On Some Principles of Nomadic Computing and Multi-Access Communications

Leonard Kleinrock, Nomadix, Inc.

ABSTRACT

In this article we identify some of the key problems one encounters when thinking about multi-access systems. We begin with a general discussion of nomadic computing and move on to issues of multi-access in a distributed environment. We then specialize to the case of wireless systems, and identify some of the key considerations and algorithms which must be addressed in that environment. Lastly, we identify some of the higher-level issues and principles one should properly keep in mind when investigating the design and behavior of these systems.

INTRODUCTION

The issues involved with untethered communications are extensive. This article discusses some of these from a high level with the purpose of addressing the considerations that extend across many of the detailed analyses and discussions. No single multi-access scheme is considered in detail. Instead, the discussion centers around more general themes.

We begin by discussing the larger field of nomadic computing, and the issues of mobility, access, and service that arise as a result. We then specialize to those considerations that arise in the particular case of shared media and multi-access. Following that we specialize further to issues unique to the wireless environment. We conclude by examining some general metrics and insights the author has developed over many years of research in this area.

NOMADIC COMPUTING

The combination of portable computing with portable communications is changing the way we think about information processing. We now recognize that access to computing, communications, and service is necessary not only from one's "home base," but also while one is in transit and when one reaches one's destination.

These ideas form the essence of a major shift to *nomadicity* (nomadic computing and communications). The focus of nomadicity is on the system support needed to provide a rich set of capabilities and services to the nomad as he/she moves from place to place in a transparent and convenient form [1–5].

Of concern are those capabilities that must be put in place to support nomadicity. The desirable characteristics for nomadicity include independence of location, of motion, of computing platform, of communication device, and of communication bandwidth, with widespread presence of access to remote files, systems, and services. The notion of independence here refers not only to the quality of service one sees, but also to the perception of a computing environment that automatically adjusts to the processing, communications, and access available at the moment. For example, the bandwidth for moving data between a user and a remote server could easily vary from a few bits per second (in a noisy wireless environment) to hundreds of megabits per second (in a hard-wired ATM environment); or the computing platform available to the user could vary from a low-powered personal digital assistant while traveling to a powerful supercomputer in a science laboratory.

Today's systems treat radically changing connectivity or bandwidth/latency values as exceptions or failures; in the nomadic environment, these must be treated as the usual case. Moreover, the ability to accept partial or incomplete results is an option that must be made available due to the uncertainties of the informatics infrastructure.

The concept of mobility in a nomadic environment takes on many meanings. One component of mobility is the desire to gain access to the informatics infrastructure when one arrives at different destinations; an example might be when one transports one's laptop from the corporate office to one's home or a hotel. In this case, issues of seamlessly managing the configuration settings of the nomad (TCP/IP address, netmask, Web proxy setting, domain name server, gateways, etc) must be managed. Another component of mobility is the desire to gain access to the informatics infrastructure while in motion. In this case, issues of tracking, handoff, connectivity, and so on, as well as managing configuration settings come to mind and must be addressed.

The need for nomadic computing support has been recognized recently, and an entire industry has begun to arise in response to this need. We have seen the rollout of broadband access (e.g., DSL, cable modems, wireless access) from a number of suppliers (Copper Mountain, Paradyne, SBC, GTE, Lucent), and the rollout of subscriber management systems (e.g., Nomadix, Redback, Shasta). Most recently, it has become clear that we must shift from a "connection-centric" view (connectivity is now almost a commodity in terms of innovation and scarcity) to a "service-centric" view in which the focus is on providing services over those broadband pipes. No longer is it the "fat pipe" that matters, but what goes through it. Local service is no longer simply printing and e-mail; it is movie tickets, games, and pizza. Indeed, the Internet is becoming much more than connected networks and computers — it is becoming a service access and delivery system.

The ability to automatically adjust all aspects of the user's computing, communication, storage, and service functionality in a transparent and integrated fashion is the essence of a nomadic environment.

Shared Media and Multi-Access

In dynamic environments such as that of the nomad, it is necessary to find ways to share resources in an adaptive fashion. Indeed, we must deal with the problems of multiple users attempting to access common resources in a competitive fashion (i.e., multi-access). Not only are we faced with the queuing problems that arise from the stochastic nature of the demands; we are also faced with the issue of allocating resources to a geographically distributed (and possibly mobile) set of demands. Were we not in this distributed environment, queuing theory [6] would provide us with the ultimate mean response time–throughput performance profile. However, we have additional loss of resources due to the cost of organizing the separated demands into some kind of cooperating queue, which permits intelligent access to the available resources.

