McCormick

Northwestern Engineering

CCITT · Center for the Commercialization of Innovative Transportation Technologies

Integration of Real-Time Mapping Technology in Disaster Relief Distribution

February 12, 2013

Irina Dolinskaya

Karen Smilowitz

Jennifer Chan

CCITT | 600 Foster Street | Evanston, IL 60208
Tel: 847-491-7287 | Fax: 847-491-3090
http://www.ccitt.northwestern.edu/

NORTHWESTERN UNIVERSITY

Disclaimer

The contents of this report reflect the views of the authors, who are responsible for the facts and the accuracy of the information presented herein. This document is disseminated under the sponsorship of the Department of Transportation University Transportation Centers Program, in the interest of information exchange. The U.S. Government assumes no liability for the contents or use thereof.

Acknowledgments

This work was funded by the Center for the Commercialization of Innovative Transportation Technology at Northwestern University, a University Transportation Center Program of the Research and Innovative Technology Administration of USDOT through support from the Safe, Accountable, Flexible, Efficient Transportation Equity Act (SAFETEA-LU).

Integration of Real-Time Mapping Technology in Disaster Relief Distribution

CCITT Final Report

Irina Dolinskaya, Karen Smilowitz, and Jennifer Chan

February 12, 2013

Contents

1 Project Overview

Vehicle routing for disaster relief distribution involves many challenges that distinguish this problem from those in commercial settings, given the time sensitive and resource constrained nature of relief activities. While operations research approaches can improve the effectiveness of relief routing, these challenges must be addressed in routing models in order to realize the potential of the approaches. There have been many promising advances in the literature on relief routing, and aid organizations have been collaborating with academic researchers to increase the practicality of such models. Increases in the availability and use of information technology in the wake of disasters can further the effectiveness of routing models for aid distribution. Currently, challenges still remain to make routing models more applicable to humanitarian assistance delivery and more integrated with new streams of imagery, mapping, and crowdsourced real-time data.

This project focuses on dynamic routing models for the distribution of relief supplies and services in humanitarian settings. We focus on the potential to improve these models, and thus improve the effectiveness of humanitarian relief, by using new applications of mapping technologies and real-time information to mitigate the effects of dynamic changes during humanitarian crises and disasters and the significant uncertainty that exists in these settings. Our work evaluates the improvements from these technologies for relief organizations in the field and develops a set of test cases for the research community to better design and test their routing models and solution approaches. In the present work, we take urban search and rescue operations as our study case, while maintaining a focus on generalizability to other post-disaster operations. To facilitate wide implementation and potential commercialization of our work, a developed test case is available online to practitioners and academicians, through a server dedicated to Humanitarian and Non-Profit Logistics at Northwestern University.

1.1 Needs Statement

Dynamic environments and uncertainty are the norm during disasters. Many of the datasets currently used in emergency response become critically out of date in a fast changing environment. As a result, models and routing pathways often fall short of their intended goals to provide timely transfer of goods and services to organizations and disaster-affected communities.

Whether it is routing metric tons of food during the 2012 Horn of Africa Food Crisis or routing ambulances in the Northeast in the aftermath of Superstorm Sandy, disasters require real-time knowledge of changing environments. The information flows in such settings emerge from a complex changing environment, and the humanitarian assistance world is left with persistent questions about potential new data sources. These sources come from complex, interdependent systems and questions cannot be answered in isolation from each other, nor can they be resolved with the operations and logistics models currently available. Routing and path planning models must be redesigned to accommodate real-time information, while integrating specific characteristics of the evolving technology and data in a disaster relief setting.

Recent years have seen rapid growth in the application of technology to humanitarian relief operations. Updated orthogonal imagery is made freely available by government agencies and commercial sources, sometimes within hours of a disaster, so that damage to infrastructure can be quickly assessed. Geographic information system (GIS) data layers on flood extents, earthquake damage, medical dispensaries, settlement camps, and other spatial data can be made available for download and integrated into online map visualization tools. Mobile phones play a growing role in aid coordination thanks to the high penetration of mobile technology throughout the world and the ease of using Short Message Service (SMS) technology for data transmission. Mobile units are increasingly used to send georeferenced updates on road networks, movement of displaced persons, stocks of supplies, and other time-sensitive information, and this use is expected to continue to grow.

However, the potential for technology implementation in humanitarian relief still has a long way to go before it is fully met. In 2011, the United Nations Office for Coordination of Humanitarian Affairs (UNOCHA) and the Harvard Humanitarian Initiative investigated the humanitarian relief sector's struggle to accommodate the "information fire hoses" offered by new information sharing tools and systems in the wake of the Haiti earthquake. The report recommended a framework for addressing this shortcoming, including increased attention to innovation, experimentation, academic evaluation, and coordination between the academic, technical, and practitioner communities [14].

1.2 Solution Approach

In this project, which received cost share from a grant through the Google Research Awards Program[11], we consider the incorporation of real-time data elements as "inputs" for routing models to improve humanitarian relief routing. In a recent study [5], we investigate gaps in existing routing models through a review of operations research models for the transportation of relief goods and interviews with aid organizations, ranging from government agencies to non-government organizations (NGOs) and commercial partners engaged in disaster relief. Importantly, both the nature of uncertainty (e.g., the need for relief supplies, the availability of resources to address needs, and the impact of the disaster on infrastructure) and the mechanisms available to mitigate this uncertainty are very different from what is found in commercial settings. For example, travel times may vary in a commercial setting due to congestion; however, one can model this uncertainty reasonably well with a bounded distribution of travel time. In a disaster setting, the very existence of a path between locations may not be known, or the lack of security of the path may restrict its use by response agencies.

Our group began by researching dynamic models of humanitarian relief routing that explicitly address the unique nature of this uncertainty in disaster relief and the feasible mechanisms to respond to uncertainty. In this project, we focus on addressing uncertainty through using newly available information to reduce it, in conjunction with routing models that are designed to accommodate uncertainty. We address the question: **how can operations research models exploit advances in mapping technologies and real-time information to improve the distribution of humanitarian relief?** Specifically, we investigate how the identified sources of uncertainty in relief routing can be mitigated through available real-time information about the affected region, taking urban search and rescue (USAR) operations as a motivating case and a starting point for modeling. To accomplish this, we

1. develop a testbed of routing problems designed specifically for humanitarian relief routing, to be used by the research community to evaluate new modeling and solution approaches;
2. quantify the benefits of technology for agencies engaged in humanitarian relief through improved operations research models that incorporate this technology; and
3. launch an online server housing the developed testbeds with the capabilities for other academicians and practitioners to use as well as contribute to the database.

1.3 Research Team

Irina Dolinskaya is an Assistant Professor of Industrial Engineering and Management Sciences, and holds the Junior William A. Patterson Chair in Transportation. Her research interests include optimal path finding in a direction, location and time dependent environments, and path planning with limited information about the region. Applications include problems in humanitarian logistics, optimum vessel performance in evolving nonlinear wavefields, and autonomous navigation for amphibious vehicles. Dr. Dolinskaya is currently working on a number of projects with the Office of Naval Research studying optimum vessel performance in evolving nonlinear wavefields and autonomous navigation for amphibious vehicles.

Karen Smilowitz is an Associate Professor of Industrial Engineering and Management Sciences. Her research focuses on freight transportation systems and non-profit and humanitarian logistics. Recent projects have analyzed the opportunities and challenges of introducing operational flexibility in distribution systems. Dr. Smilowitz has worked with a range of collaborators from industry and non-profit organizations, including UPS, Coyote Logistics and the Mobile C.A.R.E. Foundation of Chicago.

Jennifer Chan is an Assistant Professor of Emergency Medicine at Northwestern Memorial Hospital and an Associate Faculty member at the Harvard Humanitarian Initiative. Her research focuses on crisis mapping and the integration of emerging technologies into humanitarian operations. Recent projects include program evaluations of open-source technology organizations and co-authoring *Disaster 2.0*, a UN Foundation report that analyzed the interface between humanitarian agencies and volunteer and technical communities. Dr. Chan has worked with UN and NGO agencies, including Oxfam America, UNOCHA, UNDP, and IRC providing technical and public health support during humanitarian conflicts and natural disasters.

This research effort has included a number of undergraduate and graduate students, including current Northwestern PhD students Luis de la Torre and Zhenyu (Edwin) Shi. The undergraduate students engaged in the pilot study described in Section 2 are Alex Huang, Alex Ma, Sara Schmidt, Nancy Xu, and Brandon Zhang.

Lewis Meineke is the Humanitarian and Non-Profit Logistics Research Coordinator at Northwestern and provided technical assistance and data research on this project. In prior work at the University of Chicago, Meineke managed spatial database creation and analysis for a National Institutes of Health-funded research center in the fields of economic history and health economics.

1.4 Organization of this Report

This report is organized as follows: Section 2 describes a pilot project continuing from the outcomes of the previous year's CCITT project and demonstrating the feasibility of combining real-time mapping technologies and crowdsourced data collection with operations research modeling to better inform vehicle routing in the immediate aftermath of a disaster. This proof-of-concept project describes a stylized model of a workflow for urban search and rescue operations, but is intended as motivation for subsequent work and does not represent a final product. Section 3 elaborates on the data processing techniques employed to generate the testbed data used in the project described in Section 2. It provides a concrete discussion of technical considerations for generating usable routing data from currently available data pipelines as well as summary data documentation for potential users of the testbed. Section 4 describes our mathematical formulation of post-disaster vehicle routing as an adaptive orienteering problem as well as solution approaches under development. Finally, Section 5 describes the next steps toward eventual realization of a new post-disaster routing technology.

2 Test Scenario: Haiti 2010

2.1 Motivation

The response to the 2010 Haiti earthquake involved an unprecedented use of digital technology to collect and disseminate information. The affected population made extensive use of cell phones, not only for personal communication but as a means of mass communication and coordination. Responders used aerial and satellite imagery, and digital maps derived from it, to gain situational awareness in a poorly mapped environment. However, without prepared protocols for integrating these novel data sources into their ground operations, relief distributors and search and rescue teams were unable to make full use of them.

The years since the earthquake have seen a deepened understanding of the roles new technologies can play in disaster relief, as well as the severe challenges involved in adequately putting these technologies to use. In this pilot project, we combine recommendations from our previous study with recent developments in the use of new technologies in disaster relief to assess the viability of real-time data for urban search and rescue (SAR) team routing, using the response to the Haitian earthquake as a test case. This study's aim is to demonstrate proof-of-concept in developing potential workflows to incorporate new data sources into vehicle routing and to scope the need for future work, rather than to quantify true potential gains from a fully deployable set of tools. We also motivate the future development of stochastic routing algorithms tailored to post-disaster settings and data conditions.

De la Torre et al. [5] review the real-world problems related to vehicle routing for delivery of goods and services in disaster-affected regions, analyze the representation of these problems in current operations research models, and identify next steps for modeling, focusing on elements of humanitarian relief distribution that make these settings differ from standard vehicle routing model settings. These elements include, among others, uncertainty in demand and supply; incomplete knowledge of the local road network and changing road conditions; and the need for routes to change to accommodate new information. A distinct need is identified for models that incorporate simultaneous uncertainties in supply, demand, network connectivity, and travel times in a dynamic setting that allows operations to be modified in response to new information.

In this report, we describe how data sources of the type that appeared in the aftermath of the Haitian earthquake could be incorporated into these future models, using a simplified preliminary routing algorithm as a starting point. Section 2.2 introduces our data sources and processing workflows. Section 2.3 details two parallel solution approaches for using these data sources to generate SAR routes, one based in a traditional static framework and one that incorporates real time information. Section 2.4 discusses some results of the exercise and charts a course for future extensions.

