Beyond-4G Cellular Networks: Advanced Radio Access Network (RAN) Architectures, Advanced Radio Resource Management (RRM) Techniques, and Other Enabling Technologies

Halim Yanikomeroglu
Carleton University, Canada

Jietao (Tony) Zhang
Huawei Wireless Research Department, China

ABSTRACT

4G has been the driving force behind a number of global research initiatives in the last few years (such as the WINNER Project). However, at the preparation of this paper (October 2008) we have formal documents outlining 4G objectives (such as IMT-Advanced), and tangible standardization efforts towards achieving those objectives (such as LTE-Advanced and 802.16j/m). Therefore, 4G is moving from the research phase to the development phase.

This paper discusses ***i)*** *advanced radio access network (RAN) architectures,* ***ii)*** *advanced radio resource management (RRM) techniques, and* ***iii)*** *other enabling concepts, to facilitate the provision of very high data rates with virtually ubiquitous coverage in beyond-4G wireless networks.*

I. Fundamentals

The data rate that can be supported by a single isolated base station (BS) is determined according to the following fundamental relation[1]:

$$R = nW\log_2(1+SNR) = nW\log_2\left(1+\frac{P_S}{N_0W}\right),$$

where

- R: rate (bits/sec),
- n: the minimum of the number of antennas at BS and wireless terminal (WT),
- W: bandwidth (Hz),
- SNR: signal-to-noise ratio,
- P_S: received signal power (Watts), and
- N_0: noise power spectral density (Watts/Hz).

Clearly, the rate can be increased by increasing n and/or W and/or SNR. Although increasing the rate through the MIMO architecture looks very attractive, it is not feasible to deploy a high number of antennas, especially at WT, due to a number of practical limitations. Increasing the transmit power to achieve a higher SNR is not very profitable due to the logarithmic relation, in addition to many other prohibitive factors associated with high transmit power levels, once again, especially at WT. Finally, bandwidth is scarce, and the licensed portion of it is very expensive. Besides, increasing the transmission bandwidth results in a linear decrease in SNR; therefore, even if the available bandwidth happens to be very high, P_S will also have to be very high to guarantee a sufficient SNR. This, in turn, means that high path-losses can not be tolerated; as such, WT and BS can not be too apart.

It is clear that the data rate, dynamics of which is governed by the above equation, is not unbounded in practical scenarios. If higher and higher rates are required in a given area, deploying a dense network of BSs with a very dense channel reuse scheme becomes inevitable. In that case, the multiple-access interference (co-channel interference) becomes a major concern and its impact has to be accounted in the rate equation as follows:

$$R = nW\log_2\left(1+\frac{P_S}{N_0W+P_I}\right),$$

where P_I denotes the interference power (Watts).

II. Cross-Layer Cooperative Communications

Wireless networks are evolving; one upcoming major architectural advance is the integration of the multihop capability. As a matter of fact, the concept of multihop communications has a long history in the context of infrastructure-less packet radio and, more recently, ad hoc networks. The main emphasis in this line of research has been on connectivity, i.e., establishing a route between any two WTs in an infrastructure-less network. Computer scientists and researchers from the wired network community played important roles in the conceptualization of packet radio and ad hoc networks.

Research on pure ad hoc networks predates internet; as such, the infrastructure-less nature of the pure ad hoc networks is not appealing anymore (except for certain isolated applications such as emergency & rescue and tactical communications). In this proposal, we focus on infrastructure-based multihop networks (such as cellular, WiFi, or WiMax, or even sensor networks).

[1] This well-known channel capacity theorem (revised here to incorporate the MIMO architecture) is for AWGN channels. The theorem constitutes an upper bound for the rate in wireless fading channels provided that fading is smoothed out through means such as diversity.

The area of "communications" has a long and rich history. In the early days the research community's attention has mainly been on the link level problems in wired (AWGN) channels. The communication and information theorists often transformed the physical layer problems into isolated mathematical problems through a high level of abstraction, and in most cases were able to solve them through elegant mathematical techniques.

Then, wireless communications became the area of focus. The research community's approach to wireless problems in the beginning was often similar to the earlier approach. In many cases, the classical problems of wired communications (such as coding, modulation, equalization) were analyzed in the wireless context (such as coding for fading channels) as pure physical layer problems.

However, the research community very quickly realized that there was more to wireless communications than merely the physical layer. It became apparent that "multiple access" and the associated "assignment (BS, power, channel)" and "radio resource management" issues were much broader and tangled cross-layer problems. Clearly, the goal had to change from optimizing a single link to optimizing the entire network; cross-layer design was at the centre of this approach.

