KUNGL TEKNISKA HÖGSKOLAN
Royal Institute of Technology

On the feasibility of unlicensed operation in shared spectrum

Olav Queseth

Radio Communication Systems Laboratory

KUNGL TEKNISKA HÖGSKOLAN
Royal Institute of Technology

On the feasibility of unlicensed operation in shared spectrum

Olav Queseth

TRITA-S3-RST-0307
ISSN 1400-9137
ISRN KTH/RST/R—03/07--SE
Radio Communication Systems Laboratory
Department of Signals, Sensors and Systems

Abstract

Whether unlicensed operation is a feasible mode of operation in future infrastructures depends on a number of factors. Here a comparison is made between systems with exclusive licenses and cases where more than one system uses the same radio spectrum. The comparison is made in terms of capacity, the ease with which a new operator can establish itself in the market and if there are benefits in breaking the etiquette rules.

Here two systems coexist in the same geographical area. Various system designs are evaluated utilizing different multiple access schemes (DS-CDMA, frequency hopping and dynamic channel allocation) as well as different traffic types (voice and data). Given specific requirements for user satisfaction and fraction of satisfied users the load each system can carry is evaluated. The systems using orthogonal access schemes perform significantly better due to the near-far effect.

To get better performance in technical terms compared to traditional spectrum splitting by using spectrum sharing seems difficult. The technical solutions used here do not perform better. However the technical performance is not significantly worse either which opens up for gains in other areas than the purely technical ones.

Acknowledgements

This work was funded by PCC for which I am most grateful.

In addition I would like to thank advisors, colleagues, friends and family for support.

Contents

1. Introduction

This thesis is a part of the large puzzle in Swedish research called Personal Computing and Communication (PCC). The vision of PCC is "Personal multimedia communication to all at the same cost as fixed telephony today" and one of the overall aims is to strengthen Swedish industry in the mobile area. To be able to do worthwhile research we used scenarios to be able to understand future developments in this industry. One of the areas that we identified as important is radio communication using unlicensed spectrum.

The demand for wireless connectivity is increasing. The data rates available in the fixed networks increase which enables wider spread of bandwidth hungry applications. At the same time people are more mobile and they want to have the same services available everywhere. As more people want to use a wireless connection more people are also interested in providing wireless connectivity. For example, people want to set up their own home wireless LANs, owners of airports may want to provide their customers with wireless internet access in gates and lounges. Internet cafes may want to extend their offering to include "bring your own (computer)".

The scenarios also state that unlicensed operation is the main mode of communication. The implication of this is that there are firms that make money from providing communication. Note that it is not necessarily the end users that pay for the transmission directly. They may pay for services, but somehow in the value network there is a firm that make money from providing infrastructure. There are initiatives for providing infrastructure where individuals share their infrastructure on a voluntary basis and in return they get the possibility of using infrastructure belonging to other individuals [1]. However this kind of sharing rarely becomes more than something enthusiasts do.

The necessity of firms determines the focus of this thesis. It is not enough to find out if unlicensed bands can be used. It is also necessary to determine if money can be made. This makes this thesis integrative in nature.

Currently most operators use spectrum that is licensed in some way. A license is a permission from the (national) licensing authority to operate a radio transmitter. The license is typically time limited and specifies the characteristics of the transmitter and the purpose the transmitter is used for

[2]. National borders delimit the geographical areas licensing authorities have jurisdiction in. But national borders do not stop radio waves and international cooperation is necessary.

When radio communication started to be used for more than experiments problems with interference quickly rose. The remedy at that time was careful planning. Radio transmitters were only allowed to use specific frequencies at specified locations. The operations license was born. Since then technological advances have allowed us to utilise more radio spectrum, i.e. user higher frequencies, and also make better use it. However the main tool to ensure that communications is reliable and to avoid interference is planning in advance. The demand for radio spectrum is almost always larger than the "supply", but constant negotiations and compromises still make it possible limit the influence of interference.

However there are problems with the current handling of licenses. One problem with the current method of licensing is the planning procedure, or rather the effects of the planning procedure used. It is complex to handle and as more and more services should be considered when planning the complexity increases. As the telecommunication markets are deregulated the complexity is further increased. The reason is simply that instead of letting one operator provide a service a number of operators should provide the same service and each operator has to be given an individual license. The current situation on the market is that there are a few actors. If competition is to increase there will be more actors and this will make the licensing procedures even more complex.

The licensing procedure is slow compared to the changes in the telecommunications markets. The payback time on infrastructure is often fairly long. This means that the in order to make it economically feasible to build infrastructure the operator must be able to utilise the radio spectrum for a long time. Thus the license has along running time. Which of course means that there is usually takes time before a piece of spectrum can be put to other uses.

When changes in the telecommunications market were not as quick this was not a large problem. However with the more rapid changes on the market there is a mismatch between usage and allocation. A successful operator may make a lot of use of his spectrum while another who has failed to attract customers may not be utilising his spectrum at all. This situation is wasteful since the spectrum is just sitting there and could be put to better use by the first operator. If the allocation mechanism can follow the changes in use this situation will not occur, but since markets change quickly the licensing authorities cannot follow.

If we should give out more licenses and there is a fixed amount of spectrum available that means that there will be less spectrum for everybody. The problem is that in this case the sum of the parts is less than the whole. On other words: the total traffic that can be carried over the pieces is less than what could be carried over a large chunk of spectrum. This effect is known as trunking losses and has to do with the randomness of the traffic. Since the traffic varies there is always a certain extra capacity to make sure that the traffic peaks can be handled. When many users share a resource the variations among them will even out thus allowing a smaller safety margin. Another problem is that high data rates require large pieces of spectrum that can be used concurrently.

The advantage with licensing is that some planning and handling of interference can be done beforehand. Thus actually using the system becomes simpler. Which in turn means that the end user equipment becomes simpler and thus cheaper. This was a sound engineering decision when end user equipment was expensive and changes could only be made manually. However it is now possible to produce complex devices with a lot of intelligence at a low cost. Thus some of the planning can actually be done later. The benefit is of course reduced complexity at the planning stage. This type of design choice is used for example in DECT and IEEE 802.11 WLANs.

Planning beforehand has usually been made to ensure high quality communications almost anytime. The result is a lot of resources that are just sitting there most of the time. Even if this is not a problem specifically related to planning beforehand the planning procedures have been influenced to a great extent by the quality requirements. This has been perfectly reasonable when the only service was voice telephony and the service quality was compared to fixed telephony. However with more intelligent terminals and different types of services without strict real time requirements, e.g. email downloading, the requirements on the quality of service from the actual communication link may not be as strict as before. The terminals are able to compensate to provide the same perceived quality. This may allow for a different trade-off between quality of service and installed infrastructure. By reducing the quality of the communication links less infrastructure is needed which results in lower costs. There is nothing in the licensing procedure that by necessity requires a high quality, but the paradigm may be hard to change.

Unlicensed spectrum offers a remedy to many of the problems. Since there are no licenses the problems with long and complicated license procedures disappears. But even if there are benefits with unlicensed operation there may be other problems. Since the technology is fairly new there may be a reluctance to use new unproven technology by the operators.

Lately the trend has been to release more spectrum for unlicensed use. It seems like the regulators believe that unlicensed spectrum can provide lower prices and better use of the spectrum.

1.1 Problem

In this thesis we study the feasibility of using unlicensed spectrum or license exempt spectrum for providing communication services.

To operate a transmitter in an unlicensed or license exempt frequency band no license is required. However the transmitter has to adhere to certain rules. For unlicensed bands the rules are simple and limit things like transmission power or transmission time, these rules are known as etiquette rules. In license exempt bands the rules are more complex and typically require a transmitter to adhere to a specific standard. Furthermore we also mean that a transceiver can be connected to some kind of fixed infrastructure. A number of transceivers connected to the fixed infrastructure then form a network that makes it possible for a user to connect to the fixed infrastructure without using wires.

We assume that in the fixed infrastructure there are a number of services that the user wants to get access to. In general this could be anything, but examples such as real time voice connections, video and file downloading come to mind. To be able to do this data bits has to be transferred between an access point and the user. Providing communication services is the same as making it possible for the user to get his bits.

That unlicensed spectrum can be used for communication is reasonably established. This is shown by the wide spread of IEEE 802.11b wireless networks. However so far the networks are only deployed on a small scale to achieve coverage over small areas. At the same time most of the time there is only one network in any given area. It is not clear what will happen when the networks expand or what will happen when they overlap.

In the end we want to determine if it is possible to provide coverage over large areas using technology that operates in unlicensed spectrum. We believe that a network with fairly large area coverage cannot be built using voluntary non-profit contributions. There must be an operator that provides users with communication services and get paid for that. To be able to do this the operator either builds access points himself or he rents capacity in other networks.

In this thesis unlicensed operation is considered feasible if an operator has a fair chance to be successful in the marketplace given that he chooses to utilise technology using unlicensed spectrum. Unfortunately this is not easy to do. In fact in this thesis we will not even try to have a look at the whole picture. However we know that the available technology and the technology choices influence the operator performance.

In this thesis we study systems that coexist in the same geographical area and in the same spectrum. There are numerous things that can be studied, but in this thesis we pick three performance measures in an ad-hoc fashion. The important aspect when picking the performance measures is that they should be able to interpret in an economic context. The results could then be used as input when studying the behaviour of the operators in the marketplace. However studying the behaviour of the operators is left out in this thesis.

The measures we pick here is the cost of providing services both. We study both the influence of traffic load for the operators on the cost and also the influence of relative size of operators on the provision costs. In addition we study if it is possible to gain benefits (at the expense of other operators) by not following the established rules.

1.2 Thesis contribution

There are three main areas where this thesis makes a contribution.

The first is to study the capacity of two networks interfering with each other. Previous studies have mostly focussed on the influence of interference on one network. The total capacity, i.e. how well spectrum is utilised by two non-cooperative networks has not been studied before.

The interpretation of the results is done in a novel way. Instead of just determining the capacity of two networks the combined behaviour of these two systems are taken into account when determining how firms will act in the marketplace. Thus we make the relation between behaviour of a technological system and operator behaviour explicit. To be able to do this cross-disciplinary study the methodology is also developed which is a contribution as well.

In almost all studies the systems are assumed to follow the rules set for their behaviour. However in here we take a look at what happens when these rules are broken. This thesis outlines a framework for analysing this rule breaking behaviour and also gives some examples of what happens when one does.

This thesis is based on the following publications:

"On competiton in unlicensed networks" 57th IEEE Vehicular Technology conference, April 2003, Jeju Korea.

"On the performance of unlicensed data access systems" Nordic Radio Symposium 2001, April 2001.

"On the performance of Coexisting Spread Spectrum Systems" IEEE PIMRC 18-21 Sept 2000, London, UK

"Coexistence in Spread Spectrum Systems" PCC Workshop 99 Lund, Nov 1999

In addition the scenario work has been reported in the following publications:

"4th-Generation Wireless Infrastructures: Scenarios and Research Challenges" I EEE Personal Communications Magazine, December, 2001.

"An approach to 4th generation wireless infrastructures Scenarios and key research issues" VTC 99 Houston, TX, USA, May 1999

"Telecom Scenarios 2010 - a wireless infrastructure perspectives" S3/KTH, Stockholm16 Nov 1998

“Key Research Issues in 4th Generation Wireless Infrastructures” PCC Workshop 98, Stockholm1998

“Telecom Scenarios for the 4th Generation Wireless Infrastructures” PCC Workshop 98, Stockholm1998

“Scenarios-A tool for starting a research process” PCC Workshop 98, Stockholm1998

1.3 Related work

Three groups of papers that study unlicensed systems can be distinguished. One group of papers studies the performance and behaviour of these systems. The research has typically been done during standardisation and development of equipment. However many times these studies does not take into consideration more than one system operating at one time.

There is another group of papers that study coexistence in a more general context. This type of research is typically done before a piece of spectrum is released for unlicensed operation. A common question is how system behaviour is affected by etiquette rules.

A third group papers do not focus on coexistence aspects of radio systems. However the problems studied are similar to the ones studied in coexistence research. The results in these papers can be interpreted in a coexistence context.

1.3.1 System related research

DECT (Digital European Cordless Telecommunications) is a system designed to provide short-range voice communication. Typically a DECT system is deployed in an office environment or at home with one access point and a few handsets [3].

DECT is designed to be able to operate without the intervention of licensing bodies. There is a specific band allocated for DECT systems, no other systems are allowed to use that particular band. However there is no need to obtain a license to operate a DECT system. This makes DECT a license exempt system. In order to solve the frequency allocation problem DECT uses a dynamic channel allocation.

The performance of DECT and the DCA algorithms has been studied. The focus is to study the quality of service for a given traffic load. The traffic is assumed to be voice traffic and the quality of service is measured as blocking probability and signal quality for ongoing calls. The traffic is assumed to be voice traffic. However only one system is studied, i.e. there is no interference from other systems and a user can connect to all access points. Various algorithms for the dynamic frequency allocation have been tried. For example autonomous reuse partitioning (ARP) where a call is assigned to the first available channel that passes certain quality tests. Another algorithm is the least interfered channel algorithm (LIC) where the channel with the highest SIR is assigned to a call. The LIC algorithm provides higher quality of service i.e. SIR, but the ARP algorithm can support higher traffic loads [4] and [5].

Personal Handy Phone (PHS) and Personal Access Communication System (PACS) are two systems that are similar to the DECT system. The envisioned usage is voice communication and as the technical solutions are similar to DECT [6]. These systems also rely on some form of dynamic channel assignment to avoid interference and the need to plan. When searching the literature it is possible to find performance studies similar to those for DECT. But there seems to be no studies where interference from other systems is considered.

Hiperlan/2 is a standard for wireless LANs. It is intended for indoor or short-range communications with data rates up to 54 Mbit/s [7]. It operates in the license exempt band around 5.2 GHz. To avoid interference an adaptive channel allocation algorithm is used.

There have been a number of investigations of the performance of Hiperlan/2. The focus is to find the throughput for individual users and for an entire Hiperlan/2 system, e.g. [8] and [9]. However these studies only take into account one system. I.e. how well two systems perform together is not considered.

Part of the band allocated for Hiperlan/2 is also allocated for radars of various kinds. How these different systems coexist has been studied. The results indicate that Hiperlan/2 does not suffer any major performance degradation because of radar interference [10].

The radar community is worried that Hiperlan/2 systems will create disturbances to the radars. Studies have been performed to see how much radars are interfered. Studies have also been performed to see how the Hiperlan/2 standard should be changed to detect and avoid interfering with radars [11].

Bluetooth is a system for short-range radio communication with data rates up to 1 Mbit/s [12]. The usage is mainly intended for connecting a few devices together in an ad-hoc fashion. These devices may for example be a mobile phone, a laptop or wireless headphones. These devices are connected together in something called a piconet. Piconets are physically small and consists mainly of personal devices.

Bluetooth operates in the unlicensed 2.4 GHz band. Frequency hopping is used in combination with a 2/3-rate block code and/or selective retransmissions to combat interference. The performance of Bluetooth networks that interfere each other have been studied e.g. [13] and [14]. Here the throughput in the piconets has been studied for two different traffic cases, fully loaded piconets and for WWW traffic. The scenario studied is a room where there are a number of piconets scattered.

The conclusion is that the performance is not affected until there are a high number piconets in the room. The results may be explained by the way the devices in a piconet are located compared with the location of the piconets. Generally the distance between the transmitter and receiver is much smaller than the distance to an interferer.

IEEE 802.11 is a standard for wireless LANs, or to be exact a family of standards. The most popular is 802.11b. The devices operate in the 2.4 GHz unlicensed band. To combat interference DS-CDMA is used. Although there have been a number of studies of the performance of IEEE 802.11 networks [15], [16] and [17] there have not been any studies where there are more than one network operating in the same geographical area. The implicit assumption is that all the access points in a geographical area belong to the same network. This may be a reasonable assumption since this is the way wired infrastructure is laid out. There is rarely more than one wired network in any area. If there is a requirement for privacy that is assumed to be solved at higher protocol layers.

