Spectrum Without Bounds, Networks Without Borders

Linda Doyle, Jacek Kibilda, Timothy K. Forde and Luiz DaSilva
CTVR, The Telecommunications Research Centre,
Trinity College, University of Dublin, Ireland

Abstract

The purpose of this paper is to present a vision for future mobile and wireless networks. The vision, which we call Networks without Borders (NwoB), is based on a market-place of virtual network operators which construct networks from a pool of shared resources (e.g. basestations, spectrum, core network components, cloud resources, processing capabilities etc.). The resources will be sourced from traditional industry players as well as crowdsourced from individuals. The paper describes this approach from a value-chain perspective. The proposed value-chain is substantially different from the value-chain models that are currently used to illustrate mobile and wireless networks. The economic imperatives and innovation drivers for this approach are discussed. Early work showing the promise of this vision is presented. This work focuses on diverse examples which advocate the removal of traditional and historical restrictions on spectrum and infrastructure and move towards more dynamic use of shared resources. In the first example we look at how frequency-division-duplexing (FDD) and time-division-duplexing (TDD) restrictions on spectrum usage can be relaxed; we remove the borders between TDD and FDD. In the second example we look at the aggregation and pooling of corporate infrastructure which uses exclusive spectrum and removes the borders between different mobile operators. Finally we look at the aggregation of user-deployed or crowdsourced infrastructure that opportunistically uses spectrum and removes the borders between independently deployed hotspots. These are starting points and the full realization of the vision will involve more dynamic access to spectrum and more extensive sharing of infrastructure. Hence, the final part of the paper describes the resulting research challenges.

Index Terms

Dynamic Spectrum Access, Sharing, Value-chain, Virtual Network Operator, Service-driven Networks
Spectrum Without Bounds, Networks Without Borders

I. INTRODUCTION

Moving from static forms of spectrum access to more dynamic scenarios which redirect spectrum resources to where they are needed and which heavily leverage spectrum sharing coupled with the ability to compose networks from a pool of shared infrastructure can have a transformative effect on how mobile and wireless networks are owned and controlled. The purpose of this paper is to present one future vision which is predicated on this greater freedom of access to spectrum and more inclusive view of network infrastructure.

Our vision, which we call Networks without Borders (NwoB), is based on a market-place of virtual network operators which construct networks from a pool of shared resources (e.g. base-stations, spectrum, core network components, cloud resources, processing capabilities etc.) [1]. The resources will be sourced from traditional industry players as well as crowdsourced from individuals. The new forms of service provider will be facilitated by the virtual network operators. These operators will control virtualised slices of the physical resources aggregated from the pool to create customized virtual networks on which to offer specialised services to end-users. There will be a high level of dynamism involved; unneeded resources will be returned to the pool for redeployment by other service providers. This model broadens the definition of infrastructure provider through significantly leveraging user-deployed infrastructure in a systematic manner. This model extends the sharing economy deeper into the mobile network through extensively embracing sharing of all types of resources, especially the spectrum. The model fundamentally redefines the term mobile network operator as a highly virtualized entity using heavily shared heterogeneous resources. The idea of an over-the-top service no longer exists, as the service and virtual network over which it is delivered are interconnected, which allows for the service provider to extract value from the network while compensating the virtual operator for their services.

Our proposed architecture recognizes the pressures on wireless providers to substantially increase their ability to support mobile data, while simultaneously keeping costs to customers...
fairly constant, pointing to the need for a new, sustainable business model that involves major changes in the way networks are provisioned. The vision is transformative because it challenges the conventional notion that exclusive ownership of spectrum resources over large geographical areas is needed to justify large-scale investment in infrastructure in the first place, leading to subsequent tight control of both the spectrum and the infrastructure in order to recoup the investments. More fundamentally, the vision recognizes the fact that a network cannot continue to be a costly set of pipes from which value is difficult to extract.

The future vision of networks is built up over the paper. The paper begins by exploring the mobile network value-chain, identifies the changes that have taken place and are taking place in that chain and discusses the economic imperatives and innovative drivers of the change. Section III draws attention to the fact that these changes do not substantially alter the structure of the value-chain. In particular section III focuses on the fact that sharing of infrastructure and of spectrum is restricted and, as a result, so are the subsequent economic gains that can be achieved. In response to this, a new value-chain is suggested and presented in section IV. This chain starts with the services. The network is composed from a pool of resources in response to this service need, encapsulating the ideas presented in the opening paragraphs of this paper. Section V goes on to detail some early work on some of the possibilities for rethinking networks as suggested and provides details of examples in which spectrum is more dynamically accessed and infrastructure is pooled. Section VI lists the outstanding challenges in the field from a technical and regulatory perspective, and lays the foundation for future research. Section VII draws conclusions.

II. THE CHANGING MOBILE NETWORK VALUE-CHAIN

Figure 1 shows an illustrative value-chain associated with a mobile or cellular network. The value-chain includes all the activities that exist as a direct result of usage of the cellular network. The purpose of creating the chain is to understand where the costs are incurred and the revenue is generated [2]. In the early years of cellular networks, all aspects of the value-chain were typically controlled by one entity, namely the mobile network operator (MNO). In the past decades we have seen many changes that have challenged this approach.

