Distance Learning Technologies for Basic Education in Disadvantaged Areas

Junwen Lai Elisha Ziskind Fengzhou Zheng Yilei Shao Chi Zhang
Ming Zhang Nitin Garg Sumeet Sobti Randolph Wang
Department of Computer Science, Princeton University
Email: {lai,eziskind,zheng,yshao,chizhang,mzhang,nitin,sobti,rywang}@CS.Princeton.EDU

Arvind Krishnamurthy
Department of Computer Science, Yale University
Email: arvind@CS.Yale.EDU

Abstract: *Basic education plays a crucial role in uplifting disadvantaged areas from the grip of poverty. Distance learning systems are a promising option to deliver education to children who have access to none today. The main problem with the large scale deployment of such systems is either their reliance on broadband network access or their lack of interaction among students and teachers. This paper describes a software architecture that turns the global postal system into a generic digital communication channel that enables relatively sophisticated distance learning systems for disadvantaged areas. We also explore techniques to compliment this communication channel with storage elements and the traditional Internet, where it exists, to achieve a high bandwidth low latency hybrid network on top of which a wide array of other applications may be built.*
Keywords: distance learning, basic education, storage system, peer-to-peer system, postal system, networking, asynchronous communication.

1 Introduction

In a modern society that is increasingly driven by the exploitation and development of science, technology, and information, there has never been a greater need than there is today for well trained and educated people. Basic education is the very foundation upon which a nation's future rests. The building of more traditional brick-and-mortar schools is a natural way of investing in basic education. However, the high illiteracy rate in developing countries demands huge and continued resource commitments that is unlikely affordable by these countries.

Our project aims to utilize an innovative blend of the "high-tech" and the "low-tech." The low-tech includes approaches such as using the transportation of mobile storage devices via the postal system to provide high-bandwidth interaction without relying on a well-developed networking infrastructure. The high-tech approaches seek to exploit an integration of multiple types of communication channels (including both data transmissions through the Internet and storage devices carried by the postal system) in a peer-to-peer system that bridges the space and time gap among geographically distributed teachers and their students. If successful, we believe that the proposed effort will not only dramatically improve the basic education landscape, but also put in place a digital communication infrastructure that will serve the needs of a wider array of applications in health care, commerce, information dissemination, and entertainment.

The rest of this paper is structured as follows. Section 2 motivates our effort by briefly examining the challenges faced by traditional approaches, and the promises of a distance-learning approach. Section 3 describes our main technical innovations. Section 4 examines our effort in the context of related work and Section 5 concludes.

2 Motivation

A key difficulty faced by traditional approaches for improving basic education is "scalability." By some estimates (CIA, 2003), the population of the illiterate reaches nearly 200 million in China, and nearly 400 million in India. The condition of the female population is worse. There is also likely to be a sizable gap between literacy and a level of basic education that can open up reasonable job opportunities.

Due to the massive demand for basic education, addressing the education delivery issue using traditional means is likely to continue to face serious difficulties. The building of traditional brick-and-mortar schools is costly and slow. The sheer number of the illiterate demands huge and continued resource commitments over many years. There is a lack of teachers, especially those that are well trained, well qualified, and highly motivated. There are many needy remote regions, which may have difficulty attracting and retaining good teachers. Indeed, even developed nations are experiencing severe strain on educational resources. The U.S., for example, currently spends $880 billion per year on education; yet many believe that the U.S. basic education is inadequate.

As networking and other digital technologies continue to advance, one possible solution to the problem of basic education dissemination is distance learning. It holds many promises. Content that is developed once can be reused at many places many times. Without the costs such as those needed for stationing many specialized teachers at many local schools, distance learning may be more cost-effective. It may also be more scalable if the content reaches a large audience. Distance learning may allow children who need to help their parents with farm work to learn at a more flexible pace and schedule. For teachers who are enthusiastic about helping needy children, but are reluctant to live in less developed regions, distance teaching offers an attractive alternative. The system also gives volunteers the option of more flexible time commitment.

A positive byproduct of a digital distance learning undertaking, if well executed, is a software and hardware infrastructure that can serve other causes for the targeted disadvantaged areas. Some possible examples are: rudimentary health care, promotion of local commerce, and richer social interactions and entertainment. In general, it allows disadvantaged areas to be better connected to the larger world, and fosters innovative economic and social development. At the dawn of the industrial age, physical infrastructures in the form of railways, roads, and interstate highways served as powerful catalysts to bring progress to isolated parts of the world. In our current information age, digital infrastructures should serve equally critical roles. Distance learning provides a compelling initial impetus for the development of such an infrastructure.

