A message oriented phone system for low cost connectivity

R.J. Honicky Omar Bakr Michael Demmer Eric Brewer
UC Berkeley
{honicky, ombakr, demmer, brewer}@cs.berkeley.edu

Although mobile phone usage has proliferated in urban areas of developing countries, many rural or poor users remain completely unconnected, particularly in Africa. In this paper, we advocate an alternative "voice message mostly" mobile phone architecture to reach the next *billion users. Using initial analyses of mobile phone usage and rural power data, we outline the potential benefits of switching to an asynchronous model including improved utilization, increased effective coverage, better perceived service, and (most significantly) cost reduction. By leveraging these benefits, we believe a system based on voice messaging can make remote rural villages and the urban poor viable telecommunications markets.*

1 Introduction

Cell phones are by far the most successful modern technology in developing regions, with hundreds of millions of users in both India and China. Africa, which has wired telephony rates below 2% [9], is now the world's fastest growing cellular market with about 9–10% penetration. With continued adoption of current technologies, one can expect roughly 3B cellular users within this decade.

Yet even this remarkable expansion has its limitations. Many people living in rural areas have limited or no access to cellular coverage. Other users, both rural and urban, may have access to services, but only with unaffordable fees. Our aim in this project is to address these users and expand cellular usage by one more tier, to the *next* billion.

The first critical factor to this expansion is better penetration into rural areas. The current cellular infrastructure is a largely urban and semi-urban phenomenon, mostly due to economic reasons. The combination of limited purchasing power and low user density makes deployment of rural base stations uneconomical [12]. As a case in point, even the best known "rural" cellular system, Grameen Telecom [7], avoids purely rural base stations. Instead, exploiting the high population density of Bangladesh, it relies on road and rail base stations intended for middle class users to cover rural areas.

Cost is the second key factor limiting the reach of current wireless technology. Although low-cost phones have dropped below US$25, air time remains expensive, particularly in Africa, where rates are typically US$0.15–0.50 per minute (e.g. see MTN Uganda [13]). For consumers with a household income of no more than a few dollars per day, these costs are well out of reach.

Additional evidence of the impact of airtime costs is the phenomenon of "beeping," in which a poor user calls someone, lets it ring once, then hangs up. The receiver then gets a "missed call" notice with the caller's number and by convention knows to call back [4]. Since most carriers do not charge for incomplete or received calls, the original "callee" ends up paying for the call. Thus even when users can afford to own a phone, they make some effort to avoid paying for airtime. Still other users own only a SIM card and borrow a phone to make calls (typically for beeping).

We argue that neither better rural penetration nor significantly reduced airtime costs are likely to happen in the near future simply through continued refinements to current technology. Therefore, our central idea is to build a phone that is "voice message mostly," by which we mean that although the phone can make normal calls, its normal usage is to send and receive voice messages. Because such a system is based on asynchronous messaging, it addresses the two key barriers of penetration to the rural and urban poor:

First, it can extend the *effective* coverage range by queuing messages on the phone and leveraging user mobility to carry them into and out of connectivity, thereby reducing the required radio coverage area. Also, it should enable significantly lower airtime charges, as communication can be scheduled during times that would otherwise be idle. For existing deployments, this shift means that voice message traffic has little or no marginal cost to the carrier. For new deployments, it means that designers can tailor power and channel capacity for the average case instead of the peak case, and thereby reduce the scalability requirements for the infrastructure. In both cases, this savings can be passed on to the consumer.

2 Feasibility

Perhaps the most immediate question regarding the feasibility of our proposed system is how readily users will adopt an asynchronous voice-messaging model. Other asynchronous modalities such as E-mail and SMS are in widespread use in the industrialized world, so it seems clear that this type of communication has benefits, even when other modalities are readily available and affordable. It is reasonable to suppose that voice messaging could fit in alongside those other approaches. In fact, for users with limited literacy or technical competence, voice messaging is likely to be much easier