2.2 Innovation: Integrating new data sources into relief routing

The chief innovation introduced in this pilot project is the integration of recently introduced real-world data streams into a workflow for humanitarian routing in real time. We identify two data streams that, together, provide sufficient inputs (graph structure, demand, and travel time) to conduct routing analysis. We develop a simple algorithm to use these data sources as new data arrive and use the development of this workflow to pose questions about the adaptation of these data sources for this purpose. In future work, this data integration will allow us to benchmark the performance of newly developed routing algorithms against real-world scenarios using historical data.

2.2.1 Data sources

Although a variety of data sources, such as aerial imagery and social media services, played significant roles in the aftermath of the Haiti earthquake and hold the potential to be integrated into routing workflows, we restrict the present focus to two independent sources: Mission 4636 and OpenStreetMap.

Mission 4636 was a mass communication initiative launched by a quickly organized team of Haitians, academics, and members of the international humanitarian community shortly after the earthquake struck. The system allowed Haitians affected by the earthquake to send informational updates, including requests for help, via text message (SMS) to the shortcode 4636. This system was announced over radio, one of the chief means of mass communication in the disaster's aftermath. The messages were translated into English, categorized by message type, and, when the message contained location information, assigned geographic coordinates by Kreyòl-speaking volunteers, primarily members of the Haitian diaspora, with additional contributions from nonprofits and activist groups like Ushahidi Haiti [22]. This workflow produced information that could be shared with response teams in near-real time–the median lapsed time from receipt of a message to translation, categorization, and submission to the data consumption stream was less than five minutes, and over 80,000 messages were processed in total.

OpenStreetMap (http://www.openstreetmap.org) is a freely editable online mapping platform with a wiki-like model designed to publish map updates from its broad user community within minutes. Volunteers use data from a variety of sources, such as satellite and aerial imagery, public datasets, and personal knowledge, to add information to OpenStreetMap's (OSM) spatial database. OSM employs a flexible data structure that allows volunteers to "tag" spatial data elements (points, lines, and polygons) with essentially unlimited information in the form of key→value pairs. Because both rendered maps and the underlying spatial data are made available, this rapidly updated data source can be used directly by viewing maps or as input data for computer modeling.

In disaster settings, OSM has shown that it can generate a more complete and up-to-date map of the affected area than would otherwise be possible. Since existing road maps of the affected area in Haiti were incomplete, inaccurate, and insufficiently detailed for routing in the post-disaster phase, or difficult to access if they were available at all, for many responding agencies OSM data quickly replaced these maps, providing critical situational awareness for international volunteers and local organizations. Its data on transportation networks and points of interest expanded dramatically in the days following the disaster and became a frequently used source of geographic information for many relief organizations.

2.2.2 Data processing

In this study, data for the non-real time model constitute a base to which real time data are added; data inputs for the real time model are thus a superset of inputs for the non-real time model. The base dataset comes from OpenStreetMap data from the Port-au-Prince metropolitan area on road networks and locations of buildings with high estimated demand for search and rescue, such as hospitals and schools. This dataset was chosen both because it offered a ready point of comparison between real time and non-real time data processing from equivalent sources and because, although little mapping had been done outside the core of Port-au-Prince by the day of the earthquake, OSM still would have constituted a relatively good source of information in the Haitian setting, which faced especially poor data availability because of the destruction of baseline maps and data in the earthquake. The real-time data add two additional and independent streams: text messages sent to 4636 and post-earthquake road network updates from OpenStreetMap, both described above.

Figure 1 summarizes the procedure for processing non-real time data. We use a data layer of building footprints from OpenStreetMap as it existed shortly after the earthquake, after an initial period of spatial data entry had occurred. These building footprints are primarily sourced from digitally traced ("heads-up digitized") aerial imagery. We first filter buildings to include only sites

Figure 1: Data processing workflow for non-real time model

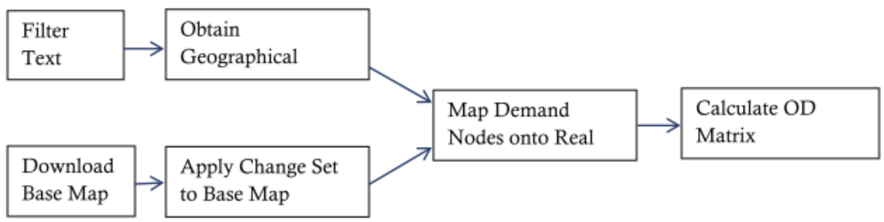

Figure 2: Additional data processing workflow for real time model

likely to have high demand and vulnerable populations, such as hospitals, schools, and orphanages, within Port-au-Prince. This filtering results in 224 demand nodes. We use footprints to calculate building square footage to generate a rough estimate of demand at a site. We further restrict this set to the 133 nodes with highest estimated demand for computational ease. Because the rest of the process requires point data rather than polygons, we convert the data to points snapped to the road network. Figures 3 and 4 show a portion of the final building dataset in map and tabular form.

2.2.3 SMS processing

Messages sent to 4636 were routed via an internet service to online volunteers and translated into English, categorized according to a quickly purpose-built categorization scheme, and manually geolocated using whatever location information was present in the message (the SMS protocol does not send geographic coordinates, so the message itself was the only source of this information available). Both the categorization and the geo-location were conducted by volunteers on each message individually and involved no automated data extraction. The categorization scheme allows rapid filtering by message type. We use only messages 1) with non-missing geographic coordinates placing the message within Port-au-Prince, 2) categorized as "People trapped," "Person trapped," or "Collapsed structure," and 3) sent during the five-day study period, 1/17/2010 through 1/21/2010. This filtering reduces the number of text messages from over 30,000 to 22; most of this vast reduction comes from the restriction to non-missing coordinates.

2.2.4 Road processing

For a more detailed description of the network data creation procedure, see Section 3.3.

To construct a retrospective set of road networks that reflects the state of information on each day of the study period, we begin by making a data extract from an archived copy of OSM as it existed immediately before the earthquake. We then apply each daily "changeset," a record of all data updates performed over the course of a period, in sequence, generating a full copy of OSM's data for Haiti for each day of the study period.

We then design network configuration rules for creating crude but serviceable network datasets that can be used for routing. The most critical component of this configuration for our purposes is the construction of a travel cost variable. The mandatory "highway" tag contains the road class of a segment (as determined by the volunteer entering the data). Travel cost on each road segment is estimated by the segment's total length multiplied by a parameter c depending on this variable: $c = 1$ for primary roads, $c = 1.2$ for secondary roads, $c = 1.4$ for tertiary, residential, and unclassified

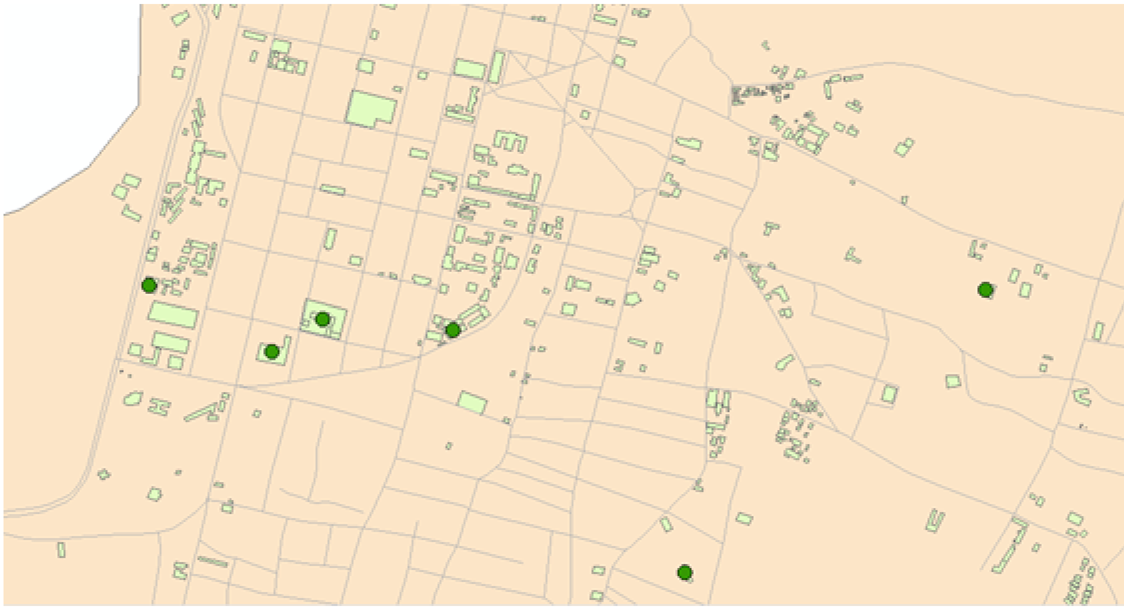

Figure 3: Locations of interest (dots) among major buildings in a portion of Port-au-Prince

Top133_nonrealtime

	OID	osm_id	name	type	X	Y	Absolute_A	Demand
	0	48215556	CENTRE HOSPITALIER D	hospital	-72.3258	18.539	1623.282749	4
	1	48276805	PORT AU PRINCE HOSPI	hospital	-72.3455	18.5373	5145.92958	12
	2	48276806	Hôpital St. François de	hospital	-72.3441	18.5382	12783.853292	31
	3	48285471	Hotel Caraibe Haiti	hotel	-72.3644	18.5308	805.262259	2
	4	48288054	Hôtel Villa Créole	hotel	-72.2921	18.5224	2695.510115	6
	5	48304467	Centre d'Apprentissage	school	-72.331	18.5533	1258.335792	3
	6	48330852	Gheskio - National HIV/	hospital	-72.3489	18.5391	1915.24413	4
	7	48345364	Hôtel Royale Haitien	hotel	-72.3728	18.5314	1225.444607	3
	8	48370607	Ecole Evelyna Levy	building_concret	-72.5382	18.2341	1354.174095	3
	9	48371950	Savanne Christ	building_concret	-72.6926	19.4502	9066.309462	22
	10	48374633	Hospital Frances	hospital	-72.3405	18.5379	1151.332058	2
	11	48417965	Nursing School (FSIL)	school	-72.6312	18.5151	1512.098967	3
	12	48417966	Nursing School (FSIL)	school	-72.6318	18.5155	1625.56616	4
	13	48423089	École Mixte de Sainte Th	building_concret	-72.3323	18.5772	930.43622	2

Figure 4: Portion of building table after calculating estimated demand

roads, and $c = 1.6$ for other minor roads. The parameters used are essentially arbitrary and serve as a stand-in for empirically determined parameters derived from the literature or from ground experience; travel costs associated with different road types are likely to vary by locality. Since these parameters simply scale travel cost based on shape length, they are functionally unitless and have only relative meaning.

Volunteers used OSM's key-value data structure to store information about road traversability (e.g., *impassable* $\rightarrow$ '*yes*'). While it would have been possible to include this information in the network dataset creation, we elected not to in the present work because comparing the results of the real time and non-real time models would have required penalizing the non-real time model for sending routes over non-traversable roads; this is a straightforward extension left for future work.

Each of the constructed network datasets corresponding to one day of the study horizon is used together with the corresponding day's list of demand nodes to construct an origin-destination (OD) matrix storing pairwise travel times. These OD matrices are generated using the OD matrix solver built into the geographic information system used in the project, ArcGIS. This solver is a proprietary implementation based on Dijkstra's algorithm for finding shortest paths and is integrated directly with the ArcGIS network data structure. Since roads were being added and corrected very rapidly over the study period, the estimated travel time between two points could change dramatically in a short time.

2.3 Solution approach

We model the scenario as an orienteering problem. The decision-making unit, a single mobile team, is given a list of demand nodes, each with an associated demand (score) and location. The team moves along a graph consisting of vertices connected by edges, each with an associated travel cost (considered here to be travel time). Demand nodes are located on these vertices. In addition to travel time, each node imparts a stopping time that is assumed to be an increasing linear function of demand at that node. The objective is to maximize the total demand served over the study period by selecting a set of nodes to visit, and the order in which they are visited, each period, subject to a maximum time expenditure per period. Furthermore, travel time, stopping time, and local demand are assumed to be normally distributed random variables; however, the team is not able to change the selected nodes or routes based on additional information obtained while traversing the route. The team services each stop locally rather than transporting people or goods to a central location.

