ON-DEMAND TRANSIENT DATA STORAGE AND BACKUP IN MOBILE SYSTEMS

Jeffrey Hemmes, Christian Poellabauer, and Douglas Thain
Department of Computer Science and Engineering
University of Notre Dame
Notre Dame, Indiana 46634
Email: {jhemmes, cpoellab, dthain}@cse.nd.edu

***Abstract*—In a crisis response scenario, the availability of information to first responders can be greatly enhanced through the use of mobile computing systems. However, such systems are limited by available storage space and battery life. Substantial degradation of their utility over the course of a prolonged operation is likely as batteries wear down and storage space fills up. While frequent offloading or backup of data can ensure persistence, an indiscriminate approach can accelerate the consumption of resources and further shorten the system's availability.**

This paper presents a method for transferring data on demand between cooperating mobile systems based on system state and both current and projected geographical location, taking into account the observed mobility of neighboring nodes. Peer selection is accomplished by first rating potential candidate machines to receive the data transfer, then further evaluating each with a utility function that conservatively estimates a predicted window of opportunity over which the transfer can occur.

I. Introduction

Mobile computing devices have become increasingly ubiquitous over the last several years. Such devices have been employed in a wide array of applications, exemplified by such disparate information systems as cell phones, UHF radios, and PDAs. While using such devices for recreational purposes is commonplace, critical applications such as coordination of search and rescue and emergency response operations are also becoming widespread. Using such a system for military or paramilitary command and control purposes places a high premium on data availability, security, and internode communication. Simply put, without efficient data sharing in a crisis environment, mission accomplishment is impossible.

While mobile information systems enable such data sharing, mobile network topologies are inherently unstable, posing a significant challenge to providing reliable data access. Network partitions and churn are a certainty in any mobile network of nontrivial size. Furthermore, over time a particular node may become unavailable due to diminishing system resources. Because of the significant likelihood of device and network failures in a deployment scenario and the potential criticality of data to operational success, it is necessary to have a reliable means of data backup and retrieval. Additionally, the criticality of much of the data in a crisis response environment is based on location. For instance, the threat posed by specific hazards may be localized and of concern to operators in the immediate vicinity, but may not be significant enough to warrant a broadcast to all users of the system, particularly in light of limited computing resources.

It is well established that more data is not necessarily better. In typical mobile networks, devices cannot simply broadcast all data recorded by all participants to all other devices within range. An indiscriminate data transfer model such as this has several serious drawbacks. First, such an approach can consume tremendous amounts of bandwidth and storage space. Second, it places an undue burden on other devices and operators to sort through extraneous data, posing a performance penalty on computing devices or hindering the effectiveness of human operators. Third, heavy resource utilization has substantial system availability costs due to shortened battery life on each mobile device. Conversely, data which may not be needed right away may in fact be needed later as the situation on the ground changes, necessitating its preservation. The purpose of this work is to develop a cooperative data sharing protocol that uses context for data transmission decisions and thus limits the total amount of broadcasting required.

The contribution of this paper is a new method for context-aware cooperative data transfer and recovery in which context is described by available system capacity and geographical positioning and relative motion of peer nodes. This paper is organized as follows: First, we present our method for context-aware peer selection using location-based routing in mobile ad-hoc networks, to include the initial scoring process and the final selection through mobility prediction. Next we discuss issues related to data recovery. We evaluate the method through simulation trials and present our results. Finally, we discuss related works and some concluding remarks.

II. Context Aware Data Replication

Mobile networks are frequently assumed to be homogeneous, consisting of a collection of identical hardware platforms with relatively similar capabilities. In practice, despite the similar configurations, it is common for devices to be deployed with varying states such as battery level or available storage space. We assume any underlying system would consist of dissimilar hardware, such that each node may be either a laptop computer, PDA, or even a stationary desktop workstation, each of which with varying capabilities in terms of computing power, storage space, and wireless radio range. Furthermore, as with any ad-hoc network, a deployed system can be expected to experience a substantial amount of churn as operators enter and leave network partitions, either by traveling beyond maximum range of individual wireless radios, entering shielded buildings, experiencing device failure, or as other such events occur. Device failure can occur either through physical destruction of the device in certain circumstances or through gradual loss of battery power. The potential loss of power is significant as continuous operations in austere locations may preclude swapping batteries when required. Despite differences in capability, what is important is that these devices are connected to a wireless ad-hoc routing network and are able to receive positioning data, perhaps through a portable Global Positioning System (GPS) receiver.