The contemporary interest in cooperative communications started in late 1990’s. Once again, so far most of the research in this area has been performed by communication and information theorists in the physical layer (such as cooperative diversity protocols, diversity-multiplexing trade-off, and distributed space-time coding). “Cooperative communications” almost became synonymous with “cooperative diversity”, yet through cooperation so many other benefits can be achieved in addition to diversity against multipath fading. The real challenging issues, and the corresponding remarkable opportunities, exist in the MAC and networking layers. These challenges and opportunities have not received enough attention until recently. And, most of the recent interest has been originating from the computer science and networking researchers who import ideas from the ad hoc networks research; we feel that at least some these imported ideas are not relevant in the context of the envisioned infrastructure-based multihop networks. The communications and information theorists often do not have the wireless network perspective, and the computer science and networking researchers are often not sufficiently familiar with the physical and MAC layers. An “interdisciplinary” approach is needed in the study of cross-layer cooperative communications.

III. Advanced Radio Access Network (RAN) Architecture

There are at least three major issues associated with the deployment of a dense network of BSs:

1) Cost: A dense network of conventional BSs may be prohibitively expensive.

2) Coverage: Ubiquitous very high data rate coverage is an extremely challenging problem with the conventional radio access network architecture (such as that used in 3G networks) in which case rates decrease substantially at the periphery of BS coverage regions (the well known cell edge coverage problem).

3) Efficiency: The conventional cellular design with fixed (a priori) radio resource allocations and assignments is inefficient; this inefficiency becomes even worse in a dense network due to the increased interference.

The solution to the above three problems is to adopt an advanced **Radio Access Network (RAN)** coupled with advanced **Radio Resource Management (RRM)** algorithms which will enable an advanced RAN to operate at its full potential towards achieving cost-efficient and virtually ubiquitous very high data rate wireless networks.

The envisioned advanced RAN has the following elements (refer to Figure 1):

- Central Stations (CSs)
- Base Stations (BSs)
- Distributed Antenna Ports (DAPs)
- Fixed Relay Stations (FRSs)
- Terminal Relay Stations (TRSs)
- Wireless Terminals (WTs)

BS: This refers to the conventional BS which does not need any further elaboration.

CS: A group of neighboring BSs may be connected to a CS to facilitate the BS coordination/cooperation; as such, the presence or absence of a CS depends on the nature and extent of the BS coordination. That is, there may not be a need for an explicit CS if the information exchange between the BSs is sufficient (this will be the case if the corresponding BS coordination algorithms are distributed), or, if one of the BSs acts as a CS.

DAP: This can be considered as a low-cost micro- or pico-BS with limited functionality (and low transmit power) wired to a full-fledged BS where many decisions related to a set of DAPs are made. A DAP can be considered as a wired relay; "radio-over-fibre" is also often used in the literature to refer to the concept of deploying limited-functionality micro-/pico-BSs connected to a capable BS. How much functionality a DAP has is a design consideration; at the logical extreme, all the signal-specific processing (including detection) may be performed at the BS, in addition to the RRM decisions, provided that there is sufficient merit to justify this scenario (such as the creation of a network MIMO).

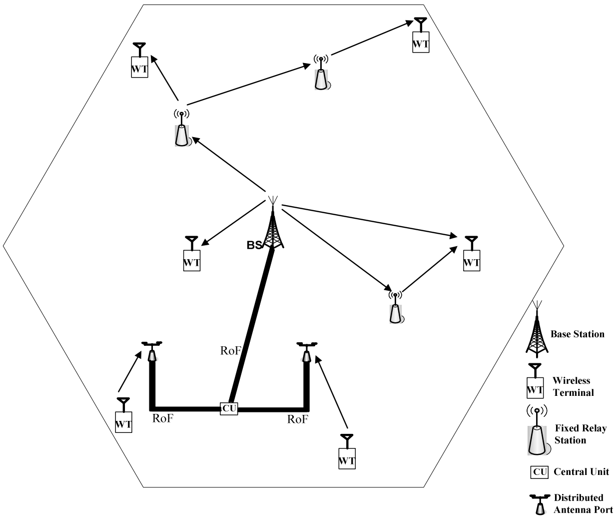

Figure 1: A typical cell demonstrating interaction among the network entities (DAP, FRS, WT, BS in the envisioned advanced RAN)

FRS: This refers to a relatively low-power relay deployed by the operator, preferably in a strategic location. An FRS, by its very definition, does not need a wired connection to the network (it does need, however, a wired DC source, unless some sort of an alternative energy source (such as solar) is employed). An FRS may communicate with other FRSs and multiple BSs; therefore, the topology is of mesh type.

