Hierarchical Spectrum Market and the Design of Contracts for Mobile Providers*

Pablo J.C. de Carvalho *campop@rpi.edu*
Aparna Gupta *guptaa@rpi.edu*
Koushik Kar *koushik@ecse.rpi.edu*

In this paper, we propose a hierarchical market structure and contracting mechanisms that could be used by wireless/mobile service providers to trade spectrum and spectrum-derived services with each other. This market advances beyond the current contracting mechanisms between service providers that, due to availability of only long-term contract choices, do not lead to efficient utilization of spectrum resources and service innovation for the customers. We describe the spectrum market structure in terms of the role of the providers and the different spectrum contract types that could be traded in the marketplace. Finally, the role and function of present day mobile virtual networks operators (MVNO) and the mobile network operators (MNO) are discussed from the context of the proposed market structure.

I. Introduction

In recent years, there has been tremendous growth, both in terms of coverage and speed, in wireless voice and data services derived from the use of licensed spectrum. This trend is expected to grow unabated in the near future, or even escalate, as the number of broadband multimedia applications continue to increase. Despite this growth, the current spectrum access and services environment suffers from significant limitations.

Today licensed spectrum services are offered by a few large providers with a national presence, and market competition is realized only at a national level. Large, national providers may not have the motivation to extend to the 'far end' if there is not a critical mass of demand. Extensions into far reaching areas for a small population may not be well justified from economic or management perspectives. The typical provider-customer contracts today are relatively long term, with high penalty for early termination, and only a few standard contract choices that do not vary much across different providers. This does not lead to the best value realization for the customers. From an economic perspective, few contract options and flat pricing structure result in inefficiencies in resource utilization, deficiencies in investments for technology deployments/upgrades, leading to socially sub-optimal allocation, and lost opportunities/revenue for providers. For instance, users of one provider that is overloaded at a certain time are denied spectrum access, while another provider serving the same geography may have excess spectrum at the time. These issues arise from missing market mechanisms to allow dynamic exchange of spectrum services between providers.

Presence of a well structured spectrum market is necessary to ensure that: i) the spectrum available at any time is used efficiently; ii) spectrum allocation is 'fair' to individual subscribers, and is 'socially optimal'; iii) pricing for spectrum usage is done appropriately through free-market competition; and iv) effective management of risk-return tradeoff in spectrum usage is possible in response to demand risks. Principles of economics and finance imply that trading of spectral resources – not only as raw spectrum (bandwidth), but also of the different kinds of service contracts derived from the use of spectrum, is essential for accomplishing the above goals.

I.A. Licensed versus unlicensed spectrum

Since spectrum is a limited resource, its effective management is vitally important to meet the growing demand. The spectrum available for public use can be broadly categorized into the *unlicensed* and *licensed* zones. Any wireless device is allowed to transmit on the unlicensed spectrum, as long as it adheres to certain rules or protocols. To use the licensed part, however, a license must be obtained from appropriate governmental authority, for instance the Federal Communications Commission (FCC) in the United States. Typically, exclusive right to transmit in a certain block of the spectrum over a time period may be obtained for a fee. A large part of the licensed spectrum is leased out to large providers, like AT&T, Verizon Wireless, Sprint, T-Mobile, who provide voice and data/multimedia services to the end-users (customers) over these licensed bands, through 3G/4G cellular technologies.

*This work was supported by the National Science Foundation through awards CNS-0916958 and EARS-1247958.

I.B. Debate on spectrum regulation

While spectrum management in licensed bands has mostly been controlled by responsible government bodies, the need for bringing market based reform in spectrum trading is being increasingly recognized [27, 11, 15, 20]. Both economists and engineers believe that continued spectrum allocation solely by licensing by the FCC would lead to grossly inefficient use of spectrum resources [11, 15, 20]. While more flexible spectrum licensing and trading may be necessary for economic efficiency, the recent financial crisis has demonstrated the importance of appropriate regulatory monitoring and restrictions for the proper functioning of such markets.

Economics of spectrum allocation, auction mechanisms and debate of spectrum commons dominate the literature. Peha [27] states that the three most used methods of spectrum allocation are hearings, lotteries and auctions, with the latter being the most used. Valletti [34] argues that the current centralized model of spectrum management is highly inefficient and should be replaced with decentralized solutions. Peha [29] points to the trade-off between quality of service and efficient use of spectrum. Licensing spectrum guarantees better quality of services, but leads to inefficient use, while unlicensed bands promote sharing at the expense of quality of service. Etkin *et al.* [10] claim that the choice of a legal regime for allocating spectrum is a choice between markets and regulation.

For consistency in terms, Lehr and Crowcroft [22] divide the models of spectrum management in three categories: i) command and control approach, where a regulatory agency acts as a central manager of spectrum; ii) property-rights approach, where a license is assigned to a user for a specific use and may be traded in a secondary market; iii) the spectrum commons (open access) where the right to use the spectrum is shared among user and controlled by protocols that define spectrum management.

Current policy makers allocate a fixed spectrum band for each technology seeking to minimize interference, which leads to unassigned spectrum bands [30]. While the command and control approach is very similar to the current state of affairs of spectrum management, decision-making is decentralized in the other two approaches, with end-users, manufacturers and service providers interacting to determine spectrum allocation [22]. Peha [28] argues that there are a variety of models available and different models can be used for different spectrum bands considering the particularities of the applications. Lehr [22] points out that while all regimes offer some type of property rights, unlicensed does not mean unregulated. Moreover, open access is available to only those in conformity with the unlicensed spectrum access protocol.

Spectrum sharing games and/or pricing issues have also been considered in the literature. Daoud *et al.* [7] consider the problem of a primary license holder who aims to lease its spectrum within a certain geographic region. They argue that the pricing approach that yields better profit charges the buyer per admitted call, in proportion with the interference it generates. Nie and Comaniciu [26] propose a game theoretic framework to analyze the behavior of cognitive radios for distributed adaptive channel allocation. They show that cooperation-based spectrum sharing improves the overall network performance at the expense of an increased overhead required for information exchange. Halldorsson *et al.* [14] argue that spectrum sharing is an inherently distributed problem, with no central authority to coordinate and arbitrate channel allocation. Sahasrabudhe and Kar [31] study bandwidth allocation at base stations of a wireless network as a non-cooperative game between different wireless users, and study the fairness properties of the Nash equilibrium.

