ANNENBERG RESEARCH NETWORK ON
INTERNATIONAL COMMUNICATION

Spectrum licensing and spectrum commons—where to draw the line

Martin Cave
Centre for Management under Regulation
Warwick Business School

William Webb
Warwick Business School and Ofcom

A research report prepared for the International Workshop on Wireless Communication Policies and Prospects: A Global Perspective, held at the Annenberg School for Communication, University of Southern California, Los Angeles, October 8th and 9th 2004.

“Wireless Communication Policies and Politics: a Global Perspective”

October 8-9 2004

University of Southern California

Spectrum licensing and spectrum commons – where to draw the line

Martin Cave and William Webb[*]

Revised version, September 2004

Martin Cave
Centre for Management under Regulation
Warwick Business School
University of Warwick
Coventry, CV4 7AL
Martin.Cave@wbs.ac.uk

---

[*] Warwick Business School and Ofcom. This paper was written while William Webb (who is now head of R&D at Ofcom) was at PA Consulting. It represents the authors’ views only.

SUMMARY

In this paper, we consider the role of unlicensed spectrum and in particular address the question as to how to determine whether there should be more or less unlicensed spectrum.

We start with an economic analysis of the situation which suggests that spectrum should be unlicensed where there is little probability of congestion. We then note that despite arguments about the ability of "spectrum commons" to alleviate congestion, congestion across key parts of the spectrum is likely for the foreseeable future. But congestion is unlikely where short range communications are used and can be made less likely by regulatory insistence on eg politeness protocols. This leads us to conclude that there should be a mix of licensed and unlicensed spectrum with the unlicensed approach restricted to bands and applications where congestion is unlikely.

This conclusion implies that some entity has to determine the likelihood of congestion for each band on a regular basis as circumstances change. We would prefer this to be the market, and have put forward a mechanism whereby a band manager might buy spectrum under auction and turn it into a private commons. However, we have concerns that difficulties in collecting revenues might render this suboptimal. In this case, the responsibility to determine whether a band is likely to be congested falls to the regulator.

Regulatory intervention is always a matter of judgment, but we suggest the process set out in Figure 1 which might go some way to guide the regulator.

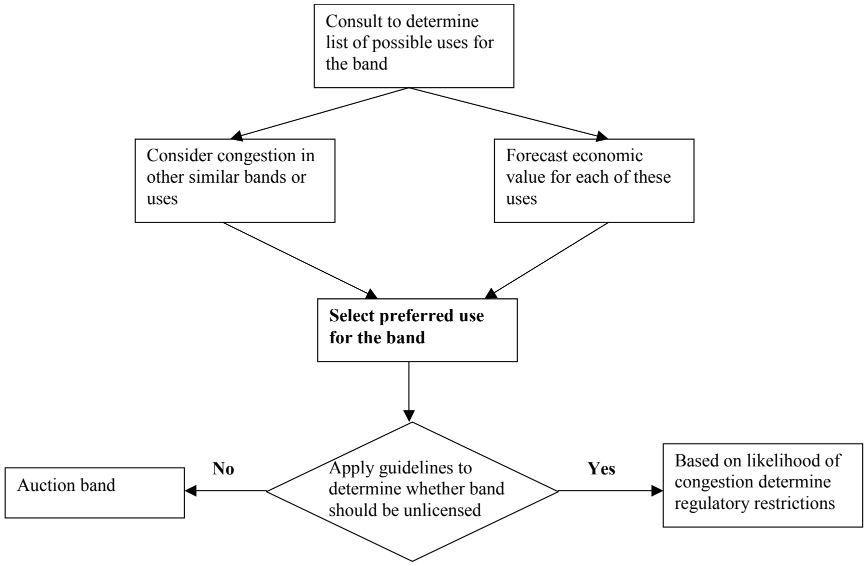

Figure 1 : Process to guide the regulator in determining whether spectrum should be unlicensed

We provide more details on each of the stages of this process below.

1. INTRODUCTION

Unlicensed[1] spectrum was until recently of little interest. However, in the last five years it has been debated more widely. This has been caused by the following developments:

- Deployments of new technologies in the 2.4GHz band, particularly W-LANs have been commercially successful, leading many to ask whether further unlicensed allocations would result in more innovation and deployments.
- The development of ultra wideband (UWB) and the promise of software defined radio (SDR) has led some to question whether these technologies can overcome historical problems with unlicensed spectrum.

The debate around the role of unlicensed spectrum has been particularly intense in the US, where the term “spectrum commons” has come to be used to advocate an approach where much more of the spectrum is unlicensed.[2] Advocates have suggested concepts such as radios seeking temporarily unused spectrum, making short transmissions and then moving onto other unused bands. Some of these concepts extend beyond unlicensed spectrum and into property rights and we discussed easements in the context of property rights in a previous paper in this series.

In general, the debate has suggested that there are three key options in the management of the radio spectrum:

- The current approach, sometimes termed “command and control” where the regulator decides what the spectrum is used for.
- A trading approach whereby owners can trade spectrum with others and change its use.
- A “commons” approach where spectrum is unlicensed.

There is general agreement that the “command and control” approach should be used as little as possible, mainly in cases such as for public safety or military usage where market structures might not generate an appropriate result. However, there is little agreement as to the relative amount of spectrum assigned to trading and unlicensed usage. There are also many hybrid suggestions. For example, Noam[3] has suggested that spectrum be unlicensed but users have to pay a fee to access it depending on the current level of congestion. Alternatively, Faulhaber[4] and Farber have suggested that all spectrum be licensed but that license holders be able to create “private commons” allowing a form of unlicensed access which they charge for in some form.

---

[1] Regulators often prefer the term “licence exempt spectrum”, but because “unlicensed” is in more common usage this is the terminology we will adopt for this paper.

[2] See especially, interesting papers by Benkler, Lehr, Werback and others.

