

Towards an Architecture for Efficient Spectrum Slicing

Suman Banerjee¹, Arunesh Mishra¹, Vladimir Brik¹, Vivek Shrivastava¹, Paramvir Bahl²

¹University of Wisconsin-Madison, ²Microsoft Research

Email: {suman,arunesh,vladimir,vivek}@cs.wisc.edu, bahl@microsoft.com

Abstract

With the increased demands for wireless spectrum, dynamic spectrum sharing is emerging as an important and powerful concept. Most research in this domain is being conducted in design of cognitive radios and on specific PHY and MAC layer challenges associated with them. However, for a dynamic spectrum sharing architecture to be viable, research is needed to resolve many other challenges, e.g., in the context of real-time spectrum management and enforcement. This paper is the first to present a study of some such important architectural considerations, driven by our ongoing design and implementation of a spectrum sharing system, called Spark. We propose some promising approaches to address these challenges, and enumerate the need and opportunities for significant future research in this domain.

I. INTRODUCTION

Efficient allocation and use of spectrum, a central problem for all mobile and wireless communication systems, is regulated today by governing bodies such as the FCC in the US and the Ofcom in the UK, using two different approaches. (i) A *spectrum licensing* approach in which exclusive use of a frequency band is conferred through the sale of a license, e.g., the PCS band, and (ii) the *commons (or unlicensed)* approach where users are allowed to share the spectrum without any licensing requirements (see [1] for a detailed discussion), e.g., the 2.4 GHz ISM band.

Both the licensed and commons approaches have their own advantages. For example, the commons approach can be viewed as a ‘bazaar’ model of operation that has naturally spurred innovative spectrum sharing technologies leading to efficient spectrum utilization. Similarly, the licensed approach, with its exclusivity of spectrum use, is particularly attractive to users with strong quality of service and interference protection requirements. From an economic standpoint, it is likely that even as regulatory bodies make more spectrum available using the commons model, they will have to be sensitive to the impact this has on users who have made significant financial investments in purchasing spectrum licenses.

Under-utilization in licensed bands: Spectrum licensing under the current models is done statically and at a fairly coarse granularity over time and spatial domains. For example, a typical spectrum license issued by the FCC in the PCS band in the US recently span multiple counties (sometimes even multiple US states), and is valid for a 10 year period. Recent studies, e.g., the Shared Spectrum Company (see www.sharedspectrum.com) have shown that such static and long-term spectrum licenses leads to significant underutilization — even in densely populated urban areas of the US, spectrum occupancy rarely exceeded 25%. To address this limitation, the FCC in 2004 legalized secondary markets for spectrum — a primary licensee is now allowed to sub-lease spectrum access to other secondary incumbents [2].

Motivated by these developments, there has been, and continues to be, significant research effort devoted to efficient spectrum utilization strategies, primarily through design of cognitive radios. (Cognitive radios [3] are radios which can adapt their operating parameters by sensing and learning about their environments.) Most of such endeavors address physical and MAC layer challenges associated with these radios, including efficient and dynamic sensing of spectrum [4], coordinating use of common frequencies [5], and managing spectrum contention [6]. Assuming that these issues can be reasonably addressed, for dynamic spectrum sharing to be viable, we still would need to resolve many other challenges, especially in the context of real-time spectrum management and enforcement. This paper presents a critical study of these important architectural considerations (unexplored in previous work) that were discovered through our ongoing design and implementation of a dynamic spectrum sharing system.