Discussions and recommendations for transition to spectrum markets and secondary markets for spectrum trading are ongoing. Valletti [34] points that spectrum trading requires an initial allocation of property rights. Since licences offering similar services have been assigned using very different mechanisms, this disparity has to be addressed before a market for the spectrum is activated. Caicedo and Weiss [2] note that issues such as how to deal with interference in a multi-provider environment, and determining the elements and architectures for feasible implementations of spectrum trading markets, are still unsolved. Kwerel and Williams [20] argue that the FCC should reallocate restricted spectrum to flexible use; conduct auctions of spectrum voluntarily offered by incumbents together with any unassigned spectrum held by the FCC; and provide incumbents with incentives to participate in such 'band restructuring' auctions. Peha [29] argues that before enabling a secondary market, a regulator must make sure that it is able to enforce regulatory constraints. Nevertheless, a clear design of the market structure has yet to emerge, along with contract types and delivery mechanisms for this transition.

I.C. Secondary markets

Primary and secondary markets exist in capital markets and perform a fundamental role for efficient allocation of resources in the financial, as well as the real, economy. Primary market is where new securities originate, both equity and debt, while the secondary market allows those securities to be available for a broader participation of agents in the financial economy, as well as provide liquidity for primary market participants. In the process, funds are allocated where they are most needed and efficiently utilized, and agents are rewarded for their role in resource relocation. In this article, we utilize this thematic similarity between primary-secondary segment of capital markets and the design of primary-secondary market for spectrum to develop the definition of spectrum markets, contract types and agent roles.

There are, however, important differences between spectrum markets and capital markets, which will substantially influence the development of spectrum markets. Contracting in spectrum markets has an associated spatial dimension and necessitates consideration of wireless channel interference in posing and solving the market design question. Contracts for a part of the spectrum do not interfere with each other if they are for regions that are geographically apart; while contracts for the same or neighboring regions using bands close to each other in frequency can result in interference, thereby deteriorating service quality and increasing risk of contract violation. Such spatial multiplexing and interference issues, an exact analog of which does not exist in financial markets, must be accounted for in the design and pricing of contracts in spectrum markets.

I.D. An example of secondary spectrum market players: MVNOs

One consequence and benefit of a secondary spectrum market initiative is the emergence of secondary cellular providers. In today's market setting, this corresponds to what are commonly called Mobile Virtual Network Operators or MVNOs. Traditionally, cellular telephony services are provided through network operation or service provision channels. Offering services based on ownership of network operations requires access to spectrum, which is a scarce resource [36]. In recent years, some national regulatory bodies have allowed mobile operators that do not possess their own frequency spectrum and infrastructure to lease the network facilities from Mobile Network Operators or MNOs [19]. MVNOs are a current ad hoc market solution for limited supply of spectrum, high demand for mobile services and regulatory desire for increased competition.

MVNOs first appeared in the US in May 1996 and in Europe in 1999 [37]. From the customers' point of view, services provided by MVNOs have mostly been indistinguishable from those provided by MNOs [1]; however from a technical perspective, MVNOs rarely have ownership of operating base station infrastructure [36, 21]. An MVNO establishes an agreement with an MNO for use of the latter's spectrum, while avoiding the constraints of radio communication infrastructure [4]. There are different architectures for the MVNO business model in use today, ranging from a full MVNO, with its own infrastructure, to a service re-seller with no infrastructure [36]. An MVNO can choose to participate at any point in this range [23], but from customer perspective it must deliver the same interfaces as an MNO.

The MVNO business model is breaking the traditional mobile value chain in offering retail customers an expanded set of service options [9], and are actually competing with the MNOs for retail demand [17]. This ability and attraction of MVNOs can be severely diminished if MNOs, who effectively control facilities and infrastructure and share available frequencies with MVNOs, exert monopolistic prices for their resources [36]. The MVNO-MNO relationship is a valuable cooperative competition relationship, with the strong potential of creating tele-traffic that may otherwise not exist [35], and new business opportunities with acceptable profit margins [36].

The MVNO business model gained much popularity; however, there are signs of drastic reduction in participation by new players. The main deficiency of the current MVNO business model is that it drives the MVNOs into a corner of the market, where the MNOs have no incentive to explore. In the current state, it can be argued that MVNOs have a systematic competitive disadvantage, as the underlying MNOs persistently hold the cost advantage [32]. Moreover, even though MVNO have enhanced competitiveness in retail mobile services segment, competition at the wholesale level has not yet emerged or matured. In absence of such competition, it is argued that entry by MVNOs and re-sellers of mobile services is likely to be impeded, and therefore some regulatory intervention is necessary [1] for modalities and rates of supplying termination services, and typologies of agreements between MNOs and MVNOs [4].

In this article, we propose a hierarchical spectrum market structure designed to help improve efficiency of spectrum usage, and define the relationship between the different spectrum market players. Section II presents the players and the contracts supported in this structure and section III discusses possible contracting strategies that arise from the interactions between providers

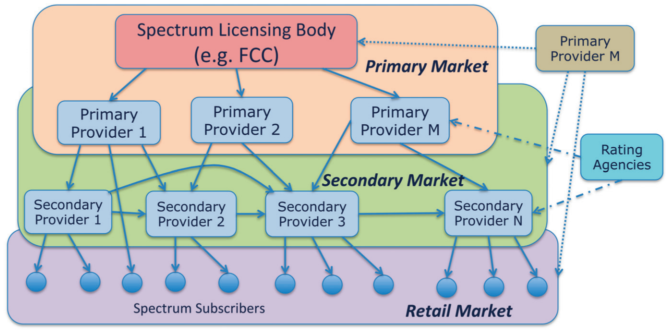

Figure 1: Primary, Secondary, and Retail Spectrum Markets.

and subscribers. The concluding section IV integrates MVNOs into the proposed spectrum market structure.

