A Concept of Communication Distance and Its Application to Six Situations in Mobile Environments

Jeffrey V. Nickerson, *Member, IEEE*

Abstract—Wireless networks combined with location technology create new problems and call for new decision aids. As a precursor to the development of these decision aids, a concept of communication distance is developed and applied to six situations. This concept allows travel time and bandwidth to be combined in a single measure so that many problems can be mapped onto a weighted graph and solved through shortest path algorithms. The paper looks at the problem of intercepting an out-of-communication team member and describes ways of using planning to reduce communication distance in anticipation of a break in connection. The concept is also applied to ad hoc radio networks. A way of performing route planning using a bandwidth map is developed and analyzed. The general implications of the work to transportation planning are discussed.

Index Terms—Mobile computing, wireless communication, decision aids, transportation planning, emergency response, mobile ad hoc networks, route guidance, contingency planning.

◆

1 INTRODUCTION

1.1 Overview

MOST of our communication is instant, relying on electronic networks. Yet, this has not always been so—communication has usually involved travel. Now, wireless technology is again creating a mix of travel and communication, with special characteristics. We can move as we communicate. We can know the physical locations of the people we are talking to. We can pass in rapid succession through areas of no connectivity and areas of high bandwidth. We are more likely to encounter sudden breaks in communication.

This paper sets out to define a new set of problems which occur in the context of mobile communication. By focusing on the time it takes to communicate a message, including the associated travel time, we can gain insights into a set of problems surrounding mobility, intermittency, and productivity. We call this concept communication distance and describe a set of mobility situations this concept applies to.

Our claim is that this concept is a useful way to frame mobility problems. It may have application in work on route guidance [1], as well as in emergency response, contingency planning, and transportation planning. In order to make our points, we discuss six situations. First, we look at a simple emergency response situation and show how communication distance can frame a solution. We then discuss a similar situation involving secure communication. Third, we look at planning in disaster situations. Fourth, we discuss mobility in relation to ad hoc networks. Finally, we analyze two situations in which

mobility and communication are closely intertwined and propose an algorithm for route planning. Extensions to the approach are discussed to handle dynamics and uncertainty.

1.2 The Concept

In Fig. 1, two people can communicate by physically meeting. In order to do so, both may need to travel. The communication distance shown in the figure is 300 seconds, which might represent the amount of time it took for them to convene in a conference room and exchange a simple message.

With telephones, A and B can exchange a simple message practically instantly, as shown in Fig. 2. We use epsilon to represent this practically instant communication.

In Fig. 3, we combine movement and telecommunications. It might be that A's telephone service goes out. In such a case, if A is in an urban setting, A might walk to a phone booth to talk to B. The communication distance is the sum of the two links.

Fig. 4 shows a cell phone connection, with both A and B mobile—if their radii of communication are within the vicinity of cell towers, the towers effectively bridge communication across a wired network, and the communication distance is epsilon.

In Fig. 5, we show an example from a mobile ad hoc network. A and B, in order to communicate, may actually have to move to accomplish the communication. They do not have to meet, as in Fig. 1, but their surrounding circles, representing their radio range, must meet. In such a case, we can measure their communication distance as the amount of time it takes them to establish the electronic connection. Usually, A and B will connect through many intervening nodes, but A and B need to move into the range of each other or nearby intermediaries.

● *The author is with the Stevens Institute of Technology, Castle Point on Hudson, Hoboken, NJ 07030. E-mail: jnickerson@stevens.edu.*
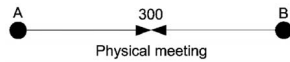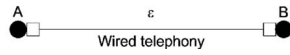
Fig. 1. Meetings.

Fig. 2. Telephone conversations.
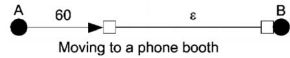

Fig. 3. Combining movement and wired telephony.
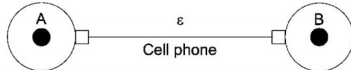
Fig. 4. Cell phone communication.

Fig. 5. Ad hoc network communication.

Fig. 6. A command and control team send an emergency response team toward an accident, and then lose contact with them. Assuming the locations of other responders are known, these responders might be able to intercept the response team.
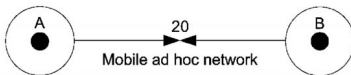
Communication distance as we describe it is related to the concept of latency broadly viewed. We can describe the communication distance between two entities as the minimum possible latency for a particular message size using all available forms of movement and technology. Latency is usually defined as the sum of propagation, transmission, and queuing [2]. When we are communicating electronically, propagation delays are defined in relation to the speed of light. When we communicate by walking to meet someone, we can think of propagation as the amount of time it takes for us to get there. We are delivering the message packet by a very slow link.

Carrying a message involves a long propagation time and practically infinite bandwidth, in the sense that we can carry computer tapes, CD ROMS, and other forms of mass storage. When we hand over the message, we are performing transmission. Then, the message somehow needs to be read—we can allocate the time to parse the tape to the queuing component of latency. In the broader context of vehicle movement, when we model traffic congestion, we model it with queues. Why is it that we don't usually think about travel as part of communication? Part of the issue has to do with where we draw the boundary around the communication act. If I decide to walk down the hallway to talk to a colleague, I might define the start of the act as the moment at which I begin to speak. Or, I might define the start of the act as the moment I begin to walk.

The other reason we don't naturally combine the two concepts is that they compose differently. As a physical package moves from place to place, we can compose the movement by adding the travel time. But, with electronic communication, once we have enough data to fill a set of pipes, the time is constrained by the size of the smallest pipe. Transportation is, in our sense of the term, propagation-focused, and electronic communication is bandwidth-focused.

In speaking of transportation and communication, Cooley writes:

> There is good reason for the separation of the two in the unlikeness of their processes ... Space—distance—as an
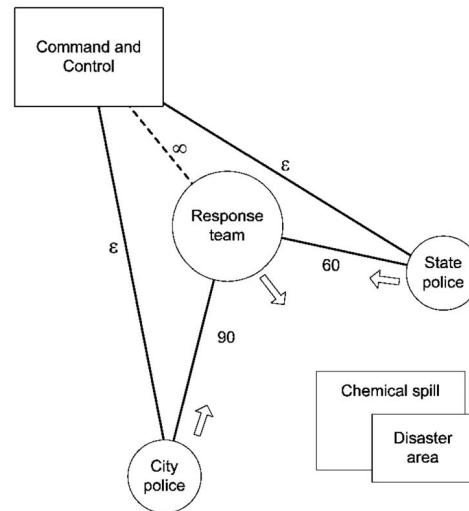
obstacle to communication has so nearly been overcome that it is hardly worth considering [3].