The overall procedure is diagrammed in Figure 5. We start by building the network graph with real-world data as described above, using a standard origin-destination (OD) matrix solver based on Dijkstra's algorithm to find pairwise shortest paths, and passing the resulting matrix to a MATLAB program as a model input. The elements of this matrix are modeled as the means of normally distributed random variables with a known variance. The program employs a simple greedy algorithm that selects a demand node to visit one at a time, based on a local maximization, and then passes the node and its associated parameters (demand and service time, both random variables subject to variance) to a standard traveling salesperson problem (TSP) genetic algorithm to solve for the shortest path connecting all selected nodes; this selection process repeats until the sum of estimated travel times and service times exceeds the time allotment for the period, and the algorithm returns an ordered list of nodes to visit. This list is passed back into ArcGIS and visualized as a route with real-world paths (Figure 6).

When the real time version of the algorithm runs, this process is modified by updating the model inputs at the beginning of each period: the set of candidate nodes is augmented based on information received over the course of the previous period, and the OD matrix is replaced with a new one generated with an updated graph and set of nodes.

We assume that local demand D_i, travel time c_{ij}, and service time S_i are normally distributed random variables with standard deviations estimated as a fraction of the sample means. We assume this fraction to be 0.1 for real time data and 0.2 for non-real time data, reflecting our assumption that real time data provide more reliable information on current conditions than data collected before the onset of the disaster. We assess model performance by estimating true values for D_i, c_{ij}, and S_i from their observed values, repeating this procedure over 50 iterations.

2.3.1 Optimization model

Notation	
0	Depot designation
i, j	Demand node indices
t	Period index
D_i	Demand at node i
c_{ij}	Travel time from node i to node j
S_i	Stopping time at node i; linear function of D_i
T_{max}	Total time available per period
N	Set of candidate (unvisited) demand nodes
N_t	Set of new candidate nodes in period t, from 4636 data
V_t	Set of nodes visited in period t
D_{V_t}	Total demand at nodes visited in period t
$W_i(N) = \frac{D_i}{\sum_{j \in N} c_{ij}}$,	Demand divided by total distance to all other nodes
$W_i(V_t) = \frac{D_i}{\sum_{j \in V_t} c_{ij}}$,	Demand divided by total distance to nodes visited in t
$OptRoute(V_t)$	Optimal ordering of nodes in V_t; output from TSP
$T(V_t)$	Total time spent on route in period t; output from TSP

We begin by localizing the problem and creating a scaled down model with minimal data inputs to be solved in AMPL, a linear programming solver. The simplified model solves a single-vehicle problem maximizing demand served within a deadline.

Variables:

X_{ij} — 1 if team travels from node i to j, 0 otherwise

Y_i — 1 if team visits node i, 0 otherwise

Objective function:

Maximize:

$$\sum_i Y_i D_i \tag{1}$$

Subject to:

$$\sum_j \sum_i X_{ij} c_{ij} + Y_i S_i \leq T_{max} \tag{2}$$

$$\sum_j X_{ij} = \sum_j X_{ji} \tag{3}$$

$$\sum_j X_{ij} = Y_i \tag{4}$$

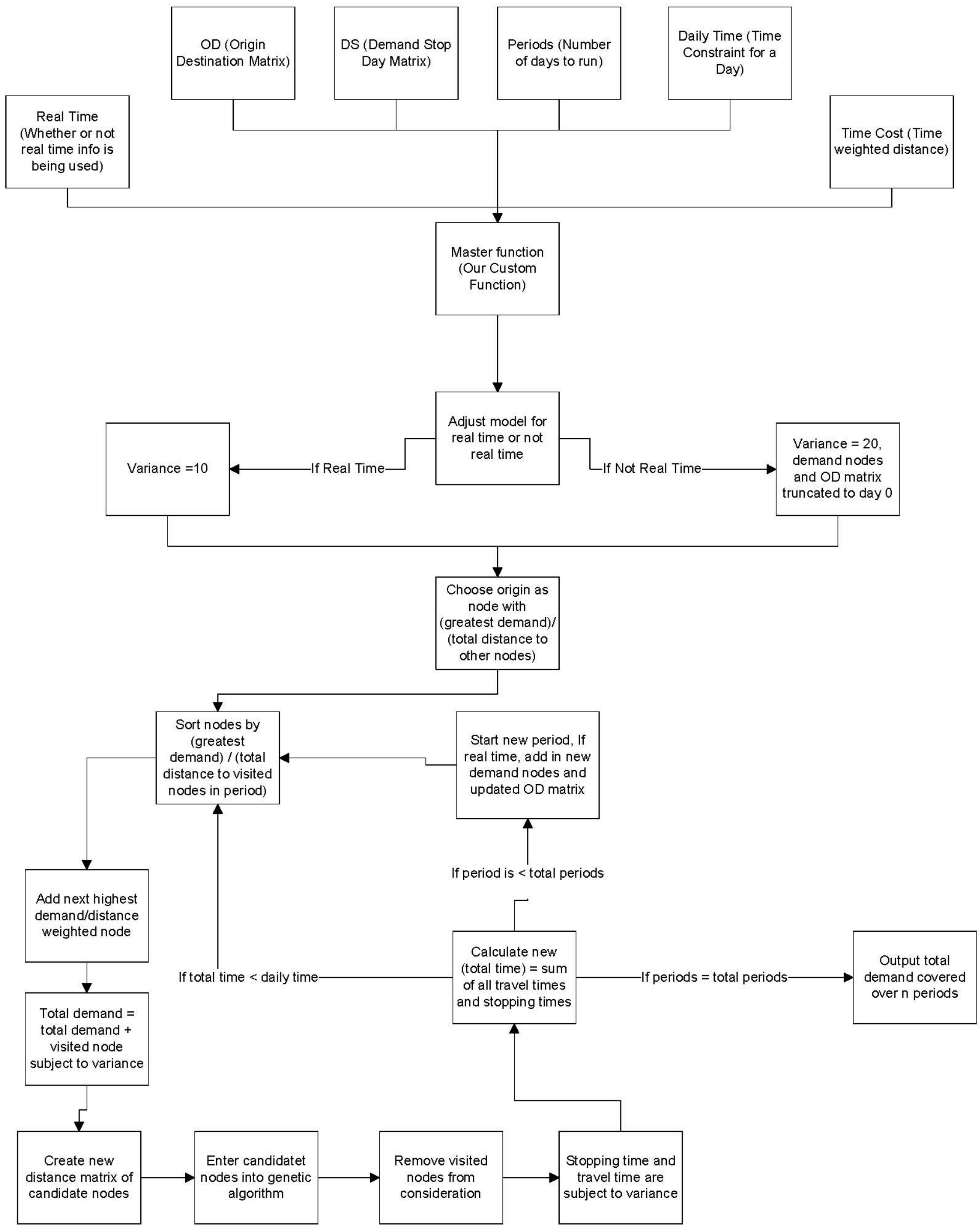

Figure 5: Visual representation of the solution procedure

Although the AMPL model is able to efficiently solve the basic problem, it is unable to accommodate the following intricacies:

- For both the real time and non-real time model, we require a constantly updating model over a course of time (five days). AMPL would only be able to solve for one day at a time and we would need to manually update all aspects of the model after each day, i.e., removing nodes visited and updating with real time OD matrices.
- One motivating factor behind the inclusion of real time data in the model is the likelihood of reducing variance in demand, stopping time, and travel time. AMPL would require datasets to be generated ahead of time to be fed into the model. Since we planned on running this simulation at least 50 times, creating 250 ($50 * 5$ days) datasets would be highly inefficient.
- In TSP modeling, subtour eliminating constraints are needed to ensure that all nodes are visited in one single trip with a single starting and ending point. In this model, the constraint that every node must be visited is relaxed since the objective is based on maximizing demand satisfied in a limited time, but subtour constraints remain. The number of subtour eliminating constraints needed is 2^n, where n is the number of nodes. Even with only 10 nodes, we would already require 1024 subtour constraints.

Algorithm 1 **Master algorithm**

Input N from OpenStreetMap building data
Input network dataset from OpenStreetMap
Apply ArcGIS origin-destination algorithm to N to generate $OD = [\mathbf{c}_{ij}]$
Rank nodes by $W_i(N)$
Set depot: $0 = W_{max}(N)$
Set variance of D_i, S_i, c_{ij} to 20% of sample means
for $t \leftarrow 1$ to 5 **do**
  if realtime **then**
    $N \leftarrow N \cup N_t$ and update ranking of nodes by $W_i(N)$
    Rebuild network dataset with new road data
    Recalculate OD with new network dataset and candidate nodes N
    Set variance of D_i, S_i, c_{ij} to 10% of sample means
  end if
  Initialize $V_t = \{0\}, T(V_t) = 0$
  while $T(V_t \cup \{W_{max}(V_t)\}) \leq T_{max}$ and $N \neq \emptyset$ **do**
    $V_t \leftarrow V_t \cup \{W_{max}(V_t)\}$
    $N \leftarrow N \setminus \{W_{max}(V_t)\}$
    Recalculate $W_i(V_t)$ and select new $W_{max}(V_t)$
    Apply TSP genetic algorithm to $V_t \cup \{W_{max}(V_t)\}$
    return $T(V_t \cup \{W_{max}(V_t)\})$
  end while
  return V_t ordered according to TSP and display route in ArcGIS
  return D_{V_t}
end for

2.3.2 Implementation

To address these needs, we develop a heuristic in MATLAB. Figure 5 is a schematic representation of the MATLAB procedure. The entire procedure described in the master algorithm is executed in two distinct sections: ArcGIS is used to prepare spatial data inputs and to run the OD matrix solver, and MATLAB uses these inputs to execute the algorithm and call the TSP genetic algorithm as a subroutine. Data preprocessing is accomplished with a variety of tools, including several small

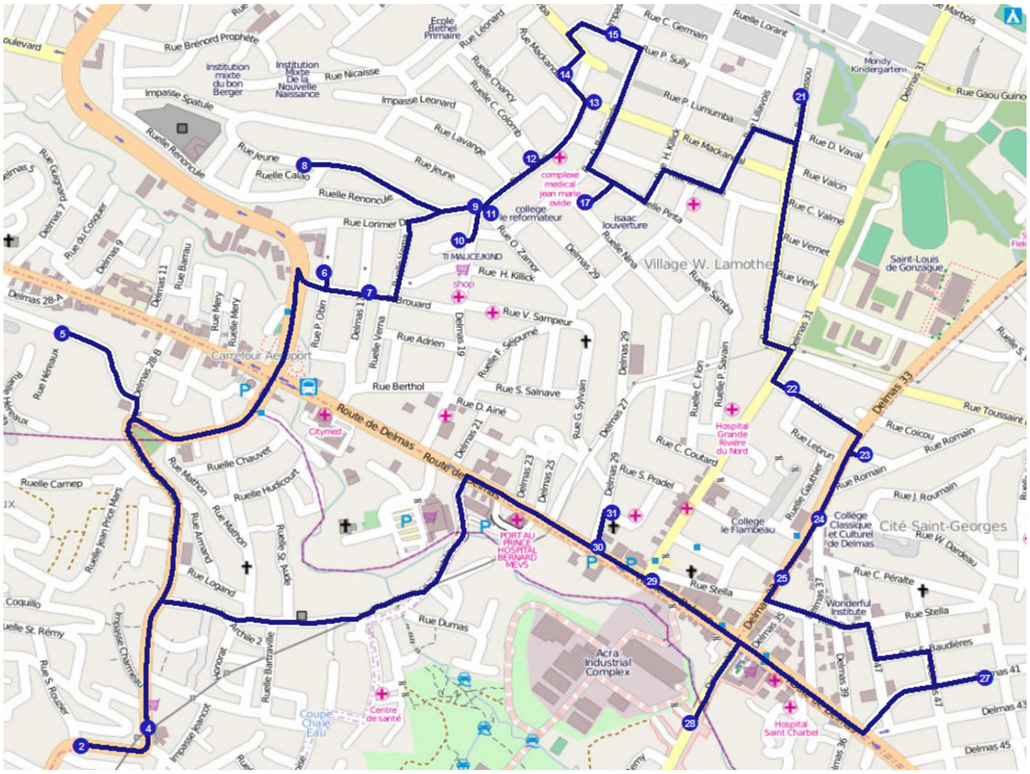

Figure 6: An example of a recommended route. Map source: OpenStreetMap.

programs written by the OpenStreetMap community to facilitate extraction and manipulation of OSM data. On a computer with a quad-core 2.6 GHz processor, execution time for both real time and non-real time models together is approximately 15 minutes.