3 Our Approaches

Much of the existing distance learning effort has focused on the delivery of higher education in relatively resource-rich environments. There are at least two challenges to overcome if we are to successfully adopt distance learning as a means of providing basic education: (1) the need of providing sophisticated modes of *interaction*; and (2) the need of adapting to *resource constraints*; in particular, communication bandwidth constraints. To a large extent, these are conflicting goals, and existing approaches tend to favor one at the expense of the other.

There are two simple extreme approaches. On one extreme, content is disseminated via TV broadcasts or on storage media (such as CDs or DVDs) that can be delivered by the postal system. Content that is less bandwidth-intensive may be downloaded via the Internet. The amount of interaction afforded by this extreme is typically either non-existent or very limited, therefore, such a self-guided approach may not work well for young learners, who may need closer supervision and more personal and immediate interaction with teachers. At the other extreme of the spectrum in terms of the amount of interaction, an existing approach to distance learning is similar to teleconferencing: a teacher and his students can engage in real-time video and audio interaction. The disadvantage of this approach is its consumption of a large amount of communication bandwidth—this cost and scalability handicap makes the approach unsuitable for resource-constrained environments in developing nations.

Our technical approaches serve two high-level purposes. One is to provide meaningful interaction without depending on traditional high-bandwidth network communication. The other is to maximize the reach and productivity of the human teachers. Our first technique seeks to turn alternative asynchronous high-bandwidth communication channels, such as the postal system, into generic digital communication mechanisms, on top of which sophisticated and bandwidth-intensive applications such as distance learning can be built. Our second technique, a distributed peer-to-peer system for supporting educational interactions, seeks to create a distributed "marketplace" where the supply and demand of the various educational services (such as the grading of homeworks) can be matched up. Such a marketplace not only may create a "network effect" akin to the open source model of development, but may also encourage increased specialization among the distributed teaching staff, which may lead to increased efficiency.

3.1 Providing Interactivity Using Alternative Asynchronous Communication Channels

In most rural areas, the availability or the bandwidth of Internet connectivity is severely limited. This situation is unlikely to change drastically in the near future since the improvement of wide-area bandwidth is constrained by factors such as how quickly ditches can be dug to bury fibers in the ground or how quickly satellites can be launched. Magnetic storage density, however, has been increasing at the annual rate of 60% to 100% for many years, and it is likely to continue in the foreseeable future. Shipping storage devices by the postal system, therefore, provides a High Latency High Bandwidth (HLHB) channel. In comparison, the Internet can be viewed

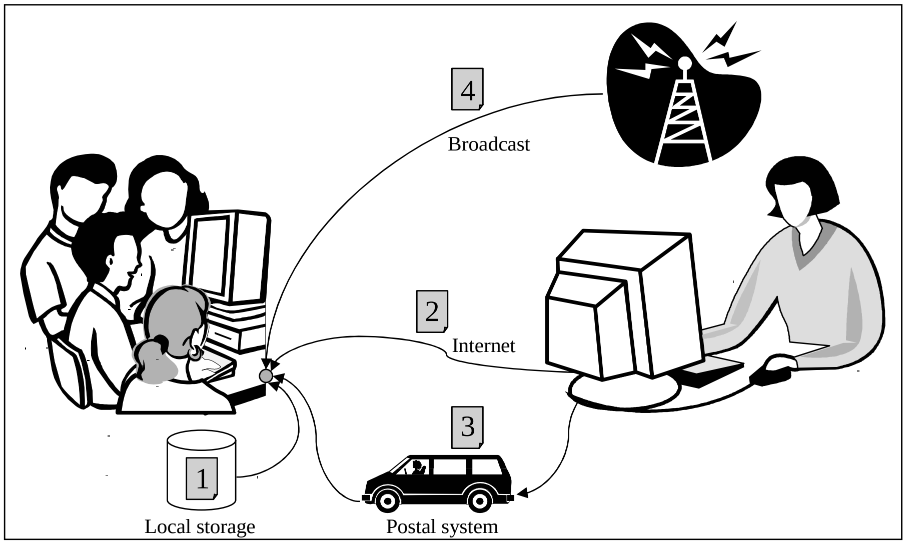

Figure 1. The integration of multiple downlink communication channels.

as a Low Latency Low Bandwidth (LLLB) channel. Our aim is to turn these alternative channels, such as the postal system, into a generic digital communication mechanism that can support a wide array of educational, commerce, informational, and entertainment applications. One challenge that we face is to integrate these alternative HLHB channels and the existing LLLB Internet into a seamless whole.