It is notable that this comment came in 1894, when the locomotive, the telegraph, and the telephone were widely seen as annihilating space. We are arguing that this annihilation is incomplete and that, by viewing problems of mobility through the concept of communication distance, we may gain insights. For example, by looking to minimize communication distance in a situation in which a connection has been cut, we will be more likely to consider alternative transportation and communication links. The possibility of other means of communication may not always be obvious to us, and these means can be important, as illustrated by the following scenario.

2 THE SITUATIONS

2.1 Communicating with Someone Who has been Disconnected

Situation 1. *An emergency response team is sent out toward a disaster site by a command and control center. The team is approaching the disaster site when command and control learns that the site is chemically contaminated. Command and control tries to notify the emergency response team, but communication has been unexpectedly lost. How can communication be reestablished in the fastest manner?*

Before the connection broke, the communication distance with the team was practically instant. With the wireless communication broken, the direct communication distance between the command center and the car is effectively infinite, as shown in Fig. 6.

There is an indirect way to communicate. If the State Police are reachable by radio, they might be able to overtake the emergency response team and flag the team down. They are closer to an alternative, the city police, as measured by the amount of time it will take them to physically intercept the team and deliver a simple message.
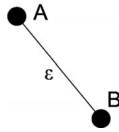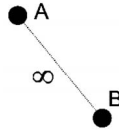
Fig. 7. Wireless communication.
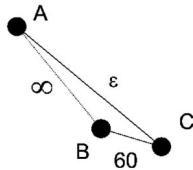

Fig. 8. Broken wireless communication.

Fig. 9. Working around the break.

Fig. 10. A person A wished to talk to B—but wishes to do so privately. Movement may be necessary to achieve this—either A can move to a secure phone or A can use D to courier a message.

Initially, if we are in contact with a resource using wireless communication, we have a sense of closeness due to the immediacy of contact (see Fig. 7).

When a connection is broken, our perception is we cannot communicate with B (see Fig. 8).

But, in a situation in which there is a larger set of resources, there probably is an indirect way to communicate—by calling someone else, C, who may attempt to intersect B, with guidance from A as to B's location.

We have labeled the link between A and C with an epsilon, indicating instant communication. C can intersect B in 60 seconds. We have mapped mobile resources onto a graph structure so that communicating to B becomes a shortest path problem (see Fig. 9).

The above situation can be described more generally as involving the concatenation of different modes of mobility and communication. There are many situations which fall into this class—for example, if we needed to communicate with someone on a ferry, then, in the absence of radio contact, we would need to either wait at a dock or use another boat to intercept it. In emergency situations, there may be a variety of modes of transportation and communication that might be combined in many ways—by mapping the problem to a graph, we increase our chances of finding a novel solution.

2.2 Finding the Shortest Private Route

In an adversarial situation, we may wish to communicate privately and this may change the route we use. We can specify different levels of privacy for different modes of communication and then we can look for the shortest communication distance at a particular level of privacy.

Situation 2. *Someone has the option of communicating in either a secure or insecure way—through couriers, over a cell phone, or over a secure line. What is the shortest secure way to communicate?*

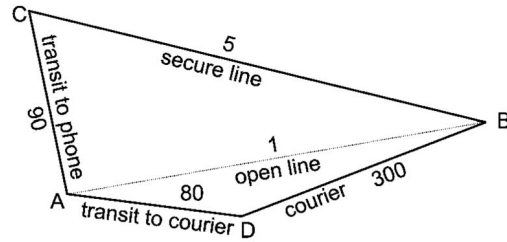For example, consider Fig. 10. A can communicate with B in one second over a phone at a low level of privacy or A can move to location C and can call from a secure telephone. Or, perhaps most privately, A can courier the message to B. In order to differentiate private from public communication paths, we can add an attribute to each edge indicating a level of security. Then, we can find the shortest path at a particular level of privacy by creating a new graph in which we first subtract edges at lower levels of privacy before performing a search.

2.3 Reducing Communication Distance through Planning

In Situation 1, we discussed a scenario in which someone unexpectedly lost a connection. But, in planning for emergencies, we may actually assume that we will lose connections.

Situation 3. *An institution anticipates it may encounter a major disaster in which the communication and transportation infrastructure may be damaged. How might the institution plan now for a response in the future?*

Some institutions create emergency plans which call for those who are disconnected to assemble at the closest of several prespecified locations, called the rally points [4]. Such plans seem like good ideas and there is anecdotal information that they are good, but there has been little analysis of why they are effective—and how to optimize such plans. Communication distance provides a way of discussing the problem. In an emergency situation, every member of a team may become temporarily isolated. There is a point at which a certain fraction (C) of the team reconvenes (say 1/3) and the organization can continue to function. We call such a fraction a communication critical mass. Reconvening can involve bringing people together in one location—or, in a set of electronically connected locations—so measuring communication distance can determine if critical mass has been achieved.

We need a measure for the overall communication distance of an organization of people, P. The communication between people can be represented as a graph G. In order to measure cohesion, we can use the diameter of the graph, which is the maximum node-to-node distance (in our case, person-to-person communication distance) [5]. We can pick a communication threshold D (say 60 seconds). Then, we can say the organization has communication critical mass when there exists $Q \subseteq P$, with Q's links shown by the subgraph $S \subseteq G$, such that $diameter(S) < D$ and $|Q| > |P| * C$.

It is obvious that, without any plan, the reassembly of a team of people will tend to be chaotic—for, without a
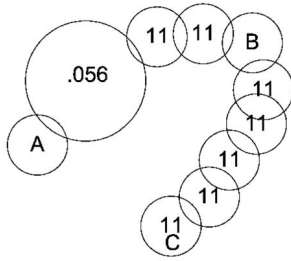
Fig. 11. A moving convoy communicates through an ad hoc network. The circles represent the area of bandwidth coverage, in this case either 56Kbps or 11Mbps. A can talk to B through the top chain or move to the bottom chain and communicate faster.

Fig. 12. A simplification of Fig. 11.

Fig. 13. The communication distances of Fig. 11.

location, many will proceed home, where they may find themselves without any easy way of communicating with other team members. Immediately after a disaster, without planning, all the edges of a communication graph go to some large number, perhaps a sizeable fraction of a day.