2.4 Results and analysis

We run the simulation under four different sets of conditions, each with a different set of features included in the real time component of the simulation. These combinations are summarized in Table 1. In each of the first three simulations, one of the enhancements in the real time model was omitted. Simulation 4 includes all differences between the two procedures. Each simulation is run over 50 iterations, with resulting boxplots in Figures 7–10.

The current problem formulation assumes that real time demand data is essentially equivalent to non-real time demand data except that it is subject to lower variance in local demand, travel time, and stopping time. This assumption does not permit an assessment of the relative merit of new data sources; because the real time model simply has lower variance and more data points from which to draw a solution, it generates at least equivalent model performance essentially mechanically. That said, a comparison of the two models' output does allow us to examine how much the additional real time data contribute to solutions.

In all four simulations, there is little performance difference between the models. This is unsurprising given the relatively modest number (22) of additional demand nodes in the real time model and the conservative estimated demand at those nodes. That said, in all cases, the performance of the real time model relative to the non-real time model improves noticeably after the first period. The real time model is replenished with new data points after the first period and gains a more detailed road network as time goes on, which tends to increase the difference in data inputs between

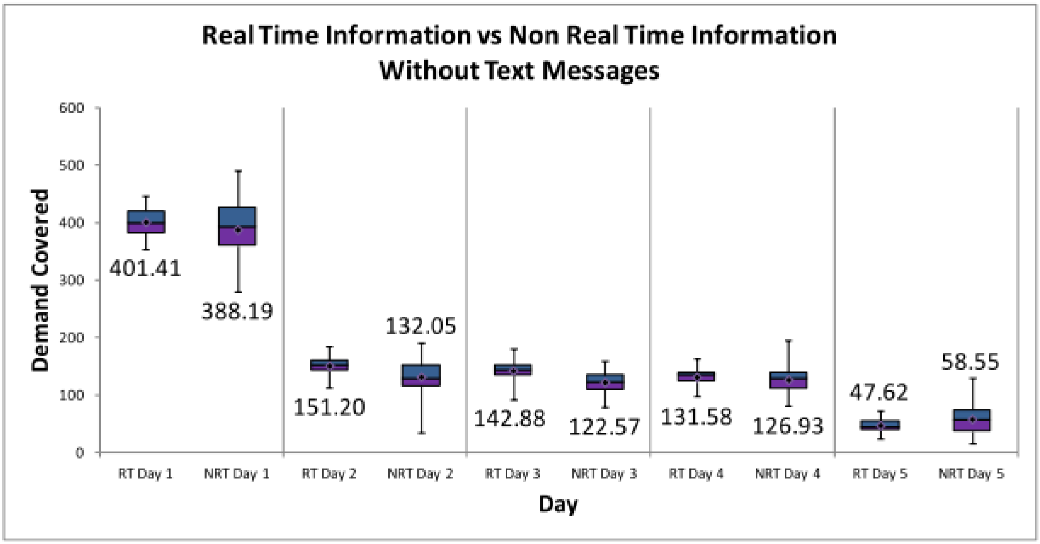

Figure 7: Simulation 1

the two models with time.

	New nodes	Map changes	Different variance
Simulation 1		X	X
Simulation 2	X		X
Simulation 3	X	X	
Simulation 4	X	X	X

Table 1: Four sets of simulations

2.4.1 Next steps and recommendations

The work presented here is only a preliminary step in the development of a set of deployable relief routing models and should be regarded as a demonstration of the feasibility of using the data sources described at an organizational level and a motivation for the work described in the following sections of this report. Among the most immediate limitations of the approach demonstrated here are as follows:

- We reduce the problem here to a single mobile team. Expanding the scenario to multiple depots and teams within this algorithm requires determining stop order using algorithms modeling the vehicle routing problem rather than the traveling salesman problem.
- The model presented here assumes that all data inputs have already been validated and all additional information is beneficial. Assessing the validity of crowdsourced information in crisis settings is a highly active area of research, and determining appropriate methods of filtering data before sending it to end users (that is, into a routing algorithm such as this one) is a critical element of any usable tool.
- Several parameters, including the variance of real time and non-real time data and the cost multipliers for each road type, are essentially arbitrary in the present work and should be replaced with empirically derived numbers.

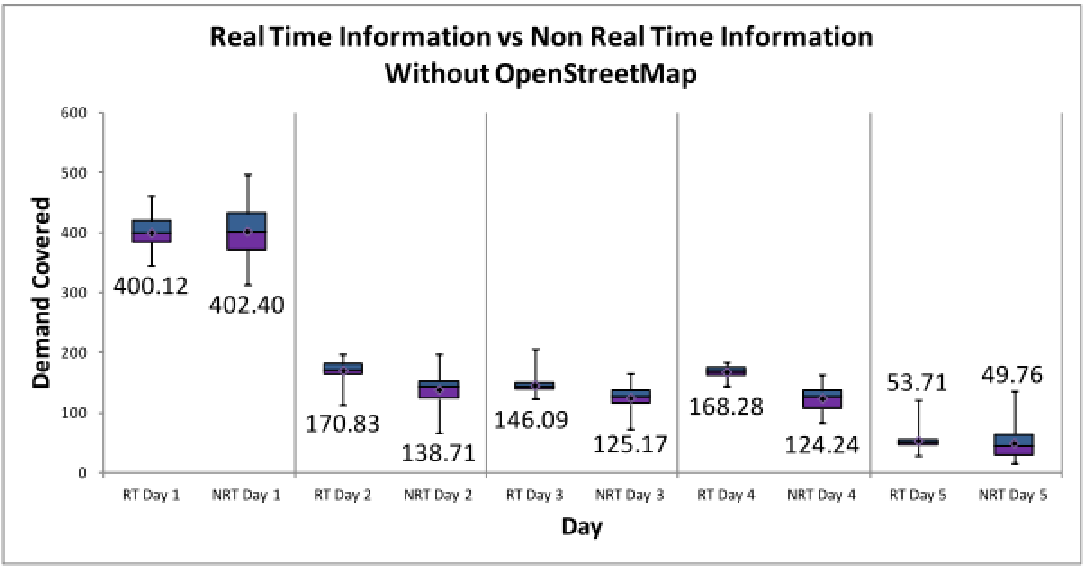

Figure 8: Simulation 2

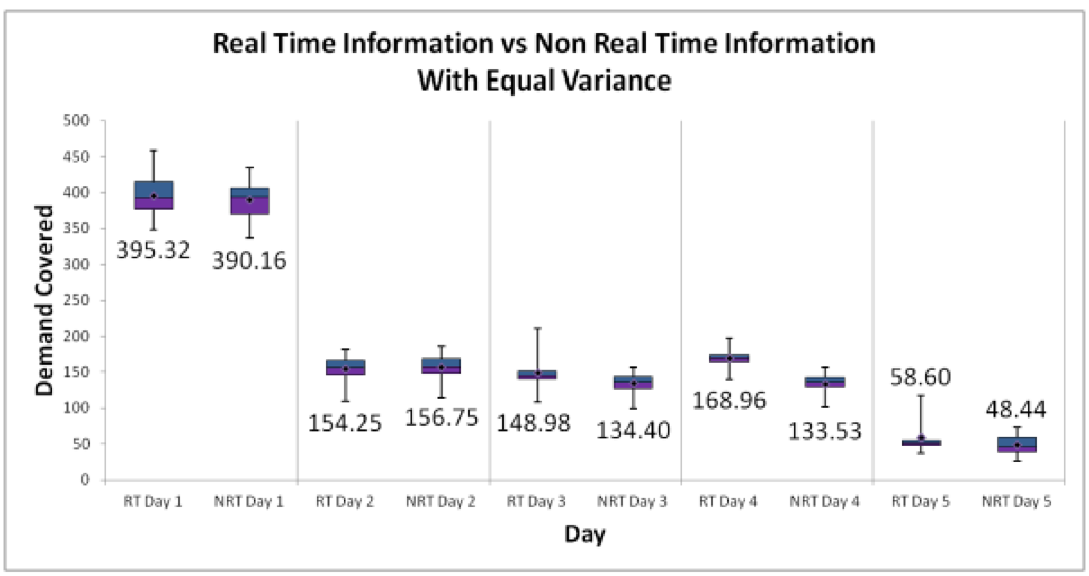

Figure 9: Simulation 3

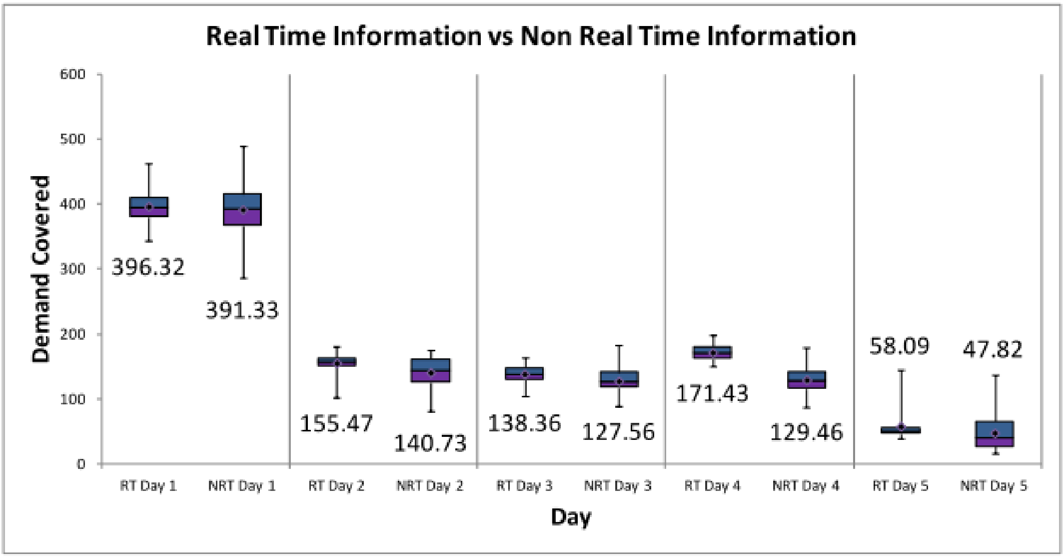

Figure 10: Simulation 4

- We model service time as a linear increasing function of local demand with a normal distribution. More appropriate estimates of service time depend on the setting.
- Future work should take into account "spoke" traveling where either the full SAR team or a portion of it must return back to the base node after visiting certain nodes to drop off individuals requiring medical attention or to replenish supplies.
- The real time data processing workflow here does not incorporate information on non-traversable roads, which would allow the solver to avoid road segments that are not traversable due to flooding, debris, damage, and the like. This is a straightforward extension of the network dataset creation step; assessing the relative usefulness of this inclusion requires adding a realistic penalty to the non-real time model for designing routes over roads that the team is unable to traverse.
- The current optimization model is updated with new data on a daily basis. To make the model more reactive to real time data, more frequent updates should be considered.
- Finally, the SAR team itself gathers important information about demand and road conditions. More detailed models should capture this information in addition to exogenous updates.