II. Spectrum Market Architecture

II.A. Different spectrum providers

We propose a hierarchical spectrum market structure (see Figure 1) with the following players: i) primary spectrum (wireless service) providers; ii) secondary spectrum (wireless service) providers; iii) subscribers of wireless services (users).

A licensing body, like the FCC, would classify providers into the two broad categories of primary and secondary through specific registrations with the regulator, where the primary providers obtain licenses for different parts of spectrum from the regulatory body, and possess exclusive usage rights over the bands they have licensed, while secondary providers acquire spectrum contracts from the primaries. An additional classification of spectrum providers is by their geographical presence, such as global or regional providers.

An interesting and critically important component of our architecture is that secondary providers can also trade bandwidth contracts with each other, which is facilitated in the secondary market. Primary and secondary providers sell service contracts to subscribers in the retail market, using their respective retained or acquired spectrum resource. We note that our notion of 'primary' and 'secondary' spectrum providers must be distinguished from similar terms often associated with users themselves (subscribers in our case) [28].

We first discuss the design and pricing of spectrum contracts between primary and secondary providers. Primaries can sell or lease the parts of the bands they are not using in the secondary market. The primaries determine the types of contracts they will choose to offer in the secondary market, as well as the service, financial and duration terms of the contracts. The secondaries select the primaries they would trade with and the type of contracts they would obtain based on their customer-base and geographical presence, prices and terms of the contracts offered.

The primary providers maximize their revenue both from the spectrum contracts they sell, as well as from providing services to their own subscribers. Therefore, they must take into consideration the spectrum demand among their own subscribers and the demand variability. On the other hand, secondary providers will seek to minimize the cost of purchasing contracts (from other providers) to provide services for their subscribers, as well as maximize the revenue from selling contracts they acquired previously to other secondary providers. Coping with demand risk justifies the existence of *guaranteed-bandwidth* and *opportunistic-access* contracts. Guaranteed contracts on a channel provide guaranteed access to the channel bandwidth, possibly at a higher per unit price, while opportunistic access contract offers stochastic access upon availability, possibly at a discounted price. A key problem for a provider participating in the secondary market is to determine the amount of guaranteed and opportunistic contract units to buy in order to satisfy its stochastic user demand.

The role played by regional providers, and their regional contracts and services, is a mechanism for localized spectrum access and services. The regional providers, who obtain secondary (sub) licenses, or buy guaranteed and/or opportunistic spectrum contracts from the primary (national) providers, use the spectrum to serve customers in a locality, say a city, a couple of counties, or a state. Like electricity markets, trading of spectrum (and contracts derived from it) would take place in regional wholesale spectrum markets.

Despite the analogy with electricity (energy) markets, and that of other 'commodities' and commodity-derived services for which wholesale markets exist, it should be noted that there is a major difference between how spectrum is used as compared to other commodities and services. Spectrum users are increasingly more mobile, and demanding access to spectrum while on the move. Therefore, when a user is outside the coverage area of its regional provider (traveling out-of-state, say), provisions must exist such that there is no disruption in service and the user can avail of reasonable coverage and competitive prices. A 'soft roaming' agreement with the primary provider from whom the user's regional provider buys spectrum contracts, can provide such flexibility.

Existence of regional providers would imply some spatial sensitivity to spectrum service prices. A user who accesses spectrum in a region that is more congested (or equivalently, under-provided, like some densely populated cities) may have to pay more (subject to some regulations, of course) for spectrum access, thereby prompting more competition in that space and likely more investment. In other words, the presence of regional providers and regional spectrum markets would imply that the supply-demand matching game is played at each region individually, rather than at the national scale. Similarly, through more user-adapted flexible services managed by the regional provider, a user who resides in a densely populated city but uses spectrum mostly in low demand areas (being frequently on travel, say) would pay lower fees for spectrum services than an all-the-time city user. Thus a spectrum user would pay in accordance with the overall cost it imposes on the system.

In this sense, the relationship between primary and secondary providers gains complexity with the regional aspect. The core question here is what are the economic benefits/incentives for primary providers to sell bulk spectrum contracts/access rights to regional providers at wholesale prices. In other words, what are the economic reasons behind the profitability of both businesses (those of the larger and the smaller providers) if they co-exist. The answer to this question, depends on whether the regional provider has base stations (infrastructure) of its own, or has to rely on the base stations of the primary provider to provide direct service to the user. For the latter type of regional providers, we believe that the answer lies in benefits to the primary providers in terms of their modified risk-return tradeoff by letting off some of the risk. Specifically, through the wholesale selling, the larger provider is able to pass on some of its demand risk to the regional provider, in return for a slightly reduced (but more steady) revenue/profit. For the former type of regional providers who have base stations of their own, the use of their base stations can, in addition, provide better reachability/coverage and service to users, and in the process indirectly improve the primary provider's revenue by allowing access to a larger user (revenue) pool.

II.B. Contract types

We divide the possible contract types in this market into two categories: *spot contracts* and *derivative contracts*. A spot transaction refers to a contract which is settled immediately, say on the same date or hour. In general, it refers to a contract which is acted upon and utilized immediately by the transacting parties to meet the current demand. A derivative contract, on the other hand, is a contingent claim and derives its value from future realizations of an underlying resource, in this case, spectrum bandwidth. Among other reasons, they may be used for hedging the risk of fluctuations in the demand for bandwidth and price of spot contracts in the future.

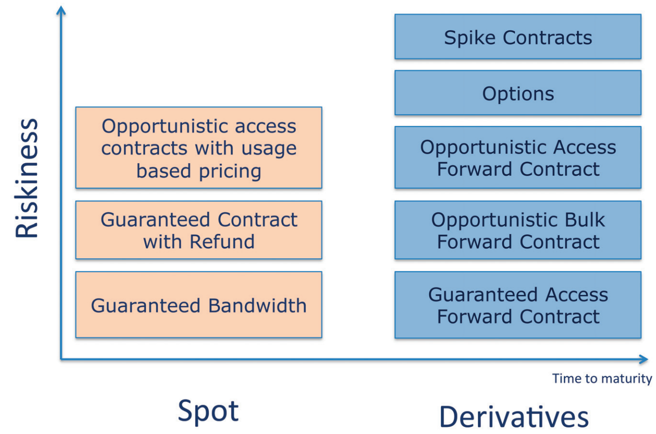

Figure 2: Different Contracts Types on Spectrum.

