Personal Travel Assistants and the World Wide Web

Jeremy Padmos and David Bernstein
Princeton University

August 1997

New Jersey TIDE Center
Directed by Prof. Louis J. Pignataro
New Jersey Institute of Technology
Newark, NJ
pignataro@admin.njit.edu
www.njtide.org

Personal Travel Assistants (PTAs) are hand-held computers designed to provide travelers with information of various kinds. Interest in PTAs has increased rapidly in recent months. This is in part because of improvements in the underlying Personal Digital Assistants (PDAs) on which they are based, and in part because of the realization that they can play a very important role in improving multimodal travel.

PTAs can be used in a wide range of situations. For example, at an airport a PTA could supply you with information about departures and arrivals, the location of airport services (e.g., restaurants, baggage claim areas, ground transportation), the availability of vacant spaces in different parking lots, and much more. On a city street, a PTA could supply you with information about points of interest, bus and subway routes/schedules, the current position and/or anticipated arrival times of buses and trains, the location of pedestrian walkways, and traffic congestion levels. At a commuter rail station a PTA could provide information about schedules, delays, alternative routes, and station services.

In general, as is in these examples, PTAs can either "stand alone" or be connected to external sources of information. However, there is little doubt that they are most valuable when connected to external sources of information. In our opinion, it is only then that they have a significant advantage over paper books and maps. Hence, much of the PTA development effort has been devoted to making this connection.

One very important issue in this regard is to ensure that PTAs can be taken from location to location and always remain connected to whatever external information is available (i.e., a "universal" PTA). As a result, attempts are now underway to develop the standards that would make this possible (1). However, we are somewhat pessimistic that such standards can be developed quickly enough. Indeed, when related technologies were introduced (e.g., electronic toll collection, in-vehicle navigation systems) many manufacturers used proprietary components, making it very difficult to develop standards.

Therefore, as part of Princeton University's Large-Scale Automobile Routing (PULSAR) Project, we have begun exploring ways in which the existing protocols being used on the Internet and the World Wide Web (WWW) can be used to develop a "universal" PTA. This paper presents some of our initial findings.

It begins with a discussion of the Internet and the WWW (focusing on the various protocols involved), and then considers whether these protocols are robust enough to meet the needs of PTAs. Some of the analysis is qualitative and some is quantitative. The qualitative analysis compares the kinds of information that PTAs need to manage and the kinds of information that the WWW has been used to disseminate. The quantitative analysis is based on a simulation of PTA use by pedestrians in a generic transportation terminal (e.g., an airport or rail station). Finally, it concludes with a discussion of future areas for future research.

THE INTERNET AND THE WWW

The words "Internet" and "World Wide Web" mean many different things to many different people. This makes it somewhat difficult to ask, in a generic way, whether the Internet and or the WWW can be used to develop a mobile PTA. Hence, we begin by explaining what we mean by these terms.

For the purposes of this paper, the Internet is a set of interconnected computers that communicate with each other using the *Internet Protocol* (IP) and the *Transmission Control Protocol* (TCP). In IP, datagrams are transmitted and routed between sources and destinations that are identified by a fixed address (called the IP address). The datagrams themselves may be transmitted from the source to the destination through intermediate computers, and each datagram may be broken up into *packets* that are sent along different routes. Interestingly enough, by itself, IP does not guarantee the delivery of datagrams. Hence, TCP is often used on top of IP. TCP essentially provides for the retransmission of any datagrams that are lost or corrupted. In TCP, the sender retransmits data until the receiver acknowledges that it has been received correctly. Hence, TCP/IP is inherently a bi-directional protocol.

The WWW is simply one application of the Internet. Most importantly, the WWW is the use of the Internet to transmit multimedia documents written in the *Hypertext Markup Language* (HTML) using the *Hypertext Transfer Protocol* (HTTP). HTML is a *markup language* that is used to describe the format of multimedia documents. HTML documents are displayed using a *browser*. A browser running on one computer can request an HTML file from another computer on the Internet using an HTTP session. The two computers at either end of an HTTP session play distinct roles. The *client* generates requests and receives responses. The *server*, on the other hand, receives requests and generate responses. For the purposes of this paper, clients request HTML documents and servers provide them.

