BBN Technical Report No. 8410

XOSA: eXploiting Opportunistic Spectrum Access for Wireless Ad Hoc Networks

Job No. 86563005

December, 2004

Prepared for:

DARPA
3701 North Farifax Drive
Arlington, VA 22203-1714 USA

Prepared by:

Internetwork Research
BBN Technologies
10 Moulton St.
Cambridge, MA 02138-1191 USA

XOSA: eXploiting Opportunistic Spectrum Access for Wireless Ad Hoc Networks

C. Santivanez S. Ramanathan D. Ugarte R. Krishnan M. Condell
S. Polit C. Partridge *

Internetwork Research
BBN Technologies
10 Moulton St.
Cambridge, MA 02138-1191 USA
csantiva@bbn.com

December 2004

Abstract

Current rigid and inefficient allocation of spectrum has lead to spectrum underutilization and an apparent shortage. A new paradigm, Opportunistic Spectrum Access (OSA), has emerged as the potential solution to the apparent spectrum scarcity problem. In OSA, radios identify unused portions of licensed spectrum, and utilize that spectrum without adverse impact on the primary licensees. OSA promises a significant improvement on spectrum utilization. However, while conceptually simple, OSA turns out to be a very complicated concept to realize, specially under a dynamic mobile ad hoc network. Harnessing OSA potential requires key issues to be addressed in a systematic fashion. In this paper we present XOSA ("eXploiting Opportunistic Spectrum Access"). XOSA is a highly modular suite of mechanisms for opportunistic spectrum access on mobile ad hoc networks. XOSA consists of several mechanisms: opportunity identification, neighbor discovery, opportunity dissemination, and opportunity allocation, all working cohesively to provide the first complete system solution. In addition, we present high fidelity simulation results showing that an order of magnitude increase on capacity can be achieved for realistic scenarios. Based on our experiments, we also identify and discuss the most promising areas for future research: topology control, efficient MAC and IDLE channel allocation algorithms, and underlaying. In addition, our simulation results may guide policy makers on the determination of suitable policies that trade off a licensee's rights and the commonweal (i.e. increased capacity enabled by efficient spectrum usage).

*This work was supported by the DARPA XG program under contract number F30602-03-C-0139.

1 Introduction

Currently most of the spectrum is rigidly allocated. A particular frequency is to be used by a particular service, and only by those licensed to provide that service. Not surprisingly, this approach to spectrum allocation results in large portions of the spectrum being left underutilized. Spectrum can be underused spatially: an assigned spectrum may only be used in certain geographical areas (e.g. certain television channels are active in some cities and idle in others). Spectrum can also be underused temporally: an assigned spectrum may only be used intermittently. For example, emergency radios have special frequencies allotted to them according to region, but are only used occasionally. Another example of intermittent spectrum use are earth observation satellites that communicate with ground stations only when they pass over at fixed periodic intervals and durations determined by their orbit.

Opportunistic spectrum access is an approach to improving spectrum utilization. The core idea is that a device first examines the portion of the spectrum it wishes to use and characterizes the presence, if any, of the primary users. (*Primary users*, also called *incumbents* are those who hold the rights or license to the portion of spectrum being examined.) Based on its understanding of how the portion of spectrum is being used, the device identifies communications opportunities. These opportunities are frequencies, times, or even codings that can be used without causing unacceptable interference with the transmissions of the primary users.

While conceptually simple, opportunistic spectrum access turns out to be a very complicated concept to realize. One (of the many) complicated questions is level of dynamism one wants to achieve. As a simple example, consider two possible scenarios. One is the use of unused television channels. The allocation of the television spectrum is fairly static (while new channels pop up on cable routinely, new broadcast channels are rare). A radio could safely check a database of television use in its area once a day and know which channels were safe for (re)use. In the other example, consider the case of opportunistically sharing with an active cellular telephone provider. When cellular traffic traffic is low, one or more of the cellular frequency bands may be free for reuse – but the provider wishes to be sure that should load increase, the opportunistic users will immediately stop using bands. As cellular traffic can be quite dynamic, this requirement means opportunistic users must be continuously monitoring the state of the frequency bands.

In this paper, we are interested in situations similar to the second example: environments where opportunistic users in a (possibly mobile) ad-hoc network seek to make use of a highly dynamic portion of the spectrum. Our contribution is a system called XOSA (eXploiting Opportunistic Spectrum Access for ad hoc networks).

2 Related Work

Recently, the subject of Opportunistic Spectrum Access has derived considerable attention, motivated by the FCC release of its Spectrum Policy Task Force (SPTF) report [1] and a Notice of Proposed Rule Making (NPRM) on the matter of "Facilitating Opportunities for Flexible, Efficient, and Reliable Spectrum Use Employing Cognitive Radio Technologies". Both documents recommended changes that greatly support the vision for opportunistic spectrum access. At the same time, the Department of Commerce, by order of the President of the United States, has started a review of the governmental sector spectrum usage [2]. And in 2002, DARPA launched its neXt Generation (XG) program – on which much of this paper is based. XG goals are to "develop both the enabling technologies and system concepts to dynamically redistribute allocated spectrum along with novel waveforms in order to provide dramatic improvements in assured military communications in support of a full range of worldwide deployments [3]."

Most current solutions for spectrum sharing are not general but 'ad hoc'. For example, dynamic frequency selection (DFS) - a method being developed by the IEEE 802.11h subcommittee, provides a harmonized set of rules for Wireless LANs to share the spectrum with protected users (mostly military radar) in the 5 GHz band. DFS detects other devices using the same radio channel and switches to a new "clean" channel if required. The protocol has mechanisms for the access point to instruct the terminals to switch to the new channel.

The works in [4, 5, 6] address etiquette protocols designed so that different technologies (e.g. 802.11.x, 802.15.x, Bluetooth, Hiperlan etc.) can coexist on the same band. A spectrum etiquette is a set of rules to be followed by all users of the spectrum so that fair and conflict-free access to the radio resource is enabled. For instance [4] proposes the use of a common coordination channel located at the edge of the unlicensed band and defines a protocol for the announcement of radio and service parameters.

The IEEE 802.18 Study Group is working toward a standard for opportunistic access in the TV band, especially those that will fall vacant in the transition from analog to digital TV.

The XG program, on the other hand, is developing a general ontology and representation format for machine-readable policies. The XG framework [7] allows reasoning about policies. The work is unique in that it decouples behaviors from policies and allows real-time control of behaviors through machine readable policies (policy agility).

While not indispensable for policy-driven operation, the concept of cognitive-radio is closely related to it. Indeed, cognitive radios may represent the more sophisticated instantiation of a policy-driven opportunistic spectrum access device. First developed by Mitola [8], cognitive radio refers to a device that has knowledge of its capabilities, internal state and the radio environment. Further, the knowledge is represented in a form that allows for automated model-based reasoning to satisfy the needs of the user. It allows expressive negotiations among peers about the use of radio spectrum across fluents of space, time, and user context [9]. A language for representing

radio-domain knowledge, called RKRL, is given in [8]. A cognitive radio is self-aware and "knows that it knows." In its extreme, the concept accommodates adaptation through learning.

A radio platform called the adaptive spectrum radio (ASR) that demonstrates the principles for dynamically accessing the spectrum is described in [10]. Based on a Software Defined Radio (SDR), the ASR adapts its frequency and modulation to exploit spectrum gaps both in frequency and time. The ASR uses an adaptive form of Orthogonal Frequency Division Multiplexing (OFDM) that exploits spectrum gaps through the use of non-contiguous carriers.

Finally, previous works have focused on issues of interpreting/understanding policy, or addressed particular mechanisms for individual users to become aware of their environment and exploit spectrum gaps. This is the first work describing a complete networking solution where nodes need to self-organize on a network and coordinate the exploitation of spectrum opportunities. Besides providing a proof-of-concept of Opportunistic Spectrum Access, our experiments allowed us to identify the main system-level issues affecting performance.

3 Preliminaries

In this work, we consider two types of spectrum users: protected and opportunistic. Protected users are similar to current spectrum licensees in that they expect some interference protection from the regulatory entity. Opportunistic users, on the other hand, are given no protection or guarantees but are granted permission to opportunistically use portions of the spectrum provided they obey some regulatory policies. These policies are designed to avoid inducing harmful interference over the protected users and may not always result on opportunistic users fulfilling their bandwidth needs or protecting them against interference from other opportunistic users.

In the next subsection we present some important concepts about interference and allowable transmit power. We then describe the class of policies and other assumptions on which this work is based. Finally we present a formal definition of the problem.

3.1 Interference Mitigation and Spectrum Underlaying

One fruitful way to think of the problem of spectrum reuse is to consider it a problem of managing interference: of ensuring that any opportunistic use does not compromise the correct transmission from a protected transmitter to a protected receiver.

There are two ways to manage interference. One is *interference prevention*, where an opportunistic user is simply prohibited from using a portion of the spectrum where a protected user (transmitter or receiver) is present. The alternative is *interference mitigation*, where an opportunistic user may cause interference, but the level of interference is restricted to levels that permit transmissions between protected users to continue without disruption. (Talk about FCC regulations for noise temperature...).

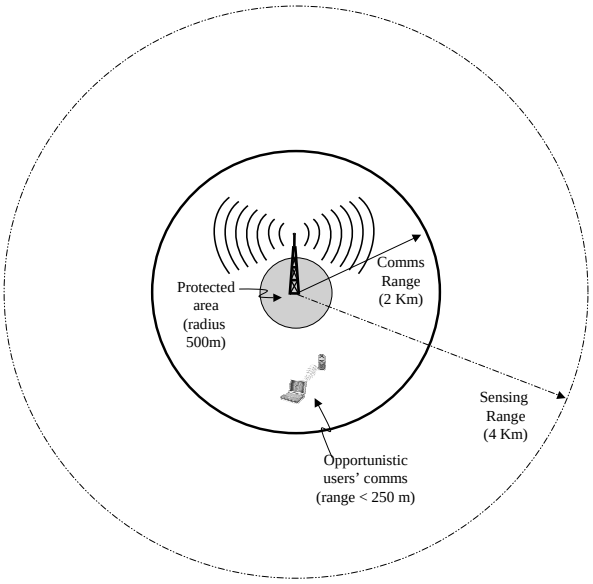

Figure 1: Opportunistic spectrum access using spectrum underlay.

Clearly interference mitigation is the more flexible approach and enables greater effective use of the spectrum. One way to implement interference mitigation is a technique known as *underlaying*, that is, opportunistic users employ low-power transmissions that do not increase the protected users' interference level above a predefined threshold.

An example of spectrum underlaying is shown in Figure 1. A protected transmitter is sending over a 100 kHz band in the 2.3 GHz range, with a transmit power of 24dBm, to some protected receiver (not shown) within a range of 2 Km. The receiver has been designed to operate with a minimum SNR of 9 dB. At normal temperature (300 K) the background noise level is about -124 dBm. Assuming an exponential power decay factor of 4 (i.e. received power at distance $d = \Theta(1/d^4)$), the protected node's signal will be received with a power of -112 dBm at a distance of 2Km.

The opportunistic users want to transmit at -12 dBm, and since they are designed to work with a SNR of 12 dB, their communication range is at most 250m. If we enforce a "protected area" of 500m around the protected receiver by requiring opportunistic users to transmit (at -12 dBm) only if they receive the protected node's signal at a level equal or below -88 dBm, then the opportunistic node's signal will reach the protected node with a signal of at most -124 dBm, which will increase its noise floor by 3 dB to -121 dBm. Since the protected node needed a SNR of at least 9dB, he'll need to receive another protected node's transmission with at least -112 dBm. Therefore, his communication range will be within the 2Km goal and there's an opportunity for underlaying.

Now, to illustrate the potential of underlaying, suppose there was a network of protected nodes within our 2Km radius. Let d be the number of neighboring protected nodes (i.e. less than 2 Km apart) using the same frequency band. Then, the total sum of protected areas inside the 2Km

radius will be (assuming non-overlapping protected areas) at most $(1+d)\pi(0.5)^2 Km^2$. In other words, even when the protected nodes' signal (in the respective band) occupy the entire area, a fraction of no less than $1 - \frac{1+d}{16}$ of the area will not be considered "protected."

But, what is a typical or sensible value of d?. The answer depends on the application. If the protected nodes are members of an ad hoc network, and this network is well designed then a node degree of 6-8 can be expected, meaning that about 50% of the space will be free. Alternatively, if the protected node network is an analog (AMPS) cellular network with a cell site coverage of 2Km., since the system assigns a small band to each cellular user we have that $d = 1$ and more than 87.5% of the space will be available for underlaying.

The discussion so far has only considered the interference experienced by the protected nodes, as ensuring communication among protected users is assumed to be the foremost policy concern. However, the opportunistic users' transmissions will experience interference from the protected users. In the example, opportunistic nodes close to the boundary of the protected area will experience interference of 36 dB above the background noise level, which is high. Thus, in order for the opportunistic users to take full advantage of the underlaying opportunities, they will have to be able to remove the interference from the the protected nodes' signal, possibly using knowledge of features (waveform type, etc.) of the protected node's transmissions and applying multi-user communication techniques such as simultaneously decoding both their signals and the signals from the protected nodes.

For most of this paper we assume that the opportunistic users will seek to exploit the spectrum as much as possible, including the capability of decoupling the protected nodes' signals from their own.[1]

3.2 Sense-Transmit Policies

So far, we have discussed two situations when the opportunistic users may access the protected users' spectrum: there is no protected user around, or underlaying is possible. Both situations are similar in the sense that they both grant spectrum access based on the amount of protected users' signal measured by the opportunistic user, and that the spectrum usage restriction – the maximum transmit power allowed to the opportunistic user (O_{xmit}^{max}) – is a function of the signal measured by the opportunistic user (O_{sensed}), and the protected node's (minimum) transmit power (P_{xmit}^{min}) and interference limit ($P_{interference}$).