Firstly, different parts of the value-chain have come under the control of different players. A key example of this has been the emergence of the mobile virtual network operator (MVNO).
An MVNO is a wireless communications services provider that does not own the spectrum, radio access network and backhaul elements of the value-chain. Instead, an MVNO enters into a service level agreement with a mobile network operator to obtain bulk access to network services at wholesale rates and can subsequently set its own retail prices independently [3]. The MVNO incurs no significant capital expenditure on spectrum and infrastructure and does not have responsibilities associated with building out extensive radio infrastructure. The service level agreements in place are simple. The MVNOs bulk-buy minutes or data and typically there is no differentiation between the MNO or MVNO customer on the physical network. Different flavours of MVNOs exist and are defined according to how much of the value-chain they control. For example, some MVNOs merely focus on branding and marketing and therefore control the distribution channel only, while others may operate their own core networks, provide value-added services, back-office services and define the offerings as well. This latter value-chain is labelled as 1 in Figure 2.
Secondly we have seen the emergence of competitive/complementary value-chains. WiFi is a good example of this and labelled 2 in Figure 2. WiFi uses license-exempt spectrum and is a user-deployed technology which is configured to offer hotspots of connectivity. In the past number of years WiFi has moved from what was once a competing wireless infrastructure to a complementary one, playing a central role in supporting data off-loading from the mobile networks\(^1\). The value-chains of the mobile operator and the WiFi hotspot are becoming increasingly intertwined. Relationships exist between mobile operators and WiFi providers, and protocols such as Passbook, Hotspot 2.0 and IEEE 802.11u will make it easier to move from one wireless infrastructure to the other. There are of course value-chains that are still to fully emerge. The machine-to-machine (M2M) value-chain is still evolving. Solutions exist that build this into the value-chain of the mobile operator, and others exist to support M2M on completely separate networks. Hence, how competitive or complementary these value-chains may be is yet to be determined.

Thirdly elements of the value-chain are being bypassed by new players. The emergence in the last decade of what are known as over-the-top services (e.g. Skype and WhatsApp) make use of the radio spectrum, radio access network, backhaul and core network of the mobile operator but do not contribute to operator income. In other words no service level agreements exist with the mobile operator. These services are seen as sabotaging value or extracting revenue that would otherwise be generated by the MNO from traditional voice and data services. Content and value-added services are increasingly designed, built and operated by third parties that use this mode of operation. In essence the parts of the value-chain which generate revenue are under the control of different entities than those who control the parts of the value-chain which incur great costs. This is labelled 3 in Figure 2.

The increasing complexity and expense of new technologies, the increasing speed at which these new technologies emerge and are in-demand by users, the challenge in providing coverage in areas that are unlikely to generate sufficient revenue to justify heavy infrastructure investment, all have led to significant challenges in the provisioning of networking infrastructure. The loss of revenue and value to other players via over-the-top services, reducing the physical network to

---

\(^1\)Forecasts predicting data offloading of in excess of 60% of mobile data traffic have been made by CISCO and others. Examples of CISCO predictions can be found [http://www.cisco.com/en/US/solutions/collateral/ns341/ns525/ns537/ns705/ns827/white_paper_c11-520862.html](http://www.cisco.com/en/US/solutions/collateral/ns341/ns525/ns537/ns705/ns827/white_paper_c11-520862.html).
a set of expensive pipes, compounds this further. To respond to the large costs associated with provisioning of networks we see a fourth major change occurring in the value-chain, and that is a growth in the sharing of ownership and control within the infrastructure elements of the chain (e.g. the RAN, backhaul and core network elements), i.e. the areas in which most costs are accrued.

From a technical perspective, sharing is often described as either passive or active [4]. Passive sharing refers to situations in which resources are shared without the need to change the radio access network or the core network. For example the sharing of sites and masts between mobile operators has been ongoing for a long time. Active sharing usually involves some level of virtualisation which tends to have implications for the underlying architecture. Active RAN sharing, for example, refers to the sharing of the basestation between different operators. There are also options for mobile operators to share the core network. In fact 3GPP\(^2\) has identified two architectures for network sharing. In both architectures, the radio access network (RAN) is shared. The first of these is known as the Multi-Operator Core Network (MOCN) configuration and in this case only the RAN is shared [5]. In the Gateway Core Network (GWCN) configuration, elements of the core network are also shared between operators [5]. The implementation of these configurations involves the virtualisation of the network entities so that resources can be shared among operators. Active network sharing is still in its infancy and is not widely deployed. Other architectures that are suited to sharing are also emerging as viable. One such example is the Cloud RAN [6], [7]. In this case the RAN is split between remote radio heads and baseband processing, the latter residing in the cloud.

We also see sharing within the elements of the value-chain emerging in other forms. Through the rise in popularity of the femtocell\(^3\) we also see the sharing of ownership of the basestation between the mobile operator and the customers. Many of the femtocell business models involve the customer purchasing a femtocell for their home and connecting to the core network via the customer’s own fixed network (i.e. via the customer backhaul). The ownership of the RAN from both a capital expenditure (CAPEX) and operational expenditure (OPEX) perspective is

---

\(^2\)The 3rd Generation Partnership Project (3GPP) is a collaboration between groups of telecommunications associations which come together to develop standards for mobile networks.

\(^3\)A femtocell is a basestation that is intended to be deployed by the end user within their home or work premises. Once in the vicinity of a femtocell, handover from the macrocell to the femtocell occurs.
therefore split between the service provider and the end user.

III. THE RESTRICTION OF CHANGE

The changes in the control of the value-chain and the increased focus on sharing are driven by the different factors which have been alluded to already in the paper. The major driver is the economic imperative to find ways of reducing cost and maximising return. Change is also driven by new players seeking to insert themselves within the chain as part of what is widely understood as the disruptive innovation process [8]. However, while on the one hand there appears to be a lot of change or disruption, on deeper analysis we see that change is quite restricted. Phrased differently, there are borders around different parts of the value-chain that limit change. The consequence of this, we maintain, is reduced potential for cost savings, reduced potential for addressing the challenges of sabotage of value and reduced options for innovation. Two significant borders, and the basis of our analysis, are those that surround infrastructure and spectrum.

Sharing of infrastructure is a topic of increasing focus, with active sharing of networks promising cost savings between operators. However as would be expected the 3GPP solutions are very much aimed at supporting what we might call the traditional value-chain or the current business models underpinning the mobile operator. Sharing in the case of the MOCN and GWCN solutions described above is understood to mean sharing between two or more existing mobile network operators. Even in solutions which push the virtualisation deeper into the network [9], in all of these approaches to sharing the network itself is centrally planned and rolled-out/upgraded as before under the control of the mobile operator.