In contrast, if a rendezvous location has been chosen ahead of time, then the edges of the graph might be assigned the expected time of each team member to travel (by foot, if the transportation infrastructure is damaged) to the rendezvous location. It becomes possible to now reason about the situation. We might model the arrival times of all participants and project a time at which critical mass might be achieved. We can call the time from the disaster to the reestablishment of communication the *time to communication critical mass*—and, we can optimize alternative disaster plans by minimizing this measure. This problem can be seen as similar to optimization problems related to the placement of emergency and service facilities [6], the difference being that we are interested in minimizing communication, rather than travel, distance.

2.4 Making Movement Decisions in an Ad Hoc Network

In many emergency situations, the underlying infrastructure may be absent. Mobile ad hoc networks (see Fig. 11) are attractive because they can be rapidly set up. These networks present some interesting problems.

Situation 4. *A wishes to communicate to B over an ad hoc network, as shown in Fig. 11. Should A move down toward the tail of the 11 megabit chain to A's southeast, to the location C?*

We note that, in ad hoc networks, our link between A and B is only as fast as the slowest link in between, in the current situation 56Kbps. Since travel times combine through addition and bandwidth times combine through parallel flow, in modeling the problem, we need to represent the entire channel to B. We are assuming a situation in which, perhaps to avoid disrupting the convoy, one cannot overtake vehicles directly ahead—the options are to proceed as usual or move laterally to the end of another chain.

We can abstract the diagram in Fig. 12.

Our current link between A and B shows a 56Kbps link and no way to physically link up with B, as might be the case if A and B are moving at the same speed. Our options are to either stay put or travel over to C. This traversal will
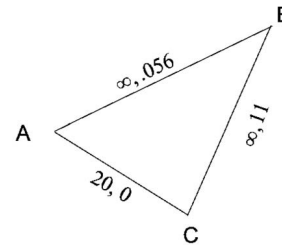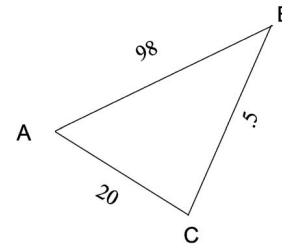
take 20 seconds, and there will be no connectivity during the travel time. For a given message size, we can compute the communication distances. Picking a message size of 5.5 megabits, we can transform Fig. 12 to Fig. 13.

The shortest path is 20.5, through C, which suggests moving to C is optimal.

However, there is uncertainty—moving over to C may result in a higher expected bandwidth—but might also incur more risk, given that the path from C to B involves more links and, perhaps, more opportunity for failure. We are leading toward a more thorough discussion of risk—for now, we note that, with the proper utility function, the risk-averseness of the ad hoc mobile user might be factored into the decision above.

2.5 Deciding to Move to a Higher Bandwidth Connection

Situation 5. *A person (A) working from home over a dial-up modem has a need to transfer a large 100 megabyte file to a colleague (B), who is working in an office. A is willing to travel and is near two coffee shops offering Internet access. What should A do to get the file to B the fastest?*

Note there is an underlying assumption that makes the situation a problem—that the bandwidth of network connections at different geographic locations may vary drastically. If every network connection in the world only supplied 56Kbps access, then the solution is always easy—find the nearest connection. But, it is currently the case, and probably will be for a long time, that network bandwidth varies greatly depending on where one is physically. The above situation is similar to a minor quandary that the reader may have faced when traveling—deciding whether or not it makes sense to move from the slow dial-up connection of a hotel room to the high-speed communication of a nearby coffee shop. There are analogous situations with larger social impact—for example, an emergency responder who needs to receive a map or photograph in the middle of a disaster.
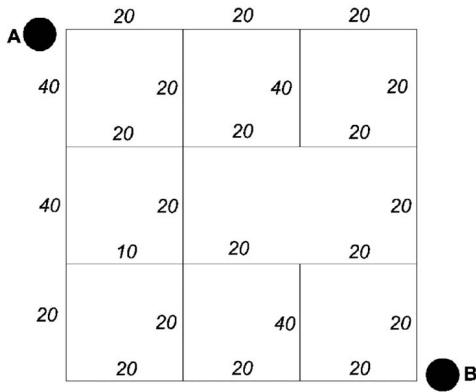
Fig. 14. A road map. A is moving, B is stationary.

Fig. 16. Each hotspot is connected through a broader network such as the Internet.

In order to solve this and the more complex problem that will follow, we will, through a series of transformations, construct a graph which, when traversed using a shortest path algorithm, will yield our solution. We are making simplifying assumptions—that we have available accurate expected automobile traffic speeds and expected bandwidth. Later, we will discuss the issues of dynamics and uncertainty.

First, we can create a spatial graph G_S representing roads in the area. Associated with each vertex are map coordinates. Associated with each edge is a measure of time to traverse the road in seconds. Mobile nodes such as A can move over the road network.

On the road map (Fig. 14), we can imagine placing a set of stationary communication hubs, Z, in this case, a home (w), an office (z), and hotspots (x, y), through which we can communicate over a shared network such as the Internet (see Fig. 15).

We can develop another graph, showing expected bandwidth between each point of the network, shown in Fig. 16.

We are really looking for the bandwidth to z from w, x, and y. So, we can transform this bandwidth graph again, shown in Fig. 17.

Now, we can combine the graphs. For a particular message size M, we can transform the edge of the label of the bandwidth graph into the seconds to transmit the file,
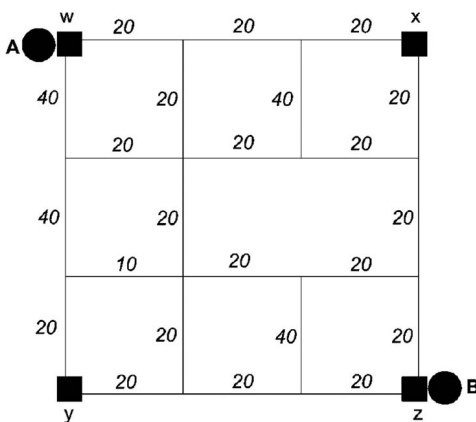
yielding a graph we name G_Z. Then, we can create a graph $G = G_S + G_Z$. Fig. 18 shows G for $M = 100$ megabytes. Each edge now is labeled with a communication distance.

The paths in G can be added. We can inspect for the shortest path—in Situation 5, with a 100 megabyte file, the solution is to drive along the path shown in Fig. 18. For a different size file, the graph will be different, as will the solution. For very short files, A is best off staying home. For an extremely large file, A is best off delivering the file in person to B.