For the purposes of this article (and at no loss of generality), we assume that the shared resource is a communication channel and that the demands are message sources which require transmission of their messages over this shared channel. We are faced with controlling access to this channel from these distributed message sources in which the control information must pass over the same channel being controlled (or over a control subchannel derived from the data channel).

We characterize the classes of multi-access schemes into three categories, and observe that the above-mentioned cost of organizing these sources into a cooperative group also falls into three categories. The three classes of multi-access control schemes are:

- Fixed (i.e., static) control
- Random control
- Dynamic control

The three sources of lost resources are:

- Wasted (idle) resources
- Collision of resources
- Control overhead

In the first class of fixed control we include all multi-access schemes that rigidly assign a portion of the channel to each source. Examples include time-division multiple access (TDMA), frequency-division multiple access (FDMA), and so on. Fixed control is extremely easy to implement. However, the typical price paid for such a rigid assignment is that the channel assigned to a given source is wasted (i.e., lies idle) whenever the source has nothing to send in its assigned portion of the channel.

	Idle resource	Collisions	Control overhead
Fixed control	Yes	No	No
Random control	No	Yes	No
Dynamic control	No	No	Yes

Table 1. *The price of distributed resources.*

At the other extreme from the rigid approach taken with fixed control is that of no, or minimal, control, namely the second class, random control. Examples include Aloha, carrier sense multiple access (CSMA), and so on. Random control is also relatively easy to implement. However, the typical price paid here (for poor or no control) is that of collisions which occur when two sources attempt to transmit in the same portion of the channel (the definition of portion could be time, frequency, code, space, or some combination).

With the third class, dynamic control, the channel resource is apportioned on a demand basis according to the needs of the sources. This control can take the form of polling (where a source waits to be asked if it needs channel access) or active requests from the source in the form of a request for access, such as asynchronous time-division multiplexing (TDM) or reservation Aloha or others. In all such cases, the typical price paid is that of the overhead to send the control signals over the channel. Table 1 summarizes these trade-offs.

Of course, in the general case, one can combine some of these access control methods and then suffer a mix of their forms of channel cost. An example of mixed access control would be that of a dynamic control scheme in which reservations for data slots are made using a random access request control channel (say ALOHA), and where the data slots themselves are allocated on a fixed TDMA basis as long as the source has data to send, after which the data slots are assignable to other sources as the demand arises.

The Issues of Wireless Access

It is clear that a great many issues regarding nomadicity arise whether or not one has access to wireless communications. However, with such access a number of interesting considerations arise that we discuss in this section.

Access to wireless communications provides two capabilities to the nomad. First, it allows the nomad to communicate from various (fixed) locations without being connected directly into the wireline network. Second, it allows the nomad to communicate while traveling. Although the bandwidth offered by wireless communication media varies over an enormous range as does the wireline network bandwidth, the nature of the error rate, fading

Access to wireless communications provides two capabilities to the nomad. First, it allows the nomad to communicate from various (fixed) locations without being connected directly into the wireline network. Second, it allows the nomad to communicate while traveling.

behavior, interference level, mobility issues, and so on for wireless are considerably different from wireline networks, so the algorithms and protocols to support wireless access are far more complex than those for the wireline case. Whereas the location of a user or device is a concern for wireline nets as described above, the details of tracking a user while moving in a wireless environment add to the complexity and require rules for handover, roaming, and so forth [1].

The cellular radio networks so prevalent today have an architecture that assumes the existence of a cell base station for each cell of the array; the base station controls the activity of its cell. The design considerations of such cellular networks are reasonably well understood and are being addressed by an entire industry [7]. We discuss these no further in this article.

There is, however, another wireless networking architecture of interest that assumes no base stations [8]. Such wireless networks are useful for applications that require "instant" infrastructure, among others. For example, disaster relief, emergency operations, special military operations, and clandestine operations are all cases where no base station infrastructure can be assumed. In the case of no base stations, maintaining communications is considerably more difficult. For example, it may be that the destination for a given reception is not within range of the transmitter, in which case some form of relaying is required; this is known as *multihop* communications. Moreover, since there are no fixed location base stations, the connectivity of the network is subject to considerable change as devices move around and/or the medium changes its characteristics. A number of new considerations arise in these situations, and new kinds of network algorithms are needed to deal with them.