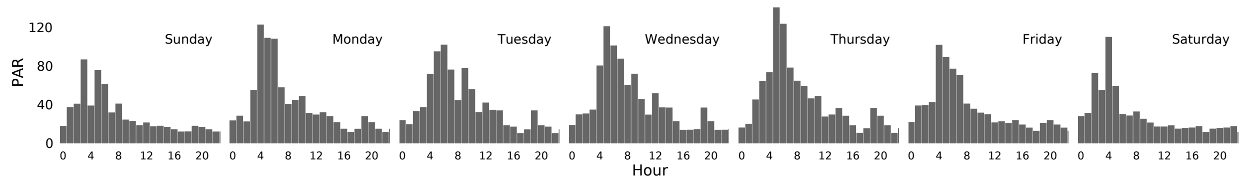

Figure 1: The Peak-to-Average Ratio (PAR) of calls in the MIT study, grouped by hour of the day. The average PAR during the busiest hour (17:00) was 13, while the during the least busy hour (5:00) it was 107. Only about 2% of all the hours in the study (about one hour for every two days) had a PAR less than 4, and the overall PAR for the entire study was 35.6, showing the large amount of unused system capacity.

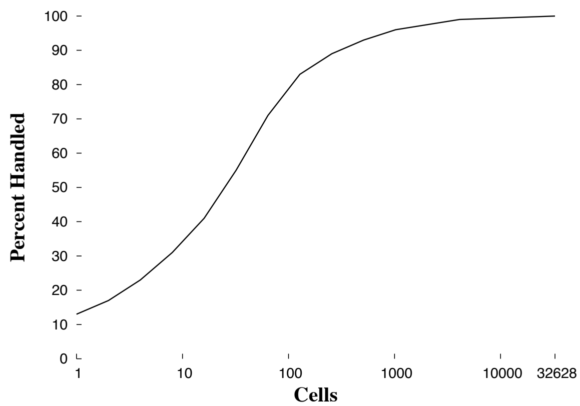

Figure 2: Cumulative distribution of the connections handled by a given number of cells (logarithmic axis).

to use and more intuitive than text-based communication.

Yet perhaps the most compelling reason why we believe in the viability of this proposal is that our target user population is currently unserved by existing communication services. It seems highly doubtful that this is due to lack of demand, given the widespread growth of cellular systems to other populations, hence it must be due to lack of availability and/or affordability. We now turn to two observations of user behavior and mobile phone usage that motivate why we think a voice messaging model can help address these barriers.

2.1 User Locality

The first observation is that for many groups of users, both in developing countries and elsewhere, there are a few key locations where most (in some cases all) individuals within the group travel to and from on a regular basis. In the case of rural populations in developing regions, such locations are likely to include markets, churches, government offices, and village centersxf [14].

To gain further insight into this pattern, we analyzed cell phone usage data gathered by the Reality Mining project from MIT [5]. This study logged 89 students' phone activity between January 2004 and June 2005, including records of which cell phone tower each phone was connected to over time. We ranked all towers based on the number of connections they handled, and show the cumulative percentage of connections handled by various-sized subsets of the towers in Figure 2.

Although the tower that a phone connects to depends on several factors, it is useful as a rough approximation of users' locations. To that end, these data support the hypothesis that people tend to be in the same general locations most of the time, as over half of the phone connections were handled by only a few dozen towers.

This type of mobility pattern leads to the opportunity to provide only intermittent coverage but still reach the majority of users. Moving to asynchronous messaging enables a store and forward message routing system that can queue messages when disconnected, waiting for the sender or receiver to travel into range. This means that if a large number of users regularly go to same set of key locations, then providing connectivity at those places effectively extends the coverage to all the users that visit.

2.2 Unused Capacity

The second observation is that traditional communications systems are designed to handle peak usage levels, yet few systems run at peak load all of the time. As such, there is often a significant amount of unused system capacity.

Figure 1 shows a plot of the peak-to-average calling ratio (the maximum number of simultaneous calls divided by the average for a given period) for all calls from the MIT trace, grouped by the day of the week and hour per day. The data show that usage is bursty, with 10–100x higher peaks than the hourly average.

Since there is a direct correlation between system capacity and cost, this graph shows how much extra expense carriers endure to provide service. For systems targeted towards voice calls, there is little one can do to reduce this burden, as there is no way to predict when users will decide to make a call, and the system must to be prepared to handle it.

On the other hand, basing the system on voice messaging opens up a degree of freedom in terms of when data is transferred, since a scheduling algorithm can wait for an idle period before transferring a message. This means that voice message traffic can be sent in otherwise unused periods, with little (if any) added cost to the carrier. Also, by evenly distributing the workload throughout the day, then a system with much lower peak capacity can handle the same total amount of traffic, at much less cost.

3 Proposed System Model

We now describe our proposed system in more detail.

