Design Challenges of Open Spectrum Access

Konstantinos V. Katsaros, Pantelis A. Frangoudis, George C. Polyzos
Department of Computer Science
Athens University of Economics and Business
Athens 104 34, Greece
{ntinos,pfrag,polyzos}@aueb.gr

Gunnar Karlsson
School of Electrical Engineering and
Linnaeus Center ACCESS
Royal Institute of Technology (KTH)
SE-100 44, Stockholm, Sweden
gk@ee.kth.se

***Abstract*—The use of licensed spectrum for wireless communication is driven by the need to control interference between different operators. However, with this mode of regulation, spectrum utilization is far from efficient and the growth of wireless networks is hindered by the shortage of free frequency bands and the vast investments for the acquisition of a license. In view of this situation, we present an alternative evolution path for the unobstructed growth of wireless networks and the efficient use of spectrum. The proposed architecture is based on the use of unlicensed spectrum and the open access of users to all public networks without prior contracts with operators. We highlight and discuss the inherent technical challenges that must be tackled before the proposed solution can be realized. Special attention is paid to the inherent need for alternative interference mitigation strategies.**

I. Introduction

The great advances in the fields of wireless communications and networking have resulted in a multitude of wireless technologies. The proliferation of the resulting wireless services and the corresponding devices has inevitably led to an increased demand for radio spectrum. The necessary sharing of this finite resource has traditionally been regulated by governmental agencies. Spectrum is split into bands, mostly licensed, with a few unlicensed ones. Licensed bands are allocated long term to particular radio standards or applications and are further divided into assignments to individual licensees which thereby hold exclusive access rights on them. Access to unlicensed bands is unrestricted and this (together with the low deployment cost) is an important reason for the proliferation of certain wireless technologies, such as the popular IEEE 802.11 family of protocols.

This mode of spectrum regulation presents significant problems that hinder the seamless growth of wireless networks. With most of the spectrum being already allocated, it is nowadays difficult to find a vacant frequency band for the deployment of a new wireless service or the enhancement of an existing one [1]. Even if a frequency band is available, the competition for its lease is usually high, leading to vast investments for the corresponding license. In consequence, new operators face a high barrier to enter the market and established operators face the burden of the long payback time of their investment. Hence, inter-provider competition is limited, leading to noncompetitive service offerings and slow decrease of usage tariffs. At the same time, customers are tied to exclusively access the network of their provider lacking the flexibility of dynamically choosing the service with the best price-quality relationship.

Furthermore, studies show both temporal and geographical underutilization of the spectrum [2]. This means that, in certain areas and/or periods of time the available spectrum is only partially occupied by the license owners and their subscribers, or not occupied at all. On the other hand, the increasing popularity of technologies for unlicensed parts of the spectrum, such as IEEE 802.11a/b/g, has led to dense urban deployments incurring significant interference problems (Fig.1) .

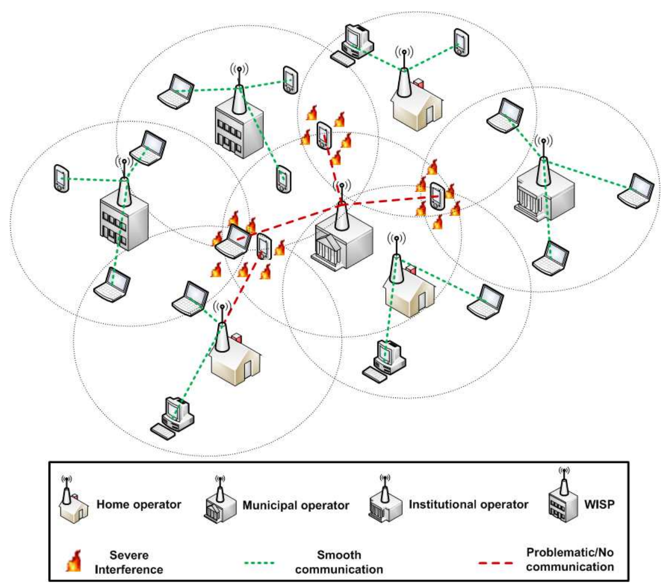

Fig. 1. Dense urban deployments suffer from significant interference.

