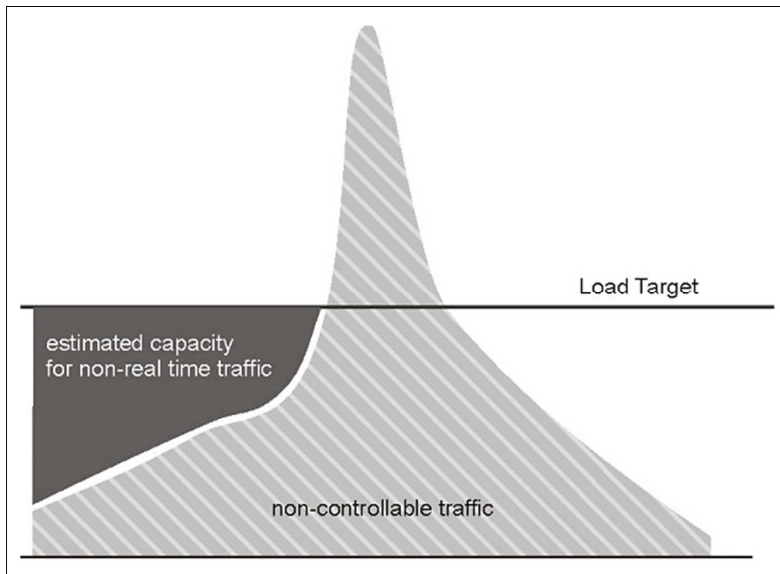
very steep narrow peaks represent a twofold problem. They either lead to blocking of real-time traffic which directly negatively impact customer satisfaction (Lee, 2010) (overload area Fig. 16), or they require very high capex to meet peak capacity needs, which, however, is utilized only for a very short time. This in turn provides an extreme low payback for a given investment.

## 4.2 Current Data Pricing Concepts

As demonstrated in Fig. 12, early mobile data uptake between 2001 and 2007 has been extremely low. It has been established that the key reason therefore was not so much the pricing of mobile data but the lack of suitable hardware. It has been further documented that the launch of 3G coincided with the trend towards pricing voice services flat. These two factors were instrumental for the fact that a variety of countries (e.g., US) offered unlimited usage or very high data bundles (5 GB) for a relative low price. Given the small size of 3G cells, the initial capacity of the network was quite high and usage very low. Hence, the underpricing of data had no capex effect. This changed dramatically in 2007.

With the smartphone becoming popular, growth in mobile data grew exponentially as shown in Fig. 12. Over time, another phenomenon became apparent: very few extreme heavy users consumed a large part of the bandwidth. As described above, the US is the leading mobile data market. Interestingly, it was the number one smartphone carrier AT&T, which experienced severe service problems in its metropolitan area first. Fig. 17 describes systematically the impact of blocking. Most users consider the non-real time services such as accessing the email or social media sites or general websites as much more important as streaming a YouTube video. As will be shown later, they are also willing to pay much more money per Mb for these services than video streaming. However, in a flat data or excess data bundled pricing scheme, consumers will make abundant use of streaming. Since streaming is real-time traffic, it will be prioritized for any delay leads to an almost total loss of quality.

After the initial capacity of a given 3G cell (load target) was reached, the cell has zero capacity for sending the non real-time traffic (email, etc. dark grey area in Fig. 17), which actually is more valuable to consumers. Eventually, with more consumers engaged in streaming, the cell would become overloaded providing no streaming satisfactory streaming service whatsoever. One remedy against overloading networks is the so called throttling policy: the top x% of users would be sharply reduced to 2G data speed for the following month, respectively their access speed was lowered when they reached the end of their bundles. Since this does not provide immediate network capacity relief, AT&T, for example, started to remove the heaviest mobile data network users under the flat pricing scheme. In fact, AT&T claims that 2% of its top user accounted for 60% of the traffic.



*Figure 17 – The Relationship between Real-Time (non-controllable) and non-Real Time Traffic and Load Target*

### 4.3 Towards a New Mobile Data Framework

It is clear that the throttling down policy cannot meet the conceptual challenge posed by network loading in the medium to long term future. Other mechanism must be found to ensure good network quality and access justice (Zhao, et al., 2012). A fundamentally new concept is needed to balance the capacity and device and type of service. The pricing should be balanced as to be acceptable to the consumer on the one hand and allowing the carriers to recoup its investment on the other hand.

