Scarcity of licensed Spectrum for Mobile

Timo Ali-Vehmas
Timo_Ali-Vehmas@hut.fi

Availability of spectrum for mobile communication is an open question. The number of different players with different agendas, emerging new technologies and recently, extremely high auction prices paid by some mobile operators have made the issues more complex and also by far more public. Radio spectrum can be seen as an enabler for different businesses but it may also be considered an important factor impacting also indirectly to the overall prosperity in national or even global perspective.

Keywords: Spectrum management, Spectrum efficiency, Technology evolution, Mobile, Wireless.

1. Introduction

Radio spectrum has been utilised for 100 years by now. Since the first days of operation there has been more or less vivid dispute who and on what grounds should manage this valuable asset. Spectrum management has been under special attention by the governments in all countries because of its importance for security, emergence and military use but also because it has been tradition to regulate all telecommunication matters anyway.

Growth in mobile communications have exposed the scarcity of radio spectrum but it is not clear how serious this demand really is and especially, what is the future growth rate of service demand, taking into account e.g. that voice penetration in most OECD countries is saturating. Also it may be debated how efficiently the current spectrum allocations are in use.

Availability of radio spectrum for highly valued purposes will have significant impact to overall economical prosperity. Some studies have analysed the issue and they have found that macro economic impact may be of the same order of magnitude as the micro economic impact, i.e. contribution of spectrum to Gross Domestic Product (GDP) in Denmark. Macro economical impact includes also production of products and equipment utilising the radio spectrum. This is dependent on the economical structure of the country. Because the market of radio based telecommunication products and equipment is global, the macro economical impact can be evaluated only based on the global availability of radio spectrum and analysing the overall business potential of related products and equipment including components and software etc. Therefore it is useful to limit the discussion here to the micro-economical approach. [2]

![Direct contribution from Radio Spectrum to Danish GDP](image)

Based on the figure 1 it seems that spectrum used for mobile communication is the most important contributor to GDP. Therefore one conclusion is that radio spectrum should be reserved primarily for mobile traffic, including mobile reception of broadcasting. Any non-mobile traffic should be provided using cables and other means, which do not consume the precious asset, the radio spectrum.

In the following sections spectrum management and regulative issues as well as service and technology evolution in radio based mobile communications are discussed and some future prospects debated.

Commercial and strategic aspirations impact heavily the debate about spectrum management but the focus of this paper is a bit more scientific rather than political and therefore this type of argumentation is not discussed here.

2. Spectrum management and regulation

Spectrum management and regulation is a national matter. The national regulators, like Federal Communications Commission (FCC) in the United Stated and Viestintävirasto, Finnish Communications Regulatory Authority (Ficora) in Finland are responsible both on the efficient use of the spectrum (Spectrum management) and the licensing of the spectrum (Regulation) in their domains. National Regulators however coordinate their activities on EU level and globally in ITU in order to harmonise the use of the spectrum. [2, 3, 4]
Although the scope of this study is the microeconomic issues in spectrum allocation, the economical impacts are not small. The overall goal for spectrum regulator is to maximise social welfare created by the radio spectrum. The regulator has to balance between e.g. the public interest in spectrum auction income and commercial viability of the operators to provide high quality service for consumers. Too high investment in spectrum license will reduce the capability of the operator to invest in the network. On the other hand price of the spectrum should be high enough to discourage less serious or capable operator candidates. Poor knowledge and unrealistic plans will also end up with bad service quality.

Regulators have to consider also many difficult technical details in order to justify the best possible use of spectrum. Spectrum fragmentation as well as too wide spectrum blocks make implementation of the products more expensive because of challenging filtering and other implementation requirements, technology independent allocation leads to narrowband domination, which on the other hand is not superior in supporting new wide band services. Spectrum efficiency of different radio systems is very difficult issue to estimate. Small-scale test networks provide only some rough estimate because the real figures can be measured only in fully loaded network with practical mobility of commercial mass production terminals, indoor coverage and other factors. It is important to note that spectrum efficiency, maximum bit rate capability and other features are used in marketing and political messages of the candidate operators in quite liberal way.

Finally the regulators must anyway secure some spectrum for emergence and administrative use, including military. In many countries at least part of the broadcasting operations is under political control, which naturally prohibits use of that spectrum according to commercial logic.