IEEE has stared a group to study the impact of Bluetooth on IEEE 802.11b WLANs and vice versa. The results from the working group has not been reported in the literature to a great extend. However some studies that study the impact on a single WLAN link of a Bluetooth network [18] has been performed. The indication is that there will be a substantial loss in the capacity of a WLAN when Bluetooth is used in the same area. However studies of the system level capacity has not been reported (or at least not found by the author of this thesis).

Recently there has been an increasing interest in a multiple access technique called pulse position radio or impulse radio [19], it is also known as ultra wideband (UWB). It is a form spread spectrum communication. There are some studies on the capacity of such systems, but they generally consider only one cell with perfect power control [20]. With these assumptions thousands of users can be accommodated in a cell.

It is not easy to judge from these results how pulse position would perform in a scenario with more than one operator. Since the bandwidth used for communication is really large there is a lot of radio resources available. There are also claims that the system can be designed to handle near-far effects. Thus the performance in systems that coexist may be good.

1.3.2 Coexistence from a more general perspective

There are a number of papers that deal with the usage of unlicensed spectrum. But the papers do not focus on a specific standard. For example the etiquette rule listen before talk is studied in the paper "An evaluation of Traffic throughput in the Asynchronous UPCS band" [21]. Both voice and data traffic is studied. The results indicate that it is hard to guarantee the quality of voice connections, but moderate data rates can be achieved. Another example is the paper "On the feasibility of a CDMA Overlay for Personal communication Networks" [22]. Here the authors study if a cellular CDMA system can coexist with microwave links in the frequency band around 2 GHz. The authors find that coexistence is indeed feasible.

Yet another paper studies quality of service i.e. blocking probability in a cellular system where there four operators that coexist [23]. The mechanism for avoiding interference is dynamic channel allocation. The results show that the capacity of the system is approximately the same for the case when there is only one operator and four operators. Although in the 4-operator case four times as many access points are used.

Although the research mentioned here does not focus on a specific standard the papers in some sense looks how to design a system given that there are a set of rules already in place that governs how a piece of spectrum may be used. Thus the focus here is to evaluate various technical solutions of spectrum sharing. The constraints in assumed to be given already. This could be called the engineers view of unlicensed operation.

However there is also a body of research that has been performed in a different context. Regulators make decisions on how spectrum should be allocated. Although there may be many interests the regulator has to consider at least there are interests and they do not overlap with the interests of other actors on the scene. To be able to make well-informed decisions there are studies performed with the interest of the regulators in focus. These studies try to determine if unlicensed operation is feasible or they try to determine how the etiquette rules should be designed. A survey of some of the issues facing a policy maker is the paper "Spectrum Management Policy Options" [24].

One problem that has been identified in unlicensed operation is the problem the "tragedy of the commons". In short this is the problem that greed benefits a single user. If one user is greedy and for example uses higher transmitter power or keeps a channel even if there is no communication that user will benefit at the expense of other users in the system. But if all users behave in the same manner everybody looses. In unlicensed spectrum there is no incentive to not be greedy since using the spectrum is free. Thus if the spectrum etiquette rules are not designed in a proper way there is a risk that greedy users are rewarded. In the papers "Performance Of Unlicensed Devices With a Spectrum Etiquette" [25] and "Etiquette modification for Unlicensed Spectrum: Approach and Impact" [26] the authors discuss this problem and also proposes various solutions. In general the solutions are based on some kind of penalty or cost for using the spectrum, for example a device must wait a certain time after transmitting before it can transmit again, this waiting period increases when a device transmits longer time. The conclusion is that a penalty function can discourage users to be greedy, but there is some performance loss.

1.3.3 Research on similar problems

The technical problems that arise when studying the behaviour of coexisting systems are similar to other problems in the area of radio resource management. One thing that makes unlicensed operation problems unique is that there may be more than one system operational in the same geographical area; another is that the interference experience by a system cannot be controlled. These kinds of situations are also found in other radio resource management studies. Thus we can look at other studies that focus on other issues, but where the results can still be interpreted in an unlicensed operation context.

Hierarchical cell systems have been studied to some extent in the context of current cellular systems, which are mainly aimed at voice services. In a hierarchical system there may be more than one layer of access points covering the same area. For example there may be one layer with large cells that create coverage over a large geographical area. In addition a second layer with smaller cells may be used to satisfy high local capacity demands.

Some of the problems studied are the interference between the layers [27]. Another problem is how frequency spectrum, should be shared between the layers. This can be viewed as two systems that coexist. The difference between this problem and the problem in this thesis is that one layer (i.e. system) has a much higher access point density than the other layer. In addition it is possible to coordinate the radio resource management in both layers e.g. by making handovers between the layers. However the methodology used here to depict the capacity of both layers have been used in this thesis as well.

In military applications there is always a risk that the enemy will try to block communication by creating interference. To mitigate this, spreading techniques e.g. DS-CDMA or frequency hopping are utilised. This situation is similar to unlicensed operation in the sense that the interference cannot be controlled.

1.3.4 Previous work in relation to this thesis

In this thesis we use many different systems for performing our experiments. The designs of these are all taken from the literature. References to models and algorithms are given in chapter 4. The idea has been to pick elements that perform reasonably well over a wide range of cases. At the same time we avoid designing systems that need excessive amounts of computation.

To make system design choices a lot of inspiration has been received from the system designs that have been made to operate in unlicensed spectrum, for example Bluetooth, Hiperlan/2, IEEE 802.11x, DECT and so on. But other methods for handling interference have also provided inspiration, e.g. frequency hopping is borrowed from military communications and GSM.

Parts of the thesis rely on microeconomics to explain the business behaviour of the operators. The methods and measures used are taken more or less straight from the textbook, i.e. from "Microeconomics" [28]. The theory here is not new, but the application of it is novel.

The method with feasible regions is borrowed from the studies of hierarchical cellular systems, e.g. from [27].

2. Unlicensed operation

2.1 Regulation

Marconi could probably not have imagined the bureaucracy that has emerged 100 years after his first radio transmission from Cornwall to Newfoundland [29] in 1901. Following his first transmission radio communication quickly became popular both for maritime communication and for broadcasting. But increased use also resulted in increased interference. To remedy the situation national legislation was passed. However national borders do not stop radio waves, and it quickly became apparent that some sort of international cooperation was needed. In 1927 the International Radio Consultative Committee (CCIR) was established. The same year frequency plans were made to ensure greater efficiency of radio operations [30].

In 1932 CCIR was merged with a number of organisations, notably the International Telegraphic Union (CCIT) and other consultative committees on different aspects of telecommunication to form the International Telecommunication Union (ITU). In 1947 the Union was made a specialised agency of the United Nations [2].

Today ITU has almost 800 employees. The headquarters is located in Geneva and there are 11 regional offices. The oldest activity of the ITU is to develop internationally agree standards and to define tariff and accounting principles for international services. Nowadays ITU also facilitates radio spectrum management and to help developing nations to create infrastructure for telecommunication [31].

One of the responsibilities of ITU is maintaining the international Radio regulations (RR). The RR is a document that details how different pieces of spectrum can be used, for example certain bands are allocated for mobile communication, broadcasting or medical and scientific uses. Some bands have so called shared use. This means that different services can use the same band for example radars and Hiperlan/2 have allocations in the same band. When bands are shared certain services have priority, i.e. other services may not interfere with the service with priority.

It is the responsibility of the national regulator to decide the exact use of the frequencies. They typically give out a license, which is a permission to operate a transmitter. The license will typically specify the characteristics of a transmitter, i.e. the owner, the location, the purpose of the transmitter and the period of validity of the license [2]. For example "Post och Telestyrelsen" (PTS) in Sweden has given Vodafone the permission to use the frequencies 1920,3 - 1935,3, 1915,0 - 1920,0 and 2110,3 - 2125,3 MHz until March 31 2006 to provide third generation mobile services [32]. Radio waves do not stop at national borders though and in some cases, e.g. broadcasting, bilateral agreements between neighbouring states are made to avoid interference. The national regulators have agreed to follow the international radio regulations. This may at some times be inconvenient, however the benefits from having a global agreements are much larger.

The RR are updated and amended during a world administrative radio conference (WARC) where the national regulators participate. These are typically held every couple of years. The preparations for a WARC include preparatory conferences and possibly also technical studies. In other words changing the radio regulations is a lengthy process, which makes changes to spectrum allocations inherently slow.

The regulators usually strive to ensure that the radio communication and other services are available to the public with an appropriate quality, availability and price [2]. This may sound simple but there are a number of factors that have to be weighed against each other to do this. Currently regulators seem to believe that these goals can be achieved better by increasing competition. Thus the trend during the last decades there has been a deregulation of the telecommunications market. Traditional state owned monopolies have been split up and other actors have entered the telecommunications market.

From microeconomic theory we learn that competition generally results in lower prices and higher volumes [28]. However in a market with competition there will also be a (wasteful) duplication of resources. One example is from the telecommunication market in Stockholm around the previous turn of the century. There were two competing telephone companies that for various reasons did not want to route a call from one network to the other. The result was that many subscribers had two telephones, one for each network [33]. This is an example of duplication and a waste of resources that the regulator tries to avoid.

In a mobile radio access network it is not clear that creating competition will result in lower prices. If the regulator decides to create competition by splitting the spectrum into smaller pieces and allowing each operator to use one piece more access points are needed to provide services to the same amount of users. I.e. the production cost increases.

Consider the following simple example. In a very simplified model the capacity of a cellular network is proportional to the available bandwidth and the number of access points [34]. Assume that the available bandwidth is B and that a monopolistic operator builds P_M access points. The capacity C_M of this network is thus proportional to P_M x B.

$$C_M \propto P_M B \tag{2.1}$$

Now assume that instead we have a market with N operators. Each is allocated a part of the spectrum: B_C. For simplicity we give all operators the same amount of spectrum.

$$B_C = \frac{B}{N} \tag{2.2}$$

We want to spend the same amount of money as in the monopoly case to make comparison easy, in this simple example this corresponds to building the same total amount of access points. Thus each operator builds PC access points.

$$P_C = \frac{P_M}{N} \tag{2.3}$$

However the capacity of the network C_C of one operator becomes:

$$C_C \propto P_C B_C = \frac{P_M}{N}\frac{B}{N} \tag{2.4}$$

The total capacity of all networks is then:

$$C_{CT} = C_C N \propto \frac{P_M B}{N} = \frac{C_M}{N} \tag{2.5}$$

This can also be expressed so that the total cost of a specific capacity is proportional to the number of operators.

There are many things this simple analysis does not cover. For example an access point capable of using a larger bandwidth will probably cost more, trunking effects are disregarded, there are probably volume discounts on access points and so on.

Despite this the regulators have most of the time chosen to award more than one license for cellular radio networks.

2.2 Unlicensed operation and license exempt operation

The general rule set by government agencies is that in order to operate a radio transmitter the owner of the transmitter must have a license, i.e. permission from the government. However there are exceptions to this rule. In some portions of the radio spectrum it is possible to operate a radio transmitter without permission. But even though there is no need to have a license there are a number of rules to follow. These rules are different for different parts of the spectrum. [35]

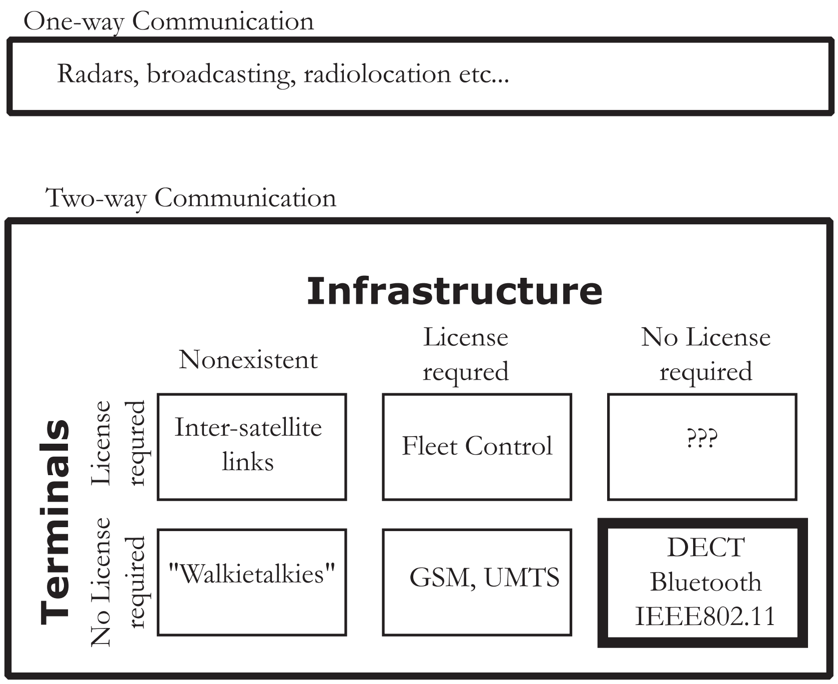

Figure 2-1 Overview of different licensing regimes. For two-way communication some systems rely on fixed infrastructure while others do not. A license may be required to run infrastructure or use terminals. The focus of this thesis is systems where no license is required to run either the infrastructure or to use the terminals.

When one talks of radio communication the general understanding is that there is some form of two-way communication. However there are other uses of the radio spectrum. For example radars usually have the transmitter and receiver collocated. Another example is the broadcasting, radiolocation and paging services. Here there is one transmitter that transmits to a number of receivers, but there is no two-way communication. Generally these types of transmitters require a license. There are also the class of applications where the important aspect is not the information embedded in the radio waves, but some other aspect. For example the energy contained in the radio waves is the main interest in microwave ovens. This kind of transmitter usually does not require a license. Anyhow, the focus here is two-way communication and we leave these types of systems out of the studies.

One way to characterise a system is if there is any infrastructure and if a license is required to run it. In this context infrastructure is a fixed transceiver that may be connected to other types of fixed networks, for example telephone or IP networks. Another distinguishing characteristic is if it necessary to have a license to operate the mobile terminal.

If we look at figure 2-1 we see that there are many combinations where it may not be necessary to have a license to operate a transceiver. First look at the case where there can be no infrastructure, i.e. it is prohibited to use fixed stations. There are both licensed and unlicensed systems. For example for satellite-to-satellite links a license is required and in Europe there are the PMR446 unlicensed walkie-talkies.

Now have a look at the case where a license is required to run the infrastructure. When a license is required to run the mobile station as well we find fleet control systems, aircraft and ship communication systems etc. The main example of the case when no license is required for the mobile station is the mobile phone networks, e.g. GSM and UMTS.

Finally there is the case when no license is required to run the infrastructure. To have no license requirement for infrastructure, but require a license for running the mobile stations does not really make sense. But the case where no license is required to run either the infrastructure or the mobile stations is interesting. In fact it is the focus of this thesis. In this (shaded) box we find systems such as WLAN, cordless phones and so on. It is this type of systems that generally is considered to be unlicensed.

As we noted previously there are different rules for each piece of spectrum. We can make another division here. There are pieces of spectrum where the rules specify that only one type of system be allowed to operate. Examples are the spectrum allocated for DECT or for HiperLAN/2. This kind of operation is also called license exempt. There is another kind of rules that are much more simple. They only specify a few things that a transmitter must obey, for example that they must listen before transmitting and have a limited transmission power and transmission time. An example of such a band is the band 2400-2483 MHz where both IEEE 802.11 and Bluetooth devices operate. An interesting side note is that the ISM bands were originally created as a place to locate equipment that creates a lot of interference. Thus the other parts of the spectrum should be free from these noisy applications.

This work is mainly focused on rules that allow one transceiver to be fixed and where there is no license required to run the fixed or the mobile transceiver.

2.3 Environments for unlicensed operation

There are many places where radio communication is used today. The distances where radio communication is used ranges from halfway around the earth to replacement for the cable between a computer mouse and the computer itself. In all these applications there may be a possibility for using unlicensed operation.

For all radio communication the resource allocation problem has to be solved. That is all transmitters have to be assigned a frequency, a transmission power, a waveform and the time of transmission. The solution to the problem depends on other active transmitters and the propagation conditions and possibly quality requirements on all transmissions. In addition a common requirement is that the planning should be made so that as many users possible can communicate. It is easy to realise that the problem solution will be different for different environments.