We believe that a new evolution path is required to promote the growth of wireless networks in terms of geographic coverage, service offerings, number of users and traffic volumes. To this end, we explore an alternative spectrum utilization model based on the following two basic premises. (1) We assume a model based on the use of unlicensed spectrum. Thus, spectrum allocation becomes a non-issue and in principle anyone may become an operator by offering coverage over an area. (2) A user has open access to all public networks without subscription or any other form of prior contract with operators. This permits users to negotiate and buy wireless network access in small quanta, as by seconds or kilobytes.

Along these lines, a continuing wireless deployment will arise from the fact that operators in competition will position themselves, both geographically and market-wise. The wireless Internet composed of these operators will therefore be heterogeneous and will be able to evolve steadily in functionality when services move from the high-end market to the mass market. The network will also expand in coverage as operators move into new territory in order to escape competition and to gain first-mover advantages.

In this paper, we highlight and discuss the inherent technical challenges that must be tackled before the proposed model can be realized. Special attention is paid to the fact that the lack of regulation introduces the need for alternative interference mitigation strategies.

The remainder of this paper is organized as follows. In Section II we provide an overview of the proposed architecture together with an illustrative case study scenario and a careful investigation of its functional requirements. In Section III we focus on dynamic spectrum access, discussing various approaches. Section IV provides a short summary and a roadmap to our future work in the area.

II. The proposed architecture

A. Design aspects

The proposed architecture is designed along the following three major axes. The common denominator is the focus on the user perception of the network operation.

1) Utilization of client-supplied information: Most existing interference mitigation schemes are Access-Point-centric in nature. Such apporoaches utilise spectrum usage information as measured by the deployed Access Points (APs), thus exposing interference experienced solely at APs' sites. The conditions at the exact client locations are not revealed. This *hidden inteference problem* [3] can be more efficiently solved with client feedback.

Roaming users can record network coverage information and report them to a central or distributed coverage database. The reports of covered parts can include estimates of interference levels and congestion, detected service capabilities of the network entities and their pricing structures. The number and content of reports from uncovered areas indicates the traffic demand that has not been satisfied due to the lack of a network infrastructure. Network operators can consult the coverage database to determine good locations for deploying new cells. This reporting system together with the human agents that may deploy the access points (APs) form an *outer feedback loop* of the overall architecture.

2) Adaptive wireless infrastructure: The *inner feedback loop* concerns the operation of the deployed wireless infrastructure. Since the deployment is organic and based on user reports and some modicum of externally provided information, it may occur that a newly installed AP "‘causes"’ interference to the installed base and suffers from it. Such interference can be detected by the APs themselves and through client reports. The more nodes affected, the more reliable are the estimates of the interference and the sooner it will be reliably spotted and resolved. The inner loop builds on these reports to resolve the contention between the established base and the new APs through channel selection, power control, directional antennas, time sharing, or a combination thereof. Radios might be software defined to adapt their coding and modulation and in general we expect the mobile nodes to be adaptive with respect to frequency and air interface.

3) Service discovery, negotiation and handovers: The third and final aspect of the proposed model relates to the behavior of mobile users. A mobile node must be able to discover the network service and to negotiate access with the provider of its choice at a given location. Then it must be able to hand over from one provider to the next due to mobility, service dissatisfaction or opportunitstically (e.g. availability of lower price service or improved QoS etc.). Coverage maps constructed from user reports may serve here as well. A mobile node may query maps to plan ahead for expected handovers. Vertical handovers are also possible, assuming nodes with multiple or reconfigurable air interfaces.

B. A simple motivating case

Imagine a square in a densely populated city. Nearby wireless APs, connected to fixed broadband lines, cover areas of the square using directional antennas and provide various differentially charged Internet and location-based services to passers by. They belong to a heterogeneous crowd of operators that may include the following:

- Residential WLAN owners who share their WLANs on an altruistic or on a for-profit basis.
- Fixed ISPs or 3G operators who "lease" residential APs to extend their coverage and provide a richer set of services to their subscribers.
- Municipalities offering broadband services to citizens.