However, it is when spectrum is added to the discussion that the restrictions become more obvious. Spectrum is at the heart of all wireless networks and, in general, is a tightly controlled resource, much more so than infrastructure. The International Telecommunication Union (ITU) coordinates the shared global use of the radio spectrum. At regional level there are different bodies at work depending on location. In Europe, for example, the European Union issues directives to member states, including directives on spectrum policy. As a non European example CITEL is the Inter-American Telecommunication Commission. At national level the national regulatory authority is in charge. The allocation of spectrum to particular services happens at international and regional level, though this can be further refined at national level when more
specific band plans are created. The assignment of spectrum to different companies and agencies happens at national level.

Extremely tight borders are drawn around spectrum. At the spectrum allocation stage, not alone is spectrum coupled with a service (e.g. mobile services, satellite services, TV broadcasting service), spectrum bands are also demarcated as Time Division Duplex (TDD) or Frequency Division Duplex (FDD). At the assignment stage there are different procedures in place. Broadly speaking there are two classes of spectrum, licensed and license-exempt. The former grants exclusive usage of the spectrum to the licensee (usually for long periods of time, e.g. 20 years). Licensed spectrum is typically assigned via auction or using some kind of administrative process. The latter defines bands in which any interested party can operate, provided simple rules, such as limitations in power levels, are followed. The ISM bands, in which WiFi operates, are examples of a licensed-exempt scenario.

From the perspective of the operator, licensed spectrum tends to be tightly coupled with the ability to provide a quality service, long-term security and the subsequent justification for large-scale investment in infrastructure. Because of the power restrictions in these bands, license-exempt use is synonymous with small cells and short ranges. It also tends to be synonymous with best effort and no expectation of blanket coverage. In addition, the networks that operate in these bands are typically user-deployed and tend to grow organically, though in recent years mobile and fixed operators have been actively deploying WiFi networks to complement their cellular networks.

There has been much work over the past decade on the academic front to break down this view of the spectrum world [10], [11], [12], [13], [14], [15], [16]. The work that has sought to change how spectrum is assigned and accessed broadly falls under the heading of dynamic spectrum access or spectrum sharing. Dynamic spectrum access (DSA) refers to a mode of accessing spectrum that happens on a dynamic rather than static basis. Spectrum can be accessed temporarily on exclusive or shared basis and on a temporal or spatial basis depending on the specific technologies in use. The perceived scarcity of spectrum was the initial driver for this challenge. Measurements emerged showing low-usage levels of much of the assigned licensed
spectrum and hence an opportunity to use the idle capacity\textsuperscript{4}. Other drivers were linked with the barriers to innovation and development of new services [17], [18]. The static and exclusive means of assigning licensed spectrum is slow. Once assigned it is hugely difficult to clear and retrieve. The turning-off of analogue TV took 28 years from the inception of the idea to the final switch-off and, in fact, analogue TV is still not universally switched off. There are numerous examples of allocated spectrum targeted at services that never emerged. In addition to academic work, there has been much work on the policy side to promote less static approaches to dynamic spectrum and an increased focus on sharing\textsuperscript{5}.

Despite all of this work, advancement has been slow. It remains the case that we continue to designate spectrum as TDD and FDD irrespective of which technologies emerge as most popular. It remains the case that FDD blocks are in the main still assigned symmetrically (equal bandwidths for the up and downlinks) irrespective of traffic demands. It remains the case that early areas of opportunity that have been identified for deployment of more dynamic access to spectrum such as in the TV bands (i.e. the so called TVWS), are yet to take off, despite tests and trials on-going internationally\textsuperscript{6}. And while the value of unlicensed spectrum has been more explicitly acknowledged [19], especially in the context of WiFi networks and data off-loading, this has not substantially altered the wider perspective on unlicensed spectrum. Hence, as stated above, it remains the case that licensed spectrum is coupled tightly with the ability to provide good quality of service, long-term security and the subsequent justification for large-scale investment in infrastructure. It remains the case that license-exempt is considered inappropriate for services that are expected to be reliable. This underpins the notion that user-deployed infrastructure is limited and scale ultimately is only reached through central planning. The borders remain tightly drawn.

\textsuperscript{4}http://www.sharedspectrum.com/papers/spectrum-reports/ are examples of a range of measurements that were taken as part of these campaigns.


\textsuperscript{6}There have been trials in the USA, Singapore, Finland, UK, Ireland and other countries. Currently the UK regulator Ofcom is coordinating significant trials in the UK.
IV. THE REMOVAL OF RESTRICTIONS & A NEW VALUE CHAIN

It is our belief, that as long as the changes which are taking place are restricted by the more traditional view of the mobile operator and traditional views of license-exempted systems, the real cost savings of sharing cannot be achieved and the challenge of over-the-top sabotaging of value cannot be addressed. With this in mind we consider a new approach we call Networks without Borders. This approach is about re-conceiving networks without the restrictions described above and from a perspective of significantly more fluid access to spectrum and all of the resources that comprise the network. It leads to a reconstituting of the value-chain. Figure 3 attempts to capture what this value-chain might look like. There are a number of points that are worth noting.

Firstly the service to be delivered is the driver of the network. The Networks without Borders approach envisages a network as an entity that ‘responds’ to a service, a service-driven network. This value-chain draws more from that of the Internet [20], with the physical networks providing the connectivity for whatever service is required, than from the value-chains described earlier. The key point in Figure 3 is that the connectivity is tailored for the service. To understand this the term service must be unpacked. MNOs tend to define services along traditional lines such as voice, text and data. Our view of a service is a much more nuanced view than that and we think about services in a much more fine-grained manner. So, for example, a Netflix-like service specifically focused on streaming of video will require a specific kind of connectivity while a
bursty M2M service would require a very different kind of connectivity. The service will be able to select from a menu of virtual network operators, each offering different kinds of coverage and capacity. The idea of an over-the-top service no longer exists, as the service and virtual network over which it is delivered are interconnected, which allows for the service provider to extract value from the network while compensating the virtual operator for their services. The notion of a service bundle is also introduced in Figure 3 to take into account the possibility that services can also be grouped (e.g. entertainment and billing services).