The logical evolution of such a pricing framework is described below and must comprise static and dynamic pricing elements:

*Tariff structure and device segmentation:* clearly mobile data pricing has to become a function of the characteristics of the device with clear distinction between smartphone/tablets and laptops.

*Ensuring high quality basic service:* In order to ensure good quality service for smartphone users, carriers must focus on guaranteeing high quality non-real-time services (dark grey area in Fig. 17). Hence, a new pricing framework must clearly perform price segmentation between RT and NRT traffic.

*Regain control over network load by reducing data bundle size greatly:* As a consequence on these two demands, carriers will have to reduce bundle size sharply to the average usage of smartphones (Zheng, et al., 2015). A smartphone uses about 350 MB per month, while a tablet requires about 1 GB per month. Furthermore, they will have to raise prices. Bundle size must be sufficient to

allow for only very occasional streaming. Even the largest bundle size should not facilitate unrestricted use of streaming. As shown in Fig. 18, the countries such as Japan and the USA feature already much higher data prices to get network capacity in line with demand.

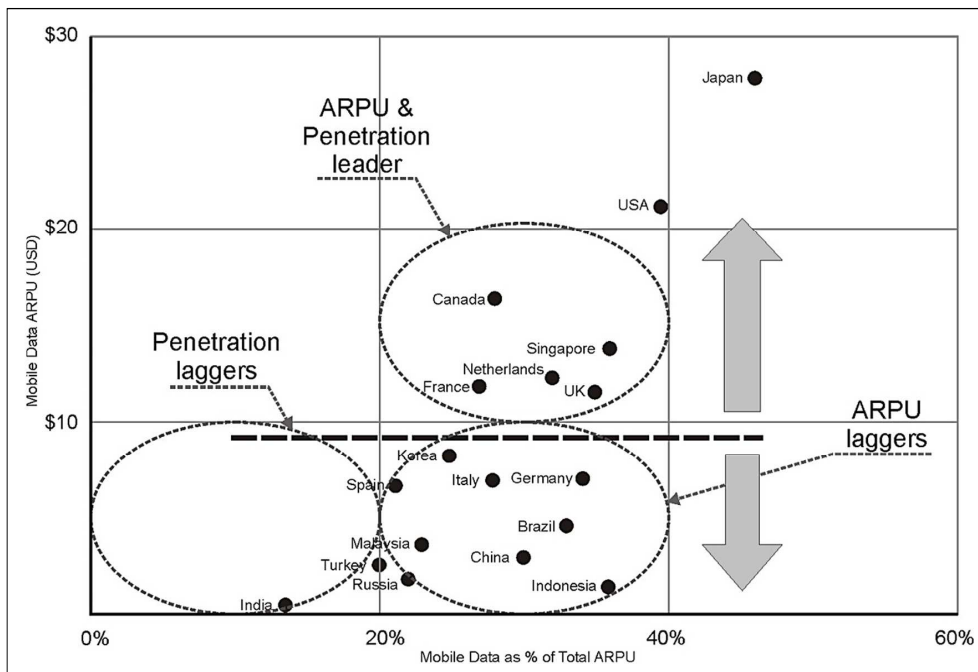


Figure 18 – Mobile Data ARPU – Absolute and as % of Total ARPU (Sharma, 2012a)

After the initial network load was used (first investment required to achieve 3G coverage), it became clear that the true cost of accommodate mobile data traffic is indeed very high. AT&T and Verizon both introduced small bundles with only 200 MB for smart phones for 15-20\$ and a second bundle capped at 2GB. Both carriers stopped selling unlimited packages.

*Static versus dynamic pricing elements:* Above mentioned small mobile data bundles for smartphone and tablets represent static pricing elements. Since the bundle sizes are small and the traffic characteristic is non-real time centric, a high degree of quality of service and predictability can be achieved.