One way to solve the problem is by planning in advance. That is the idea behind licensing transmitters and behind planning cellular radio networks. For unlicensed operation planning in advance is not possible. The reason is that little is known about other transmitters operating at the same time. Thus a cornerstone in any system using unlicensed bands is the ability to adapt to changing propagation conditions and most important changing interference situations. The adaptations have to be made in real time so that the communicating users do not notice the changing radio conditions.

We have to distinguish between transmitters belonging to the same system and those that belong to different systems. Within a system it is possible to have coordination among the transmitters and thus they can solve the resource allocation problem so that they do not interfere with each other. However across systems that is not possible. Transmitters belonging to different systems may not have the same coordination functions and they may not want to coordinate their transmissions with other systems.

For really long distance communication unlicensed operation is probably not feasible on a larger scale. The reason is the number of possible interferers. For communication half way around the world interferers may be located almost anywhere.

For systems that cover fairly large outdoor areas, for example a city there may be interferers, i.e. other systems, located in the same area. The reason may be that there is more than one operator active in that area. In addition there may be interference coming from other systems indoors. It is reasonable to believe that there is more than one operator active at the same time. So the probability that there is interference from other systems in an outdoor setting is large.

For shorter range communication, for example in indoor settings it is not as likely to encounter interference from other networks, at least not to a great extent. It is not likely that there will be many networks operating in the same area within a building. However it is possible that many networks will be located in a building, but in areas that are only partially overlapping. For example there may be different networks on different floors, but the attenuation in the floor will make the mutual interference lower.

For really short range communication there may also be interferers located close by. One can imagine a person carrying many devices that use different standards for communication, standards that are incompatible and thus cause interference to each other.

For this study we are not only trying to determine if it is technically possible to use unlicensed operation. The spread of IEEE 802.11b wireless LANs is evidence that it is technically possible to use unlicensed bands for communication. The other aspect we are trying to find out is if it makes sense from a business perspective. The underlying assumption is that an operator takes care of the users communication needs. The environment where that may occur is in the cases where the communication distance is a little bit larger. In the really short distance communication it is not likely that there will be an operator involved. The user buys the equipment, puts it in his pocket and then expects it to communicate. On the other hand using unlicensed bands for really long distance communication is probably not feasible from an interference point of view.

2.4 Technology choices

One of the problems with radio communication in an unlicensed environment is interference. There are various strategies for dealing with it.

In licensed systems the main strategy to deal with interference is to plan in advance to make sure that a user never gets too high interference levels. The positive side is that planning can be done once and for all and the planning does not have to be made in real time. The drawback is that the plan must by necessity be based on a statistical traffic load, i.e. there is no way to adapt to changing traffic conditions.

Another strategy to deal with interference is to avoid it. If there is too much interference we switch to another frequency where the interference is less or we simply wait to the interference level has dropped. Another way is for a user to move to another spot where interference levels are lower. Dynamic channel allocation and retransmission schemes are examples where this strategy is used.

A third strategy to deal with the interference is to "live with it". The idea is to send the same information on several different frequencies and/or at different times. Even if some of the information is lost hopefully another copy of the same information arrived safely at the destination. Examples of this strategy are frequency hopping, hybrid ARQ schemes and direct sequence CDMA.

2.5 Problems with unlicensed operation

The main problem with unlicensed operation is the interference situation. Interfering transmitters may have approximately the same transmission power as the other transmitters in the system and they may be located in physical proximity. This makes the interferers potentially strong, especially in the outdoor environment. The interfering transmitters are not transmitting all the time. This makes the interference quickly varying. Finally since interferers belong to another system it is not possible to control them. Thus the interference can be characterised as strong, quickly varying and not controllable.

To handle the interference situation it is necessary to make transceivers that automatically adapt. There is a benefit with this. The configuration of the transceiver becomes automatic. Thus there is no need to make a frequency plan, since it is not possible to make such a plan anyway.

One potential problem is that operators are not willing to trust unlicensed operation. They believe that it is difficult to provide any quality guarantees to their customers. The reason is the interference situation. Other operators are using the same frequency spectrum and causes interference. This is of course only valid when there is more than one operator, which makes this argument valid mostly for outdoor systems.

There are also problems that are of a more economical business nature. Operators have recently bought the licenses in many countries for third generation systems. The license fees have been large (approximately 100 billion euro) [36] and the operators want to get return on their investment. Thus they may not want to give up any spectrum for use for unlicensed operation or try new modes of operation. This is especially true for the infrastructures that cover large areas, and thus are costly.

Another problem with the large license fees is that they may make operators and equipment suppliers reluctant to put any effort into alternative technologies. Nobody is interested in making investments in something that may reduce the value of their already made investments. Since everybody is in the same situation there is no incentive to develop new technology.

Unlicensed operation opens up the possibility to have more competition. It is also possible to imagine a system where the user instantaneously chooses the best operator. This makes it more difficult for an operator to keep customers. Operators may not be willing to loose their customers and it may be more difficult to build customer loyalty in such situations. However it is also possible that operators think they are so good at providing services that the competition is a good thing.

It is possible to envision systems where one system provides capacity in hotspots and another system is used to create full coverage. At the target date for the scenarios the third generation systems will be operational. It is possible that there is no need for another system that provides large area coverage since the systems already in place provide enough capacity.

Regulatory bodies, i.e. governments, also have interests in the licensing procedure. To ensure that areas with few customers are covered an operator has to agree to cover a large percentage of the country even if some areas are not deemed to be profitable[36]. This is the case for example in Sweden. The procedure can be viewed as a tax built into the license. If regulators allow unlicensed operation they loose this possibility. This argument is of course only valid for large areas with little population.

2.6 Benefits with unlicensed operation

There are many problems with unlicensed operation. However there are also a number of arguments that suggest that unlicensed operation is a mode of operation that may provide competitive edges for operators.

The technology that enables the use of unlicensed spectrum also facilitates automatic configuration. This makes it possible for operators to avoid the labour intensive task of configuring all access points in a network. The result is less errors and lower costs of running the infrastructure. It should be noted that the same techniques could also be used in networks using licensed spectrum to achieve the same benefits.

Unlicensed spectrum also opens up possibilities for new business models. Since configuration by technically trained staff is not required and there is no license required for building a network it is possible for the end user to build the network. Here the end user does some of the work for the operator. This offers an increased degree of freedom to implement new business models. Actually this shift in responsibilities is one of the factors behind the success of IKEA [37].

The rules governing the use of unlicensed spectrum are usually not as complicated as those governing the use of licensed spectrum are. This should be compared to the standards for systems that use licensed spectrum. These standards are usually more elaborate and detailed. If there are not too many rules it is easier to take advantage of new technical innovations. Thus it should also be possible to use the available resources more efficiently.

The frequency spectrum is not split when using unlicensed operation. This avoids losses caused by trunking inefficiencies [23]. There are also additional gains to be made from the fact that more bandwidth is available for use by each access point. This may be an advantage for unlicensed systems that can provide more efficient operation.

Telecommunication operators see benefits with unlicensed operation as well. It becomes easier for them to enter a market since they don’t have to go through the process of applying for and possibly vacating frequency spectrum.

One of the reasons that unlicensed operation is believed to become one of the dominating modes of operation is that it better supports the needs of operators. Using unlicensed operation may then provide a competitive edge for the operator that makes it successful in the marketplace. However there is also a possibility that unlicensed systems are more efficient than systems operating using licensed spectrum. The higher efficiency, from a technical perspective, can then be translated to a lower cost of providing communication services.

3. Method

The main question in this thesis is if unlicensed operation can be a used to provide profitable communication services. In this chapter we look further into the problem and into the methods used for attacking the problem. The multidisciplinary approach used in this work makes it important to be careful when choosing methods. The important things studied in radio systems research may not necessarily make sense when understanding how a firm is affected and vice versa. It is not enough to use the tools provided by traditional radio systems research. In addition economic theory that deals with the firm must also be used.

3.1 High-level problem overview

When we approach the problem we will first have a high level view of the whole problem. We can imagine one instance where there are two operators that provide wireless communication to users using unlicensed bands. Figure 3-1 depicts some of the interactions that take place in this case.

In this example there are two operators that provide communication services, but there may be more. However they all use the same radio spectrum to be able to provide the services. Each operator owns or rents the infrastructure necessary. Users pay operators to receive services.

The regulator may put rules on how the radio spectrum is to be used. These rules may be more or less complicated. The span is from simple rules, for example a limit on the maximum transmission power, to complex rules for example those specified in a standard. A standard usually specify a lot of things that transmitters and receivers must adhere to. If a regulator then requires that equipment must follow a specific standard it is equivalent to setting a large number of rules for spectrum use.

In this scenario there are also equipment manufacturers that supply both users and operators with devices and get paid for that. There may also be other operators that provide the same kind of communication services, but the spectrum used is different.

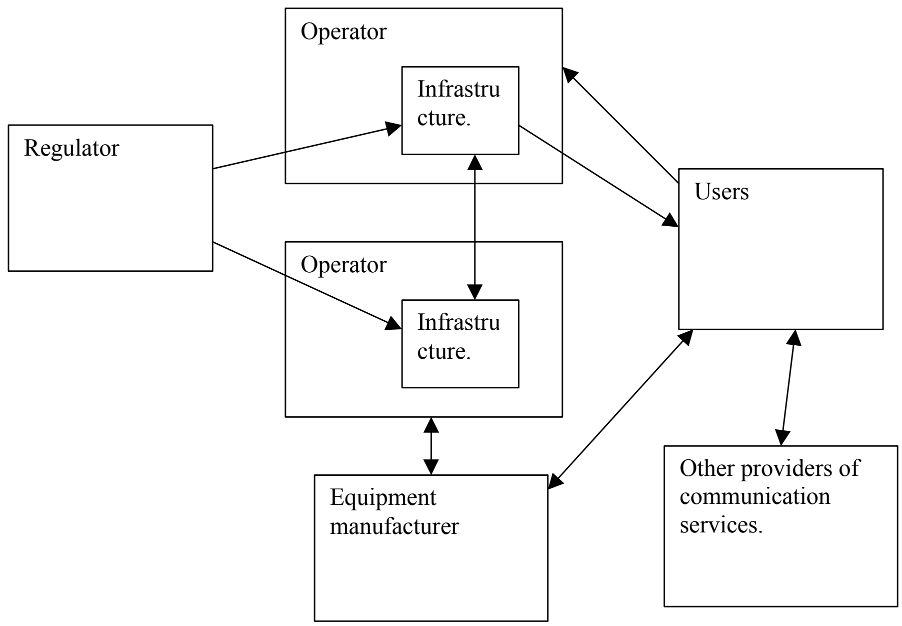

Figure 3-1 Overview of interactions taking place in a case where multiple operators provide communication services to users.

The focus of this thesis is the infrastructures and the interference between them. If we look further into the case in figure 3-2 we can see that there are some parameters that are unique for each infrastructure, e.g. power control scheme, modulation format, link adaptation method, MAC scheme, admission control, scheduling and so on. In addition each system has users that have a specific traffic pattern, user location and quality requirements. There are also a number of parameters that are common for all infrastructures e.g. things relating to the physical environment, e.g. radio propagation conditions.

We can now define the set of all possible instances where systems coexist. For each element in the set we define a specific number of systems that coexist. Each system has its own design parameters. In addition there are a number of parameters that specify the common features for all infrastructures.

We would like to check if a combination of parameters constitutes a feasible system design. To do this we set a number of performance criteria that must be met. Selecting these criteria is not trivial, but for the moment let us assume that we have a number of criteria that we can compare the performance with.

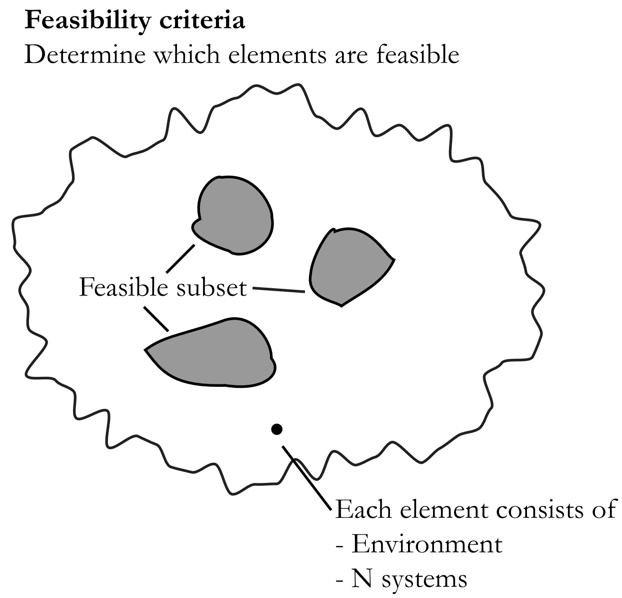

Figure 3-2 In the set of all possible combinations of environment, and a number of systems that coexist there will only be a certain set of cases where operations is feasible.

To demonstrate feasibility it is, in the strict sense, only necessary to find one case that satisfies the feasibility criteria. But there is always the risk of selecting a case that is of little practical relevance and thus demonstrating feasibility may not be of large interest. However we want to be able to extend the conclusions from the investigation to unlicensed systems in general. We want to understand some of the workings behind unlicensed systems. This knowledge will be helpful in making good design choices when creating systems for unlicensed operation.

We can group the elements into subsets where all elements represent a feasible case. If we determine differences between the elements belonging to feasible subsets and the others we can understand the characteristics that make unlicensed systems feasible.

3.2 Feasibility criteria

We want to investigate if unlicensed operation is a feasible mode of operation in future infrastructures. Whether unlicensed operation is feasible depends on a lot of factors. We can see that the factors span a large number of topics, from economical and technical issues to regulatory and marketing issues. Figure 3-1 shows some important facets of this problem.

The infrastructures are part of a larger system and there is interaction between infrastructures and the rest of the system. Thus it is not enough to study the technical characteristics of the infrastructures in isolation.

Feasibility is measured outside the studied object, i.e. the infrastructures. We want to determine if the operators can stay in business and we study the performance of infrastructures.

This thesis is limited in scope and the whole system outlined in figure 3-2 is not studied. This is done mainly from practical reasons. The thesis was made in the context of radio systems research and as such the focus was put on technical issues.

The approach used here is to use three factors that we think influence the whole techno-economic system. These are such that the design choices in the infrastructure influence them. As such it becomes possible to understand how the design choices in the infrastructure affects the whole techno-economic system.

The three parameters we select are the service provision costs, how the size of operators influence the service provision costs and finally if it is possible to gain benefits by breaking rules set for the cooperation.

3.2.1 Cost of service provision

In any business the income must on average be larger than the expenses. This indicates that an important factor when determining if it can be done is the amount of income that can be generated by a network (capacity) and the cost for building it. Communication services provided in unlicensed spectrum will most probably face competition from similar services provided using licensed spectrum. Thus it becomes important to compare networks using licensed and unlicensed operation in terms of cost and capacity.

If unlicensed infrastructures are to be successful it is important that the cost of providing a user with services is of the same order as for current systems. Probably the largest cost is the cost of building the infrastructure. A very simple measure is the number of access points, or rather how large fraction of an access-point, is required to support one user. This can be viewed as the cost of supporting one user.

There are really two comparisons that we are interested in. The first is how the operators using the same spectrum influences each other. This could then be used to understand how competition among operators would evolve. The other comparison is between operators using unlicensed spectrum and those who do not. This would then aid the understanding of the competitiveness of unlicensed systems.

Of course the cost of providing a service is only one side of the equation that determines if the infrastructure is economically feasible. The other side is the price user is willing to pay for the service. However it is not likely that services in an unlicensed infrastructure is so much better that a user is willing to pay more.

3.2.2 Ensuring competition

Regulators and policy makers are interested in ensuring that there is competition in the marketplace. It then becomes important to determine if more than one operator can be successful in the marketplace at the same time. If there are significant advantages for a large operator or the one who enters the market first regulators may want to avoid that situation by changing the rules for using unlicensed spectrum.