We further assume the radio transmission range for a mobile device to be a predetermined system parameter. This is not a new assumption, and was argued previously in [2]. Despite variations in RF signal propagation, a conservative estimate is all that is required for our approach, so estimating the approximate wireless range is not unreasonable.

The transient data may be arbitrary depending on the application and scenario, but generally can be assumed to be logs of events recorded internally by the device or externally by either peripheral sensor devices or the operator. As data is collected, eventually that which is deemed critical may need to be replicated on a neighboring node to ensure persistence.

A. Peer Selection

Selection of peer nodes depends on several factors. The first step in performing a data transfer is determining the amount of data that must be moved. Once the data size is known, the next step is finding a location with sufficient available space. Because criticality of data is application or situation dependent, the specific data set is best determined with policy and the volume of data is best measured with an internal resource monitor that periodically records the total size of the files expressed in the policy specification. With this information and knowledge of the system state of neighboring nodes, it is trivial to determine whether its peers have sufficient available storage which may be used as temporary scratch space. Gathering this information requires advertisements of available storage space from each peer node, which may be included in a routing protocol or transmitted separately.

In addition to determining the size of data which must be preserved, a resource monitor task periodically polls the battery and available storage space on its own device and records it. Such state information is transmitted to its peers as data fields in the packets in the routing protocol, and is used in conjunction with other state information such as physical location determined via a connected GPS receiver.

As routing packets are received from remote devices, each node constructs a routing table that contains the battery state, available free storage space, and GPS coordinates of all connected peers. At the time a data replication operation is required or requested, the current view of the network using the routing table is used to select a peer within range of a single hop. Selecting a closely connected peer is desirable for two reasons. First, it is extremely challenging to determine with any reasonable degree of certainty the properties of the wireless connection between two arbitrary remote nodes. More importantly, however, in many scenarios moving the data is time-critical and storing it in a geographically proximate location may be more beneficial to recovery. In other words, placing data at an arbitrary connected node that may only be reachable across several hops may not be advantageous despite the potential availability of greater system capability at that node. This does not even consider the higher probability of failures and retransmissions introduced when moving data over additional, intermediary nodes described in [6]

Our approach consists of a two-part method for peer selection that can augment existing routing protocols in ad-hoc networks. Whichever routing protocol is employed would need to include, or be modified to include, system state and geospatial location information about each node, but such information could easily be inserted

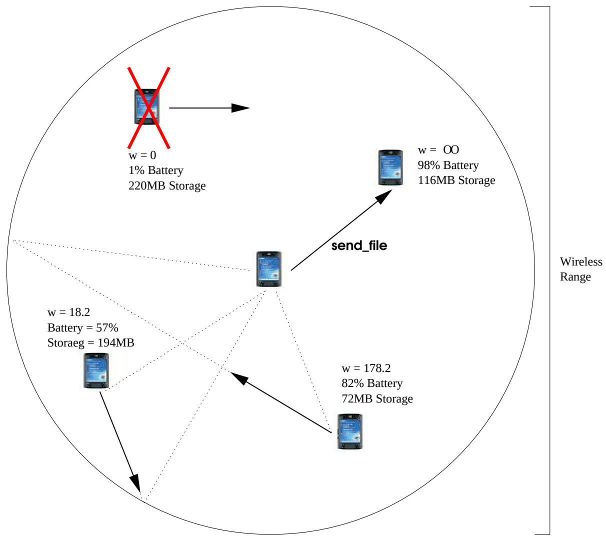

Fig. 1. Candidate Peer Node Selection

This figure illustrates the selection of a peer node for data transfer. Nodes with insufficient remaining battery life or available storage space to complete the transfer are disregarded immediately. The remaining nodes are scored based on the expected length of time the node will remain within wireless range in addition to overall system state. The selected node may then receive the data transfer.

into many types of routing packets, as long as the mobile devices themselves have some notion of location built in, e.g., are affixed to a GPS receiver.