II.B.1. Spot contracts

A variety of contracts can be created for the participation of providers in the secondary market to support an efficient usage of bandwidth. This would facilitate the providers' efficient use of resources, while profitably satisfying their respective customer demands. We identify three types of contracts that would lend to efficient spectrum usage, extending the notion of guaranteed and opportunistic-access contracts defined in Section II.A: **Type I:** Guaranteed-bandwidth contracts; **Type II:** Guaranteed contracts with refunds; and **Type III:** Opportunistic-access contracts with usage based pricing.

Under the guaranteed bandwidth contracts, or Type I contracts, a guaranteed amount of bandwidth for a specified duration of time must be made available for a price agreed upon irrespective of whether the bandwidth is used or not. Guaranteed contracts with refunds, what we label as Type II contracts, are similar to Type I contracts except that under these contracts either the primaries (sellers) or the secondaries (buyers) or both, would have the option of terminating the contract (before the agreed upon end-of-term) by retracting/releasing the bandwidth in lieu of incurring refund penalties, which are agreed upon upfront. The penalty will depend upon how close the retract/release decision epoch is to the start or the end of the term of the contract. The bandwidth freed up from a prematurely terminated contract can be re-sold by the primaries in the secondary market. Under opportunistic-access contracts, or Type III contracts, a secondary provider enrolls with some selected primaries after paying some basic membership fees, thus acquiring the privilege of obtaining bandwidth from the primaries on an opportunistic basis, paying for the service only per its usage if and when it becomes available. We now focus on the research questions underlying the definition and pricing of the three contract types.

Guaranteed-bandwidth contracts are the basic contract type for delivery of deterministic service. If we look for capital markets equivalents of these contracts, the deterministic feature of Type I contracts, along with their fixed time-duration, makes these contracts resemble risk-free bond instruments in capital markets. Therefore, the definition and pricing of these contracts will take advantage of this similarity. Under this contract, the primary provider will make a fixed level of bandwidth available to the secondary for immediate usage for a fixed period of time for a fixed fee, received either as a lump-sum or as a periodic payment through the duration of the contract. The spot prices of the guaranteed-bandwidth contracts is fundamental to the development of all other contract types.

A guaranteed-bandwidth contract with the possibility of retracting or releasing the bandwidth during the duration of the contract gives the flexibility for more profitable use of the resources, should such opportunity arise. As Figure 2 points out, these contracts are designed for variability of the available bandwidth around its mean, $\bar{A}_{[0,T]}$. When the primary (secondary, respectively) provider retracts (releases, respectively) the bandwidth, it must refund part of the contract money, and pay a penalty for recalling the contracted bandwidth (receive the contract price minus a refund penalty, respectively).

For the seller an option to retract bandwidth associated with a guaranteed contract will provide it the flexibility of reallocating the retracted bandwidth to its user base or in guaranteed (Type I) contracts for a higher profit. This would also allow it to offer higher number of such contracts than that permitted by its available bandwidth if it estimates that it is unlikely that all the secondaries that purchase the contracts will use the entire bandwidths offered (the sale of air tickets provides an analogy here: airline companies routinely sell more tickets than the number of seats). For the secondary provider, the release option allows it to refund the bandwidth it does not need, in cases when it ends up overestimating its traffic demand.

Type II contracts with retract (release, respectively) options resemble callable bonds (puttable bonds, respectively) in capital markets which have lower (higher, respectively) risk for the seller, and therefore must be priced lower (higher, respectively) than plain-vanilla bonds with comparable terms (Type I contracts in our case). The pricing approaches employed for callable/puttable bonds (Type II contracts) as an extension to pricing vanilla bonds (Type I contracts), depending on the terms of the retract-release features, will need to be utilized for pricing this class of contracts.

As Figure 2 points out, opportunistic-access contracts with usage based pricing are designed for utilization of available bandwidth at the highest percentiles, as and when it becomes available. Clearly, bandwidth is available at this level unpredictably and sporadically, therefore no guarantees can be provided. Instead, the terms of the contract will be defined by 'if-then' clauses. For instance, if bandwidth becomes available in a duration of time, it can be used by the secondary for a pre-determined usage-fee per unit time, and the secondary may continue to use it until the time it stays available.

In capital markets, this contract resembles a barrier-type bond option, providing the option to buy the bond at a certain price during a specified period of time, conditional on the bond price or another reference quantity (here the available bandwidth, A_t) reaching a certain target level (here a pre-chosen high percentile of the available bandwidth distribution, a). This privilege is acquired for an upfront premium, or as we labeled it earlier - a membership fee. This membership fee may be determined using option-pricing framework based on the definition of the contract - its maturity, set availability threshold a, price of guaranteed contracts and the usage-fee per unit time set in the opportunistic-access contract.

II.B.2. Spectrum derivatives

The contracts developed this far are designed for an immediate transaction of bandwidth, along with additional clauses, hence we call them spot contracts. We now turn to the construction of derivative contracts. In order to manage risks in spectrum markets arising due to variability in spectrum prices and demand uncertainty, buyers and sellers can benefit from utilizing derivative contracts. The optionality embedded in spot contracts, specifically Type II and Type III contracts, is valuable, however dedicated derivative instruments for managing future risks have been found useful in other infrastructural domains. Motivated by the electricity and the weather markets, where the stochasticity of the underlying is anticipated to be similar to that in spot spectrum prices, we propose designing several derivative contracts: Forward contract with (i) Guaranteed Access, (ii) Opportunistic Access, or (iii) Opportunistic Bulk Access; Option Contracts; and, Spike Contract.

