ON THE SCALABILITY OF RELAY BASED WIRELESS NETWORKS

Adrian Florea and Halim Yanikomeroglu
Broadband Communications and Wireless Systems (BCWS) Centre
Dept. of Systems and Computer Engineering
Carleton University, Ottawa, Canada
E-Mail: {aflorea, halim}@sce.carleton.ca

***Abstract*[*] — In this paper, we argue that while relays will surely be an essential component of future infrastructure-based wireless networks, their indiscriminate use is an inefficient way to allocate the limited radio resources available. The most efficient method to allocate the radio license is when the number of relays is minimized, while still achieving the desired coverage objectives. Under very general assumptions, we show that there exists an optimal number of relays which maximizes the aggregate spectral efficiency for wireless connections within a base station coverage area.**

Keywords-relays; wireless networks; mesh networks; wireless networks; multihop networks; resource management.

I. INTRODUCTION

The architecture of the present day cellular networks is not suitable for meeting the stringent requirements envisioned for 4th generation (4G), or beyond IMT-2000, wireless systems. Economically feasible solutions are likely to be based on some form of multi-hop relaying allowing uniform coverage at very high data rates and reducing the required number of expensive cell sites [1]-[3].

It is envisioned that relays can be implemented as low cost low footprint self configuring devices, which could be inexpensively deployed in large numbers. In this paper we investigate at a very general level the system architecture limitations of a wireless network deployed using potentially large number of relays.

We consider a possible system architecture for a 4G cellular network, as shown in Figure 1. The network is organized into two overlaid systems as follows:

- The *wireless access system* (WAS) has the role of providing wireless connectivity at the required bit rate and grade of service to the mobile user terminals; this is the mobile air interface. The access system is composed of mobile user terminals (MT), fixed radio stations (relays and base stations) directly facing and connected to the mobile user terminals, and radio resources allocated to the wireless access system.
- The *wireless feeder system* (WFS) has the role of transporting the traffic generated by the access system elements to / from the wired backbone network. The WFS is composed of fixed relay stations (FRS), base stations (BS) and radio resources allocated to the wireless feeder system. It should be noted here that, regardless of the implementation method, logically, WFS is an essential component present in any wireless network using multi-hop links.

As defined above FRSs and BSs have a dual role: on one side, they have an “access” component which supports the last-hop mobile connectivity – the mobile air interface; on the other side, they also include a “feeder” component which supports the first-hop(s) fixed link(s) between BSs and FRSs. Logically, WFS carries traffic data between the WAS elements and the gateways to the wired backbone. For the purposes of this study, the logical difference between a BS and a FRS essentially is that the BS has a wired connection to the backbone network (the BS includes a gateway to the wired Internet), while the FRS does not. The BSs and MTs are terminating points for the wireless traffic, while the FRSs have a “bridging” role.

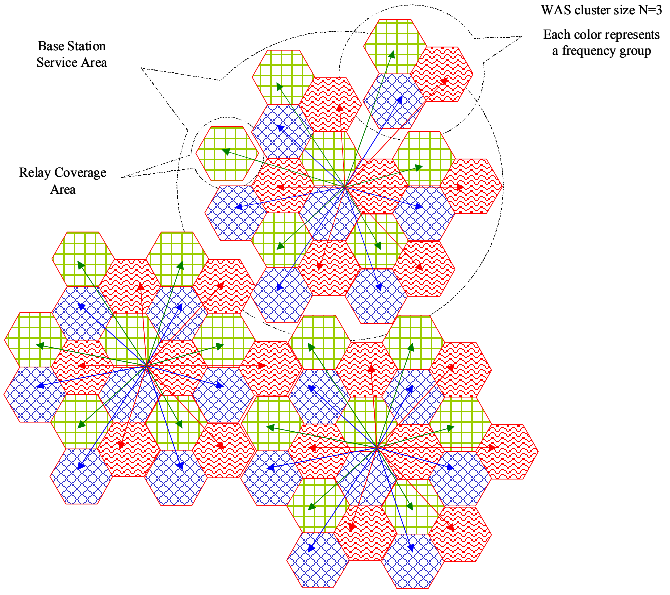

Figure 1. Proposed system architecture for a 4G cellular system

---

[*] This work was supported in part by the Natural Sciences & Engineering Research Council of Canada (NSERC) under participation in project WINNER (Wireless World Initiative New Radio) - www.ist-winner.org

This full text paper was peer reviewed at the direction of IEEE Communications Society subject matter experts for publication in the WCNC 2006 proceedings.