3.1 Voice Messages

Voice messages are stored on a per-user card in the phone (such as an SD card). Therefore they can be recorded and played back without support from the infrastructure, and thus without requiring the phone to be in connectivity. These cards also function as SIM cards in that they contain a unique ID, imply a permanent number, and offer the ability to add minutes using pre-paid scratch-off cards. Although these cards will require more capacity to store messages than traditional SIM cards do, a few megabytes per user should be sufficient.

To check for new messages or to send queued ones, the phone (not the user) periodically polls the infrastructure. There are two different polling rates: polling for a signal and polling for messages once in coverage. Polling for a signal requires little power, so once per minute should still allow for long battery life. Assuming the phone initially checks for messages when it detects a signal, the polling rate for messages can be lower, perhaps every five minutes. Because the phone does not need to be always on to await a potential incoming call, this low duty cycle should lead to much longer battery life.

Typically, messages are routed from a phone to the infrastructure, where they are queued awaiting delivery to another phone. In some cases though, phones may be able to make an ad-hoc connection to each other and transfer messages directly, without assistance from the network. Also, in cases where one party is using a traditional phone system, messages would be routed to and from a gateway where they would be played out as a voice call, sent as an MMS, or converted to an email attachment.

3.2 Services

The phones in our system should also be able to make normal voice calls, for three main reasons: First, the marginal cost of supporting calls on the phone device is likely zero, since the capabilities are deeply integrated with any chipset on which we would build. Second, this ability allows the handset to be marketed as an enhanced normal phone, not a degraded one; historically, poor users are reluctant to accept low-cost but crippled devices due to reasons of dignity. Finally, there are important cases in which live calls are worth the cost, such as emergencies.

Explicitly supporting the "beeping" model also has a few advantages. Fundamentally, it clearly conveys to the device and the carrier the *intent* of the user. For the device, this means that it can enter "standby" mode to await an incoming call, and potentially alert the user if they leave coverage before receiving the expected call-back. For the carrier, the explicit intent enables a lower-cost signalling mechanism than a missed call would, and leads to more accurate usage statistics. Of course, it is critical that this service be easy to use and free of charge, otherwise users will revert back to their original behavior.

There are also opportunities for other value-added services. Carriers could provide "high priority" messages with faster delivery, "return receipt" services, news and information subscriptions (such as cricket or football scores), and timed delivery. Delivery at a particular time is useful when leaving a message for a normal phone, since the message will play out over a live call. Since many calls from the rural poor are to relatives abroad, it is somewhat important that these calls are made at a time presumed to be convenient for the receiver.

3.3 Deployment Scenarios

We envision two complementary deployment scenarios with different cost implications:

Leverage existing base stations: With software-only upgrades to current base station hardware, we can add a voice messaging service to already-deployed cellular networks and thereby extend the functionality and profit of the systems. The only new equipment required should be the gateway system (installed only in the central office) to convert between voice messages and live calls or e-mail, and to store messages in transit. It is in this context that we refer to "zero-marginal cost" deployment, as the system can schedule the movement of messages during otherwise idle base station time, increasing its utilization and profit without needing additional capacity.

Voice messaging base station: We can also extend the coverage area by developing new low cost base stations that support voice messaging and a few concurrent live calls. This design results in significant cost savings due to the lower capacity, lower power (enabling solar), reduced backup power, and less complexity due to reduced need for high availability. These systems should enable carriers to incrementally add rural coverage; as demand grows, these micro base stations may be upgraded to full base stations, and then redeployed on the new fringe. This path is thus a middle ground between the first approach (new users for existing base stations) and the existing industry approach (new full base stations for new areas).

3.4 Related Work

The idea of using voicemail as an alternative to synchronous voice communications is not new. Besides voicemail and related technologies like "BubbleTalk" [3] and MMS, the ComBadge from MERL [11] records and stores messages on the device until they can be uploaded. This device, however, focused on voice-only interaction, and did not explicitly consider infrastructure architecture.

Asynchronous voice messaging has also been studied from a human interaction perspective [15]. There has been significant work on delay tolerant networking [6], including applications for developing regions [2]. There has also been significant work on cellular usage [18], and on the use of missed calls for communication ("beeping" or "flashing"), which is prevalent throughout the developing world [4]. We build on all of this work.

4 Potential Benefits

We now briefly outline the key benefits that we think can be obtained by moving to a voice message based system.