The way all of this fits together is illustrated in Figure 1. The top half of this Figure shows the information flows associated with a user's request and the bottom half shows how how the information flows when that request is satisfied. Initially, most transactions on the WWW involved the client *pulling* information from the server as shown in Figure 1. Now, there is considerable interest in transactions that involve the server *pushing* information to the client and appropriate protocols have been and are being developed for this purpose.

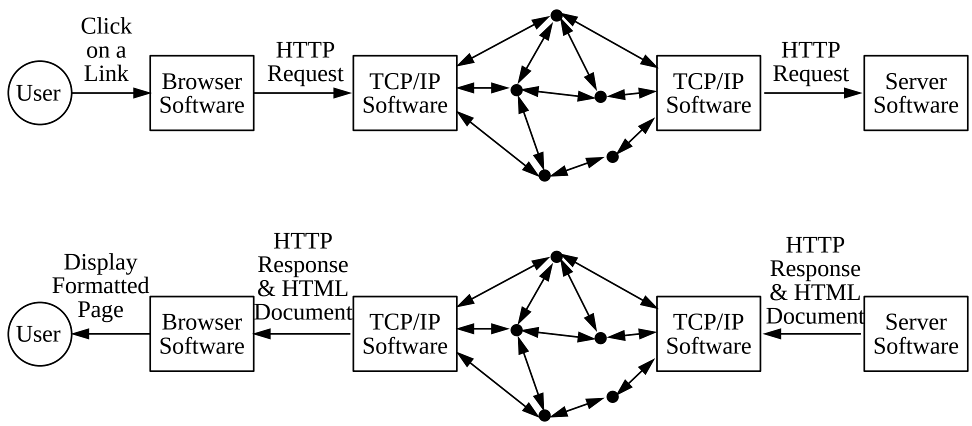

Figure 1: One View of the Internet and WWW

With this description of the Internet/WWW in mind, two questions need to be addressed. First, can the Internet/WWW be used to disseminate the wide variety of information that is necessary in a transportation context? This is really a qualitative question about the flexibility of HTML. Second, can the Internet/WWW be used to disseminate information to mobile PTAs? This is both a qualitative question about the flexibility of TCP/IP and HTTP, and a quantitative question about the "capacity" of such a system. We will now attempt to address each in turn.

TRANSPORT INFORMATION ON THE WWW

The number of ways that PTAs might be used in practice is almost limitless. In fact, there are so many potential applications that it is impossible to list them all. However, we can categorize the different applications based on their information requirements. Several examples are presented in Table 1.

First, multimedia is more than just a buzzword in the PTA context. Given the amount of information that travelers want/need, it is essential to consider the most appropriate medium. In general, four media are required – *text*, *graphics*/*maps*, *audio*, and *video*. Further, both text and graphics may need to be either fixed (i.e., tables and figures) or animated (i.e., tickers/teletypes, time-varying maps).

	Static	Quasi-Static	Dynamic
Text	Awareness Campaigns	Road Closings	Arrival/Departure Boards
Graphics	Street Maps	Traffic Condition Maps	Vehicle Location Maps
Video	Aerial Photos	Traffic Snapshots	Weather Reports
Audio	Road Closings	Traffic Conditions	Reservation Systems

Table 1: A Categorization of Some Sample Applications

Second, PTAs require *static*, *quasi-static*, and *dynamic* information. Static information does not change. For example, the information on this page is static. Quasi-static information changes, but infrequently. Hence, from the standpoint of a particular user at a particular point in time, it is essentially static. Finally, dynamic information changes rapidly.

In order for the WWW to be a viable option for connecting PTAs to external sources of information, HTML must be able to satisfy all of these requirements and it can. In fact, as discussed in (2), it is already being used in all of these ways. In addition, new ways of better meeting these requirements are being developed everyday. Hence, it is clear that the Internet/WWW be used to disseminate the wide variety of information that is necessary in a transportation context. All that remains is to consider whether the Internet/WWW be used to disseminate this information to *mobile* PTAs.

THE WIRELESS WWW

There are a number of ways [e.g., Point-to-Point Protocol (PPP), Indirect TCP (I-TCP)] to connect a mobile computer to the WWW using a two-way communication system such as a cellular telephone/modem and the tradeoffs involved are not trivial. Nonetheless, in this paper we are not concerned with the details of this connection (for a discussion of these issues see 3 or 4). Instead, we are concerned with whether such connections can handle the demands placed on them in a PTA context.