[3] Noam, “The fourth way for spectrum”, FT, 29 May 2003

[4] Faulhaber and Farber, “Spectrum management: Property rights, markets and the commons”, http://rider.wharton.upenn.edu/~faulhabe/SPECTRUM_MANAGEMENTv51.pdf

In this paper we examine the issues that underlie the debate and draw our own conclusions as to the most appropriate path for regulators to follow with respect to unlicensed spectrum. We do so as follows:

- In section 2 we provide some history to the commons and set out the frequencies considered.
- Section 3 sets out the economics of the choice between licensed and unlicensed.
- Section 4 looks at the likelihood of congestion in the radio spectrum, a key component of determining whether unlicensed usage can succeed..
- Section 5 sets out guidelines for the regulator in determining whether spectrum should be unlicensed.

2. HISTORY

A detailed history of the development of unlicensed spectrum is the US is provided by Carter *et al*[5] . Broadly the same history applies in other countries. In essence, in the 1920s most spectrum was unlicensed. The confusion and interference this caused, especially among broadcast stations in the USA, led to a licensed approach being adopted in the 1930s, although some spectrum was still set aside for unlicensed use.

Over time, the main unlicensed bands were those designated as industrial, scientific and medical (ISM). These were bands where there was non-communications use of spectrum, for example for heating purposes. Because this use generated interference, the ISM bands were generally not licensed. Hence, they were often made available for unlicensed usage. Table 1 shows the currently unlicensed bands in the UK.

Generic Frequency Band	**Application**
9 kHz to 30 MHz	Short Range Inductive Applications
27 MHz	Telemetry, Telecommand and Model Control
40 MHz	Telemetry, Telecommand and Model Control
49 MHz	General Purpose Low Power Devices
173 MHz	Alarms, Telemetry, Telecommand and Medical Applications
405 MHz	Ultra Low Power Medical Implants Devices
418 MHz	General Purpose Telemetry and Telecommand Applications[6]
458 MHz	Alarms, Telemetry, Telecommand and Medical Applications
864 MHz	Cordless Audio Applications
868 MHz	Alarms, Telemetry and Telecommand Applications
2400 MHz	General Purpose Short Range Applications, including CCTV and RFID. Also used for WLANs including Bluetooth Applications.
5.8 GHz	HiperLANs, General Purpose Short Range Applications, including Road Traffic and Transport Telematics
10.5 GHz	Movement Detection

---

[5] Carter, Lahjouji and McNeil, "Unlicensed and unshackled: A joint OSP-OET white paper on unlicensed devices and their regulatory issues", http://www.fcc.gov/osp/workingp.html

[6] Note: This band is to be withdrawn by December 2007

Generic Frequency Band	**Application**
24 GHz	Movement Detection
63 GHz	2nd Phase Road Traffic and Transport Telematics
76 GHz	Vehicle Radar Systems

Table 1 : Unlicensed bands in the UK (Source: Spectrum Strategy Document, published by the RA)

In 2000, 9% of the spectrum (up to 60GHz) in the UK was licence-exempt – the same proportion below 3GHz (‘prime spectrum’) as above it.

By far the most important band in terms of economic value is that at 2.4GHz. The reasons why this band has proved so valuable are:

- it is available worldwide as an unlicensed band.
- the band is relatively large (83MHz wide).
- it falls within one of the preferred frequency bands, having a useful range and relatively low cost equipment.

The development of this band is mostly fortuitous and based on the fact that the resonant frequency of water molecules is 2.45GHz. This makes the frequency optimal for many heating applications including microwave ovens. This in turn forced the allocation of the band as ISM and hence unlicensed. The same issues apply worldwide, hence the nature of the allocation. There appears to be no other bands below 100GHz where a similar physical property has resulted in another world-wide allocation.

In future, as shown by the 5GHz band allocation process, a widespread unlicensed allocation will likely require co-ordinated regulatory activity with all the inherent problems and risks involved.

3. THE ECONOMICS OF THE CHOICE BETWEEN LICENSED AND UNLICENSED

Initially we assume known technologies and demands for spectrum-using services, and on these undoubtedly unrealistic assumptions establish when in principle spectrum should be rendered tradable (and priced) property and when it should be a commons. The assumptions are then relaxed.

3.1 CASE 1: A SINGLE OUTPUT

Suppose a frequency has been allocated to a single service, and the only issue is whether spectrum has to be rationed. Suppose initially that output requires spectrum in fixed proportions, and is competitively supplied with constant returns to scale.

The benefits of each regime are shown in figures 2 and 3. The net benefit (NB) curve shows willingness to pay for the service minus marginal/average cost of production including all complementary inputs. Supply is given by the vertical supply curve S. If enough spectrum is available to take output up to Q*, then the equilibrium price of spectrum is zero. There is no need for rationing and a commons, which avoids the administration cost of a property and trading regime, is preferable. This is illustrated in figure 2, where social welfare is shown as the area under the net benefit curve (A). In figure 3, by contrast, marginal net benefit remains as positive up to Q***, yet spectrum constraints limit output to Q**. The equilibrium price of spectrum is OP**. Welfare comprises the area (B) of consumer surplus and the Area (C) of spectrum revenue, which is a transfer normally to the government. Treating spectrum as a commons in these circumstances will lead to overuse, congestion and harmful interference which will reduce the value of the service and move the net benefit line towards the origin (eg NB^1), reducing social welfare. If the supplies of spectrum in figure 3 become a profit maximising monopoly, the monopolist may restricts supply further and may leave some spectrum unused. (A spectrum monopolist would behave in the same way in the conditions described by figure 2)