APs advertise information regarding their operation and services (e.g. spectrum bands that they can operate on, the type and price of the offered services etc.). A passer-by, now, wishes to place a VoIP call through his dual-interface (3G/WLAN) soft phone. The device will scan for advertisements from APs in range and combine them with user and application (i.e. VoIP) profiles to pick the optimal AP, also considering service prices. Then, it will set up a "mini-contract" with the selected AP for using the service via appropriate protocols.

The mobile node also periodically scans for wireless presence and reports his sensings to the reporting system. APs consult it to get updated information about spectrum usage in their vicinity and adapt their operating parameters for optimized performance. Each time a new prospective provider is detected in the area, APs self-organize to reach a state where interference is minimized. They may need to adjust their transmission power, change their operating frequency or tune their directional antennas to cover a different part of the square. Or, they may engage in negotiations to share spectrum on a time basis.

C. Functional requirements

The system we envisage is composed of mobile nodes, APs and the reporting system/spatial database. They are depicted in Fig. 2 along with the message paths for interference control and reporting and queries of coverage. It should be clear that the functions that we are designing pertain to the control and management plane of the system components.

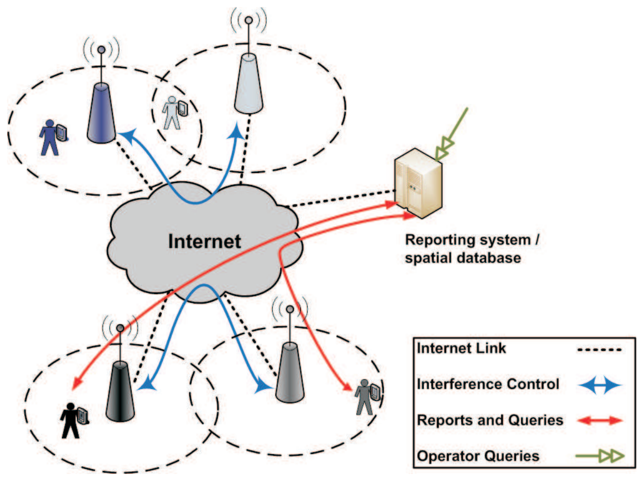

Fig. 2. Interference is controlled in a distributed manner by the APs; mobile nodes report experienced interference, coverage and service availability to a server. The coverage maps in the server could be queried by roaming mobile nodes and by operators.

1) Mobile node: Mobile nodes can be mobile phones, laptop computers, PDAs and other hand-held computers, cameras and media players with wireless communication capabilities. The roles that a mobile node must handle are the following.

- The node must be able to perform spectrum sensing and service discovery in order to identify wireless presence and service offerings at any given location.
- The node should report the above information to a spatial database. Absence of service is particularly important to report[1] since it helps identifying white spots with respect to coverage and the node's communication abilities.
- It must be able to "tune in" to a selected service offering. The node's spectrum profile, which encodes its capabilities and requirements, will be matched with service advertisments by APs to come to the smartest AP choice. The more spectrum agile the node is, the more service offerings will be available to it.
- If desired, the mobile node user must be able to authenticate himself and certify that he will pay for the access provided.
- The mobile node must be able to perform handover. The most rudimentary form of handover is to allow an association to terminate (usually due to mobility) and to perform a new service discovery. With handover, we shall therefore mean a preparation for a new association while still being associated.

[1]These reports can be stored on the mobile node and submitted to the spatial database when connectivity is (re-)established.

2) Access point: This is the name we choose to use for the infrastructure nodes and it does not imply that they must be Wi-Fi or Bluetooth nodes; hence an AP may run any and several air interfaces and data-link protocols. An operator has one or more APs in order to offer service to mobile nodes in the places where it has decided to be active. The roles that an AP must handle are the following.