TRS: This refers to a wireless terminal acting as a relay.

WT: This refers to the terminal of an end user. Some WTs may act as TRSs when need arises.

Comparisons between the RAN Elements

- A BS and a DAP have the common feature that they are connected to the backhaul through the wired medium (fibre).

- On the other hand, a BS and a DAP are different in the processing capability and functionality; we perceive DAPs as wired relays with limited functionality (in comparison to full-fledged BSs) in order to achieve cost-effective deployment.

- In contrast, FRSs and TRSs do not have wired connections.

- An FRS is likely to have much more functionality and capability (signal processing, MIMO, security, RRM, etc.) in comparison to a TRS.

- It is also important to highlight that an FRS has a continuous DC power source from an outlet unlike a TRS which is battery-powered.

Co-existence of DAPs and FRSs

The main argument regarding FRSs has been the deployment cost advantage (in comparison to, for instance, micro-BSs) since FRSs do not require the wired backhaul. On the other hand, fiber penetration is relatively high in certain parts of the world (such as South East Asia), and this penetration will be at a much higher level in the years to come. In such cases, it makes sense to utilize low-cost DAPs in addition to FRSs. Smart RRM algorithms will enable the concurrent operation of DAPs and FRSs.

IV. Dynamic Radio Resource Management – Partly Centralized and Partly Distributed

In conventional cellular networks each WT is assigned to one BS; in such networks handoff is an undesirable event implemented through hysteresis to prevent the ping-pong effects.

Moreover, in the conventional cellular design, a channel is not supposed to be used more than once in a cluster. In the conventional non-CDMA networks, the cluster size is almost always greater than one. In order to avoid cell planning, there has been an increasing interest in adopting the single-frequency concept in OFDMA networks as well which will allow the reuse of a channel in every cell whenever the conditions allow, i.e., the cluster size can be as low as one (as a matter of fact, the resources can be reused in every sector).

We envision a very ambitious wireless mesh network architecture to support very high data rates with highly bursty and geographically non-uniform traffic patterns. The considered air interface is OFDMA-based. Clearly, the use of the radio and network resources in the most efficient way is of paramount importance. Any fixed assignment or routing association will be inefficient as it will not be able to fully exploit the dynamic conditions in the network. Towards that end, in the envisioned mesh network there is no, or minimal (only whenever necessary), a priori channel allocations, no a priori radio resource assignments (for instance, a channel may be reused multiple times in the vicinity of a BS whenever possible), and no fixed routing associations; all such decisions are made dynamically and opportunistically (in other words, everything is for grab).

A WT's data may be routed through different DAPs or FRSs to a BS or even to different BSs (this mesh architecture may be considered as the evolved version of the CDMA soft handoff concept); moreover, a WT's data over multiple subchannels may be sent through multiple routes to the BS or CS. A great level of diversity gain against shadowing (and, if the RAN is advanced enough, against multipath fading as well) can be achieved, and load balancing against congestion (which may occur as a result of the highly variable and bursty traffic) can be attained through opportunistic routing.

It should be noted that the notion of cell becomes rather fuzzy in the articulated advanced RAN. Through the mesh topology, a WT's signal may be routed through FRSs to a further away BS or DAP whenever there is merit (such as avoiding congestion).

A salient feature of the envisioned advanced RAN is that while the activities of some of its elements (BSs connected to a CS, and DAPs connected to a BS) are expected to be controlled by highly centralized RRM algorithms, some of its other elements (mainly TRSs) are likely to rely on highly distributed RRM schemes.

a) Centralized RRM Algorithms in Relay Networks for Coordination

RRM in OFDMA-based cellular multihop networks is a very active research area due to the potentially high performance returns. The "coordination" dimension brings in new opportunities as well as various new challenges.

As discussed earlier, advanced RRM algorithms are essential to get most out of the advanced RAN architecture considered in this paper. The advanced RAN and advanced RRM are inseparable concepts. If there are no advanced RRM algorithms to support the advanced RAN architecture, then there is even no point in deploying such complex RANs. However, the advanced nature of RAN makes the associated RRM a highly complex task; but as stated before, the potential benefits from the proper operation of these RRM algorithms are truly substantial.