There are many ways to define the use of radio spectrum. The most common framework is allocation where a chunk of radio spectrum is allocated to some operator or other party. Spectrum allocation may be defined in more detail in technical terms such as maximum emitted power and spectrum masks. Spectrum allocation may also be defined more detail in terms for what it is allocated. This is called spectrum assignment. In this case the allocated radio spectrum is further on assigned to certain purpose. For instance it may be defined that a broadcasting company may use its spectrum allocation for broadcasting but not for mobile communications purposes. National or any borders do not stop radio signal propagation. Therefore there must be rules how radio spectrum is used geographically. This is called spectrum allotment.

The value of radio spectrum depends strongly on the frequency range. The low frequencies (below 3GHz) are considered the most valuable part. This is because the propagation characteristics of the low frequency radio waves enable wide area coverage. Also the availability of RF components for low frequencies is better because of mature technologies. However, the frequencies below 1 GHz require larger physical size of antennas and other components. Therefore the best frequency band for small size portable wireless devices is in practise between 1 and 3GHz. For very small cells frequencies up to 10 GHz can be used supporting high capacity but propagation through walls and behind the corners is getting poor as well as the RF technologies for high frequencies are more expensive. High frequencies are also suitable for unlicensed operation because the poor propagation is in this case an advantage, which reduce the need for cell and frequency planning.

The spectrum license often includes elements of assignment and allotment. Unlicensed use of spectrum is in best case not restricted by neither the services it is used for, nor by the availability of the spectrum in geographical terms. Unlicensed use of spectrum is typically limited by technical characteristics of the transmission (maximum output power) and therefore it is not very suitable for wide area networks.

**Licensed use of Mobile spectrum**

The current allocations can be put into at least four different baskets

- First generation systems around 400 MHz. This band is almost obsolete for mobile communications today and is in the process of being re-farmed to other purposes.
- First and Second generation systems on 800-900 MHz and on 1700-1800 MHz. This band is heavy use and is almost all upgraded to all digital 2G.
- Third generation systems on 1900-2000 MHz. This band is now being deployed using 3G except in Region 2 (mainly USA) is used for 2G.
- Other bands at 2400 and around 5000 MHz. Currently used or planned to be used for 3G extensions or for other radio technologies, which may be applicable also for mobile communications. Licensing conditions vary.

![Figure 2. Mobile communications spectrum after WRC2000. Source: UMTS Forum [6]](image)
The World Radio Conference (WRC2000) decided that current mobile communications spectrum allocations at 800-900 MHz should over long period of time be re-farmed (upgraded) to 3G. There is also a new potential mobile communications spectrum available between 2500-2680 MHz. [5]

Depending on the region and country it may be concluded that the current allocation for mobile communication is roughly 350 MHz and it may be extended to about 590 MHz during the next 10 years.

Unlicensed Mobile Spectrum

There is also about 80 MHz allocated for ISM (Industrial, Science and Medical) which is currently used e.g. by Wireless Local Area Network (WLAN) systems, also called Wi-Fi. This spectrum is not licensed for any particular use or for any particular operator but is free for anybody who only complies with certain technical requirements. Additionally 200 MHz at 5000 MHz range is being allocated for similar purpose. There is heated debate on the role and value of traditionally non-mobile (in wide area sense) wireless systems such as Wi-Fi. Because of the recent fast growth rate of these technologies and services and because their capability to offer local area mobility, it may be useful to consider these options also in the context of licensed spectrum for mobile communications.

Art of Spectrum licensing

There is no single global way to decide the spectrum licensing but the situation varies strongly in different countries. In case of the 3G spectrum licenses there may be three main approaches visible:

Fully administrative process means that regulator simply without any external influence decides the operator and the conditions of the spectrum license. The advantage of this approach is that the government fully owns the spectrum and any reason is a good reason to re-decide if the outcome was not satisfactory.

Beauty Contest is a process where the candidate operators present their plans to the regulator and regulator selects the best proposals. Typically it seems that best proposal is the one having the best coverage and best service plans with credible investment plans.

Spectrum Auction is a process, where the candidate operators offer higher and higher bids for the spectrum license and the operator with the highest offer will win. This is a straightforward approach but it may lead to unrealistically high price of the spectrum.

Administrative decisions were used in all countries before the liberalization during 1990 but for the 3G spectrum licenses beauty contests were used for instance in Sweden and in Finland and spectrum auctions were used in most of the EU countries and in the USA.