- The AP must announce its service to mobile nodes in the form of a spectrum portfolio. This includes the AP's transmission capabilities and its service offerings.
- It must authenticate a mobile node and verify proofs of payment, when desired.
- An AP must accept interference feedback from the reporting system. The AP must also report interference that it can detect directly.
- It must engage in interference control in order to limit any occurrence of interference in which it is involved.
- The AP must also be able to communicate with mobile nodes to warn them about imminent (possibly vertical) handovers when the interference control causes the AP to change frequency band, air interface or perform any other disruptive action. Such handovers may lead to a re-association with the same AP or a cause a new association with a neighboring one.

3) Reporting system/Spatial database: This system component handles reports from mobile nodes and APs regarding indications of interference. It also collects information about the services that mobile nodes detect in the field. All received information is compiled into a database that can be queried. The system may be centralized (as in Fig. 2) or distributed among confederated operators. It has the following roles with respect to the other system components.

- It must be able to aggregate reports of interference and available services from mobile nodes and APs, identify the involved transmitters and record the information into the spatial database.
- It must be able to reach APs and provide as detailed information as possible from the aggregated reports to the interference control mechanism.
- The system must monitor the control system for compliance with previous requests for control actions and log detected violations.
- The database answers mobile node and operator queries on service availability and spectrum usage in an area.

It must be noted at this point, that in the functionality of the reporting system, entities are assumed to be location-aware. This can typically be achieved via GPS, which tends to become ubiquitous in modern handheld devices. However, exact positioning information is not always necessary. One's location can be inferred by the signals received by known reference points (e.g. registered APs[4]). Exploiting location-awareness is important for optimized AP configuration and more efficient interference mitigation.

Another, important aspect of the reporting system relates to truthful reporting. Having proposed a client-assisted interference mitigation scheme, one of our major concerns is to avoid fake information reporting. The straightforward approach to the problem is to accept only authenticated client reports. If we assume a central authority responsible for creating user accounts and managing a *Public Key Infrastructure (PKI)*, this task is simplified. In some centralized environments, this approach comes in handy. Digital certificates can be used by users to identify themselves and digitally sign their reports. This excludes unauthorized users who wish to launch false information or denial of service attacks to the system. Also, it helps track down registered users who repeatedly report false information; if the majority of users are truthful, misbehaving nodes' fake reports can be filtered out. Then, misbehavers can be excluded from the system or their reports can simply be ignored.

The above discussion reveals another non-trivial requirement for the interference reporting scheme, namely, (loose) time synchronization among reporting nodes and the spatial database. The significance of this requirement is more obvious considering the dynamicity in spectrum usage in the environment that we envisage.

III. Dynamic Spectrum access

Several research approaches have emerged in an effort to alleviate the aforementioned deficiencies of the current spectrum management regime. These approaches present significant differentiations in various aspects, which are crucial in the design of an interference mitigation mechanism. We identify and discuss these aspects in the following. [2]

A. Spectrum sharing dimensions

Spectrum sharing can be performed in all dimensions of the *spectrum space* i.e. frequency (including code), space and time. Note that, interference is caused when more than one users' spectrum access coincides in all dimensions.

In the frequency dimension, the focus is on the efficient and fair *channel* allocation. Spectrum is divided in non-overlapping slices (channels) and proposed schemes aim at selecting the appropriate channel for each transmitting entity so that interference is minimized. In this context, spectrum allocation is usually reduced to a graph coloring problem [7]. Unfortunately, the global, i.e. network wide, optimization problem is NP-hard and therefore approximating algorithms have been proposed [8].

In the space dimension, spectrum sharing consists in power control techniques and/or the use of directional antennas. In the first case, the target is to dynamically adjust the transmitting power of a node, effectively controlling its transmission range to minimize interference with neighbor nodes using the same frequency [9]. In the second case, directional antennas allow for the use of a certain frequency band in a well specified area improving this way the frequency reuse factor.

In the time dimension, spectrum access is dynamically controlled on a time basis. Here, access is based on the concept of spectral leases [10], [11]. It is noted however, that this type of sharing is performed at a lower granularity than that of a frame/packet since the latter falls within the functionality of the superjacent Medium Access Control (MAC) protocol (e.g. TDMA).

We believe that a unified framework considering all dimensions will provide the necessary flexibility to achieve interference mitigation in a highly competitive environment where multiple operators and end users attempt to access the unlicensed spectrum.