4.1 Better resource utilization

Figure 1 showed that most of the resources of a synchronous communication system are idle most of the time. This likely holds true for any resource in the system which must be provisioned to account for the worst case; power, energy, cooling, interconnect bandwidth, radio bandwidth, spectrum, processing power, memory, etc.

For existing deployments, flexibility in scheduling message transmission means that we can avoid interfering with higher priority services, such as live calls. The new service fits within the gaps of the existing service, and hence has zero marginal cost.

For new deployments we can avoid dramatic over-provisioning: we can instead provision for average utilization, and schedule message transmission based on available resources. This implies an order of magnitude reduction (or more) in infrastructure cost. In some senses, this is a similar benefit to that obtained by moving from circuit-switched to packet-switched, in that the latter allows more flexibility in assigning transmissions to available resources.

4.2 Increased Effective Coverage

As discussed in Section 2.1, we expect that an asynchronous system can leverage the fact that people regularly travel to the same locations to extend the effective coverage range. To gain initial insights into how effective this form of communication would be, we ran some simulations based on the Reality Mining Data.

In these experiments, we measured the delay to deliver simulated traffic on a synchronous voice call and an asynchronous voice message system. We model voice messages by translating a call from the original trace into a voice message of the same duration. Because the call recipients were removed from the data for privacy, we randomly assign a recipient for each call from the users in the study. In the asynchronous case, once a message has been "recorded," it is queued for transmission to a base station, where it waits for the recipient's phone to come into connectivity. In the synchronous case, we "retry" the call until both the sender and receiver are in range. We modelled different coverage scenarios by disabling cells

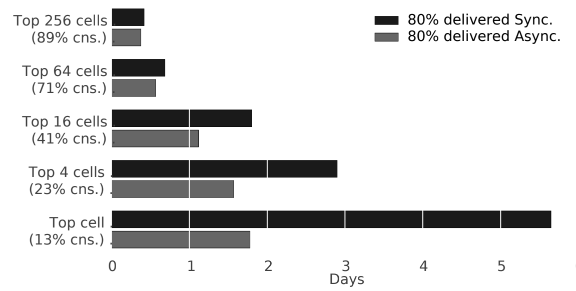

Figure 3: A comparison of communication delay for synchronous and asynchronous communication.

from consideration, and ignore transmission delay, as it is negligible compared to the average queuing delays.

Surprisingly, as shown in Figure 3, in cases of poor coverage, moving to an asynchronous model can in fact *improve* delivery time. This is because in areas of poor coverage, the probability that *both* the sender and the recipient are in radio range simultaneously is low. As we would expect, this effect diminishes as coverage improves (as in the 256-cell scenario). This plot shows the time needed to deliver 80% of the messages, but results are similarly shaped for other percentiles. These experiments also showed that (in the MIT case at least), the store-and-forward messaging model can deliver most messages within a reasonable delay, even in scenarios with only a few percent of the thousands of cells in the study.

4.3 Better Perception of Service

Many people in developing countries have probably had the frustrating experience of needing to retry failed call attempts over and over until a connection can be established. An added benefit to the voice messaging system is an improved user experience due to a better perceived system availability in areas of poor coverage. By virtue of the asynchronous model, short system outages or congestion intervals can be hidden from the user, as the phone will simply delay communication until it can successfully transfer a message. Unexpectedly long outages would still trigger a notification that a user's message has not yet been delivered, similar to e-mail delivery failure notifications we see today.

Figure 4 shows a cumulative distribution of the duration of grid power outages, gathered from mid-2005 to early 2007 by Arrow Networks, a wireless network operator in Ghana, West Africa. The data was collected at five different rural backbone microwave repeaters by a device that sampled the state of the AC power supply every ten minutes, before being conditioned by a UPS system. Although power data for developing countries are often hard to find, these data give us a snapshot of the availability of grid power in several areas of southern Ghana.

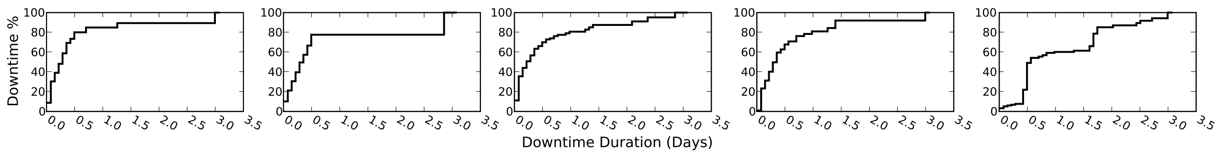

Figure 4: Cumulative distribution of downtime duration for five different microwave repeaters in rural Ghana.

