Psychic Routing: Upper Bounds on Routing in Private DTNs

Jonathan Anderson and Frank Stajano
University of Cambridge
firstname.lastname@cl.cam.ac.uk

July 6, 2011

Abstract

We present early work investigating a one-way delay-tolerant communications channel which affords its users perfect unobservability at the price of a limited bitrate. We suggest an unrealizable protocol, *Psychic Routing*, against which we can compare the performance of concrete Delay-Tolerant Networking routing schemes. We then use Psychic Routing to evaluate the performance of routing in our perfectly unobservable channel.

1 Introduction

Today's social networking sites provide users with fast, convenient access to shared user experiences, but they also share private information more widely that users may intend [3, 4]. We have previously described the Footlights social networking system, albiet under a different name [1], which seeks to provide users with a service whose availability and performance are even higher than today's services, but which does not blatantly violate user privacy.

However, whatever the good intentions of the authors, such a system provides an opportunity for service providers to link users to encrypted data and, from there, to each other. In order to subvert such detection, going beyond privacy and into the realm of providing anonymity properties, Footlights will also provide users with a low-bitrate communications channel that is *perfectly unobservable*: even a global adversary is unable to determine whether or not the channel is being used, despite detecting all communications in the network.

This system relies on users constructing a Delay Tolerant Network (DTN) [7] and forwarding a certain amount of traffic on others' behalf. Routing in such a network is very difficult, so we have attempted to define the criteria of success via the concept of *Psychic Routing*, an unrealizable protocol which provides us with something that we believe is currently missing from the literature: an unattainable upper limit, *à la* Shannon limit, with which we can compare new protocols.

In this early work, we explore the concept of Psychic Routing and its relevance to routing in private DTNs.

2 Footlights and Perfect Unobservability

In order to put users in control of their private information, we have proposed an architecture for a privacy-enabling social networking system [1]. This system, called Footlights[1], provides confidentiality and integrity properties through cryptography, while relying on centralised infrastructure such as Content Delivery Networks for availability.

In this system, private data is stored as fixed-length encrypted blocks in a centralised, highly-available store. Blocks are content-addressed—therefore immutable—as in the Venti archival store [16], and are organized in a directed acyclic graph, as in the Git version control system [15]. Explicit linkages between blocks are revealed only by plaintext, so the presence of ubiquitous encryption will prevent services from intentionally and explicitly revealing user data to data miners [4] or the world at large [3]. We have mitigated the most obvious privacy attacks that affect real-world systems today but, in anonymity terms, our use of a centralised block store has introduced a global adversary [5] into the system.

We pessimistically assume that our block-oriented communications substrate—which in practice will be based on a CDN and backing store from a service provider like Amazon—is able to uniquely identify every user of the system by the IP address that they connect from. Thus, the operator of the block store can identify who has uploaded any particular encrypted block as well as who is downloading it. Given this assumption, we can safely model our system as a set of one-way messages: if Alice stores a 4 kB block that is later read by Bob, that is effectively the same as Alice sending Bob a 4 kB message through a medium that is being observed by a global adversary; this duality is illustrated in Figure 1. If Bob then stores a block of his own on the server, which is later read by Carol, the adversary can observe that Alice has talked to Bob and Bob has talked to Carol, but the contents of those messages are only known to the communicants.

It is worth noting that, in such a system, the graph of users (nodes) and messages (edges) may be disjointed: Alice, Bob and Carol may form a clique that does not communicate with the rest of the user population. The system can observe such communication patterns, but the presence or absence of cliques does not change any of the analysis that follows.

For the sake of plausible deniability[2], we use blocks which can take on a

[1] So named because, like a strip of theatrical footlights, it helps users to define the interfaces between themselves and their diverse audiences, *à la* Goffman [10]. The name "Footlights" is also traditionally associated with a Cambridge comedy troupe, although our project has no affiliation with said ensemble.