3.1.1 "Downlinks"

"Downlinks" refer to the delivery of teaching material from a remote teacher to children. Let us consider the example illustrated in Figure 1. Suppose a class that has been offered in a previous semester is being taught again. Because it has been taught before, much of the course material is already stored on a local disk at the school (data block 1 in Figure 1).

The teacher may determine that a two-minute segment of an existing lesson needs to be replaced. This two-minute segment might be small enough that we may choose to deliver it over the Internet (illustrated by data block 2 in the figure). The teacher may also need to deliver a new two-hour segment that contains either new content or homework feedback. Due to the large size of the data, it is placed on a mobile storage media and shipped by the postal system (data block 3 in the figure) or TV broadcasting (data block 4 in the figure).

The simultaneous use of these multiple channels may seem straightforward enough to be managed manually. The reality can be much more complicated. To facilitate actions such as identifying and replacing a two-minute segment, a lesson needs to be encoded into a large number of fine-grained data elements that can be reassembled, rearranged, or replaced. Manually tracking these data items is not an easy task. When a number of new data snippets are created, the author has to assign IDs to these snippets, choose one of the channels on which to send them, prepare instructions for data receivers, manually copy data onto mobile storage media for shipment, worry about whether and when the content has reached the destination and finally respond to data loss. A data receiver needs to decipher the instructions, arrange for data acknowledgements, manually copy data out of the arriving mobile storage media, worry about out-of-order delivery, answer status queries and finally schedule the delivery of a lesson. It is clear that the need for all these manual interventions might severely limit the benefit that we can derive from these multiple communication channels.

3.1.2 Automating Communication Across Asynchronous Communication Channels

One of our goals is to automate away as much manual effort as possible. Using this system, after authoring a new segment of an existing lesson, a remote teacher only needs to provide some hints such as when it is expected to be used. The system decides which of the LLLB and HLHB communication channels to use depending on how

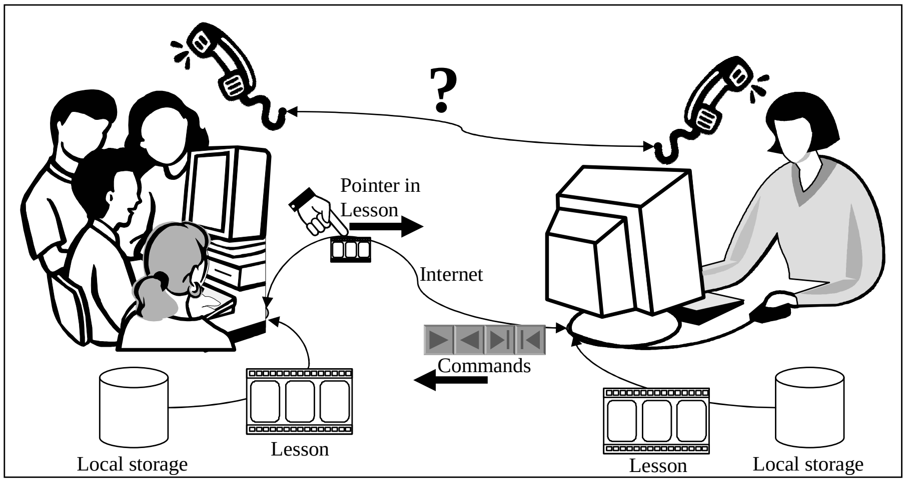

Figure 2. Synchronous interaction.

much data is involved and how much time is available. The system may also use multiple channels in parallel or ship multiple copies of the data (spaced out in time) via the postal system to provide added reliability.

The system also takes care of initiating communications. If the postal system is to be used, at the end of a teacher's working day, for example, the system would automatically prepare a mobile disk that contains the new content to be picked up by a mailman. The copying of data to disks can happen in the background, to reduce the user-perceivable delay at the end of a day. Due to the relatively large capacity of storage devices, the system can be more aggressive in copying data, even if it has only a very small chance of being used later.