Indeed, consider that protected nodes are half duplex. Let $pathloss(t)$ be the total signal attenuation from the protected node P to the opportunistic user O at time t, which is the time the

[1]The class of protected users, on the other hand, will likely be composed by the legacy systems, incumbent, current licensees, and are not expected to add new hardware capabilities but maybe add an extra margin when doing their link budget computation (even this is not necessary if the interference limit is below the background noise level, although this would reduce the achievable throughput).

opportunistic user senses the medium to detect the protected users' presence. This pathloss is the total result from different attenuation sources as propagation pathloss, antenna gains, connectors attenuation, etc. And let $pathloss(t + \Delta t)$ be the total attenuation at time $t + \Delta t$, when the node O intends to transmit. Also, let $P_{xmit}(t)$ be the transmit power used by node P at time t, $O_{xmit}(t + \Delta t)$ be the transmit power used by node O at time $t + \Delta t$, and $P_{rcvd}(t + \Delta t)$ the opportunistic user signal strength (interference) experienced by the protected user at time $t + \Delta t$. We have:

$$\begin{aligned} O_{sensed} &= P_{xmit}(t) - pathloss(t) \\ &\geq P^{min}_{xmit} - pathloss(t) \quad (1) \\ P_{rcvd}(t + \Delta t) &= O_{xmit}(t + \Delta t) - pathloss(t + \Delta t) \\ &\leq O^{max}_{xmit} - pathloss(t + \Delta t) \quad (2) \end{aligned}$$

If we assume that for small Δt, $pathloss(t) \approx pathloss(t+\Delta t)$ and setting $O^{max}_{xmit} = P_{interference} + P^{min}_{xmit} - O_{sensed}$, we obtain:

$$\begin{aligned} P_{rcvd}(t + \Delta t) &\leq P_{interference} + P^{min}_{xmit} - O_{sensed} - pathloss(t) \quad (3) \\ &\leq P_{interference} + P^{min}_{xmit} - O_{sensed} + (O_{sensed} - P^{min}_{xmit}) \quad (4) \\ &= P_{interference} \quad (5) \end{aligned}$$

Where (4) comes from applying (1) in (3). Thus, as long as $O_{xmit} \leq O^{max}_{xmit}$ the opportunistic user's transmission will not cause interference on the protected user above its tolerable level. The dependence of O^{max}_{xmit} on the sensed protected node's signal (O_{sensed}) is linear. The lower O_{sensed} the higher the allowed transmit power. However, the transmit power cannot grow unlimited, since O_{sensed} cannot decreases below node O's sensing device's sensitivity threshold $O_{threshold}$. Basically, when the protected node's signal level at node O decrease below the sensor sensitivity, then node O will fail to detect the protected node presence. Thus, for the case where no protected node is detected, we have:

$$\begin{aligned} O_{threshold} &> P_{xmit}(t) - pathloss(t) \\ &\geq P^{min}_{xmit} - pathloss(t) \quad (6) \end{aligned}$$

By comparing inequalities (1) and (6), and setting $O^{max}_{xmit} = P_{interference} + P^{min}_{xmit} - O_{threshold}$, we get that if $O_{xmit}(t + \Delta t) \leq O^{max}_{xmit}$, then $P_{rcvd}(t + \Delta t) < P_{interference}$.

Summarizing, an interference limiting policy guaranteeing that a single opportunistic user do not induce an interference level above $P_{interference}$ will limit the opportunistic node transmit power to O^{max}_{xmit}, defined as:

$$O^{max}_{xmit} = \begin{cases} P_{interference} + P^{min}_{xmit} - O_{sensed} & \text{Protected node's signal detected} \\ P_{interference} + P^{min}_{xmit} - O_{threshold} & \text{otherwise} \end{cases} \quad (7)$$

Figure 2 show an example of the variation of the opportunistic user's maximum allowed transmit power (O_{xmit}^{max}) in the protected band ($100KHz$ as in the previous example) versus the sensed level of protected node's signal for three different values of protected node's transmit power (P_{xmit}^{min}). We may see how a higher-powered protected node may result on a higher transmit power for opportunistic users. For example, when no protected node is detected in a region, the maximum transmit power that an opportunistic user can employ will be 6dBm, 18dBm, or 30 dBm, if the protected node associated with that band is known to transmit at 0dBm, 12 dBm, or 24 dBm, respectively (assuming a sensor sensitivity of -124 dBm, that is, at the background noise level for 100 kHz and 300K). Let's consider the case where opportunistic users want to transmit at 10dBm, since it is the power required to close links of up to 1Km using QPSK. As we can see, when the class of protected users transmit at 0dBm, it will not be possible for opportunistic users to transmit at full rate using this band even if no protected user is detected. Instead, they will have to add a processing gain of up to 4dB to its waveforms in order to close their links. Of course, if the opportunistic users' improve their sensor capabilities by setting its sensitivity to -128 dBm, they will increase their sensing range enough as to be able to determine when their 10dBm-transmissions does not violate the protected node's interference limit. Thus, it is not surprising that efforts are being made to increase the sensing capabilities of candidate opportunistic spectrum users (see for example the work under DARPA's XG program [7]). For example, sub-noise detection may be achieved by exploiting knowledge of features of the protected node's signal, and matching the sensor to it. Finally, it should be noted that when P_{xmit}^{min} is either 12dBm or 24dBm and the opportunistic users want to transmit at 10dBm (and the sensor sensitivity is not a limiting factor) then the opportunistic users will be able to successfully transmit at full rate if the sensed protected node's signal is no more than -116 dBm and -104 dBm respectively. However, since the corresponding pathloss between protected and opportunistic node is the same (128 dB) there will be no difference on the "protected area" induced by the two types of protected nodes. However, detecting a signal at -104 dBm requires a less sophisticated (and less expensive) sensor than detecting a signal at -116 dBm.

Figure 2 shows the allowed transmit power when the pathloss remains constant over a (small) period of time and there is only one opportunistic node transmitting at any given time. In practice, however, the allowed transmit power need to be reduced by a margin in order to account for:

- Short-timescale pathloss variations due to mobility. At frequencies around 3GHz, with a wavelength $\lambda = 0.1m$, a displacement of $\lambda/4 = 0.025m$ will be enough to cause a multi-path signal to switch from being received with cumulative interference to be received with destructive interference. If the relative speed of the opportunistic nodes is $5m/s$ this means that in only 5 msec the received signal strength (and the corresponding pathloss) can change dramatically.

- The cumulative effect of several simultaneous opportunistic nodes' transmission. Consider the

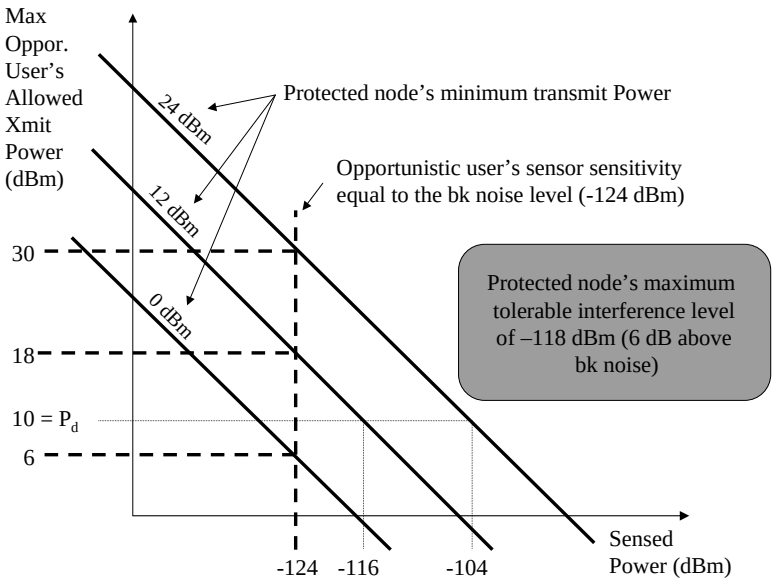

Figure 2: Opportunistic users' maximum allowed transmit power.

example discussed in the previous subsection and illustrated in Figure 1. If two opportunistic nodes are located at opposite sides of the protected area, that is, more than 1000m apart, and assuming they transmit at -12dBm and we are using the same rate of exponential power decay (4), we get that they will receive each other signal at a level below -136 dBm, that is, 12 dB below the background noise floor. Thus, they will not be able to sense each other and if they are using a CSMA based MAC they may decide to transmit simultaneously. The interference level induced at the protected node will be equal to the sum of the interference induced by each opportunistic node individually. In this setting, with the carrier sensing range of the opportunistic nodes being the same as the protected area's radius, the worst-case interference is achieved when 6 opportunistic nodes are located at the boundary of the protected area, in the vertices of a regular hexagon, and the rest of opportunistic nodes are located in the vertex of the regular triangular lattice that is the expansion of the 6 regular triangles formed when drawing lines from the protected node to the vertices of the aforementioned hexagon (i.e. the 6 closest opportunistic nodes). Thus, grouping the opportunistic nodes in order of their distance to the protected node in hexagons of side length $l, 2l, 3l, ...$, where $l = 500m$, we get that the first 6 opportunistic nodes are at a distance l. The next 12 nodes are at a distance between $2\frac{\sqrt{3}}{2}l$ and $2l$, and in general there are $6k$ nodes at a distance between $k\frac{\sqrt{3}}{2}l$ and kl. If I_1 is the interference caused by any of the 6 closest opportunity user, then I_T, the total interference induced by all the opportunistic nodes is at most $I_T \leq I_1(6 + 6\frac{2}{(\frac{2\sqrt{3}}{2})^4} + \ldots + 6\frac{k}{(\frac{k\sqrt{3}}{2})^4} + \ldots$, which after some simplification becomes $I_T \leq 6I_1(1 + \frac{16}{9}\sum_{k\geq 2}\frac{1}{k^3})$. It is easy to show that the summation converges since it is bounded by $\frac{1}{8} + \int_{x=3}^{+\infty} dx/(x-1)^3 = 0.25$, therefore $I_T \leq 8.66I_1$. Thus, in this case, adding a margin of 10dB to sense-transmit policy will be enough to prevent

the interference to increase above an acceptable level.[2]

So far, the above discussion has centered on a particular band. It can be generalized to cover any set of bands assigned to protected nodes. In such cases we can speak of a vector of maximum transmit power per band based on the signal sensed at each band. Moreover, when protected nodes split its assigned frequency in sub-bands (using for example FDMA) we may apply the same policy on a per-subband basis, and if the sub-bands are small enough (or there are some constraints about uniform power distribution over a band) we may translate the sense-transmit policies into a power spectral density function bounding (as an envelope) the opportunistic node's spectral transmission profile.

Summarizing, the sense-transmit class of policies results on a valid power spectral density function that is inverse linear with respect to protected nodes' signal power sensed at the different frequencies. Since sensing errors (due to noise, signal accumulation, etc.) will typically result on an overestimation of the signal measured, the resulting effect will be a reduction on the allowable power spectral density function and consequently a lower interference level at the protected nodes (conservative).

3.3 Working Assumptions

In this work, we study the performance obtained by a network of opportunistic users deployed in a region where some protected users are present. Given the multiple different ways that opportunities can be exploited, as indicated by different governing policies reflecting different protected users behaviors, it is necessary to specify our working assumptions. These are:

A.1 Protected nodes are given privileged use over some (wide) frequency bands. This assignment is fixed.

A.2 Opportunistic nodes access to the bandwidth outside their coordination channel (i.e. likely assigned to protected nodes) is governed by sense-transmit policies (see previous section). Sense-transmit policies restrict the maximum power used in a frequency band (and not any other communication parameter, as for example waveform type, transmission time, etc.) based only on the level of protected nodes' signal sensed at the opportunistic node. Protected nodes may use any subset of their assigned frequencies at any point in time.

[2]Actual pathloss values do not perfectly match the $1/d^4$ curve, which only represent and expected – or averaged – value. For example, walls or other obstacles may increase the pathloss between opportunistic nodes close to the protected area boundary and result on nodes less than 500m apart not being able to sense each other. In this case, the number of opportunistic nodes simultaneously transmitting close to the protected area may be more than 6. Actually, there is no limit to it, since we can always built a topology where nodes in the protected node boundary are isolated between them (by metal walls, etc.) but free to interfere with the protected node. However, this situation is unlikely in practice and the 10dB value presented here is a good estimate for an effective margin.

These two assumptions seek to establish a simple regulatory regime. There are protected users and basic rules about how opportunistic users may reuse spectrum. The next few assumptions seek to refine this regime into an environment where we can clearly define operating rules:

A.3 The opportunistic nodes are assigned (i.e. are privileged users of) a small frequency band. They use this band as their coordination (or bootstrapping) channel.

A.4 The minimum transmit power of the protected nodes is known.

A.5 Protected nodes support a known level of interference.

A.6 Opportunistic nodes have a sensing mechanism that distinguishes between their transmissions and those of protected nodes. This sensing mechanism reports the protected nodes' signal level in the different frequency bands.

These assumptions define some key features. First, there is a way (A.3) for opportunistic nodes to safely communicate with each other.[3] Second, the necessary information is available (A.4 and A.5) to compute the maximum transmit power available to the opportunistic nodes according to Equation 7 in Subsection 3.2. Finally, the ability to sense and distinguish (A.6) is critical for opportunistic nodes to collectively exploit unused spectrum. Otherwise, as soon as an opportunistic node starts transmitting over a frequency band, the rest of the opportunistic users will detect its signal, mistake it by a protected nodes transmission, regard the frequency band as occupied and stop using it. Distinguishing between protected and opportunistic nodes can be done by exploiting knowledge of the corresponding waveforms.

The last four assumptions are refinements, intended to define the operating environment crisply enough that we can define and simulate protocols, and yet not lose much generality (and none of the complexity).

A.7 Protected nodes' use of a frequency is half duplex, i.e. a node alternates between transmit and receive. Protected nodes, while using a frequency band are not silent for more than a predetermined amount of time.