2.6 Moving and Communicating at the Same Time

In the previous situation, we assumed a file could not be transferred while driving. Here, we assume we can transfer a file while driving. We assume we can maintain our connection and our bandwidth just increases or decreases when we reach the hotspots. In other words, we are assuming that there is some underlying mechanism in place which provides seamless mobility. This is rarely possible

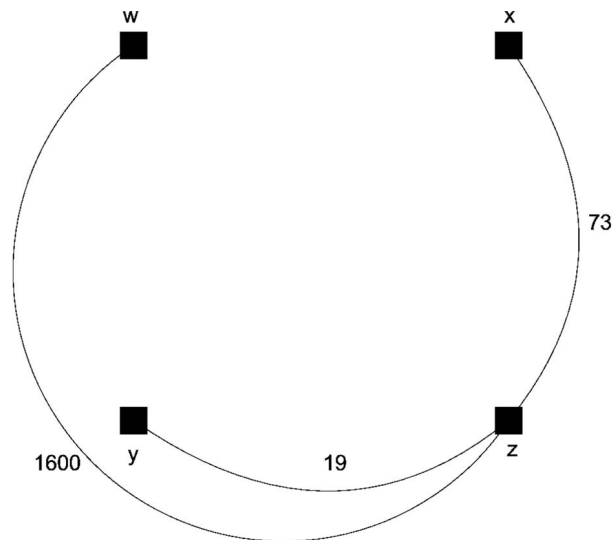Fig. 15. Four hotspots are situated on the corners of the map.

Fig. 17. Bandwidth from Fig. 16 is converted into seconds to transmit a message of a particular size to the destination.

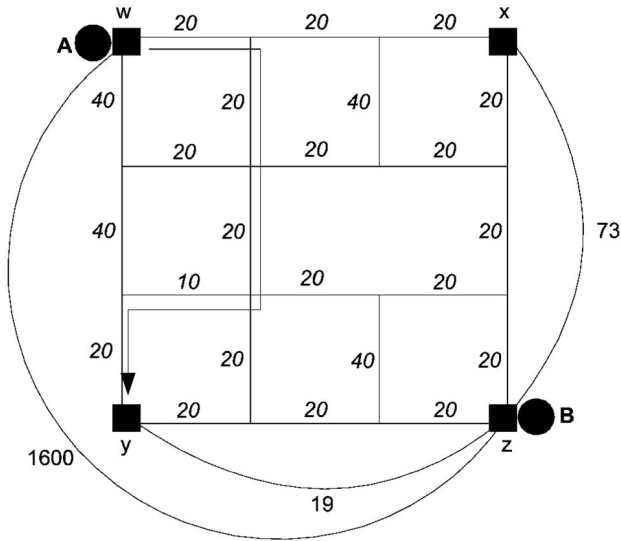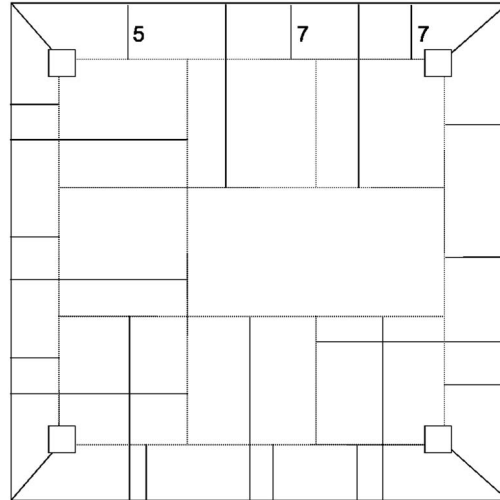Fig. 18. The graphs are combined; the edge labels are all times.

Fig. 19. Every route on the map has associated with it an expected bandwidth—which is really the expected bandwidth of communication up to and across a communication backbone.

today, but we are anticipating that a combination of in-car communication systems and WiFi roaming standards will converge in the future to create such an infrastructure [7].

Situation 6. *A person (A) has a mobile device and realizes he needs to transfer a large 100 megabyte file to a colleague (B), who is working in an office. A is willing to travel and is near two coffee shops offering Internet access. These coffee shops, the road network, A's home, and B's workplace, all have compatible technologies so that A can maintain a connection while traveling. The fixed locations all provide higher bandwidth than the road network. What should A do to get the file to B the fastest?*

We need to augment our map because we want to know what our bandwidth is as we drive. Let us assume we have a map of the expected available bandwidth along each road. This is a futuristic assumption—such maps are currently rare. However, they can be built [8]. We are assuming that the bandwidth along each edge represents the available bandwidth out to a high-speed backbone link. Fig. 19 shows this—each of the top segments is labeled with the bandwidth connection out to the backbone.

To simplify our diagram, we can label each edge in the graph with two numbers—the time to travel across the edge and the rate of communication while on the edge, shown in bold in Fig. 20. Note that, although this diagram looks similar to network flow representations in which flows and constraints are shown for each edge, it is not the same. The rate of communication is not a flow between the intersections, but instead represents bandwidth out through a backbone to a destination, in this case B.

As we drive, we are transmitting and the file size diminishes as we move. As we reach each intersection, our accumulated time has increased and our remaining file size has decreased. If we approach a particular intersection along two different paths, it is possible that one path will have a smaller accumulated time, but a larger leftover file. For example, in Fig. 20, consider the approach to intersection q. If we approach down the left side, we will spend

80 seconds and will have transmitted 120 megabits (40 seconds * 1Mb + 40 seconds * 2Mbps). If we approach from the path indicated on Fig. 18, we will have spent 70 seconds but will have transmitted 70 megabits. Which path is shorter? The answer is dependent on the bandwidth of our hotspot for, if the final bandwidth is high enough, we should choose the second path, with the lowest accumulated time, over the first. We develop this more formally, creating the structure we will need to solve Situation 6 and, along the way, making some more formal observations about Situation 5.