Because the postal system typically provides a datagram-level service, attempts to turn it into a generic digital communication mechanism that can support higher-level semantics must deal with problems such as loss or damage of mobile storage devices, out-of-order delivery, duplicates, and the cost of copying data among different devices. An important strategy is to exploit the simultaneous use of an LLLB Internet (when one is available) so, for example, control messages such as data receipt acknowledgements can be transmitted on the LLLB Internet while bulk data traverses the HLHB postal system.

3.1.3 "Uplinks" and Interactivity

So far, we have focused on the downlink direction. We now consider the uplink direction, and how teachers may provide feedback to pupils using asynchronous communication channels.

Some of the interactions may have relatively low bandwidth demands and may occur in real time via telephone or the Internet. For example, as illustrated in Figure 2, a pupil can ask a specific question about a subject matter in the middle of the content stream that is being played off a local disk. The teacher needs to "see" that part of the content. This type of interaction in a traditional distance learning system is accomplished through two-way video conferencing. In our system, since both the teacher and the students have access to the same locally stored data, the system only needs to send "pointers" into the content stream along with questions in real time. The teacher can send either voice feedback or commands over the Internet to control the "playback" of the lesson at the students' site or even a pre-canned response to an anticipated question.

It is not always possible to provide real-time interaction for every child with limited communication bandwidth. Our system also exploits indirect interaction by allowing children to benefit from observing the interaction between the teacher and other students.

Let us consider the example Gantt chart in Figure 3. In this figure, there are two groups of students: group *A* and group *B*. These two groups may or may not be co-located. The two groups perform the same exercises at time steps 1 and 3, respectively. The exercises are graded at time steps 2 and 6. And the resulting feedbacks are played in front of the students at time steps 4, 5, and 8. The time scale illustrated in the figure can range from a few minutes to several days, depending on the amount of data that is involved and the communication channel that is used.

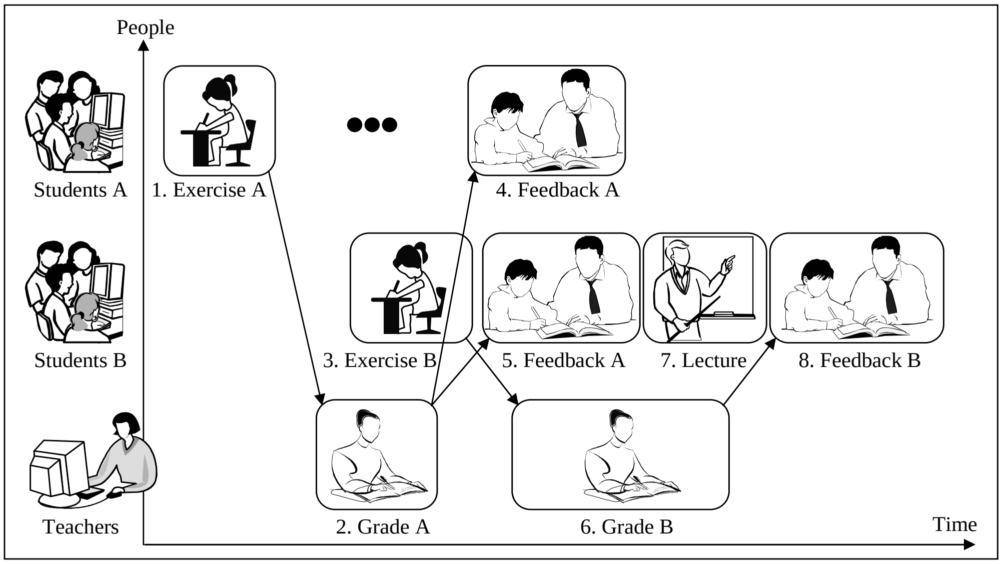

Figure 3. An example Gantt chart of providing interactivity using asynchronous communication channels. The arrows denote communication. The numbers denote the order of events.

A child in group *B* should see an uninterrupted stream of carefully prepared material. The stream includes instantaneous feedback (although the feedback consists of the teacher's interactions with other kids in group *A*) and specific feedback tailored for him (although the feedback is delayed). Consequently, the stream of events observed by a kid in group *B* should not be very different from what he would see if he were in a conventional classroom.