We find it convenient to articulate some of the issues and algorithms with which one must be concerned in the case of no base stations by decomposing the possible scenarios into the following three.

Static Topology with One-Hop Communications — In this case, there is no motion among the system elements, and all transmitters can reach their destinations without any relays. The issues of concern, along with the needed network algorithms (shown in italics), are as follows:

- Can you reach your destination: *power control*
- What access method should you use: *network access control*
- Which channel (or code) should you use: *channel assignment control*
- Will you interfere with another transmission: *power and medium access control*
- When do you allow a new "call" into the system: *admission control*
- Is there sufficient bandwidth for your application: *capacity assignment*
- For different multiplexed streams, can you achieve the required QoS (e.g., bandwidth, loss, delay, delay jitter, higher order statistics, etc.): *multimedia control*
- What packet size should you use: *system design*
- Is there a need for systemwide synchronization: *global control*
- How are errors to be handled: *error control*
- How do you handle congestion: *congestion control*
- How do you adapt to failures: *degradation control*

Static Topology with Multihop Communications — Here the topology is static again, but transmitters may not be able to reach their destinations in one hop, so multihop relay communications is necessary in some cases. The issues of concern, along with the needed network algorithms (shown in italics), are all of the above plus:

- Is there a path to your destination: *path control*
- Does giant stepping [9] help: *power control*
- What routing procedure should you use: *routing control*
- When should you reroute existing calls: *reconfiguration control*
- How do you assign bandwidth and QoS along the path: *admission control* and *channel assignment*

Dynamic Topology with Multihop — In this case, the devices (radios, users, etc.) are allowed to move, which causes the network connectivity to change dynamically. The issues of concern, along with the needed network algorithms (shown in italics), are all of the above plus:

- Do you track or search for your destination: *location control*
- Which network reconfiguration strategy should you use: *adaptive topology control*
- How should you use reconfigurable and adaptive base stations: *adaptive base station control*

These lists of considerations are not complete, only illustrative of the many interesting research problems that present themselves in this environment. Indeed, in this section we have addressed only the network algorithm issues, and have not presented the many other issues involved with radio design, hardware design, tools for CAD, system drivers, and so on.

SOME GENERAL CONSIDERATIONS

In the foregoing, we have listed a number of issues and problems that arise in the case of nomadic and untethered communications. In this section we choose to focus on some overriding principles and observations that are useful across this set of issues and problems. These principles and observations are simply a subset (albeit an interesting one) of guiding principles of value when discussing multi-access in a distributed environment.

ON THE LATENCY PARAMETER *a*

In many distributed communication systems, there are three parameters that interact:

C = capacity of the communication channel (say in megabits per second)

L = length of the channel (say in kilometers)

b = length of the data unit transmitted (say in bits per packet)

These three can be combined into a single key system variable, latency, which we denote by a, and which is defined as the propagation delay (time for a bit to travel the length of the channel) divided by the time it takes to transmit a packet into the channel. It turns out that the system performance of many communication channels is closely tied to the latency. If we assume it takes 5 µs for energy to travel through 1 km of the channel (this is approximately the value for a wireline channel), the latency is simply [10]

$$a = 5LC/b$$

since $5L$ is the propagation delay (in microseconds) through the channel, and b/C is the time (also in microseconds) to transmit a packet. It is interesting to observe the range of values taken on by a for some characteristic systems; these are shown in Table 2.

	Bandwidth (Mb/s)	Packet length (bits)	Propagation delay (µs)	Latency (a)
LAN	10.00	1000	5	0.05
WAN	0.05	1000	20,000	1.00
Satellite	0.05	1000	250,000	12.50
Fiber link	1000.00	1000	20,000	20,000.00

Table 2. *Latency for some common systems.*

The thing to note from this table is the enormous range over which the key parameter, latency, varies (six orders of magnitude!). If the channel is a wireless channel, and if the access method uses some version of CSMA, it is well known that the performance of the channel critically depends on the latency a. Since we have such a wide variation in a, it is very important to consider its effect on performance.

Power

In evaluating the performance of a system we often compare the mean response-time (T) with the throughput (γ) of a system. This profile typically looks like that shown in Fig. 1.