We define an attributed spatial graph $G = (V, E)$. Vertices in V can represent both intersections and moving nodes such as people. Included in V are mobile nodes, often labeled in our examples as A and B. Edges in E are pairs of vertices $V \times V$ and can represent roads or communication links between two vertices. We access attributes associated with E using a traversal function $t: E \rightarrow S$ and a bandwidth function $bw: E \rightarrow B$, where t yields the number of seconds
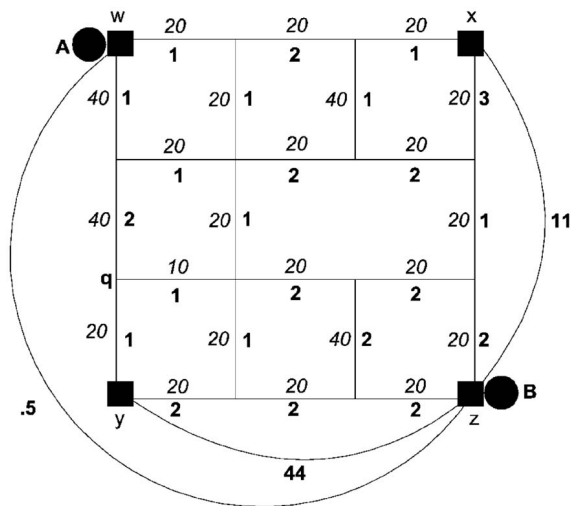

Fig. 20. The graph is labeled with two costs for each edge—one a physical travel time, one a wireless communication rate.
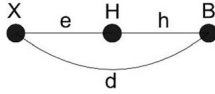
Fig. 21. We will move from X to H only if the speed increase we get from this makes our overall message transmission time less than if we stay where we are and transmit over link d.

to physically traverse edge e and bw yields the expected bandwidth in bits per second over the length of edge e. It is apparent that, from this structure, we can derive the expected amount of bits transmitted while traversing an edge by computing $t(e) * bw(e)$. For problems such as Situation 5 in which each edge can be used for physical or data movement, but not both, we can define a function

$$communicationDistance(e, M) = \min\left(t(e), \frac{M}{bw(e)}\right),$$

where M = the message size in bits to be communicated.

By extension, we define the communication distance between A and B in a graph G for message size M as the shortest path in a graph labeled with communication distances at each edge.

When M is small and there is an electronic connection, then we have the cases discussed earlier in the paper where we described the communication distance as epsilon, representing an instant connection.

The problem is more complex when we wish to both traverse and send data at the same time. Starting from a base A, for a hotspot H, is it worth traveling to get there? We can consider a hotspot reachable with respect to a message size M if, along the shortest path p, we will transmit less than M bits. Otherwise, by the time we arrive at the hotspot, the message will have been sent.

There are two variations possible—today's version, in which we would need to restart file transmission when we arrived at a hotspot, and tomorrow's version, in which we can continue seamlessly. In the variation in which we would need to start our file transmission over from the beginning when we reach a hotspot, we need to first check its reachability. If it is reachable, we need to compare our time to transmit from the present location with the time to travel to the remote location and send the file.

Assuming a current position X, a hotspot H, and a destination B, with edges e, h, and d as shown in Fig. 21, then we move only if $t(e) + \frac{M}{bw(h)} < \frac{M}{bw(d)}$.

But, if we are in a situation in which we can maintain connectivity, then the problem is more difficult—for, as we drive toward our hotspot, our file is getting smaller. The path we should choose is dictated by the bandwidth of the hotspot we are driving toward. If the hotspot is really fast, then we are willing to drive low bandwidth links which get us to the hotspot quickly. We can define a spatial graph $G_S(V, E)$, a start and end point A, B in V, and a hotspot graph $G_Z(Z, F)$, where, for all f in F, all e in E, $bw(f) > bw(e)$. In other words, the bandwidth at the hotspots is higher than the bandwidth of all the other edges. We will consider each hotspot in turn, so, for each H in Z, with a corresponding edge f in F, we create a new graph $G_H = G + H + f$.

Consider that we are trying to optimize a path made up of two parts—a path p from A to H and an edge h from H to B, in which we take the remaining file and send it at the terminal bandwidth, $bw(h)$. The last edge weight is the second term in (1), in which we calculate the transmission time at the hotspot by subtracting the bits sent so far from the message size and dividing by the hotspot's bandwidth.

$$pathLength(p_H) = \sum_{e \in p} t(e) + \frac{M - \sum_{e \in p} t(e) * bw(e)}{bw(h)}. \quad (1)$$

In order to make sense, the bandwidth at location H must be higher than bandwidth elsewhere. For, if there is a road along the way with higher bandwidth, A should just pull over and transfer the file from the side of the road.

In comparing paths, we need a way of taking into account a potential high transfer rate at the end of the journey. We do so by creating a new graph G' with an edge time measure t$'$

$$t'(e') = t(e) - \frac{t(e) * bw(e)}{bw(h)}. \quad (2)$$

The edge from the hotspot H to B is now:

$$t'(h) = \frac{M}{bw(h)}. \quad (3)$$

Essentially, (3) is the time to transmit the entire message at the terminal bandwidth. Equation (2) offsets the time traveled along an edge by the future time value of the bits transmitted on each edge. Adding (2) and (3), the total path length is:

$$pathLength(p_H) = \sum_{e \in p} \left(t(e) - \frac{t(e) * bw(e)}{bw(h)} \right) + \frac{M}{bw(h)}, \quad (4)$$

which is equivalent to (1). We can view the traversal time t as a cost and the number of bits transferred as a benefit—the benefit is relative to the terminating bandwidth. We do not want to add many seconds to a journey if we are transmitting at a rate much lower than the terminating rate—in such a case, we want to bias toward traveling to the hotspot as quickly as possible.

Equation (4) can be used to create a graph traversable with an efficient shortest paths algorithm and all edge weights except the last are independent of the message size. We show this transformation in Fig. 22 for our example, using the bandwidth of y, 44Mbps.

More formally, we claim that the shortest path in the transformed graph G' is the same as the shortest path in the Graph G. Our proof is by contradiction. For, if the shortest path is not the same, there must exist at least one pair of paths p in G, p' in G', with corresponding edges e, e', such that $pathLength(p) <> pathLength(p')$. For, if all the paths are the same length in both graphs, then the shortest paths would also be the same. But, for all p, $pathLength(p) = pathLength(p')$ for, as we have noted, (1) = (4):
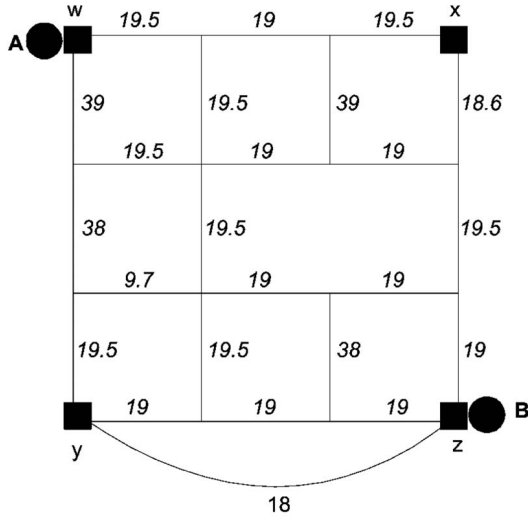
Fig. 22. Travel and bandwidth times are converted into a generalized cost.