3 Testbed Development

This section describes the process of generating road network datasets ready for routing analysis using freely available, collaboratively updated spatial data from the OpenStreetMap (OSM) project. We begin by briefly describing the context and motivation for this work (Section 3.1) and the particular need for processed routing data for academic research (Section 3.2). We then detail the procedure used for generating the network data employed in the pilot project described in Section 2 (Section 3.3). These test case data are available from the Humanitarian and Non-Profit Logistics server, `ftp://curie.iems.northwestern.edu`. Due to privacy considerations, only road network data, not geocoded text messages from the 4636 project, are part of this data release.

3.1 Current and future landscape of data streams

The success of creating field-appropriate dynamic stochastic routing models for humanitarian logistics and transport is tightly related to the information flows and data that act as inputs into the models. New information streams (e.g., text messages, Twitter, online mapping), often created by newer technologies and social media, can create a real-time picture of disaster and humanitarian environments. The processes required to filter, structure, and make sense of this large body of information are changing to match these new streams.

The datasets themselves present new challenges. New information streams are generated by various groups (e.g., affected communities, organizations, satellite providers) and in different formats (e.g., SMS, Twitter, image files). These disparate pieces of information, often generated simultaneously, are processed by individuals and groups (e.g., specialists, volunteer groups, computers). As each new disaster unfolds, datasets change in structure, size, and language. Nonetheless within these information streams exist important road data, requests for food and services, and, in some settings, time-critical requests for SAR. Research to date reflects the growing interest in these new humanitarian datasets, and practitioners themselves bring light to the potential impact of these new information streams on humanitarian response. A recent report by the Harvard Humanitarian Initiative [14], coauthored by PI Chan, presents a broad overview of the evolving landscape, highlighting new volunteer-generated online maps that influence operational response, and also presenting challenges and potential impact on future emergencies.

Mission 4636 and Ushahidi Haiti (see Section 2) pioneered approaches to systematizing the use of SMS messages generated by crisis-affected communities using human translators and manual information processing. Collaboration between those groups and first responders showed a glimpse of the potential impact of using crowdsourced information to assist disaster-affected individuals [21, 24]. Recently published research shows the increasing academic attention to these large swaths of information and how this information can be used for various humanitarian response activities [23, 2, 31, 4, 29].

3.2 From raw data to routing data

Despite this increased attention, academic researchers in many fields face significant challenges in analyzing much of the data already in existence from prior disasters. This is particularly true for researchers interested in routing, an application requiring highly specialized data types and a large amount of pre-processing. While routing models can be tested on simulated data representing a stylized picture of the problem scenario, it is much more desirable to assess the salient characteristics of the disaster response setting using real-world data; furthermore, routing models must be applied to real data if they are to be useful, and this requires designing workflows to accommodate the kinds of data streams commonly seen in disaster settings.

In general, data generated during disasters are most immediately available either in raw form or as consumption streams targeted toward particular kinds of use. For example, to pick a simple case, data on road blockages may be available either as a set of tagged points in the OpenStreetMap master database or as a visualization layer that displays those points on a digital basemap in a web browser. This map is the most immediately useful way for end users to interpret the data.

Although (in this case) free and open, neither of these forms make this useful information available for routing applications without additional processing and decision-making. There are additional steps requiring both technical and conceptual work. On the conceptual side, questions for this simple road blockage dataset include: How were these data points collected, and how precisely are they located? How many different types of road obstruction (sinkhole, landslide, debris, etc.) are recorded and how should we decide how to modify the characteristics of a road segment in our network database based on each type of reported blockage–for example, should we increase the cost of traversing the segment, or perhaps mark the entire segment as non-traversable, or only traversable by some kinds of vehicles or in one direction? The answers chosen for these conceptual questions must then be translated into rapidly deployable technical procedures for using the raw data to generate a database ready for solving routing problems.

3.3 Road network data in real time: Haiti 2010

The project described in Section 2 centers on incorporating real time information on road conditions and demand into vehicle routing algorithms. The project uses data from OpenStreetMap (OSM) and Mission 4636 as inputs into vehicle routing models that are modified specifically to accommodate these real-time data streams.

The datasets described below are used to generate an origin-destination (OD) matrix, in which each matrix element is the cost of traveling along the road network from an origin to a destination, where travel cost in this case is estimated travel time. OSM is the sole source of data used to construct the road network and the weights used to estimate travel time.

3.3.1 Raw data collection

The first step is to collect the raw OSM data. Archived OSM data are available at `http://planet.osm.org/` in the form of snapshots of the entire database at a point in time, called *planetfiles*, and additional files describing changes made to the database over a given span of time, called *changesets*. We clip the study area (a bounding polygon of Haiti) from the 1/13/2010 planetfile and download all daily changesets from 1/13/2010 to 1/31/2010, using a script to clip each changeset to the study area, apply it to the previous day's .osm file, and save the new .osm. This process gives us daily OSM extracts from immediately after the earthquake to the end of January. (These extracts are available at `ftp://curie.iems.northwestern.edu/haiti_OSM`.)

There are two major potential methods for using these daily extracts to generate routable networks: ArcGIS, a commercial geographic information system, or PostGIS, a spatial extension of the open source PostgreSQL relational database management system. Each choice has advantages: ArcGIS enjoys industry standard status and is more easily interoperable with the data formats most commonly used in the humanitarian field, while PostgreSQL, in addition to being freely available and open source, supports OSM's data model based on key-value pairs, making interaction with OSM potentially more efficient. In this project, we use ArcGIS, mainly due to user consideration: porting the process to PostgreSQL is generally straightforward, but wrapping the resulting network data in a user-friendly and flexible system is an undertaking for future work.

We convert each .osm in its entirety to an Esri geodatabase using ArcGIS Editor for OpenStreetMap 1.1. While this is not the fastest approach and would not be suitable for real time deployment, it involves the fewest data processing steps during which information could be lost. A recently released update to this interoperability software may make this data translation step significantly faster in production environments.

3.3.2 Creating a network

Network Structure. Before creating a network dataset, we conduct some exploratory data analysis to characterize the data, beginning with basic elements of network topology. We note that, since road segment elements in OSM do not necessarily begin and end at intersections, the network requires *vertex connectivity* rather than endpoint connectivity, and that the network must be built with *global turns* (i.e., assuming turns in any direction are possible at each intersection) rather than a turn source. While the absence of robust data on turn rules at intersections is a potential issue for many routing applications, it is likely to be a minor concern in post-earthquake Haiti.

Network datasets must also keep track of roads that pass above and below each other without intersecting. This is commonly accomplished with elevation fields. For OSM Haiti data, this is not an option due to missing data. Nor do the *bridge* and *layer* tags give consistent and reliable elevation information suitable for routing in this context; while we build the networks without taking grade separation into account on the grounds that there are relatively few grade-separated road segments in the study region, this is not in general an acceptable solution. This limitation is still common in many geographic areas, but as OSM's database grows and users continue to improve tags used for routing, elevation data will become more usable for more areas of the globe.

Road segments in OSM have a non-nullable (mandatory) tag, *highway*, which records road type. We delete segments with *highway* values indicating that they are not traversable by vehicles, including "path," "footway," and "pedestrian."

The final steps in characterizing road networks are to define the time cost variable and to determine whether to include barriers. The number of tags recording road barriers (including blockages due to debris, damage to roadbed, etc.) grew precipitously over the study period, although the majority were recorded using the "impassable→yes" tag. For the current work, we elect not to use barriers due to modeling considerations (see Section 2.2.4), although it is straightforward to add them back to the completed network datasets.

Travel Cost. Besides network structure, the other fundamental component of a network dataset for our purposes is travel cost, the cost (in our case, travel time) incurred by a vehicle traversing a road segment. Travel times can be estimated with increasing precision as data availability increases, starting with rough estimates based on road class, to estimates based on historical traffic patterns and logged travel speeds recorded by GPS devices, to real-time data collection from GPS devices such as mobile phones. While fine-grained data of the latter types are frequently used in commercial settings and are being innovated and tested in open data settings, for our study setting, it would only have been possible to estimate travel times based on time-invariant road characteristics.

To characterize the cost of traversing a road segment, we place segments into categories based solely on *highway* tag values. We create a new road class variable and map *highway* tag values "primary," "primary_link," and "motorway" to 1; "secondary" and "secondary_link" to 2; "tertiary," "residential," and "unclassified" to 3; and all others to 4. We then define *time cost* along a segment as the length of the segment multiplied by a parameter depending on road class, ranging from 1 for road class 1 to 1.6 for road class 4. The parameters used are essentially arbitrary and serve as a stand-in for empirically determined parameters derived from the literature or from ground experience; in both cases, the relative (and absolute) travel costs associated with different road types are likely to vary by locality. Since these parameters simply scale travel cost based on shape length, they are functionally unitless and have only relative meaning. We consider our travel cost definition a distinctly second-best option, but it does have the desired effect of prioritizing travel along more robust roads. Users may calculate their own travel cost variables on this set of network datasets.

With the above approximations and compromises, it is possible to generate a series of network datasets with surprisingly sound results. In the project described in Section 2, the team uses these network datasets to calculate OD matrices of points from processed 4636 data, which are in turn used as inputs for routing algorithms implemented in MATLAB.

4 Model

This section presents the mathematical model of the search and rescue operation problem. Section 4.1 presents the problem, emphasizing its specific characteristics and related literature. The preliminary mathematical models are presented in Section 4.2, and their analysis is discussed in Section 4.3. In Section 4.4 we detail our ongoing and planned work, including solution approaches for the preliminary models and other models we plan to develop.

4.1 Problem Overview

Consider the setting of search and rescue (SAR) operations in Port-au-Prince following the 2010 earthquake. Numerous SAR teams arrive in Haiti from around the world to assist in rescuing survivors. To facilitate this operation, the affected area is divided into regions, each assigned to an SAR team. The goal of each team is to search collapsed structures within the area to rescue as many survivors as possible within a specified deadline (usually 48-72 hours in such settings). Each SAR team must commit to a set of structures to search and communicate this information to other teams at the beginning of the time horizon. This ensures coverage of areas by the multi-team SAR operation and avoids redundancy and inefficiency.

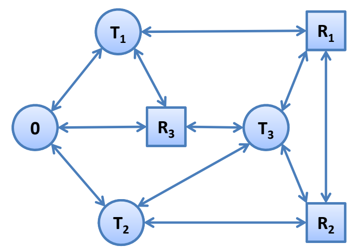

Figure 11: Example of Adaptive Orienteering Problem network with depot node (0), transition nodes (T_1, T_2, T_3) and reward nodes (R_1, R_2, R_3).

From the perspective of a single team, the decision problem can be formalized as an adaptive orienteering problem as follows. Given is a network $G = (N, A)$ of nodes $i \in N$(see Figure 11) with a specified *depot node* where a specific SAR team is housed, $i = 0$, a subset of *reward nodes*, $i \in N_R$, (e.g., collapsed buildings with potential survivors) and a set of *intermediate/transition nodes*, $i \in N_T$, (e.g., corresponding to road intersections) that define paths among the depot and reward nodes, where $N_R \cup N_T \cup \{0\} = N$. The set of arcs is denoted by A. The problem models operations of a single team with **three sets of decisions**: **(1)** which reward nodes to commit to visit, **(2)** the order in which to visit these reward nodes, and **(3)** the path to follow between each consecutive pair of reward nodes, defined by transition nodes. Other problem characteristics are:

Reward node commitment. The team commits to a set of reward nodes at the beginning of the planning horizon, and must visit those nodes and cannot visit other reward nodes. This problem characteristic directly impacts the available path choices.