We note that at low throughput we get good response time, and at high throughput we get poor response time. So a natural question arises regarding the most effective trade-off between these two. Some years ago, we proposed a single performance measure (P), *power*, which combines these two. We define power as [11]

$$P = \gamma/T.$$

It turns out that power is maximized at that point on the response time–throughput profile where a straight line from the origin first becomes tangent to the profile; in Fig. 2 we denote the optimum throughput operating point by γ^*.

What is amazing about this result is that for all $M/G/1$ queuing systems [6], this point occurs where $E[N]$, the average number of messages in the system, is *exactly* one! What makes this interesting is that it is intuitively the correct operating point for deterministic systems [11].

Giant Stepping

In the case of packet radio systems, a totally different consideration leads to exactly the same result we just quoted earlier. Let us consider an ideal multihop packet radio system where the power in every radio is adjusted so that each hop covers exactly a radius R. Further assume that the total distance a message must travel is $D >> R$. Now let $T(R)$ be the mean response time experienced by a message in traveling one hop, due to interfering traffic from other radios. It is clear that if we choose R to be large, $T(R)$ is large (more interference), but the number of hops is small, and vice versa for a small value of R. Thus, we desire to find that value of the step size R which uses an appropriate transmission power in a way that balances these two effects. We call this *giant stepping* [9]. This is a typical consideration in ad hoc network design. If we assume that there is a continuum of radios in the plane, the total mean delay along the path is clearly $T(R)$ $[D/R]$. Differentiating this function with respect to R, we find that the optimum value of R which minimizes the total mean delay is such that $dT(R)/dR = T(R)/R$. It is interesting to note that this solution is at exactly the same point as that shown in Fig. 2.

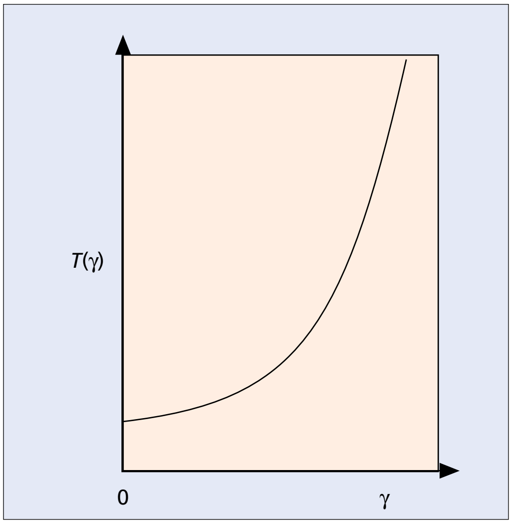

Figure 1. *Typical response time–throughput trade-off.*

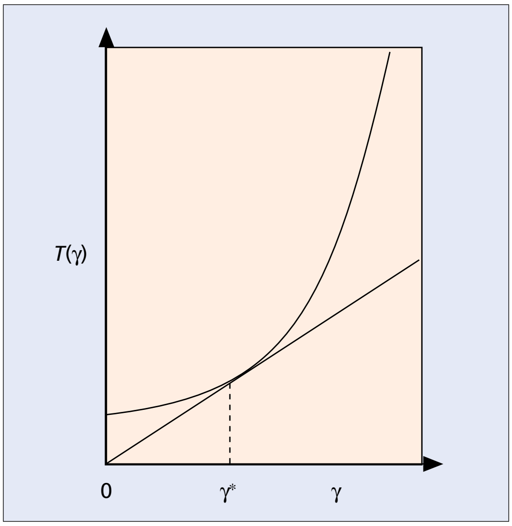

Figure 2. *The operating point at maximum power.*

as a complement to detailed systems analysis, it is important to look for overriding principles of operation that transcend the specific systems and layers being studied.

A number of refinements can be made to this simplified model to account for some of the practical constraints of the problem (e.g., discrete locations for radios instead of a continuum). However, the beauty of this idealized result is that it provides a guideline for system design and enjoys the intuitive interpretations given in the previous section; for example, it says that at optimality, the load on a system feeding a pipeline should be such that, on average, there should be one packet per hop. It is this kind of intuition that leads to a higher level of thinking and understanding about these very complex systems.

CONCLUSION

The systems of interest we studied in this article were nomadic computing, multi-access systems, and wireless systems, these systems sharing a common need for control, access, and allocation. A key conclusion from this article is that in order to gain understanding of the behavior of complex systems such as these, it is often helpful to look for overriding principles of performance. For example, we pointed out that one way or another, one must pay a price for organizing distributed demands in their competition for use of a shared resource; it turns out that the price will be some mix of idle resources, collisions, and control overhead. Another factor we studied was that of latency, and we pointed out that it often has a dominant effect, depending on a small set of basic system parameters. We characterized two very different optimization problems using very simple and different models, and found that the same underlying principle applied: the notion of finding the knee of the appropriate curve as the optimal operating point.