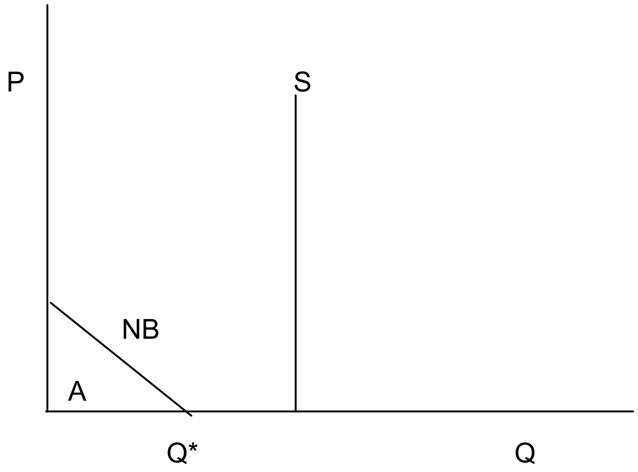

Figure 2 – A commons

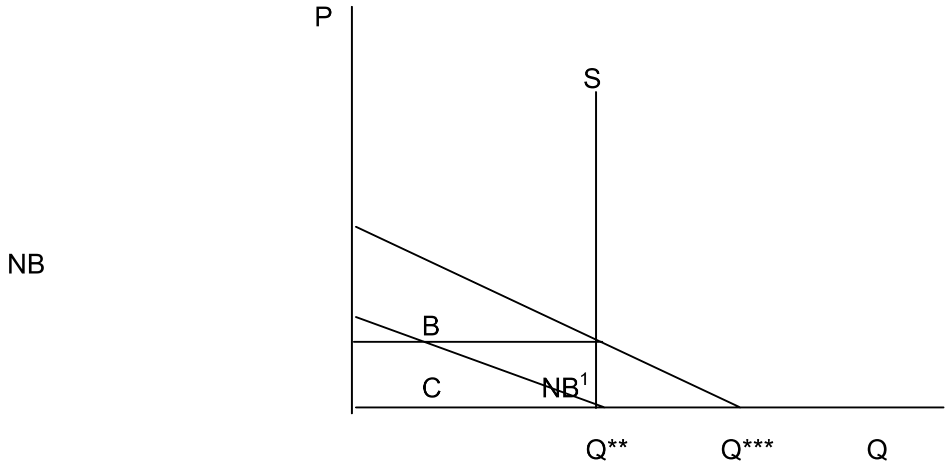

Figure 3 - Scarcity

If other inputs can substitute for spectrum, a positive spectrum price will provide incentive to cut back on spectrum use, and this will tend to increase the maximum output attainable. Under a commons regime, however, firms faced with a zero price for spectrum will have no incentive to economise, except to the extent that doing so improves their competitive position.

The results so far show that where there is an excess supply of spectrum, a commons works best. This is hardly surprising.

3.2 CASE 2: COMPETING OUTPUTS

Suppose now that a band can be used for either of two purposes, (but not both in combination) in one of which it is appropriately a commons, as in figure 2. In the other, shown in figure 4, there is excess demand for spectrum. Clearly, auctioning the spectrum to competitive firms producing the output shown in figure 4 will yield positive returns.

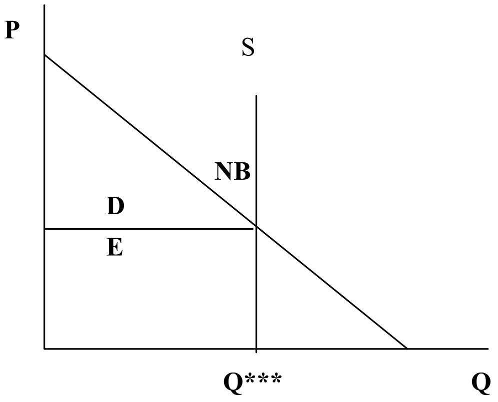

Figure 4: Demand from a competing industry

Is this the better use of the spectrum? Social welfare in figure 4 is shown by the sum of the areas D and E. Do they together exceed area A in figure 2? In this case they do, but clearly they need not. If the net benefit curve in figure 4 were much flatter, the alternative ‘commons’ use might be preferable.

This shows that, if spectrum is intended to maximise social welfare, auctioning is not enough. A prior decision has to be made on to whether to designate bands as a commons.

3.3 CASE 3: MULTIPLE BANDS

Suppose numerous bands are available simultaneously. In the traded sector, in the absence of ex-ante allocation decisions, firms can be expected to sort themselves out through an auctioning or trading process to match services and frequencies optimally, but as discussed in Case 2, commons may be under supplied. If several commons exist, they may not be used with optimal technical efficiency, but this does not matter if there is no congestion in any band. One possible outcome would be for all spectrum prices to drive to zero, permitting a universal commons, but this looks highly utopian in the foreseeable future.

3.4 CASE 4: GROWING OR UNCERTAIN DEMAND

Reverting to Case 1, now suppose that demand for spectrum is currently less than supply, but this will change –or may change- in the future. If a commons is created, that will serve for now but the harmful consequences of congestion will or may occur later.

These effects can, of course be mitigated by making more spectrum available later. If not, there seems little alternative to introducing property rights at the outset, in view of the difficulty of retrieving unlicensed spectrum. Initially, the spot price for access to spectrum would be zero, but the asset price would reflect likely future scarcity.

3.5 CASE 5: REGULATION OF USE OF THE COMMONS

It is possible to postpone or avoid the effects of congestion by imposing limits on use of unlicensed spectrum with respect to i) use, including use to provide a service to the public, ii) equipment permitted, iii) the power at which equipment may be used, iv) the enforcement of politeness protocols. This will reduce net benefits but be preserve the viability of the commons for longer.

3.6 CASE 6: SPECTRUM OWNERSHIP VS SPECTRUM ACCESS?

It has been suggested, by Noam and others, that a system of spectrum ownership or exclusive licensing be replaced by one of spectrum access, so that spectrum users should buy access to spectrum only for the period that they need it. The difference is analogous to that found in telecommunications access, where a firm might either lease transmission capacity from another operator (equivalent to an exclusive spectrum licence) or by pay for traffic throughput.