It is imperative that the developed RRM algorithms must be efficient and must have reasonably low radio overhead. To the best of our knowledge, there is no RRM literature for the type of advanced RAN considered in this paper which includes CSs, BSs, DAPs, FRSs, and TRSs.

b) Distributed RRM Algorithms in Relay Networks for Cooperation

In addition to the requirements listed above for the centralized RRM algorithms, the distributed algorithms must also be highly robust; this is an essential requirement from the network point of view. Otherwise, an unstable algorithm may cause great damage not only to a particular TRS or WT but to the entire network.

c) Advantages and Disadvantages of Centralized and Distributed RRM Algorithms

In centralized resource allocation algorithms, assignment decisions are made at a centralized controller which may have partial or full knowledge of the link conditions in some segment of the network. In the logical extent, a central controller may instantaneously know the conditions (channel-state information, CSI) for all links in the entire network; this will constitute a *genie-aided* network to be considered for determining the performance bounds.

Due to the spectrum overhead, signaling and computing requirements, and latency, centralized algorithms may not be feasible in certain networks and/or for certain applications (especially for those which are highly delay-sensitive). This problem is exacerbated in a fast-fading environment as it is very difficult to acquire and assemble link information from all nodes at a central unit due to the rapidly changing channel conditions. In addition, if the collected link information quickly becomes outdated, this may result in a highly inefficient operation.

The advantages and disadvantages of purely centralized algorithms may be summarized as follows:

- Advantages
    - Potentially superior performance

- Disadvantages
    - Computational complexity and cost
    - Spectrum overhead (for CSI transmission)
    - Wired backbone overhead
    - Latency
    - Scalability

On the other hand, in distributed schemes each transmitting node can make resource decisions (frequency carriers, time, power, modulation and coding level) based on the limited information it has. This information may be acquired by the decision making node explicitly through monitoring the control channels.

Here are the advantages and disadvantages of purely distributed algorithms:

- Advantages
    - Low latency
    - Low complexity
    - Low overhead
    - Scalability

- Disadvantages
    - Potential performance loss
    - Robustness, stability, and convergence concerns

Fully-centralized and fully-distributed algorithms constitute the two ends of the spectrum. In many cases, there will be merit in using some hybrid algorithms; for instance, limited CSI may be broadcasted in control channels, and some entities in the network may make autonomous or semi-autonomous decisions based on this information.

V. Enabling Concepts and Analytical Tools

Optimization: As well known, optimization is a major tool in every advance RRM scheme, in both centralized and distributed domains. Since this is already clear, there is no need to elaborate on it any further.

There is also no need to elaborate on the other conventional concepts and analytical tools used in layers 1, 2, and 3. Next, we highlight the less conventional concepts and tools which are mainly relevant in the context of distributed RRM.

Cognitive Radio: TRSs utilizing distributed RRM algorithms should have enough cognition and intelligence to make a number of important decisions including

- with which TRSs to cooperate, to what extent, and in which capacity;

- when to transmit, at which subcarriers, and at what power levels;

- which DAP or FRS or BS to connect to.

Cognitive radio capabilities will enable a TRS to make the above decisions either autonomously or with minimal assistance from FRSs and BSs.

Game Theory: The employment of some game-theoretic concepts in RRM has been addressed in the literature during the last ten years. Game theory is mainly applicable to scenarios in which a set of uncoordinated WTs make autonomous RRM decisions. Since robustness is one of the most important requirements for any distributed RRM algorithm (in particular in the licensed spectrum bands), the incorporation of non-cooperative game-theoretic algorithms (used in the presence of selfish users) may especially be important in preventing any catastrophic situation from the interference point of view.

Machine Learning, Dynamic Feedback Control Theory, and Artificial Intelligence: It is clear from the above discussions that learning the environment and making good decisions (often with very limited information) in a dynamic and adaptive manner are the key concepts in distributed RRM. It is interesting to observe that some of these concepts have been studied extensively in very different contexts in the fields of machine learning, dynamic feedback control, and artificial intelligence. This potential synergy should be explored further to determine whether some of the existing results in these different fields could be utilized with minimal touches in developing distributed RRM algorithms.

It is worth noting that the identification of such synergies, if they do exist, will be an important achievement by itself as this will help in bridging a number of research areas. Currently, there is almost no literature in employing concepts such as machine learning, dynamic feedback control, and artificial intelligence, in RRM problems.

ACKNOWLEDGEMENT

The authors would like to thank Dr. Abdulkareem Adinoyi (Carleton University) for fruitful discussions during the preparation of this paper.