This assumption is necessary in practice to ensure there are no gaps or overlaps in service among the teams.

Deadline. There is a deadline, D, for collecting reward from visiting a node; each reward node $i \in N_R$ is associated with a reward value r_i received for visiting before the deadline and a penalty e_i for arriving past the deadline. The problem objective is to maximize the expected profit, equal to the collected reward minus the incurred penalties.

In SAR operations, finding a survivor after the initial 48 hours is highly unlikely. If a team commits to a building to search but does not arrive on time, there is a penalty for not giving another team an opportunity to rescue people at that location.

Stochastic Travel Times. Travel times on the network arcs are stochastic and are realized after a team traverses that arc. Once the travel time is realized, that realization remains fixed for the remainder of the time horizon. This is an important fact since one might need to traverse an arc multiple times during the operation. In the preliminary model presented here, we assume there are no exogenous information updates about the network during the problem time horizon. In addition, travel times on the arcs are assumed to be independent of each other. These assumptions are relaxed in ongoing work described in Section 4.4.

Following an earthquake and other disasters, little information about the transportation infrastructure is known. One has to travel through the network to learn the network.

A large class of humanitarian logistics problems falls within a class of previously studied problems in optimization called the orienteering problem (OP). The orienteering problem has been extensively studied in the existing literature and is closely related to the Selective Traveling Salesman Problem (TSP) and the TSP with Profits. The overview of the orienteering problem and related variant problems can be found in review papers [10, 30]. The objective of the OP is to visit a subset of available customers within a specified deadline in order to maximize the total reward we collect from visiting the customers. To account for the transportation network and infrastructure uncertainty encountered in humanitarian settings **we focus on a variation of the orienteering problem with stochastic travel times**. These problems are less studied variations of the OP. Ilhan et al. [17] study the OP with stochastic profits and deterministic travel times, focusing on selecting a subset of demand nodes in order to maximize the probability of collecting reward above a given threshold before a deadline. Settings of the orienteering problem studied by Campbell et al. [3] are most closely related to our proposed work. They assume stochastic travel and service times, with deterministic rewards and deterministic penalty. They focus on finding the a priori path (subset of customers and visiting order) that maximizes the expected profit. In the proposed research we extend the work by Campbell et al. [3] to include path finding components of the problem and evaluate adaptive, as opposed to a priori, solutions. In the most recent work, Gupta et al. [12] discuss approximation algorithms for the stochastic orienteering problem and propose an a priori solution to approximate the optimal adaptive policy with a guaranteed lower bound on the worst performance. While we are not able to find any work on adaptive OP models, there is a significant body of literature on adaptive versions of related problems: shortest path, traveling salesman, vehicle routing, and knapsack problems, which we discuss next.

In addition to stochastic travel times, **we introduce a path finding component to the problem** by defining a subset of intermediate nodes without an associated reward that represents alternative paths among reward nodes and the depot. Hall [13] was the first to discuss the shortest path problem with stochastic and time-dependent travel times, where the arc travel times are realized after traversal, and the next route choice is conditional on the location and time of arrival at that node. Hall shows that the routing policies are better than a priori paths in the stochastic and time-dependent network. Dynamic programming approaches are used to find those optimal polices. The papers that followed propose more efficient algorithms to solve variations of the problem. Miller-Hooks [19] studies the adaptive least expected time paths in the stochastic time-varying network by generating hyperpaths for different departure times. In the following paper, Miller-Hooks and Yang [20] studied a problem where real-time information is observed within a small neighborhood of the vehicle. An important conclusion emerges from their analysis: knowing only a small neighborhood of information yields decisions that are equivalent to decisions made with wide information. It is important to emphasize that inherent dependence of the paths between two distinct pairs of reward nodes in our problem prevents us from direct application of the algorithms developed for adaptive shortest path problems. It may be more beneficial to find a path that puts us into a better position for visiting additional nodes in the future rather than the shortest path between two reward nodes.

Finally, **we allow our model to react to the dynamic nature of uncertainty and information updates** by developing adaptive models and solving for an optimal policy, as opposed to an a priori path. Thus, the decisions regarding customer (reward node) sequence and path choice affect the information gathered and require exploration versus exploitation analysis. Adaptive and reoptimization models for TSP and vehicle routing problem (VRP) have been studied in the literature by integrating model reaction with updated information. The probabilistic traveling salesman problem is introduced by Jaillet [18], where some customers can be absent with a known probability distribution. An a priori tour is built that visits all the customers, and the absent customers are skipped, while continuing along the chosen tour. Secomandi and Margot [26] study the VRP with stochastic demand via reoptimization approaches, where the tour is reoptimized after the stochastic demand is realized from visiting a customer. Tang and Miller-Hooks [28] study the selective TSP with stochastic service and travel times. In their problem, the authors use a chance constraint model to find an a priori tour to visit a subset of customers that maximizes expected profit.

We present adaptive path planning methods to take advantage of dynamically updated data, combining OP and optimal path finding into a single model, and propose to study a new class of problems that we call the Adaptive Orienteering Problem with Stochastic Travel Times (AOPST).

4.2 Optimization Model

The focus of the proposed work is integration of continuously updated data into humanitarian relief logistics models. The development of adaptive optimization models will allow us to achieve this goal, as well as to assess the value of available information. However, just like the dynamic nature of a disaster itself, the sources and types of data that are available at one region or time period versus the next one might differ. In addition, it is hard to anticipate the level of adaptivity and changes in relief operations that would be feasible to implement in various settings. Thus, we propose to study a variety of adaptive techniques, degrees of operational flexibility, and levels of data updates available on the ground. Initially, we introduce two levels of adaptive models for AOPST, differentiated by the subset of decisions that have to be made at the beginning of the planning horizon and the decisions that can be dynamically adapted to the realized information.

Level 1 model ($L1$). The adaptive orienteering problem can change the paths between consecutive reward nodes, maintaining the fixed order and subset of customers.

Level 2 model ($L2$). The adaptive OP can change the order of visiting reward nodes and the paths between those nodes, but maintains the selected subset.

In addition, a level 3 model that can update all its decisions (subset of customers, order, and paths) at any time during the operation is introduced later as a benchmark model.

First, we introduce additional notation for the formulations of AOPST. Let X_{ij} denote stochastic travel time on arc (i, j), which is assumed to be discrete, where m_{ij} is the finite number of realizations of X_{ij}. Then x_{ij} is the realized travel time on arc (i, j), and H corresponds to a set of previously traversed arcs and their realized travel times, e.g., $H = \{(i_1, j_1, x_{i_1 j_1}), (i_2, j_2, x_{i_2 j_2}), ...\}$.

Our problem decomposes into a two stage model. The first stage (*master problem*) decisions are to select a subset of reward nodes with the highest expected profit; note this subset must be ordered in the Level 1 model. Then, for a given subset of reward nodes, the corresponding second stage (*subproblem*) finds the optimal routing policy between the reward nodes. As we allow our second stage model to dynamically adjust its path when more information about the network is learned, we model the second stage as a dynamic programming (DP) problem.

Master Problem

The master problem can be described as an optimization problem,

$$(L1): \quad \max \quad U(y_1) \qquad \text{s.t} \quad y_1 \subseteq_{ordered} N_R \qquad\qquad (L2): \quad \max \quad U(y_2) \qquad \text{s.t} \quad y_2 \subseteq N_R.$$

In the optimization models above, y_1 represents an ordered subset of reward nodes, while y_2 is an unordered subset. Then $U(y_i)$ corresponds to a function of y_i for $i = 1, 2$, whose value represents the optimal expected profit achieved from visiting all the nodes in y_i. Notation $y_1 \subseteq_{ordered} N_R$ denotes the elements in ordered set y_1 corresponding to a subset of N_R. For the level 2 model, we simply have $y_2 \subseteq N_R$, since the order can be selected and changed in the second stage of the problem, based on travel time realizations. To compute the value of $U(y_i)$ we solve the DP subproblem.

DP Formulation of Subproblem

Given a set of reward nodes from the master problem, we develop a dynamic programming model to evaluate the expected profit. Our DP model makes sequential decisions on which node (be it transition or reward node) to travel to next. Note that in both models we are restricted to only visiting the selected reward nodes. Furthermore, in $(L1)$ we also must follow the specified order of visiting those reward nodes. This, in turn, limits the set of feasible decisions. The stochastic travel time on each arc is realized after traversal. Then, based on the observed information, we choose the path that has a higher probability of visiting the remaining reward nodes before the deadline. Recall that all the reward nodes selected in the master problem have to be visited. Unlike the a priori orienteering problem, the optimal routing policy in AOPST provides a set of paths between consecutive nodes, i.e., an optimal policy, rather than a single path. The details of the DP model follow.

Given an (ordered, for $(L1)$) subset of committed reward nodes, N_C, where $N_C \subseteq N_R$, we develop a dynamic programming model to calculate the optimal visiting policy and the corresponding optimal expected profit. Assuming $|N_C| = k$, we rewrite $N_C = (n_1, n_2, ..., n_k)$.

DP state: $s = (i, t, H, I)$, keeps track of the current location ($i \in N_T \cup N_C$) and current time ($t \geq 0$), as well as the set of realized arc travel times previously traversed (H) and the set of already visited reward nodes (I).

DP action: $a = j$, corresponds to the decision to move next to node $j \in \mathcal{A}(s)$, where given the current state $s = (i, t, H, I)$, the action space $\mathcal{A}(s)$ is

$$\mathcal{A}_{(L1)}(s) = \begin{cases} j \in N_T \cup_{i=1}^{|I|+1} n_i & \text{for } I \subset N_C \\ \emptyset & \text{for } I = N_C \end{cases}, \qquad \mathcal{A}_{(L2)}(s) = \begin{cases} j \in N_T \cup N_C & \text{for } I \subset N_C \\ \emptyset & \text{for } I = N_C \end{cases}.$$

Observe that in the case of the $(L1)$ model, a reward node becomes equivalent to a transition node once it is visited. Also note that the sets $\mathcal{A}(s)$ are different for the two models since we relax the visiting order of the reward nodes in $(L2)$.

DP state transition: $g(s, a)$, is the state transition function that returns a state to which the system transitions when action a is chosen in state s. For state $s = (i, t, H, I)$, when action $a = j$ is chosen, the model transitions to state $s'_k = (j, t + x^k_{ij}, H'_k, I')$ with probability $P(X_{ij} = x^k_{ij}|H)$ for $k = 1, ..., m_{ij}$, where

$$H'_k = \begin{cases} H & \text{for } (i,j) \in H \\ H \cup \{(i, j, x^k_{ij}\} & \text{for } (i,j) \notin H \end{cases} \quad \text{and} \quad I' = \begin{cases} I & \text{for } j \notin N_C \setminus I \\ I \cup \{j\} & \text{for } j \in N_C \setminus I \end{cases}.$$

Note that if the arc (i, j) has been previously traversed and element $(i, j, x^k_{ij}) \in H$ for some $k \in (1, ..., m_{ij})$, then $P(X_{ij} = x^k_{ij}|H) = 1$ and $P(X_{ij} = x^{k'}_{ij}|H) = 0$ for $k' \neq k$. Otherwise, $P(X_{ij} = x^k_{ij}|H) = P(X_{ij} = x^k_{ij}), \forall k \in (1, ..., m_{ij})$ (due to the preliminary independence assumption).

DP action cost: $R(s, a)$. For state $s = (i, t, H, I)$, when action $a = j$ is chosen, $R(s, a)$ denotes the expected profit gained from traveling from node i to node j, starting at time t, given travel time information H and set of already visited nodes I.

$R(i, t, H, I, j) = (r_j + e_j)P(X_{ij} \leq D - t|H) - e_j$ for $j \in \mathcal{N} \setminus I$, and 0 otherwise.