Such alternative access arrangements are clearly more valuable if operators are using agile technologies which can use a variety of frequencies, as they can then respond to dynamic price signals which guide them to the cheapest points.

While this approach encourages spectrum efficiency it may not eliminate scarcity. In other words, the dynamic market clearing price may not always be zero. Accordingly, if the price were zero ‘throughput’ would be subject to the same congestion and degradation of service quality as outlined above. We do not, therefore, regard this as a ‘middle way’ between a licensed and unlicensed approach.

3.7 CASE 7: INDIRECT REVENUE GENERATION FROM UNLICENSED SPECTRUM

Is it ever possible that a firm would bid for spectrum against other users and make it available as a ‘private commons’? It might be a viable strategy if another form of revenue were available – for example through the sale of equipment required to make the unlicensed use. Analogously, before advertising-financed broadcasting was developed, radio programmes were provided as a ‘commons’ by equipment manufacturers.

This is clearly a possibility, and the revenue available by this means would reflect the willingness to pay of consumers for the relevant final services. But individual manufacturers would have an incentive to free ride by failing to contribute the cost of spectrum, as would individual purchasers of equipment.

3.8 IMPLEMENTING THE NET BENEFIT CALCULATION

When a service is widely diffused, it is relatively straightforward to compute the net benefits. What we seek is the area under the demand curve, or the cumulative willingness to pay, of customers. The net benefit is this magnitude, minus the non-spectrum costs of supply.

By way of illustration, we consider a recent paper on the demand for wireless Internet access in the United States[7]. This service uses licensed spectrum, but the same approach would be employed for a service which used unlicensed spectrum

Like many telecommunications services, wireless Internet involves an access charge. A consumer will buy the service provided her surplus from usage (the amount by which her willingness to pay exceeds the price) is greater that the access charge. This enables the investigator to estimate the distribution of consumer surplus from responses to questions about willingness to pay for access of the kind ‘what is the most you would be willing to pay on a monthly basis for wireless access to the Internet?’

Several other studies, including some carried out by the former UK Radiocommunications Agency (now part of Ofcom), have used this method of establishing people’s willingness to pay to estimate the value of spectrum. There is continuing debate about whether such statements about preferences are consistent with actual behaviour, but a growing consensus that, if appropriate questions are asked, the result can be relied on.

The difficulty, however, relates to new products. Respondents may genuinely be unable to express a willingness to pay for something which they do not understand. This would have restricted, for example, the application of this technique to Wi-Fi until recently.

---

[7] P. Rappoport, J.A. Alleman and L.D Taylor, Demand for Wireless Technology: an empirical analysis. Presentation to the 31st Annual Telecommunication Policy Research Conference, September 2003

3.9 CONCLUSION

This analysis has emphasised the difficulty of optimising the division of spectrum into licensed and unlicensed components. It does, however, suggest some general conclusions:

- Spectrum should be unlicensed where there is little probability of congestion; this applies predominantly to short-range applications.
- Restrictive allocations make the problem more acute.
- It is necessary but impossibly difficult to look ahead.
- Consideration has to be given to what happens if forecasts fail and congestion emerges.

Based on the overall assessment that the key to determining whether spectrum should be licensed is the probability of congestion we now go on to a more general discussion of the issue of congestion in spectrum.

4. THE LIKELIHOOD OF CONGESTION IN RADIO SPECTRUM

4.1 INTRODUCTION

Section 3 suggests that spectrum should be unlicensed where it is unlikely to be congested. This conclusion is widely supported by the literature summarised in Section 1. The logical next step is to determine the likelihood of congestion.

At present, congestion is generally defined as a situation in which there is more demand than the available supply. Using this measure, congestion is dependent on frequency and location. It is also time-variant, growing in some bands, decreasing in others. However, many of the commentators have suggested that this is an inappropriate definition because those holding the spectrum may be using it inefficiently. This judgement is made on the basis of measurement activity[8] showing that some fully-licensed bands have apparently little usage. Hence, they have concluded that if the licensing regime were changed to allow usage of the apparently unused spectrum then the pool of available spectrum would grow and the probability of congestion decrease. Easements and spectrum commons are both based on this logic.

In determining the most appropriate regulatory policy regarding unlicensed spectrum it is necessary to determine:

- Whether there is spectrum which is currently uncongested, can be expected to remain uncongested, and so could become unlicensed.
- Whether there is spectrum which is congested, but only because of inefficient usage, and where changing the management policy to unlicensed usage would remove the congestion.

In this section we do not seek to perform a band-by-band analysis, but rather set out the principles by which such an analysis could be performed. Firstly, we set out the general causes of congestion in order to understand which bands in general might be less congested. Secondly, we address whether moving towards unlicensed usage will reduce congestion. Finally, we suggest mechanisms whereby congestion can be reduced in unlicensed bands.

4.2 KEY FACTORS WHICH LEAD TO CONGESTION

There are many factors that influence congestion. Some of these are caused by suboptimal allocation policies and can be expected to be gradually alleviated by the introduction of trading. Others are caused by the nature of the radio spectrum. In essence, frequencies below around 100MHz have limited application because they propagate too far, preventing effective reuse. Frequencies above around 5GHz are also less desirable because propagation is too short. In between these areas, and particularly in the bands around 500MHz to 2GHz there is the greatest level of congestion.

---

[8] FCC Spectrum Policy Task Force, Report of the Spectrum Efficiency Working Group, November 2002, http://www.fcc.gov/sptf/files/SEWGFinalReport_1.pdf

There is little that the regulator can do to affect the relative desirability of these bands. However, there is one factor that the regulator can control which has a significant effect on congestion. This is the maximum transmit power.