Spectrum license include also a lot of additional terms and conditions such as nominal duration of the license, minimum service levels, roaming requirements, conditions if service is not provided as planned, possibility to re-sell the license, possibly sharing of infrastructure and so on. Both the EU and the USA regulators have started recently to look for more flexible ways to allocate spectrum but there are a lot of issues unsolved. [7, 8, 9, 10]

3. Service evolution

Service evolution is very important guideline to the regulators. Today the main service used in mobile communication is still the basic voice service. The main characteristics of voice include narrow bandwidth, relatively loose bit error rate but demanding i.e. short delay requirement. Most of the new advanced services, such as data access, email, video streaming etc. require order of magnitude wider band and better bit error rate but can handle delay constraints better. Additionally, most of the new services will use Internet Protocol, which originally is developed for cable-based transmission.

Service evolution is driven also by the capabilities of the mobile terminal devices. Currently voice quality is limited by the transmission rather than the technology constraints of the terminals. For video and data processing the limitations may be opposite. The always connected, always on –characteristic of mobile devices allow optimisation, which limits the maximum needed bit rate to the level that can consumed real time by the terminal and the user. *)

Finally the service evolution is clearly limited by the spending capabilities and willingness of the subscribers. The average ARPU has been around 50 € with large variations in different countries. The overall spending for communications has grown from 1 % to 3 % level of the overall spending in OECD countries during the last 10 years. Taking into account that in most OECD countries the penetration has reached saturation level it may be estimated that the actual demand is growing only at modest pace for the next years. [11]

4. Technology evolution

Technology evolution will be quite fast in the radio technologies for the next decade. This is due to many factors.

*) (Kind of 1000 horse-power cars are not very popular in cities or even on motorways with less maximum speed limit of 120 km/h. Many people are dreaming about them but few actually drive one.)
Transition to third generation technologies is in short term the most important factor. Evolutionary step from 2G technologies to 3G technologies is at the same time both smooth but also significant. Strong desire to avoid service discontinuity has lead to some compromises in the 3G -technology selection process. The mandatory requirement to support circuit switched voice has increased the complexity but flexibility of wideband radio supports well also IP based services. Spectrum efficiency of 3G technologies, like WCDMA and EDGE, is significantly better than that of traditional 2G technologies, such as GSM/GPRS. But EDGE is fundamentally narrow band technology and therefore its possibility to support wide band services is limited.

For all 3G technologies maturing period is still in progress and there will be a lot of improvements identified during the first years of the operation, both in understanding how to deploy the networks and especially how to implement 3G products.

The long-term factor however will be the well-planned evolution of 3G technologies to 3.5G and further on towards 4G. Technology innovations such as HSDPA for packet optimised radio traffic, IMS for IP based service support and MIMO, MUD and other low layer radio technologies all impact to the spectrum efficiency and flexibility as well as compatibility with IP networks.

The original WCDMA (as defined in ETSI in 1998) has a spectrum efficiency of about 200 kbit/s/MHz/cell in loaded network. This is probably close to the level the WCDMA networks are able to deliver during the first years of operation. With all the enhancements taken in use this figure will be 2 to 4 times better or even more. But it is important to note that performance of radio network is extremely dependent on the radio environment. This actually creates higher uncertainty to the performance than the technology evolution is able to improve in average during a life cycle of one technology generation. [12]

5. Estimating Spectrum needs

The ultimate goal in spectrum allocation is hence to optimise the social welfare while enabling the service evolution under constraints of consumers spending capabilities as well as evolution capabilities of radio and other terminals.

Simple Example

We can look at the service and technology evolution by estimating what would be the cell radius under some scenarios using WCDMA and some imaginary service mix in different environments [13].

Taking a bit forward looking assumptions we may define the service specification for instance as follows [14]:

- Mobile Voice service: 200 mErl, 10 kbit/s coding (like in GSM EFR, MOS better than 4)
- Mobile Video conferencing: 100 mErl, 144 kbit/s (Minimum would be around 64 kbit/s but this will provide reasonable voice and video quality)
- Mobile Multimedia: 10% of consumers using information from the Internet, 256 kbit/s (This bit rate can support real time high quality video stream for the screen size of portable devices, less than 5 inch)

We may estimate the service penetration to be 100%, which is not far from the current situation in EU. We can estimate that the spectrum needs are highest in the urban areas and specifically in some hot spots like at the airport. We may hence define two cases for instance as follows:

- Case 1: London metropolitan area, Population 10M, area 3000 km²
- Case 2: London Heathrow Airport, Average distance between people 2 meters.