Recursive equation

Let $V(s)$ denote the maximum expected reward the system could obtain, starting from the current state $s = (i, t, H, I)$ until the end of the problem horizon. The following DP recursive equation evaluates the value of $V(s)$,

$$V(s) = \max_{a \in \mathcal{A}(s)} \left\{ R(s,a) + \sum_{k=1}^{m_{ij}} P(X_{ij} = x^k_{ij}|H)V(g(s,a)) \right\}. \tag{5}$$

Boundary conditions: $V(i, t, H, N_C) = 0, \forall i, t, H$.

Given the (ordered) subset N_C, we find the optimal policy to visiting the reward nodes in N_C and the corresponding optimal expected profit by recursively solving the equation (5).

When we find the optimal value for $V(s_0)$ with $s_0 = (\{0\}, 0, \emptyset, \emptyset)$, $U(N_C) = V(s_0)$.

4.3 Analysis of Models

4.3.1 Path Planning Component

Unlike the OP variations studied in the literature (where every node is a reward node and the network is a complete graph), our network is not complete and a significant number of nodes are transition nodes that do not have associated reward. As a result, one chooses the optimal path between each pair of committed reward nodes. It is important to highlight the level of complexity this brings to the problem. In the traditional OP, the travel times between each pair of reward nodes are independent of each other, static, and independent of the reward nodes we have visited so far. In AOPST, those **travel times are not independent of each other, nor of the set of visited nodes** since we can traverse the same arcs multiple times, thus taking advantage of the information learned about a given arc travel time the first time it has been traversed. As a result, the value of each potential path between a pair of reward nodes is two-fold: (1) earlier arrival to a reward node delivers greater expected profit, and (2) realization of stochastic travel times for the traversed arcs puts us into a better position to make the following sequence of routing decisions. These two benefits are often conflicted, and our model has to find the appropriate tradeoff between the two. This relates to a widely studied analysis of exploration versus exploitation in the reinforcement learning literature (see [32] and [1] for examples in path finding settings and [25] for more general discussion).

4.3.2 Challenges of the Two Stage Model

Recall that a subset of our decisions (set of committed reward nodes to visit, as well as visiting order for $(L1)$) is made at the beginning of the trip and stays fixed for the duration of the problem horizon, while the remaining decisions are dynamically updated enroute when more information about the network is gathered. This characteristic of the problem leads to a two stage model: (1) master problem and (2) DP subproblem.

The master problem is a variation of the **knapsack problem**, where we choose a subset of items to maximize the total reward. A number of adaptive stochastic knapsack problem variations have been studied in the literature, with random size or reward of the items. The adaptive models sequentially select the items to add to the knapsack, while uncertainty is realized after the item is inserted. Ilhan et al. [16] study the adaptive stochastic knapsack problem with deterministic size and stochastic rewards. Their problem objective is to find a sequential inserting policy to maximize the probability of the reward exceeding some threshold value without violating the capacity constraint. Dean et al. [6] study the adaptive stochastic knapsack problem with items of deterministic reward and stochastic size. Their goal is to maximize expected value while fitting all the items in the knapsack. The authors demonstrate the benefit of the adaptive policy and provide an approximate adaptive policy. In our problem, we also assume deterministic reward; however, the arrival time to a reward node depends on the previous nodes (i.e., items) visited. This dependency adds complexity beyond the adaptive stochastic knapsack problem, which is shown to be PSPACE-hard [6]. Furthermore, unlike the knapsack problem, the AOPST does not have the additive property of the value each selected item contributes, and the value of each selected item depends on what other items have been selected, and in $(L1)$, the order of adding items further impacts each item's value.

The master problem models are computationally demanding, due to the fact that the total value of any given subset of reward nodes selected does not satisfy the properties currently exploited in the existing literature. Thus, if we want to add one more reward node to a currently selected subset, to evaluate the additional value of that node one has to **resolve the dynamic programming subproblem from scratch** for this new subset of reward nodes. This pure enumeration approach is extremely inefficient. To overcome this challenge we propose to study the underlying structure and properties of our problem to guide the master problem to an optimal (or near optimal) solution.

Furthermore, the DP subproblem to evaluate each feasible solution to the master problem is negatively impacted by the curse of dimensionality [25]. Thus, DP computational time grows exponentially with the size of the network, deadline value and discrimination of travel time distribution (since a longer deadline allows us to consider larger set of reward nodes to commit to). In Section 4.4 we propose a solution approach to manage the computational challenges of our models.

4.4 Ongoing Work

Our current work on solution approaches for the preliminary models described above is discussed in Section 4.4.1. We will build on these solution approaches in the development of new routing models that integrate exogenous information sources (Section 5.1) and ensure compatibility with characteristics of the data sources (Section 5.2). These new models will be driven by outcomes of ongoing work focused on the data aspect of our project (Section 5.3).

4.4.1 Solution Approach for Proposed Models

Due to the lack of structure of the master problem, pure enumeration of feasible solutions might be required to find the optimal solution. Furthermore, to evaluate each feasible solution one needs to solve a computationally demanding dynamic programming model. Note that if we assume $|N_R| = n$, there are $O(2^n n!)$ number of possible subsets to be considered in $(L1)$, and $O(2^n)$ in $(L2)$. We propose a two-stage search-based heuristic to find a near optimal solution. The first stage is focused on analyzing the structure and properties of the DP subproblem to identify a promising neighborhood of feasible solutions to the master problem. Then, in the second stage we perform local search of the neighborhood to find a local optimal solution. A number of neighborhoods might be identified and explored to improve performance of the heuristic.

Stage 1. Subproblem Properties

Dynamic programming models are often affected by the curse of dimensionality; however, there are a number of techniques to reduce computational demand. Analysis of the problem structure allows us to identify the properties of the problems and their optimal solutions, subsequently delivering necessary conditions for optimality, state dominance and other rules that decrease the computation demand by pruning the decision and/or state space. PI Dolinskaya has successfully employed these tools on similar problems of real-time dynamic vessel navigation [9, 7, 8].

To enhance the implementation speed of the AOPST models presented above, we state a number of preliminary results. (We present Propositions 1 and 3 without proofs, as these are straightforward results.) Proposition 1 allows us to significantly reduce the number of the DP states considered in the model, since this state dominance restricts our algorithm from considering the paths that circle back and forth between nodes. Note that this property is specific to the path-planning component of our model that allows to traverse the same network arcs multiple times.

Proposition 1. $V(i, t_1, H, I) \geq V(i, t_2, H, I), \;\; \forall t_2 > t_1, i, H, I.$

Proposition 2.

$$P(X_{ij}(h) \leq D - t|h) \geq P(X_{ij}(h) + X_{jm}(h) \leq D - t|h) \quad \forall t > 0, (i,j) \in A, (j,m) \in A, h.$$

Proof: Since we assume non-negative arc travel times, we know $X_{jm}(h) \geq 0$, $\forall (j,m) \in A, h$. Then it is easy to show that $P(X_{ij}(h) \leq D - t|h) \geq P(X_{ij}(h) \leq D - t - X_{jm}(h)|h)$, $\forall t > 0, (i,j) \in A, (j,m) \in A, h$. □

This proposition demonstrate that at the system state (i, t, h, I), the probability to arrive to node j from node i at time t with travel time information h is always at least as large as the probability to arrive to node m by the deadline from node i by passing node j with the same set of real time information. While we do not apply this proposition directly during implementation, it leads to further propositions presented below.

The following propositions do not apply to all the adaptivity models, but rather just a subset of them.

Proposition 3. $V(i,t,H,I) = -\sum_{k\in\mathcal{N}\setminus I} e_k, \quad \forall t \geq D, i, H, \mathcal{N}, I \subset \mathcal{N}.$

Proposition 3 prevents further exploration of the paths once the current time is past the deadline and, in turn, prunes a set of states to be evaluated. Thus, without finding the exact path to visiting each of the remaining reward nodes, we can compute the value function $V(i,t,H,I)$.

The left-hand side of Proposition 4 is part of equation (5), and as a result, the proposition provides an upper bound estimate of the expected future reward we could collect starting at the current state (i,t,H,I) when action j is chosen. This upper bound estimate might allow us to prune the action j, without reaching the terminal state of the problem and computing the exact value of the reward.

Proposition 4.

$$\sum_{k=1}^{m_{ij}} P(X_{ij} = x_{ij}^k|H)\cdot V(j, t+x_{ij}^k, H_k', I')\} \leq \sum_{n\in\mathcal{N}\setminus I'} [(r_n+e_n)P(X_{ij} \leq D-t|H) - e_n].$$

Proof: Recall, $g(i,t,H,I,j) = (j, t+x_{ij}^k, H_k', I') = s_k'$ with probability $P(X_{ij} = x_{ij}^k|H)$ for $k = 1, ..., m_{ij}$. Then, when at node j we chose the next node to visit $m \in N$, we can bound the probability of arriving to m before the deadline as

$$P(t + x_{ij}^k + X_{jm} \leq D|H_k') \leq \mathbb{1}(t+x_{ij}^k \leq D), \forall t > 0, (j,m) \in A.$$

Then, the maximum profit collected starting at state s' is bound by the maximum reward minus the penalty one collects from all the remaining reward nodes, $\mathcal{N}\setminus I'$. That is,

$$V(j,t+x_{ij}^k, H_k', I') \leq \textstyle\sum_{n\in\mathcal{N}\setminus I_k'} [(r_n+e_n)\mathbb{1}(t+x_{ij}^k \leq D) - e_n].$$

The above expression is true for each realization of x_{ij}^k with probability $P(X_{ij} = x_{ij}^k|H)$ and we calculate the expectation of those values:

$$\sum_{k=1}^{m_{ij}} P(X_{ij} = x_{ij}^k|H)\cdot V(s')\} \leq \sum_{k=1}^{m_{ij}} \left(P(X_{ij} = x_{ij}^k|H)\cdot [\sum_{n\in\mathcal{N}\setminus I'} [(r_n+e_n)\mathbb{1}(x_{ij}^k \leq D-t) - e_n]] \right).$$

Exchange of the summation in the right side results in the desired statement. □

We assume the system state $(j, t+x_{ij}^k, h_k', I')$ transits from (i,t,h,I) by action j with state transition relationship $h_k' = h \cup \{i,j,x_{ij}^k\}$, if $(i,j,x_{ij}^k) \notin h$. $I_k' = I \cup \{j\}$, if $j \in \mathcal{N}$ and $j \notin I$, otherwise, $I_k' = I$.