We study the influence of the operator size on the cost of service provisioning. If the average cost decreases when the operator size increases it is an indication of a situation where there eventually will be one operator who survives in the marketplace.

3.2.3 Reliability

When an operator decides if it should invest in new technology one important measure is what the performance of the new technology is. Actually there are two parts of this measure. First there is the expected capacity that can be expected from the new system, but there is also the reliability of the expected value, i.e. how much it can vary.

It is very difficult to quantitatively determine how the correctness of performance measures will influence the investment decisions of an operator. The reason is that there are so many factors weighed in when an investment decision is made. Thus how well a technology performs only plays a small role. However it seems reasonable to assume that the more precise performance predictions can be made the more likely an operator is to invest in new technology. Thus a desirable characteristic is that the performance is predictable and similar under similar conditions.

Another aspect an operator looks into is the performance of a system for the competitors. If a technology is expected to perform much better for a competitor it is not likely that the firms invests. Thus to ensure competition it is desirable that the chosen technology is fair, i.e. the performance should be similar for all operators.

In this thesis no attempt is made to quantify how predictable system performance is and how that influences the risk evaluation of operators. Instead we try to at least get an overview of the issues involved and the mechanisms behind.

We look at two things. First we look at stability, i.e. how predictable system performance is under similar operating conditions. One thing that is special for unlicensed operation is that the behaviour of other operators influences the operator. Thus it is possible for instabilities, i.e. actions by one operator causes reactions from the other operators, which in turn creates more actions etc. Thus we study if the operating point converges.

The rules for using the spectrum, the etiquette rules, are made by the regulator and they are (hopefully) made to ensure fairness between the operators. We also look at the possibilities to get better performance by breaking the etiquette rules.

3.3 Selecting relevant cases to study

The problem domain is large and to investigate all combinations of system designs is at best a very tedious task. Thus we pick only a subset of cases to investigate. Determining which cases to study is non-trivial. This is a task requiring intuition in addition to the factual knowledge about this kind of systems. A number of criteria for picking cases can be conceived.

Cases that represent probable future scenarios
It is good if the cases investigated correspond to probable future scenarios. That way the results of the investigations can be used when designing systems of the future. This also makes the research more relevant in terms of usefulness.

Cases related to current technology
There are already a number of systems that use unlicensed bands. It is good to select system design combinations that are similar to these technologies. One reason is that the current systems already work. It seems reasonable to assume that similar system designs will be feasible. Another reason is that the current systems will probably be around in ten years time, which makes the results more relevant in future scenarios.

Results that are easy to interpret
We want to understand what factors of the system design that makes coexistence feasible. In order to do this we need to select cases that make interpretation of the results easy. The question is how to select cases that give results that are easy to interpret without knowing the results. It is not an easy question, but it seems reasonable to believe that these cases are characterized by relatively simple system designs.

Representative cases for a class of systems
In many cases it is possible to find a subset of points that have very similar characteristics. An example is all cases where the systems that coexist are the same, but the environment is different. In these subsets one case that is representative for the whole subset should be chosen. Then it may be possible to extend the results to the rest of the cases in the subset by deduction without actually doing the experiments.

Cases related to previous studies
There have been some studies already performed in the area of unlicensed operation. These have to be considered when selecting the cases to study in this thesis. One obvious criterion is that studies should not be redone since it

is much easier to just read about the results. But the cases selected should not be too different from the cases already studied. That way it is easier to interpret the results from the studied cases together with previous results. In addition it is possible to check the relevancy of the results. I.e. if the results differ a lot from previous studies there is something strange going on.

3.4 Chosen cases

Now we get back to the point where we should pick the cases to investigate.

One thing that we know is that unlicensed operation is already used in a number of systems and it seems to work fairly well. The environments that come to mind is DECT telephony systems deployed in small or large offices or systems covering a plant or other company premises. Another case is wireless LANs that also cover similar areas or wireless LANs that are used for providing public access to the Internet.

One thing to note is that generally there is usually only one system of each type deployed in the same geographical area. When there is more than one system in one area the aggregate performance is not well known. In addition we want to know something about competition and the behaviour of two coexisting systems. Thus we must pick cases where there are at least two systems that interfere with each other.

One design parameter that needs to be decided for any communication system is the multiple access scheme. The choice is not obvious, something that the large number of proposals for the third generation system shows [38]. For systems intended for use in unlicensed spectrum it is important to select a scheme that can cope with interference from external sources, interference that can not be controlled. Various systems intended for unlicensed operation has made different design choices. For example DS-CDMA is used in IEEE 802.11b, slow frequency hopping in Bluetooth and dynamic channel allocation in HIPERLAN/2. Since there are many possible candidates and since there is no scheme that is "the obvious choice" we pick a number of multiple access schemes that has been used before and thus may be promising candidates.

Depending on the type of traffic (voice, data) the system carries the system design and thus performance will be different. Systems designed mainly for voice have been around for quite some time. The behaviour of these systems is fairly well known. In addition the amount of computations required when doing numerical experiments is comparatively small. Thus the results from computational experiments should be easy to interpret and quick to obtain.

The scenarios indicate that a large amount of traffic will be data. The behaviour of this type of systems is not as well known, and there is no easy relationship between the performance of voice and data systems. Unfortunately the amount of computations required when doing numerical experiments with data traffic is larger. But to be able to say something about future systems we need to make experiments on data systems.

To summarise: we look at three different multiple access schemes in systems that carry either data or voice. That gives us in total six different possible combinations.

There is also a question of how many systems that we should have coexisting in the same area. There must be more than one, but there is no natural upper limit. We decide to look at two simply because it is simpler to run numerical experiments with only two systems. Also judging from the situation today where there is only one operator in most areas, two seems like what will be the next step.

Then there is the question to determine if the two systems should be designed in the same way or in a different way. The argument to keep the design the same is that numerical experiments will be simpler to perform. The results will probably also be easier to interpret. The reason for keeping the systems different is that in the future there will not only be one system design around. But to save time, complexity and to keep results easier to interpret we use the same system design for both systems.

3.5 Comparison issues

In this study we compare systems with similar design both for licensed and unlicensed operation. In one sense the comparison is a little incorrect. In licensed systems it is possible to use more advanced radio resource management methods. The reason is that an operator has more control over the interference generated by the users. Thus it is possible to design systems that give better performance, i.e. lower cost of supporting a given number of users. However to avoid the influence of different system design parameters, e.g. radio resource allocation algorithms, on the performance we use the same scheme for all systems that we compare.

For someone running a mobile communication system on a commercial basis it is important to have customers using the system and it is important that those customers are satisfied enough to pay the bill. The number of users that can successfully be served by a system is dependent on many factors. Here we focus on available system bandwidth, the number of access points and the user quality requirements.

For systems that achieve full coverage the number of users that can be served is proportional to the available system bandwidth and the number of access points [34]. There are a few things worth noting. Ultra wideband (UWB) systems have been reported to be able to serve a large number of users without disturbing other radio services. This is accomplished by spreading the radio signal over a very large bandwidth, several GHz. Thus other narrowband radio services using the same spectrum experiences just a small amount of increased interference. This may seem like system capacity is created out of thin air. But the same limitations apply to UWB systems as to other communication systems. If there are a lot of users in the UWB system eventually it will become interference limited. To make fair comparisons the spectrum available must fixed for all systems.

Quality is a complicated issue. From an economical point of view the quality experienced by a user is what determines his willingness to pay for a service. The user satisfaction is difficult to measure. For voice services the blocking probability has been used both for fixed and mobile networks. For data communication the measure is usually throughput and delay. Giving users increased quality generally lowers the capacity of a system. E.g. to reduce blocking probability less users can be served or to accommodate higher data rates fewer users can use the services.

How the system is designed, i.e. which technology is chosen affects the number of users that can be served. However when the system design has been made the influence of technology on capacity can be viewed as a constant.

To summarise: The capacity of a system is an important measure since it reflects the revenue of an operator. Capacity of a system depends on the amount of available spectrum, the number of access points and the user quality requirements. When making comparisons between different systems these factors must be accounted for.

3.6 Research process

Today technological evolution is perceived to be immensely fast. People think that new products come out every day and we can read about new findings every day in the papers. The speed of technological evolutions is increasing and increasing. There has even been a new word invented “internet time” to signify that things are happening faster nowadays.

This may to some extent be true. For example web development cycles can be as short as 3 to 6 months compared to 2 to 3 years for software projects ten years ago [39]. New mobile phone models with new features arrive every six months or so.

However in some areas technological development is not as fast, infrastructure is such an area. For example standardisation of GSM begun in 1982 and GSM networks have been operational in Sweden since 1992 [40]. It is only now that a new infrastructure generation, known as the third generation, is about to be deployed. Building infrastructure is expensive and the time required for payback thus becomes large, which means that infrastructure typically have long (must have) lifetime.

Now when the third generation of mobile infrastructures is in the initial phase of deployment it is time to start to think about what will come after. If one counts all licensing fees, development costs and expected deployment costs the third generation is said to be the most expensive project in the history of mankind. This means that the time until the third generation has covered its costs is expected to be fairly long, and there will be some years before there is money available for other infrastructure projects in the mobile communications world. Studies of the Swedish telecom market indicate that it will not be until around 2010 some operators have recovered the costs for infrastructure investments [41]. We may have to wait until 2010 or 2015 until it is economically feasible to run another infrastructure project.

When doing research on infrastructures that will be built in 15 years time we want to do research on relevant issues. There are a number of reasons for this. From a personal point of view it is more fulfilling to do something that is useful. Also from the viewpoint of society it is more relevant to spend effort, i.e. money, on something that can be used and that benefits society as a whole. But there is also an interest from the one who funded this work to get something back.

The research presented in this thesis is performed within the context of a large Swedish research project, the PCC project. One of the intentions with PCC – Personal Computing and Communication is to help Swedish industry to maintain leadership within the telecommunications area. From this viewpoint it is important to do research on something that can be used by industry.

The problem is to find relevant questions to research. It is a nontrivial task considering that the object of study (infrastructures) will not be deployed for another 15 years. The work presented here has been performed in the context of a subproject within PCC: the 4GW (4^{th} Generation Wireless infrastructures) project. In this project we have used a scenario technique as the basis for determining the relevant research issues.

The research process is outlined in figure 3-3. From current trends, ongoing research efforts and other predictions of the future a number of descriptions of what the telecommunication business may look like in 2010. These descriptions are also known as scenarios. It is worth noting than a scenario is one possible future, not the only possible future.

A couple of factors have a large impact on the direction the technological evolution will take. Some factors are more important than other in determining the direction of the development. These factors are codified into a set of background and working assumptions.

The details of how the scenarios were created and the detailed description of them can be found in [42]. A quick overview is given in appendix A.

From the working assumptions research projects are created that investigate one or a couple of issues. The results from these research projects can then be fed back into the process to make the scenarios more refined and to update the working assumptions.

The working assumptions are of different types. Some working assumptions describe a state of technology that is common to all scenarios.

The reason for doing research in this area is that the results can be used in future infrastructures. The focus is to develop the technology that the working assumptions describe.

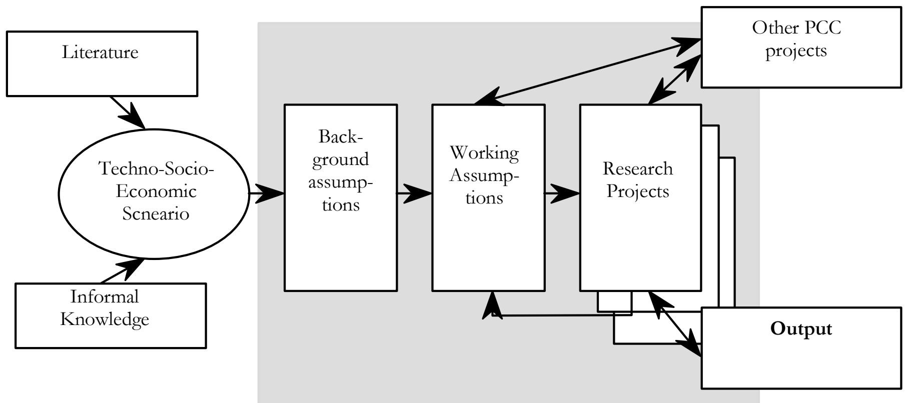

Figure 3-3 A schematic overview of the research process within the 4GW subproject.

Another group of working assumptions describe technology that is only relevant in one of the scenarios or a technology that is a key component of the wireless industry. Which scenario we end up in depends if this technology is feasible or not. To do research on these issues is relevant if one wants to understand the way technology will evolve.

In this thesis one specific working assumption will be targeted. The assumption is that licensed operation dominates. Unlicensed operation is only a key component in one of the scenarios, so this research aims at understanding what possible directions the technological evolution can take. Even if this thesis only studies one of the assumptions, the assumption is still really wide, and complex. Thus in order to be able to study the issue we select some relevant aspects of the working assumption and state that in terms that can be measured.

4. Evaluation Models

It is possible to trace most of the model choices made back to a cellular voice type paradigm. There are many reasons for this. One reason is that the available tools were geared towards research on large cellular systems. Most of the computational experiments were made with the help of a MATLAB toolbox called RUNE [43]. This toolbox was originally designed for evaluating GSM networks. Over time additions have been made to include packet traffic, CDMA etc. and for this thesis other models have been added as well. But still the GSM origin shines through.

Another reason is that large-scale cellular systems have mainly been intended for voice or data services with fairly low data rates.

One important factor for the relevancy of this study is how well this research paradigm matches the future infrastructures that will be built. There are two parts to this question. First of all how large the difference is between future systems and the models used here. The second question is what the influence of the difference will be, i.e. how much more difficult the interpretation will be and what aspects of future infrastructures will not be possible to say anything about.

Current technologies as well as the scenarios focus indicate that data rates will be higher than what current cellular systems can achieve. There is also a difference in the coverage of the systems. The trend seems to be local coverage instead of large area coverage. But as always predictions are difficult.

Cellular systems tend to scale rather well. I.e. results obtained for large-scale systems can be used to understand the behaviour of small systems. The increased data rates influences the carrier bandwidth so that fewer carriers can fit within the same spectrum. Also the statistical properties of the traffic may be influenced. A user may either have a lot to transmit or nothing at all. This is something that can influence the resource allocation algorithms.

And with all that said, these are the assumptions that were made in this study.

4.1 Traffic and environment

In this work we only study the traffic in the downlink. That is the traffic from an access point to a mobile user. In future systems we expect most of the traffic to be in the downlink [44] and that will probably be the direction limiting the system performance.

4.2 User distribution

We assume that the users are uniformly distributed over the area. Of course it is possible to imagine any number of user location schemes. However by using a uniform distribution there is one parameter less that influences the results.

4.3 Voice traffic

Even though we do not expect voice traffic to be the only source of traffic in future wireless networks it is still useful to model voice only traffic. Since the traffic model is simple it is easier to understand the underlying mechanisms. As an additional benefit it does not require a lot of computations to find results.

Users arrive to the system at random locations and have a call duration that is exponentially distributed. This is the classical traffic model used when modelling mobile radio networks. Using a two-dimensional Poisson distribution simulates it.

We only look at one specific time instant when doing the numerical calculations, i.e. snapshot simulation. This method does not capture problems that occur with handover of mobiles. It does not capture effects to changing pathgain over time for the mobiles either. However we are not interested in studying those effects.

4.4 Packet traffic

Recently there has been an increasing interest in mobile Internet, as well as wireless LANs. There are many driving forces behind this novel interest, but the relevant point in this context is that the radio link will be used more and more for transmitting data. To make the results of this study relevant it is necessary to be able to model the traffic of such networks.

It is difficult to predict the characteristics of the traffic in future wireless systems since the use and application or the application requirement is not known. This is a very complex issue spanning a wide range of subjects from image compression techniques to economic issues. It must be pointed out that determining the traffic in future wireless systems is not the task of this study.

One approach commonly chosen is to assume that the characteristic of the traffic will be similar to the traffic patterns seen in the wired computer networks of today. The traffic in these networks has been studied and models for this kind of traffic have been made.