In order to calculate the spectrum needs we can first calculate the bit rate demand per geographical area as follows:

- Case 1: 140 Mbit/s/km²
- Case 2: 10500 Mbit/s/km²

These figures are downlink specific, because the uplink may be less loaded because of the Internet traffic is mainly downlink.

Now if we take into account the allocated spectrum for all operators (60 MHz) and spectrum efficiency of WDCMA as stated above, we can calculate minimum average cell radii as follows:

- Case 1: \(r = 167 \text{ m}\)
- Case 2: \(r = 19 \text{ m}\)

In this example the London metropolitan area can be supported by the WCDMA with cell structure similar to the voice oriented GSM networks today. But the very dense user population at the airport introduces too heavy load and forces the cell radius to small. This will be partly compensated by very low mobility of the end users and we may also assume the base stations being located inside the terminal. Further on we may assume that part of the traffic is actually supported by the evolved 2.5 G systems or even re-farmed 3G (hence total downlink spectrum is 150 MHz). If we further assume that future evolution of WCDMA radio technology will provide up to 4 times better spectrum efficiency, maybe more the case 2 is actually more realistic.

- Case 2, revised: 85 m
Assuming the likely re-farming of current mobile systems to 3G and its future evolutionary features it seems that even very high demand of services may be satisfied with demanding but realistic cell structures. The question however remains, if the service concept used above is realistic. If more users would like to use high bit rates services, it is possible that the current spectrum for mobile services can’t satisfy the needs.

It is also important to note that the simplified study above is not WCDMA specific, it only assumes the spectrum efficiency of WCDMA as one parameter. The approach is valid for any radio system if only the spectrum efficiency of the loaded network and available spectrum allocation are known.

Fundamentally there is no capacity limit in the cellular systems, if only smaller and smaller cells can be implemented. The devil is as usual in details. The implementation constraints, especially the dynamic range of signal processing (analog and digital), define the minimum cell radius as a function of maximum cell radius. On the other words, a system may be designed for long range or high capacity, but it is not possible to achieve the both ultimate targets at the same time.

The operational point of any cellular system is defined by the investments needed to build the network. The cost of the spectrum will impact the operational point and in practise limit the maximum range – or make large cell network unprofitable by default.

We may describe the situation with following qualitative drawing.

![Figure 3. Relationships of capacity and cost of a cellular network as a function of cell radius](image)

The three continuous curves indicate capacity of the network as a function of cell radius. The capacity is limited, however, with the minimum specified cell radius, which is a function of minimum power level of the products (and also function of the carrier frequency used). Different radio systems define the minimum power level differently. In the figure there is an assumption that minimum power levels of all systems is the same, i.e. the minimum cell radius is the same. This is often the case because of the backwards compatibility requirement of evolitional specification process.

Maximum power level defines the maximum cell radius, which traditionally has been pushed high, in order to save network investment cost for coverage. Therefore one conclusion is that for high performance, high coverage system it may be better to split the requirements into two modes, preferably operating on different frequency ranges. This opportunity was not fully utilised in the 3G system specification phase. It could have assumed that 900 MHz band will be re-farmed to high coverage mode and 2 GHz band will be used only for high capacity mode. Power levels of each frequency band would have been defined accordingly. This assumption would have been possible only if there had been some assumptions how the spectrum licences would be allocated.

The dotted curves indicate the cost of the network as a function of fixed cost and variable cost. Fixed cost include non scalable costs, such as spectrum licence cost and variable cost include cost elements such as base station costs, which are somehow relative to cell radius.

In the Figure 3 there are two important points:

- With high spectrum cost only high performance (an often more expensive) technologies can be used because low performance system can never support high enough number of users and services.
- With high spectrum cost, there is a maximum applicable cell radius, which can be considered, because the capacity is a function of cell size and large cells can never support enough high number of users to create the needed service revenue.

As an intermediate conclusion we can therefore state that it should be known already in the specification phase of a new cellular system, whether spectrum licensing costs are going to be a significant factor. If so, then focus in the technology development should be pushed towards the high capacity system and especially push the minimum power level of the base stations and terminals lower.

In order to connect the discussion to some real world case, some average cell size targets have been added to the graph. Those estimates are based on the 3G beauty contest documents of the licensing process in Sweden. The potential match of the curves with the beauty contest cases is purely accidental but may still support some general discussion on the topic. [15]

6. Planned extension bands

The current spectrum allocations are likely to limit the service evolution in long term if the popularity of wide
band services, both in demand and in supply, will grow significantly. Therefore it is important to seek some additional spectrum for mobile communication already today. There are some potential new bands identified but before these bands are allocated and especially assigned for any purpose it is important to consider some new aspects.