To help answer this question, we simulated PTA use by pedestrians at a generic transportation "terminal" (e.g., a rail terminal or airport) illustrated in Figure 2. In this simulation, pedestrians (some with PTAs and some without) entered a 10 meter wide “terminal” from the left, walked 100 meters, and exited to the right. Users of PTAs were able to request and/or receive information at any point in the transmission zone.

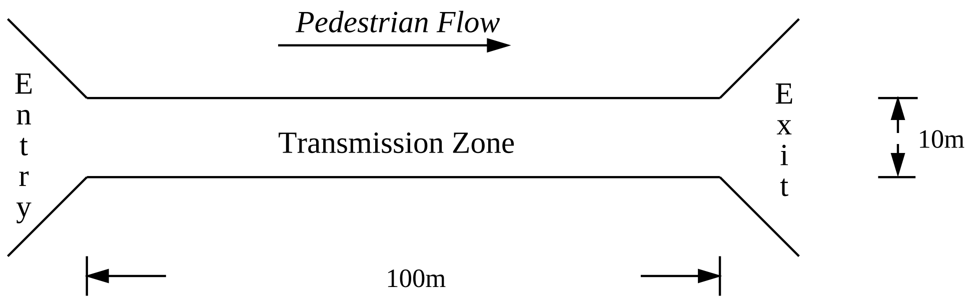

Figure 2: The Simulated Terminal

Pedestrian arrivals were assumed to be Poisson distributed (over one hour intervals). The specific parameters used were based on actual passenger arrival rates at Newark International Airport. Arrival rates change each hour, with peaks at 9:00AM, noon, and 6:00PM.

Walking speeds, s, were modeled using a two-regime model of the form:

$$s = \begin{cases} 60.83\exp(-d/4.166) & \text{if } d \leq 1.075 \\ 36.78\ln(4.32/d) & \text{if } d > 1.075 \end{cases} \tag{1}$$

where d is the pedestrian density (measured in pedestrians per square meter) and s is measured in meters per minute (5).

The simulation was run 100 times for each scenario. For each run, the total simulation time was 4800 seconds. This included a 600 second warm-up period and a 600 second cool-down period. Thus, data were collected for only 3600 seconds (i.e., one hour).

The parameters of the simulation included: the number of HTML pages, N, which represented the total amount of information available; the transmission time per page, T, measured in seconds; the number of simultaneous connections that could be made (i.e., the number of requests that could be handled at one time), S; the balk time, B, which is the amount of time that a user is willing to wait for a particular page before making another request; and the probability that a particular user will make a request of the system, P. The typical range for these parameters are shown in Table 2.

Parameter	Typical Range
Number of pages (N)	5 - 10
Transmission time per page (T)	3-15 seconds
Max. number of connections (S)	8-128
Balk time (B)	15-30 seconds
Probability of a request (P)	0.1 - 0.7

Table 2: Simulation Parameters

Several different measures of performance were recorded for each run. These measures included: the user waiting time (between request and receipt), the PTA (or unit) waiting time (which may be different from the user waiting time when the PTA caches incoming pages), the number of balks, the number of unrequested deliveries (in pages), the amount of unsatisfied demand (in pages), and the advanced receipt time (for PTAs that store incoming pages).

The results of the simulation were quite interesting and showed that the WWW is a viable means of connecting PTAs to external sources of information. However, the simulation also showed that care must be taken when designing such a system.

Perhaps the most important performance measure is the average user waiting time. Figure 3 shows the average user waiting time for a system with 32 simultaneous connection and various values of P and T. As you can see, there can be considerable variation in average waiting times. When $T = 3$ there is virtually no waiting time. However, when T increases to 5 or 10, the average waiting time increases to a minute or more. Hence, relatively fast servers may be required in practice.

Another important performance measure is the number of unsatisfied requests. Figure 4 shows the average number of unsatisfied requests for various different parameter values (where the number next to the 2BV1 label indicates the number of simultaneous connections, S).

As you can see, even when the transmission time is as low as 3 seconds per page, there are a significant number of unsatisfied requests (on average) when the system can only support 8 simultaneous connections. Hence, a fairly large number of connections may be required in practice. This is because PTAs can generate a large number of similar requests in a short amount of time.