1) Initial Scoring: In order to find a suitable peer node, it makes sense to quickly eliminate from further consideration those devices which are obviously not appropriate candidates based on available system capability. For instance, if a peer has insufficient battery power to receive, for example, a 1 MB file transfer, then it should not be considered after the initial examination of its system state. The first step in peer selection, then, is to make a rough initial evaluation of all nodes in the routing table, intended more to eliminate clearly unsuitable nodes than to find an optimal node, and assign a score to each that reflects whether its capability warrants further consideration. Any node with very low battery power or available storage space less than the amount required to complete the transfer is assigned an initial score of zero. Nodes that are more than a single hop away are also scored zero. All remaining nodes are assigned an initial score of one. These nodes are the candidates for possible data transfer.

2) Availability Estimation: When a data transfer is requested or required, the candidate nodes, indicated in the routing table with an initial score of one, are rescored based on an estimated window of opportunity. The window of opportunity is determined by extrapolating the peer's location history over time to some approximate point at which it will move out of wireless range, rendering any further direct data transfer impossible. The window of opportunity is illustrated in Figure 1. Once the window of opportunity is computed for each possible candidate peer, the node requesting the transfer does a greedy selection and chooses the peer with the largest window of opportunity.

This method ensures that nodes near the fringes of the wireless range are less likely to be selected than those nearby if they are moving away from the requesting node, but more likely if they are moving towards it, and that nodes which are stationary relative to the requesting node are the most likely to be selected as their windows of opportunity are considered infinite. Figure 1 provides an illustration of the peer selection method. In this scenario, a node with insufficient battery power is eliminated from consideration in the first step, as indicated with the large X. Other candidate nodes are then considered based on the predicted window of opportunity, illustrated with arrows and dotted lines, which indicates the estimated length of time of availability. In the figure, two candidate nodes are in motion, so they have a limited window, and a third is relatively stationary. The stationary node is the one selected.

We use a very simple distance equation to determine the window of opportunity between the peer's current location and the intersection point of the wireless radio range. Because wireless range is generally not isotropic, any empirically or analytically determined range can be used; as long as the boundaries are roughly known, the intersection points can be computed. For each candidate node c_i,

$$score_{c_i} = \frac{\sqrt{(x_{int} - x_i)^2 + (y_{int} - y_i)^2}}{t_i - t_{i-1}}$$

where (x_{int}, y_{int}) is the intersection point between the candidate node's extrapolated current path and the estimated limit of the wireless radio range.

B. Data Management and Recovery

Nodes that store data on behalf of another use whatever scratch space is available on its local storage device to do so. Eventually, however, that stored data might need to be either retrieved or purged. The node that originated the data must either fetch the data or inform the peer that any of the replicated files it stores no longer need to be maintained. This might result from replacing a battery or the operator's return from a potentially hazardous location. In a wired network with fixed infrastructure, control of replicated copies is not difficult. On the other hand, in a wireless network, particularly an ad-hoc network, data recovery is a nontrivial problem. If a node transfers data just before leaving the network, the only metadata that might be known to the transferring node is the initial location of the transfer, but even that is hardly guaranteed. The node performing the temporary storage may have further transferred its data to a subsequent node. Since most peer-to-peer networks assume either a fairly static locality of data or multiple replicated copies of data, traditional approaches to data recovery are unsuitable for this system.

In this approach, a node attempting to rejoin the network will check for the presence of the peer to which its data was transferred. If it is present, then either retrieval of or purging the copy is trivial. Otherwise, the node must transmit a request to the other nodes to locate the data. While peers that are reachable only through multiple hops are not initially considered for transferring data, it is possible that over time, multiple transfers by several peers may position the data more than a single hop away from the originating node. Currently we assume all transfers are point-to-point, but because of the likelihood of such a scenario and the difficulty inherent to forcing data to reside within a specific geographical boundary, such a multi-hop file transfer mechanism would be required, at least for recovery. We do not discuss recovering migrating data in this paper.