The main cost associated with a BS is assumed to be the cost of the wired data connection to the backbone or Internet (this involves a potentially considerable initial capital cost, as well as a monthly recurring cost). This is the reason of employing FRSs instead of BSs, whenever possible, as a means to reduce the total capital and operational costs of the network.

The WAS and the WFS could use similar physical/ MAC protocols and I/F interfaces, in which case the entire system is *homogeneous*, or use dissimilar physical/ MAC protocols and I/F interfaces, in which case we have a *heterogeneous* system [4].

A number of FRSs is connected to one BS, which could be either omnidirectional or sectorized. For simplicity and without reducing the generality, in the following discussion the BSs are considered non-sectorized. The results obtained below can be easily expanded to the sectored BS scenario. The maximum possible radius of the BS is determined by the link budget of the WFS.

Also, in the present study the FRSs are assumed to be non – sectorized (beside the directional antenna(s) used for the feeder system, each FRS has one antenna for access, which can be directional or omnidirectional, as dictated by the specific local conditions). The FRS maximum coverage radius is determined by the WAS link budget.

In the following section, we analyze the radio resource allocation for the system architecture introduced here, particularly the division of available resources between the feeder and access portions of the network. By resources we understand bandwidth and/or time slot allocated to particular network elements. Below, we have used interchangeably the terms “bandwidth” and “radio resource”.

II. RADIO RESOURCE ALLOCATION TOACCESS AND FEEDER SYSTEMS

A. Radio Resource Partitioning

For simplicity, we assume that the feeder link is implemented in one single hop (direct radio link between each FRS and BS). For the purposes of this discussion, the single hop assumption does not limit the generality of the result.

The following notations are made in the context of the system architecture shown in Figure 1.

N = cluster size for WAS ($N = 3$ in example in Figure 1.)

η_a = average spectral efficiency for the WAS

η_f = average spectral efficiency for the WFS

n_r = number of FRSs “fed” from one BS

b_a = amount of resources (bandwidth) allocated for access portion of each FRS

b_A = total resources (bandwidth) allocated for access for all FRS serviced from one BS

b_f = amount of resources (bandwidth) allocated for feeder portion to each FRS

b_F = total resources (bandwidth) allocated for feeder for all FRS serviced from one BS

Λ = total available licensed radio resource, (bandwidth x time)

C = aggregate throughput to be provided within service area of each FRS

For each FRS, the maximum data rate at the link layer must be the same for both the access and feeder links (otherwise very large buffers would be required at FRSs):

$$b_a = b_f \frac{\eta_f}{\eta_a}. \tag{1}$$

The total frequency spectrum required for the entire system is composed of the spectrum required for WFS and spectrum required for WAS. While for the access portion the frequencies can be reused with a reuse pattern of N, the same is not valid for the WFS, where all FRSs communicate with a single transceiver at the BS, so no reuse is possible. This creates a *bottleneck* which occurs at the BS (where the gateway to wired backbone is located), regardless of the particular implementation of the WFS. This can be expressed as:

$$\begin{aligned} b_F &= n_r b_f, \\ b_A &= N b_a, \\ b_F + b_A &= \Lambda. \end{aligned} \tag{2}$$

From (1) and (2) we obtain the expressions for the bandwidth allocated for feeder and access portions of each FRS

$$b_f = \frac{1}{N\frac{\eta_f}{\eta_a}} \cdot \frac{\Lambda}{1 + \frac{n_r}{N\frac{\eta_f}{\eta_a}}}, \tag{3}$$

$$b_a = \frac{1}{N} \cdot \frac{\Lambda}{1 + \frac{n_r}{N\frac{\eta_f}{\eta_a}}}, \tag{4}$$

respectively. Then, the total spectrum allocations for WAS and WFS are

$$b_A = N b_a = \frac{\Lambda}{1 + \frac{n_r}{N\frac{\eta_f}{\eta_a}}}, \tag{5}$$

$$b_F = n_r b_f = \frac{\Lambda}{1 + \frac{N\frac{\eta_f}{\eta_a}}{n_r}}. \tag{6}$$

respectively.

Observations:

1. As defined, $b_A + b_F = \Lambda$.
2. For $n_r = 0$ (no relays): $b_A = \Lambda$, $b_F = 0$; i.e., all resources are allocated for WAS.
3. In the example of a case where n_r=25 (access radius aprox. 5 times smaller than the feeder radius), $N = 3$, $\eta_f/\eta_a = 1.5$, the resulting resource allocations are $b_A = 0.35\ \Lambda$ and $b_F = 0.65\ \Lambda$; That is, only about one third of the radio resources is used for access (the rest for feeder).

B. *Radio Resource Partitioning Dependency on the Number of Relays*

In this section we study the variation of the bandwidth allocated to WAS and WFS function of the number of relays in the BS, and we show that there exists an optimal number of relays for which the aggregate network spectral efficiency is maximized.