This leads to the second point, that virtualisation is key. The virtual network operator, as understood in Figure 3, is something very different from the MVNO depicted in Figure 2 which simply leverages infrastructure and works on the basis of simple bilateral agreements. It is an entity that manages specific virtual slices of a selection of resources to create a network that is highly tuned to a specific service. For example, the virtual network operator may control an extremely low-latency network or a high power efficient (green) network. The virtual network operator may chose to operate on a substrate of exclusive resources or shared resources. Hence the virtual network operators can work at different price points. Virtualisation has been well studied in the area of fixed networks. More recent work has seen a focus on wireless networks [21], [22], and while virtualisation over a set of physical resources that change, or that are not exclusively owned, needs much work, the emerging solutions bode well for exploiting virtualisation to allow extensive sharing of resources.

Thirdly, and perhaps most significantly, the resources which make up the network are considered to come from a pool of all and every resource. This means that the borders between licensed and licensed-exempt spectrum or professionally and user-deployed infrastructure are removed. Spectrum is spectrum and infrastructure is infrastructure. Whatever types of resources are available or, more importantly, whatever types of resources that are appropriate for different kinds of services are used. This notion of a pool of resources lays the way open for a wider range of contributors to the pool, and a broader means of encouraging investment in infrastructure. Given the vast need for small cells to deliver the capacities for future data consumption, it makes great sense that small cells, or indeed other resources, can be crowdsourced from the public, for example. The ‘bring your own’ concept can be extended. Bring your own device (BYOD) (also known as bring your own technology (BYOT), bring your own phone (BYOP), and bring your own PC (BYOPC)) refers to the policy of permitting employees to bring personally owned
mobile devices (laptops, tablets, and smart phones) to the workplace. These devices can then be used to access privileged company information and applications. We can extend this to be not just about the devices that hang on the end of the network but to include the devices/resources that constitute the network. Bring your own spectrum (BYOS), bring your own basestation (BYOBS), bring your own processing power (BYOPP) etc. are examples of how this can be extended.

Following on from the idea of a pool of resources, a fourth point is worth stressing. This approach opens the way for making significant progress on how spectrum is accessed. At least a decade of academic work has taken place which has explored a myriad of ways in which spectrum can be accessed in a dynamic manner. There is technical work on how radios can sense white spaces, use databases for identifying spectrum, use dynamic auctions of all kinds in spot-markets or over longer time-scales, trade spectrum and much more. There is also much work on the licensed or licensed-exempt constructs in which these more dynamic approaches can reside. However, as mentioned in the previous sections of the paper, this has not yet translated into significant regulatory action or change. The mobile network operator continues to prefer exclusive spectrum because, as is very evident from the original value-chain presented in Figure 1, from spectrum all else flows. The mobile operator continues to try and understand the value of spectrum from this viewpoint and tends to ask questions that attempt to value shared spectrum within this confined framework. Figure 3 aims to shift this perspective to the idea that the appropriate connectivity for the specific service is needed and more extensive virtualisation can create the illusion of infinite resources that is needed to fulfil demands [23]. Hence, Figure 3 is as much about new technology as it is about a mindset change.

Fifthly this value-chain allows for new players to come on board. In fact, most of the entities in the value-chain in Figure 3 can be conceptualised as services. Hence, the value-chain provides a service-centric view with the virtual operators and resource aggregators becoming new forms of service innovation opportunities. Focusing on the resource aggregator for a moment provides some examples of this. This entity classifies and organises physical resources so a subset of those resources can be gathered together and subsequently ‘sliced’ into virtual pieces and made available to the virtual network operator. The aggregator may garner the subset of physical resources using old-fashioned negotiation mechanisms, using real-time online auctions or very much in the manner that cloud resources are currently garnered for streaming video services. The aggregated subset of resources which make up the physical substrate on which the virtual
operators reside may remain as part of an aggregated subset for long or short periods of time. Hence, the idea that resources are returned to the pool when not needed, i.e. a recycling service, is supported.

Finally, it is worth noting that at a glance the value-chain of Figure 3 is a much fatter value-chain than that in Figure 1, emphasizing the opportunity for parallel co-existing activities, the opportunity for big and small players and the opportunity for niche as well as mainstream activities.

V. SOME EARLY WORK ON NETWORKS WITHOUT BORDERS

The ideas presented in the previous section cover an extremely wide range of topics and areas for research. Our early work on Networks without Borders has focused on the right hand side of the value-chain presented in Figure 3. It has concentrated on the removal of restrictions that are placed on key resources and on the aggregation of those resources in response to specific drivers. What follows here are three different examples of this work. In the first example the focus is purely on spectrum. In particular we look at how FDD and TDD labels on spectrum can be more dynamically allocated in response to service and market needs rather than labelled in advance in a static manner. We essentially remove the borders between TDD and FDD. In the second example we look at the aggregation and pooling of corporate infrastructure which uses exclusive spectrum. Here we remove the borders between different mobile operators. Finally we look at the aggregation of user-deployed or crowdsourced infrastructure that opportunistically uses spectrum. In this last example we remove the borders between independently deployed hotspots. A rounded solution would contain elements of all this early work and much more, especially in the area of virtualisation and services. The purpose of presenting the early work is simply to demonstrate some of the concepts that are part of the Networks without Borders vision.

A. Spectrum without Bounds - Removing Restrictions on Spectrum

The straight-forward, yet complex example, we start with is connected with the notion that spectrum must be designated as TDD and FDD in advance of use of that spectrum. A good example of this relates to the LTE or 4G spectrum bands. E-UTRA has defined 35 different LTE bands. There are 24 paired bands to frequency division duplex (FDD) mode, ranging in width
from 10 to 90 MHz. There are 11 bands that have been identified for time division duplex (TDD) LTE systems with bandwidths ranging from 10 to 200 MHz. This is typical of how spectrum is managed. The FDD, TDD division is drawn irrespective of whether it is appropriate, whether one technology or another will be more popular, whether services are more suited to one rather than the other, and whether uplink and downlink traffic patterns are symmetric for all services. In [24] we explore a spectrum assignment process through auction that is not bound by these restrictions. While the paper fully explains this auction process, some key details are reproduced here in the context of Network with Borders and, in fact, *Spectrum without Bounds*.