III. Evaluation

In this section, we demonstrate that our context-aware selection method outperforms selection by either random choice or geographic proximity. A selection is considered superior if it generally remains within wireless range for a longer period of time using various patterns of motion. We evaluated the effectiveness of our approach through a series of discrete event simulations using both linear and random movements among remote peers. The simulations were conducted in two parts, discussed here separately. The first part is evaluating the quality of the initial choice, based on the criteria needed to conduct data transfer. The second part evaluates the quality of the selection over a longer period of time to determine whether a context-aware approach yields a selection with a higher availability, as determined by system state and connectivity.

Unless otherwise specified, in these simulations, 50 nodes were placed at random within the node of interest's wireless range of 300 meters, with uniform distribution, and assigned random speeds and directions. Because having nodes remain stationary relative to the node of interest does not yield interesting evaluation results, as the availability of any such node would not be restricted by movement, we limit our evaluation focus to nodes actually in motion. For our simulations, we selected a minimum speed of 1 meter per second and a maximum speed of 4 meters per second.

A. Initial Selection

Table I compares the ability of our approach to initially choose an optimal peer node moving linearly with constant speed to that of the random and geographic selection methods. For initial selections, a successful selection is one in which the chosen peer has ample battery life and storage space to complete a 50 MB data transfer, and lies within a 300-meter wireless radio range. Because all nodes are initially within range, a selection in this case is considered a failure only when the selected node has insufficient system state. Optimal selections are those with the largest composite evaluation score among all nodes, and for the random and nearest selection methods, an optimal selection is one that was also chosen by the context-aware approach. Context-aware selection will always choose a node with sufficient system state for the size of the data; hence the 100 percent success rate shown in the table. These results are shown for simulations of 1,000 separate runs.

Method	Success Rate (%)	Failures	Optimal Selections
Random	74.27	186.50	10.43
Nearest	84.47	112.50	5.79
Context Aware	100.00	0.00	725.00

TABLE I
INITIAL SELECTION (LINEAR MOTION)

This table shows the initial success rates for each selection method among 725 neighboring nodes with varying system capabilities. Selections are considered failures if the selected node has insuffi cient battery and storage space to complete a 50 MB fi le transfer. Failure and optimal selection numbers are averages over all trials.

Table II further compares our approach to random and geographic selection when peer nodes move randomly in both speed and direction. Other simulation inputs are the same as in the linear model.

For both patterns of motion, even using a random approach can produce reasonable results, with an initial

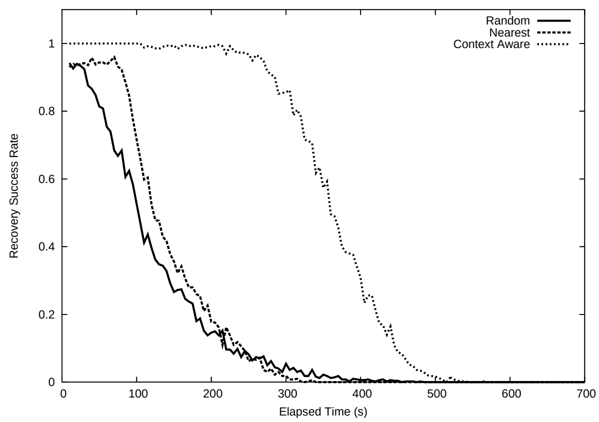

Fig. 2. Recovery With Linear Motion

This figure illustrates the ability of the node of interest to locate the node previously selected using a random algorithm, a strictly nearest neighbor approach, and our context-aware method. Our approach improves the recoverability of data by choosing nodes with higher availability than either a random or geographic approach, even as the time between selection and data recovery increases.