*Dynamic pricing elements:* Given what has been established above, it is clear that the load/remaining capacity of a cell can vary hugely according to time are coverage area. While very “light” streaming is facilitated by the data bundles (static pricing element), extreme heavy streaming is prevented. Key streaming applications such as watching a movie, use of Skype or You tube can be controlled by allowing consumers to purchase vertical packages. These vertical mobile packages are priced according to actual or predicted traffic load in the coverage area at the price of purchase (low/medium/high price) to grant access

for a certain time. It is therefore possible that watching a movie on a tablet PC depends on the time of the day.

*Sourcing pricing elements:* It has been established that video and browsing is currently already accounting for up to 75% in mobile networks. In the future, a large part of the data growth will also stem from cloud based storage and/or applications. This could lead to situations in which at a given busy time in a given area critical business data cannot be accessed readily because the bandwidth is absorbed by streaming of leisure videos. Since both user bundles are priced identically, time, intensity and especially economic value are not priced. Given that increasingly traffic originates from large cloud networks such as more leisure centric clouds like YouTube, Netflix as well as more commercial cloud offerings such as Microsoft, SAP cloud, etc. tariffs could be structured as to allow network carrier to block access to some clouds for a limited time during the day.

Aside there is little doubt that fixed and mobile broadband will eventually compete head on. This kind of “fixed” mobile traffic can accommodate by using two dynamic elements: First fixed-mobile DSL packages are tied only to certain cells outside the metropolitan areas and can be reduced during peak hour by 25% interval steps, transparent to the customer. This is to ensure that the mobile RT and NRT traffic enjoy priority over the fixed mobile traffic.

## 5 SUMMARY

The paper documented the astounding rise of mobile communications. While mobile voice was thought to be an add-on the fixed switched public network, it eventually eclipsed the fixed telephony service: the vast majority of voice traffic is handled today by the mobile networks. It is clear that mobile data is following the same evolution path. However, it has been pointed out that the characteristics of mobile data are very different compared to voice.

The small cell size and the fact that the capacity of such a cell is not scalable confine available capacity to very small areas. A variety of devices absorbing very different RT/NRT traffic at certain peak times, pose another big challenge that cannot be addressed by static data bundles alone. The problem is further aggravated by the fact that the traffic with the highest demand on network capacity (i.e. streaming) has generally the lowest economic value to user. The paper therefore argued that moving forward, pricing of mobile data services have to comprise static and dynamic elements. Importantly, the pricing scheme has to segment NRT/RT traffic sharply, ensuring that especially the quality of service of NRT traffic meets customer expectation. Occasional or “light” streaming can be performed, but since this data bundles are small, consumer will not engage in extreme streaming activities. The RT traffic is segmented by first introducing vertical bundles (e.g. watching a video at a train station) for specific streaming services priced according to network load and second making specific services

such as “mobile DSL” only available in non-peak coverage areas. Only such a pricing framework appears suitable to address the specific challenges arising from meeting customer needs for high quality reliable service on one hand and control network cost on the other. Further research is needed to understand how various price setting factors that are effective in controlling the described complex network load balancing between various types of devices and content within the network and its impact on price perception, overall customer satisfaction and data usage on part of the consumer.

## ACKNOWLEDGMENTS

This work was supported by the Slovak Research and Development Agency under the contract No. APVV - 14 - 0797.

## REFERENCES

- Albon, R., 2006. Fixed-to-Mobile Substitution, Complementarity and Convergence. *Agenda*, 13(4), pp.309-322.
- Banerjee, A. and Ros, A.J., 2004. Patterns in global fixed and mobile telecommunications development: a cluster analysis. Growth in mobile communications. *Telecommunications Policy*, 28(2), pp.107-132.
- Baochun, L., Nahrstedt, K. and Xue, Y., 2006. Optimal resource allocation in wireless ad hoc networks: a price-based approach. *IEEE Transactions on Mobile Computing*, 5(4), pp.347-364.
- Basar, T. and Srikant, R., 2002. Revenue-maximizing pricing and capacity expansion in a many-user regime. In: *IEEE INFOCOM, Twenty-First Annual Joint Conference of the IEEE Computer and Communications Societies*, New York, New York, 23-27 June 2002. Piscataway, USA: IEEE Operations Center.
- Calhoun, G., 1988. *Mobile Radio Telephony*. Norwood, MA: Artech House.
- CTIA, 2005. *Semi-Annual Wireless Industry Survey*. [online] Available at: <<http://www.mindfully.org/Technology/2005/Wireless-Industry-SurveyJan05.htm>> [Accessed 01 January 2016].
- Ernst & Young, 2013. Metrics transformation in telecommunications. [online] Ernst & Young Global Limited. Available at: <[http://www.ey.com/Publication/vwluassets/metrics\\_transformation\\_in\\_telecommunications/%24File/metrics\\_transformation\\_in\\_telecommunications\\_eF0117.pdf](http://www.ey.com/Publication/vwluassets/metrics_transformation_in_telecommunications/%24File/metrics_transformation_in_telecommunications_eF0117.pdf)> [Accessed 01 January 2016].
- Fitkov-Norris, E.D., 2003. *Optimal dynamic pricing strategies for mobile communication networks*. Ph. D. University of London.