3.2 Supporting Peer-to-Peer Interactions

The distance learning application we envision requires the involvement of a wide range of geographically distributed participants that include students, teachers, graders and teaching assistants. We consider some ways in which the system could support such peer-to-peer interactions.

3.2.1 Centralized Routing

Figure 4(a) illustrates an approach that "routes" all interactions through a centralized server so, for example, a student would submit homework to this central server, which would then delegate the grading job to a grader. Note that the communication could utilize any of the aforementioned communication channels (including the Internet or the postal system).

This approach has several disadvantages, including unnecessary routing delays and extra costs incurred by the postal system as data is always forced to traverse through the server, server bottlenecks, and the costs of setting up and running the server. However, this scheme has an important advantage: for communication effected by mobile storage devices delivered by the postal system, the central server performs "gather-and-scatter" operations—it copies data from incoming storage devices to outgoing devices, so that each site only receives one storage device and sends one for each mailman visit.

3.2.2 Direct Peer-to-Peer Routing

Figure 4(b) illustrates an "opposite" approach. The role of the central server is limited to the coordination of routing decisions: it does not participate in data forwarding. In this figure, the student site *A* consults the central server to determine which teachers should grade *A*'s submission. Then site *A* prepares two "packages," one for teacher X and one for teacher Y.

Data routing is potentially much more efficient than that in Figure (a). On the other hand, when mobile storage devices are used for communication, a site could receive or send a large number of storage devices per mailman

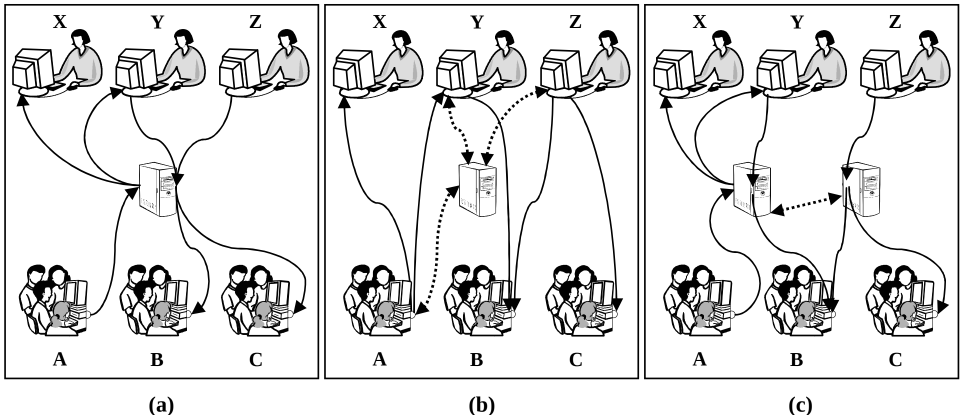

Figure 4. Alternative routing architectures. A solid arrow denotes data communication that is carried either by the Internet or other asynchronous communication channels, such as a mobile storage device carried by the postal system. A dotted arrow denotes routing information (which is most likely carried by the Internet). (a) Centralized routing via a single data distribution center. (b) Peer-to-peer routing. (c) Routing via multiple data distribution centers.

visit, which could become an administrative and cost burden. For a modest-sized system, however, this approach can be an attractive approach as it demands the least from a shared infrastructure.

3.2.3 Multiple Distribution Centers

Figure 4(c) illustrates a compromise alternative. We employ multiple geographically distributed centers. As is the case in Figure (a), when mobile storage devices are involved, each site only needs to send one storage device toward the closest data distribution center per mailman visit. Data is recopied among devices at the distribution centers. Each site may receive multiple devices per mailman visit, as many as the number of distribution centers. (Or alternatively, a site may send multiple outgoing packages but receive only one incoming package per mailman visit. Or alternatively, a site may employ a mixture of these approaches and send and receive multiple packages per mailman visit. In all cases though, the number of packages involved per mailman visit is limited by the number of distribution centers.)

As is the case in Figure (b), the geographically distributed distribution centers allow for some degree of geographical awareness in routing decisions. The distribution centers communicate among themselves to coordinate routing decisions. The latency achieved under this alternative, however, is likely to be worse than what is possible under the alternative illustrated in Figure (b) due to the extra hops through the distribution centers. It is possible to allow the coexistence of the alternatives illustrated in Figures (b) and (c), so occasional latency-sensitive packages can be routed directly to their destination without passing through data distribution centers, while less urgent packages are routed through the data distribution centers to minimize the number of packages.