We can rewrite (2) as

$$\Lambda = b_A + b_F = b_A\left(1+\frac{b_F}{b_A}\right) = b_A\left(1+n_r\frac{1}{N}\frac{\eta_a}{\eta_f}\right). \qquad (7)$$

The expression (7) shows that the total required radio resources Λ (entire radio spectrum license) grows linearly with n_r, while, as shown in Figure 2, the amount of the radio resources allocated for access to all FRSs, b_A, remains constant, since the cluster size N does not change. In essence, the difference between the variation of resources allocated to WAS (constant) and WFS (growing linearly) is determined by the fact that no reuse is possible for the feeder system, since all FRSs in BS service area communicate with the same BS transceiver.

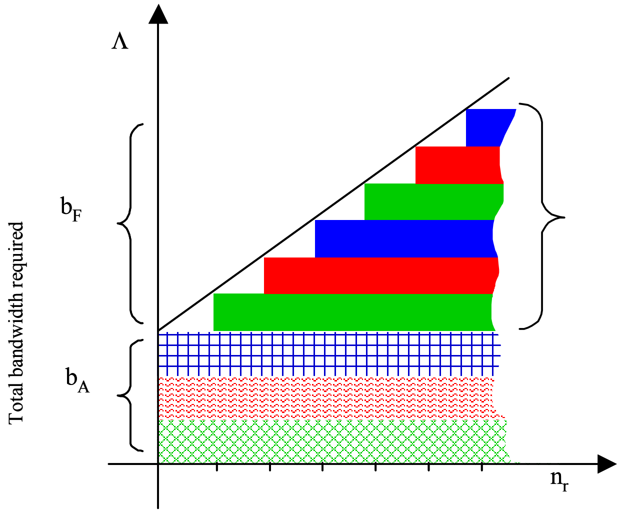

Figure 2. Required license Λ versus the number of FRS in each BS

When the number nr of FRSs in a cell increases to large values (tens), the most part of available resources are allocated to the feeder system, while the access system is allocated only a small proportion of the total system resources, as plotted (using the same assumptions as in Observation 3 above) in Figure 3. Given that this small fraction still has to be sufficient for the high access data rates envisioned, the result is a large increase in total resources required for the entire system (most of which are used for feeder). On the other side, if the number of relays is too small, there will be gaps in the coverage, where communication at the required data rate will not be possible. A larger number of relays means a smaller coverage area for each relay, leading to better average SNR for the access link, which can be exploited through higher order modulation and coding to obtain a better spectral efficiency for the WAS. Increasing the number of relays beyond the level at which all coverage gaps are eliminated, (which depends on the system parameters) will not be accompanied by an important increase in the spectral efficiency for the WAS, not fully compensating for the bandwidth "lost" to additional allocations to the feeder system. This scalability issue can only be alleviated, but not eliminated, by using sectorization at the BS, due to the limited frequency reuse which can be employed between collocated sectors. In conclusion, increasing the number of relays beyond a certain point is inefficient. This is plotted[1] in Figure 3, where the average net spectral efficiency (defined as the ratio between the access throughput and the total system bandwidth for access and feeder) at the edges of the relay coverage areas is shown for BS radius R = 800 m and R = 500 m. We can see that for a smaller BS radius, the number of relays required to fully cover the entire service area is reduced, and the achieved spectral efficiency is higher. Of course, these advantages do not come without costs: an extra 156% additional base stations with R = 500 m are needed to cover the same area covered by base stations with R = 800 m.

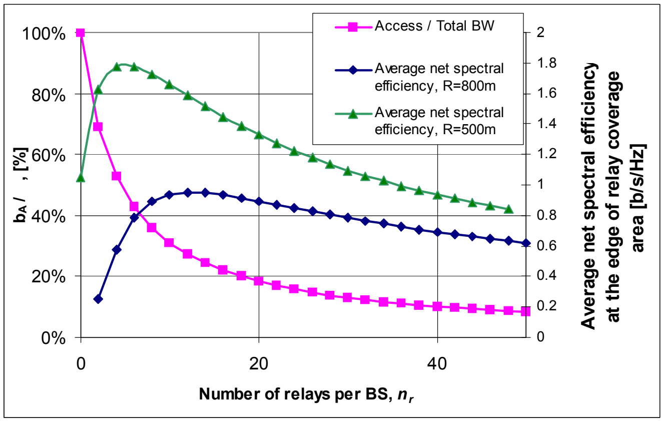

Figure 3. Bandwidth allocated for access vs. number of relays in a BS, n_r

---

[1] The system parameters used to plot the average net spectral efficiency are: two-slope propagation model with d_0=2m and path loss exponent 4; the link spectral efficiency depends on link SNR following the Shannon limit formula; channel bandwidth 25MHz, transmit power 1W; transmit and receive antenna gain 10dB; receive noise figure 5dB.