B. Opportunistic vs. Open Spectrum Access

A major body of research has been devoted to the Opportunistic Spectrum Access model [5] which focuses on the sharing of licensed spectrum. In this approach users are categorized with respect to their spectrum access rights. *Primary* (or licensed) users have exclusive access rights to the spectrum (when active), and *secondary* (or unlicensed) users have only opportunistic spectrum access rights given that they do not interfere with primary users. In this asymmetric model the technical goal is to identify *spectrum holes*, i.e. frequencies not occupied by a primary user at a particular time and geographic location, and fill them with secondary user access. The Open Spectrum Access model focuses on sharing *unlicensed* spectrum. Here, no priorities exist since users have equal rights on spectrum access. This results in a symmetric model where the target is to provide either a fair or an efficient (or both) solution to spectrum access for all users.

We believe that focusing on the Open Spectrum Access model presents a significant advantage towards the wide deployment and growth of wireless networks by enabling new (micro-)operators entering the market. To this end, our effort targets at the alleviation of the inherent technical challenges stemming from the non-differentiated access rights and the resulting contention.

C. Centralized vs. distributed

An important aspect of a dynamic spectrum access scheme refers to whether decisions on spectrum access are taken based on complete or partial information about the current utilization of the spectrum, perceived interference and access demand. In this respect, spectrum sharing approaches can be classified into two main categories, namely Centralized and Distributed.

In centralized approaches (e.g. [12]) a central entity is responsible for gathering spectrum utilization information from end-devices (clients, APs), for allocating spectrum and for controlling spectrum access based on global information. Although their performance is optimal due to the acquisition of network wide information, reporting to a central server may not be practical due to communication and/or computation overhead.

In distributed approaches (e.g. [7]) no centralized infrastructure is assumed, and spectrum access is controlled by each entity locally based on own or common policies/rules (imposed by a MAC protocol or a higher level mechanism

[2] It is noted that an exhaustive review of the research area is out of the scope of this paper. Readers are referred to [5] and [6] for further information.

such as pricing) and neighbor information. Distributed approaches reduce reporting overhead, but achieve suboptimal performance due to the lack of complete information.

We believe that gathering spectrum related information constitutes an invaluable tool both in terms of an outer and an inner feedback loop (See Section II-A). Hence, a significant challenge concerns the intelligent design of a low overhead reporting system (See also Section II-C3).

D. Cooperative vs. non-cooperative spectrum sharing

Special attention must also be given to the degree of cooperation assumed between the participating entities, especially in the case of unlicensed spectrum sharing, where no intrinsic access priorities are enforced. Cooperation is usually expressed in the form of interference information exchange, compliance with predetermined spectrum policy rules and/or willingness for individual performance degradation.

On the one hand, there are approaches (e.g. [11]) that assume full cooperation of the entities sharing the spectrum. The ultimate goal is the optimal or near-optimal network wide spectrum utilization. On the other hand, non-cooperative approaches (e.g. [8]) assume selfish entities pursuing the maximization of individual utility, regardless of the efficiency of spectrum utilization in the network as a whole.

For the case of information exchange, in centralized approaches, the involved network entities cooperate, for instance, by aggreeing to provide truthful interference information required for the operation of the spectrum sharing scheme to a single central entity. In distributed environments, an expression of cooperation is the exchange of information for the achievement of a local solution among interested parties.

We believe that an important requirement in the design of a dynamic spectrum access scheme for the unlicensed spectrum, is to design incentives that will lead to a high degree of cooperation between competing spectrum users.

E. Game theoretic modelling of spectrum sharing

Going a step further, the absence of cooperation in scenarios where the available spectrum is shared among selfish entities (such entities are also referred to as rational, or self-interested) has resulted in approaches employing Non-Cooperative Game Theory. In this context, a well-studied problem in the literature is power control to limit interference (e.g. [13]). The typical case analyzed comprises a set of nodes transmitting at the same frequency, each of which has to decide on its transmission power level; a high such value increases the node's Signal-to-Interference Ratio, but on the other hand increases the node's cost and the interference it causes to others. In other approaches [14], cooperative game theory (exact potential games) is used to address the problem of joint channel selection and power control in cognitive radio networks. The relevant works often use specialized tools of Game Theory, such as potential games, supermodular games, etc.