The auction described in the aforementioned paper combines the allocation of FDD and TDD bands with spectrum assignment to the end-user rather than treating it as a two-stage process involving a band planning and subsequent assignment stage. In other words the resulting division between FDD and TDD mainly lies in the hands of the bidders, who are making bids on the basis of their needs, rather than on a pre-designated structure imposed by regulators. In addition, in the case of FDD the user can bid for non-symmetric blocks of spectrum to suit whatever type of service the network wishes to facilitate. At the heart of the solution is a combinatorial clock-auction that is highly *expressive*. Combinatorial auctions typically allow bidders to ask for ‘A AND B’. This for example would be needed to acquire paired spectrum for FDD services. The block size of the bid is also set in advance - i.e. the number of MHz per block. In our auctions we allow far wider requests including all combinations of ‘AND, OR and XOR’ bids as well as bids that allow the bidder to make statements like ’ANY X MHz from the frequency range Y MHz to Z MHz’. This is what we mean when we say the auction is expressive. In addition, no restriction is imposed on the size of the basic bid (i.e. number of MHz requested) and the only restriction that exists is the minimum duplex spacing required between FDD blocks. Bidders can ask for what they really want.

The paper details the price-discovery phase and the winner determination algorithms used in this auction process. Results are presented to highlight the value of this approach. The results are based on a wide range of simulations which are carried out for different bidder profiles and different auction parameters. As expected if the TDD bidders value the spectrum more than the FDD bidders, then all spectrum ends up being allocated to TDD bidders. In the case where the opposite is true the bulk of assignments go to the FDD bidders, allowing of course for the fact that some duplex spacing is need. Overall the increased freedom in the proposed
auction mechanisms result in an improved spectrum allocation efficiency - i.e. the requirements of the bidders and the assignment of spectrum are well matched. This is in contrast to traditional clock auctions which may result in a bidder getting more spectrum that it needs because the only options are too much or too little or, worse still, having no option as the wrong kind of spectrum is available (e.g. a TDD block instead of FDD blocks). The asymmetric FDD options also ensure little wastage. Moreover some of the decision variables currently chosen by the auctioneer through consultation emerge as outcomes of the market demands.

Returning to the terminology in Figure 3, we can say that the aggregator services consist of an expressive combinatorial auction and that the spectrum which is auctioned is from a pool of unallocated (from an FDD and TDD perspective) resources. The removal of these bounds has made the spectrum ‘go further’.

B. Corporate Infrastructure and Exclusive Spectrum Aggregation

As part of Networks Without Borders the physical infrastructure, coming from various sources, ranging from large-scale operators to individuals, becomes primarily a means for the transport of bits (packets) and an inter-connection point to the cloud of services, formed from the virtualized network control processes and functions. The physical resources, control processes and functions, when turned into services, are traded among different parties of the newly formed mobile communications value chain, which can be represented as in Figure 3. Selections of these virtualized network resources, or otherwise services, are aggregated to create networks that meet a specific end goal. The desired end goal may have a variety of forms, from the basic quality of service aspects such as coverage or capacity, to more complex ones such as indoor coverage or low latency, or a combination thereof. We refer to the process of aggregating and selecting virtualized network resources that meet the desired end goal as network shaping. Being part of the NwoB ecosystem, the network shaping process itself becomes a resource aggregation service which can be offered to over-the-top service providers, for example, video streaming provider (Netflix) or low-rate voice service provider (Skype).

One possible way of looking at network shaping is to analyze construction of a network that will purely rely on virtualized network elements belonging to the existing mobile network operators. In [25] we have looked at a model of constructing a mobile network when a border, the border between a mobile network operator and mobile network operator, was removed. To a
certain extent we already see this happening as part of inter-operator network sharing. However, as far as Networks without Borders is concerned, instead of assuming bilateral agreements between the selected operators, we consider the resources of all the MNOs as a collective pool and analyzed the resulting efficiencies, which we recall in the following along with any relevant argumentation.

Inter-operator network sharing is a well recognized form of network-related cost optimizations [26], [27], [28]. The most widespread form of inter-operator network sharing is site sharing (passive sharing), where the operators share the physical site and its basic installations, such as mast, cooling equipment and power supply. Passive sharing has been practiced by MNOs at least since the introduction of 3G systems [29]. A more complex form of inter-operator network sharing is sharing of active network elements, such as base stations, which has recently gained its momentum as more and more operators world-wide have decided to share their existing RANs or undertake joint greenfield deployments, such as Newco, the joint venture of two Danish MNOs [30], or NetWorkS!, the merger between two Polish units of T-Mobile and Orange [31], just to name a few. Networks without Borders goes a step further than the aforementioned inter-operator network sharing models. NwoB offers a new way of thinking of network sharing, which allows for unbounded customization of a network towards meeting the demand of an over-the-top service. From the view point of the existing MNOs only, such unbounded customization may help to improve network resource utilization and increase cost efficiency, while potentially increasing service quality, due to coverage and capacity enhancements [28].

In our initial models we consider basic QoS requirements only, hence, in [25] we draw a parallel between the network planning and the network shaping processes. Contemporary mobile networks are dimensioned, planned, deployed and optimized towards meeting the demand of a conventional set of mobile services, i.e. voice and data services [29]. These services can be directly translated into basic QoS requirements of coverage and capacity provisioning. Each of the MNOs deploying its own mobile network will follow these requirements, leading to almost identical network roll-out principles, which, presumably, will rely on the same (or similar) dimensioning assumptions. Effectively, mobile networks of today have similar coverage and capacity efficiency, and, conversely, offered QoS. Part of the idea behind NwoB is to open up a market for many virtual operators that will offer network access which has distinctively different QoS, even if only coverage and capacity are considered. This, however, requires pooling and
splitting the virtualized network resources (including spectrum) without constraints imposed by fixed service level agreements and roaming agreements.