[2] It is important for users to be able to hide very small amounts of data in blocks that allow the same block to be interpreted differently by different recipients. If all blocks are of a pre-specified size, there is a place for random padding, which may be purely random, or may in fact be ciphertext saying, "there's more to read over here." Further details are available in

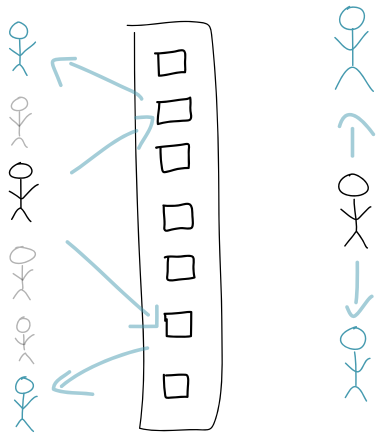


Figure 1: Footlights communication model.

limited number of sizes, perhaps even one fixed size of e.g. 4 kB. Because of this, we expect to find "spare" bits in these communications which could be used to carry covert traffic. This spare capacity can be regarded as a Delay Tolerant Network (DTN) [7], which could be used to route data around the network via intermediaries such that the global adversary cannot observe the hidden communications. In the above example, Alice might use the fact that she is sending Bob a message to also say, "Bob, the next time you talk to Carol, please tell her something for me." This message routing is *perfectly unobservable*—from the adversary's perspective, Alice has sent Bob one message, and Bob has sent Carol another message, but the number, size and timing of messages is unchanged whether the cover traffic is included or not. Only the content of the messages will change, so assuming that we use a good cipher, *zero bits of information* are conveyed by the presence or absence of covert traffic.

Having developed a DTN substrate upon which covert communications can occur, we must consider another problem: how to route Alice's message to Carol.

3 Psychic Routing

Delay Tolerant Networking is an umbrella term which describes a heterogeneous collection of network types, ranging from interplanetary networks to opportunistic Bluetooth contacts [7], and many routing schemes have been proposed for these various types of DTN [19]. When communications opportunities are fully deterministic, as in the case of the interplanetary network, models can be built and deterministic routes selected [8, 11]. In stochastic networks, two protocols are commonly used as benchmarks: Epidemic Routing, a form of controlled network flooding [18] and PRoPHET, a probabilistic scheme which keeps track of the likelihood that a node will have contact with other nodes [14].

Anderson et al. [1].

The trouble is, when developing a routing protocol for our one-way opportunistic network, it is unclear whether "delivers 15% more data than PRoPHET" is really very good; perhaps the existing protocols perform miserably on our network because their assumptions (e.g. near-instantaneous two-way sharing of routing probabilities) are not met, and thus we have no cause to celebrate being slightly better than them. What we need is something like a Shannon limit [17], a theoretical maximum that can never be reached, but which progressively better routing protocols can approach more and more closely. Such a limit would provide a real sense of both how far we've come *and* how much further we might yet go.

We propose that such a limit can be expressed via *Psychic Routing*: the most efficient routing of data that could be done today if one had full knowledge of future events and a protocol which imposes no communication overhead. Clearly, such a scheme is impossible to implement in practice, but a protocol with good probabilistic estimation of future events, based on statistics of past events, could begin to approach the upper bound imposed by Psychic Routing.

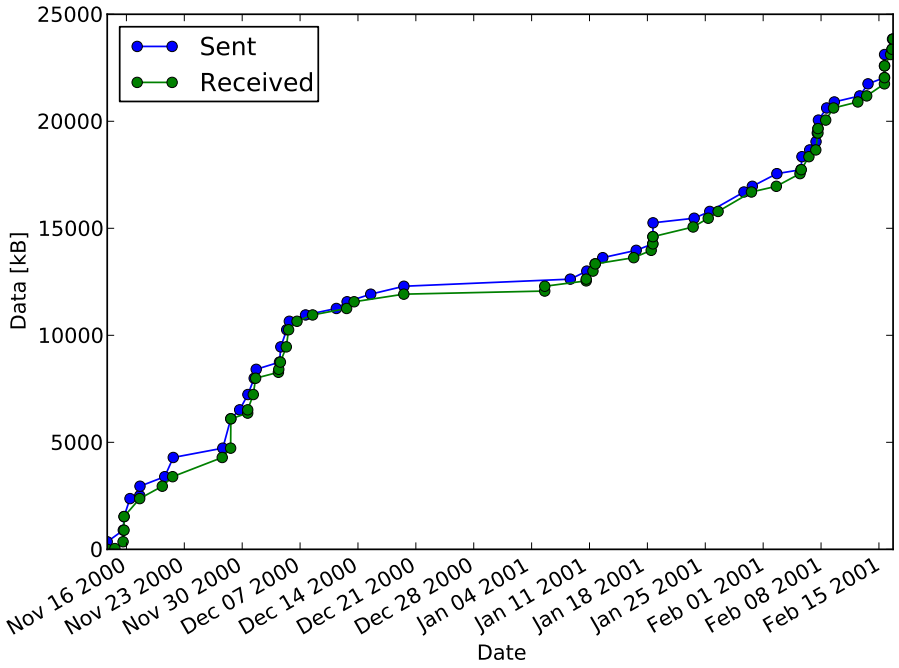


Figure 2: An example of maximal data transfer between two nodes in a DTN.