$$\sum_{e \in p_H} t(e) + \frac{M - \sum_{e \in p_H} t(e) * bw(e)}{bw(h)} =$$
$$\sum_{e \in p'_H} \left(\left(t(e) - \frac{t(e) * bw(e)}{bw(h)} \right) + \frac{M}{bw(h)} \right).$$

Now, we discuss how to find the shortest path in G'. Note that there are paths in the graph in which we will have exhausted the file length before we even arrive at the terminating node. Depending on the application, we may want to reject such paths—in label-based shortest path algorithms such as Dijkstra's, we can keep, at our node labels, the size of the file remaining and searches which go beyond the message size can be truncated.

Our algorithm then is the following: For a graph $G(V, E)$, a source and destination node A and B, and a graph $G_Z(Z, F)$ containing a set of hotspots, we can define a transformation $G \times Z \to G'$ mapping the edges of G and G_Z into weights using (2). We can then look for shortest paths through each hotspot in turn:

$$\forall H \in Z, G'_H := \ transform(G, H); p_H := shortest\_path(G'_H).$$

Then, we can take the shortest of the shortest paths through each of the hotspots. The size of the set Z in practice will be small and the algorithm complexity is bounded by the shortest path algorithm, which can be any label-based one such as Dijkstra [9] since the edge costs are all positive. Implementations of Dijkstra's algorithm differ in complexity according to the data structure used, which in turn may depend on the nature of the weights. If the weights are positive and bounded by some number W, as they are in the situations we have discussed, and we can approximate the costs to integers, as would be reasonable in these situations, then we can use a Dijkstra implementation [10] with complexity $O(|E| + |V|\sqrt{\lg W})$. With such an implementation, the complexity of this algorithm would be $O(|Z|(|E| + |V|\sqrt{\lg W}))$.

While the situation we have discussed is a new problem to best of the author's knowledge, there is a large catalog of problems that can be transformed into shortest path problems [11], [12]. What we have done is created a generalized cost function; not all two-dimensional problems have such solutions—if the criterion are truly separate, then the problem may prove intractable [13], [14].

3   EXTENSIONS UNDER MORE COMPLEX ASSUMPTIONS

3.1   Dynamics

In most environments, both travel times and bandwidth available will change over time. In some situations, we may know exactly the changes that will occur. For example, in the case of train schedules, we know how long the wait time for the next departure and transit time between stops will be.

The model we presented can be extended to consider these dynamic cases by using Space-Time Networks [14]. The basic idea is the following—we quantize the time under study into discrete increments—for example, every five minutes. We then create a vector of time delays and generalized costs according to this quantization. Next, we create a new graph in which each node is replicated according to the number of time increments being studied. Edges between the time increments correspond to time delays.

Such a technique is frequently used to model train timetables—and provides us a way of modeling communication distance in such settings. It is straightforward to map the technique of the previous section into a Space-Time Network. A shortest path algorithm can still be used on the new graph. The complexity of the algorithm goes up as a function of the new edges created, which is, in turn, determined by the number of time periods under examination. Space-Time Networks could also be used to model automobile traffic—but, there is greater uncertainty and the techniques discussed next may be more applicable.

3.2   Uncertainty

In planning an automobile trip during rush hour, our ability to predict the future will be limited. Ideally, we would like the user of a decision aid to be able to indicate a level of risk adversity, either in general or per situation, and then recommend paths consistent with that level. Loui [15] analyzes several methods for creating and then finding shortest paths in graphs with stochastic weights. He points out that, given a graph where the weights on each edge are independent random variables with known distributions, it is possible to construct a utility function which will map the edge weights in such a way that risk is considered. The new graph can be solved using shortest path algorithms such as Dijkstra. The actual utility function must be either affine linear or exponential—in other words, it must be absolute risk adverse [15], [16], [17].

Whether or not such an approach can be used on the communication distance problem in a particular situation is a subject for future research. The utility function might prove too restrictive. And, the initial assumptions about independence may not hold. Still, Loui's approach offers a method for performing experiments on risk averseness as it relates to communication distance. Loui also points out that, with certain assumptions about distributions, a risk dimension can be traded for two deterministic dimensions. In some settings, communication distance may be just one of many dimensions being considered (for example, fuel cost might be important) and, in these

cases, a multidimensional approach (with its associated higher complexity) may be necessary and the mapping of risk into deterministic dimensions would make sense. Other researchers have attacked the shortest path problem in stochastic graphs using a variety of techniques and heuristics [1], [18], [19], [20], [21].

Are there easier ways? One common way of dealing with uncertainty is to acknowledge that costs are random variables—but immediately take the expected values of the variables and proceed deterministically. When we earlier defined the bandwidth measurement we used as the expected bandwidth, we were implicitly invoking such a strategy. The advantage of doing so is, of course, simplicity in both analysis and algorithm design. However, by taking the expected value, we are ignoring the nature of the distribution.

In our situations, when presented with two paths, one slightly shorter than another in terms of communication distance, we imagine a typical commuter will pick the shorter one. But, if we were to tell the commuter the shorter path involves driving through a downtown area at the beginning of rush hour, that commuter might opt for a path that is slightly longer, but less likely to trap the commuter in a traffic jam. Similarly, a decision aid might suggest a route that would take a commuter through an area notorious for spotty wireless coverage—the commuter might prefer a longer but safer path.

The reader might speculate that a simple heuristic which points a commuter in a certain direction might work. Such heuristics are difficult to construct as shortest routes are not straightforward—they often involve backtracking to find highway entrances [22], [23]. However, heuristics will be important at a higher level. When information proves inaccurate—or disappears—through sensor malfunction or failure, then it is clear that risk has increased. If, for example, one finds while traveling that one's GPS unit is inaccurate, then one probably should use a heuristic to decide on whether or not to continue using the decision aid. If the mislocations are greater than the distance between blocks, the heuristic might be to immediately head toward a rally point, perhaps the closest visible hotspot. Such a heuristic might be most important for emergency responders—who might rely both on the algorithm discussed in Situation 6, as well as the contingency planning discussed in Situation 3.

