On the Importance of Translucence for Mobile Computing

(position paper)

Maria R. Ebling and M. Satyanarayanan
Carnegie Mellon University
Computer Science Department
Pittsburgh, PA 15213
email: `{mre,satya}@cs.cmu.edu`

Introduction

Mobile clients experience a wide range of network characteristics, and this situation is likely to continue for the foreseeable future. This range of characteristics includes fast, reliable, and cheap networks at one extreme and slow, intermittent, and expensive ones at the other. The demand for mobile connectivity has created an active area of research. As new technologies become available, mobile clients will have to choose between competing network providers offering different levels of service. In fact, mobile clients will eventually be capable of seamlessly switching from one network to another depending on current needs. Thus, mobile clients will need to choose a network provider dynamically.

The decision regarding which network provider to use involves trade-offs that include both energy and financial components. Our position is that mobile clients cannot balance these tradeoffs well without assistance from the user. The challenge is how to gain enough assistance to make wise choices without imposing undue burden on users. We believe that a key to meeting this challenge will be *translucence*. A translucent system exposes critical details of the system to the user in order to improve the system's ability to service the user's needs, while hiding non-critical details from the user to minimize the imposed burden.

Energy Trade-Off

An important tradeoff, inherent in wireless connectivity, is between battery power and communication. As one researcher characterized it, sending a packet over a wireless network is like "spraying a piece of your battery into the air." The fundamental question here is how a system that supports wireless communication should balance this tradeoff.

In order to communicate, mobile clients must expend battery power. The scarcity of this resource depends on the environment in which the mobile client finds itself, including factors such as the availability of replacement batteries, the ability to recharge batteries, and the expected length of isolation from energy sources. Thus, in deciding whether and how much to communicate, mobile clients must consider the energy consumption required.

When using a wireless network, the cost of communication has an added complication. Communicating over longer distances requires more energy than communicating over shorter distances. Under ideal conditions, the laws of physics dictate that communicating twice as far requires four times as much energy. Real world conditions, however, are less than ideal. Cellular phone companies use a rule of thumb that battery consumption increases as the third or fourth power of the distance. This complication means that different routing algorithms may require vastly different energy expenditures. An algorithm that uses a few long hops may require substantially more battery power than one that uses many more hops each covering a shorter distance. By increasing the number of hops however, the algorithm is also likely to increase the service time. Thus, in order to choose which network to use, mobile clients will need to balance energy and performance.

How does a system decide which network to use from the perspective of battery power? The choice is easy if the user has said that minimizing energy consumption or maximizing performance is the primary consideration, but this simple choice is not likely to satisfy users all the time. This *energy-performance trade-off* complicates decisions regarding whether and how much to communicate, as well as what network to use.

Monetary Trade-Off

Two important characteristics that will differentiate competing network providers are performance and cost. The performance characteristics offered will differ in terms of bandwidth, latency, and reliability, but they will also differ in terms of cost. Today, service providers charge a flat monthly fee or by units of connect time. Future service providers may well charge in units of kilobits or packets.

The cost of transferring a piece of data can thus be estimated based on its size. Because mobile clients will be able to seamlessly switch from one network to another, they will be able to choose the network provider best able to service each data transfer. These network providers will compete for traffic based upon cost and performance. Mobile clients, then, will need to balance cost and performance to choose a service provider.

How can a mobile client decide which network to use from the perspective of financial cost? Once again, the choice is easy if the user has said that minimizing cost or maximizing performance is the primary consideration. Unfortunately, as before, this simple decision is not likely to serve all users all the time. Just as people don't typically send all of their physical documents via a single class of delivery service (e.g., first class mail or overnight delivery), mobile users will want to have different classes of network service available to meet different needs at different times. Although a user may want to minimize cost most of the time, certain data may be so critical to the user's work that she may be willing to pay substantially more money for it to be delivered quickly. Thus, an important trade-off that mobile systems will face is the *cost-performance trade-off*.

Need for Assistance

Balancing the financial and energy costs of mobile communication is a very difficult problem, one that early systems are unlikely to solve satisfactorily without assistance from their users. The key challenge is that users' requirements change based upon current needs and the importance of the data being transferred. Systems, in general, have little knowledge regarding the user's current needs or the importance of a piece of information to the user's work. In order to make wise decisions regarding network usage, systems will need more information—information that is available only from the user.