In the very first formulation of the network shaping process, we simplify things and restrict our attention to MNO base station sharing and coverage provisioning only. Hence, we can state the following problem [25]: given a set of geographical points in an area, called a universe, and a set of base stations (with their corresponding configurations), each of which covering a subset of elements of the universe (service areas), find a minimum number of base stations such that the union of their service areas will contain a specified fraction of the elements of the universe to meet a pre-specified service reliability at a minimum cost. Figure 4a depicts a multi-operator wireless network prior to coverage shaping, while Figure 4b shows a minimum cost instance of the same network. It is worth noting that the aforementioned problem formulation can be seen as an extension to a classical combinatorial optimization formulation, the set cover problem (SCP) [32].

In our first attempt to evaluate large-scale corporate resource aggregation we turn to look at the efficiency of the inter-operator infrastructure sharing only. We evaluate this efficiency based on real network deployment data, and a randomly generated dataset [25]. In order to find the network that represents the minimum cost subset of the pooled network elements which yield the desired coverage under a pre-specified reliability level we model network shaping as a mixed integer program (MIP) which we solve using commercially available exact MIP solvers [25].

It is important to understand that the achieved gains are context and localization dependent, and as such should be treated as indicators of the possible efficiencies given certain deployment
density. Figures 5a and 5b show the shaping efficiency for the two power allocation schemes in the case when full coverage is required for the whole geographical area. What we observe is that the highest efficiency improvements can be achieved for the areas with higher deployment density. In the densest real case (Warszawa) the gains of shaping were as high as 98%. This means that a single operator, in order to provide full macrocell coverage, requires only 2% of all the existing base stations in the area (i.e. 16 out of 934). This gain is also very high in comparison to the cost of a single-operator network for both homogeneous and heterogeneous power allocations, which shows how much redundancy is introduced to the network in order to secure the capacity and possibly ensure indoor macro-, microcell coverage. For the lowest density case (Świdnica), the benefits of sharing arise as a result of the overlaying network topologies, that are deployed most probably according to similar dimensioning and planning assumptions, which may justify potential benefits of infrastructure sharing in less urbanized areas. In this case a single operator may not need to roll-out all the base stations required to fully cover the considered area; instead, infrastructure sharing among several operators may make it economically feasible to provide service to such typically under-served areas. Analyzing the differences in results between homogeneous and heterogeneous power allocations, we can also observe that the efficiency is reduced with the power allocated to the system. This reduction is inversely proportional to the deployment density.

As mentioned previously the achieved results are likely to be context dependent, therefore we have decided to model the base station deployments as a Poisson point process (PPP) where the distribution of points across space is uniform, and which has been widely used in research to model spatial base station distribution [33]. When comparing real and random efficiency results (see [25]) for high deployment density cases we can note that the efficiency follows nearly the same trend, which may suggest that the PPP model may provide an approximate upper bound on the efficiency for a generalized high density base station deployment (above approximately $5 \text{ bs/km}^2$). However, if we compare the same results for low deployment densities (below approximately $5 \text{ bs/km}^2$) the gains for random and real deployments diverge. This may be due to the fact that real multi-operator deployments were planned to provide coverage for the given area following more or less the same coverage pattern for every operator (hence, resulting in redundancies), while the random deployments are more spread, resulting in lower redundancy. In this case it is hard to assess the feasibility of the PPP model, and a better analytical model that
more accurately describes base station deployment for a multi-operator shared network needs to be developed, which is part of our future work.

The presented analysis indicates that there are significant gains to be made from considering mobile network infrastructure as a collective when only coverage provisioning is considered. In the following, we plan to expand this analysis on spectrum sharing and capacity provisioning. However, there are already some initial studies which report that inter-operator spectrum sharing may produce significant gains. In [34], after analyzing radio spectrum occupancy for two Chinese MNOs in a few areas of China, the authors conclude that there are significant imbalances in the usage of radio spectrum bands belonging to the analyzed MNOs, i.e. while one of the operators has low spectrum usage, the other experiences increased spectrum usage. These imbalances arise as a consequence of a mismatch between dynamic spectrum demand and system bandwidths allocated to this specific areas. The study shows that if spectrum sharing is applied, operators experiencing a peak in spectrum usage could re-use the spectral resources of the other operators, hence, temporally increasing the capacity of their networks. In [35] spectrum sharing is applied to a network of operators that have high spatial separation (low inter-cell interference levels), in such case the reported gains in terms of average user throughput can be as high as 100%. When, as in [36], the capacity of the whole LTE or HSDPA system is considered in a typical simulation scenario, along with a spectrum auctioning mechanism, these gains are scaled down but the spectrum sharing gain remains. Finally, the spectrum sharing gain occurs for both competitive and cooperative scenarios (with the cooperative scenario allowing for higher relative gains) when spectrum sharing between operators is modeled as a game of power control [37]. All these studies, though initial and perhaps lacking generality, indicate that if the correlation of demand across operators is low enough, spectrum sharing gains may be produced both in rural areas, where base station separation may allow for spectrum re-allocations between operators, and in highly populated areas, due to the spectrum usage imbalances among operators.

Operators are sharing infrastructure, more and more, as with the example of Newco7 or

---

7Document available at http://ec.europa.eu/competition/ecn/brief/02_2012/dk_mobile.pdf summarizes a network sharing joint venture (NewCo) between TeliaDenmark and Telenor A/S. The sharing agreement involves sharing of the network infrastructure and radio frequencies for all cellular technologies, i.e. 2G, 3G, LTE and potentially LTE-Advanced.
NetWorkS\textsuperscript{8} . We envision that the trend will continue, towards NwoB, which will necessitate more sophisticated understanding of how the aggregator services can combine wireless network resources coming from various infrastructure providers to instantiate virtual networks. What we have shown with our initial model is that the removal of the bounds between mobile network operators may significantly increase the flexibility and cost efficiency of running an instance of a virtual network.