Dynamic spectrum sharing through secure secondary licenses: In this paper, we present the design of a spectrum sharing architecture, called *Spark*, that realizes the recently legalized secondary markets for spectrum. The two main entities in Spark are the spectrum buyer, i.e., the secondary user, and the spectrum seller, typically the primary licensee (spectrum owner) but can also be a secondary user interested in re-sale of previously bought spectrum. As part of a Spark transaction, the seller issues a secure *Secondary license* (or *Slice*) to the buyer¹. We advocate a *flexible slice use policy* — once issued, the seller imposes no restrictions on how the buyer uses the spectrum slice, as long as they adhere to the slice parameters. Such flexibility typically promotes innovative spectrum use — it allows buyers freedom to develop new (PHY and MAC) mechanisms that utilize their slices in the most efficient manner. It also means that the secondary user can, in turn, re-sell a portion of this slice to other potential buyers. This explicit design goal, thus, facilitates both technological and economic competition in the secondary marketplace ultimately leading to greater innovation and efficient spectrum utilization. Such a secondary licensing mechanism is particularly beneficial to *transient and mobile users of spectrum* who prefer relatively interference-free, short-term spectrum allocations in a fixed location, and then move to a new location. In particular, we believe that this concept can become powerful enough that *all future mobile devices can get enhanced to take advantage of such secondary licensing mechanisms*.

In order to promote seller confidence in this dynamic spectrum sharing architecture, we advocate *enforcement* strategies that validate buyer conformance to slice parameters. Owing to the flexible slice use policy, such enforcement mechanisms cannot be performed using centrally managed sensing architectures. Instead, we argue that verification of slice parameters needs to be performed *in-band*, i.e., on the transmit/receive path of a Spark radio, assisted by specific tamper-proof hardware.

To meet the above objectives, Spark uses the following constructs: (i) *Hardware-based*: We propose the design of a Spark-capable radio interface, which combines a Software-Defined Radio (SDR) with a tamper-proof hardware. While the SDR facilitates flexible spectrum use, the tamper-proof hardware implements slice enforcement. We posit that in order to leverage such opportunistic use of spectrum in licensed bands through secondary market mechanisms, future SDR-based interfaces will need to be specifically designed with a tamper-proof component. Proper operation of both these components are ensured through a certification process, e.g., by the FCC; (ii) *Software-based*: This includes protocols and algorithms for real-time spectrum management and utilization, as well as software components for slice enforcement.

Challenges: Through the design of Spark, we focus on three important challenges that arise in any dynamic spectrum sharing approach based on secondary markets. They are: (i) *real-time spectrum management*: mechanisms to allocate and de-allocate spectrum at short timescales (analogous to real-time IP address configuration and management in Internet hosts which is conducted using DHCP), (ii) *slice enforcement*: mechanisms to provide a reasonable validation of the buyer being restricted to their slice specifications in time, space, and frequency (we believe that to a large extent, popularity of secondary markets will depend on the confidence of sellers that buyers will not misuse their temporary usage rights), (iii) *spectrum fragmentation*: mechanisms to mitigate fragmentation, and consequent under-utilization of spectrum, that results from the continuous allocation and de-allocation of this five dimensional resource, three dimensions of space, and one dimension each of spectrum frequency and time.

Related work: The notion of secondary spectrum markets, itself, is not new and were previously proposed by Peha et. al. [7] and in Dimsumnet [8], both in the context of cellular networks. Peha et. al. [7] had focused on the modeling and analysis of such a market and its positive impact on network performance of users, and not on architectural or system design issues. Dimsumnet describes secondary market methods that are closely controlled and implemented by cellular phone manufacturers. They also advocate an out-of-band method for spectrum sensing (requiring a separate spectrum sensing infrastructure). Apart from the fundamental design difference of *in-band* spectrum of Spark, this paper introduces other spectrum sharing challenges in real-time spectrum management and its impact on spectrum fragmentation. Opportunistic spectrum use [9], [10] and the etiquette-driven spectrum sharing methods [5], [6] are other spectrum sharing approaches which

¹We use the term Spectrum Slice to indicate a part of the spectrum resource made available to a buyer through a secure secondary license.