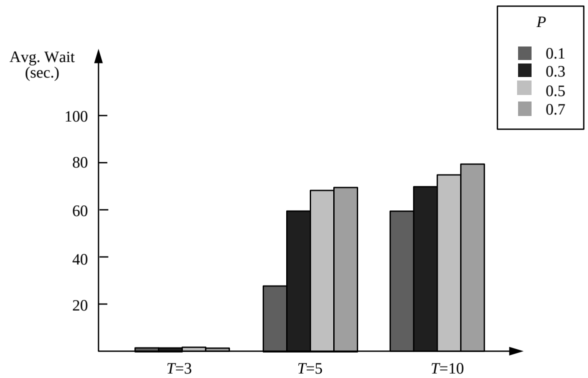

Figure 3: Average Waiting Time

USING ONE-WAY COMMUNICATIONS

In situations where the size of each request is small it is possible for PTAs to be *indirectly* connected to the WWW using one-way communications technologies. Of course, with one-way communications users do not actually make requests of the server. However, they can be made to think that they do. That is, a local host connected to the WWW can continuously broadcasts information to all PTAs within its range. When a user makes a request, the PTA either waits until the appropriate information is broadcast or it displays information that it has already cached.

A variety of possible technologies can be used to implement this kind of system including FM sideband radio data systems, diffuse infrared data links, and broadcast satellites. These kinds of systems are relatively inexpensive and can support an unlimited number of users simultaneously. Hence, it is important to compare these systems to those with a true two-way connection to the WWW. To that end, we used the same simulation model discussed above to consider five one-way systems divided into two categories.

Category 1A systems continuously broadcasted information on a fixed, sequential schedule. In type 1A version 1 (1AV1) systems, the PTA could receive and temporarily store everything that was transmitted. The user was presented with the first complete page that the PTA received, and all pages contained a complete index of links to other pages. Type 1AV2 systems were similar to 1AV1 systems except the PTA did not display any information until it received the first/home page (which contains a complete index of links to the other pages).

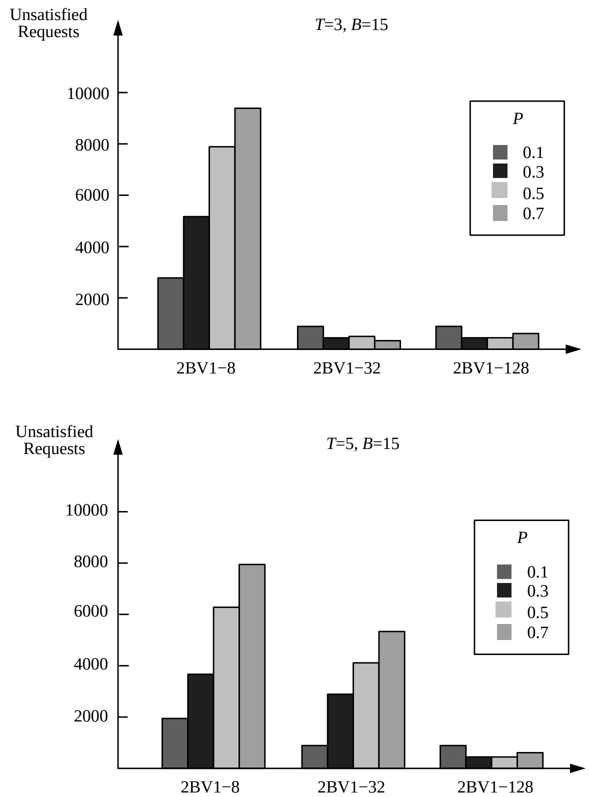

Figure 4: Unsatisfied Requests

Type 1AV3 systems were different from the other 1A systems in that the PTA had limited memory (and hence would be less expensive to produce). Specifically, type 1AV3 systems could only store one page of information. As a result, users of 1AV3 systems requesting a particular page had to wait until the next time it was broadcasted. On the other hand, users of 1AV1 and 1AV2 systems may not have waited for a page at all, since their PTA may have already received and stored it.

Category 1B systems differed from 1A systems in that the transmission schedule depended on the relative popularity of the pages. That is, more popular pages were transmitted more often. In type 1BV1 systems, dependencies across pages were used to determine transmission schedules. In other words, the probability that a particular page was transmitted was conditional on the page that was last transmitted. In 1BV2 systems these dependencies were ignored.