3.2.4 Indirect Peer-to-Peer Routing

Under this approach, a small number of movable storage devices arriving at a site may contain data intended for other sites, and this site would need to copy such data onto a small number of outgoing storage media, which would be transported by the postal system to their next-hop destinations.

Unlike the approaches illustrated in Figures 4(a) and (c), by distributing the data forwarding task among the end-participants, this approach does not rely on an infrastructure of data distribution centers. Unlike the approach illustrated in Figures 4(b), the number of mobile storage devices sent or received by each site per mailman visit under this approach can be much smaller. For example, if we organize all the participants into a ring topology, so that each participant always receives mobile storage media from an upstream neighbor and sends to a downstream neighbor, one can eventually receive from and send to any participant. The simple ring topology, however, has poor latency characteristics, and topologies with much better latencies exist. Simultaneously limiting the number

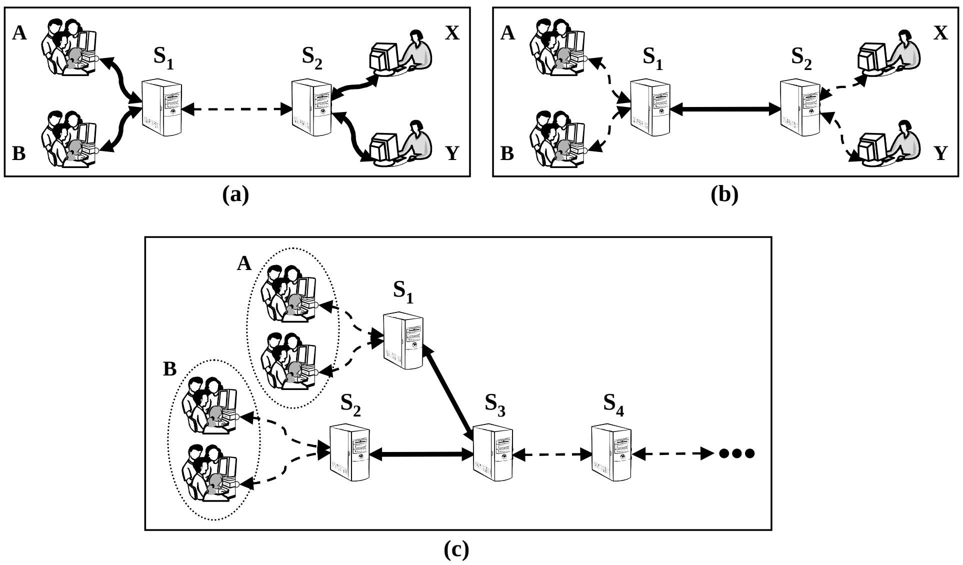

Figure 5. Examples of sequentially composing a communication path with hops of different technologies. The solid arrows denote communication effected by mobile storage devices carried by the postal system; the dashed arrows denote communication carried by the Internet.

of storage devices handled per mailman visit and the end-to-end postal system forwarding latency, and doing so in a dynamic manner that can adapt to daily traffic patterns, are research topics that we are actively pursuing.

3.2.5 Concatenating Internet and Postal Forwarding Hops

Figure 5 shows some scenarios of composing end-to-end communication paths by sequentially concatenating Internet and postal forwarding hops. For example, in Figure 5(b), S_1 and S_2 are village-level servers. Due to its remoteness, S_1 communicates with the rest of the world via mobile storage devices transported by the postal system. An inexpensive local area network (that is possibly wireless) may cover several local sites within the village. This arrangement allows a certain amount of communication and storage infrastructure (represented by S_1), including the postal system-based communication effort (for handling mobile storage devices by mailmen), to be shared across multiple local sites (represented by A and B).

3.2.6 Application-Specific Routing Decisions for Asynchronous Communication

In traditional routing, one may be interested in optimizing for metrics such as latency. On the other hand, in the context of routing homework submissions to teachers, for example, a more useful optimization metric might be minimizing the time elapsed between the moment when a homework assignment is submitted and the moment when the graded homework is returned to the students. In the examples shown in Figure 4, where multiple equally qualified teachers can all be candidates for receiving homework submissions (potentially via the postal system), the routing mechanism needs to account for factors such as predicted amounts of time needed by the teachers to grade homeworks, their current load (or “queue length”), and the postal system delay incurred to reach these teachers.