A.8 Opportunistic nodes have two transceivers. One is constantly listening to their assigned OSA-coordination channel (see a.2). The other transceiver is frequency agile and may dynamically adjust its transmission frequency and bandwidth, as well as its processing gain and signal waveform.

A.9 The transceiver used for application data communication is able to simultaneously distinguishing between protected nodes' transmissions and one opportunistic node's transmission (the

[3]It should be noted that a dedicated frequency band is not absolutely required. For example, the distinction between coordination and application data channels can be made in the code domain (or even the time domain) instead that in the frequency domain.

one the receiver is currently "tuned to"). All other transmission from opportunistic nodes are effectively noise.

A.10 Protected and opportunistic nodes are equipped with exactly one antenna. This antenna is omnidirectional.

We need some way to sense the presence of protected nodes. Assumption A.7 puts us in one of the more challenging environments, where presence is dynamically sensed without any explicit coordination. (For instance, access to bands could be regulated by a 'take it' and 'leave it' beacon signals.[4]).

Assumptions A.8 and A.9 define the capabilities of the opportunistic node. The assumption that there are two transceivers simply makes the world easy to understand, without losing generality. (As noted earlier, the coordination channel may simply represent different coding and could be decoded in parallel through a single transceiver). The second assumption (A.9) to treat all but the desired opportunistic transmission as noise is again, an attempt to simplify without losing complexity. Clearly the transceiver could have enough processing power to support multi-user communication among opportunistic nodes. Making the transceiver capable of receiving from only one node at a time makes scheduling transmissions (to achieve good channel use) more challenging and also increases the noise level the signal must overcome.

Finally, assumption A.10 makes the simulation of the various nodes easier, while maximizing the interference among nodes. It should be noted that the use of directional antennas and antenna arrays with null-steering capabilities will increase the effectiveness of underlaying methods. However, designing and simulating algorithms that are cognizant of directional beams and developing networks that support directional neighbor discovery and directional media access and other features (see, for instance, [11]) is hard and not needed to demonstrates the benefits of spectrum re-use.

3.4 Opportunistic Spectrum Access (OSA) Problem Formulation

The Opportunistic Spectrum Access (OSA) problem is one of power allocation for different communication pairs and frequency intervals subject to power-limiting (policy) constraints. In addition, there is a practical limit on the maximum available power at a node. Power is allocated to communication pairs/frequencies in order to achieve the transmission rates fulfilling some optimization objectives. For a given power allocation schedule, the transmission rates achievable are determined by Shannon's capacity formula for the single user receiver, where the received power corresponding to simultaneous (co-channel) transmissions are regarded as noise.

[4] A 'take it' beacon is a tone authorizing the use of a band, say a public safety band, when the band is not in use by the public authority. The opportunistic users will have to vacate the band as soon as the beacon stops. In addition, a 'leave it' beacon may provide local overrides as for example when an ambulance is going through an area covered by a 'take it' beacon.

Let $V = \{1, 2, \ldots, N\}$ be a set of nodes, $F = \{f_1, f_2, \ldots, f_M\}$ be a set of frequency intervals, $\mathcal{G} = \{\Gamma^1_{NxN}, \Gamma^2_{NxN}, \ldots, \Gamma^M_{NxN}\}$ be a set of pathloss matrices such that $\gamma^k_{i,j}$ represents the attenuation experienced by signals originated at node i arriving at node j when using the frequency interval f_k. Let P^{Policy}_{NxM} be a matrix which $i-th$ row represents the vector of maximum transmit power per band discussed at Subsection 3.2, that is, $P_{i,k}$ represents the maximum power node i can transmit over frequency interval f_k. And let P^{max}_{Nx1} be a column-vector representing the maximum available power at each node, that is, $P^{max}_{i,1}$ is the maximum power available at node i (independent of policy or frequencies being used).

Definition: *Valid Power Assignment Set*

Let $\mathcal{P} = \{P^1, P^2, \ldots, P^M\}$ where P^k is a NxN matrix where $P^k_{i,j}$ represents the (useful) power assigned to the communication pair $i \rightarrow j$ (from node i to node j) over the frequency interval f_k (this implies that node j's role is that of a receiver and it is tuned (only) to node i. We say that $\mathcal{P}$ is a *valid power assignment set* (of matrices) iff:

$$
\begin{array}{lllll}
(VPA.1) & P^k_{i,j} \geq 0, & & & \forall\ i,j,k \\
(VPA.2) & P^k_{i,i} = 0, & & & \forall\ i,k \\
(VPA.3) & \text{If } P^k_{i,j} > 0 \Longrightarrow & P^s_{r,i} = 0, & \text{and} & \\
 & & P^s_{j,r} = 0, & & \forall\ i,j,k,r,s \\
(VPA.4) & \text{If } P^k_{i,j} > 0 \Longrightarrow & P^s_{r,j} = 0, & & \forall\ i,j,k,r \neq i,s
\end{array}
$$

(VPA.1) follows since the transmit power cannot be negative. (VPA.2) follows since it does not make sense to have a node transmitting to itself. (VPA.3) Comes from the fact that at any given time (schedule) a node may assume only one of three roles: either it is a transmitter, it is a receiver tune to one (and only one) transmitter, or it is in idle. Thus, if a node i is transmitting at a given frequency, he cannot receive at any frequency (therefore $P^s_{r,i} = 0$). Similarly, if node j is a receiver tuned to node i, it won't transmit at any frequency ($P^s_{r,i} = 0$). Additionally, (VPA.4) follows since if node j is tuned to node i's waveform, it will not be able to decode any other signal from any other transmitter and therefore it would be of no use to assign a non-zero power to any $r \rightarrow j$ communication pair (i.e. $P^s_{r,j} = 0$). The above does not mean that node j will not experience interference from other ongoing communications, but that this interference will result from transmissions to other receivers. In this sense, the *valid power assignment set* represents a conflict-free scheduling: there are not two different transmissions with the same destination. On the other hand, we are not preventing a node from simultaneously transmitting to two different destinations. While a (single-user) receiver needs to be tuned to one and only one waveform, a transmitter can generate (mix) several waveforms simultaneously (e.g. by using OFDM, or FDM modulation). However, due to the characteristic of the shannon capacity formula for single-user receivers(see $VTS.3$ below), in any optimal solution a transmitter will not transmit to 2 or more receivers over the same frequency interval at the same time.

Definition: *Valid Transmission Schedule*

Let R_{NxN} be a matrix of nonnegative elements (i.e. $R_{i,j} \geq 0$). R_{NxN} is said to be a valid transmission schedule under $\{V, F, \mathcal{G}, P^{Policy}, P^{max}\}$ if and only if there exist a *valid power assignment set* $\mathcal{P} = \{P^1, P^2, \ldots, P^M\}$ such that:

$$\begin{array}{lrcll} (VTS.1) & P^k \ \mathbf{1_N} & \leq & P^{Policy} \ \mathbf{e_k} & \forall\ k \\ (VTS.2) & \sum_k P^k \ \mathbf{1_N} & \leq & P^{max} & \\ (VTS.3) & R_{i,j} & \leq & \sum_k |f_k| \ log_2(1 + \frac{P^k_{i,j}\gamma^k_{i,j}}{N_o|f_k| + \sum_{(u,v)\neq(i,j)} P^k_{u,v}\gamma_{u,j}}) & \forall\ i,j \end{array}$$

Where $\mathbf{1_N}$ is the all-1 column vector of size N, $\mathbf{e_k}$ is the $Nx1$ column vector of all-zero elements except by the $k-th$ row element which is equal to 1, $|f_k|$ represent the bandwidth (in hertz) of the frequency interval f_k, and the expression $A \leq B$ in (VTS.1) and (VTS.2) denotes element-wise inequality between matrices A and B of the same dimensions (i.e. $A_{i,j} \leq B_{i,j} \ \ \forall\ i,j$).

(VTS.1) states that a node's transmit power (for all possible destinations) over a given band f_k cannot exceed the maximum allowed by policy (which in turn depends on the protected energy sensed locally, as discussed in Subsection 3.2. (VTS.2) states that a node's total transmit power (over all possible frequency bands) cannot exceeds the node's total available power (hardware-related, policy-independent). Finally, (VTS.3) represents Shannon's capacity limits for the given power assignment set. Any transmission rate smaller than the right-hand value of (VTS.3) is achievable, provided that sufficient long codewords are allowed, or, equivalently that the channel conditions are assume to be stable over a sufficient long time period.

Based on the above definition of a Valid Transmission Schedule, several useful Opportunistic Spectrum Access (OSA) problems can be defined. Two of them are the Maximum Throughput (MT) and the Fair Time-Frequency Traffic Scheduling problems.

Definition: *Maximum Throughput problem*

Let S be the set of *Valid Transmission Schedules* under $\{V, F, \mathcal{G}, P^{Policy}, P^{max}\}$, the Maximum Throughput (MT) problem can be stated as follows:

$$\max \quad \mathbf{1_N}^t R \ \mathbf{1_N} \quad s.t.\ R \in S$$

(where A^t represents the transpose of matrix A). The MT is the simplest of the OSA problems. It returns the maximum possible throughput achievable by a network employing OSA. Since the schedule needed to achieve this maximum value will typically imply that only those communiation pair with the smallest pathloss are able to transmit – unfairly depriving spectrum access to the rest of the network users – this maximum throughput will sendomly be achieved/used in practice. However, knowing this value will give us an idea of the limits achievable by an OSA systems and will guide us in our design decisions. This value, then, is the benchmark we use to compare our systems performance in Subsection 8.1.

It is useful to distinguish two subclasses on the MT problem, depending of whether $P^{Policy} \ \mathbf{1_N} \leq P^{max}$ or not (once again, '$\leq$' here denotes element-wise comparison). In case the above inequality

holds, we say we are in the policy-limited regime. If it does not, we say we are in the power-limited regime, in which we could increase the achievable thorughput by providing the nodes with more powerful transmitters. When we are in the policy-limited regime, the MT problem present two useful properties:

- Independence between different frequency intervals. So, each frequency band can be optimized independently, provided we choose the same transmitter/receiver pairs.

- The right hand of (VTS.3) does not present any local maximum for valid power assignments (i.e. greater than zero) but only local minumums, therefore the optimal solution for a node i transmitting over a frequency band f_k will be that either it does not transmit on that band or it transmit to only one node, say j, with the maximum possible power (i.e. $P^k_{i,j} = P^{Policy}_{i,k}$ and $P^k_{i,j'} = 0 \ \forall \ \ j' \neq j$).

This above properties greatly reduces the MT problem complexity, although it is still exponential on the number of nodes. Therefore, heuristics and approximate solutions will be needed when the network size increases above the tens of nodes.

Finally, it should be noted that alternatively we could defined the Maximum Throughput-Distance (MTD) problem by replacing the optimization goal by $1'_N(R \bullet D)1_N$, where '$\bullet$' denotes element-wise multiplication and D is a NxN matrix of nonnegative elements where $D_{i,j}$ represents the distance separating nodes i and j. Similarly as the MT problem, the MTD problem will result on only the shorter, faster communication pairs being given access to the spectrum.

Definition: *Fair Time-Frequency Traffic Scheduling problem*

Let S be the set of *Valid Transmission Schedules* under $\{V, F, \mathcal{G}, P^{Policy}, P^{max}\}$ and let $S^* = \{R : R = \lambda_1 R_1 + \lambda_2 R_2 + \ldots + \lambda_n R_n, \ R_k \in S, \ \lambda_k \geq 0, \ \sum_k \lambda_k \leq 1\}$ be the smallest convex set containing all the elements in S and the 0_{NxN} (all zeros) matrix. Let T be NxN matrix of non-negative elements where $T_{i,j}$ is the traffic required by node i to transmit to node j, the Fair Time-Frequency Traffic Scheduling (FaTiFreS) problem can be stated as follows:

$$\begin{aligned} \max \quad & \mathbf{1_N}^t R^* \ \mathbf{1_N} \\ \textit{subject to} \quad & \\ & R^* \in S^*, \ \textit{i.e.} \ \exists \ \{\lambda_k\} \ \textit{s.t.} \ R^* = \textstyle\sum_k \lambda_k R_k \\ & \min_{T_{i,j}>0} \{\tfrac{R^T_{i,j}}{T_{i,j}}\} \geq \mathcal{K} \\ & \mathcal{K} = \max_{R \in S^*} \min_{T_{i,j}>0} \{\tfrac{R_{i,j}}{T_{i,j}}\} \end{aligned}$$

The $\{\lambda_k\}$ set will define the time schedule, and for each time slot associated with λ_k, R_k (and its associated P^k) will determine the frequency schedule. The FaTiFreS solution will maximize the achieved throughput provided that no communication pair is given a service level below the minimum level $\mathcal{K}$ that can be guaranteed for all.

Note that more stringent fairness condition may rquire that no communication pair is given preference over another, that is, that is not possible to increase the level of service given to a communication pair without affecting another communication pair with a *lower* level of service. In other words,

$$\begin{aligned} \max \quad & \mathbf{1_N}^t R^* \ \mathbf{1_N} \\ \textit{subject to} \quad & \\ & \text{If } \exists \ \hat{R} \in S^* \ s.t. \ \frac{\hat{R}_{i,j}}{T_{i,j}} > \frac{R^*_{i,j}}{T_{i,j}} \Longrightarrow \exists \ T_{u,v} > 0 \ s.t. \ \frac{\hat{R}_{u,v}}{T_{u,v}} < \frac{R^*_{u,v}}{T_{u,v}} < \frac{R^*_{i,j}}{T_{i,j}} \end{aligned}$$

Finally, it should be noted that the definition of a Valid Transmission Schedule (VTS) used in defining the above problems differs from typical literature for ad hoc networks, that define a network topology as a duple $\{V, E\}$, in that we have made no reference to links, neighbors, or a topology defined by $\{V, E\}$, where E is the set of links connecting some nodes in V. However, in a OSA system, the bandwidth, waveform, and rate that a node may use to reach another is not fixed, but it may be determined on-the-fly. Therefore, in an OSA system any communication pair is possible, assuming the proper communication rate is used, as defined by (VTS.3). Therefore, in an OSA system the determination of a set of links and neighbors is part of the problem solution searched. Only if a particular waveform is imposed, say $QPSK$ with a processing gain of 0 dB, we would be able to discard some communication pairs that are not feasible due to a high pathloss between the nodes. We could only then talk about 'links' between some nodes. However, it would be wrong to assume that interference can only be induced by nodes connected by these 'links', since the aggregate effect of several nodes that are far enough as to not be able to communicate directly (at a given bandwidth/rate) may result in harmful interference. Thus, modelling an Ad Hoc network as a $\{V, E\}$ graph will result on a overestimate on the system performance. On the other hand, given the high computational complexity involved in solving the OSA problem, a reasonable heuristic may be to temporarily model the network as a $\{V, E\}$ by imposing minimum limits on the desired transmission rate between two nodes then to use known heuristics for ad hoc networks (e.g. the use of Conflict Free Sets (CFS)) to determine, say, a good TDMA transmission schedule. And finally, we may go back to the original VTS problem are recompute the actual achievable rates for that schedule by using (VTS.3). We will not obtain the optimal assignment but at least (hopefully) a good one.