The previous section already introduced advantages of the multimode and multiband approach, which can at the same time optimise the systems for high capacity and high coverage scenarios. Earlier goal for single radio system technology operating only on one global band therefore is not the best approach.

Until now the service mix is thought to be symmetrical. It should be studied further what will be the main direction of the traffic. If the majority of future services will be client-server type information consumption it may be realistic to allocate more capacity for the downlink. It may be useful to note that in most systems the uplink performance is better because of possibility to use more signal processing at the base station than in the terminal. On the contrary, heavy peer-to-peer traffic, as experienced today in the Internet will reduce the asymmetry. There are two fundamental concepts of allocating asymmetrical traffic to spectrum. One can use TDD operation, which is also able to dynamically change the direction on demand. Another option is to simply allocate more downlink spectrum and use variable duplex distance in FDD systems [16].

Whatever concepts are used, it is important to note that spectrum fragmentation for number of incompatible systems in any geographical area will lead to lower overall spectrum efficiency. This is due to additional guard band requirements. Impact of interference is not the same for all systems but some system technologies suffer more than the others. This is generally function of the system bandwidth.

Novel technologies, such as Ultra Wide Band (UWB) radio systems are highly tolerable to alien signals as long as the dynamic range of the radio is not exceeded. But these new systems are vulnerable to near far effect also, and high power narrow band transmitter (like GSM terminal) in the near vicinity will kill the UWB receiver and therefore it is difficult to provide guarantee of service using secondary licensed spectrum even with advanced technologies.

7. Future of Spectrum regulation

It is extremely challenging to propose any new rules to spectrum allocation process, because the issue is so complex and complicated.

As a first step the old pit falls should be avoided. Spectrum fragmentation, commercially too high price of spectrum license without possibility to resell the license, significantly delayed deployment due to unsuccessful beauty contests, blocking of new technologies due to claimed interferences and in general very slow progress of spectrum licensing have been the claimed problems in spectrum allocation. [ ]

New innovative solutions should be investigated. New technologies may be used to enable commercially viable operation also on unlicensed bands. It may be possible to create technologies, which enable real time charging of spectrum use. This would allow concept similar to toll highways. This was briefly considered already in the early phase of 3G technology development but it was seen too difficult technically and politically to implement. May the timing is better today. The real time trading may be extended even further to allow the consumers to gain profit when they allow other users to utilise their products as one hop link in future mesh and other adhoc radio networks. The trading could include some spectrum charge also.

8. Summary

Spectrum management is complex and complicated matter, where high system intelligence is needed. There seems to be enough radio spectrum allocated or becoming available also for mobile communication but if not efficiently managed and utilised there will be scarcity of spectrum during the next 10 years.

Governments are seeking ways to maximise social welfare. It seem that mobile communications is likely to convert the economic value of radio spectrum to contribution to GDP. Re-allocation of all non mobile traffic to cable systems or limited by strict allotment to limited geographical area is one of the most important goals, which will release the most valuable spectrum below 3 GHz.

There should be enough competition and other regulative measured such as trading of the spectrum licences and possibly secondary use of licensed spectrum to speed up the deployment of the advanced technologies as quickly as possible. It should also be noted that any product put into service will stay there for quite some time and it will keep on consuming more spectrum resources and often disable more advanced products to gain benefit, too. This requirement is valid for both product technologies as well as for network planning and deployment.

Dedicated Mobile communication systems may not be the most efficient in broadcasting the unidirectional services. Therefore radio broadcasting should be used especially for services with mobile relevance. Other broadcasting services should possible be provided by cable and other non-radio technologies.

Based on this study it seems quite likely that future radio communications systems should always include...
multimode and multiband capabilities to address both requirements for high coverage and capacity at the same time. This is also the most likely goal of future generations, since the 4G of mobile communication is assumed to be a combination of existing and new technologies.

Spectrum fragmentation should be avoided as much as possible. Because of political reasons this may not be possible in all cases. In order to achieve scale of economy advantages the minimum requirement is global core bands for coverage and capacity networks.

Finally the spectrum management process should be streamlined to be able to provide spectrum for high value services in promptly manner. Economical value of the radio spectrum can be utilized by many different approaches, like auctions, beauty contests or by allocating the spectrum for open unlicensed use. Whatever the method may be, it should support the use of radio spectrum in efficient way for services, which require mobility.

References