C. Crowdsourced Infrastructure and Opportunistic Spectrum Aggregation

As an alternative to a network constructed from entities that are controlled by corporations we have also carried out some work that has looked into how networks might be crowdsourced. A WiFi network can of course be seen as a crowdsourced network but our work looks beyond access points that are confined to certain frequencies and to limited ranges. Instead we look at a more futuristic version of user-deployed infrastructure that is based on cognitive radio and that can dynamically access whatever spectrum it can find. This work is presented in [38] and the relevant aspects are recast here in the context of the re-constituted value-chain.

The user-deployed access point, as mentioned already in this context, is a cognitive radio [18] with a frequency agile frontend. The cognitive radio scavenges for available spectrum. In our work the database paradigm is deployed and the radio access point contacts a white spaces geolocation database to ascertain which frequencies are available\(^9\). The geolocation database has been much discussed in the literature and is currently the preferred approach to TV White Space usage [39],[40]. Our access points also retrieve rules about power levels and emission characteristics from the database in the form of a spectral mask. Using a variety of waveform shaping techniques we have developed [41],[42], the radio sculpts the emission profile to comply with the retrieved mask, ensuring that harmful interference to systems in adjacent channels does not occur.

The above description allows for one communication system based on white space frequencies to be deployed and to serve a set of users. The scenario of interest is one in which individual neighbouring user-deployed systems aggregate to form a network which is available to all users of the individual systems. Figure 6 captures this idea. In Figure 6(a) a number of these advanced access points exist. As the deployments become denser we look at how the individual access points can be aggregated. Figure 6(b) shows two separate aggregations and Figure 6(c) shows one very large aggregation.

![Figure 6: The Formation of Coalitions](image)

In this early work we use the concept of the *cyclostationary signature* as a mechanism for coalition definition and as the basis on which a handover mechanism between neighbouring nodes

\(^9\)A geolocation database is a database which contains information such as incumbent deployment patterns, propagation models and terrain data that can be used to determine which frequencies are unused in a given geographical location.
in the coalition is supported. Sutton et al. introduced the idea of the cyclostationary signature in [43]. A cyclostationary signature is a feature intentionally embedded in the physical properties of a digital communications signal. The set of aggregated access points wishing to belong to the same coalition, all use the same signature to signify membership of that coalition, and each access point embeds the signature in its transmitted signal. The signature is not only useful as an identifier of the aggregated group but is also useful as a means of frequency rendezvous.\footnote{Frequency rendezvous is the term used to describe when a transmitter and receiver both synchronise to the same frequency. In traditional systems this is easily done. In dynamic spectrum access networks it can be complex to ensure that rendezvous happens as so many frequencies of operation are possible. Hence efficient techniques are needed.}

The spectral frequency of an embedded cyclostationary signature is a function of the carrier frequency of the signal and the properties of the subcarrier sets used in its generation. For example, mapping a subcarrier set onto a second set, equidistant from the carrier frequency, results in a cyclostationary feature that is centered upon that carrier frequency. So even though all cyclostationary signatures of the aggregated group will appear at the same cyclic frequency, each signature will reveal the individual carrier frequency of that access point. Hence a user in the coalition simply locks onto the cyclic frequency and just like in more traditional handover scenarios locks on to the strongest one. This is what is known as a terminal initiated handover and it is considered to be lightweight as central coordination is not needed. The cognitive terminal reconfigures itself each time it locks on to a different access point to the appropriate frequency. Tallon et al. have shown this approach to be viable and created working prototypes of it in action. The scenario envisages some form of signature authority which manages in a lightweight fashion the distribution of signatures to the members of the coalition.

Returning to the new value-chain in Figure 3, we can say that the resources which are aggregated in this scenario comprise spectrum and radio access networks. The resource aggregator services consist of the signature authority and the database for spectrum occupancy and rules of operation.

\textbf{D. Virtualising over the Aggregated Resources}

The aggregated corporate resources of course contain basestations of high quality and a constant supply of spectrum. The aggregated crowdsourced resources are of a different type and the bandwidth of the network might vary over time. It is possible to envisage that a virtual
network operator running over the corporate resources may be called on for one kind of service (e.g. one that demands high QoS), while a virtual network operator running on the crowdsourced system may provide a different kind of service (e.g. one that demands best effort). It is also possible to imagine how a virtual network operator can use virtual slices from both sets of aggregated resources. The operator can balance costs, for example, by drawing on a mix of the ‘cheaper’ and ‘more expensive’ resources.

VI. TECHNICAL, MARKET, AND REGULATORY CHALLENGES

The vision of Networks without Borders that we offer here implies deep changes to how wireless services are provisioned and sold and how virtual and physical networks are designed, built, and managed. In other words, it opens up a plethora of technical, economic, and public policy questions. In this section we outline and discuss some of those challenges.

A. Technical Challenges

The first step, and one that we have started to undertake, is to develop a better understanding of the potential efficiency gains and limitations of sharing in existing mobile networks. As discussed earlier, this sharing has to date taken the form primarily of bilateral agreements between large operators. We propose to take sharing further, enabling virtual networks that are instantiated from a selection of resources belonging to multiple operators, large and small, and a first question is how to select the optimal subset of these resources that would meet a desired quality of service for the network. Optimal, in this context, can be thought of as the set of resources that lead to the lowest cost to the virtual network operator. An initial formulation, considering coverage goals, was discussed in the previous section. The next logical step is to re-formulate the problem in terms of capacity goals. This requires models of demand for cellular services, a separate challenge in and of itself and a moving target. The assessment of efficiency gains due to sharing also requires a better understanding of the correlation of demand across space and time, and between operators. Clearly, the higher the correlation of the demand that multiple operators experience at the same location and time, the lower the gains from treating resources as belonging to a common pool. Fortunately, some early preliminary studies indicate that the correlation is low enough that sharing is expected to bring significant gains, even in highly
populated areas. In rural areas, where there is typically substantial idle capacity for any given operator, the gains are expected to be even higher.