Now, we focus to describe the particulars of our solution.

4 XOSA system overview

XOSA is an interacting set of modular mechanisms for opportunistic spectrum access. It consists of several new mechanisms working cohesively to implement the first complete opportunistic spectrum access (OSA) system. In designing XOSA, our objectives were:

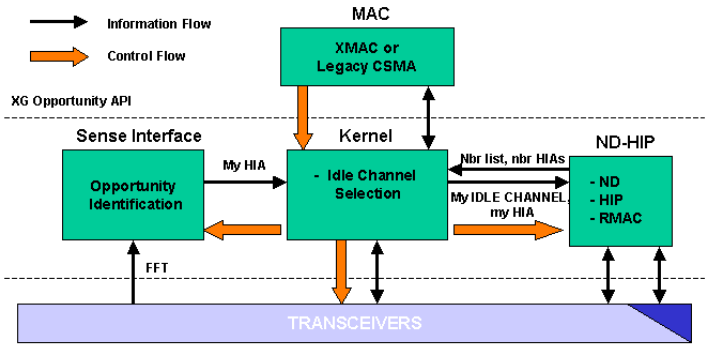

Figure 3: XOSA system architecture.

- To develop a complete system solution, capable of exploiting OSA even if the upper layers are not OSA-aware (e.g. using legacy MACs).

- Simplicity. The goal was a straightforward realization of OSA that forced us to develop all the essential mechanisms and none of the optimizations. A simple system is a good benchmark against which to test optimizations, and provides a clear starting point.

- Modularity. We have sought to divide up the functionality such that adding new features or optimizations is easy, without destroying the general architecture.

Figure 3 shows the XOSA system architecture. From XOSA's perspective, the node is divided into three layers: physical, XOSA Management, and MAC-and-above layers.

The physical layer consists of two transceivers (following the assumptions in section 3.3). The primary transceiver is frequency agile and can be tuned to different frequencies and waveforms by the XOSA management layer. The second transceiver is tuned to the XOSA control channel. Each transceiver is capable of successfully decoding (receiving) one packet at a time (again, following the assumptions in section 3.3). In addition, the physical layer includes a sensing device that periodically reports to the XOSA management the protected nodes' energy over the frequency bands of interest.

Above the physical layer sits the XOSA management layer. Among its functionalities are:

- Estimating the available frequencies in the node's neighborhood (opportunity identification).

- Discovering other opportunistic nodes in the neighborhood and selecting those of them that are appropriate neighbors.

- Informing other opportunistic nodes of this node's frequency observations.
- Selecting a default set of frequencies to become the node's IDLE channel. The IDLE channel is the frequency band(s) and waveform(s) that the node is tuned to when it is not transmitting, receiving, or preparing for reception of a pre-schedule packet.
- Sharing this opportunistic node's IDLE channel identity with its neighbors so that neighbors know how to reach it.
- Serving as a front-end for the MAC layer. Beyond simply transmitting and receiving, the interface between XOSA and the MAC layer allows the MAC layer to take an active role in channel selection and management.

The XOSA management layer divides these functions across three modules: XOSA kernel, Sensing Interface, and Neighbor Discovery and Hole Information Protocol (ND-HIP). The XOSA kernel module centralizes this layer's decisions and presents a well-defined interface to the MAC layer. It also handles most physical layer functions. The Sensing Interface module is in charge of determining the presence of transmission opportunities. And the ND-HIP module is in charge of disseminating OSA information among opportunistic nodes.

In the experimental results in section 8 we assumed the MAC layer was unaware of the XOSA layer below it (e.g. a legacy MAC) and XOSA was entirely responsible for finding spectrum to meet the MAC layer's needs. However, we expect that in many cases the MAC will be XOSA-aware and that the MAC will indicate to the XOSA kernel the capacity the MAC needs.

Figure 3 also shows the information flow in XOSA. The physical layer reports measured energy in various frequency bands. The Sense Interface module receives this information and converts it into a Hole Information Array (HIA). The HIA is a vector where each entry corresponds to a frequency band and indicates if the band can be opportunistically used.[5] The Sense Interface reports the HIA to the XOSA kernel.

Concurrently, the NDHIP module exchanges HIA information with other opportunistic nodes and provides the HIA information from the node's neighbors (and perhaps peers farther away). Combining its local HIA infomation with that of its neighbors, the XOSA kernel develops a map of how the spectrum is being used in its area, and selects the node's IDLE channel(s). XOSA kernel provides the NDHIP module with both its HIA and IDLE channel information for the NDHIP to propagate it to other opportunistic nodes.

The next few sections describe these processes in greater detail.

[5]One should view the HIA as a joint product of the physical sensors and the Sense Interface, as the physical layer may have to be programmed by the Sense Interface to detect transmissions

5 Sensing Interface

The Sensing Interface is in charge of Opportunity Identification based on the physical layer measurements.

We assume that the physical layer capable of distinguishing protected transmissions from all other signals.[6] and it only reports energy received from the protected nodes to the Sensing Interface.

The XOSA system divides the frequency of interested into 'frequency slots' of length Δf. The frequency slot is the minimum frequency unit of our system. Choosing the right frequency slot size is a challenge. No opportunity smaller than a frequency slot can be used, so we generally want a small slot size. Unfortunately, the slot size is inversely proportional to the sensing interval – the smaller the slot size, the longer we must sense to characterize it. Expressing this point another way: if the goal is a swift response to changes in usage by protected users, then the sensing interval must be short, and the slot size relatively large. Pragmatically, considering sensing intervals of the order of miliseconds, frequency slots of length 1 kHz or more make sense.

The Sensing Interface receives from the physical layer sensor an array of receive power values for the discrete frequency slot intervals.[7] The Sensing Interface can then determine the maximum allowed transmit power by applying Equation 7 to the sensed energy. If there is more than one type of protected node in the band associated with the frequency slot, the maximum allowed transmit power is the minimum over all protected node's types.

This process results in an array of varying transmit powers by slot. Unfortunately, an array of variable power values is probably the wrong data structure (at least to start with). It is big (and thus requires large amounts of bandwidth between opportunistic nodes). It changes frequently (as the frequency values are likely to vary from interval to interval). And it requires a highly programmable transceiver to tune to every nuance and change in available frequency.

Thus, following the goal of keeping XOSA simple, we chose to use use the same maximum transmit power P_d over all the frequencies slots. A frequency slot is considered an opportunity if an opportunistic node can transmit with power up to P_d without interfering with protected transmissions. This simple approximation allows us to reduce the slot array to a bitmap, with bit values of 1 indicating that opportunistic use is permitted in the relevant slot.

There is obviously a tradeoff in the choice of the value of P_d. A small value will limit the effective transmission rate, since a processing gain will need to be added to close the links. A larger value increases the geographic area that must be protected from opportunistic interference and

[6] In fact, distinguishing traffic may well require the Sensing Interface to program the physical layer sensor (or its DSP). Also, it may turn out to be simpler to recognize opportunistic traffic (whose form is known by the node), and assume that all non-opportunistic traffic is protected.

[7] One way to build this array is by applying the Fast Fourier Transform (FFT) to the time samples measured by the sensor.

reduce the number of available slots. In the long term, it seems likely that a node (or set of nodes) will determine a P_d by examining the actual spectrum measurements and selecting the value that seems likely to yield the greatest bandwidth. However, again for simplicity, in our experiments we set the value of P_d *a priori* to match the NDHIP module's neighbor discovery range. That is, P_d is set as to guarantee that links to any neighbor can be closed with a bandwidth efficiency of 1 bps/hertz (i.e. no processing gain is necessary), if possible. If not, P_d is set to the power allowed by the sensor's sensitiviy.

Finally, to account for short time-scale path loss oscillations as well as protected nodes silence times, we keep memory of the past measurements up to a *sense_window*. The actual energy measurement we use to compare against O_{cut_off} is the maximum value observed over the previous '*sense_window*'. A more elaborate solution would be the use a Kalman-based Linear Predictive Filter as the one used in [11] to determine the protected nodes' signal energy level.

6 Neighbor Discovery and Hole Information Protocol

The Neighbor Discovery and Hole Information Protocol (ND-HIP) module's main function is to exchange OSA information among opportunistic nodes. Among the information exchanged is reachability (at a given rate), opportunity availability (i.e the nodes' HIA vector), and pathloss to nearby nodes. The level of propagation dissemination (both fidelity and scope) is dependent on the XOSA-management and MAC layers requirements. Thus, the design is general allowing for multiple instantiations to accommodate diverse MAC implementations. Also, ND-HIP uses a reserved OSA-coordination channel to bootstrap communication to neighboring nodes (see assumptions A.2 and A.8). This reserved coordination channel is likely a scarce resource (we assume a 200 kHz channel in our experiments) and should be used wisely. To this end the Rendezvous MAC (R-MAC) was developed. R-MAC efficiently uses the OSA-coordination channel. In the next subsections we describe the HIP protocol in more detail, the neighbor discovery criteria, the HIP dissemination technique, and R-MAC.

6.1 Neighbor Discovery and Hole Information Protocol - Basic Operation

ND-HIP is a simple protocol and similar in many respects to a link state routing protocol. This section describes the basic workings of ND-HIP.

Essentially, ND-HIP works by having each node periodically broadcast a Hole Information Protocol (HIP) packet to its neighbors, who in turn relay the HIP packet to their neighbors. Packets have a time-to-live (TTL) field which is decremented at each hop and is used to control how far each HIP packet is disseminated. (See section 6.2 for more discussion). HIP packets also contain a sequence number that nodes use to suppress duplicate and old HIP packets.

While it is described as a Hole Information Protocol packet, the packet actually may not

actually contain hole information, and it serves other functions (mostly related to maintaining state with neighbors). We describe each set of functions in turn.

First, ND-HIP is used to keep track of the state of the local XOSA environment. The packet contains a transmit power field (in dBm) which each hop's transmitter sets to its power level. Each receiver notes the received power and compares it with the transmit power to compute path loss. [8] (To distinguish between sender and per hop [re]transmitter, the HIP packet has both a source and a transmitter field. The source is where the packet originated, while transmitter is the most recent hop).

ND-HIP is also used to distribute information about the source. The packet contains information about the source's IDLE channel (see section 7.1), and the source's radio capabilities such as maximum available power, its sensor maximum sensitivity, and its noise level. For simplicity, much of the discussion in the paper so far has assumed that all equipment has the same capabilities, but in practice they will differ from device to device. A source may also indicate that it wishes to reserve certain channels for its use and thereby warn off other nodes from scheduling transmissions over those bands.[9].

There's also a list of neighbors, where the source lists each radio it has discovered, the pathloss from the radio to the source, and a flag indicating whether the radio is considered a neighbor. (As noted in section 6.2, not all nearby radios are considered neighbors as they may not even be part of the same network or organization – but their presence may need to be recognized to determine the necessary power or processing gain in the area).

And there's a list of rendezvous times. This feature is designed to reduce collisions on the control channel. Each source indicates when it expects to next transmit a HIP packet, and also when it next expects to hear a HIP packet from each of its neighbors. This information is used by the neighbors to try to coordinate transmissions and avoid collisions. The rendezvous times are only transmitted one hop (to the source's immediate neighbors) and are not retransmitted.

Finally, the HIP packet may contain the hole information array (HIA), which is a bitmap of sensed spectrum information, indicating which channels have been determined to be in use, and which ones are free (following the approach described in section refsense). Because the HIA may not change very much in certain environments, its presence is not required in every HIP packet. The HIA may be compressed to reduce its size and there are a variety of options for how the compression may be done. A compression method field is used to specify the compression in use.

[8]For the opportunity availability information to be useful when scheduling transmissions in a contention-less fashion – e.g. TDMA – information about the interference matrix between links is necessary. For a system with omnidirectional antennas and per-packet power control, the interference information can be derived from the matrix of pathlosses between nodes. This pathloss matrix can also be used for topology control.

[9]The ability in an opportunistic environment to reserve channels may sound strange, but there are situations where it is useful. For instance, different organizations – running different MACs – can use this feature to coordinate their sharing/splitting of the available spectrum, and in a world where there may be one sender but many receivers, it allows the sender to advertise on behalf of the passive receivers

6.2 Neighbor Discovery and Topology Control

In traditional networks, the term *neighbor* refers to nodes that can be reached directly with a certain (high) success probability. Reachability is determined by the maximum available power, modulation type and transmission rate. In OSA networks, however, things are not so simple. OSA nodes can use different modulation schemes, transmission rates, and frequency bandwidths. A node that is not reachable at a given rate/band can be reached at an smaller one. In theory, any node is directly reachable provided a transmission rate lower than the link capacity (Shannon limit) is used. Thus, we may ask *what is the meaning of declaring a node a 'neighbor'?*.