The shorter the range of transmission, the lower the probability that there will be two users in range of each other that might interfere. For example, at one extreme, a person using a garage door opener with a range of 20m is highly unlikely to find another user of a similar device within the coverage area and operating their device simultaneously. At the other, in a cellular system with a cell covering a busy town, it is almost certain that there will be more than one person in the same cell transmitting at the same time during the peak hours. For a short range device with a maximum range of, say, 100m, the coverage area, and hence the probability of congestion, is only 0.04% of a cellular phone with a range of 5km.

Therefore, if only short range devices were allowed to use a particular piece of spectrum, the probability of congestion would be lower than for more general purpose spectrum. This would tend to favour unlicensed usage. Broadly, this has been the regulatory policy to date, with unlicensed spectrum having a maximum transmit power which tended to limit the range to around 100m.

The other factor influencing congestion is the bandwidth and time of transmissions. These mostly depend on the usage. For example, the garage door opener only needs to transmit a short burst of narrowband data and only on a few occasions each day. A W-LAN base station might transmit broadband data almost continuously. The probability of congestion is proportional to this time-bandwidth product or information rate. Historically, most short range devices have also had a low information rate, but more recently W-LANs and BlueTooth have changed this trend. If the unlicensed bands were restricted to products with a low information rate then congestion would be lower. However, it is quite difficult for the regulator to restrict the information rate in an unlicensed band – the only feasible way to influence this is to ban equipment with a broad transmission bandwidth.

Hence, the main tool at the disposal of the regulator in controlling the level of congestion and the suitability for unlicensed use is the maximum transmit power, which equates to the range. By enforcing a low maximum transmit power, the probability of interference is reduced. Further, the amount of usage will also likely be reduced as some applications will not be viable with short range transmissions. Regulators might have a number of different bands with different transmit power limits to offer users a different levels of range and congestion. Or alternatively, as an unlicensed band became more heavily used the transmit power might be progressively reduced to new entrants in order to maintain the congestion at an acceptable level.

Mesh networks have been proposed where the signal from a user is relayed by other users before eventually reaching a base station. It is not immediately obvious whether mesh networks qualify as short or long range communications systems since each “hop” might only be 100m or so, but the overall distance between the user and the base station might be 1km or more. In practice, they lie somewhere in between. A transmission of 1km made by 10 hops of 100m in length will result in a coverage area of 0.3km^2. A single hop transmission of 1km would result in a coverage area of 3km^2. By comparison, a single 100m transmission would have a coverage area of 0.03km^2. In practice, it is hard to see how to prevent mesh usage in unlicensed spectrum since it will operate within the transmit power limits. Equally, mesh usage has proved difficult to realise to date, especially in a mobile environment and it is far from clear that it will ever provide a cost-effective communications method.

4.3 SPECTRUM COMMONS IS UNLIKELY TO SIGNIFICANTLY ALLEVIATE CONGESTION

Historically, the number of applications and users of radio spectrum has grown faster than the ability of technology to accommodate them. Hence, congestion has increased over time. However, it has been argued that if a “spectrum commons” approach were widely adopted, then this would reduce the overall levels of congestion. This section considers whether this is likely.

It has been observed that for much of the time, some of the spectrum apparently goes unused. This has led to the proposal that radios be allowed to locate and hop onto temporarily unused pieces of spectrum and remain there until the owner of the spectrum wishes to make a transmission. We discussed this concept elsewhere, observing that in its simplest form it would not work because of the hidden terminal problem and that as a result some form of central management was required to tell terminals whether the spectrum was free and to grant them access[9]. Hence, we do not believe that such hopping behaviour, sometimes termed “SDR”, will work without band management, which in turn implies some form of band ownership, rather than unlicensed use. In this section we explore the concept of band management further.

The basic concept of band management would be to create a large pool of spectrum. Owners of a piece of spectrum that was put into the pool might have guaranteed access to an equivalent amount of spectrum. They might also receive some payment for the additional usage that occurred on their spectrum.

Pooling of spectrum is effective under two conditions:

- Individual holders of spectrum have insufficient spectrum to achieve good efficiency of usage.
- Different holders of spectrum have demand patterns that peak at different times.

Efficiency of use. The number of radio channels needed is generally calculated according to the Erlang formula. This shows that a few more radio channels are needed than the average demand would suggest in order to allow for demand peaks. The number of additional channels needed as a percentage of the overall number of channels falls as the total number of channels rises. This is shown in Figure 5, where the efficiency is compared with the number of channels for a 2% probability of blocking (a measure often used for cellular systems). The efficiency is the percentage of channels used on average compared to the total number of channels that needs to be set aside.

---

[9] M.Cave and W.Webb Designing Property Rights for the Operation of Spectrum Market, August 2003 http://users.wbs.warwick.ac.uk/group/common/publications/spectrum2

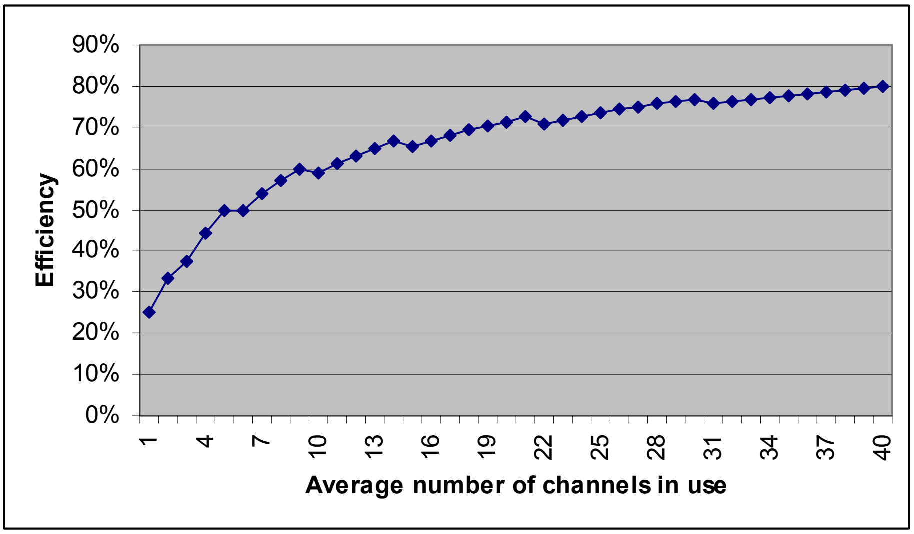

Figure 5 – Efficiency of use of radio channels

The slightly jagged nature of this chart is caused by the fact that there can only be integer number of radio channels. Although there is no clear point of inflection, the figure suggests that if the average number of channels in use falls below around 10, there might be significant efficiency gains from pooling channels with another operator.