Not surprisingly, there were rather dramatic differences between the different

System	Avg. Waiting Time When $T = 3$	$T = 5$	$T = 10$
1AV1	0.996	2.170	5.629
1AV2	6.000	10.036	20.060
1AV3	6.000	10.036	20.060
1BV1	0.004	0.003	0.031
1BV2	0.004	0.003	0.031

Table 3: Avg. Waiting Times for One-Way Systems

systems. To provide some idea of these differences, we will first focus on a particular set of parameter values, namely $N = 5$, $T = 5$, $B = 20$, and $P = 0.1$. With these parameters, the average user wait time for 1AV1 tended to be in the 1.0-1.2 second range regardless of the page. However, for 1AV2 systems, users had to wait about 10 seconds for page 1, and about 1 second for the other pages. 1AV3 systems had even worse performance. Users had to wait about 10 seconds for page 1 and between 6 and 8 seconds for the other pages. In addition, 1AV3 resulted in a significant number of unsatisfied requests, unlike 1AV1 and 1AV2 which resulted in none. Systems 1BV1 and 1BV2 had very similar performance characteristics, and both outperformed the other one-way systems. User waiting times for pages 1-3 were all less than 0.4 seconds, and for pages 4 and 5 ranged between 1 and 1.2 seconds.

Of course, the performance of the different systems varied with the simulation parameters. Table 3 illustrates how the average user waiting time increased with T for $B = 15$ and $P = 0.3$. It also shows how it varied across the different systems.

In general, 1AV1 and the 1B systems performed quite well. Though not shown, these systems also had small numbers of balks and few unsatisfied demands. Hence, it does seem possible to connect PTAs to the WWW indirectly using one-way communications.

CONCLUSIONS AND FUTURE RESEARCH

It is clearly both possible and advantageous to connect PTAs to the WWW. HTML is certainly robust enough to handle most, if not all, of the information needs of PTAs. However, the question of how best to make the connection remains open. Two-way communications technologies have obvious advantages. Unfortunately, the capacity requirements of such systems may be quite large. One-way communications can be used to indirectly connect PTAs to external information sources like the WWW. This means that, for many applications, it may be possible to develop PTAs that are inexpensive to purchase and operate. The problem here is that such systems will require at least some new protocols. While IP can be used for one-way communications, it is not guaranteed to be error-free.

Of course, further research is still needed. For one thing, specific applications need to be simulated in greater detail. In particular, the "terminal" that we simulated had only one entry point and one exit point. Adding additional gates/tracks and more complicated travel patterns could qualitatively change the results we obtained.

In addition, more complicated two-way systems should be explored. In particular, two-way systems can, in principle, receive information using both narrowcast and broadcast technologies. That is, general information could be broadcast and user-specific information could be narrowcast. Although this would increase the purchase cost of the PTA, it could significantly decrease its operating cost.

Finally, a considerable amount of work still needs to be done to determine how best to disseminate static, quasi-static and dynamic information. For example, there is no technical reason to transmit a complete street network when only some attributes of the network change (e.g., the travel time). This same kind of issue arises in the transmission of Java applets. There is no particular reason to retransmit an entire applet when only some of the objects or methods in that applet have changed. Some work is already underway in this area (e.g., the Castanet system), but much more clearly needs to be done.

REFERENCES

1. Davies, P. G. Trenta, and J. Booth. "Development of Standards for Traffic/Travel Information Exchange", *Proceedings of the 1995 Annual Meeting of ITS America*, Vol. 1, pp. 275-284, 1995

2. Guensler, R. and D. Bernstein. "Transportation Resources on the Internet", *ITE Journal*, Vol. 66, No. 4, pp. 42-47, 1996.

3. Myles, A. and D. Skellern. "Comparison of Mobile Host Protocols for IP", *Journal of Internetworking Research and Experience*, Vol. 4, 1993.

4. Bakre, A. and B.R. Badrinath. "I-TCP: Indirect TCP for Mobile Hosts",

Working Paper, Rutgers University Department of Computer Science, DCS-TR-314, 1994.

5. Virkler, M.R. and S. Elayadath. “Pedestrian Speed-Flow-Density Relationships”, *Transportation Research Record*, Vol. 1438, pp. 51-58, 1994.