Psychic Routing does not provide general, closed-form equations like the Shannon limit in an AWGN channel; rather, it provides an upper bound on the pairwise performance of a routing protocol in a particular context, with particular parameters. For instance, Figure 2 shows the psychic limit for hypothetical communication using the Footlights system from Section 2 as a communications

substrate, between two people in the Enron e-mail corpus [12] who have been selected at random. This is an absolute maximum: no routing scheme will be able to transfer more data in the same time, or the same data in less time.

Psychic Routing only considers the maximum flow of data from one source node to one sink, it only provides pairwise maxima; it is not a global network-wide optimisation. The model can assume that e.g. some of the channel's capacity has been consumed by other traffic, but global optimisation for network-wide properties such as "fairness" is firmly in the realm of future work.

3.1 Calculation

Figure 2 is generated by walking "backwards" from the destination node (here denoted Bob), applying a mark to all e-mails that Bob received in the Enron corpus. We then walk backwards from the senders of *these* messages, marking all messages that could have "influenced" a message sent to Bob. We continue walking backwards, recursively, building a set of possible routes to Bob, stopping when we discover cycles in the graph. We then discard all routes that do not contain Alice, and throw away any portion of a route the precedes Alice.

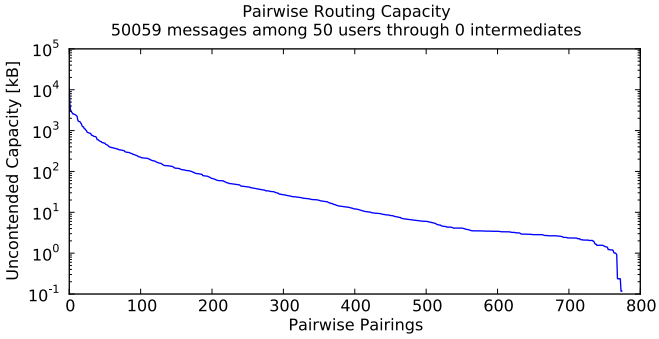
We then calculate a capacity of "spare" bytes—the number of bytes required to pad the message to a fixed block size, in this case, 4 kiB—for each message in a potential route. Each of these point-to-point messages, and associated capacity, can be seen as an edge in a *flow network*; Psychic Routing corresponds to pushing the *maximum flow* [9] from source Alice to sink Bob.

More precisely, for each instant x on the X axis, there will be a specific flow network, formed by the messages that emanated from Alice and its descendants up to that time; and the y value for that x will be the solution of the maximum flow problem for that flow network. The meaning of that (x, y) data point is that, if every participant had cooperated and used the spare capacities of the available messages in the most favourable way towards that goal, the maximum amount of data that could have been transferred from Alice to Bob by time x (without altering the observable pattern of messages that were to be sent, regardless of this covert communication) would be y.

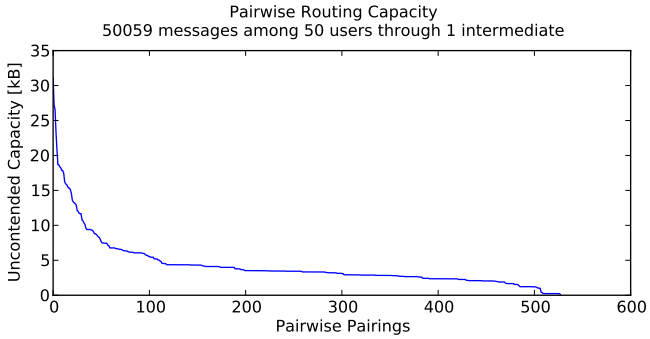
3.2 Related Work

The use of unattainable maxima as comparison points is well-established in the communications and computer science literature. The oft-cited Shannon limit—the most popular of which, applied to an Additive White Gaussian Noise (AWGN) channel, is sometimes treated as synonymous with "Shannon limit"—specifies the most error-free information transmission that is possible over a noisy channel [17]. Real communications systems cannot reach the Shannon limit, but it provides a useful absolute comparison point: we can say that an error correction code comes within 0.3 dB of the Shannon limit, rather than "10% better than code X."