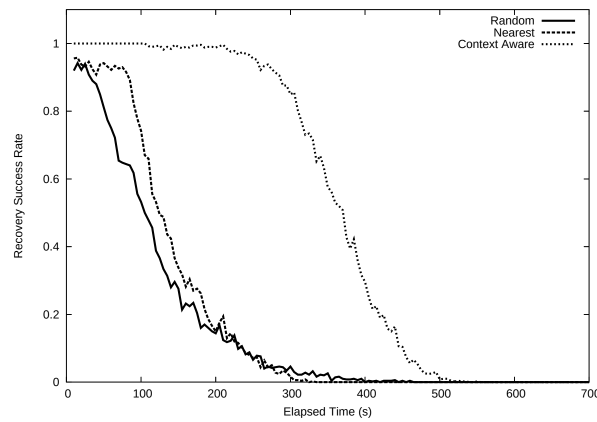

Fig. 3. Recovery With Random Motion

This figure shows the likelihood of maintaining contact over time with nodes moving randomly. Results similar to that of the linear model can be achieved with a random pattern of motion. Taking a snapshot of the network topology and evaluating peers based on estimated availability increases the success rate for recovery over approaches using random selection or geographic proximity.

Method	Success Rate (%)	Failures	Optimal Selections
Random	73.11	194.79	9.43
Nearest	84.30	113.79	5.64
Context Aware	100.00	0.00	725.00

TABLE II
INITIAL SELECTION (RANDOM MOTION)

This table shows initial success rates for each selection method among nodes moving with random velocity. By ignoring nodes with critically low system state, a context-aware method has a higher initial success rate than either random or geographic selection.

success rate of about 74 percent in simulations where the initial system states among nodes are randomly generated with a discrete uniform distribution. A random selection method generally chooses a higher number of unsuitable nodes compared to selecting the most geographically proximate node, but this is more likely due to the randomly generated system state than. By eliminating the possibility of selecting unsuitable nodes, our approach doesn't experience the failures experienced in the other selection methods. It is always possible in a given scenario for no suitable peers to exist, but no selection algorithm would work in such a case, and as in the case of stationary nodes, is a much less interesting simulation scenario.

Due to the greedy nature of the selection algorithm, it is possible that multiple peers may select a single stationary node for data backup simultaneously. To avoid a single node becoming a bottleneck for multiple transfers, a policy mechanism can be used to select other nodes in such an event. Similarly, an access control policy could be used to prevent unauthorized placement or retrieval, but such policy and access control schemes have not yet been explored within the context of this work.

B. Recovery

After the initial evaluation and selection, the selected nodes were reevaluated after increasing periods of time to measure the quality of the selection. While it is certainly true that nodes out of wireless range of a specific device may still be reachable over multiple hops, the likelihood of failure increases significantly with the number of hops [6], so ideally we would like to select nodes that can later be reached directly. This section evaluates the ability of our approach to do so.

To evaluate the longer term suitability of the selections, we simulated both random and linear motion patterns among 50 nodes over a period of 700 seconds, and at 5-second increments, reevaluated the availability of the selection. Batteries discharge at a constant rate with a 3-hour lifetime. For each selection time, we ran 1,000 independent simulation runs and recorded the mean success rate for each method.

Figure 2 shows the effectiveness of our approach as compared to a random or strictly geographic approach