The data show that for this environment at least, most outages are fairly short in duration and can therefore be masked from the end user in an asynchronous system. In contrast, these failures can not be masked for live calls and thus must be avoided via expensive backup power. Note that some outages are very long (multiple days) and as such are prohibitively expensive to avoid by backup power systems. Thus in either system, these outages would result in a user-visible system failure.

4.4 Lower Power Requirements

The power amplifier (PA) is the most power hungry component of most cellular base stations deployed today [19], and is governed by the peak capacity and coverage area. Increased power leads to increased heat dissipation, which influences the size and power requirements of the associated cooling system (CS) (the exact relationship depends on other factors such as the characteristics of the environment, ambient temperature, etc.). For traditional cellular base stations that support 30–100 simultaneous voice calls, the combined power demand from the PA and the CS can be on the order of kilowatts [1].

As discussed above, by moving to an asynchronous model, we may be able to achieve the same total capacity with a peak load of only a few simultaneous calls. This would reduce the PA power requirement by an order of magnitude, impacting the overall system cost in several ways. First, if the power is reduced, the size and cost of the CS also goes down, and beyond a certain point, an active cooling system may not even be necessary. Also, smaller base stations give network administrators a lot more flexibility in packaging and mounting, leading to further savings in mast and tower costs.

With this reduced demand, power consumption may be reduced to tens of watts, at which point portable solar power becomes a viable option [12, 17]. This removes dependence on the (unreliable) power grid or other sources like diesel generators, which require regular maintenance and additional infrastructure (e.g. roads).

4.5 Cost Reduction

The benefits listed so far (better resource utilization, increased effective coverage, better perception of service, and lower power requirements) all contribute to lowering the cost of deployment and operation for carriers, and should in turn reduce the cost of communication for users.

For existing deployments, using otherwise idle capacity can provide additional revenue streams for carriers with only minimal investment. Because voice messaging has effectively zero marginal cost, we expect that carriers can set a very low price and still be profitable. The per-minute charge must be low enough to create demand, but not so low as to cannibalize live-call revenue. Voice-mail mostly phones may also have higher live-call charges than normal phones.

For new deployments, a given user base can be covered for voice messaging with fewer, cheaper base stations than would be required for existing systems. Fewer base stations also means fewer masts, power systems and cooling systems, and ultimately fewer truck rolls to difficult-to-access rural areas. Based on our experience and that of network operators we talked to in Ghana and India, these external factors are by far the most significant deployment costs in hard to reach areas, and all contribute to making remote rural markets inaccessible. Similar factors drive the cost of cellular networks in the industrialized world [10], although better infrastructure improves viability.

5 Challenges

Despite our confidence in the ultimate feasibility of the proposed voice message system, there are several challenges — both technical and non-technical — that will need to be addressed for this system to be viable.

User acceptance: Voice messaging has rarely (if ever) been deployed as a first-class communication mechanism on a large scale. As such, we do not have any quantitative data to suggest how much consumers will be willing to pay for voice messaging as opposed to live calling, nor whether or not they will readily accept the voice message communication model. As mentioned in Section 2, the fact that our target user population is not being served by current communications media suggests that voice messaging may fulfill an unmet need. Also, the fact that our proposed platform can still function as a normal phone should overcome the perception that it is a "dumbed-down" device for poor users. Still, the economic benefits of this proposal depend on users' general acceptance of the voice-messaging communication modality.

Carrier acceptance: The other half of the chicken and egg question is how readily carriers will accept the voice messaging model. We think that the low marginal cost of our proposal should translate to little adoption risk for carriers. The added benefits of being able to access a large new market of customers who are priced out of contemporary offerings should bolster this benefit. At the same time, the CEO of one African cellular company was concerned that the availability of this service might actually draw customers away from the higher-rate voice services, though the opportunity to better utilize the rural fringe was attractive [8]. Still, some early adopter(s) will have to take the initiative to try out this proposed new service model.

Scheduling and routing messages: To conserve energy, phones should have a low duty cycle, but they need to be on for long enough and at the right times to deliver messages during periods of connectivity. Thus, new routing and scheduing algorithms need to be developed to balance between communication delay and energy consumption. This mechanism also needs to consider the impact of delaying a message based on available resources and the message's history. Adding the option of routing directly from phone to phone reduces demand on the infrastructure, but introduces a new set of technical challenges of discovering appropriate peer-to-peer routes.