Traffic on an Ethernet has been found to exhibit a self-similar property [45]. One consequence of this is that aggregated traffic exhibits the same "burstiness" as the individual traffic streams. This has an impact on the delay performance and the trunking gains that can be expected from the system. A model for modelling WWW traffic is proposed in [46]. This model is picked in order to be able to accurately model both throughput and delay of wireless systems. This model has been used earlier (with some modifications) to model Bluetooth-type of networks [13], wireless LANs [47] and packet traffic in cellular systems [48].

The traffic of one user is modelled in the following way: The communication of each user is grouped into sessions. In each session the user transmits a number of packets of a specific length and with a specified time between each packet. This is illustrated in the figure below.

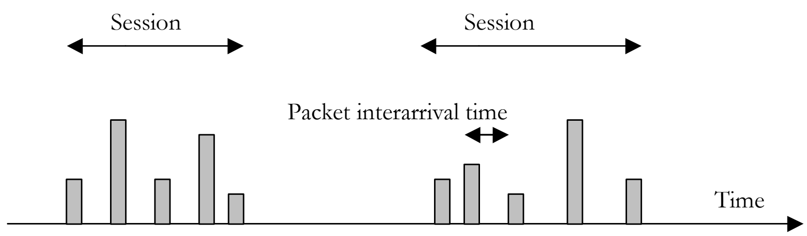

Figure 4-1 The traffic is modelled as a number of sessions with a exponentially distributed interarrival time. The packet size is normal distributed, the number of packets in a session are binomially distributed and the interarrival time is Pareto distributed.

Session inter-arrival time is modelled as an exponential distribution. Which makes the number of sessions that arrive in a specific time interval Poisson distributed. The arrival rate is the only parameter that is used to control the load of the system.

The number of packets in a session is geometrically distributed with mean 10 packets.

The packet inter-arrival time is considered to have a truncated Pareto distribution with shape parameter α=1.2. The minimum is 0.84 s and maximum 333.3 s.

The packet length is lognormal distributed with mean 5 Kbytes and variance (σ) 15 Kbytes.

4.5 System layout

The location of the access points is one of the factors that determine the performance of the wireless system. A model where the access points are located in hexagonal patterns is a commonly used model since it is simple. There are limitations to the usefulness of the model though. For example in downtown areas with lots of high buildings the radio waves tend to propagate along the streets, but not around corners. This has resulted in other ways of modelling the access point location, e.g. the Manhattan model. [49]

In this thesis the hexagonal layout is used since it is simple.

However since there is more than one operator in the same area the relative location of access points is also important. Two extreme cases can be imagined. One possibility is that all access points are collocated; the other possibility is that access points belonging to different operators are as far apart as possible.

In the coexistence problem we can see that the problem occurs when there is a lot of interference that cannot be controlled. When studying the downlink this happens when a user is far from its own access point and close to another operator's access point. Thus it is easy to realise that the worst-case scenario is when the access points are located as far apart as possible.

Here we choose to have hexagonal cell layouts for each operator. The cell patterns for each operator are shifted one cell radius to get access points as far apart as possible.

When the access point density is equal for both operators each operator has 16 cells with a radius of 1000 m.

In the cases where the access point density is different operator 1 still has 16 cells with a radius of 1000 m and operator 2 has 36, 48 or 64 cells with radii according to table 4-1.

Ratio	Number of Cells	Cell radius
1:1	16	1000 m
1:2.25	36	667 m
1:3	48	577 m
1:4	64	500 m

Table 4-1 The number of cells and cell radius for different access point density ratios.

4.6 Propagation

The propagation loss is modelled as a sum of three components. First a constant to account for the antennas used etc. Second a distance dependent component and a lognormal component to account for shadowing etc.

The path loss can then be written as:

$$L = -21 + 35\log(r) + 8X \tag{4.1}$$

Where X is a normal distributed variable with variance 1.

This is the propagation model implemented in RUNE [43]. One of the reasons for choosing this model was convenience. Another reason is that it gives a reasonable approximation of a number of different scenarios, e.g. outdoor cellular systems, indoor or mixed.

The receiver noise is set to –118 dBm, which corresponds to a reasonably well-designed narrow band receiver. However the influence of the receiver noise is not significant since the studied systems are mostly interference limited.

4.7 System models

There are three groups of systems considered in this work. The three groups are based on the multiple access technique used. The design choices for the various groups are described in this section.

4.7.1 FH system design considerations

We have designed two different systems that use slow frequency hopping as access method. One system carries voice and the other system is intended for data traffic, i.e. packet traffic. Both system designs have a number of features in common.

The hopping sequence is random over the whole set of available channels. There is no coordination between access points. However since we study the downlink it is reasonable to assume that the hopping sequences are orthogonal for users that are connected to the same access point.

We assume that the channels are completely orthogonal. I.e. there is no adjacent channel interference. This may be a somewhat optimistic assumption. However the adjacent interference is most pronounced between two channels next to each other in frequency. If the frequency difference is larger the adjacent channel interference is not as pronounced. If the number of channels is large and the number of users is small the probability that two users end up at frequencies next to each other is reasonably small. Thus this assumption is probably not overly optimistic.

The output power has been set to 30 dBm.

The hopping sequence is random. That is, each user selects one channel out of all the available with an equal probability in each hop. But there is one exception. Users that are connected to the same access point do not select the same channel in a hop.

In the voice system there is no admission control. I.e. all calls that arrive are admitted even though there may be an interference problem.

In the system designed for data traffic each user has one queue. A packet that arrives is immediately transmitted if there are no packets in the queue. If the user is already transmitting a packet the new packet is put in the queue.

The packet is split into equal sized blocks. One block is transmitted each hop. Blocks that are erroneously received are retransmitted using an ARQ scheme. We assume that the acknowledgements are perfect and instantaneous. If a block is not filled with data the block is padded to fill the complete slot.

We assume that the whole system is synchronized. There is also synchronization between operators. This assumption is done since it makes implementation of numerical simulators easier. The assumption may result in a slight overestimation of the available capacity since collisions occur less frequently. Under this assumption a frame either collide and is completely lost or it is received correctly. But in a practical system there is a larger probability that a frame collides with two frames. Parts of these frames are then lost. Depending on the coding and interleaving schemes used this may result in more errors.

4.7.2 DS-CDMA design considerations

Direct sequence CDMA is a technique that has been used to suppress interference in both military (stealth) and civilian applications, e.g. IEEE 802.11b wireless LANs.

The interference from one user to another user depends on the cross correlation between the codes used. It also depends on the time difference, the multipath propagation conditions and the design of receivers. Here we use a simplification: Interference is suppressed by the processing gain for users that are not connected to the same access point. The users that are connected to the same access point are assumed to have perfectly orthogonal codes and thus not interfere with each other.

Since the channels in a DS-CDMA system are not orthogonal a power control scheme is usually employed to ensure that all users experience approximately the same amount of interference. In this system we use a SIR balancing algorithm known as DCPC [50]. The algorithm is iterative. The transmitted power is updated according to the following function:

$$P_i = \min(P_{\max}, P_{i-1}\frac{\Gamma_T}{\Gamma}) \quad (4.2)$$

The idea is that all users should have a specified sir (Γ_T). If a user is below that the power is increased and if he is above the power is decreased. However there is a maximum allowable transmit power due to physical constraints.

In the voice system the algorithm is run for several iterations to ensure that the powers have converged (in most cases). The idea is to find a set of power values that will support as many users as possible. In the system for packet traffic the algorithm is run once every slot. By varying the slot time the influence of the update rate of the power control algorithm can be studied.

One problem with a DS-CDMA system with power control is that if there are too many users active at the same time there will be too much interference and a lot of users will suffer, i.e. nobody will be able to communicate. Thus the system has to somehow limit the number of users that are active in the system, i.e. removing some of the users.

In the voice system a removal algorithm is used:

1. The DCPC algorithm is run for a number of iterations.
2. If there are users that are unsupported one of the unsupported users are removed. The one to remove is randomly selected.
3. The DCPC algorithm is run again. The process is repeated until all remaining users are supported.

This algorithm is similar to the SMIRA algorithm [51]. It has been shown that this algorithm performs almost as well as one that searches all possible combinations of removed users and maximises the number of supported users [52].

Packets are split into smaller blocks that are transmitted one at a time. The blocks that fail are retransmitted. Here we also assume that the acknowledgements are error free and instantaneous.

There is a problem with users that transmit at maximum power, but that cannot reach the quality target. They cause a lot of interference to the other users in the system and thus they should stop transmitting. The problem is to determine when they should start transmitting again. An exponential back-off algorithm is used in various systems to control the traffic in the system. The one used here is similar to the one used in the 802.11b standard [53]. Whenever a user hits the maximum power target he waits for a number of slots until he transmits again.

The number of slots he waits is a random number from 1 to 2^n. Initially n is 1. But if the user tries to transmit and fails n is increased by one. The maximum allowed n is 8. Upon a successful transmission n is reset to 1 again.

4.7.3 DCA system design

Dynamic channel allocation has been used for automatic frequency planning as well as for avoiding interference in unlicensed systems. For example DECT uses DCA.

The problem is to determine which channel to use for communication. There have been a number of suggestions on how this can be done. However here we use a method known as minimum interference [54]. A user listens to all channels and measures the interference, i.e. signal power, on all channels. Then he selects the one with the lowest interference for communication. If during the communication the communication quality becomes too low he repeats the process to find another channel.

In the packet system reassignment is performed every slot if necessary.

4.7.4 System properties - Summary

The various design choices for the various systems can be found in the table below.

	System 1	System 2	System 3	System 4	System 5
Multiple access method	Slow FH	DS-CDMA	Slow FH	DS-CDMA	DCA
Channel selection	Random	N/A	Random	N/A	Least interfered
Transmission power	Fixed	DCPC control	Fixed	DCPC	Fixed
Traffic type	Voice	Voice	WWW	WWW	WWW
Congestion control	N/A	Removal	N/A	Delay transmissions	N/A

Table 4-2 Summary of the key features of the various system designs

5. System performance

5.1 Performance measure - System capacity and cost of communication

One underlying assumption rarely articulated in radio systems research is that the capacity of a system is important. With a limited amount of infrastructure we want to have as many satisfied users as possible. This limitation is what causes all problems (and research opportunities) since it is easy to have happy users given an infinite amount of resources available. But since infrastructure must be paid for and somehow the users of the system are the ones who pay there must be a trade-off between the number of happy users and the amount of infrastructure.

Here we study the cost per user for providing services, i.e. the amount of infrastructure required to support one user. We also look at the inverse measure, the amount of users that can be served by a given amount of infrastructure, also known as capacity. There are two comparisons that we make here. We compare the per-user cost for one operator to another operator where both operators utilise the same spectrum. This should then help us understand how the operators perform on a market. In addition we can compare the per-user cost for one operator utilising unlicensed spectrum to one that uses licensed spectrum. This gives an indication of the competitiveness of using unlicensed spectrum.

In order to do this there are two things we need to determine. First of all what is the total cost of providing users with service. The second important thing is to find out how many users we can serve. First we have a look at the cost.

We make the bold assumption that the cost of infrastructure provision is directly proportional to the number of access points. The main advantage is that it is simple to measure and we can disregard a number of things in the analysis. This is a simplification of course; there are many things that contribute to a cost of providing a service. There are things related to the infrastructure such as purchasing the necessary equipment, deploying the access points and when the network is in place there are operation and maintenance to take care of. These costs may be proportional to the number of access points, however there are many things that are disregarded. For example: volume discounts on access points, sites may be more difficult to obtain, i.e. expensive, when the number of sites increase, operation may be more effective when there are many access points to handle etc. In addition there are other costs not directly related to the infrastructure such as marketing, administration and finance.

The second important part when determining the cost is the number of users served. The amount of users that can be in the system depends on the requirements they have to become satisfied. If they had no requirements we could easily fit any number of users into the system. Thus the number of users and the user requirements are tightly related.

If the requirement is that all users should have the specified service quality the system capacity will become really low. The reason is that there is almost always at least one user in a really bad position, with a lot of interference or high pathloss or both. Another problem is that the system capacity becomes very hard to measure. Since the location of a user is stochastic and the system reaches capacity when one single user is located in an unfavourable position it is easy to realise that the system capacity becomes dependent on one single event. If we rely on numerical experiments to find the capacity we need a lot of experiments to average over. To avoid these problems we require that at least 95% of the users in the system fulfil the quality requirements. To summarise: The cost of service provision is measured as the amount of access points per user with as many users in the system as possible, while keeping the service quality at or above the specified level.

To determine when a user has met the quality requirements is not easy. What we really want to find out is when the quality of the service is perceived so high that the user is prepared to pay the current price for it.

For voice traffic what is commonly done is to measure the (average) signal to interference ratio. By making certain assumptions about the variations of the signal it is possible to map the signal to interference ratio to a frame erasure probability. To determine if a specific frame erasure rate gives satisfactory voice quality listening tests are employed. Here the simplification is used that a user in system 2 (DS-CDMA) is satisfied if the SIR is above 11 dB.

In system 1 (frequency hopping) the SIR will vary for each hop. The SIR is assumed to be constant for the entire frame. If the SIR is above 11 dB we consider the frame to be correctly received and if the SIR is below the frame is considered to be lost. Coding and interleaving is used to remedy the variation. A user is satisfied if 70% of the frames are correctly received.

WWW browsing is more difficult to create an accurate quality measurement for user satisfaction. How a user perceives a data service depends on so many thing. However quality requirements are usually expressed in terms of delay and throughput requirements. Here we focus on the throughput per user. It is calculated as the total number of bits transmitted by a user divided by the total time the user is transmitting. A user is considered to be satisfied if the average throughput is at least half of the maximum channel rate (10 kbit/s).

It can be discussed if this is an appropriate measure to determine if a user is satisfied. But since there is no admission control algorithm implemented the difference between a system where most of the users are satisfied and one where most users are not satisfied is not that large. Thus it does not matter that much which criteria is selected for determining if a user is satisfied.

We define load as:

$$L_i = \frac{N_i}{B_i \eta} \tag{5.1}$$

Where N_i is the average number of users active for operator i, B_i is the number of access points in the network for operator i. Finally η is the number of available channels in the allocated spectrum.

We define cost as:

$$C_i = \frac{1}{L_i} = \frac{B_i \eta}{N_i} \tag{5.2}$$

This measure may seem a little bit strange since it involves the number of available channels. However we may see the channels as a resource that has to be spent in order to serve a user. Even if unlicensed spectrum is usually free to use there may be a fee related to using a channel.

Ns_i is the number of satisfied users for operator i. For all load combinations L_i we can find the fraction of satisfied users, with K operators, as:

$$S = \frac{\sum_{i=1}^{K} Ns_i}{\sum_{i=1}^{K} N_i} \tag{5.3}$$

$$O = 1 - S \tag{5.4}$$

The feasible region is all load combinations that satisfies: $O < 0.05$.

5.2 Cost of providing communication

The performance of system 1 and system 2 are outlined in figure 5-1 and 5-3.

The capacity plots outline the feasible region. The axes and the curve limit the region. Note even though only a part of the graph is plotted the rest of the graph can easily be constructed since operator 1 and 2 have the same cell layout and the same system design. Thus there is symmetry around a y=x line. For any load combination in the feasible region the quality requirements are met. The curve can also be interpreted as the maximum load for operator 2 given a specified load for operator 1. In the cost plots the feasible region is located above and to the right of the curves. These can also be interpreted as the minimum achievable cost for operator 2 given that operator 1 operates at a specified cost.

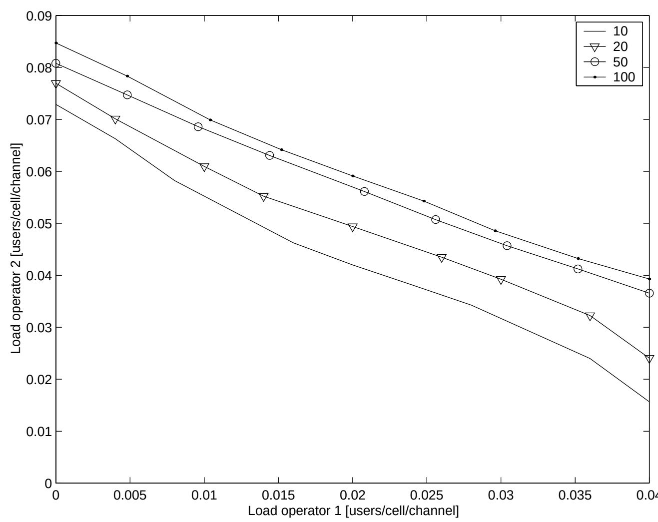

Figure 5-1 Performance of system 1 (frequency hopping with voice traffic) for different available bandwidth, i.e. number of channels.