Similarly, Belady described an unimplementable page replacement algorithm that real paging algorithms can be compared against [2]. Belady's MIN algo-



(a) Direct messaging.



(b) Routing through exactly one intermediary (data laundering).

Figure 3: Routing capacity over two years (no contention).

rithm, much like Psychic Routing, relies on full advance knowledge of what Virtual Memory (VM) pages will be required by the system in order to make optimal decisions about which pages to swap out of memory. A realizable algorithm such as Least-Recently-Used (LRU) that approaches the performance of Balaly's MIN algorithm in a variety of VM workloads can be deemed appropriate for concrete systems.

4 Measurement

In order to test the applicability of Footlights' covert DTN to real-world traffic, we have driven a Footlights model with data from the Enron e-mail corpus [12]. We first took a subset of the most "interesting" e-mails in the corpus[3]: 50,059 messages among the 50 users that communicated the most in the period 1999-

[3]This reduction was performed in the interest of computation time: it would simply take too long to process all 300k e-mails in the corpus, many of which are single messages to or from e-mail addresses outside of Enron (e.g. an invitation to one Enron employee to attend a University graduating class reunion).

2002. Each e-mail in the corpus was then translated into a Footlights message with a spare capacity of $4096 - l \bmod 4096$, where l is the length of the e-mail, including SMTP headers.

Figure 3 shows the routing capacity of the network; it is able able to support some limited communication. In these graphs, we see how much data Alice can send to Bob—for all possible values of Alice and Bob—either by sending it directly to Bob (Figure 3a) or via some intermediary (Figure 3b). The x axis represents ranked pairings of Footlights users: pairing 0 is the "couple" with the best routes from one to the other, pairing 1 the next best, etc., and the y axis is how much data can be transmitted covertly between 1999 and 2002. Clearly, the ability to communicate kilobytes of data over a period of years does not make for a general-purpose communication system, but as a means of sharing keys, perhaps in order to establish other channels, it could be quite useful.

Implicit in Figure 3 is the assumption that there is no contention for any of the network resources that Alice wishes to use. This is clearly an inaccurate assumption, but nonetheless we can see that *the network is capable of bearing some traffic*—the question is how much.

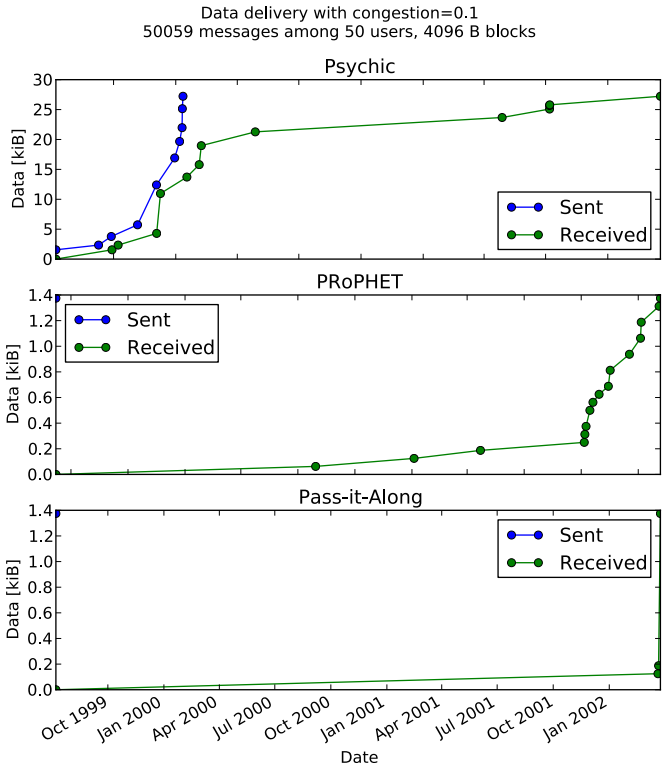


Figure 4: Routing messages via Footlights in the Enron e-mail corpus.

This question is answered in Figure 4, which shows a comparison of simulated routing performance by several routing algorithms—Psychic Routing, PRoPHET and the "Pass-it-Along" strawman—routing data through one intermediary, for pairing 0 in Figure 3b.