In practice, most operators have many more channels than this. For example, GSM operators in the UK have around 1,000 voice channels each. Even allowing for re-use across cells this is still more than 100 per sector. Hence, in general, we do not expect to see strong efficiency gains.

Differing demand patterns: If one operator had peak demand in the morning and another in the evening then there would clearly be scope for improved efficiencies from sharing spectrum. The best way to understand whether demand patterns differ is to look at measurements of spectrum usage over time.

There are few published results of spectrum usage. The material here is taken from the FCC Spectrum Efficiency report[10]. This report made limited measurements and noted that many channels were lightly used, especially in the lower frequency bands (although not totally clear from the report, probably the bands below 500MHz). The report also noted that public safety usage of radio channels is often below 10% utilisation but demand can rise to over 100% of capacity (ie blocking occurs) during peak usage periods corresponding to major incidents. The report noted that its measurement results were preliminary and that more work was needed but suggested that they indicated that pooling spectrum was likely to bring benefits.

The report itself notes that its measurement method will tend to understate the usage of the spectrum because:

- A measurement made at a certain point might not be able to detect a nearby transmission if that transmission is behind a building. This becomes

---

[10] FCC Spectrum Policy Task Force, Report of the Spectrum Efficiency Working Group, November 2002, http://www.fcc.gov/sptf/files/SEWGFinalReport_1.pdf

particularly difficult when CDMA technology is used which results in relatively low signal strengths.

- Some operators use repeat patterns to avoid interference so that a particular frequency might not be in use in a cell, but might be used in neighbouring cells. This gives the impression of unused spectrum, but were the spectrum to be used then interference might result[11].

For all these reasons, measurements of usage are likely to return low results and need to be treated with some caution.

Even if the results are reasonably accurate, they suggest that spectrum is mostly under-utilised in the bands between around 200MHz and 500MHz. In the bands above this, corresponding to broadcasting and cellular, utilisation is much greater. This is unsurprising. In the 200MHz – 500MHz band there is a mix of military use, public safety use and private use. Because much of the private use is not trunked it is known to be relatively inefficient. Also, private use is tending to decline somewhat year-on-year with the result that the spectrum is gradually becoming less used. The inefficient usage may change once spectrum trading is introduced. The economic incentives might result in more trunked usage and perhaps spectrum being used for different applications.

Even if there is under-utilised spectrum in this band, it is of limited value to the operators in the bands that are more congested. For example, these lower frequencies cannot be used efficiently by the cellular operators in congested areas because the propagation is extensive, tending to generate interference across multiple cells and hence provide very little additional capacity. They cannot readily be used for broadcasting since broadcasting requires quite broad bandwidth channels (eg 8MHz wide for TV transmissions) which cannot be easily accommodated in the fragmented allocation pattern in the band.

It is also likely that much of the demand is correlated across operators. For example, all operators of cellular networks are likely to see similar demand patterns. Even emergency networks might see some correlation in the case where an incident causes disruption which in turn triggers mobile phone calls. Broadcasting networks have constant usage and so little advantage from pooling spectrum. Air traffic control, taxis, and other similar operators are likely to see peak usage around the same times as the cellular operators. Only, perhaps, military use might see uncoupled demand patterns.

There will be additional costs associated with pooling spectrum. User equipment will need to operate over a wider frequency range. Multiple control channels may be needed to inform the user equipment as to which frequencies are available which will themselves use spectrum and will require a constant stream of information to be passed between the various operating networks.

In summary, pooling of spectrum is a complex issue that merits a much deeper study. *Prima facie* the spectrum is likely to be more heavily utilised than measurements suggest and the difficulties in pooling are considerable. Hence, it remains unclear as to whether pooling is likely to be successful in reducing congestion. Our view is that at present there is insufficient evidence for the regulator

---

[11] This is a somewhat complicated issue. For example, it might be possible to use cellular transmissions in the unused TV transmitter bands because the cellular transmissions are at a much lower power level.

to embark on a process of converting spectrum currently considered as congested to unlicensed spectrum.

4.4 IN UNLICENSED BANDS, REGULATORY RULES HAVE AN IMPACT ON THE LEVEL OF CONGESTION

In addition to transmit power, there are some other rules which can impact the probability of congestion. These are:

- Restricting the type of equipment which can be used, which will tend to prevent the band being used for certain applications.

- Making the equipment more efficient so it uses less of the spectrum resource in transmitting its message.

- Making the equipment "polite" so that it does not transmit if doing so would interrupt on-going transmissions.

The first approach essentially blocks a particular application from unlicensed spectrum, or from some of the unlicensed bands. Such an decision would need to be made on the basis that allowing this application would likely reduce the overall utility from the band. In practice, it would be an engineering-based judgement that allowing the application would result in a high probability of congestion, or excessive interference to existing users.

The last two approaches will tend to make the equipment more expensive for no apparent gain for the end user and so will require regulatory intervention in the form of type approval or similar. Even so, there may be enforcement problems, particularly if the increase in the price of the equipment is substantial. In this case users may be tempted to acquire simpler, non type-approved equipment which might perhaps be legal in other countries. Because of the short-range and short-duration nature of most of the transmissions in these bands, enforcement could be difficult.