A forward contract is a standard derivative contract allowing to buy (or sell) an asset at a certain price at a specified future date. In the context of spectrum markets, a forward contract may be defined corresponding to all the three spot contracts discussed earlier. For instance, a guaranteed access forward contract will allow the contract buyer to take delivery, from the seller, of a specific bandwidth access at a predetermined price on a future delivery date. The contract allows its buyer to transfer the future price uncertainty of spot spectrum licenses to the seller of the contract, by agreeing upon a price at contract initiation. The seller reduces its demand uncertainty for bandwidth at the delivery hour, although it must ensure delivery of the bandwidth at contract maturity under a guaranteed access forward contract. In order to do so, the seller must make provisions to secure availability of spectrum at forward contract maturity. This may be done through reserving bandwidth for this purpose, contracting with other spectrum holders, or by buying spot licenses at the time of maturity depending on the need.

An opportunistic access forward contract can also be similarly defined, and allows buying/selling of an opportunistic access contract starting at a specified future date at a predetermined price. An extension of this is an opportunistic bulk forward contract that allows the buyer to acquire a fraction of the remaining bulk capacity from the seller at a predetermined price. These contracts would allow the buyer providers to plan in advance towards offering services to their customer base based on cheaply acquired opportunistic bandwidth.

An option contract on spectrum bandwidth provides the right, but not the obligation, to buy or sell spectrum licenses (services) in the future at a predefined price, called the exercise price. A call option grants the buyer (of the option) the privilege of buying spectrum bandwidth at a given price, while a put option gives the buyer (of the option) the right to sell bandwidth at a certain price. This right, but not an obligation, to purchase/sell bandwidth at specified price in future is acquired by paying a premium upfront, unlike for forward contracts.

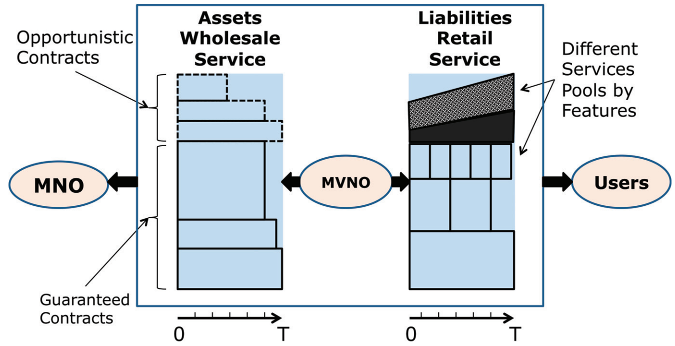

Figure 3: Asset-liability management by secondary providers (MVNOs).

Option contracts thus additionally help limit the downside price risk, and may be defined for all the three types (Type I, II and III) of spot contracts discussed earlier.

In wireless demand traffic, the peak demand-hour periods are very important. During the peak hours, sudden large increases, or spikes, in traffic can translate to sharp rises in spot prices of spectrum licenses. Since the sharp rises in spectrum demand exposes providers to the risk of congestion and high spectrum prices, we propose the design of spike derivative contracts that pay off if there are unexpected spikes in spot spectrum prices during specified periods of time.

III. Spectrum Contracting Strategies

We now focus on the challenges arising due to the interactions between the providers and subscribers in the spectrum retail market. Providers need to efficiently utilize the spectrum portfolio they have acquired so as to minimize the operational costs subject to delivering the desired quality of service (QoS) to the subscribers, and thereby attract more subscribers and maximize their profits. Spectrum utilization, or scheduling, involves determining which subscribers will be served at any given time, the corresponding service rates, infrastructure (e.g. base station, access points, mesh points) and time-frequency allocations to the subscribers. Providers need to price their offered wireless services to maximize their profit. The pricing will clearly depend on, and in turn determine, provider spectrum portfolio, subscriber demand for bandwidth, and therefore the scheduling of the subscribers.

III.A. Primary provider

A primary provider can enhance its revenue and reduce its risk by spatially multiplexing different kinds

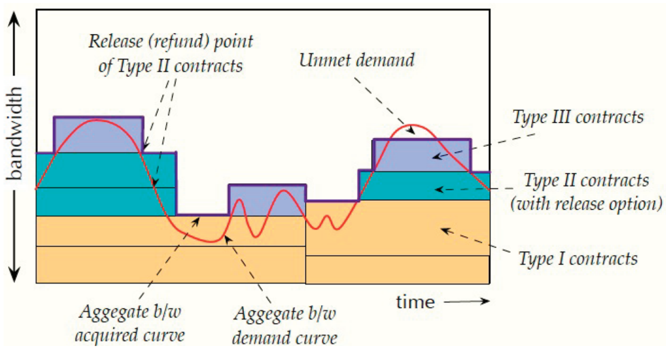

Figure 4: Contracting strategy of a secondary provider (buyer perspective).

of contracts in different locations. For example, a primary can sell a large number of Type I contracts in a region, and mostly Type II and III contracts in the same band in the same or neighboring regions. If all the secondaries that purchase the Type I contracts simultaneously transmit in the bands (i.e., use the bandwidth promised), the primary can accommodate the resulting high demand by retracting the Type II contracts and not offering bandwidth for the outstanding Type III contracts. This would avoid the interference for the Type I contracts and thereby satisfy the bandwidth promised in the Type I contracts already sold. Again, if the sale of a particular contract in a certain region (say region R), prevents the sale of some other lucrative contracts in other (neighboring) regions, owing to interference constraints, then contracts in region R will have high prices.

Design and pricing of contracts for maximization of revenue and minimization of risks of contract violation through exploitation of interference constraints will require an effective combination of channel allocation and contract pricing strategies. Both of the above issues have been studied extensively but separately in the respective communities (wireless networking and financial economics), but their aggregation has not been attempted so far, which would require future research. Once the three contract types are developed and introduced in the secondary markets with a pricing scheme constructed for them, a primary provider will need to determine the mix of contract types it should offer.