Gartner Group, 2009. *Web usage vs. operating system*. [online] Available at: <<http://demo.tizra.com/Morgan-Stanley-Economy-Internet-Trends-October-20-2009/39>> [Accessed 01 February 2016].

GSMA, 2013. *The Rise of Connected Devices Will Drive Mobile Operator Data Revenues Past Voice Revenues Globally by 2018*. [online] Available at: <<https://www.gsma.com/newsroom/press-release/gsma-the-rise-of-connected-devices-will-drive-mobile-operator-data-revenues-past-voice-revenues-globally-by-2018/>> [Accessed 10 February 2017].

GSMA, 2014. *The Mobile Economy 2014*. [online] Available at: <<https://www.gsmaintelligence.com/research/?file=bb688b369d64cfd5b4e05a1ccfc9cb48&download>> [Accessed 04 February 2016].

GSMA, 2016. *The Mobile Economy 2016*. [online] Available at: <[https://www.gsma.com/mobileeconomy/archive/GSMA\\_ME\\_2016.pdf](https://www.gsma.com/mobileeconomy/archive/GSMA_ME_2016.pdf)> [Accessed 04 February 2017].

Kessler, S., 2012. *Apples's App Store hits 25 Billion Downloads*, [online] 03 March 2012. Available at: <<http://mashable.com/2012/03/03/app-store-25-billion-downloads/>> [Accessed 01 February 2016].

Khanfir, A. and Fitkov-Norris, E.D., 2000. Dynamic pricing in mobile communication systems. In: IEEE Xplore, *First International Conference 3G Mobile Communication Technologies*. London, UK, 27-29 March 2000. London: Institution of Engineering and Technology.

Lee, H.S., 2010. Factors Influencing Customer Loyalty of Mobile Phone Service: Empirical Evidence from Koreans. *Journal of Internet Banking and Commerce*, 15(2), [online] Available at: <<http://www.arraydev.com/commerce/jibc/>> [Accessed 01 August 2016].

Lee, W., 1976. *Mobile Communications Engineering*. New York: McGraw-Hill.

Levin, H.J., 1971. *The Invisible Resource: Use and Regulation of the Radio Spectrum*. Baltimore: John Hopkins Press.

Malik, O., 2012. *Monthly Traffic volumes in 3G by type of equipment*. [online] Available at: <<http://gigaom.com/broadband/smartphones-ipads-the-state-of-the-mobile-internet/>> [Accessed 05 February 2016].

Paetsch, M., 1993. *Mobile Communications in the U.S. and Europe: Regulation, Technology and Markets*. Boston: Artech House Mobile Communications.

Pelkmans, J., 2001. The GSM standard: explaining a success story. *Journal of European Public Policy*, [e-journal] 8(3), pp.432 - 453. <http://dx.doi.org/10.1080/13501760110056059>.

Portilla-Fugueras, A., Salcedo-Sanz, S., Hackbarth, K.D., López-Ferrerias, F. and Esteve-Asensio, G., 2009. Novel Heuristics for Cell Radius Determination in WCDMA Systems and Their Application to Strategic Planning Studies.

*EURASIP Journal on Wireless Communications and Networking*, 2009(314814), pp.1-14.

Rodini, M, Ward, M.R. and Woroch, G.A., 2003. Going mobile: substitutability between fixed and mobile access. Competition in Wireless: Spectrum, Service and Technology Wars. *Telecommunication Policy*, 27(5-6), pp.457-476.