The capacity of two frequency hopping systems with voice traffic is outlined in figure 5-1. Although the whole feasible region is not outlined we can easily determine the rest of the curve. Since the systems of operator 1 and 2 are identical the axes could be interchanged and thus we could obtain the rest of the curves. We could view this as a mirroring in the line Y=X.

We can note here that the lines are almost straight. Since they cross the axes at the same point (for symmetry reasons) the lines could easily be described as $L_1 + L_2 = C$. In other words the total capacity of both systems is constant for a given allocated bandwidth.

The effects of trunking gains can also be seen. In systems with fewer available channels the per access point and per channel capacity is lower. For example if there are 200 channels available 0.084 users can be server per channel and access point, but if there are only 20 channels available the corresponding number is 0.073.

There are two reference cases we can compare the results to. One case is where all the access point forms one system. The other case is where the spectrum split in two and each part is allocated to each operator. In this specific example there are 50 access points, 25 belongs to operator 1 and 25 belongs to operator 2. In this little example we assume that there are 100 channels allocated. Now assume that both operators have the same traffic load.

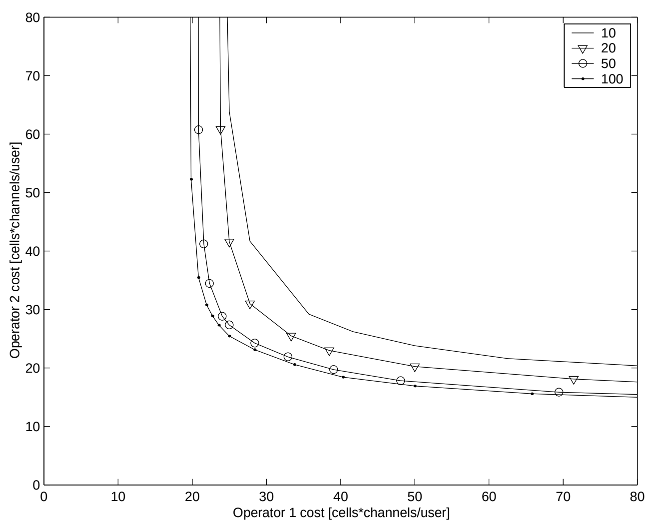

Figure 5-2 Cost per user for system 1 (frequency hopping with voice traffic

If we use shared spectrum we can support approximately 200 users. If we make a static split approximately 190 users can be supported. The reason for the difference is the trunking losses. If the split ratio does not match the load ratio for the different operators the number of supported users will be lower. For example if operator 2 has twice as many users as operator 1 only 140 users can be supported. Here we see the advantage with unlicensed operation. The traffic mix does not influence how many users we are able to serve. However if all the access points had been part of one system we could have supported 400 users.

The cost curve for system 1 (frequency hopping with voice traffic) is outlined in figure 5-2. There are some observations we can make from this figure. The curve converges to some value that corresponds to the cost of a single operator, if he can use the entire spectrum himself. It would seem like the cost approaches infinity, but that is of course not the case. If there are no users in a system at all the per-user cost is not defined and when there is one user the cost is finite.

We can also note that the lowest per user cost for both operators occur when the load is shared equally among them. We also note that the minimum cost only occurs at maximum load. The large variation in terms of cost per user as the number of users in the system varies makes the per-user cost difficult to use.

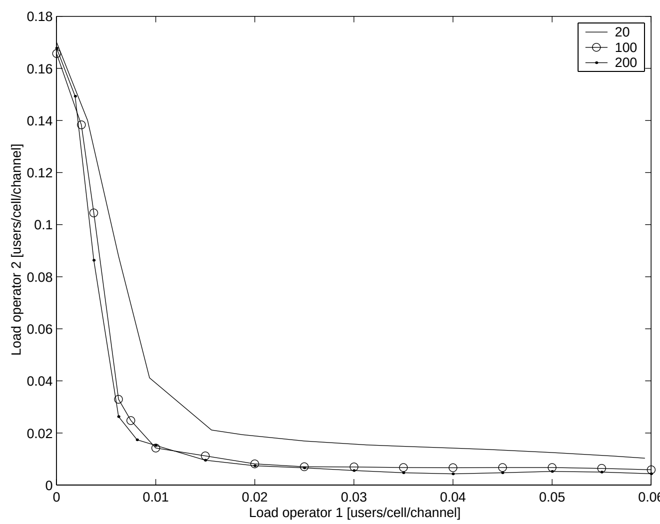

Figure 5-3 Performance of System 2 (DS-CDMA with voice traffic) for different processing gains, i.e. available bandwidth.

When looking at the figures outlining capacity (figure 5-1 and 5-3) one of the most striking differences is the shape of the curves.

There are a number of observations that can be made from the result plots. One observation that can be made is that the orthogonality of the channels is important. Problems occur when users from operator 1 are close to access points belonging to operator 2 and when users from operator 2 are close to access points belonging to operator 1. Since the pathloss to the interferer is small and the pathloss to the wanted signal source is large the adjacent interference suppression becomes important. In system 2 (DS-CDMA) the “adjacent channel” interference is on the order of 20 dB to all other “channels”. In the system 1 (frequency hopping) we have assumed no adjacent channel interference. This is an optimistic assumption and we can expect the performance to degrade when there is adjacent channel interference but the amount of adjacent channel interference is not as large in FH systems.

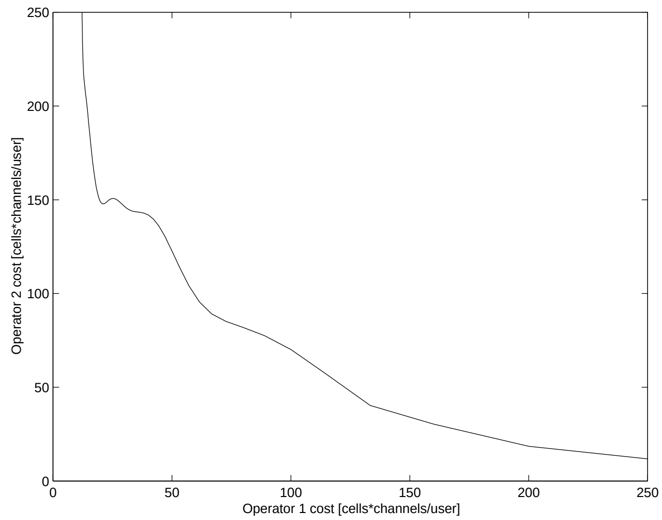

Figure 5-4 Cost per user for System 2 (DS-CDMA with voice traffic) for processing gain 100.

Another observation that can be made for system 2 is that the performance degrades if there is more bandwidth available. To understand this behaviour we need to consider the following scenario. One user belonging to operator 1 is far from the access point and close to one access point belonging to operator 2. This means that operator 1 has to increase the power to overcome the interference. But if operator 1 increases the power that means that there is more interference created for those users belonging to operator 2. So operator 2 has to increase the power on the access points and so on. The closer a user is to the operator the more severe this problem becomes. Now if the processing gain is larger that means that more users can fit in the system. This in turn means that it is more probable that there will be a user far away from the access point and close to the other access point of the other operator. This is what explains the degraded performance when there is traffic in both networks.

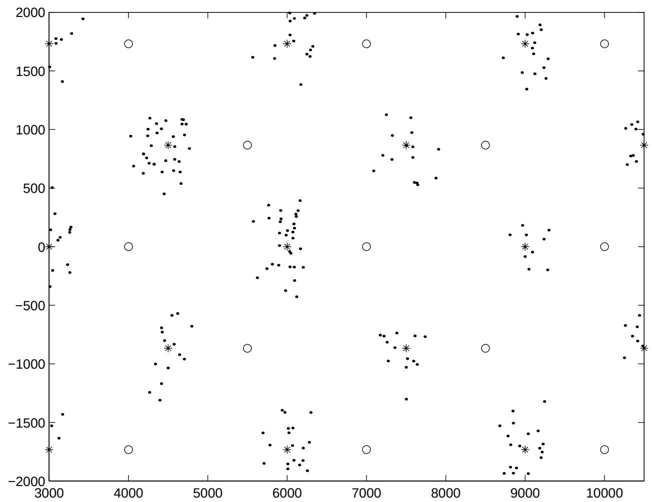

Figure 5-5 Location of removed users in system 2 (dots) belonging to operator 1. The circles are base stations belonging to operator 1 and stars are the base stations of operator 2.

We can also note that the performance characteristics in system 2 are similar to hierarchical cellular systems where both the micro and macro layer use DS-CDMA [27]. The assumptions made for a hierarchical system are a little bit different, the micro layer has a much higher access point density and the power targets are different for both the micro and macro layer. However in these systems we can also see a dramatic decrease in performance when there is traffic in both layers.

In systems 3, 4 and 5 the traffic is data. We can se the same behaviour as in systems 1 and 2. Frequency hopping and DCA gives similar performance as static splitting, but DS-CDMA gives lower total throughput when there is traffic in both networks.

One thing that we can notice is that the degradation in capacity is not as severe for a DS-CDMA network in the case when the traffic is data compared to the case when the traffic is voice. The reason is the back-off algorithm implemented in the system for data communication. This algorithm gives the system a flavour of time division. When there is a lot of traffic the users do not transmit continuously, instead they wait for some time until there is less traffic (hopefully). This behaviour can actually be seen as a form of time division multiplexing, which explains why there is less degradation in the data traffic case.

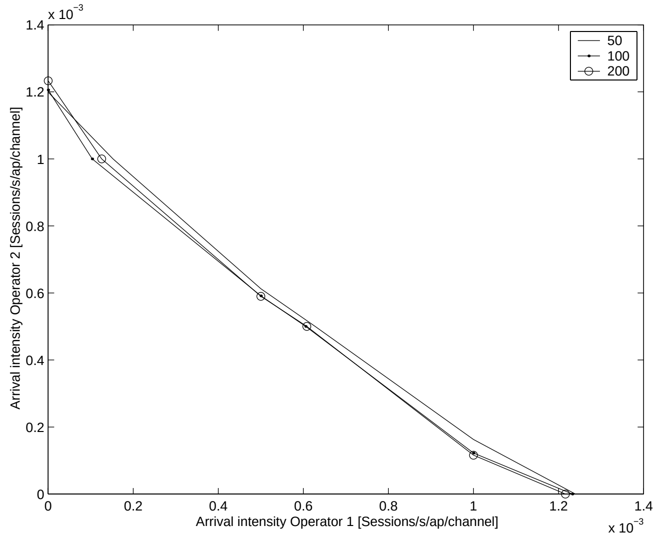

Figure 5-6 Performance of system 3 (Frequency hopping with data traffic) for different number of available channels

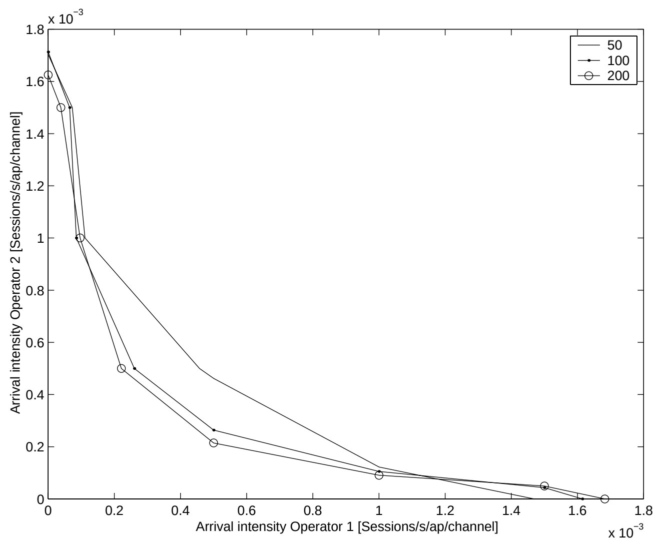

Figure 5-7 Performance of system 4 (DS-CDMA with data traffic) for different available bandwidths, i.e. processing gains.

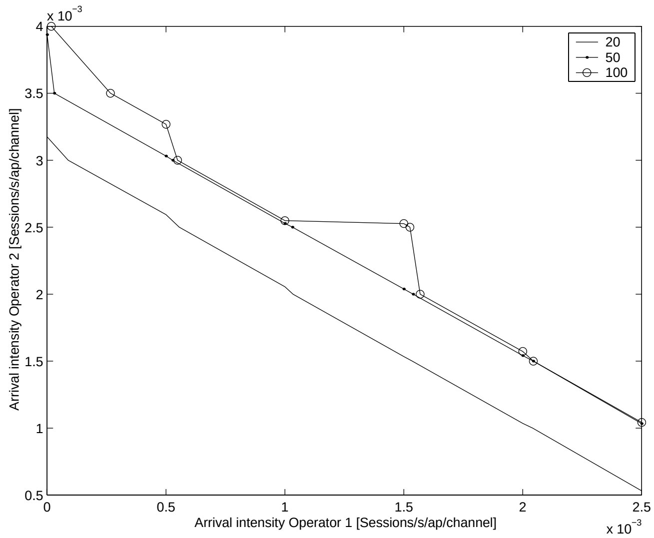

Figure 5-8 Performance of system 5 (Dynamic channel allocation with data traffic) for different amounts of available channels.

We cannot see any trunking gains. This is an effect of the characteristics of the traffic patterns. It has been pointed out previously that the burstiness of data traffic makes aggregated data streams bursty as well. This means that there are no trunking gains to get since the same fraction of spare capacity is needed to be able to maintain the quality requirements.

We can also see that the no system performs better than static splitting although the frequency hopping and DCA systems are close. It is worth noting that the static splitting is static in nature. It is not possible to handle variations in traffic loads, for example if operator 1 has a lot of traffic in the network while operator 2 does not have anything the static split will perform worse.

5.3 Discussion

Here we have studied downlink traffic only. In a practical system there will probably be uplink traffic as well. Depending on how frequency spectrum is allocated it is possible that uplink traffic will occur in the same frequency spectrum as the downlink traffic. Depending on the amount of traffic we can expect a slightly different behaviour from the system. Let us consider a user belonging to operator 1. The interference from operator 2 will now be more spatially spread. It is not only concentrated around the access points, there are some other small “islands” of interference elsewhere. This may actually improve performance since interference is more equally shared between the users of operator 1 and more users may be able to achieve the performance target. On the other hand the area where the interference is high becomes larger, which reduces performance.

In this work we have considered a propagation environment where the propagation is fairly nice. I.e. the distance between mobile user and access point has a large influence in the attenuation of the signals. There are other environments where this is not the case. One example is the Manhattan environment where the propagation along the streets is good, but there is a large attenuation around street corners. When two operators coexist in the same area we can see the same type of interference problems. However the areas where interference is too high will have a different shape. They will follow the streets rather than be close to the access points.

6. Operator size influence on cost

The regulators have been interested in ensuring competition in the telecommunication market. There are a lot of factors that determine if there is a monopoly. Here we do not attempt to cover all the aspects. One class for monopolies is one known as natural monopoles. It is characterised by diminishing average costs. The models used for systems here can be used to calculate average costs and we will look at this in order to see if this factor points towards monopolies or not.

Cellular networks have a different division of cost than most examples in microeconomic textbooks. The part of the cost associated with producing a specific phone call is small. Most of the handling of a phone call is automatic and there are few items that are consumed when a phone call is made. However the fixed costs are rather large. There are large investments in infrastructure and so on. Since the fixed costs are much larger that the variable costs the average cost tend to be decreasing. At the same time the marginal cost is almost zero. This is true when the infrastructure is already in place. However when the capacity of the infrastructure is reached there is a need to build more, which raises cost. Here we make two assumptions. First that the network is always fully utilised and second that the cost of the infrastructure is variable. If the time-period we study is long these assumptions are not that much of a limitation.