3.3 Pragmatics

One might ask what evidence there is as to our ability to predict either traffic patterns or bandwidth. Wireless mobility technology is new and changing and data about expected bandwidth is scant. In contrast, automobile traffic modeling is a venerable field. However, drawing from the automobile literature, one can make a case either way. In looking for predictors of trip duration, some researchers have found that, counterintuitively, present measurement of traffic flow does better than a projection based on time of day—particularly in rush hour [24]. This would support a strategy of continually rerunning a deterministic shortest path decision aid based on real-time sensor data. On the other hand, most modelers of traffic report prediction success [25] and advocate a stochastic approach.

One might note that, if everyone has a decision aid with the above algorithm, that the aids may be guiding all participants to the same center, which would both increase traffic and reduce bandwidth.

The extent to which this might happen has a lot to do with the configuration of hotspots. If multiple high bandwidth centers are located on the edges of a town, for example, then the algorithm might neatly distribute drivers. If, in contrast, there is one or a small set of centers all located along a particular highway, then congestion might occur, in which case, the simple model discussed here could be extended to consider overall network capacity.

The technology that might create this problem could also solve it—for, if users of the system were willing to share their intended destinations with others on the road, then the decision aid could take into account the projected loads on the communication and transportation network.

4 RELATED WORK

The 1894 work of Cooley mentioned earlier is closely related to this work. Even though he argued for the separation of transportation and communication concerns, he wrote that "a market is the area over which competition extends and, though it has no sharp boundaries, it is clearly determined by the facility of transportation and communication" [26]. He noted that a central group of wholesalers could support a 100 mile radius of retailers if they had both a telephone and a way of getting goods to the retailer in a few hours. So, he described how transportation and communication can be used together and can broadly affect market behavior.

Moving to the present, ad hoc network protocols try to find the quickest path to communicate—and are more complex than fixed networks as mobility generates constant change. Location information is used in many protocols, but mainly as a way of limiting link status update frequency [27]. Some researchers have begun to consider explicitly directing where mobile nodes travel. In recent work, two alternative relay schemes are discussed —one in which a set of nodes acts like a snake, maintaining fixed distances, and another where the nodes act like runners between different parts of the network. In their analysis, runners work better [28]. In the same vein, others have discussed how, with knowledge of relative location, messages can relayed through the active movement of members of the network [29].

At a higher level, there has been a body of work focused on using mobile networks for the distribution of large data files. Research into the creation of infostations and other mobility caching systems, all recognize the transportation aspect of mobile communication [30]. In a very specific scenario, researchers describe a system in which individuals who walk by associated hotspots might receive data as they walk, both from the hotspot and from information intentionally cached on the PDAs of other pedestrians [31]. The authors point out that most ad hoc network simulations assume no knowledge of the node's future movement path and that this assumption is not valid in many situations.

Work on infostations has been described as an instantiation of the trade-off between capacity and delay [32]. In other words, if we are willing to hold off on a communication, then an architecture of small radius, very high bandwidth locations may provide us greater capacity.

While the original work on infostations contemplated a stationary network, some recent modeling has shown that peer-to-peer networks can perform a similar function—data can move quickly between two passing mobile nodes which connect and disconnect—then the data can hop again so that in areas of high node density, data diffuses rapidly [33]. We observe that this idea can be seen in the following way as related to communication distance. For large messages, one might actually be better off with an intermittent, but high bandwidth connection, as envisioned in infostation work. Communication distance gives a possible way of evaluating this trade off—given a set of expected messages sizes, travel sources/destinations, and alternative, equal-cost configurations of infostations, one could decide on infostation placement by minimizing the communication distance for a set of mobile users.

The concept of a personal mobility advisor is developed in a system called Mobiplan [34]. This system analyzes both costs and environmental impact. The work described here might be used to integrate communication benefits and costs into mobility advice. Other research has looked at personal transportation decisions in depth [35], [36]. Recently, high-speed train owners have expressed plans to add WiFi capabilities as a way to increase ridership [37]. Research suggests that commuters are willing to travel longer if they are more productive while they are moving [38]. Our scenarios involving both driving and transmitting at the same time had WiFi in mind. Our assumption that these networks might allow roaming is not yet a reality, yet it is being worked on [7].

5   IMPLICATIONS

The prevalence of GPS and wireless communication creates a new set of situations concerning our movement through space. We have detailed six of these situations and shown how to map these situations into simple graph representations. One feature of our graph representations is a mixing of transportation time and bandwidth time, leading toward a single measure which we call communication distance.

We have shown how this measure may be useful in dispatching situations. We suggest that individuals will not think to use all of the potential resources at their disposal to reduce communication distance. If this is true, then the decision aids we describe and have prototyped may have a positive impact on emergency dispatching. The measure may also prove useful in planning for disasters in which communication infrastructure may be lost.

We have cited evidence that the use of bandwidth maps may become a consideration in personal route planning. If this is true, there are some interesting public policy implications. If individuals are more likely to take public transportation if it provides more bandwidth than private transportation, then building free or inexpensive public bandwidth might be used to change transportation patterns. In a similar way, road system traffic might be altered by improving connectivity or increasing bandwidth along certain corridors, either permanently or dynamically.

ACKNOWLEDGEMENTS

REFERENCES

[1] L. Fu, "An Adaptive Routing Algorithm for In-Vehicle Route Guidance Systems with Real-Time Information," *Transportation Research, Part B: Methodological,* vol. 35B, no. 8, pp. 749-765, 2001.

[2] L.L. Peterson and B.S. Davie, *Computer Networks: A Systems Approach,* third ed. Boston: Morgan Kaufmann, 2003.

[3] C.H. Cooley, *The Theory of Transportation,* pp. 70-71. Baltimore: Am. Economics Assoc., 1894.

[4] DOD, "US Army Survival Manual: FM 21-76," US Dept. of Defense, 1992.

[5] F. Buckley and F. Harary, *Distance in Graphs.* Redwood City, Calif.: Addison-Wesley, 1990.

[6] R.F. Love, J.G. Morris, and G.O. Wesolowsky, *Facilities Location: Models and Method.* New York: North-Holland, 1988.

[7] S.J. Vaughan-Nichols, "The Challenge of Wi-Fi Roaming," *Computer,* vol. 36, no. 7, pp. 17-19, July 2003.

[8] T. Kamakaris, "A Bandwidth Map for the Stevens Campus," http://www.stevens.edu/winsec/resource/throughput.jpg, 2003.

[9] E.W. Dijkstra, "A Note on Two Problems in Connection with Graphs," *Numeriche Mathematik 1,* pp. 269-271, 1959.