Seybold, J. S., 2005. *Frontmatter: Introduction to RF Propagation*. Hoboken, NJ: John Wiley & Sons.

Sharma, C., 2012a. *Data Revenues in bn \$ as of Q4/I*. [online] Available at: <<http://www.chetansharma.com/MobilePatentsLandscape.htm>> [Accessed 05 February 2016].

Sharma, C., 2012b. *Development of Desktop/Laptop shipments (blue line) vs. smartphones/tablets (red line) 2009 to 2014*. [online] Available at: <<http://www.chetansharma.com/MobilePatentsLandscape.htm>> [Accessed 05 February 2016].

Sharma, C., 2012c. *US Mobile Data Growth 2006-2011*. [online] Available at <<http://www.chetansharma.com/MobilePatentsLandscape.htm>> [Accessed 05 February 2016].

Statista, 2017a. *Global voice service revenues from 2010 to 2015 (in billion U.S. dollars)*. [online] Available at: <<https://www.statista.com/statistics/218607/global-mobile-voice-service-revenues-since-2010/>> [Accessed 10 February 2017].

Statista, 2017b. *Mobile data service revenues worldwide from 2010 to 2015 (in billion U.S. dollars)*. [online] Available at: <<https://www.statista.com/statistics/218609/global-mobile-data-service-revenues-since-2010/>> [Accessed 10 February 2017].

Steinbock, D., 2005. *The Mobile Revolution - The making of mobile services worldwide*. London: Kogan Page.

Vogelsang, I., 2010. The relationship between mobile and fixed-line communications: A survey. *Wireless Technologies: Information Economics and Policy*, 22(1), pp.4-17.

White Paper, 2011. *TD-LTE: Exciting Alternative, Global Momentum*. [online] Available at: <<http://www.tdia.cn/test/en/downloa/20111214.pdf>> [Accessed 10 February 2016].

Wireless Intelligence, 2012. *From 2001 to 2010, mobile calls increased 10 times*. [online] Available at: <<http://gigaom.com/2011/09/01/from-2001-to-2010-mobile-calls-zoomed-10-times/>> [Accessed 02 February 2016].

Yaipairoj, S. and Harmantzis, F.C., 2004. Dynamic pricing with “alternatives” for mobile networks. In: IEEE, *Wireless Communications and Networking Conference*. Atlanta, Georgia, USA, 21-25 March. IEEE.



Zhao, L., Lu, Y., Zhang, L. and Chau, P., 2012. Assessing the effects of service quality and justice on customer satisfaction and the continuance intention of mobile value-added services: An empirical test of a multidimensional model. *Decision Support Systems*, 52(3), p.645-656.

Zheng, L., Joe-Wong, C., Tan, C.W., Ha, S. and Chiangs, M., 2015. Secondary markets for mobile data: Feasibility and benefits of traded data plans. In: IEEE, *IEEE Conference on Computer Communications (INFOCOM)*. Kowloon, Hong-Kong, 26 April-01 May 2015. IEEE.

---

## ABOUT AUTHORS

**Michael Paetsch**, University of Economics in Bratislava, Faculty of Commerce Department of Services and Tourism, Dolnozemská cesta 1, 852 35 Bratislava, Slovakia, e-mail: paetsch@online.de.

**Peter Dorčák**, University of Economics in Bratislava, Faculty of Business Management, Dolnozemská cesta 1, 852 35 Bratislava, Slovakia, e-mail: peter@dorcak.com.

**František Pollák**, University of of Presov in Presov, Faculty of Management, Konštantínova 16, 080 01 Prešov, Slovakia, e-mail: frantisek.pollak@unipo.sk.

**Lubomír Štrba**, Technical University of Košice, Faculty of Mining, Ecology, Process Control and Geotechnologies, Department of Geo and Mining Tourism, Nemcovej 32, 042 00 Košice, Slovakia, e-mail: lubomir.strba@tuke.sk.

**Branislav Kršák**, Technical University of Košice, Faculty of Mining, Ecology, Process Control and Geotechnologies, Department of Geo and Mining Tourism, Nemcovej 32, 042 00 Košice, Slovakia, e-mail: branislav.krsak@tuke.sk.



© 2017 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC-BY) license (<http://creativecommons.org/licenses/by/4.0/>).