To date, the key regulatory mechanisms have been to restrict the equipment that can be used and to demand politeness. An extreme example of the former is the DECT band where only DECT equipment is allowed to operate. An example of the latter is the 5GHz unlicensed band where European regulators have required that equipment using this band has dynamic frequency selection (DFS) which seeks a lightly used frequency within the band before transmitting.

Without any regulatory intervention there will be a tendency for none of these mechanisms to be used. Equipment will only be made efficient or polite to the extent that it is necessary for that piece of equipment to operate reliably and not for the greater good of all the users of the band. An example of this is BlueTooth and W-LAN in the 2.4GHz unlicensed band. BlueTooth has been designed to use frequency hopping which reduces the impact of interference on its operation. However, it also tends to increase the interference generated to systems which do not frequency hop like W-LAN. Studies have suggested that if both technologies operate in proximity then the BlueTooth system will work whereas the W-LAN system may stop functioning. Only with regulatory intervention has the BlueTooth standards body agreed to a modification in the standard whereby if a BlueTooth device senses the presence of a W-LAN transmitter it will not hop onto the frequencies currently being used by the W-LAN node.

By regulating the usage of the band the onset of congestion can be postponed but at the cost of increased equipment prices. From a theoretical point of view the optimal point is that at which the increased value of the usage of the band less the increased equipment cost is maximised. Practical factors related to enforcement also need to be factored in.

A further complication is that the increase in cost may depend on the device. For example, adding additional frequency hopping rules to a BlueTooth device which is already built around a complex integrated circuit will have minimal cost impact. Adding the same rules to a garage door opener which does not have sophisticated electronic circuitry could require a complete redesign with a much larger integrated circuit, significantly increasing the price. Equally, the potential for the garage door opener to generate interference is much less than the BlueTooth device because of the relatively infrequent usage of garage door openers. This might suggest different levels of regulatory intervention for different classes of devices, depending on the likely usage and cost increase.

By analogy, roads use some of these mechanisms to maximise their capacity. Certain types of vehicles are not allowed on roads, or are restricted to certain parts of the road – for example lorries are often not allowed in the outside lane of a motorway. Cars which are efficient in their use of space through being small are sometimes given tax incentives. Finally, politeness protocols are very widely applied, from deciding which side of the road to drive on through to regulating behaviour at traffic lights.

4.5 CONCLUSIONS

In the previous section we concluded that spectrum should be unlicensed if it was unlikely to be congested. As a result, in this section we considered the likelihood of congestion. We noted that:

- Congestion was most likely in the core bands of around 100MHz to 5GHz.
- There was insufficient evidence that taking bands currently considered to be congested and making them unlicensed would alleviate congestion hence this approach cannot currently be advocated.
- The probability of congestion could be dramatically reduced by restricting the range of devices through controlling the maximum transmitted power or by requiring specific behaviour such as politeness protocols.

However, there is no definitive way to predict congestion. A judgment needs to be made on the basis of the frequency band, likely use and range. The range in turn depends on the use. Hence, a key stage in predicting the congestion likely in the band is determining the most likely use. We discuss how this might be achieved in subsequent sections.

5. HOW THE REGULATOR CAN DECIDE ON UNLICENSED USAGE

5.1 INTRODUCTION

When the regulator has to make a decision it will invariably be a matter of judgment. However, we suggest the process set out in Figure 6 to help guide the decision. We emphasise that, despite the use of a flow-chart, this is not an exact science and determining the correct answer is still likely to be difficult. In outline our process consists of three key stages:

- Determine the most likely use of the band.
- Decide whether this use is best facilitated with licensed or unlicensed spectrum.
- If unlicensed, determine which regulatory restrictions should apply.

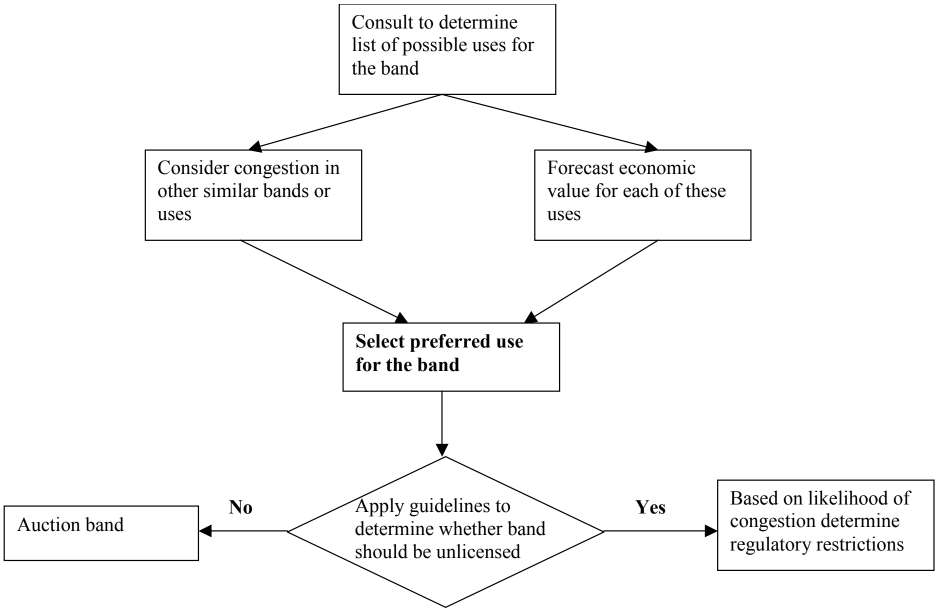

Figure 6: Outline of the process that a regulator might follow

Each of these stages is discussed in more detail below.

5.2 CONSULTATION

Consultation is already widely used. We envisage a similar process to that currently deployed. Although it may be unlikely that potential users of unlicensed spectrum would respond to the consultation, manufacturers of devices might. A specific question on the consultation document asking about unlicensed usage might yield important answers.