4.1 Psychic Routing

Psychic Routing, at the top of Figure 4, is clearly the best scheme. Using knowledge of future communications among all participants in the network, Alice is able to send approximately 27 kiB of data to Bob over the course of approximately 30 months. Such a scheme is, of course, unrealizable, but it sets our upper bound, our analog to the Shannon limit.

4.2 PRoPHET

The next graph represents the performance of the popular PRoPHET algorithm, with some tunable parameters selected from an IETF draft [13] and some chosen arbitrarily. We can see that almost 1.4 kiB of data does propagate from Alice to Bob via at least one intermediary, but unlike the Psychic Routing case, most of the transfer occurs at the end of the period of interest, rather than the beginning. Unlike Psychic Routing, a realizable scheme like PRoPHET must expend time and communications bits in order to propagate routing information—in this case, probabilistic estimates of how soon each node expects to communicate with other nodes.

In the absence of other information, we might suppose that PRoPHET's performance here is reasonable, and invest significant effort in tuning parameters to improve it incrementally. Compared with Psychic Routing, however, we can see how much further we still have to go; our time and effort might be better used in searching for a different routing algorithm entirely which better reflects the realities of the medium (e.g. does not assume that nodes are able to exchange routing information in a two-way exchange).

From a privacy perspective, we might also wish to invest in a scheme which does not require reporting accurate contact information to all peers.

4.3 Pass-it-Along

The final routing scheme depicted in Figure 4 is a straw man called "Pass-it-Along routing." As the name implies, when a node engaged in pass-it-along routing receives a packets of data, it stores the data and sends it out again attached to the next message to go out. No consideration is taken as to the suitability of the next node to finally deliver the packet, so this strawman protocol can be regarded as a constrained form of network flooding.

The performance of this routing scheme is, as might be expected, rather poor. It is only by chance that a fortuitous message succeeds in delivering over 1 kiB at the last possible moment; were it not for this one message, less than 200 B would have been delivered. Nonetheless, the comparison between Pass-it-Along

Routing and PRoPHET is telling: if we were to only compare the two schemes, we would say that Pass-it-Along is slower in delivering data, but under some circumstances, it actually delivers *more* data, since PRoPHET must consume some of the available channel in order to propagate routing information. It is only by comparing to Psychic Routing, our objective upper bound, that we see how truly atrocious the performance of Pass-it-Along Routing really is.

An interesting property of Pass-it-Along routing is that, like other flooding protocols, no source or destination addresses need be visible to routing nodes: they simply act as data mules, assuming that the intented recipient is able to recognize the packets intended for her (e.g. by decrypting them with a pre-shared key). Thus, its performance is quite poor, but it has useful privacy properties and, in this particular case, its performance is not severely worse than a scheme which requires all nodes to broadcast who they talk to and how often.

5 Future Work

There is an obvious trade-off between routing efficiency and the privacy of routing nodes: in order to improve efficiency beyond network flooding, nodes need to know about each others' communication patterns. In source-routed systems such as Tor [6], the sender of a packet does not reveal to whom she is speaking, but such a system can only function because routers have been published in a directory, and their communication graph (fully connected, over the Internet) is implicit. IP requires that destination addresses be visible to all routers, which themselves broadcast messages saying, "if you want to send packets to any of the following networks, give them to me; I am connected to them." In delay-tolerant networks with *late binding*, destination addresses may not even be fixed: routers may say, "I understand the mapping from a high-level name to a low-level one, and will forward traffic accordingly."

Having introduced Psychic Routing, an upper limit on the effectiveness of DTN routing, we now wish to study this information-efficiency trade-off, and establish a *lower* bound on how efficiently traffic can be routed in an unobservable DTN, given a certain amount of information about the communication graph the various nodes are willing to reveal to each other.

Furthermore, Psychic Routing—as it currently stands—only considers pairwise optimisation of network flows. Future work might consider global optimisations, and the incentives that would encourage all participants to "play by the rules." Even more useful might be a consideration of how local incentives—such as sender-pays vs. a *quid pro quo* arrangment of packet handling—lead to global network properties such as congestion and packet loss.

6 Conclusion

We have introduced a new feature for the Footlights social networking system: while providing high-speed, high-availability access to shared user data, Footlights will also provide users with access to a low-bitrate communications channel that is *perfectly unobservable*—even a global adversary will be unable to distinguish between users conducting normal conversation and users "piggybacking" covert traffic on their normal messages. Such a capability allows users to communicate indirectly, via other users, in a Delay Tolerant Network (DTN).