Following a 'layered' approach to communication protocols, only those nodes declared as neighbors will be visible to the upper layers. A node will only attempt to directly transmit data packets to those nodes declared *neighbors*. Thus, declaring a node as a neighbor indicates a willingness for direct (as oppose to multihop) communication. Removing a node from the neighbor set will prevent direct communication to that node, which may result on network partitions if there is not a multihop path towards that node.

Furthermore, when a node adds another node to its neighbor set it means that a communication channel (i.e. a link) needs to be setup between the two nodes. For example, each time an upper layer protocols requires a control packet to be 'broadcasted', the XOSA layer must take care that a copy of the packet reaches each node declared as a *neighbor*. This may imply that multiple retransmissions over different bands are necessary. Also, several MAC protocols rely on idling nodes (nodes not transmitting or receiving) being able to listen to transmissions from any *neighbor*. For example, a CSMA MAC like 802.11 expects the nodes to be listening for *neighbors* RTSs while in IDLE. Thus, XOSA needs to create a communication channel (namely, the IDLE channel) for the nodes to tune-in while in IDLE, such that any of their declared neighbors can contact them. Thus, each time a node is declared a neighbor there is extra overhead introduced: extra transmissions when broadcasting over the virtual *common* channel, and additional constraints limiting the bandwidth/rate of the IDLE channel.

It is clear than a topology control mechanism is necessary. As it will be discussed on Subsection 8.2, the proper choice of a topology control mechanism has a great impact in the achieved performance. XOSA's approach to topology control consists of two stages. The first stage ('discovery') consists on the ND-HIP module detecting opportunistic nodes in a nodes' vicinity and reporting their identity (as well as pathloss and other parameters contained in the HIP packet) to the XOSA-kernel module. The second stage ('selection') consists of the XOSA kernel deciding which nodes among those reported by ND-HIP to declare as neighbors (i.e. setup communication channels to them, include them on the list of 'broadcast destinations', instruct ND-HIP to include them in the list of neighbor on HIP packets, and report their identity to the upper layers if requested).

For the discovery stage, each XOSA node periodically (every T_{per} seconds) broadcast HIP

packets with TTL of 1 (acting as beacons) over the OSA-coordination channel. The HIP packets are sent at *ndhip_xmit_power*, a parameter specified by the XOSA-kernel module. It is typically much smaller than the node's available power to account for the fact that reserved OSA-coordination channel and its associated transmission rate are small compared with the spectrum that can be opportunistically accessed. If the same transmit power would be used for both types of transmissions (ND-HIP's and opportunistic) the communication range over the OSA-coordination channel would be - unnecessarily – much larger. At the receiving end, the ND-HIP module records the number and quality (SNR) of HIP packets received over a time window. The ND-HIP module reports to the XOSA-kernel module only those nodes whose transmission quality is consider good. For simplicity ND-HIP uses the 'k-out-of-n' criteria, where nodes are reported to the XOSA-kernel module if over the last 'n' time periods ND-HIP received 'k' or more HIP packets with a SNR above a given threshold. To prevent flapping ND-HIP uses hysteresis on the value of 'k', so that the threshold 'k' used to add a new node is greater than the threshold required to remove an existing one. It should be noted that more sophisticated techniques such as the use of a predictive linear filter as the one in [11] are possible. Once again, XOSA implements the simplest solution achieving acceptable performance but provides a modular framework (and data structures) that allows to add innovations as the one just mentioned.

The design of the selecting stage (at the XOSA kernel) has been considered an optimization above the simple 'core' system described on this paper. Currently XOSA control the topology by limiting the value of *ndhip_xmit_power* as to achieve a target range (pathloss or meters, for example 250 m in our experiments). Then, XOSA selects all nodes reported by the ND-HIP module. Note that the value of P_d in the previous section is also set as to accomplish the same range at full rate (1 bps/Hz). Further research is needed on good topology control mechanisms for XOSA. An starting point could be to order the set of nodes discovered by ND-HIP in function of their negative impact on the IDLE channel capacity and try removing (one-by-one) those nodes that are reachable using a multihop path. *ndhip_xmit_power* could be adjusted so that the set of links to 'discovered' nodes forms a biconnected set, to provide robustness against mobility, as suggested in [12].

6.3 HIP Dissemination

Different MACs may require different scope of dissemination of HIA information. A OSA-unaware MAC requires no HIA information at all.[10] An OSA-aware contention-based MAC may require one-hop HIA information to determine the frequency interval(s) to use to communicate with each neighbor. A distributed TDMA MAC scheduling frequency slots so that collision free transmissions are possible requires knowledge of the HIA and topology information of nodes in its 2 hops neighborhood, and a decentralized TDMA MAC requires full topology and HIA information.

[10]But the XOSA kernel module, in charge of building the IDLE channel that the node will tune in, requires one-hop HIA information.

Also, different latency on the HIA information may be tolerated depending on the distance to the node experiencing the change. For example, let's consider we are using a distributed TDMA MAC using 2-hop topology and HIA information. Furthermore, let's consider the network carries time sensitive information so that it cannot accept service disruption of more than a few milliseconds. The nodes require to quickly react to link failures or to changes on their neighbors IDLE channels. Finally, consider that the MAC performance is improved if all nodes in the network choose the same IDLE channel, reducing the number of duplicate transmissions needed when broadcasting (MAC) control packet. For this system, a solution will be to allow frequent (even event-driven) HIP packets sent locally (i.e. TTL equal to 1) so that IDLE channel re-computation upon a protected node's activation can be done as soon as possible. Information about 2-hop neighborhood may be transmitted less frequently since improper information may cause some collisions but still let the neighbors be reachable. And finally, global information can be sent sparsely to allow longer term optimizations where nodes choose a common IDLE channel among the set of globally available frequencies.[11]

And yet, besides differences in scope and latencies, there may also be different needs in granularity or precision of the information. Some algorithms may require the full HIA to make their decisions while others may trade off processing power and bandwidth consumption with HIA precision. Indeed, the time required to run an optimal scheduling algorithm time increases with the HIA size and may become so large that it forces the system to use heuristic algorithm over the complete data. Alternatively, it may pay off to run optimal algorithms over a smaller (compressed) data set. For example, when trying to built a global IDLE channel an approach may be to propagate lower resolution versions of the HIA array to allow nodes to quickly pinpoint to frequency bands of globally unused spectrum (sweet spots). Nodes will choose IDLE channels contained in this sweet spots. After that, a distributed algorithm will correct the IDLE channel values based on local opportunity information. Since the global information provided a good starting point, te resulting algorithm will present fast convergence.

The ND-HIP dissemination mechanism allows for multiple dissemination levels with different (configurable) scope, latency and fidelity, as shown in the 3-level example of Figure 4. The three HIA approximations are the result of applying a different level of lossy compression to the HIA data. They are the result of sequentially adding higher order components to the compressed data. The lowest resolution approximation (Approx 1) basically conveys the information that the first half of the spectrum is much less occupied than the second half, hinting algorithms to prefer to work on the former. The highest resolution approximation (Approx 3) is a perfect (lossless) reproduction of the HIA information. The lowest resolution data is sent to all the nodes up to R1 hops away (i.e. $TTL = R1$) every T1 seconds. The second approximation is sent with a smaller scope ($TTL = R2$)

[11] It should be noted that global information will tend to be quite stable since the presence of a protected node in a region will not depend on a particular node's position or mobility. If one node does not detect the protected node, someone else likely will.

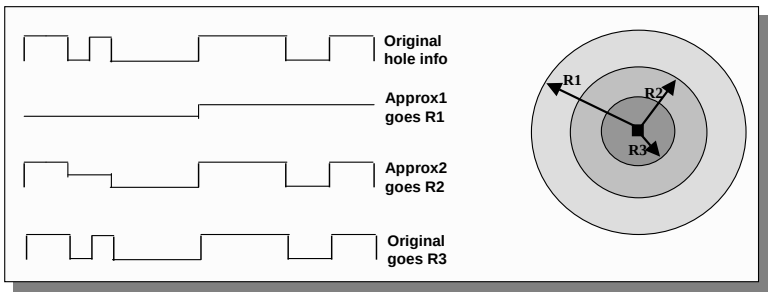

Figure 4: Example of HIA information dissemination under XOSA's HIP protocol.

but more frequently (every $T2 < T1$ seconds), and finally the lossless copy is sent at the highest rate and the smallest scope.

For the first hop, an HIP packet is transmitted by the originating node (Src ID) over the OSA-coordination channel using the R-MAC described on the next subsection. This is needed since the HIP packet carries information (HIA) needed to bootstrap the communication link among this node an its neighbors, and therefore it cannot rely on the IDLE channel or any other communication channel built based on possible obsolete HIA information. For the subsequent retransmissions (by nodes other than the HIP packet's source), the HIP packet and its HIA have no impact on the existing communication channels that XOSA created between the transmitter and the receivers (e.g. the IDLE channel), so this HIP packet is transmitted outside the OSA-coordination channel, using available spectrum opportunities. For this, ND-HIP relies on the broadcast service offered by the XOSA kernel module.

6.4 Rendezvous MAC

In this subsection we describe the Rendezvous MAC (R-MAC), designed to control the access to the OSA-coordination channel. R-MAC, an extension of our work in [13], is a simple contention-based MAC that exploit the predictability of (some of) the traffic to reduce the probability of collisions. By reducing this probability R-MACs improves on the throughput achieved by traditional CSMA-based MACs without adding much complexity or reducing the ability to handle bursty traffic.

The core idea behind R-MAC is that if a node knows, when transmitting one packet, of the time and duration of its next packet transmission, then it may append this information to the current packet transmission. Nodes receiving the packet and learning the node's intention to transmit on a future interval may choose to defer their own transmissions during that interval to prevent a collision. The transmitting nodes is not obliged to transmit at the advertised interval (for example if the carrier is busy at that time) and receiving nodes are not obliged to defer. But awareness of nodes transmission intentions and deference under certain rules reduce the chance of collisions. Ideally, when all nodes in a neighborhood receive one HIP packet – and no new node enters the communication range – they will all know when to tune for the next transmission

and continue tuning to the transmitter at the right times, resulting in collision-free receptions. In practice, hidden nodes, nodes movement, and incomplete information will result in collisions still being present, although at a smaller rate.

R-MAC uses a simple CSMA technique appropriate to the broadcast nature of the ND-HIP traffic. Before transmitting, a node senses the channel and if it is free the node transmits the packet. No RTS/CTS exchange or ACK packet is employed, since there is more than one intended destination. However, since most destinations are already waiting for the packet the chance of successful transmission is high. If the channel is busy, R-MAC enters a backoff state where the channel is reschedule for transmission a random time into the future. The rescheduling is done trying to avoid to conflict with announced transmissions, as explained below. The main differences between a node running R-MAC and one running a traditional unreliable CSMA MACs is that a node running R-MAC:

- Piggy-backs on each packet transmission its next transmission interval (if known) and a list of 'listening times', that is, time intervals where this node will be listening to someone else's transmission. The identity of the node it will listening to is also included.

- Schedules its next periodic transmission (HIP packet) as to avoid collisions with announced transmissions. Let t_{cur} be the current time and T_{per} be the ND-HIP periodic interval. The desired next transmission time will be $t_{des} = t_{cur} + T_{per} + random\_jitter$ and the transmission duration is tx_delay. [12] The HIP packet is scheduled to be transmitted following these rules:

  - Considering all the announced *transmission* and *listening* times (associated with nodes different than the one doing the scheduling) as forbidden, schedule the next transmission on the first interval $< t_i, t_i + tx\_delay > \in < t_{des}, t_{des} + T_{per} >$ that does not overlap with a forbidden interval.

  - If the above fails, choose among the intervals that do not overlap with announced transmissions (if any) the one that overlaps with the smaller number of announced 'listening intervals'.

  - If there is no interval that does not overlap with an announced transmission, schedule the next transmission to start at time $t_{des} + T_{per}$.

- Backoffs following the same collision-avoidance rules just described, trying to re-schedule a transmission attempt at a time t_{des} equal to the current time plus the MAC random backoff time. If $max\_backoff\_times$ attempts to transmit a packet failed due to the carrier being busy,

[12] It should be noted that while the transmission is perfectly known – after all, the node is doing the scheduling – the tx_delay is in general unknown. This is because the HIA is being compressed and the length of the compressed payload will depend on the information contained in it. We can, however, estimate the future value based on the previous one and adding a guard band.

then the node will enter *persistent* mode and will transmit the packet at the scheduled time regardless of the status of the carrier sensing signal.

Some implementation observations about R-MAC include:

- Instead of absolute times – which would require synchronization – each *transmission* and *listening* interval is represented by time offsets with respect to the end of the current packet transmission. Nodes receiving this information will offset the intervals based on their reception times, and therefore their record of the announced transmission intervals will already account for propagation delays. The information about the listening intervals, however, will not, and it will need to be adjusted based on the roundtrip delay between the nodes if the network has a high bandwidth-delay product. Similarly, due to different in clock skewing, a larger guard band may need to be added when the skewing differences (less than one microsecond per second even for cheap hardware) causes a time uncertainty equivalent to several bit transmissions. Since the OSA-coordination channel will typically has a small bandwidth/rate, an small guard time suffices to address both issues. However, R-MAC can be straightforwardly extended for high data rate systems.

- While the start of the next transmission interval is known – after all, the node is doing the scheduling – the tx_delay is (in general) unknown. This is because the HIA is being compressed and the length of the compressed payload will depend on the information contained in it. However, using an estimate of the future payload size based on the previous value and adding a small guard time provides good results.

- The length of the interval to look for transmission openings (T_{per} has been set to a fixed value based on ND-HIP time-sensitivity. R-MAC can be extended to work with different traffic sources/applications by setting this value adaptively to the larger inter-packet period among neighboring nodes, or to the maximum application delay tolerance. Or, it can be set to infinity for delay-insensitive applications. Also, if no collision-free interval can be found on the interval, an alternative approach can be to discard low-priority traffic instead of scheduling the transmission at the end of the looking-ahead interval.