The cost per user is obviously influenced by the actual amount of users in the system. Since the number of users that can fit in one system depends on how many users that are in the other system the cost of providing service depends on the traffic balance between the two systems. As a reference case we choose the case when two operators have the same access point densities and the same amount of traffic. Then we increase the number of access points operator 2 has while operator 1 keeps the same amount of access points and the same traffic load. Now operator 2 will be able to carry more traffic. In addition the average cost for operator 2 will be different while the average cost for operator 1 remains the same.

However it is difficult to add just one access point in an existing network since it disturbs the symmetry. Fortunately there are certain hexagonal patterns that can be overlaid on another hexagonal pattern that maintains symmetry nicely. Actually there are many different possibilities but for ratios close to 1 the number of hexagons that have to be repeated must be large. Which in turn means large simulations. Here we have used the ratios 1:1, 1:9/4, 1:3 and 1:4.

The capacity region for two different operator densities is outlined in figure 6-1. The dashed line is the reference load for operator 1. For increasing number of access points for operator 2 we see that we get a different (lower) amount of users that can be supported per access point, for operator 2. This will then correspond to an increasing average cost per user. This is outlined in figure 6-2.

These results can be interpreted in both technical and economical terms. We start in the technical domain. We can see that capacity per access point decreases as the access point density increases. Actually capacity per access point approximately halves when the access point density doubles. The interpretation is that operator 2 is not able to carry more traffic even if the access point density increases.

One possible explanation is that the interference level close to the access points belonging to operator 1 is so high so that it is not possible to support the users belonging to operator 2. Even if the pathgain is affected a little bit by increasing the access point density the difference is only on the order of a couple of dB. This difference does not make a large difference close to the interfering access point. This dead region is what limits the capacity of the system. Thus the gain that can be achieved from adding extra access points is limited.

We can also see that if there is little load in network 1 e.g. 20 % of maximum capacity per access point for operator 2 does not decrease as rapidly when more access points are built. This makes sense. If the load for operator 1 is low there is little interference to the network of operator 2. Then the network should scale nicely. I.e. the capacity per access point is fairly constant.

In the market domain we can make a few observations. There are only two operators active in this market, i.e. and oligopoly situation, which complicates analysis since we cannot use the theories valid for monopolies or for markets with perfect competition. Analysis is further complicated by the interrelation of the operators. They do interact both in the radio spectrum and on the market. The average cost for operator 2 is dependent on the actions of operator 1.

The operator with most access points has the highest average cost. Thus it seems like smaller operators have competitive advantages since the can supply services at lower cost than larger operators. This could in turn lead to a market with many actors.

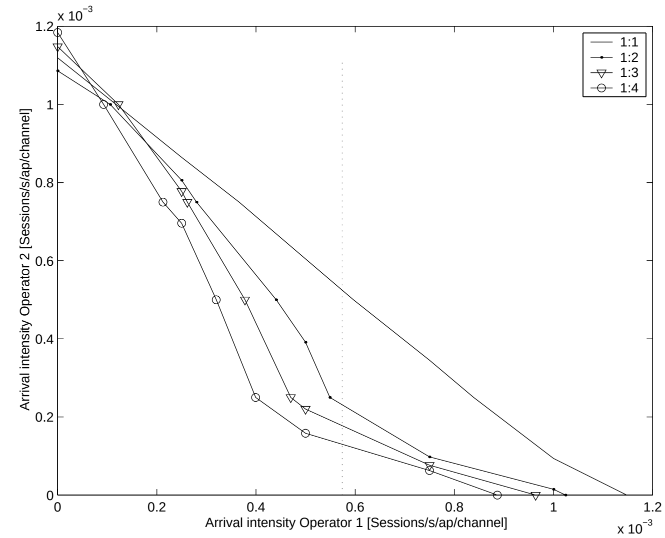

Figure 6-1 Performance of system 3 (frequency hopping with data traffic) for different densities of access points for operator 2.

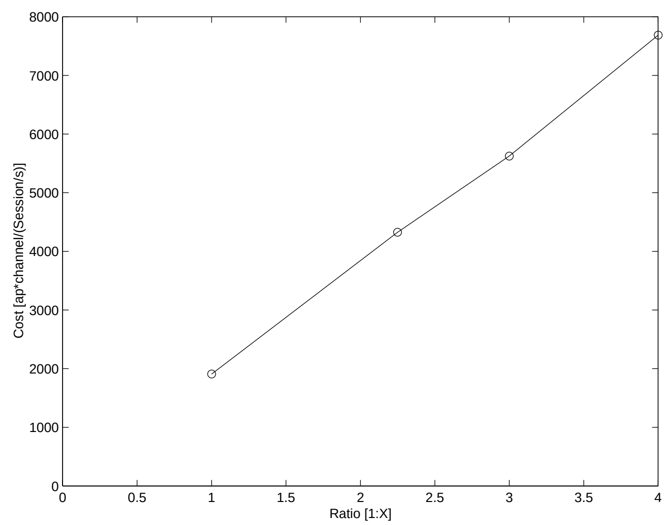

Figure 6-2 Average cost per user for operator 2 for different access point densities.

However there are things that are not covered by this analysis. By looking at a specific point (determined by the traffic load in both networks) and discussing properties in that point we implicitly assume that it is possible to steer to that point. I.e. somehow there is a possibility to jointly control the traffic in both networks. In reality that is probably difficult to achieve.

Around the access points of operator 2 users belonging to operator 1 experiences coverage holes, i.e. regions where communication is bad. Now if operator 2 builds more access points, users in network 1 will experience more coverage holes. This will probably result in a lower perceived quality for those users.

7. Stability and reliability

It is the regulator that sets the rules for using the unlicensed spectrum. This is (almost) the only way to control the behaviour of the operators using that spectrum. Since it is free for all there are no licensed that can be revoked for a misbehaving operator. It is also more difficult to control that operators follow the rules since there are more operators. At the same time following the rules become more important since the spectrum is a shared resource and the behaviour of one operator affect the other.

The regulator is interested in competition and low prices for the end user and tries to set the etiquette rules accordingly. To an operator it is important to be able to predict the performance of the infrastructure. Of course it is not the only factor that influences the willingness of operators to use unlicensed spectrum. However it is one of the factors that determine the willingness of an operator to use unlicensed spectrum.

But we do not only want performance to be predictable, it should also be stable over time. That is it should not be affected by the behaviour of other actors. Typically we want performance to be the same under the same operating conditions. It should also not be affected by the actions of other operators.

It is difficult to span the whole range of possibilities of one operator to affect the other operators. We can think of things that affect the performance of one infrastructure. First there is the possibility that the interaction that takes place between systems in the radio spectrum is such that the performance is unpredictable. This is something that we attribute to stability of the system performance.

Another possibility is that one operator is actively trying to affect the performance experienced by other operators. Actually there are two different cases to this. The first is when an operator tries to get better performance in his system and as a side effect the performance for the other operators is reduced. The other case is regular sabotage; one operator tries to reduce the performance for the other operators as a means to get competitive advantages.

There is another distinction to be made in relation to the etiquette rules for the spectrum. The operator can either abide by the rules or break them. It is reasonable to assume that operators will push the rules to the limit.

However some may also try to get performance gains by breaking the rules. We can divide this “rule breaking” behaviour in two categories. One is changing the system parameters and the other is by actually changing the system design. This line is fuzzy, but still there is a distinction in that the changing of parameters is easier than changing the system design and thus more likely.

Preferably the rules should be designed in a way that discourages breaking them. For example if the performance is degraded by breaking the rules that will probably not be done.

In this thesis we will not give all the answers to rule design. Rather this will be more of the kind of exploratory search in the area. In this chapter there are two things that we are looking for. First it is the behaviour of many systems that coexists. It is not certain that predicting the behaviour of a group of systems is trivial if one knows the behaviour of a single system. The other thing we are looking for is the changes in behaviour and ensemble behaviour if a single system is changed.

7.1 Stability

In this thesis we have defined the capacity by requiring that at least 95% of the users be satisfied with the quality they receive. However doing this requires that we have knowledge of traffic and quality in the networks of both operators. In a practical system that is not the case. One system will only have knowledge of what goes on in that system.

We have not considered any admission control. However in a practical system there will probably be some mechanism that rejects traffic to preserve the quality for the remaining users. If we introduce an admission control scheme that ensures that maximum 5% of the users that are in the system are dissatisfied we get some interesting results. (The remaining users who are thrown out are probably also not satisfied, but that is ignored here.)

If we plot the 5% outage probability level for each operator separately we get the results in figure 7-1. We can now realise that there is a stability problem. For example consider a situation where the load for operator 1 and 2 corresponds to the point marked A. In this case operator 1 experiences an outage probability of more than 5% but the outage probability for operator 2 is less than 5%. The control mechanisms in network 1 will then remove traffic while network 2 will admit more traffic. The new load point will then be located at B. It is easy to realise that we will end up in a situation where one network is fully loaded and the other has no traffic.

However by changing the control laws we can get a different behaviour. If we allow higher outages for lower loads the situation can be reversed. However designing this control law is not an easy task and probably highly situation dependent.

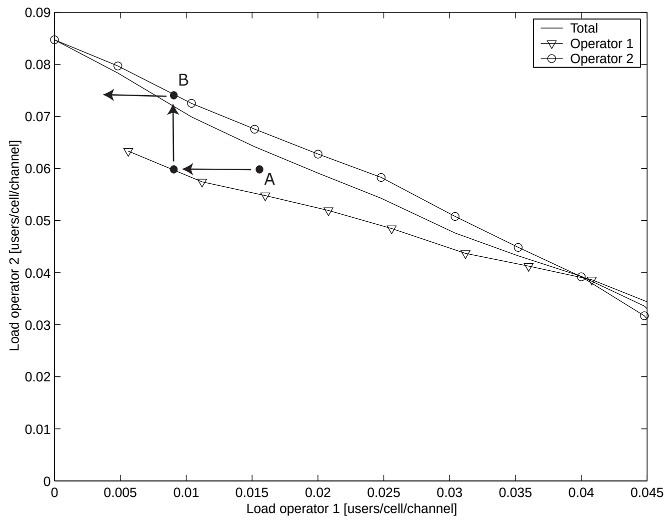

Figure 7-1 5% Outage probability in system 1 for operator 1 and 2 separated. 100 channels.

In this case we can see that there are indeed stability problems. By introducing an admission control scheme we have introduced stability problems, which will eventually only let one operator operate at a time. Thus the etiquette rules must be carefully designed to avoid this type of situations.

In practical systems these stability problems will not be as severe though. The discussion above implicitly assumes that there are always users who want service. That is only an assumption though; of course there are always a limited number of users that want service. Another observation is that users do not really care if they were completely locked out by the admission control scheme or if they have a low SIR. It is the quality they are interested in. Thus the operator must ensure that most of the users are satisfied most of the time. This can only be achieved by providing enough resources for the users. For example by providing enough access points.

It is not necessarily so that the operators will use admission control. One possibility is that all users are accepted, but in high load situations those that are expensive to support are dropped. These users are the ones far away from the access points. Thus in high load situations the coverage for one operator is reduced to patches close to the access points.

7.2 Different ways to break the rules

Here we try to make a classification of different ways to break rules. We assume that the motivation for breaking rules is to achieve higher throughput, i.e. to achieve higher rates. In principle there are two ways to achieve higher rates. One possibility is to increase the used power or by increasing the used bandwidth.

The first way is to increase the transmission power. This mode of rule breaking is fairly evident when we look at etiquette rules that are used today. Most etiquette rules focus on limiting the transmission power in various ways.

The second way to increase the rate is to increase the bandwidth utilised. The spectrum released for unlicensed use is limited, which poses a limitation. However recent releases by FCC to allow ultra wide band systems gives possibilities for using quite large bandwidth. But there may be problems with increasing the bandwidth as well. The interference is varying with frequency, so by avoiding pieces of the spectrum a lot of interference can be avoided.

Since both power and bandwidth are considered to be limited resources the etiquette rules for unlicensed spectrum usually puts a limit on how much can be used. This gives us two possibilities of breaking the rules, i.e. by exceeding power or bandwidth limits.

The third class of ways of breaking the rules is to use the (required) behaviour of other users to obtain advantages. One example is can be imagined where the etiquette rules require a back-off period when a channel is busy. For example if the channel the users must wait a random interval before sensing the channel again. Here we can think that the cheating user is waiting a much shorter period and thus gains access to the channel more often.

7.3 Breaking the etiquette rules – case 1

There are many ways to break the rules, but it is tightly interrelated with the rules that actually are used in a piece of spectrum. In this thesis we have look at examples from the various classes. Breaking the rules by increasing the power is one example that is easy to try.

In figure 7-2 we compare two instances of system 3. I one case both operators use the same power, but in the other case operator 2 has increased the power with 10 dB to 40 dBm.

We can see that there is a small gain in the case when there is a lot of traffic in network 1. But when the traffic is low in network 1 operator 2 gains almost nothing. The reason is that the interference in for users belonging to operator 2 is high close to the access points of operator 1. The interference level here is so high that increasing the output power does not significantly affect the signal to interference ratio. Also when the traffic in network 1 is low there is little to gain since both the interference and the signal power increases by the same amount for users belonging to operator 2.

This example shows that this way to cheat is not that successful.

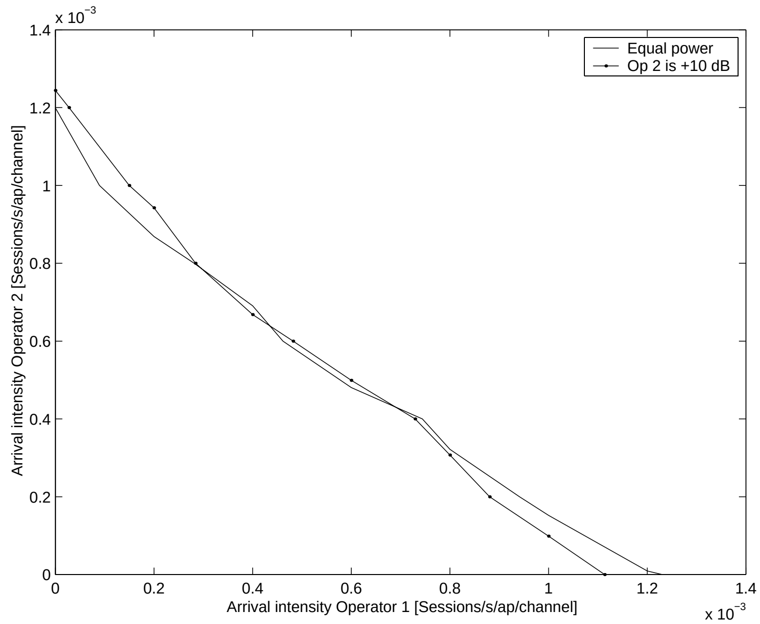

Figure 7-2 System 3 (frequency hopping with data traffic) with different output power for operator 2

7.4 Breaking the etiquette rules – case 2

Using a larger bandwidth is another way to break the rules. However in this thesis we have assumed that all users use the same channel bandwidth. This in turn affects the way the simulation software is written. To investigate this mode of breaking the rules we need to completely rewrite the simulation software. Thus we leave an investigation of this rule breaking for future studies.

Utilising the behaviour of other systems to obtain benefits can be done in many ways. For example consider a CSMA/CA system where each user listens to the channel before transmitting. If the channel is busy the user refrains from transmitting until the channel becomes free. In this system a greedy user simply keeps on transmitting even if there are no data to transmit.

Thus this user will always have access to a channel and there will be no delays caused by waiting for a free channel. It is obvious that this behaviour does not result in the best overall system performance, but for the individual user this behaviour is beneficial. To avoid this kind of behaviour a limit on maximum transmission time is set in the etiquette rules. Breaking the rules, i.e. not adhering to the maximum transmission time limit, in this case is obviously beneficial since the greedy user will force the other (rule abiding) users to back off.