C. A Different View Of The Same Problem

Under the system model described earlier, the maximum possible throughput assigned for one user would be limited by the entire bandwidth assigned for WAS, b_A. (i.e., the respective user would get assigned the entire access resources within its cluster of N FRSs, which is b_A. These resources may be channelized through one serving FRS, or they may be used on some cooperative system by multiple FRSs). The maximum throughput per user would then be

$$C = b_A \eta_a , \tag{8}$$

and we can write equation (7) as

$$\Lambda = \frac{C}{\eta_a}\left(1 + \frac{n_r}{N}\frac{\eta_a}{\eta_f}\right). \tag{9}$$

Now, in order to build a system with a maximum throughput per user C (C is assumed 100 Mb/s in [5]), the feeder system has to be able to accommodate a much larger throughput, by a factor proportional with the ratio of spectral efficiencies for the two systems and the ratio n_r/N. The BS aggregate throughput is then the maximum throughput offered by the feeder system, exceeding C (100 Mb/s) many times.

However, in [5] the aggregate BS throughput requirement is assumed to have the same value as the maximum throughput per user, i.e., 100 Mb/s. Under this assumption, if one user requests and is granted the highest possible throughput C, then no other users within the BS area will be offered any services. This requirement is less difficult to achieve, since in this case the feeder system has only to accommodate aggregate data rates up to C. In the same time, each FRS cluster is provisioned on a permanent basis with enough radio resources to handle access rates up to the required C. With only one user accessing the system at C = 100 Mb/s, all FRS clusters except the one providing service to the sole user will be idle (because the feeder system is fully loaded), which is a big waste of radio resources. The "wasted" amount is again proportional with the ratio n_r/N.

III. CONCLUSIONS

Although the arguments above are based on the assumption of a feeder system based on single-hop feeder links, the result can immediately be extended to the case of multi-hop feeder links. To do this, we consider b_A being actually the bandwidth needed by only the first hop adjacent to the BS, for all multi-hop links to all relays, and we can apply directly the same argument as before. In the case of multi-hop feeder links, the total bandwidth allocated to the feeder system is actually equal or larger than b_A, leading to a potentially even more inefficient deployment scenario. In addition, since the "bottleneck" occurs at the BS transceiver (from an architectural point of view this is the interface between the wired and wireless portions of the network), it will occur regardless of the particular relaying architecture employed; hence, the results above could also be extended to the mobile relaying scenario.

In conclusion, for an efficient network deployment making the best use of the available limited radio resources (frequency license), the number of relays employed should be kept to a minimum, while still ensuring good coverage of the service area. This suggests that advanced methods (for example MIMO systems, relay cooperation schemes, etc) need to be utilized at the physical layer in order to increase the coverage range of the wireless access system, and reduce the number of relays.

REFERENCES

[1] R. Pabst, N. Esseling, and B. Walke, "Fixed relays for next generation wireless systems - system concept and performance evaluation", *Journal of Communications and Networks, Special Issue on "Towards the Next Generation Mobile Communications"*, no. 2, pp. 1-12, June 2005.

[2] H. Yanikomeroglu, Cellular multihop communications: infrastructure-based relay network architecture for 4G wireless systems", *the 22nd Queen's Biennial Symposium on Communications (QBSC'04)*, 1-3 June 2004, Queen's University, Kingston, Ontario, Canada.

[3] R. Pabst, B. H. Walke, D. C. Schultz, P. Herhold, H. Yanikomeroglu, S. Mukherjee, H. Viswanathan, M. Lott, W. Zirwas, M. Dohler, H. Aghvami, D. D. Falconer, and G. P. Fettweis, "Relay-based deployment concepts for wireless and mobile broadband radio", *IEEE Communications Magazine*, vol. 42, no. 9, pp. 80 - 89, Sept. 2004.

[4] M. Lott, M. Weckerle, and M. Siebert, "Scheduling in wireless multihop networks", *International Symposium of Performance Evaluation of Computer and Telecommunication Systems (SPECTS'04)*, San Jose, California, 25-29 July 2004.

[5] W. Mohr, R. Luder, and K.-H. Mohrmann, "Data rate estimates, range calculations and spectrum demand for new elements of systems beyond IMT-2000", *The 5th International Symposium on Wireless Personal Multimedia Communications (WPMC '02)*, 27-30 Oct. 2002, vol. 1, pp. 37 – 46.