III.B. Secondary and regional providers

The spectrum-portfolio optimization question - that of optimally buying contracts (Types I, II or III) from different primary or other secondary providers in the secondary market - is of core importance to secondary providers. The secondary provider must periodically adapt the portfolio based on the changing market conditions and user demands. The spectrum portfolio of a secondary provider consists of contracts that it buys from other primary and secondary contract sellers (its assets), as well as contracts it sells to other secondary providers and subscribers (its liabilities). Both the assets and the liabilities, to receive and deliver bandwidth respectively, of a secondary have their inherent risks, which the secondary must manage for a commensurate reward or return. The return of a portfolio can be formally defined as the expected monetary profit that the secondary provider makes by matching the cash flows from assets with liabilities in the portfolio. Risk of the portfolio refers to the possibility of making a large monetary loss, which includes costs incurred due to contractual violations. The portfolio management goal would be to attain efficiency in risk-return tradeoffs in a Pareto-sense. A stronger and more desirable objective is to maximize the portfolio return subject to upper-bound constraints on the risk of the portfolio. Noting the risk-return trade-off objective and the dynamic nature of the spectrum markets, we discuss questions that are critical to describing the risk-profile of a secondary provider.

The first question involves determining the asset-liability spectrum portfolio strategy a secondary provider may adopt. Portfolio optimization is a fundamental, well studied question in traditional financial literature, and the version most applicable here is matching assets with liabilities while targeting to achieve optimal risk-return profile for the portfolio. These types of problems most frequently arise in financial domains, such as, in banking, pension funds and insurance, where a key service is delivered to the customers (savings accounts and insurance products, for example), and is supported by investing in appropriate assets [6]. Both assets and liabilities have associated cash-flows, term-structure of cash flows, and risks underlying the cash-flows. For a secondary provider, the primary liability is towards its customer-base or subscribers, and assets are the contract types it acquires to meet the liabilities (see Figure 3).

The total nominal bandwidth acquired by the provider through Type I, Type II and Type III contracts in Figure 4 is the total bandwidth contracted. This does not translate to actual bandwidth received since: i) there are optionalities embedded in Type II and III contracts, and ii) the actual transmission rates fluctuate due to fading and interference. The secondary provider must also determine its optimal exercise policy for its Type II and III contracts, which is the subsequent decision making required for these contracts. For Type II contracts, the buyer has to decide when to exercise the release option, and how much bandwidth to return (release) in such a case. For Type III contracts, the corresponding decision is when and how much bandwidth the secondary should seek from the seller.

With changing market and demand conditions, a secondary must re-balance its portfolio of assets to continue to support the liabilities. A strategic (long-term) solution is obtained by solving a portfolio optimization question for the secondary provider's business model under steady-state assumptions on the risk factors, such as subscriber demand and available contract types. Once a long-term strategy is available, tactical adjustments to it for local transience in the risk factors must be considered. This can be viewed as the portfolio re-balancing question, which refers to how a secondary provider dynamically adapts its portfolio with temporarily or permanently changing user demand patterns and transmission characteristics of the leased spectrum bands. An increase (or expected increase in near future) in user demand may necessitate buying new bandwidth contracts for higher return or lower risk of subscriber dissatisfaction. Similarly, a decrease in the user demand should lead to selling off some of the contracts, or replacing some guaranteed contracts (Type I) with guaranteed contracts with release options (Type II), to improve return or reduce risk of monetary loss.

Besides meeting its customers' demand, when a bandwidth surplus is generated (say through excessive purchase of guaranteed contracts, which it cannot return), a secondary provider may choose to offer some contracts in the secondary market. The provider may find that other providers, secondary or primary, may be just as desirable customers for bandwidth should the provider find a surplus of bandwidth in its portfolio. This leads to a question that is closely associated with the original asset liability matching portfolio optimization question of how should a secondary provider create (compose) its own contract types to sell its surplus bandwidth. Note that a secondary provider can potentially sell Type I contracts while it has only bought (from primary or other secondary providers) only Type II and Type III contracts; by doing so, the secondary may expect high return while at the same time taking a high risk of not delivering the contract, which may be associated with high penalties. If the user demand and the transmission characteristics are steady, then such contract composition makes sense. On the other hand, it is natural to sell Type II contracts with retract option if the secondary provider expects some surplus bandwidth to materialize; this carries almost no risk, but potentially lower returns. Thus composition of these contracts is a complex task, and depends on the desired risk-return profile.

The questions faced by regional providers are similar to those of secondary providers discussed above, except that they must be considered only in the context of the region that the provider operates in. Some regional providers will make infrastructural investments, and thus will acquire physical assets to support their services. These will need to be funded using initial capital, which is incorporated in the asset-liability assessment of the provider.

III.C. Optimal spectrum contracting

In our recent work [24], we addressed the spectrum portfolio optimization for Type I (guaranteed) and III (opportunistic) contracts over a single as well as multiple regions. The problem is formulated for minimizing the cost of acquiring spectrum portfolio subject to constraints on bandwidth shortage relative to user demand. We applied two forms of bandwidth shortage constraints, namely, the demand satisfaction rate constraint and the demand satisfaction probability constraint. The problem was shown to be convex for all density functions under demand satisfaction rate constraint, while under demand satisfaction probability constraint the problem is not convex in general. We derived some sufficient conditions for convexity in the latter case. The spectrum portfolio optimization problems are therefore solvable efficiently using standard convex optimization techniques in most cases.

In Kasbekar *et al.* [18], we addressed the question of optimal dynamic trading of spectrum contracts, for primary as well as secondary providers. In this formulation, Type I (guaranteed) and Type III (opportunistic) contracts are available to the providers to help attain desired flexibilities and trade-offs for service quality, spectrum usage efficiency and pricing. The providers create and maintain a portfolio composed of an appropriate mix of these two types of contracts. The optimal contract trading problem relates to the dynamic adjustments in the spectrum contract portfolio needed to maximize return (minimize cost) subject to meeting the bandwidth demands of its own subscribers. We formulate this question as a stochastic dynamic programming problem, and compare it with a static portfolio optimization strategy. While a dynamically rebalanced portfolio takes into consideration the current market prices of contracts and the current subscriber demand, static portfolio optimization is based on steady-state statistics of price and subscriber demand processes.

IV. Integrating MVNO-MNO into the Hierarchical Market Structure

The proposed hierarchical spectrum market will directly impact the present-day MVNO-MNO relationship. In this section, we contrast spectrum markets with financial markets, and discuss the role of regulators for enabling the proposed spectrum markets and possible consequences for the MVNO-MNO relationship.