The problem above acquires additional degrees of freedom when the virtual networks can be formed through the vertical integration of different access technologies. E.g., the pool of resources may contain points of access from LTE-A networks and networks in the IEEE 802.11 family, such as WiFi and WiGig; or, the pool of resources may contain spectrum that is licensed, licensed-exempt (such as from the ISM bands), or available through authorised shared licensing. The objective function for the optimization also becomes more complex once we consider specific goals of a service provider making use of the virtual network. The mapping from application-layer service goals to physical resources requires the selection of appropriate levels of abstraction, a non-trivial task given the different time scales involved.

Up to now, we considered the problem of how to optimally orchestrate a virtual network out of existing resources. A related question has to do with the planning and maintenance of the network infrastructure and spectrum. For existing operators, the emergence of the levels of sharing that we envision affects their decisions regarding new deployment of macrocells, small cells, and WiFi network for data offloading. It also affects planning decisions by new entrants to the market, such as a small operator that may want to deploy a dense small cell network in an area of strategic importance, and then make these resources available to the virtual network aggregators.

The implications of the wireless network architecture on the backhaul and the backbone wired network are also of key importance. The choice of an optical network architecture will affect (and, in turn, should consider) the way wireless access infrastructure will be deployed, the amount of coordination required among basestations, service locality considerations for caching, etc. Major architectural decisions for wireless and optical networks should be made in coordinated fashion.

New interfaces must be standardized to allow the virtual network to be built seamlessly out of disparate infrastructure deployed by multiple operators. Following past experience on network virtualisation, primarily done in the context of wired networks, hypervisors must be developed to abstract from the physical resources to the virtual network. A new suite of protocols must be created, or existing ones adapted, for management and control of these new virtual networks.

To take into account the large number of factors mentioned above, we envision the creation of new models of incentives and preferences for the different players in this architecture, including
infrastructure providers, virtual network architects, and service providers. Game theory, for example, has been used in the past to model network resource management problems while also incorporating economic incentive structures, and to predict the resulting decisions by the various players. Recognizing that economic considerations are critical to all major developments in telecommunication networks, we discuss those next.

B. Market and Regulatory Challenges

Network resource sharing has been driven by economic considerations, namely the increasing costs of infrastructure deployment and maintenance, coupled with the need for operators to meet rapidly growing customer demand for wireless data. The service provider-driven virtual wireless networks that we describe here are also motivated by market forces, and we expect would lower barriers to entry for new infrastructure providers, while allowing the service provider to more effectively extract value from the wireless network. However, what type of market would develop, or what kind of pricing structures, are still open questions. In this paper, we outlined some of the new roles that we envision in this market, but ours is an engineer’s view, and joint work is needed with economists to try to better predict the market consequences of such a shift in the way wireless service is offered, and to avoid potential pitfalls. For example, a first question might be whether sufficient incentives would persist for (real) operators to continue to invest in wireless infrastructure. Or would the architecture we present lead to highly concentrated markets at the wireless access level? In short, while we expect NwoB to increase competition in the wireless market and to benefit subscribers in currently under-served area, a detailed study is needed of the impact of the forms of sharing and virtualisation described here on customers, service quality, and prices.

Some amount of regulation is also needed to ensure competitive prices and service levels that, at a minimum, improve on current ones and, ideally, enable a host of new wireless services and applications. Policy changes may be required to enable the full virtualisation of the wireless network and should be based on public interest considerations. A way to frame the question is: what is the minimum amount of regulation necessary to support this level of resource sharing, and what public benefit is expected, as compared to the status quo? Similar public policy questions have been debated for the past decade in the context of dynamic access to spectrum, with concrete steps taken by regulators such as the FCC, in the US, and Ofcom, in the UK. We
advocate a broader discussion that encompasses infrastructure associated with multiple wireless access technologies, in addition to spectrum.

A change this radical in the way wireless networks are structured will bring its own market conflicts. The traditional network operators will reasonably want to protect their massive investments in network infrastructure and spectrum licenses. We expect, therefore, these changes to be largely driven by service providers, who need fast expansion of the wireless network capabilities at minimum cost to enable a new generation of services (think, for instance, of Google glasses and driverless cars). These service providers are currently best positioned to extract revenue from the network, both through traditional subscription services and by monetizing the information they can extract from usage patterns, through advertising and e-commerce. The traditional MNO will need to be nimble enough to change its business model according to the evolving market pressures, possibly by becoming both infrastructure/spectrum lease providers and virtual network service providers. This ability to venture into different parts of the value chain is not uncommon in today's network market (e.g., Netflix's recent entry into content generation, as well as content delivery).

Privacy is another area that merits careful consideration, as societies start to struggle with how much information about our network use it should be legitimate for companies and governments to collect and monitor. In the NwoB landscape, the service provider would get unprecedented access to metadata revealing user location, mobility patterns, and even health, as mobile devices add sensors to monitor heart rate and other indicators, in addition to any user provided content. Some discussions of privacy have already pointed to the emergence of the surveillance society driven by the data collecting companies [44], which in the NwoB case would correspond to the service providers. Regardless of the intent of the service providers, it is in our best interest to ensure that any future discussion of privacy envisions the possibility that the same service provider may hold data on both usage patterns (as is currently true of Google, Facebook, etc.) and rich information about user routines, health and instantaneous location. As a result, it is likely that a certain amount of regulation may be needed to provide a dividing line between sensitive and non-sensitive data, specify minimum security for storage of this data and monitor commercialization, resale, processing and usage of the data.
VII. Conclusions

The cellular network as a centrally deployed monolithically owned entity, designed and rolled-out in an organised fashion has been eroded on many levels. The increased focus on sharing of infrastructure, the role of the end-user in the deployment of small cells, the synergetic relationship with WiFi networks are all blurring the boundaries between ownership and control. The potential for disrupting further through enabling more fluid and dynamic access to spectrum and in doing so opening opportunities for new wireless services, coupled with advances in virtualisation allow us to reimagine the mobile network of the future. This paper presented one such vision that capitalises on these trends.

References