In order to write high-performance routing protocols for this DTN, we have introduced the concept of *Psychic Routing*, a routing scheme which relies on full knowledge of future communications in order to make routing decisions that maximize pairwise throughput. We have shown how this scheme could serve the role of a Shannon limit for DTN routing, rendering obsolete existing comparisons such as "15% better than an existing scheme whose properties are themselves understood relative to other schemes."

We have compared the performance of the popular PRoPHET protocol with both a straw man protocol and Psychic Routing, and have qualitatively observed an inverse relationship between pairwise routing performance and the amount of information about the network used by the protocol. In the future, we hope to explore this relationship further in order to establish a useful lower bound on privacy-preserving DTNs, so that users will be able to select routing schemes that maximize performance while respecting user-specified limits on the disclosure of contact information.

References

[1] Anderson, J., Diaz, C., Bonneau, J., and Stajano, F. Privacy-Enabling Social Networking Over Untrusted Networks. In *the Second ACM SIGCOMM Workshop on Social Network Systems (WOSN '09)* (May 2009), pp. 1–6.

[2] Belady, L. A. A study of replacement algorithms for a virtual-storage computer. *IBM Systems Journal 5*, 2 (1966), 78–101.

[3] Bonneau, J., Anderson, J., Anderson, R., and Stajano, F. Eight Friends Are Enough: Social Graph Approximation via Public Listings. In *the Second ACM EuroSys Workshop on Social Network Systems (SNS '09)* (2009).

[4] Bonneau, J., Anderson, J., and Danezis, G. Prying Data Out of a Social Network. In *the 2009 International Conference on Advances in Social Network Analysis and Mining* (2009).

[5] Diaz, C., Seys, S., Claessens, J., and Preneel, B. Towards measuring anonymity. In *Privacy Enhancing Technologies* (2002), pp. 184–188.

[6] Dingledine, R., and Mathewson, N. Tor: The second-generation onion router. In *the 13th USENIX Security Symposium* (2004).

[7] Fall, K., and Farrell, S. DTN: an architectural retrospective. *IEEE Journal on Selected Areas in Communications 26*, 5 (2008), 828–836.

[8] Ferreira, A. Building a reference combinatorial model for MANETs. *IEEE Network 18*, 5 (Sep 2004), 24–29.

[9] Ford, L. R., and Fulkerson, D. R. Flows in Networks. Tech. Rep. R-375-PR, RAND Corporation, Aug 1962.

[10] Goffman, E. *The Presentation of Self in Everyday Life*. Anchor Books, Jan 1959.

[11] Handorean, R., and Gill, C. Accommodating Transient Connectivity in Ad Hoc and Mobile Settings. In *Pervasive Computing (PERVASIVE)* (2004), pp. 305–322.

[12] Klimt, B., and Yang, Y. Introducing the Enron Corpus. In *the First Conference on Email and Anti-Spam (CEAS)* (Carnegie Mellon University, 2004), Carnegie Mellon University. `http://www.cs.cmu.edu/~enron/`.

[13] Lindgren, A., Doria, A., Davies, E., and Grasic, S. Probabilistic Routing Protocol for Intermittently Connected Networks. Internet draft (work in progress). Retrieved from `http://tools.ietf.org/html/draft-irtf-dtnrg-prophet-09`.

[14] Lindgren, A., Doria, A., and Schelén, O. Probabilistic Routing in Intermittently Connected Networks. In *Service Assurance with Partial and Intermittent Resources (SAPIR)* (Fortaleza, Brazil, 2004), Springer Berlin Heidelberg, pp. 239–254.

[15] Loeliger, J. *Version control with Git*. O'Reilly, Jun 2009.

[16] Quinlan, S., and Dorward, S. Venti: a new approach to archival storage. In *the FAST 2002 Conference on File and Storage Technologies* (Monterey, California, Jan 2002), Bell Labs, Lucent Technologies.

[17] Shannon, C. A Mathematical Theory of Communication. *Bell System Technical Journal* (1948).

[18] Vahdat, A., and Becker, D. Epidemic Routing for Partially-Connected Ad Hoc Networks. Tech. Rep. CS-2000-06, Duke University, Apr 2000.

[19] Zhang, Z. Routing in intermittently connected mobile ad hoc networks and delay tolerant networks: overview and challenges. *Communications Surveys & Tutorials, IEEE*.