Requirements

A translucent system must balance its need for information with the burden that it imposes on the user. Obviously, such a system should follow generally accepted principles of HCI design [2], but they must also follow more stringent guidelines.

User assistance should add value to the system. User assistance comes at a high price: user attention. The benefit of that assistance must be tangible. The resulting system must offer better performance, or better availability, or better usability.

User assistance should be optional. Because user assistance requires user attention, the system should not require the user to provide assistance. When no assistance is offered, the system should make decisions based upon the best information available.

Translucent systems must be unobtrusive. The user's goal is not to babysit the mobile client, but to complete his work. These systems should alert users to critical events and allow users to influence those events (when possible). The system must not demand immediate attention, must not be annoying in its interactions, and must not "cry wolf."

Interacting with the User

A key question then is how to get meaningful assistance from the user while following these guidelines. There are three components to a translucent system supporting a mobile user:

1. alerting users to important events
2. balancing demands for the network
3. choosing between network providers

These components must work together, but their solutions may differ.

First, users must be alerted to important events. One metaphor for how to accomplish this is the dashboard interface. Indicator lights present a small amount of information in a minimal amount of space. They do less well at presenting detailed information, but this difficulty is easily remedied by making the interface interactive and allowing the user to request further information when it is needed. We have used this metaphor to build a translucent interface to a distributed file system [1]. Usability testing has shown that users understand the events presented by the interface.

Alerting users to important events is, perhaps, the most important characteristic of a translucent system. This notification serves to set user expectations for system behaviors according to the current operating conditions. Systems that adapt to network bandwidth changes of four or more orders of magnitude must behave differently depending upon their current network connectivity. To avoid confusion and frustration, user expectations must match the changing network environment. Thus, even a system that offers only a notification feature presents a useful amount of translucence.

Second, users must be given the opportunity to control how limited network resources are used by the system. The user's top priority may be propagating updates back to his colleagues at home or it might be transferring data from home to his current location. The user may only need to propagate a small fraction of the updates that he has made or to transfer a small subset of the data that needs to be fetched. When network resources are extremely scarce, the system can't make these choices automatically. Translucent systems must allow the user opportunities to balance these competing demands.

Third, users must be given the opportunity to control communication expenditures (both the financial cost and energy consumed). After all, the user is responsible for paying the bill and the user must face the consequences of a squandered battery. One possible metaphor for how to present networking decisions to users is the postal delivery model. Systems could present users with different classes of network delivery service, each with a different cost. These costs would include both financial and energy components. Users might choose a default service and selectively change that service as their needs dictate.

The postal delivery metaphor, however, does not solve our problem entirely. People are generally aware of the mail they send. It is easy for them to make decisions about individual pieces of mail. In the case of mobile clients, however, users are not always aware of the network traffic necessary to service their requests. Further, they cannot become aware of that traffic in its entirety or they would never get anything done! If we apply this metaphor to our problem, we must allow users to ascribe delivery decisions to an entire class of activities.

Another interaction technique would exploit a banking model, where users have network connectivity accounts (perhaps one for money and one for energy). As the system debits these accounts, it might provide audio feedback to users. In this way, users could track their network consumption indirectly. If the accounts were being depleted too quickly, they could change their delivery specification. Users could have a simple graphical interface that would allow them to increase performance (at the expense of money, energy, or both) or minimize cost (at the expense of performance).

Conclusions

The question of how to balance the need to communicate over mobile networks with the costs involved is a difficult one. Users will require some amount of control, yet too much control will make the systems unusable. These interactions need to be reviewed in detail to balance the needs of the users with the burdens imposed. In this position paper, we have identified an important problem faced by system designers. We have also suggested some preliminary ideas for addressing that problem.

Bibliography

[1] Ebling, Maria R. *Translucent Cache Management for Mobile Computing*. Doctoral Dissertation, School of Computer Science, Carnegie Mellon University, Pittsburgh, PA, March 1998.

[2] Nielsen, Jakob. *Usability Engineering*. Boston: AP Professional, 1993.