In conclusion, R-MAC is contention-based and allow for bursty transmissions and new nodes joining the network. At the beginning or after changes due to mobility, collisions are possible as in any CSMA MAC. Over time, as nodes start to lock onto each other transmission cycles the network will converge to a schedule of broadcasts so that each node knows when to listen – collision free – for broadcasts from its neighbors. Thus, R-MAC alternate between short period of contention-based and longer period of contention-less communications, obtaining the best benefits from both: flexibility to handle bursty traffic/mobility, and high throughput.

7 Idle Channel

A node's IDLE channel is the channel used by the node's neighbors to bootstrap communication (i.e. send RTS packets) and in some cases may be the only means of communication with the node (e.g. when a CSMA legacy MAC is used). Picking and maintaining an idle channel is, perforce, an important problem.

7.1 IDLE Channel Selection.

For an OSA-enabled node the IDLE channel is not a predetermined, common channel, but rather is created based on the current opportunity availability of a node and its neighbors.

One of the goals of the IDLE channel selection algorithm is to allow for a high aggregated throughput among *all* the nodes in the network. If we looked at the problem simply between two nodes, throughput will be maximized if the transmitter used the maximum power available. The single pair scenario is *background-noise limited*. When we have a network with a large number of concurrent users, however, the situation is different. For example, as the transmission power of the different transmitters approaches infinity the achievable rate approaches a fixed value. This is sometimes referred to as the 'cocktail-party' effect: all attendees to a party, trying to compensate for the high noise, simultaneously rise their volume. Since everyone does the same, this results on the noise volume also being increased, which cancels out the speakers' volume increase. This scenario is *interference limited*.

From the above, a good transmission schedule manages transmissions so that multiple nodes, separated by sufficient distance, can transmit concurrently. Furthermore, we can limit the transmit power of those nodes already received at several dBs above the background noise level.

Once we set up the desired value for bandwidth efficiency ($r = 1$)we know the target power spectral density (psd) required to close a link to a certain distance/pathloss with that modulation, irrespective of the bandwidth being used. Not all the available opportunities will allow transmission at this target psd level. In general, we could still use this opportunities by allowing a smaller bandwidth efficiency on those frequency intervals while still using the target value (1 bps/Hz) elsewhere. However, we found that by ignoring those opportunities we could greatly simplify the algorithms as well as the amount of HIA information propagated. Thus, as previously explained in Section 5, we only consider those opportunities that allow transmitting at a psd of P_d, the power required to close the links to *all* the neighbors with a modulation with a bandwidth efficiency of 1.

There are two special cases when discarding opportunities with lower allowed transmit power can have great impact on performance: (1) there are protected nodes occupying the entire band and there is no opportunity at the P_d power level, but there are opportunities at a smaller transmit level. (2) the sensor capabilities do not allow the node to transmit at P_d or above. The first case is unlikely to appear on the short term, since currently there are chunks of unutilized spectrum,

we deferred addressing this issue to future improvements. The second case is easy to detect and may appear in practice in the short term when the opportunistic nodes want a long transmission range/power and the sensors do not have high sensitivity. In this case the transmit psd will be determined by the sensor's $O_{threshold}$ as follows: $xmit\_psd^{allowed} = P_{interference} + P_{xmit}^{min} - O_{threshold}$ (see Subsection 3.2). To be able to close the links we'll add a processing gain to the transmissions.

In addition, there is another especial case when the maximum available power at a node does not allow to transmit at full rate (1 bps/Hz) over the available bandwidth. In this case, instead of reducing the rate and spreading the signal over the entire frequency band it is better to transmit the signal at full rate (1 bps/Hz) over a smaller band, leaving the rest of the spectrum available for other opportunistic users, as in Frequency Division Multiple Access (FDMA).

We can now describe the IDLE channel selection algorithm. For each neighbor i, the kernel recorded the neighbor's HIA, pathloss ($path_i$), maximum allowed transmit psd ($xmit\_psd_i^{allowed}$, typically equal to P_d) and the maximum available transmit power (P_i^{avail}). The node also records its own background noise psd (bk_noise_psd) and is target SNR for its modulation scheme and maximum packet length (currently 12 dB).

- The available bandwidth BW^{avail} is obtained by intersecting (i.e. ANDing) the HIA's of all the neighbors and this node's. The IDLE channel will be subset of BW^{avail}, namely BW^{used}.

- For each neighbor i compute:

$$\begin{aligned} xmit\_psd_i^{desired} &= bk\_noise\_psd + TARGET\_SNR + path_i \\ xmit\_psd_i &= min\{xmit\_psd_i^{allowed}, xmit\_psd_i^{desired}\} \\ proc\_gain_i &= xmit\_psd_i^{desired} - xmit\_psd_i \end{aligned}$$

  The IDLE channel's processing gain is equal to the maximum $proc\_gain_i$ among all neighbors i, that is, $(proc\_gain = \max_{i \in \text{ Nbrs}} \ proc\_gain_i)$.

- Once again, for each neighbor i compute

$$\begin{aligned} xmit\_psd_i^{used} &= bk\_noise\_psd + TARGET\_SNR + path_i - proc\_gain_i \\ |BW_i^{used}| &= \min\{|BW^{avai}|, \frac{P_i^{avail}}{xmit\_psd_i^{used}}\} \end{aligned}$$

  The IDLE channel's frequency band size ($|BW^{used}|$) is equal to the minimum value of $|BW_i^{used}|$ among all neighbors i, that is, $|BW^{used}| = \min_{i \in \text{ Nbrs}} \ |BW_i^{used}|$. When $|BW^{used}| < |BW^{avail}|$ we choose a set of frequency intervals in BW^{avail} such that the overlapping with *known* neighbor's IDLE channel is minimized. If no overlapping exists, then we choose a subset at random.

- Finally, choose randomly among a set of known waveforms with bandwidth efficiency of 1 (e.g. QPSK) the one to use. Also, if the processing gain os greater than zero, choose the code to use. BW^{used}, $proc\_gain$, $waveform\_type$, and $code$ completely characterize the IDLE channel.

7.2 IDLE channel Maintenance

When IDLE channel it is likely leaving its neighbors unable to contact it until they update their neighborhood information with the parameters of the new IDLE channel. Due to the latency and unreliability of control packet transmissions (over the OSA-coordination channel) it is possible to induce communication shortages over otherwise healthy links. If these links are already carrying higher layer traffic (especially time-sensitive one) the chain reaction can result on significant performance degradation. Thus, care should be taken to minimize service disruption due to IDLE channel switching.

As a general rule, the IDLE channel maintenance process is performed periodically and includes some hysteresis on its parameters to prevent constant flip-flopping. The only exception to this rule is when the current IDLE channel becomes totally unreachable by a neighbor due to the activation of a new protected node. In the latter case we need to react immediately to rebuild the communication path between the two nodes.

It should be noted that the other reason the IDLE channel become inadequate, that is the pathloss between a pair of nodes increases to a point where successful communication – at the current rate– is no longer possible, is not considered critical and is dealt with periodically (TARGET_SNR already includes a margin to account for pathloss variations and besides even if the SNR becomes small still *some* packets are likely to be successfully decoded).

Regardless the IDLE channel change being event driven or signal quality driven, the procedure followed is the same: the kernel computes a new IDLE channel; it notifies the ND-HIP module who broadcast a HIP packet with the new information; the ND-HIP module notifies the kernel when the HIP packet transmission is completed; and the kernel – only at this point and not an instant before – finally switches to the new IDLE channel. The main difference is when the HIP packet is sent, since in the periodic case the HIP packet is sent in the previously announced *next transmission time*, while in the event driven case a new HIP packet is transmitted immediately (after a small jitter – msecs – to prevent collisions from several nodes reacting to the same event).

8 Experimental Results

We used OPNET to simulate our XOSA system running underneath a 'legacy' MAC system. The results are presented here. We used a CSMA 'legacy' MAC (i.e. unaware of OSA capabilities) to test that our XOSA system – while being simple – provides a complete solution that is able to

work with any existing and future MAC implementing the opportunity API. [13] For simplicity, we implemented an unreliable MAC (no ACK) when a node transmits a packet if the channel is free or backoff for a random time if the channel is busy. Nodes also backoff upon completing a packet transmission to prevent some nodes from capturing the channel. Even with this very simple MAC we were able to show a good performance enabled by using XOSA.

Table 1 summarizes the default simulation parameters for both protected and opportunistic nodes. Unless explicitly stated otherwise, our simulations consist of 60 opportunistic and 10 protected nodes placed randomly in a square area of 0.6 miles x 0.6 miles. The protected nodes are fixed and not always active. When they are active, they continuously transmit packets using the ON/OFF cycle shown, and record the period of time that they experienced interference by opportunistic nodes beyond the maximum tolerance. Their transmission range is equal to 2 Km. which is consistent, for example, with current cellular systems. The opportunistic nodes are mobile. Mobility is random (rate and direction) up to 10 mph, with nodes bouncing back each time they reach the area boundary. Their transmission range is 250m., which is consistent with current WLANs such as those based on 802.11. Opportunistic nodes generated a new packet to transmit as soon as the previous transmission is completed. This corresponds to a saturation case, particularly challenging for an unregulated CSMA MAC without feedback-based load balancing. Each packet's destination is chosen randomly among the one hop neighbors. We collect statistics on the total number of packets successfully received by each node. Thus, we collect two main figures of merit: the total aggregated one-hop throughput (sum of all bits received by each node), and the maximum interfered time (the time a protected node measured interference above its tolerance).

Finally, the physical layer modeled follows the assumptions stated in Subsection 3.3 plus some capabilities such as sub-noise detection (enabled by a DSP intensive synchronization stage). Thus, if a large processing gain is used, the sensing range could be smaller than the transmission range. The propagation model assumes a power decay factor with distance of 4, i.e. the received power is $\Theta(\frac{1}{d^4})$, where d is the distance separating two nodes.

8.1 Spectrum Utilization and Throughput relative to Capacity.

To show how our XOSA system exploits unused spectrum, experiments were conducted on a very simple network consisting of 4 nodes in a line, separated by 250m each. The protected nodes' tolerance was decreased so that all the opportunistic nodes were inside their protected area. Since we wanted to accentuate the negative impact of XOSA's coordination channel and modulation-type requirement, we increased the size of the coordination channel to 1.5MHz and the opportunistic nodes' maximum transmit power to 40dBm.

[13] For example, a legacy MAC that is sensitive to timing information tied to the packet's transmission rate will need to specify their desired transmission rate to XOSA. If this rate is achievable, XOSA will create IDLE channels with that rate.

General Parameters	
Simulation area	0.6 miles x 0.6 miles
Simulation time	200 sec.
Base Frequency	2.3 GHz
Total Bandwidth	100 MHz
Unassigned Bandwidth	5 MHz
Background noise level	-174 dBm/Hz
Protected Nodes Parameters	
Bandwidth	19 MHz
Xmit power density	-26 dBm/Hz = 2.5 Watts/MHz
Communication range	2Km
Aveg. On time	160 msec
Aveg. OFF time	180 msec
Opportunistic Nodes Parameters	
Speed	0 to 10 mph
Coordination Channel	200 KHz
Max Xmit power	23 dBm
Targeted policy allowed transmit psd	-62dBm/Hz = 0.625 mWatts/MHz
Transmission range	250 m
Target SNR	12 dB
Data power margin	10 dB
Sensing period	8 msec
Sensing (memory) window	256 msec
Sense Threshold	-138 dBm/Hz

Table 1: Default Simulation parameters.

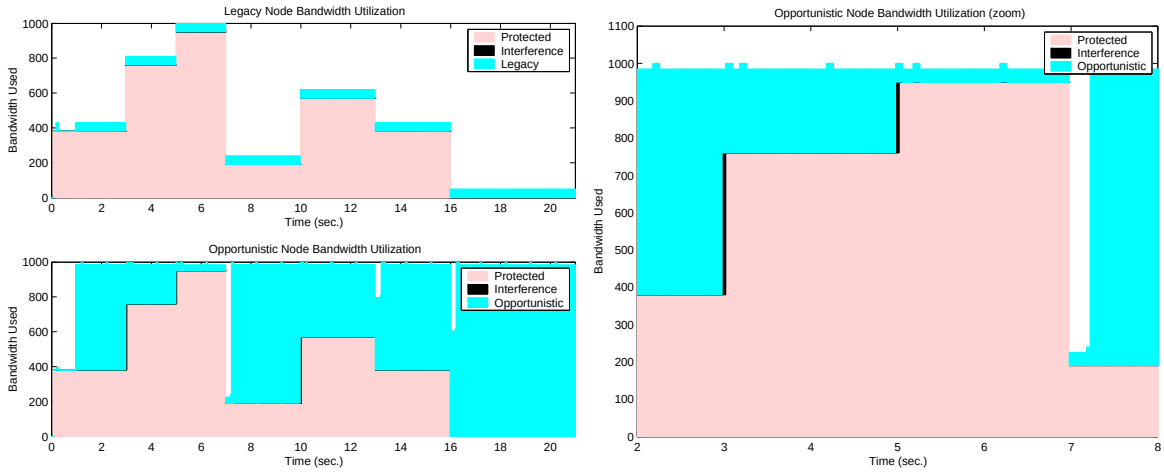

Figure 5: Legacy (upper left) and XOSA (lower left) systems' spectrum utilization for a protected node's load of 40%. A more detailed snapshot of XOSA's spectrum utilization is shown on the right.

Figure 5 shows how our algorithms take advantage of the unused portion of the spectrum. For these plots, we have divided the time into 40msec intervals and consider a frequency band used if a transmission occurred over any part of the 40 msec interval.