We believe that a game theoretic investigation of the various degrees of cooperation, expressed by the amount and quality of the available information between the players, is of major importance. In this respect, the translation of a game-theoretic model to a practical system is also considered as a challenging task.

IV. Conclusions

In this paper, we have described a new architectural model for creating a multi-access market where users can choose among different access technologies and networks and where operators compete for the users based primarily on coverage, QoS and price. Our grand goal is to form a blueprint for a new paradigm for mobile wireless networking where networks and services can be created dynamically and new technology can be deployed as soon as it is available. In this context, we have pointed out important research challenges that have to be faced in order to provide a better evolutionary path for wireless network deployment, to increase efficiency in the usage of the radio spectrum, and to promote innovation in wireless networking unfettered by today's generational model. In the context of unlicensed spectrum, these challenges focus primarily on the design of efficient interference mitigation and network planning mechanisms based on information provided by the end-users. To this end, our initial efforts focus on the detailed design of the reporting system and the investigation of the impact of sensing and reporting procedures on user-perceived application performance.

References

[1] A. Ghasemi and E. Sousa, "Collaborative spectrum sensing for opportunistic access in fading environments," in *Proc. DySPAN 2005*, November 2005, pp. 131–136.

[2] M. McHenry, "Spectrum white space measurements." New America Foundation BroadBand Forum, June 2003.

[3] A. Mishra, V. Brik, S. Banerjee, A. Srinivasan, and W. A. Arbaugh, "A client-driven approach for channel management in Wireless LANs," in *Proc. IEEE INFOCOM 2006*, April 2006.

[4] "Skyhook Wireless website." http://www.skyhookwireless.com.

[5] Q. Zhao and B. Sadler, "A survey of dynamic spectrum access," *IEEE Signal Processing Magazine*, vol. 24, no. 3, pp. 79–89, May 2007.

[6] I. F. Akyildiz, W.-Y. Lee, M. C. Vuran, and S. Mohanty, "Next generation/dynamic spectrum access/cognitive radio wireless networks: a survey," *Computer Networks*, vol. 50, no. 13, pp. 2127–2159, September 2006.

[7] "HD-MAC project website." http://www.cs.ucsb.edu/ htzheng/cognitive/HDMAC.html.

[8] H. Zheng and L. Cao, "Device-centric spectrum management," in *Proc. DySPAN 2005*, November 2005, pp. 56–65.

[9] V. Mhatre, K. Papagiannaki, and o. B. Franc, "Interference mitigation through power control in high density 802.11 WLANs." in *Proc. IEEE INFOCOM*, 2007, pp. 535–543.

[10] X. Jing and D. Raychaudhuri, "Spectrum co-existence of IEEE 802.11b and 802.16a networks using the cscc etiquette protocol," in *Proc. DySPAN 2005*, November 2005, pp. 243–250.

[11] M. Buddhikot, P. Kolodzy, S. Miller, K. Ryan, and J. Evans, "Dimsumnet: new directions in wireless networking using coordinated dynamic spectrum," in *Proc. WoWMoM 2005*, 13-16 June 2005, pp. 78–85.

[12] O. Ileri, D. Samardzija, and N. Mandayam, "Demand responsive pricing and competitive spectrum allocation via a spectrum server," in *Proc. DySPAN 2005*, November 2005, pp. 194–202.

[13] R. Menon, A. B. MacKenzie, R. M. Buehrer, and J. H. Reed, "A game-theoretic framework for interference avoidance in ad hoc networks," in *Proc. IEEE GLOBECOM 2006, San Francisco, CA, USA*, 2006.

[14] N. Nie, C. Comaniciu, and P. Agrawal, "A game-theoretic framework for interference avoidance in ad hoc networks," *in Wireless Communications (The IMA Volumes of Mathematics and its Applications, P. Agrawal, D. M. Andrews, P. J. Fleming, G. Yin, and L. Zhang, eds.)*, Springer, November 2006.