IV.A. Spectrum market design issues

The hierarchical market structure loosely parallels bandwidth contracting in the Internet and cellular networks, through nationwide, regional and local providers. It is also motivated by and closely resembles the structure of financial markets, with primary and secondary markets for securities issuance. The different kinds of bandwidth contracts are analogous, in essence, to well known financial instruments. As discussed earlier, guaranteed-bandwidth contracts can be likened to bonds, and other types of contracts can be seen as variants of bond and bond derivatives. These similarities encourage using ideas and tools from the wealth of literature in financial economics and financial engineering to address questions for the spectrum market.

Spectrum markets, however, have significant differences with financial (both stock, bond and commodity) markets. For example, bandwidth demand dynamics can be expected to be very different from the dynamics of financial asset prices. Moreover, spectrum is limited and unlike stocks or bonds, cannot always be bought or sold in large amounts as necessary, even for a very high price. Also, unlike typical commodities traded in exchanges, spectrum has a temporal dimension to it, and is wasted if not used instantaneously. Finally, wireless spectrum service contracts are associated with unique challenges in dynamic scheduling due to constraints like interference, power, signal-to-noise-ratio etc. Thus, results from financial markets cannot be directly applied to spectrum markets, making the questions on the design and operation of these markets both challenging and interesting.

Transparent pricing, ideal market design and contracting mechanisms have been actively studied and debated in other areas, e.g., electricity market [8, 16, 25], the telecom bandwidth market [12], and QoS delivery in the packet-switching Internet [13, 33]. In all the three contexts, pricing of contracts and strategies for risk management are important themes [8, 5, 25]. QoS delivery in the packet-switching Internet has an inherently stochastic nature [33, 13]. It was argued that lack of mechanisms for managing the risks in QoS delivery has contributed to the failure of QoS assured services, despite active research and development of standards [33]. Real options techniques have been used for risk management of telecommunication network services [3]. Related developments in these fields could be utilized in the spectrum markets context.

IV.B. Regulation in spectrum markets

A regulatory body must formulate rules and impose necessary restrictions on spectrum contracting counterparties to ensure integrity of contracts, efficiency and fairness in the spectrum market, and monitor enforcement and violation of the rules. Closely related to the regulatory role is the possible contribution of rating agencies, that rate different providers and contracts sold by them based on their monitoring, subscriber complaints, and other quantitative and qualitative metrics. The ratings associated with a provider would reflect the risk of contractual violations by the provider, and could potentially impact the price of contracts sold by the provider.

There should be incentives in place for the primary providers to sell their spectrum rights to those who may use it more efficiently. A regulator must additionally be engaged in detecting market failures, which must be addressed with appropriate intervention. For instance, regulatory intervention for healthy MVNO-MNO relationship is required so that MNOs don't deliberately manipulate the market or exercise excess market power.

Using MVNOs to empirically examine whether a regulatory policy promoting competition undermines the dynamic efficiency of the sector in terms of investment, [19] finds that granting market access to MVNOs may have unwanted consequences. In particular, it may lead to smaller infrastructure investment by incumbent MNOs, which may lead to undesirable results in the long-term, such as as delayed deployment of advanced networks, poor service quality, and slow technological advances. Banerjee and Dippon [1] propose that regulatory policy would be better directed at creating and helping to maintain reasonably competitive wholesale access markets by ensuring efficient retail competition, preventing the exercise of individual or collective market power. Such policy would allow spectrum trading to enable more efficient access to, and use of, spectrum resources, and apply spectrum caps to prevent anti-competitive MNO concentrations. In summary, [1] encourages a voluntary formation of MVNO-MNO relationship.

We believe that our hierarchical market structure holds close resemblance to the present-day MNO-MVNO relationship. MNOs are the primary providers that hold spectrum license, and may sell or lease idle parts of the spectrum in an attempt to get a higher yield for their investment on spectrum licenses. MVNOs are secondary or regional providers who aim to enter markets in which they may have a competitive advantage, but have no direct access to spectrum licenses. In line with Banerjee and Dippon [1], we believe that the above issues/limitations can be addressed by enabling local spectrum markets and competition, which would include participation of smaller, regional providers in providing, enhancing, or sharing wireless spectrum and services. In particular, we envision that localized spectrum access and services can be done through regional providers who obtain secondary (sub) licenses, or buy guaranteed and/or opportunistic (risky) spectrum contracts from the primary (national) providers, and use that spectrum (contracts) to serve customers in a region.

V. Conclusion

One of the key contributors to the rapidly growing demand for spectrum is the growth in mobile data services. Supply of spectrum is limited, and a market structure that allows flexible trading of spectrum contracts between different providers is necessary for efficient spectrum utilization. The present-day MVNO-MNO relationship fits in the hierarchical architecture of spectrum market proposed here. In this context, MVNOs can be seen as secondary providers without direct access to spectrum licenses, but that could use spectrum markets to buy this access.The proposed spectrum markets would enhance this relationship, by ensuring efficient pricing for spectrum usage through competition. They would also allow for dynamic rebalancing of spectrum portfolios, which in turn would enable effective management of risk-return trade-offs for both primary/national and secondary/regional spectrum providers.

References

[1] A. Banerjee and C. M. Dippon. Voluntary relationships among mobile network operators and mobile virtual network operators: An economic explanation. *Information Economics and Policy*, 21(1):72–84, 2009.

[2] C.E. Caicedo and M.B.H. Weiss. Spectrum Trading: An Analysis of Implementation Issues. In *Proceedings of New Frontiers in Dynamic Spectrum Access Networks, DySPAN 2007*, pages 579–584, 2007.

[3] G. Cheliotis. *Structure and Dynamics of Bandwidth Markets*. PhD thesis, National Technical University of Athens, November 2001.

[4] L. Cricelli, M. Grimaldi, and N. Levialdi Ghiron. The impact of regulating mobile termination rates and MNO–MVNO relationships on retail prices. *Telecommunications Policy*, 36(1):1–12, 2012.

[5] J. Crosby. A multi-factor jump-diffusion model for commodities. *Quantitative Finance*, 8(2):181–200, 2008.

[6] M. Crouhy, D. Galai, and R. Mark. *Risk Management*. McGraw-Hill, New York, NY, 2001.