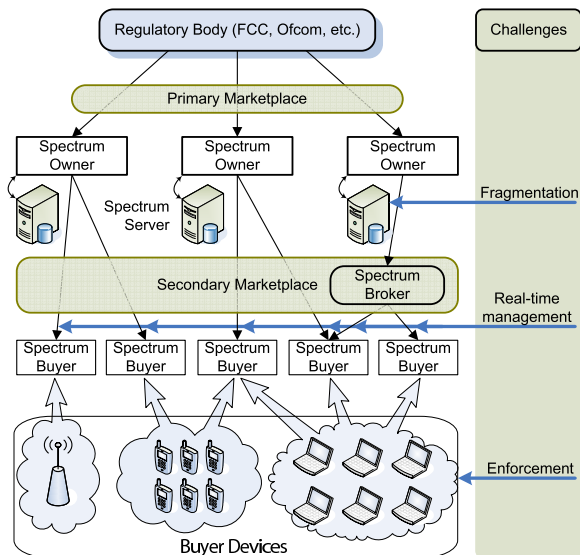


Fig. 1. The spectrum flow pyramid in Spark.

are complementary to the secondary market based model. The key difference of Spark, with respect to such approaches, lies in the use of *enforceable* secondary licences that provide some performance guarantees and rights to secondary users.

In the next section we discuss the Spark architecture and some of the unique aspects of this approach.

II. SPARK ARCHITECTURE AND COMPONENTS

We explain the different components of the Spark architecture through the *spectrum flow pyramid* shown in Figure 1. The arrows in the figure indicate how spectrum assignments move from the top of the pyramid down to end-user devices which utilize the spectrum.

At the top of this pyramid, are governing bodies, e.g., FCC or Ofcom, that regulate different aspects of spectrum policy. Through occasional auctions and sales, these regulatory bodies issue relatively long-term spectrum leases, say for a 10 year period, on contiguous blocks of spectrum to *primary spectrum owners* spanning a large geographic region, say an entire state in the US. We refer to this controlled market as the primary marketplace.

The focus of Spark is on the *secondary marketplace* through which the primary owner re-sells previously bought spectrum to potential secondary users. These could also be *spectrum brokers* who could aggregate spectrum from multiple owners and auction them favorably through open market mechanisms. Such a flexible and open model for secondary leasing allows competition and thus exposes this market to some of the forces that spurred innovation in the unlicensed spectrum domain. This is similar to the online brokers for airline tickets (such as *Priceline* or *Expedia*) which support open competition by enabling the customer to compare the services and prices alongside each other. We now discuss the three important design components of Spark that are the focus of this paper.

A. Real-time Spectrum Management

By design, Spark allows for dynamic sharing of spectrum in real-time through trading. A spectrum seller (such as a primary owner or a spectrum broker) implements its own pricing policies and algorithms in a software entity, which we call the *spectrum server* (Figure 1). On receiving a spectrum request, the spectrum server offers a secondary license to the buyer with a price based on economic and technological factors. To determine this price, it uses a local algorithm, e.g., to maximize spectrum utilization or profit.

Spectrum negotiation protocol (DSCP): The Dynamic Spectrum Configuration Protocol (DSCP) will act as a vehicle for negotiating spectrum trades and managing the spectrum access parameters. Instead of presenting protocol-specific details, we illustrate its potential operation through the following example.

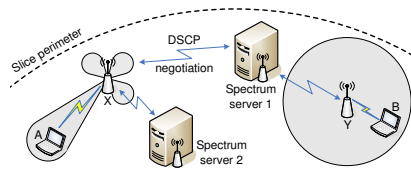


Fig. 2. Illustrating the concept of a power fence.

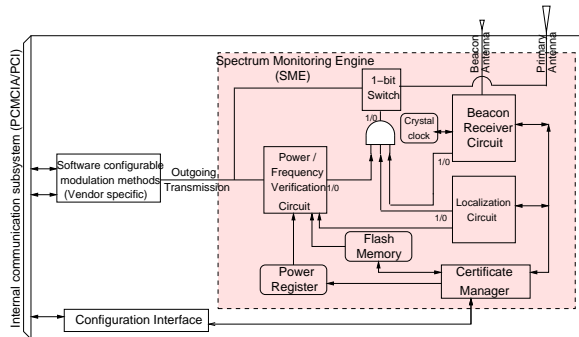


Fig. 3. A Spark radio interface illustrating the block diagram circuit for implementing the SME.