The conclusion is that, as a complement to detailed systems analysis, it is important to look for overriding principles of operation that transcend the specific systems and layers being studied. Indeed, one must span more than one layer of the communication stack when designing systems of the type described herein; for example, there must be awareness of the link layer capabilities (bandwidth, delay, error rate) by the application and transport layers lest one try to send a full motion color video to a monochromatic display on a handheld device over a limited-bandwidth wireless link. The need for proxies at various points in the communications path and at various layers can also be of use in these situations. Think globally and look for overriding principles.

REFERENCES

[1] R. H. Katz, "Adaptation and Mobility in Wireless Information Systems," *IEEE Pers. Commun.*, vol. 1, no. 1, 1st qtr. 1994, pp. 6–17.

[2] L. Kleinrock, "Nomadic Computing — An Opportunity," *Comp. Commun. Rev.*, Jan. 1995, vol. 25, no. 1, pp. 36–40.

[3] L. Kleinrock, Chair, "Nomadicity: Characteristics, Issues, and Applications," Nomadic Working Team of the Cross Industrial Working Team, 1995; also available at http://www.xiwt.org/documents/Nomadicity.html

[4] J. Short, R. Bagrodia, and L. Kleinrock, "Mobile Wireless Network System Simulation," *Mobicom '95*, Nov. 1995, pp. 195–205.

[5] A. Alwan *et al.*, "Adaptive Mobile Multimedia Networks," *IEEE Pers. Commun.*, vol. 3, no. 2, Apr. 1996, pp. 34–49.

[6] L. Kleinrock, *Queuing Systems, Volume I: Theory*, Wiley, New York, 1975.

[7] J. E. P. Padgett, C. G. Gunther, and T. Hattori, "Overview of Wireless Personal Communications," *IEEE Commun. Mag.*, Jan. 1995, vol. 33, no. 1, pp. 28–41.

[8] R. Jain *et al.*, "PC-notebook Based Mobile Networking: Algorithms, Architectures and Implementations," *ICC '95*, June 1995, pp. 771–77.

[9] L. Kleinrock, "On Giant Stepping in Packet Radio Networks," UCLA, Packet Radio Temporary Note #5, PRT 136, Mar. 1975.

[10] L. Kleinrock, "The Latency/Bandwidth Tradeoff in Gigabit Networks," *IEEE Commun. Mag.*, Apr. 1992, vol. 30, no. 4, pp. 36–40.

[11] L. Kleinrock, "On Flow Control in Computer Networks," *Proc. Int'l. Conf. Commun.*, vol. II, Toronto, Ontario, Canada, June 1978, pp. 27.2.1–27.2.5.

BIOGRAPHY

LEONARD KLEINROCK [F] (lk@cs.ucla.edu) received his Ph.D. from MIT in 1963 and has served as a professor of computer science at the University of California, Los Angeles since then. He received his B.E.E. degree from CCNY in 1957 (and an Honorary Doctor of Science from CCNY in 1997). He was a co-founder of Linkabit and is also founder and chair of Technology Transfer Institute, a Santa Monica-based conference company. He has published more than 200 papers and authored six books on a wide array of subjects including packet switching networks, packet radio networks, local area networks, broadband networks, and gigabit networks. Additionally, he recently launched the field of nomadic computing, the emerging technology to support users as soon as they leave their desktop environments; nomadic computing is poised to be the next major wave of the Internet. He is chair of a West Coast-based Internet startup, Nomadix, Inc, which he recently founded and which already provides advanced products to support the nomad. He is a member of the National Academy of Engineering, an ACM fellow, an IEC Fellow and a founding member of the Computer Science and Telecommunications Board of the National Research Council. Among his many honors, he is the recipient of the L. M. Ericsson Prize, the Marconi Award, the CCNY Townsend Harris Medal, the CCNY Electrical Engineering Award, the UCLA outstanding Teacher Award, the Lanchester Prize, the ACM SIGCOMM Award, the Sigma Xi Monie Ferst Award, the INFORMS Presidents Award, and the IEEE Harry Goode Award.