[7] A.A. Daoud, M. Alanyali, and D. Starobinski. Secondary Pricing of Spectrum in Cellular CDMA Networks. In *Proceedings of New Frontiers in Dynamic Spectrum Access Networks, DySPAN 2007*, pages 535–542, 2007.

[8] S.J. Deng and S.S. Oren. Electricity derivatives and risk management. *Energy*, 31(6):940–53, 2006.

[9] R. Dewenter and J. Haucap. Incentives to Licence Virtual Mobile Network Operators (MVNOs). *Access pricing: theory and practice*, pages 303–323, 2007.

[10] R Etkin, A Parekh, and D Tse. Spectrum sharing for unlicensed bands. *IEEE Journal on Selected Areas in Communications*, 25(3):517–528, Apr 2007.

[11] G. R. Faulhaber and D. Farber. Spectrum Management: Property Rights, Markets, and the Commons. *Telecommunications Policy Research Conference Proceedings*, 2003.

[12] T. Gros, R. Mayfield, B. K. Baulch, and S. S. Chui. The Telecoms Trading Revolution. *Risk Waters Group*, 2000.

[13] A. Gupta, S. Kalyanaraman, and L. Zhang. Pricing of risk for loss guaranteed intra-domain Internet service contracts. *Computer Networks*, 50:2787 – 2804, 2006.

[14] M.M. Halldorsson, J. Y. Halpern, L. Li, and V.S. Mirrokni. On spectrum sharing games. In *Proceedings of ACM Symposium on Principles of Distributed Computing*, pages 107–114, 2004.

[15] T. Hazlett. The Wireless Craze, The Unlimited Bandwidth Myth, The Spectrum Auction Faux Pas, and the Punchline to Ronald Coase's "Big Joke": An Essay on Airwave Allocation Policy. *Harvard Journal of Law and Technology*, 14(2):337–451, 2001.

[16] P. Joskow and J. Tirole. Reliability and competitive electricity markets. *Rand Journal of Economics*, 38(1):60–84, 2007.

[17] P. Kalmus and L. Wiethaus. On the competitive effects of mobile virtual network operators. *Telecommunications Policy*, 34(5):262–269, 2010.

[18] G. S. Kasbekar, S. Sarkar, K. Kar, P. K. Muthusamy, and A. Gupta. Dynamic contract trading in spectrum markets. In *Proceedings of the 48th Annual Allerton Conference on Communication, Control, and Computing*, pages 791–799, 2010.

[19] J. Kim, Y. Kim, N. Gaston, R. Lestage, Y. Kim, and D. Flacher. Access regulation and infrastructure investment in the mobile telecommunications industry. *Telecommunications Policy*, 35(11):907–919, 2011.

[20] E. Kwerel and J. Williams. A Proposal for a Rapid Transition to Market Allocation of Spectrum, OPP Working Paper No. 38. *Office of Plans and Policy, Federal Communications Commission*, 2002.

[21] H. Le Cadre and M. Bouhtou. Modelling mno and mvno's dynamic interconnection relations: is cooperative content investment profitable for both providers? *Telecommunication Systems*, 51(2-3):193–217, 2012.

[22] W. Lehr and J. Crowcroft. Managing shared access to a spectrum commons. In *Proceedings of New Frontiers in Dynamic Spectrum Access Networks, DySPAN 2005*, pages 420–444, 2005.

[23] A. Lillehagen, L. Armyr, T. Hauger, V. Masdal, and K.A. Skow. An Analysis of the MVNO Business Model. *Telektronikk*, 97(4):7–14, 2001.

[24] P. K. Muthuswamy, K. Kar, A. Gupta, S. Sarkar, and G. Kasbekar. Portfolio Optimization in Secondary Spectrum Markets. In *Proceedings of the 9th Intl. Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks (WiOpt '11)*, Princeton, NJ, USA, May 2011.

[25] M. Nakamura, T. Nakashima, and T. Niimura. Electricity markets volatility: estimates, regularities and risk management applications. *Energy Policy*, 34(14):1736–49, 2006.

[26] N. Nie and C. Comaniciu. Adaptive Channel Allocation Spectrum Etiquette for Cognitive Radio Networks. *Mobile Networks and Applications*, 11(6):779–797, Dec 2006.

[27] J. Peha. Spectrum management policy options. *IEEE Commun. Surv. Tuts.*, 1(1):2–8, 1998.

[28] J. M. Peha. Approaches to spectrum sharing. *IEEE Communications Magazine*, 43(2):10–12, 2005.

[29] J. M. Peha and S. Panichpapiboon. Real-time secondary markets for spectrum. *Telecommunications Policy*, 28(7-8):603–618, 2004.

[30] C. Peng, H. Zheng, and B.Y. Zhao. Utilization and fairness in spectrum assignment for opportunistic spectrum access. *Mobile Networks and Applications*, 11(4):555–576, Aug 2006.

[31] A. Sahasrabudhe and K. Kar. Bandwidth allocation Games under Budget and Access Constraints. In *Proceedings of the 42nd Annual Conference on Information Sciences and Systems (CISS)*, Princeton, NJ, March 2008.

[32] D. Shin. MVNO services: policy implications for promoting MVNO diffusion. *Telecommunications Policy*, 34(10):616–632, 2010.

[33] B. Teitelbaum and S. Shalunov. What QoS research hasn't understood about risk. *Proceedings of the ACM SIGCOMM 2003 Workshops*, pages 148–150, Aug 2003.

[34] T.M. Valletti. Spectrum trading. *Telecommunications Policy*, 25:655–670, 2001.

[35] D. Van Thanh. The Mobile Virtual Network Operator concept: Truth and Myths. *Telektronikk*, 97(4):3–6, 2001.

[36] D. Varoutas, D. Katsianis, Th. Sphicopoulos, K. Stordahl, and I. Welling. On the Economics of 3G Mobile Virtual Network Operators (MVNOs). *Wireless Personal Communications*, 36(2):129–142, January 2006.

[37] H. Ware and C. M. Dippon. Wholesale unbundling and intermodal competition. *Telecommunications Policy*, 34(1):54–64, 2010.