Consider the IEEE Hotmobile 2007 Organizing Committee (OC) as a spectrum buyer. In order to enable mobile, wireless access to conference attendees over the duration of the conference, the OC would deploy a set of wireless Access Points (APs) in and around the conference hotel. To enable wireless communication, it would request appropriate spectrum slices from one or more spectrum servers using our DSCP negotiation protocol. Some possible negotiation parameters include: (i) bandwidth, say a total of 50 MHz of spectrum, (ii) the region of interest, say 500 meters in each direction around the conference hotel, at 3800 East Sunrise Drive in Tucson, AZ, (iii) duration, say between February 26-27, 2007 and the (iii) transmit-power limitations over the different wireless APs. The spectrum server, based on availability will respond with one or more spectrum offers. Once the buyer accepts an offer, the server seals the transaction by issuing a secure slice certificate, whose authenticity is guaranteed using a public-key infrastructure. The slice certificate plays a critical role in enforcing that no buyer's device violates slice operating parameters. In this paper we advocate enabling each Spark-capable radio with a small piece of tamper-proof hardware in which the three important enforcement components — temporal, spatial, and spectral conformance — are implemented for *inline* enforcement of all wireless communication. We call this hardware component, the *Spectrum Monitoring Engine or SME*. Once a slice certificate is issued to a buyer, in our case the OC, this certificate is dynamically loaded into the SME of each wireless communicating device (APs and wireless clients of the conference attendees). Each wireless device is allowed to communicate if and only if (i) a valid slice certificate is available in the corresponding SME and (ii) the enforcement components ascertain that all parameters, as specified in the slice certificate, match the operating conditions. A possible design of slice enforcement is discussed in a longer technical report [11].

Based on our design principle of flexibility, it is important that the slice negotiation parameters do not impose restrictions on the buyer's use of (PHY and MAC layer) communication methods. An implication of this requirement is that the slice should not explicitly limit the transmission power being used in communication, but rather negotiate the the maximum signal power at the perimeter of the slice. We call this limit, the *power fence* (Figure 2(b)). Consider the case where a buyer has purchased a spectrum slice for its two communicating node-pairs ($X - A$ and $Y - B$). Let the power fence limit be -50 dBm, i.e., the maximum signal power measured at or beyond the fence should not exceed -50 dBm. The node-pair, $Y - B$ are quite close to each other and may choose to communicate omni-directionally using 10 mW of transmit power each, and stay within the power fence limit. On the other hand, the node-pair, $X - A$, are far away from each other and may choose to communicate using a higher (50 mW) transmit power. But in order to avoid violating the power fence, they may do so using directional antenna systems. In general, the appropriate choice of transmission power depends on many other communication parameters, e.g., encoding scheme used, data rate desired, hardware available, mobility patterns, etc. Hence, our design of DSCP should allow for such flexibility to the buyer. Our Dyspan 2005 poster [12] describes a possible design and reference implementation of one version of this protocol under specific simplifying assumptions (all nodes are in range of each other, 802.11 is the only communication mechanism).

B. Slice Enforcement

Enforcing a spectrum slice is important for the usability and availability of the spectrum sharing technology. Enforcement methods would ensure that the restrictions specified in a spectrum slice (over the five dimensional resource of space, frequency and time) are adhered to. Enforcement is thus a vital aspect of our architecture as it allows the spectrum leases to take effect by preventing abuse.

One way to implement enforcement is to widely deploy a spectrum sensing infrastructure (as advocated in [4] but for the purpose of opportunistic scanning of available spectrum). This has three important shortcomings. First, it is expensive to deploy a large wide-area sensing infrastructure. Second, a transmit power violation might be hard to detect as the sensing nodes will typically not be co-located with a user's device; thus, attenuation of the transmit signal might lead to inaccuracies. The third and most important reason stems from our design principle of flexibility. It is possible that a buyer's communication signal uses a physical modulation scheme that is unknown to the sensing nodes. Under such circumstances, the sensing infrastructure will merely be able to detect energy in a frequency band *without realizing its source or cause* which is practically useless.

Thus, a practical enforcement method will require a component that *resides within* the user’s communication device.