Another example of breaking the rules can be imagined based on the stability example in section 7-1. If the etiquette rules require operators to start removing users when the outage exceeds 5% an operator breaking the rules could simply ignore this requirement. Then the operator who adheres to the quality requirements will eventually be forced to shut out all users. When the networks become loaded and the quality drops one of the operators will try to remedy the situation by shutting users out. This will continue until the overall quality level is again acceptable. Now if the greedy operator has admitted more users than the infrastructure can handle, the rules abiding operator will eventually shut out all users. This is another example of where it makes sense to break the rules.

In every case where an operator is required to limit the traffic in cases of overload this problem will occur. However to be able to provide quality guarantees this kind of rules are necessary. This is a dilemma, it seems like we cannot have quality guarantees at the same time as we have rules that does not help a cheating operator.

7.5 Summary

We can conclude that there are many ways of breaking the rules and some of them can be beneficial for the one breaking the rules. This opens up two interesting research topics. The first one is rule design. The other area deals with the influence this uncertainty has on the willingness of operators to utilise unlicensed spectrum.

8. Conclusions

In this thesis we try to answer the question if unlicensed operation is a feasible mode of operation. We have assumed that there are two networks operating in the same area. There is no attenuation of the radio signals between the two layers. In this situation the mutual interference limits the performance and by using the methods used here the performance is not better than splitting the available frequencies.

We can also imagine a system with access points from both operators where the whole available frequency spectrum can be used. This would correspond to having roaming agreements between the operators and the hand-over should quick enough to create the appearance of one network. Another possibility is to have one firm running the network infrastructure while the operators rent capacity. The total cost of this network would be the same, but it would be able to carry twice the traffic a single network can carry in the spectrum sharing case. Thus the cost per user would be halved.

Here we have not seen any gains in performance compared to splitting the spectrum. However there may be other factors that are non-technical that makes unlicensed operation a good idea. One case where this may be true is where networks are not co-located. Then the whole frequency spectrum is available to both networks.

In our studies we have focused on the shape of the region where the quality requirements are fulfilled. But it is worth pointing out that the absolute values are also important. Here we have not made any effort to ensure that a comparison between the systems is fair. Thus we cannot really se if one system design performs better than another. However before designing a system a study of the absolute values should be performed.

When we compare the resource allocation problem in a conventional cellular network with the scenario with two operators the problem is similar. For example the propagation conditions, user location and traffic are similar. The only thing that differs is that in the latter case the users can only connect to a limited set of access points, thus they cannot always use the nearest access point. This would suggest that conventional wisdom could be used. However even though the difference is small we have seen that sometimes the results are unexpected. These systems have a different behaviour than systems with only one layer of access points. For example adding more access points does not necessarily increase capacity.

We have seen that the problems occur close to access points that belong to other operators. In the examples given above it seems like there is no solution to the problem that only relies on interaction in the air. The consequence is that in systems that rely on unlicensed bands there will be regions where it is impossible to communicate. This stress the need for techniques that can keep users happy even though the underlying communication link may not be there at all times. The cause of the problem is that users are locked to one infrastructure and thus cannot always connect to the closest access point. The other thing to be learnt is that network architectures together with business models that allow a user to connect to the closest access point independent of who owns the access point is important when designing systems since it improves the overall network capacity.

9. Future studies

When looking forward into things that could be studied further there are some areas that can be distinguished. The first and maybe obvious area for deepening the results in this thesis is the area dealing with further understanding the technical behaviour of unlicensed systems.

One of the most striking cases is the one where the propagation environment is different for example in downtown areas or inside buildings. The propagation conditions in these cases are "worse" in the sense that it is not necessarily the physically closest access point that has the lowest path-loss. This would make the interference conditions worse since it is possible to get a lot of interference from transmitters far away. On the other it may be possible to be shielded from a severe interferer close by. Thus simple reasoning may not be enough and more detailed studies are needed. Another case is where the system design is different in the coexisting systems. One system may be more susceptible to interference and it may simply stop working when there is interference from a system that handles interference better. It may also be useful to consider more than two operators.

Yet another thing to investigate is what happens when there is both downlink and uplink traffic. It is hard to imagine an unlicensed frequency band where a part of the band is allocated for uplink or downlink only. Since the networks are not synchronised in any way there will be access points and terminals transmitting at the same time. Thus the interference will be more spread out rather than concentrated around the access points. This could reduce performance but it is not obvious.

There is also the possibility to improve the system design. There is a large number of "tricks" that the system designer can use to improve performance and reduce susceptibility to interference. There are techniques like link adaptation, different power control schemes, improved coding schemes, multi user detection and so on. Although this may produce interesting results there is always the risk that this turns into an endless string of case studies that does not substantially contribute to understanding the feasibility of unlicensed operation.

Finally it may be worth using multi-hop systems. Since the problem are the users located far away from their access points multi-hop radio may help since there are other users located closer by that may be able to help forwarding the communication.

As we can see there are almost an infinite number of things to try within the framework of this problem. However there are also some other areas that are not addressed in this thesis, but that may be of importance in future communication systems. It seems like we cannot have continuous coverage. There will always be situations where a user is far away from an access point and close to an interference source. This calls for techniques that can hide the lack of communication links to the user application.

Another set of problems arise if we remove the constraint that a user can only connect to certain access points belonging to one operator. Removing this constraint enhances the effectiveness of the system since the troublesome cases, communicating with an access point that is not the closest one, is removed. But there are other problems that need to be solved if the constraint is to be removed. There are problems to make agreements between operators to make it possible to roam between access-points. There are also problems to implement roaming between different networks.

There is also a methodological problem when mixing the radio resource management research tradition with economical questions. For example in radio resource management the interesting thing is to find the maximum capacity or to find the limits of the operational region of the system. However it is seldom that a system will be used at maximum capacity. To determine the economical success of an infrastructure there are other measures that are interesting. Typical measures are average usage and possibly also income per user. Thus the methodology needs further work before the paradigm for research in this area becomes clear.

10. References

[1] www.elektrosmog.nu 2003-04-15

[2] D.J. Withers "Radio Spectrum management" IEE Telecommunication series 23. Wiltshire 1991.

[3] F. Gessler "The Development if the DECT standard" Licentiate thesis, Industrial Economics and Management, KTH, 2000

[4] A. Law et al. "Comparison of performance of FA and LIC DCA call assignment within DECT" Networking Aspects of Radio Communication Systems , IEE Colloquium on , 11 Mar 1996

[5] A. Law and L. B. Lopes "Effects of channel history on choice of DCA algorithm within DECT" IEEE 47th Vehicular Technology Conference, 1997

[6] O. Momtahan and H. Hashemi "A comparative Evaluation of DECT, PACS and PHS Standards for Wireless Local Loop Applications" IEEE Personal Communications, June 2000

[7] J. Torsner and G. Malmgren "Radio Network solutions for HIPERLAN/2" IEEE 49th Vehicular Technology Conference, 1999

[8] H. Li et al. "Performance of the Radio Link Protocol of HIPERLAN/2" PIMRC 2000

[9] J. Kuhn-Jush et al. "Overview and Performance of HIPERLAN type 2 - A Standard for Broadband Wireless Communications" IEEE 51th Vehicular Technology Conference, 2000

[10] "Compatibility study between radar and RLANs operating at frequencies around 5.5 GHz" European Radiocommunications Committee (ERC) Report 15, CEPT, 1992

[11] S. Bertelli "Avoiding Radar Disturbance by Radar Detection in HIPERLAN/2 systems" M.Sc. Thesis, Signals Sensors and Systems, KTH 2001

[12] J. Haartsen "Bluetooth - the universal radio interface for ad hoc, wireless connectivity" Ericsson Review nr 3, 1998

[13] S. Zürbes et al. "Radio Network Performance of Bluetooth" IEEE International Conference on Communications, 2000

[14] S. Zürbes "Considerations on Link and System Throughput of Bluetooth Networks" PIMRC 2000

[15] Z. Hadzi-Velkov and L. Gavrilovska "Performance of the IEEE 802.11 Wireless LANs under Influence of Hidden Terminals and Pareto Distributed Packet Traffic" IEEE International Conference on Personal Wireless Communication, 1999

[16] B. Bing "Measured Performance of the IEEE 802.11 Wireless LAN" IEEE Conference on Local Computer Networks, 1999. LCN '99

[17] H. S. Chhaya and S. Gupta "Performance of Asynchronous Data Transfer Methods of IEE 802.11 MAC Protocol" IEEE Personal Communications, Oct 1996

[18] S. Shellhammer "Packet Error Rate of an IEEE 802.11 WLAN in the Presence of Bluetooth". IEEE P802.15-00/133r0. Working document of the IEEE 802.15 TG2.

[19] M. Z. Win "Impulse Radio: How It Works" IEEE Communication letters, No. 2 Feb 1998

[20] O. Wintzell et al "On the capacity of a Pulse Position Hopped CDMA System" ISIT 2000

[21] D. J. Purple "An evaluation of Traffic Throughput in the Asynchronous UPCS band"

[22] L. B. Milstein et al. "On the Feasibility of a CDMA Overlay for Personal Communication Networks"

[23] H. Salgado-Galicia "A Narrowband Approach to Efficient PCS spectrum Sharing Through Decentalized DCA access policies" IEEE Personal Communications, February 1997.

[24] J. M. Peha "Spectrum Management Policy Options" IEEE Communication Suveys, Q4 1998

[25] D. P. Sapathy and J. M. Peha "Performance Of Unlicensed Devices With a Spectrum Etiquette" Globecom 1997

[26] D. P. Sapathy and J. M. Peha "Etiquette Modification for Unlicensed Spectrum: Approach and Impact" 1998

[27] R. S. Karlsson "Radio Resource Sharing in Hierarchical Cellular Systems" Ph.D Thesis KTH 2000, TRITA-S3-RST-0102

[28] R.S. Pindyck "Microeconomics", Prentice Hall, New Jersey 2001

[29] Encyclopedia Britannica

[30] http://www.itu.int/itudoc/about/itu/history/history.txt 2002-07-03

[31] http://www.itu.int/aboutitu/overview/index.html 2002-07-22

[32] Descision Post och Telestyrelsen 15 August 2001 ref: 01-14147 http://www.pts.se/dokument/getFile.asp?FileID=2376

[33] C-F. Helgesson "Making a Natural Monopoly", Doctoral thesis 1999, Handelshögskolan, Stockholm.

[34] J. Zander "On the cost structure of future wideband access", IEEE Vehicular Technology conference 1997

[35] "Föreskrifter om undantag från tillståndspliken för vissa radiosändare" Post och telestyrelsen PTSFS 1997:6.

[36] "The Introduction of Third Generation Mobile Communications in the European Union: State of Play and the Way Forward" European Comission, Brussels, 20.3.2001 http://europa.eu.int/ISPO/infosoc/telecompolicy/en/com2001-141en.pdf

[37] M. Aniander et al. "Industriell Ekonomi" Studentlitteratur 1998

[38] F. Gessler "The development of wirelesss infrastructure standards", Ph.D. Thesis, KTH Stockholm 2002

[39] R. Baskerville et al. *"How Internet software companies negotiate quality"* IEEE Computer May 2001

[40] B. G. Mölleryd "Entrepreneurship in Technological Systems - The development of Mobile Telephony in Sweden" Ph.D. Disseration, Stockholm School of Economics 1999.

[41] J. Björkdahl and E. Bohlin "Financial Net Present Value calculation of UMTS investments", ITS 12th European Regional Conference, September 2-3, 2001, Dublin

[42] M. Flament et al. "Telecom Scenarios 2010" Report, Signals Sensors and Systems, KTH, 1998

[43] J. Zander and S. L. Kim *"Radio Resource Management for Wireless Networks"*, Artech House 2001

[44] J. Zander "Trends in resource management future wireless networks" IEEE WCNC, 2000

[45] W.E. Leland et al. *"On the Self-Similar Nature of Ethernet Traffic (Extended Version)"* IEEE Transactions on Networking, Feb 1994.

[46] K. Blomquist and J.Å. Kjellberg *"A study of Self-Similar Data Traffic and Development of a WWW Traffic Model."* Master thesis, Linköping, LiTH-ISY-EX-1796

[47] E. Carlson " WLAN standard IEEE802.11a MAC scheme Linkadaptation FRIST And Comparison with ETSI HIPERLAN/2" Master thesis, Radio Communication Systems Laboratory, KTH 2000

[48] C. Johansson et al. *"Performance of Different Scheduling strategies in a Packet radio System"* ICUPC 98.

[49] M. Gudmundson "Cell planning in Manhattan Environments" IEEE, VTC 1992

[50] S.A. Grandhi, J. Zander "Constrained power control in cellular systems" VTC 94

[51] T. H. Lee et al. "Downlink Power Control Algoritms for Cellular Radio Systems" IEEE Transactions on Vehicular Technology, Feb 1995

[52] M. Andersin "Power Control and Admission Control in Cellular Radio Systems" Ph.D. thesis KTH 1996.

[53] M. Natkaniec an A. R. Pach "An Analysis of the Backoff Mechanism used in IEEE 802.11 Networks" Computers and Communications, 2000 ISCC 2000

[54] D.J. Goodman "Distributed dynamic channel assignment schemes" IEEE Vehicular Technology Conference May 1993

A. Appendix A

A.1 A short introduction to scenarios

Predicting the future has always been something that people have wanted to do. The methods have varied from looking at the stars to the remains in a coffee cup. However after the Second World War people started to attack the problem in more organized ways. One of the ways to envision the future was pioneered by Royal Dutch Shell in the middle of the seventies. This method known as scenario creation is about creating several possible futures.

A scenario is a description of how the world may look in the future.

Most of the development proceeds as expected. This explains why extrapolation is useful for predicting the future. Things that grow will generally continue to grow and things that shrink will continue to shrink. But sometimes things do not happen as expected. Some things may happen as a result of random acts, but there are also some things that happen because two things develop in a way that creates logical inconsistencies. For example exponential growth of a population will only continue to be exponential as long as there is room for an increasing population. When creating scenarios we use this kind of reasoning to reduce uncertainties.

The first step when creating scenarios is to identify things that are currently happening (trends). The trends are each evaluated to see if the trend is likely to continue, how fast it will continue and which direction it will go. From these trends scenarios can be created. In each scenario there is a different combination of how fast trends has happened, what have occurred and which way things have developed. The combination is checked for inconsistencies and then if there are none the scenarios are "filled out" by drawing the conclusions from the combination of trends.

There are several uses for scenarios. Maybe the most obvious one is to find possible development paths. Another is to make an organization more sensitive to things that are happening. Finally it is a way to make the people who actually do the work more knowledgeable on a specific area.

In this work the main focus has been to identify things that affect the direction the development of the telecommunication sector goes. There are things that at first glance may not seem inconsistent, but when researched further actually are inconsistent. We created three scenarios, one where the development really offers no surprises and two where things happened that made the development path take another way.

A.2 Anything goes

The most striking feature of the "Anything goes" scenario is the rapid development pace. The competition is fierce. The market usually sets standards as de facto standards and in combination with a liberal regulations and competition creates rapid technological development. The cost of equipment and services is low due to global markets and competition. Most people in the industrial world can afford the communication services they want. However it is the end customer who put various pieces of equipment together to fit his specific needs. This means that compatibility and interoperability are important properties when designing systems.

A.3 Big brother

In the world depicted in the "Big brother" scenario the main concern is security. Governments were forced by the public to protect the integrity of the individual. The result was a world where the regulators have a lot of power when defining and setting standards. The way standards are set and the importance they play in technological development results in a development pace that are slower than in the other scenarios. There are few operators around that provide services partially to simplify enforcement of government control and partly because the regulatory structure does not give room for more.

A.4 Pocket computing

In the "Pocket computing" scenario the development pace is rapid. However difference between "Anything goes" and this scenario is the cost of services and who can afford those. The main power lies with service providers that sell a complete service package to customers. For those who can afford it, mainly businessmen, high-speed communication is available globally for those who can afford it.

TRITA-S3-RST-0307
ISSN 1400-9137
ISRN KTH/RST/R—03/07-SE

Radio Communication Systems Laboratory
Dept of Signals Sensors and Systems
Royal Institute of Technology
S-100 44 Stockholm
SWEDEN