5.3 ECONOMIC ASSESSMENT

The optimum allocation of spectrum would be the one that resulted in the greatest economic value for the country. In general, market-based methods of allocation and assignment will provide this outcome, but as we argued above, the market may not be able to reliably allocate unlicensed spectrum. An alternative is to attempt to predict the economic value of each of the different plausible uses and then to favour the use with the highest expected value.

Economic value assessment has been widely used in the UK, with the RA issuing or updating economic value assessments on a near-annual basis. However, this assessment is retrospective, measuring the value of applications already in use. In making an allocation decision it is necessary to perform a forward-looking value assessment.

Forward looking assessments follow the same methodology as retrospective assessments. The difference is that the key input data such as numbers of users and value of equipment needs to be forecast rather than observed. Hence, the major workload becomes one of forecasting. It is in this forecasting that the inherent problems with this approach reside. Forecasting future demand for wireless is notoriously difficult because of the changing applications and technologies in this area. For an economic assessment it might be necessary to forecast more than five years out. Because such forecasts have a low probability of being accurate, this economic valuation can only be seen as approximate. Were it likely to be highly accurate it would be the only input needed into the regulatory decision but because of the likely inaccuracy it should only be one of a number of inputs into the regulatory decision.

The economic assessment must also take account the irreversibility of particular decisions. Essentially, it is much easier to covert licensed into in unlicensed spectrum than vice versa. This is because unlicensed users are anonymous and continue to use the spectrum when it has been assigned to other purposes. The economic assessment should take account of such risks. This can be done by the techniques of option pricing[12]

5.4 EXAMINE SIMILAR BANDS AND USES

The regulator might be able to learn from related occurrences in nearby bands. For example, if a potential use is fixed links, but a neighbouring band has already been allocated for fixed links and is under-used, then the use of the new band for fixed links might be given a downwards bias when compared to other uses. If usage is growing rapidly in unlicensed bands elsewhere in the spectrum then a bias towards making the band unlicensed might be appropriate.

5.5 DETERMINE WHETHER UNLICENSED USAGE IS APPROPRIATE

Based on the three processes outlined above, the regulator should be able to come to a conclusion as to the most likely use or uses for the band. The regulator does not need to impose these uses – for example if the band is subsequently auctioned there is no need to restrict its use to that deemed most likely. However, this decision will be used in the process of deciding whether spectrum should be unlicensed.

---

[12] See A. Dixit and R.Pindyck, Investment under Uncertainty, Princeton University Press,1994

Having decided on the most likely use, the spectrum should be subject to licensing where any of the following hold true:

1. The band is likely to be congested. A way to approximate for this is to assume that congestion would occur if the use would entail a wide area service (ie one covering a contiguous area greater than ~1km^2) being offered. Examples of such services are cellular and broadcasting.

2. A guaranteed quality of service (QoS) is needed. This is the case, for example, with most public safety communications.

3. International treaty obligations provide restrictions that would be breached by operation on a licence-exempt basis either now or at some known point in the future.

Each of these points is now considered in more detail

5.5.1 Wide area coverage

Interference in licence-exempt spectrum is generally acceptable because the transmitters are low power. As a result, the area that they interfere over is small, reducing the probability that there will be another user in the same area. The smaller the coverage area, the lower the likelihood of interference. Historically, transmit powers in licence-exempt spectrum have been restricted to levels that result in a coverage range of around 100m, although this is highly variable depending on propagation conditions.

Restricting the range to less than 100m would significantly reduce the attractiveness of the band as it would no longer be possible to provide systems that could conveniently cover an average home or office building with a single transmitter. Hence, we believe this is a sensible lower bound for the range.

Increasing the range beyond this would not appear to enable many new applications until the range became in excess of 1km. At this point, fixed wireless applications and city cellular systems become viable. However if a 1km range is allowed in an unlicensed band then the interference will likely become unacceptable.

Therefore, we conclude that operation on a licence-exempt basis should be restricted to a transmitter power that results in a range of up to 100m under typical usage conditions.

5.5.2 Quality of service

It is not possible to guarantee the interference levels that will be experienced in licence-exempt spectrum. Therefore, its use is inappropriate for communications that require a maximum level of interference.

5.5.3 International treaty obligations

Most spectrum bands have their allocation agreed at an international level. In some cases, this restricts the ability of any particular country to change the allocation. In general, a move to licence-exemption would be unlikely to generate interference into other countries because of the low power levels associated with this use. However, this firstly needs to be clarified to ensure no treaty obligations are over-ridden.

International agreements are often signposted well in advance. One of the well known issues with allowing licence-exempt use of a band is that it is very difficult to reclaim the band should it be required to change the allocation. This is because there is no record of the user base so it becomes difficult to inform users of the change and difficult to police interference that might result. Hence, in the case of a band where an international obligation is known to apply at some point in the future which would contradict with licence-exempt use it would be problematic to allocate it to licence-exempt use in the interim.

5.6 DECIDING ON REGULATORY RESTRICTIONS

If the band is to be unlicensed then the regulator may wish to impose restrictions as discussed in Section 4.4. The regulator will need to make a judgement as to the most appropriate level of restriction.

In outline, the greater the perceived risk of congestion developing, the more restrictions should be imposed. However, the restrictions should also take into account the likely additional cost imposed on the devices compared to the benefit that might accrue. Depending on the level of information, it might be possible to perform an economic assessment of the value of the different approaches.

For example, where imposing politeness protocols will have minimal impact on the device cost then they might be used without hesitation. Where such protocols would significantly increase the cost and where congestion is unlikely, or has little impact, then they should not be imposed.

The regulatory process is now complete. The process should be repeated periodically, and not just when spectrum is being auctioned as it might be appropriate as technology or applications change to acquire spectrum through trading and make it unlicensed.