Spectrum Monitoring Engine (SME): In Spark, we therefore perform *in-line* verification of slice parameters in a tamper-proof hardware module, called the Spectrum Monitoring Engine or SME, in each Spark-enabled radio. Figure 3 illustrates how the SME would integrate within the circuitry of the device’s radio, depicted as a PCI/PCMCIA card. The SME resides *on* the transmit path in the radio, just before the antenna element. This allows it to verify the signal’s properties before transmission and after appropriate modulation has been performed. The SME has a small amount of permanent storage (flash memory) that securely stores slice certificates purchased by the buyer and also has an interface to securely configure and manage the purchased slices.

The SME implements algorithms to detect slice violations, report them and possibly (depending on settings) take actions such as disabling the device for the remainder of the spectrum slice duration, as may be mandated in the slice parameters. Thus, from a functional perspective, the SME acts as a monitor and a sophisticated switch. The switch is turned ‘ON’ if the outgoing signal’s properties fall within acceptable limits; it is turned ‘OFF’ otherwise.

The SME implements three circuits — (i) a power/frequency verification circuit which ensures that the power spectral density of the outgoing signal meets the power fence limit and a specified ‘transmit spectrum mask’, (ii) a beacon receiver circuit that interfaces with an external source of secure time information, and (iii) a localization circuit that helps determine the location of the device with respect to the slice perimeter and feeds the power fence verification process.

Device Certification — Validating the SME: Proper functioning of the SME is critical to enforcement. We mitigate malicious tampering with the SME through a two step process: First, the hardware module is implemented in a tamper-proof casing[13], [14] such that tampering will permanently disable the transmit-receive path and also the Spark radio. Second, proper conformance to such an implementation is ensured through a rigorous certification process performed by the regulatory body. Such a certification process has been very effective with implementing frequency and transmit power restrictions on wireless devices operating in the ISM unlicensed band such as Bluetooth, Wi-Fi, etc. The task of verifying the SME would be more sophisticated but nevertheless practical.

We note that it might be possible for a user to modify the signal once it passes the SME. This is equivalent to building a wireless device that does not implement the SME altogether. These can be handled effectively through a process of certification and heavy fines for violations.

Verification requirements and threat model: The sophistication of verification algorithms required would depend on the specific threat models considered. For example, a fairly aggressive threat model in which all communication between the SME and the Spark spectrum server *pass via the adversary*. Under the assumption that the SME and the spectrum server are trusted and operate correctly, we illustrate one possible solution of slice verification — verification of power, temporal, and spatial parameters — emulated in 802.11 based hardware in a longer technical report [11]. However, many alternate and sophisticated design choices are likely to exist and is an interesting area of further research.

C. Spectrum Fragmentation

Spectrum, in our definition, is a five-dimensional resource — three dimensions of space, one dimension of time, and one dimension of spectrum frequencies. In Spark, spectrum can be allocated in quite fine granularity over all of these dimensions and hence can get fragmented. In particular, prior allocation of smaller and unfavorably positioned requests may preclude the allocation of a subsequent request to be denied even though sufficient spectrum were to be available, but just not in a contiguous manner. This is the classic problem of fragmentation which is commonly seen in memory or disk allocation algorithms.

We believe an interesting formulation of this problem can be based on popular economic mechanisms to address penalties incurred by fragmentation — requests that do not fragment the spectrum will be priced lower than those that will lead to higher fragmentation. A preliminary evaluation of this idea can be found in our technical report [11], and a detailed exploration will be part of our future work.

Spectrum resource as a malleable hypercube: We identify an interesting optimization that takes advantage of the well-known trade-off between transmit power and achieved bit-rates to ‘re-shape’ spectrum requests so as to ‘fit’ better (maximize utilization). From an end-user’s perspective, the achievable bit-rate is the real metric of interest and not the transmit-power or spectrum bandwidth. Note that the same bit-rate can be achieved by different combinations of transmission power and spectrum bandwidth — high power and narrowband of spectrum, or, low power and wideband of spectrum. Using this flexible view, a spectrum request should be considered to be a *malleable hypercube* governed by the relationship between spectrum bandwidth and transmission power (instead of our original static hypercube model). This flexibility brings in interesting new possibilities for future research on spectrum allocation algorithms and provides a new tool to mitigate fragmentation and increase utilization.