		Maximum Wireless Range (m)						
Number of Nodes	Selection Method	10	50	100	500	1,000	5,000	10,000
10	Random	**3.2**[1]	19.7	33.8	147.7	275.9	1,190.3	2,108.2
	Nearest	**3.9**[1]	31.2	50.5	217.7	426.2	1,805.7	2,998.0
	Context-Aware	**3.9**[1]	47.6	91.5	439.1	852.2	3,356.1	4,768.6
50	Random	8.6	19.2	32.1	148.8	297.2	1,219.7	1,926.8
	Nearest	10.9	31.0	53.1	221.4	434.8	1,897.5	3,127.4
	Context-Aware	11.0	67.7	133.6	633.8	1,227.7	4,377.9	5,523.7
100	Random	9.8	19.8	35.3	146.9	278.4	1,275.0	2,043.0
	Nearest	13.3	30.3	52.8	225.1	429.0	1,872.3	3,143.4
	Context-Aware	13.5	75.4	149.6	699.0	1,338.6	4,715.0	5,131.8
500	Random	10.1	19.5	34.5	152.7	303.9	1,198.5	2,020.5
	Nearest	15.8	31.2	21.7	219.2	440.6	1,920.9	3,304.8
	Context-Aware	16.4	88.3	172.4	819.8	1,545.8	5,034.0	5,041.0
1,000	Random	10.1	19.9	33.6	153.1	285.5	1,234.4	2,082.1
	Nearest	16.6	31.2	53.2	225.3	435.8	1,846.8	3,195.9
	Context-Aware	17.3	92.4	179.7	846.6	1,618.0	5,049.9	4,993.4
5,000	Random	10.0	19.6	34.5	152.8	276.1	1,231.0	2,061.1
	Nearest	17.5	31.3	54.4	227.4	433.3	1,973.3	3,179.1
	Context-Aware	18.7	99.1	192.7	903.0	1,706.2	5,038.4	5,188.5
10,000	Random	10.1	19.8	33.1	156.1	291.2	1,233.2	**1,903.0**[2]
	Nearest	17.8	31.0	53.1	228.1	432.0	1,978.8	**3,319.0**[2]
	Context-Aware	19.2	101.1	195.2	919.4	1,735.4	2,853.4	**5,356.5**[2]

TABLE III
AVERAGE TIME BEFORE FAILURE

This table shows the average availability time in seconds for each method's selection over 1,000 independent simulation runs for each node/range pair. In all cases, the context-aware method produces higher availability times, and in many cases the difference is as much as six times that of the random method.

when peer nodes move in a linear fashion, starting at random locations. Once the initial selection is made, the simulator attempts to contact the selected node again after a designated time interval. Any selected node with critically low battery power or storage space, or has moved out of wireless range, is considered a bad choice. By selecting nodes with an acceptable system state that also maximize the conservatively estimated availability window, the probability of easily retrieving data is higher than that obtained using the other approaches. Figure 3 shows similar effectiveness of the context-aware method using remote peers that move randomly.

Figure 4 shows the effect of increasing data requirements on the success rate. In our simulations, nodes are assumed to have available storage space ranging from a minimum of 50 MB to a maximum of 150MB. As the size of the data requirement increases, the success rate for all three approaches fall to zero, with our approach having a much higher success rate overall.

Finally, we evaluate the effect of the number of node and the size of the wireless radio range on the average time before each approach fails. Table III shows the amount of elapsed time, i.e., the availability period, averaged over 1,000 independent runs for each node/range pair, before each method's selection fails due to insufficient system state or movement out of range. In each case, the context-aware method produces a selection with a higher average availability time than the other methods, in many cases by as much as six times that of the random selection.

For very small scale networks, both in terms of wireless capability and number of nodes, the context-aware method doesn't provide enough of a benefit to be worthwhile, since availability is extremely limited in any case. The results indicated with a [1] are for such small-scale networks. However, as the size and capability grow much larger, as shown with a [2], the context-aware approach gives dramatically better results than either nearest-neighbor or random. This suggests that as wireless capability increases and ranges grow substantially larger, our approach may have substantial benefit for improving availability rates.

IV. RELATED WORK

While there is an abundance of research in the area of mobile networks, work specifically in the area of mobility prediction is much more limited. While a number of works attempt to predict future availability by tracking position history, we are not aware of any routing protocol

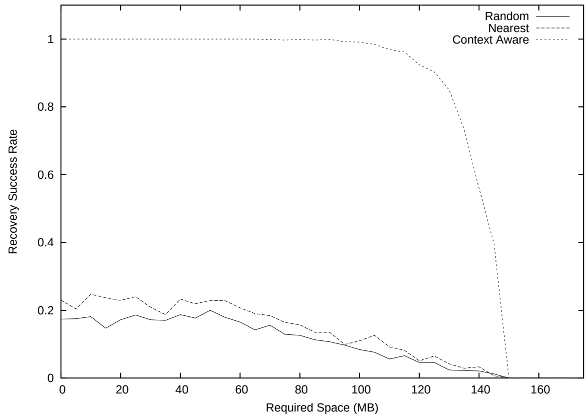

Fig. 4. Effect of Data Requirement on Success Rate

The effect of the data requirement size on success rates. Assuming nodes have no more than 150 MB of available storage, the context-aware selection method has much higher average success rates than both the random and geographic approaches, up to a data size of roughly 100 MB. At that time, success rates for all three approaches rapidly approach zero.

that incorporates system state into routing determinations, or combines system state with mobility.