The upper left plot shows the way a legacy (i.e. current, fixed spectrum allocation) system –running the same MAC – used the frequency. The protected nodes are assigned 95 MHz of spectrum, so the legacy radios can use only the remaining 5 MHz. The frequency band has been divided in frequency slots of 100 KHz each, and the number of frequency slots occupied by each - and not the actual frequencies – is shown. The lower area represents the protected nodes utilization. For example, in the interval $< 0, 3 >$ two protected nodes are active occupying 380 slots (i.e. 38 MHz). Similarly, in the interval $< 16, 21 >$ no protected node is active. Averaging over the interval $< 1, 21 >$ (i.e. after initialization) and over their assigned frequency bands, the protected nodes' present an utilization of 40%. Looking at the legacy node's bandwidth utilization, we can see the legacy nodes always use the same amount of bandwidth (the unassigned 5 MHz) and are constantly transmitting, with the exception of the interval $< 0, 1 >$ (initialization) when they are only active for short bursts sending neighbor discovery beacons.

The lower left plot illustrates the way XOSA radios exploit the available spectrum. After the one-second initialization period is completed, we see that XOSA fills the spectrum gaps almost completely, except for minute periods of time. Dark lines represent the periods where XOSA radios interfere with protected nodes. This interference is due to the latency on sensing the channel. It may take up to two sensing intervals (8 msec each) for a XOSA-enabled radio to detect that a protected node has become active, resulting on interference periods of up to 16 msec per each time a protected node becomes active. However, since our time step is 40 msec, we paint the whole 40 msec interval as 'interfered'.

The right plot is a 'zoom' of the lower left plot (XOSA spectrum utilization) for the interval

$< 2, 8 >$. This allows a more detailed examination of XOSA's behavior. First, we may notice that XOSA does not occupy the entire 1000 frequency slots all the time. Typically there will be a small gap, corresponding to the coordination channel's 15 slots. Only when HIP packets are sent (periodically every second and also event-driven upon protected nodes' activation at times 3 sec. and 5 sec.) XOSA will occupy the entire spectrum. Thus, the coordination channel and XOSA control packets reduces the DATA throughput achieved. Second, we may see that upon a protected node's activation (for example at time 3 sec.), XOSA transmissions will interfere with protected nodes until their local sensors detect the protected node signal, and before any control packet is sent they will stop transmitting on the protected node's frequency band. This implies that they will not be able to communicate with any neighbor whose IDLE channel contains parts of that band (all the 4 nodes in our example). The XOSA nodes will then recompute their IDLE channel and include it together with their new HIA information in a HIP packet that is broadcasted after some random time (to avoid collisions from all opportunistic nodes sensing the protected node's signal at the same time). Thus, in Figure 5 (right) we notice that the coordination channel usage is posterior to the ending of the interference period. So, from the time the protected node is detected until a HIP packet with new IDLE channel information is received, the nodes will not be able to reach neighbors (due to invalid IDLE channels). The fact that we don't see white gaps in the spectrum occupancy at times 3 sec. and 5 sec. indicates that the nodes recover new information and resume transmission in less than 40 msec. Thus, the periods of communication disruption in XOSA are small. Third, when at time 7 sec. protected nodes become inactive, we see that there is a period (around 300 msec.) when the system is unable to fully occupy the spectrum. Locally detecting this opportunity is not enough to use it - we need the neighbor nodes to be able to transmit over it. Therefore, XOSA nodes must wait until receiving neighboring nodes' HIA updates (in this case once every second) declaring the opportunity before recomputing their IDLE channel. In the figure, around time 7.3 sec. we see the HIP packets being sent, and only after that the IDLE channels are recomputed, another round of HIP packets with the new IDLE channel information are sent, and the XOSA nodes start using the new opportunities. Note that during this transition there is no transmission interruption since XOSA nodes tune to the new IDLE channel only after the HIP packet transmission has been completed. In between (i.e after waiting for a spreading time to avoid collision, for the coordination channel to become free, and for the transmission delay) the XOSA nodes are still tuned to their last-advertised IDLE channel. Overall, XOSA reacts fast to new opportunities. It could even react faster if HIP packets are sent immediately after detecting new opportunities, similar to the case of losing opportunities being used by someone's IDLE channel, but this is not advisable since it may lead to congestion on the coordination channel. Thus, only critical (i.e. causing communication disruption) events trigger extra HIP transmissions.

The two lower curves of Figure 6 plot the total throughput achieved by the legacy and XOSA systems for different levels of bandwidth utilization by the protected nodes. For example, the

experiment shown in Figure 5 corresponds to the 40% utilization data points. Since we set a large coordination channel of 1.5 MHz, it is expected that when the protected nodes utilization approaches 100%, the legacy system (able to transmit over 5 MHz) outperforms the XOSA network which can only use 3.5 MHz for data transmission. It can be seen that under a bandwidth utilization of 40% (consistent with current spectrum occupancy measurements, as for example [14]) XOSA gives an order of magnitude performance improvement over legacy systems, even when assuming a generous allocation badwidth allocation for the latter (5 MHZ of exclusive spectrum usage). For comparison we are also plotting the maximum achievable throughput (i.e. capacity) which was computed by solving the MT problem described in Subsection 3.4. We can see that our XOSA system is far from achieving the optimal value. In general, there are several factors preventing XOSA from achieving the maximum. Among them includes fairness (i.e. the optimal MT solution gives exclusive access to the shortest links), using energy-inefficient modulation, not using long and highly efficient but computationally expensive codes, etc.. In this example, the main reason for the difference was our use of energy-efficient modulation. We always used a bandwidth efficiency of 1 bps/Hz or less, therefore saving energy and extending the network lifetime, an important factor in mobile networks. To corroborate this, we are also plotting the maximum throughput achieved if the waveforms used are restricted to Direct Sequence Spread Spectrum (DSSS) with different codes and processing gains. As we may see, our system is within reason of the maximum achievable by an energy-efficient system for this simple topology. More complicated topologies, though, will have an extra negative impact on the simple legacy MAC.

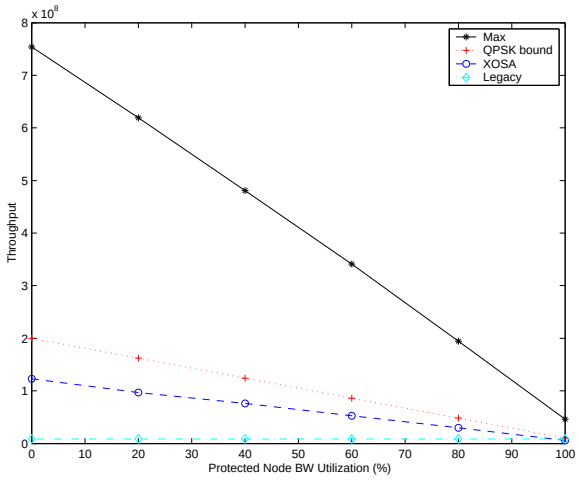

Figure 6: Comparing XOSA achieved throughput against a legacy system's, the maximum achievable (Shannon capacity), and the maximum achievable if only QSPK modulation is allowed.

8.2 Topology Control versus Throughput

Out of all the different factors affecting XOSA's achieved throughput, the one that has a greater relative impact is that of topology control. That is, choosing the right set of nodes to communicate directly with. In traditional networks, the frequency and waveform used for receiving packets is predefined. Thus, the effect of not having an effective topology control algorithm in place is to have a higher number of nodes contending for access to the same channel. In an OSA system, however, the negative effect is greater since these systems need to decide on-the-fly the channel they will listen to while in IDLE. This channel must be reachable by all the nodes chosen as neighbors, forcing the IDLE channel to be contained within the intersection of all the neighbors' opportunities. Similarly, the processing gain of the IDLE channel waveform needs to be adjusted so that the weakest neighbor signal can be successfully decoded. Thus, each node being added to the 'neighbor set' will likely decrease the IDLE channel rate. This IDLE channel rate reduction is most critical when a legacy (OSA-unaware) MAC is used. For contention based OSA-aware MACs as our XOSA-MAC, the negative impact is somehow reduced since only the RTS uses the IDLE channel. The DATA packet will be sent over a band negotiated between the sender and receiver only. Also, TDMA-based OSA-aware MAC may not use the IDLE channel for unicast transmissions at all, since the nodes always know which source is transmitting and therefore they can tune to the proper waveform (active listening). However, in both previous cases, using links with small bandwidth/low rate will deprive channel access to higher rate communication pairs resulting in lower total throughput. [14]

Figure 7 shows the throughput achieved when varying the range between neighbors, while keeping the network connected, in a network with 60 nodes, and 5 active protected nodes. For example, the second data point (range equal to 107 dB) is obtained when a node chooses as neighbors all those nodes whose signal it can received with an attenuation (pathloss) of 107dB or less. For our examples (and since we have omnidirectional, doughnut-shape antennas with a horizontal gain of 6dB) this translates to a communication range of at most 250m. Similarly, a range of 119 dB translates to a communication range (among neighbors) of at most 500m. This variation in range is accomplished by increasing the transmission power of the HIP packets sent over the coordination channel. Figure 7 also shows – for reference – the throughput achieved by a legacy (OSA-unaware) system transmitting over the unassigned frequency (5 MHz).

It can be seen that both curves (XOSA and legacy) present sensitivity to a bad choice of communication range. However, the XOSA system performance degradation is much more steep, decreasing 85% after an increase of 6dB in communication range. This points to the need of

[14] Networks using TDMA OSA-aware MAC may handle this issue at the routing layer, by choosing to send data over throughput optimal paths using high rate links only. However, it is better to handle this as a topology control problem since reducing the number of neighbors/links will simplify the processing needed at the MAC scheduler. Besides, control packets that need to be broadcasted will always use the minimum-common-denominator channel among the chosen neighbors.

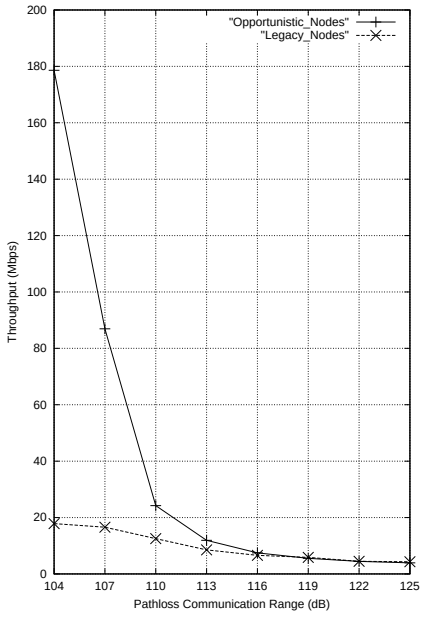

Figure 7: Throughput versus maximum communication range (in Pathloss dB).

effective topology control mechanisms to fully exploit the advantages of Opportunistic Spectrum Access. For the remaining of our experiments, we set the pathloss range to 107 dB. We prefer this value to 104 dB since for our network density, the former value is a safer bet to guarantee network connectivity under different topologies. In real life we don't have the luxury of knowing the exact nodes trajectories beforehand, so choosing a conservative value is more realistic.

8.3 Interference Tolerance versus Throughput

Figure 8 shows our throughput results versus the number of active protected nodes for several interference tolerance levels. The throughput achieved by a legacy system is also shown for comparison. The tolerance levels are labeled relative to the background noise. For example, the curve "BK_NOISE" refers to the case where the protected nodes can tolerate a total interference from opportunistic nodes equal or less than the background noise level. That is, the protected nodes tolerate a total noise that is twice the background noise value; or in other words, they can tolerate a 3 dB decrease in their SNR. Similarly, the curve "BK_NOISE - 6" ('BK_NOISE + 6") refers to the case when the protected nodes are able to handle an interference 4 times smaller (larger) than the background noise level, resulting in up to 1dB (7 dB) decrease on their SNR. The curves cover a range of SNR reduction from 0 to close to 10 dB. For each curve we set the sensor's O_{cut_off} value as to achieve the desired transmit power spectral density (psd) of -62dBm/Hz. By using a power margin of 0 dB, we have that for tolerance values of -174dBm/Hz, -168 dBm/Hz, and -180dBm/Hz (i.e. background noise, background noise plus 6dB, and background noise minus 6 dB) we need to set O_{cut_off} to -138dBm/Hz, -144dBm/Hz, and -132dBm/Hz respectively.

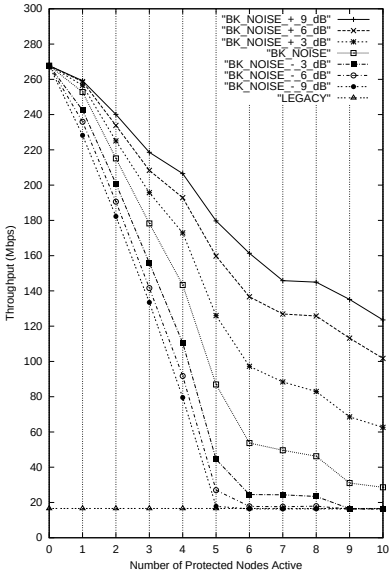

Figure 8: Throughput versus protected node's activity for several interference tolerance levels.

The protected nodes are paired (1 with 6, 2 with 7, and so forth) so that each pair is assigned a non-overlapping 19 MHz segment. Since the protected node's coverage area (of 2 Km. radius) is more than twelve times our simulation area, we may have none, one, or all 10 protected nodes inside the simulation area. It should be noted that when the distance between communicating protected nodes (say the pair 1 - 6) is on the same order as their transmission range (i.e. 2 Km.) it is unlikely that both nodes will be inside the same 0.6 miles x 0.6 miles area. So, even in the case where all the assigned frequency is actually occupied (i.e. all five pairs are active), we shouldn't expect more than 5 protected nodes present in our simulation area. So, the throughput values associated with 5 protected nodes active should be considered as a lower bound on the throughput gains expected under XOSA. In practice, spectrum measurements as the one reported in [14] show that as much as 66% of the assigned spectrum in the $< 1.4GHz, 2.9GHz >$ band is unoccupied, so much larger throughput gains can be realized.