III. CONCLUSIONS AND CURRENT STATUS

In the previous sections, we presented high level ideas for a dynamic spectrum sharing and management architecture through secondary markets. We believe that good solutions to some of these problems can play a central role in design of future mobile devices enabling them to take advantage of such markets. While these ideas are promising, as discussed, much work remains in realizing a fully deployable system.

We believe that the domain of dynamic spectrum management in general, and the Spark architecture in particular, opens a whole genre of interesting research questions ranging across basic theory, economic models, networking protocols, hardware design, wireless communication systems, and even spectrum policy design spanning different disciplines — Computer Science, Electrical Engineering, Economics, and even Law. While dynamic spectrum access, facilitated through secondary markets, is widely regarded as a path to the future, we believe that some important architectural components, such as distributed enforcement, fragmentation-aware allocation, and real-time spectrum management, remain open problems. We believe that the research challenges raised in the context of Spark are an important initial step towards an efficient architecture solution for dynamic and real-time spectrum management.

Current implementation status: As of this writing, we are waiting for the first revision of our cognitive radio hardware. When operational, this radio will have the ability to scan frequencies between 400 MHz and 1 GHz, operating in spectrum chunks between 1 and 20 MHz using OFDM modulation and running a cognitive MAC protocol (also under development). We will use this cognitive radio platform to develop the Spark architecture further and research the different challenges enumerated above.

REFERENCES

- [1] William Lehr and Jon Crowcroft, “Managing shared access to a spectrum commons,” *Working Draft, MIT CFP* <http://cfp.mit.edu/resources/>, Sept. 2005.
- [2] “Second report and order: Promoting efficient use of spectrum through elimination of barriers to the development of secondary markets,” Sept. 2004, FCC 04-167.
- [3] J. Mitola, *Software Radio Architecture*, John Wiley & Sons, 2000.
- [4] S. Shankar, N. Cordeiro, and K. Challapali, “Spectrum agile radios: Utilization and sensing architectures,” in *IEEE DySPAN*, 2005.
- [5] D. Raychaudhuri and X. Jing, “A spectrum etiquette protocol for efficient coordination of radio devices in unlicensed bands,” in *IEEE PIMRC*, Sept. 2003.
- [6] P. Bahl, A. Hassan, and J. P. DeVries, “Draft proposal for comment: Etiquette rules and procedures for unlicensed bands,” Jan. 2003, <http://research.microsoft.com/mesh/>.
- [7] J. Peha and S. Panichpapiboon, “Real-time secondary markets for spectrum,” *Elsevier Telecommunications Policy*, 2004.
- [8] M. Buddhikot, P. Kolodzy, S. Miller, K. Ryan, and J. Evans, “Dimsumnet: New directions in wireless networking using coordinated dynamic spectrum access,” in *WoWMoM*, 2005.
- [9] D. Cabric, S. Mishra, D. Willkomm, R. Brodersen, and A. Wolisz, “A cognitive radio approach for usage of virtual unlicensed spectrum,” in *IEEE DySPAN*, 2005.
- [10] T. Weiss and F. Jondral, “Spectrum pooling: An innovative strategy for the enhancement of spectrum efficiency,” *IEEE Comm. Mag.*, 2004.
- [11] S. Banerjee, A. Mishra, V. Brik, V. Shrivastava, and V. Bahl, “Towards an architecture for efficient spectrum slicing,” UW-Madison, Computer Sciences Technical Report, Oct. 2006. Available at www.cs.wisc.edu/~suman/pubs/spark-tr.pdf.
- [12] V. Brik, E. Rozner, S. Banerjee, and P. Bahl, “Dsap: A protocol for coordinated spectrum access,” in *IEEE DySpan (poster)*, 2005.
- [13] Bennett Yee and J. D. Tygar, “Secure coprocessors in electronic commerce applications,” 1995, pp. 155–170.
- [14] “First smart card ic allowing flexible memory partitioning between code and data,” www.emmicroelectronic.com/DetailNews.asp?IdNews=88.