In [4], mobility is predicted by building and maintaining a table to track estimated periods of connectivity loss, but the focus of that work is limited to disconnections due solely to mobility. While our work is similar to theirs, to include the ability to use the prediction method independent of the routing protocol, their work assumes that availability of nodes does not change due to diminishing system state.

Our method is most closely related to that proposed by Pascoe, et al [5]. Whereas in their work, the prediction is intended to be used for estimating the amount of overhead incurred in both unicast and multicast routing protocols as routes are broken due to mobility, ours is designed to select a single hop route with the largest predicted availability for purposes of storing and retrieving transient data. Like other, similar approaches, their work does not assume link breakage due to system failures.

Also closely related is the work proposed by Su [7], which uses GPS location information to predict the future location of nodes moving independently. Additionally, parameters such as radio propagation range are known a priori. Like other routing protocols, this approach does not account for availability constraints other than mobility, and doesn't address the specific problem of data recovery at all.

Other methods of predicting availability do so by measuring signal strength, and a diminishing signal portends a link disconnection. Given our experiences with the directional nature of many wireless antennas, using detected signal strength may not be the most appropriate factor for predicting mobility in many types of applications. Examples that employ a signal strength measurement for availability estimation are [1] and [3].

V. Conclusion

In this paper we propose a method for selecting peers to offload transient data that accounts for both heterogeneous and dynamic system state among nodes as well as predicted mobility. Previous work has shown that mobility prediction can improve overall availability and link longevity, but our work goes a step further and includes dynamic state information in the peer selection process. This approach has the advantage of choosing nodes most likely to remain available for recovery in applications for which transient data has high value.

Simulation results show that our approach improves the success rates of both initial selection and data recovery. For small scale networks, to include those with very limited wireless range, our approach may be of limited benefit. But as the network scales, particularly with more powerful wireless radios, using a fully context-aware selection method can significantly increase the availability of temporarily replicated data, which in turn can lead to greater mission success for operations that depend on the availability of sensor information.

References

[1] R. Chellappa-Doss, A. Jennings, and N. Shenoy. User Mobility Prediction in Hybrid and Ad Hoc Wireless Networks. In *Australian Telecommunications Networks and Applications Conference*, December 2003.

[2] J. Deng, Y. Han, P.-N. Chen, and P. K. Varshney. Optimum Transmission Range for Wireless Ad Hoc Networks. In *IEEE Wireless Communications and Networking Conference*, March 2004.

[3] T. Goff, N. B. Abu-Ghazaleh, D. S. Phatak, and R. Kahvecioglu. Preemptive Routing in Ad Hoc Networks. In *7th ACM/IEEE Conference on Mobile Computing and Networking*, July 2001.

[4] H. Larkin. *Predicting Connectivity in Wireless Ad Hoc Networks*. PhD thesis, Bond University, 2005.

[5] M. Pascoe, J. Gomez, V. Rangel, and M. Lopez-Guerrero. Modeling Duration of Routes in Multihop Ad hoc Networks. In *20th IEEE Canadian Conference on Electrical and Computer Engineering*, April 2007.

[6] C. Rohner, E. Nordström, P. Ginningberg, and C. Tschudin. Interactions Between TCP, UDP, and Routing Protocols in Wireless Multi-hop Ad Hoc Networks. In *4th ACM/IEEE Conference on Mobile Computing and Networking*, October 1998.

[7] W. Su. Mobility Prediction in Wireless Networks. In *2000 Military Communications Conference*, October 2000.