Low-Cost Base Station: A low-cost base station would largely expand the reach of this proposal. Most of the cost of existing base stations lies in making them carrier-grade, supporting high call volumes, and in the amplifier and power electronics [10, 16], all of which can be reduced for small rural deployment environments. By using standard PC hardware and software radios, we hope to be able to develop a core base station system at low enough cost to make it affordable for deployment in rural environments.

Base Station Connectivity: Pushing base station deployments to hard-to-reach areas requires connecting them to the PSTN or Internet. For routing voice message and other asynchronous traffic, DTN delivery mechanisms such as physical data transport (e.g. "sneakernet") can be used. In hard to reach locations, a data courier service could designed to leverage the existing rural transportation system (mostly periodic buses) in order to interconnect the messaging system of the various villages, and the also to connect to the core network at the nearest uplink.

6 Conclusions and Caveats

In this paper, we present some potential benefits in adopting an asynchronous voice messaging system in developing regions. Based on the data we have examined and the analysis presented above, we believe the cost of a building and operating a network based on voice messaging could be notably cheaper than a synchronous voice system, enabling a viable business model for communication services to rural and urban poor users.

A significant caveat relates to the environment and population in which the data used in the MIT study was collected (i.e. a set of urban college students), and there are surely significant differences between this environment and those in our target regions. However, we think that the general patterns of human movement and communication are similar enough to justify the high-level conclusions that we draw from this preliminary exploration.

Reaching the next billion users will need at least a partially new approach that addresses the core challenges of coverage and cost. We plan to validate the feasibility and potential benefits of this model with a trial deployment in the coming months.

7 References

[1] B. Berglund, M. Englund, and J. Lundstedt. Third design release of Ericsson's WCDMA macro radio base stations. *Ericsson Review*, (2), 2005.

[2] E. Brewer et al. The case for technology for developing regions. *IEEE Computer*, 38(6):25–38, 2005.

[3] Bubble Motion. `http://bubblemotion.com/`.

[4] J. Donner. The rules of beeping: exchanging messages using missed calls on mobile phones in sub-Saharn Africa. In *55th annual conference of the International Communication Association (Communication and Technology Division)*, May 2005.

[5] N. Eagle and A. Pentland. Reality mining: Sensing complex social systems. *Journal of Personal and Ubiquitous Computing*, 2005.

[6] K. Fall. A delay-tolerant network architecture for challenged internets. In *SIGCOMM*, 2003.

[7] Grameen Phone. `http://grameenphone.com`.

[8] M. Ibrahim. CellTel. Personal Communication.

[9] International Telecommunications Union. World Telecommunications/ICT Development Report. 2006.

[10] K. Johanssen, A. Furuskär, P. Karlsson, and J. Zander. Relation between base station characteristics and cost structure in cellular systems. In *Proceedings of the 15th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*, pages 2627–2631, 2004.

[11] J. Katz. Combadge: A voice messaging device for the masses. Technical Report TR2005-161, Mitsubishi Electronics Research Laboratory, Dec 2005.

[12] S. M. Mishra, J. Hwang, D. Filippini, T. Du, R. Moazzami, and L. Subramanian. Economic Analysis of Networking Technologies for Rural Developing Regions. December 2005.

[13] MTN Uganda. `http://www.mtn.co.ug`.

[14] J. Pal. Should the state be funding rural public internet access? The Case of Akshaya, India. In *Conference on Information and Communications Technology for Development*, December 2007. (Submitted).

[15] K. Ross. Asynchronous voice: a personal account. *IEEE multimedia*, 10, 2003.

[16] J. Sarnecki, C. Vinodrai, A. Javed, and K. Dick. Microcell design principles. *IEEE Communications*, April 1993.

[17] S. Surana, R. Patra, and E. Brewer. Simplifying fault diagnosis in locally managed rural wifi networks. In *NSDR*, 2007.

[18] H. Verkasalo. *A Cross-Country Comparison of Mobile Service and Handset Usage*. PhD thesis, Helinski University of Technology, 2007.

[19] J. Yao, T. Long, and S. Long. High Efficiency Switch Mode Amplifiers for Mobile and Base Station Applications. *Technical Report, UC Santa Barbara*, 2001.