Missteps in routing decisions for these long-latency communication networks can be especially detrimental. For example, a teacher chosen by the system to grade a homework assignment may fall ill after receiving her assignment in the postal system. By the time the system becomes aware of this problem, it may take days for the postal system to deliver another copy to an alternate teacher. One way of addressing this issue is to liberally replicate data so that the homework submission is placed on the mobile storage devices sent to many teachers. As long as there is time available to make such replicas, they can be had virtually for free due to the abundance of the storage capacity. The first teacher who chooses to grade the homework would send notification messages via a low-latency network (i.e., the Internet) to all others who have received (or will receive) replicas, so that no one else duplicates the effort. (Naturally, all the notification and cancellation mechanisms should be automated.)

If there is not sufficient time for making enough replicas to address the imperfections of routing decisions, retransmissions to alternate destinations may need to take place. In the routing architectures shown in Figures 4(a) and (c), data can be "buffered" at the data distribution centers for retransmissions, so that the initial senders do not have to be involved. One of our goals is to devise a routing mechanism that can account for not only generic routing metrics (such as the number of mobile storage devices handled per site per mailman visit), but also application-specific routing metrics (such as the ones unique to the distance learning application that we have described in this section).

4 Related Work

China has launched several ambitious distance learning efforts recently (Ding, 2003; Li, 2003; Wei, 2003). Our effort complements these endeavors well. While these current efforts focus on investing in the necessary enabling *hardware* technologies, our effort primarily focuses on the *software* aspect of the infrastructure, which, as experts have observed (Li, 2003), is just as crucial (if not more).

While the idea of sending data by using the postal system to deliver mobile storage devices is not new, and companies such as AOL.com and Netflix.com have taken advantage of this approach to ship software and movies for some time, none of these existing efforts has turned the postal system into a *generic* and *interactive* data communication channel.

Recent efforts on "Delay-Tolerant Networks" (DTNs) (Fall, 2003; Hasson et al., 2003; Shah et al., 2003) have started to examine the use of WiFi-enabled mobile elements (such as buses equipped with storage devices) to simulate "delayed" connectivity to places that have access to none today. While "postal classes of service" have been mentioned (Fall, 2003), to the best of our knowledge, the postal system has so far only been mentioned as an *analogy*—no existing system that we are aware of literally and explicitly proposes to exploit the postal system as a way of extending and complementing traditional networking connectivity. Compared to other types of DTNs, the postal network has its unique characteristics that result in unique research issues.

5 Conclusion

In this paper, we have presented our approach to turn the existing postal system into a generic digital communication medium. Digital communication is achieved by shipping data on mobile storage devices via the postal system. Such communication offers two main advantages over the traditional Internet: its global reach without additional infrastructure investment and extremely high data transfer bandwidth. We have explored the ways in which the postal system and the traditional Internet can complement each other to enable sophisticated applications such as distance learning. We have also discussed ways in which richer and more efficient interactions among children, teachers and volunteers can be achieved by building a distance learning application-specific peer-to-peer interaction system.

For a full version of this paper, please see our website at (Wang, 2003).

References

CIA (2003). The world factbook. http://www.cia.gov/cia/publications/factbook/.

Ding, X. (2003). Rural area distance learning: contributions today and benefits for all future generations. http://-www.chinaonlineedu.com/media/200314/zl3.asp.

Fall, K. (2003). A delay tolerant networking architecture for challenged internets. In *Proc. of ACM SIGCOMM 2003*.

Hasson, A. A., Fletcher, R., and Pentland, A. (2003). DakNet: A Road To Universal Broadband Connectivity. http://-courses.media.mit.edu/2003fall/de/DakNet-Case.pdf.

Li, T. (2003). Bringing about rapid improvement to elementary and middle school education in midwest rural areas in China. http://www.edu.cn/20030424/3083447.shtml.

Shah, R., Roy, S., Jain, S., and Brunette, W. (2003). Data mules: Modeling a three-tier architecture for sparse sensor networks. In *IEEE SNPA Workshop*.

Wang, R. (2003). The postman always rings twice. http://www.cs.princeton.edu/˜rywang/distance.

Wei, Y. (2003). Improving education in western provinces using information technology. http://online-edu.org/article/article/-3154.html.