From the above figure, it is evident how a small tolerance to SNR decrease by the protected nodes will result on overall significant throughput gains, making a case for policy makers to adopt OSA. Plots like the ones shown will allow policy makers to perform a trade off balance between the SNR protection to current and future licensees and the overall throughput gains (commonwealth).

Finally, it should be mentioned that since we did not apply any margin on the policy-allowed transmit power computation, there was the possibility that several opportunistic node's transmission would combine to induce interference above tolerance. However, given the boundary effects of our small simulation area, opportunistic nodes did not surround the protected nodes but at most lay on the same quadrant. Due to the carrier sensing method employed, only one opportunistic node would transmit on the protected node's frequency at any given time. Thus, the periods of

destructive interference were small. However, in the general case when opportunistic nodes may surround a protected node, a margin should be added to the power computation. This was explored in the set of experiments shown in Subsection 8.5.

8.4 To Underlay or To Not Underlay?

So far, we have assumed that the opportunistic users are using underlaying, as explained in subsection 3.1. However, the use of underlay requires superior hardware capabilities by the opportunistic nodes, as for example the ability to reject strong interference from the protected nodes. To evaluate the relative impact of underlaying in the overall system performance, we run some experiments where underlaying is disabled by setting the sensor's threshold to the same value level as the background noise. Thus, opportunistic nodes will only access the spectrum if the protected nodes energy level is equal or below the background noise level, and therefore the opportunistic nodes will only need to tolerate a noise floor level of 3 dB above background noise.

Figure 9 compares the throughput achievable when underlaying is possible and when it is not, for a protected nodes' tolerance of -174 dBm/Hz (i.e. a 3 dB decrease on SNR). We can see that when a small number of protected nodes are present underlaying is not strictly necessary to reap the rewards of opportunistic spectrum access. Indeed, when the number of primaries is less than 3 (occupying 57% of the spectrum), systems that do not use underlaying reach no less than 60% of the throughput achievable by means of underlaying. However, when the number of primaries increases to 5 or more (occupying 95% of the spectrum), underlaying is necessary to achieve good throughput levels.

It can be noted that the results obtained without underlaying are similar to the results obtained in Figure 8 for the lowest protected nodes' interference tolerance (9 dB below background noise, or -183 dBm/Hz). However, the event "X protected nodes are present" in both curves are different. In Figure 8, when underlaying was allowed and the sensing threshold was set to -147 dBm/Hz, the sensing area (for a power decay exponential factor of 4) was at most 840m, while in Figure 9, without underlaying, the sensing threshold has been set to the background noise level (-174 dBm/Hz) which results in a sensing range of 4Km. Thus, when in Figure 9, we say that "X protected nodes are present", we are not referring to our 960m x 960m simulation area, but to a area of radius 4Km (roughly, 50 times bigger). We are referring to a much smaller density of protected nodes.

Thus, *a priori* it may seem that the event "only 2 protected nodes present (in a 960m x 960m square)" is more likely that the event "2 protected nodes present in an area of radius equal to 4Km". And that the results without underlaying have no practical consequence. However, when we compare this event (only two protected nodes present, i.e., only 38% of the spectrum is occupied" with the (*a posteriori*) results of the spectrum survey conducted in [14], we see that 38% occupancy is actually *greater* than the current spectrum occupancy levels.

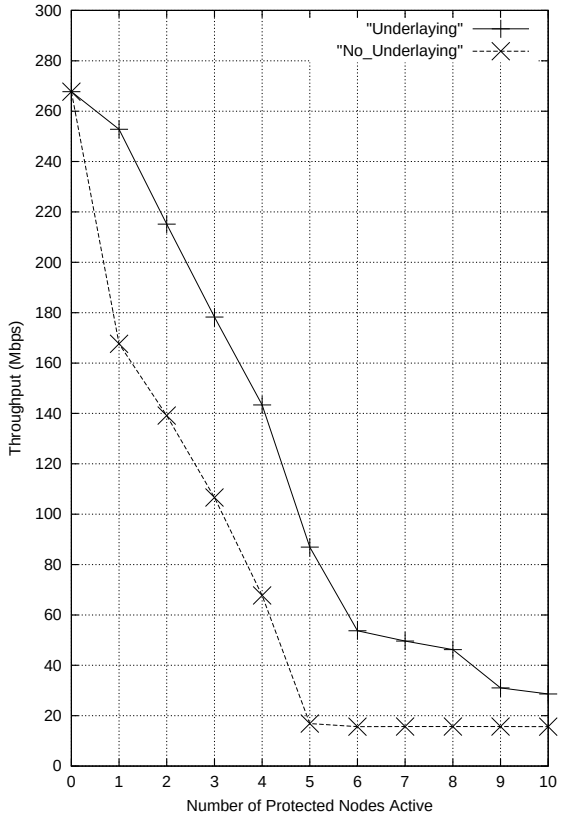

Figure 9: Throughput achieved with and without underlaying.

Therefore, we conclude that for the current levels of spectrum occupancy, solutions that do not employ underlaying may achieve a great portion (at least 60%) of the throughput achievable through opportunistic spectrum access. However, in the longer term, when most of the spectrum becomes occupied, underlaying will be the key to unlock the full potential of opportunistic spectrum access, that is, even when 100% of the spectrum is occupied (i.e. protected nodes' signal level above background noise) it is still possible for nodes with shorter transmission range to concurrently occupy the spectrum, achieving high throughput.

8.5 Policies: Carrier sensing versus interference margin

It was mentioned in Subsection 3.2 that the cumulative effect of simultaneous transmissions on a protected node's experienced interference required the inclusion of a MAC-dependent margin on the opportunistic nodes' maximum allowable power. It seemed that policy makers would need to regulate also the MAC employed by opportunistic nodes. Fortunately, for a large class of MAC mechanisms, this is not the case. Indeed, for CSMA-based MACs, it suffices to regulate the value of the power margin according to the carrier sensing threshold employed. Even for TDMA-based MACs policy-makers can avoid regulating the inside of the MAC by requiring that these non CSMA systems sense the medium before transmitting and limit their transmit power based on the amount

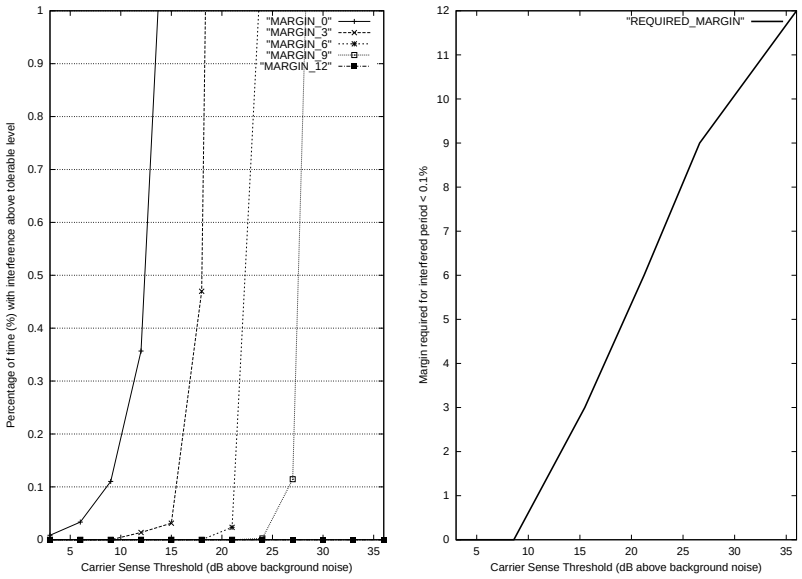

Figure 10: Percentage of time experiencing destructive interference as a function of the opportunistic nodes' carrier sensing threshold (left) and required margin to guarantee a maximum period of non-tolerable interference below 0.1% (right). Protected node's tolerance is 12dB above the background noise level.

of carrier energy sensed.

To study the dependence of the required power margin on the carrier sensing threshold used we conducted a set of experiments where a protected node is located at the center 2Km x 2Km square and it is surrounded by 240 opportunistic nodes forming a connected network. The Protected node's interference tolerance was set to 12 dB above the background noise level, that is, -162 dBm/Hz. The opportunistic nodes vary their MAC carrier sense threshold from 3dB to 36dB above the background noise level. Since the background noise increases with the bandwidth used, so will the carrier sensing threshold.

Our results are shown in Figure 10 (left), where we plot the percentage of time (if any) the protected node experienced an interference above its tolerance, which is set to 12 dB above the background noise level (i.e. -162dBm/Hz). For example, for a carrier sensing threshold of 15dB above background noise level we need to add a 3dB power margin to keep the periods of destructive interference experienced by the protected nodes to less than 0.03% of the time. If policy makers want to eliminate any period of destructive interference (likely an overkill) they will need increase the power margin to less than 6dB. From these results and for a given acceptable period of destructive interference (in particular, zero) we can derive *power margin* versus *carrier sensing* curves as the one shown in Figure 10 (right) for a destructive period of up to 0.1% and the same protected node's tolerance level. Thus, these curves can be used to choose the right policy to apply for a given propagation model. [15]

[15]In general, the lower the power decay factor the larger the required margin. The worst case being free space with a power decay factor of 2.

Finally, it should be noted that adding a power margin will reduce the maximum allowed transmit power (or increase the protected area in case a minimum power is required) resulting in a smaller throughput. In general, having a tolerance of $P_{interference}$ and a power margin of $power\_margin$ will produce the same throughput of having a tolerance of $P_{interference} - power\_margin$ and a power margin of 0 dB. Thus, we can estimate the throughput gains by looking at the results in Figure 8.

9 Summary and Conclusions

Opportunistic Spectrum Access (OSA) offers tremendous potential for an order of magnitude improvement on the capacity of ad hoc networks. To achieve this potential, however, several key issues need to be addressed in a systematic fashion. Our work, XOSA, is the first complete solution for an ad hoc network exploiting OSA.

XOSA architecture is highly modular, with its components interacting through well defined interfaces, specially the opportunity API. XOSA performs all the critical OSA functionalities allowing any API-complaint MAC, even legacy OSA-unaware ones, to be run of top of it. At the same time, XOSA provides the interfaces to allow OSA-aware upper layers to run optimized algorithms. Thus, XOSA constitutes the 'core' of an OSA system while leaving the upper layer room to perform their innovations. Our main goal in designing XOSA was to get a working system, and this often made us trade off performance for simplicity. In some cases – as for example the design of the Rendezvous MAC and the mechanism for quick IDLE channel recovery – we couldn't avoid engaging in a more elaborated design because of the expected critical impact on performance.

We reported on experiments using XOSA and a high fidelity simulator. Our results show that an order of magnitude increase on capacity was achieved for realistic scenarios using our 'core' XOSA system. We believe that there is considerable scope for further research in this area. Our experiments identified the 'stress points' affecting performance. Research on those areas will likely produce the highest benefit. We noticed that an effective topology control mechanism is critical to the system performance, and therefore implementing one should be first on the list of improvements. Also, for the short term, when spectrum occupancy is below 40%, underlaying in not critical to achieve OSA's potential, and therefore preference may be given to develop more efficient MAC and IDLE channel allocation algorithms. However, in the longer term, when most of the spectrum becomes occupied, underlaying will be the key to unlock the full potential of OSA. Finally, our results also serve policy makers on the determination of suitable policies that trade off incumbent rights and overall capacity (the commonweal).

References

[1] Spectrum Policy Task Force, "Spectrum Policy Task Force Report," ET DOCKET No 02-135, November 2002.

[2] Chris Lang, "President Bush Orders Goverment Spectrum Review," In *IEEE Spectrum Magazine*. August 2003.

[3] DARPA ATO neXt Generation (XG) Communications Program, http://www.darpa.mil/ato/programs/XG/

[4] X. Jing, D. Raychaudhari, "A Spectrum Etiquette Protocol for Efficient Coordination of Radio Devices in Unlicensed Bands," Proc. PIMRC, Beijing, 2003.

[5] S. Mangold, K. Challapalli, "Coexistence of Wireless Networks in Unlicensed Frequency Bands," Wireless World Research Forum No. 9, Zurich, Switzerland, July 2003.

[6] N. Golmie, R.E. Van Dyck, A. Soltanian, A. Tonnerre, O. Rebala, "Interference Evaluation of Bluetooth and IEEE 802.11b Systems," Wireless Networks, 9(3), pp. 201-211. 2003.

[7] R. Ramanathan, C. Santivanez, S. Polit, C. Partridge, M. Condell, and R. Krishnan, "Policy Responsive Opportunistic Spectrum Access: Vision and Challenges," Work-in-progress.

[8] J. Mitola III, "Cognitive Radio: An Integrated Agent Architecture for Software Defined Radio," PhD Dissertation, Royal Institute of Technology (KTH) Sweden, May 2000.

[9] J. Mitola III, G.Q. Maguire, "Cognitive Radio: Making Software Radios More Personal," IEEE Personal Communications, August 1999.

[10] W. Horne, P. Weed, D. Schaefer, "Adaptive Spectrum Radio: A Feasibility Platform on the Path to Dynamic Spectrum Access," Fifth Annual International Symposium on Advanced Radio Technologies, Mar. 2003.

[11] R. Ramanathan, J. Redi, C. Santivanez, D. Wiggins and S. Polit, "Ad Hoc Networking with Directional Antennas: A Complete System Solution," *IEEE Journal on Selected Areas in Communications*: Special Issue on Wireless Ad Hoc Networks, March 2005.

[12] R. Ramanathan and R. Hain, "Topology Control of Multihop Radio Networks using Transmit Power Adjustment," In *Proceedings of IEEE Infocom 2000*, Tel Aviv, Israel, Mar 2000.

[13] C. Santivanez and J. Redi, "On the Use of Directional Antennas for Sensor Networks," In *Proceedings of MILCOM'03*, Boston, MA., Oct. 2003.

[14] The New America Foundation ans the Shared Spectrum Company, "Dupont Circle Spectrum Utilization During Peak Hours", June 20, 2003, http://newamerica.net/Download_Docs/pdfs/Doc_File_183_1.pdf