[10] R.K. Ahuja, K. Mehlhorn, J.B. Orlin, and R.E. Tarjan, "Faster Algorithms for the Shortest Path Problem," *J. ACM,* vol. 37, no. 2, pp. 213-223, 1990.

[11] M. Minoux, "Structures Algebriques Generalisees des Problemes de Cheminement dans les Graphes," *Recherche Operationnelle,* vol. 10, no. 6, pp. 33-62, 1976.

[12] M. Gondran and M. Minoux, *Graphs and Algorithms.* Chichester (West Sussex); New York: Wiley, 1984.

[13] Y. Israeli and A. Ceder, "Multi-Objective Approach for Designing Transit Routes with Frequencies," *Advanced Methods in Transportation Analysis,* L. Bianco and P. Toth, eds., pp. 157-182, Berlin: Springer, 1996.

[14] S. Pallottino and M.G. Scutellà, "Shortest Path Algorithms in Transportation Models: Classical and Innovative Aspects," *Equilibrium and Advanced Transportation Modelling,* P. Marcotte and S. Nguyen, eds., pp. 245-281, Boston: Kluwer, 1998.

[15] R.P. Loui, "Optimal Paths in Graphs with Stochastic or Multidimensional Weights," *Comm. ACM,* vol. 26, no. 9, pp. 670-676, 1983.

[16] J. Pratt, "Risk Aversion in the Small and in the Large," *Econometrica,* vol. 32, pp. 122-136, 1964.

[17] K.J. Arrow, *Essays in the Theory of Risk-Bearing.* Chicago: Markham, 1971.

[18] E. Miller-Hooks and H. Mahmassani, "Path Comparisons for A Priori and Time-Adaptive Decisions in Stochastic, Time-Varying Networks," *European J. Operational Research,* vol. 146, no. 1, pp. 67-82, 2003.

[19] P. Jaillet, "Shortest Path Problems with Node Failures," *Networks,* vol. 22, pp. 589-605, 1992.

[20] I. Murthy and S. Sarkar, "A Relaxation-Based Pruning Technique for a Class of Stochastic Shortest Path Problems," *Transportation Science,* vol. 30, no. 3, pp. 220-236, 1996.

[21] R. Hall, "The Fastest Path through a Network with Random Time-Dependent Travel Time," *Transportation Science,* vol. 20, no. 3, pp. 182-188, 1986.

[22] K. Nagel, "Multi-Agent Transportation Simulations," http://www.sim.inf.ethz.ch/papers/book/, 2004.

[23] D. Park and R. Rilett, "Identifying Multiple and Reasonable Paths in Transportation Networks: A Heuristic Approach," *Transportation Research Records,* vol. 1607, pp. 31-37, 1997.

[24] S. Handley, P. Langley, and F. Rauscher, "Learning to Predict the Duration of an Automobile Trip," *Proc. Int'l Conf. Knowledge Discovery and Data Mining,* 1998.

[25] J. Sacks, N.M. Rouphail, B. Park, and P. Thakuriah, "Statistically-Based Validation of Computer Simulation Models in Traffic Operations and Management," *J. Transportation and Statistics,* vol. 5, no. 1, 2002.

[26] C.H. Cooley, *The Theory of Transportation,* p. 103. Baltimore, Md.: Am. Economics Assoc., 1894.

[27] M. Abolhasan, T. Wysocki, and E. Dutkiewicz, "A Review of Routing Protocols for Mobile Ad Hoc Networks," *Ad Hoc Networks,* vol. 2, pp. 1-22, 2004.

[28] I. Chatzigiannakis, S. Nikoletseas, N. Paspallis, P. Spirakis, and C. Zaroliagis, "An Experimental Study of Basic Communication Protocols in Ad-Hoc Mobile Networks," *Lecture Notes in Computer Science,* vol. 2141, pp. 159-169, 2001.

[29] Q. Li and D. Rus, "Message Relay in Disconnected Ad-Hoc Networks," *Proc. IEEE MASCOTS Workshop Mobility and Wireless Access,* 2002.

[30] D.J. Goodman, J. Borras, N.B. Mandayam, and R.D. Yates, "INFOSTATIONS: A New System Model for Data and Messaging Services," *Proc. Vehicular Technology Conf.,* 1997.

[31] S.B. Kodeswaran, O. Ratsimor, A. Joshi, T. Finin, and Y. Yesha, "Using Peer-to-Peer Data Routing for Infrastructure-Based Wireless Networks," *Proc. IEEE Int'l Conf. Pervasive Computing and Comm. (PerCom),* 2003.

[32] T. Small and Z. Haas, "The Shared Wireless Infostation Model—A New Ad Hoc Networking Paradigm (or Where There Is a Whale, there is a Way)," *Proc. ACM Int'l Symp. Mobile Ad Hoc Networking and Computing,* 2003.

[33] M. Papadopouli and H. Schulzrinne, "Effects of Power Conservation, Wireless Coverage and Cooperation on Data Dissemination among Mobile Devices," *Proc. Second ACM Int'l Symp. Mobile Ad Hoc Networking & Computing,* 2001.

[34] M. Kreitz, K.W. Axhausen, K.J. Beckmann, and M.M. Friedrich, "Mobiplan: An Internet-Based Personal Mobility Advisor," *Transport Policy,* vol. 9, no. 2, pp. 155-168, 2002.

[35] P. Jones, *Developments in Dynamic and Activity-Based Approaches to Travel Analysis.* Aldershot, England; Brookfield, Vt.: Avebury, 1990.

[36] T. Gèarling, T. Laitila, and K. Westin, *Theoretical Foundations of Travel Choice Modeling,* first ed. Amsterdam; New York: Elsevier, 1998.

[37] E. Nash, "GNER to Install WiFi on Trains," http://www.itweek.co.uk/News/1142986, 2003.

[38] V. Kaufmann, "Temps et Pratiques Modales. Le Plus Court Est-Il le Mieux? The Rationality of Perception and Modal Choice. Is Quickest Best?," *Recherche Transport Sécurité,* vol. 75, pp. 131-143, 2002.

Jeffrey V. Nickerson received the PhD degree in computer science from New York University. He is an associate professor at Stevens Institute of Technology. His research interests include decision aids for mobility, emergency response, human-robot interaction, and information systems design. He is a member of the ACM and the IEEE.

▷ **For more information on this or any other computing topic, please visit our Digital Library at** www.computer.org/publications/dlib.