In the subproblem of level 1 and level 2 model, given the pre-committed subset of reward nodes $\mathcal{N}$. At system state $(j, t+x_{ij}^k, h_k', I_k')$, there are still $\mathcal{N}\setminus I_k'$ reward nodes left to visit. Starting from node j at time $t+x_{ij}^k$, assuming non-negative travel times among all the arcs, we could have

$$\mathbb{1}(t+x_{ij}^k \leq D) \geq P(t+x_{ij}^k + X_{jm}(h_k') \leq D), \forall t > 0, (j,m) \in A$$

Since starting from any node m, at time $t+x_{ij}^k+X_{jm}(h_k')$, the probability of visiting any of the reward nodes will be smaller than $\not\vDash(t+x_{ij}^k \leq D)$. Based on this argument, we could say $\not\vDash(t+x_{ij}^k \leq D)$ is an upper bound probabilities of visiting the reward nodes by deadline for all the following states starting from the system state $(j, t+x_{ij}^k, h_k', I_k')$. The value function $V(j, t+x_{ij}^k, h_k', I_k')$ represent the optimal expected return the system could achieve starting from the current state $(j, t+x_{ij}^k, h_k', I_k')$ to end of the problem. Since there only left $|\mathcal{N}\setminus I_k'|$ number of reward nodes, if we calculate the expected reward of each of them by the upper bound visiting probabilities, the summation of those expectations will give the upper bound for the value function, which is

$$V(j, t+x_{ij}^k, h_k', I_k') \leq \sum_{n\in\mathcal{N}\setminus I_k'} [(r_n+e_n)\mathbb{1}(t+x_{ij}^k \leq D) - e_n].$$

This expression is true for each realization of x_{ij}^k with probability $P(X_{ij}(h) = x_{ij}^k)$, we further calculate the expectation of those values as follows:

$$\sum_{k=1}^{p(X_{ij}(h))} P(X_{ij}(h) = x_{ij}^k) \cdot V(j, t+x_{ij}^k, h'_k, I'_k)\} \leq \sum_{k=1}^{p(X_{ij}(h))} P(X_{ij}(h) = x_{ij}^k) \cdot \Big[\sum_{n \in \mathcal{N} \setminus I'_k} [(r_n+e_n)\mathbb{1}(x_{ij}^k \leq D-t)-e_n] \Big].$$

After we exchange the summation in the right side of the inequality, we have

$$\sum_{k=1}^{p(X_{ij}(h))} P(X_{ij}(h) = x_{ij}^k) \cdot V(j, t+x_{ij}^k, h'_k, I'_k)\} \leq \sum_{n \in \mathcal{N} \setminus I'_k} [(r_n+e_n)P(X_{ij}(h) \leq D-t)-e_n]. \quad \square$$

Further properties will be established in the course of the proposed project. For example, multi-level dominance properties that capture the two-fold value of a path between a pair of reward nodes (quick arrival versus gathered information) can help us eliminate some paths between reward nodes or a subset of transition nodes from consideration. The combination of OP and path planning into a single model presents unique research challenges and opportunities we plan to explore in-depth.

Approximate Values

Despite the need for evaluating the value of a selected subset of reward nodes, we aim to develop **bounds on the added value of an extra node** to an existing subset; e.g., follow the same path as before, and then add the best path we could follow to the newly added node. This will provide a lower bound on the value of the new set. Further study of the DP subproblem can deliver heuristics for efficiently searching the space of y before identifying a more promising neighborhood to explore more carefully.

Stage 2. Local Search

The first stage of the algorithm delivers an approximate solution or a promising solution neighborhood to the problem. In the second stage, we perform local search to improve this solution. Recall that a feasible solution to our problem is an (ordered) subset of reward nodes. To perform a local search we will explore the benefit of techniques such as deletion, addition, or substitution of the nodes in our subset. Some of the properties established for Stage 1 will also be beneficial in the local search stage of our algorithm, serving as guiding rules toward a faster convergence of the algorithm.

It is important to note that while Campbell et al.'s work [3] studied OP with stochastic travel and service times, their optimization problem is indeed a deterministic problem, since they only consider the a priori solution optimizing the expected value. Their neighborhood of the feasible solution is build by resequencing the route, replacing a customer on the route with the one not on the route, adding a new customer and deleting a customer. Due to independency of customers' sequencing, this neighborhood could be easily solved without resolving the problem. However, the dependency of the traveling times among different reward nodes of our models prevents us from directly applying their techniques to search the neighborhood of the feasible solution.

Model Evaluation

We evaluate the proposed models to answer the following problems.

(1) Comparison of performance and the computational demand of $(L1)$ and $(L2)$ models.

(2) Testing of proposed models versus an **a priori benchmark policy** that does not allow second stage dynamic update of the path. Since an a priori solution neglects to capture dynamically updated information, this comparison will allow us to assess the value of such real-time information. This will also allow us to asses the value of transition nodes, and how solution value relates to the number of such nodes in the network.

(3) Analysis of the **Level 3 fully adaptive benchmark model** that dynamically selects the subset of reward nodes to visit, their visiting order, and paths serves as a bound on the best case performance of SAR teams. This model (omitted here in the interest of space) does not have a first stage, and the entire model is composed of a single DP formulation, similar to $(L1)$ and $(L2)$, except that the action space is not restricted to a preselected subset (and order, for $(L1)$) of reward nodes. That is, we do not commit to a set of reward nodes at the beginning.

In practice, implementation of the Level 3 model would require ongoing coordination between SAR teams operating in the area, which is impractical or infeasible in most settings. However, using such a model as a benchmark is of benefit to practitioners considering investing into real-time communication and coordination technology in the hopes of saving more lives.

In the model evaluation and comparisons presented above, we focus on three metrics and the tradeoff between them: **(1)** ease of solution and **computational demand**, **(2)** expected profit collected, which translates to **expected number of lives saved**, and **(3) robustness** of models to data reliability. Our preliminary numerical results confirm that models with a greater level of adaptivity deliver great expected profit. More importantly, we also observe that different models have different subsets and order of reward nodes they choose to visit. More comprehensive numerical studies will be possible with improved solution methods.

5 Future Work

5.1 Exogenous information updates

The AOPST models (see Section 4) capture dynamically updated information about the network and can react to this new information. However, all the updates come from the SAR team actually traversing the network arcs. Our preliminary work establishes that there is a significant benefit to integrating exogenous information sources such as the ones discussed in Section 2. Future work on this project will focus on building new adaptive routing models capturing the information from the data sources identified as most promising as a result of our ongoing work. Unlike ($L1$) and ($L2$) models, routing decisions have little or no effect on the data updates coming from the outside sources; delaying decisions is the predominant (and in some settings the only) technique for "gathering" mode information. As a result, our transportation network explicitly becomes time-dependent, where travel time and reward node locations are functions of time. We will build on our prior path-planning in dynamic environment work [8] to efficiently handle time-dependency of the resulting models.

5.2 Model - data compatibility

Our ongoing analysis of emerging data sources in SAR and other humanitarian operations will be integrated with the developed routing models. In addition, the data characteristics will guide the direction and specifics of these models. More specifically, we will identify how various data streams and relief effort settings (e.g., international versus domestic, levels of access to communication technology) translate to different adaptive models. Data reliability will also play an important role in this evaluation, since various models have different levels of robustness and threshold for accuracy of information.

In the multi-team setting, the decisions to commit to nodes must be considered simultaneously for all teams, changing the structure of the master problem. Future work will explore new formulations and solution approaches for the master problem, building on earlier work for the single team problem. A key question will be whether or not to allow some overlap in assignments. Importantly, one must balance the need for the flexibility associated with assignment overlap to address the stochastic and dynamic elements of the problem with the need for simpler assignment structures that require less coordination among teams in the field. Similar issues of flexibility and consistency are emerging in the multi-period vehicle routing literature, such as [27].

5.3 Systematic analysis of new data sources for humanitarian logistics

Our current modeling work is oriented toward resolving the difficulties commonly faced in disaster response using information that can reasonably be made available to responders in real time. However, each disaster response scenario has its own set of challenges and requirements, as well as its own data environment.

Taxonomy creation. In the next phase of this project, we will develop a systematic taxonomy of new data sources (such as text messaging platforms, social media, satellite and aerial imagery, and participatory mapping) that organizes the data streams according to variables such as reliability, processing difficulty, context-specificity, and role as model inputs. The goals of this work are to **(1)** generate a flexible framework for matching data sources to the particular requirements of the various models described above to match a given crisis scenario; **(2)** form the raw material from which further test cases will be created; and **(3)** prioritize future research and draw links from the research models to the end users who will use them in combination with raw data.

Ongoing extensions of the project will first concentrate on datasets similar to the ones presented in Section 3. Metrics will also be explored and developed that will aid in determining the necessary processing required for model use, supply/demand reliability, and advanced skill set requirements for implementation in field settings. With frequent interactions with humanitarian practitioners and other researchers, the taxonomy will develop and integrate into the broader framework of logistics and information management in disaster settings.

Test case development. We will use the data taxonomy to extend the collection of datasets drawn from real-world events to be used as test cases for the three levels of models introduced in Section 4. These test cases will feature stochastic network connectivity, travel costs, demand, and service time in a dynamic framework in which information becomes available over time from a variety of sources; these datasets will allow model innovations to be assessed in isolation or in combination with each other. We will use these test cases to **(1)** test the performance of solution approaches and assess the relative improvements in model performance (demand served and degree of variance) from using models with progressively more dynamic and stochastic elements; and **(2)** assess the relative contribution of different data sources with varying characteristics (such as update frequency, reliability, and specificity) in improving model performance and making better decisions. The test cases thus serve the dual purpose of assessing both models and data sources. This work will occur alongside data taxonomy creation and feed results in the form of data source analysis back into the taxonomy.

Assessment of data streams. In the later stages of coming work, we will use the analyses from our set of models and test cases to deliver a systematic assessment of real-time data from new and emerging technologies in conjunction with our models. This work will be oriented toward communicating our results to stakeholders and researchers in other fields to cooperatively develop complete workflows that bring models into practice.

5.4 Education

Recent humanitarian crises have demonstrated that ready-made, deployable systems for structuring, managing, and sharing heterogeneous data streams are a critical area of work for disaster preparedness. In order to be usable, our models must be integrated into data workflows before a crisis strikes. To this end, we will offer conclusions regarding the types of data that are most important in improving model performance and that can be realistically integrated into relief operations. We also provide recommendations, workflows, and research products targeted to practitioners.

In Winter 2014, PI Smilowitz will launch a new course in humanitarian logistics in the department of Industrial Engineering and Management Sciences that will incorporate work from this project and establish case studies to be used in the course and made available to other universities.

The Northwestern University Humanitarian and Non-Profit Logistics Initiative website [15] serves as a central point for outreach, dissemination of research results, and sharing datasets and other research products. The site currently houses information about the pilot study discussed in Section 2 as well as generated network datasets.

References

[1] Y. Achbany, F. Fouss, L. Yen, A. Pirotte, and M. Saerens. Managing the exploration/exploitation trade-off in reinforcement learning. Information System Unit, Universite catholique de Louvain, Belgium, Tech. Report. Available at http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.137.9971., 2005.

[2] G. Barbier, R. Zafarani, H. Gao, G. Fung, and H. Liu. Maximizing benefits from crowdsourced data. *Computational and Mathematical Organization Theory*, 18(3):257–279, June 2012.

[3] A. M. Campbell, M. Gendreau, and B. W. Thomas. The orienteering problem with stochastic travel and service times. *Annals of Operations Research*, 186(1):61–81, 2011.

[4] C. Corbane and G. Lemoinesand M. Kauffmann. Relationship between the spatial distribution of SMS messages reporting needs and building damage in 2010 Haiti disaster. *Natural Hazards and Earth System Science*, 12(2):255–265, February 2012.

[5] L. de la Torre, I. S. Dolinskaya, and K. R. Smilowitz. Disaster relief routing: Integrating research and practice. *Socio-Economic Planning Sciences*, 46:88–97, 2012.

[6] B. C. Dean, M. X. Goemans, and J. Vondrak. Approximating the stochastic knapsack problem: The benefit of adaptivity. In *45th Annual IEEE symposium. Foundations Compuundations Computer Science*, pages 208–217, Washington, DC, 2004. IEEE Computer Society.

[7] I. S. Dolinskaya. *Optimal Path Finding in Direction, Location and Time Dependent Environments*. PhD thesis, University of Michigan, 2009.

[8] I. S. Dolinskaya. Optimal path finding in direction, location and time dependent environments. *Naval Research Logistics*, 59(5):325–339, 2012.

[9] I. S. Dolinskaya, M. Kotinis, M. G. Parsons, and R. L. Smith. Optimal short-range routing of vessels in a seaway. *Journal of Ship Research*, 53(3):121–129, September 2009.

[10] D. Feillet, P. Dejaxs, and M. Gendreau. Traveling salesman problems with profits. *Transportation Science*, 39(2):188–205, May 2005.

[11] Google Faculty Research Awards, http://research.google.com/university/relations/research_awards.html.

[12] A. Gupta, R. Krishnaswamy, V. Nagarajan, and R. Ravi. Approximation algorithms for stochastic orienteering. In *Proceedings of the ACM-SIAM Symposium on Discrete Algorithms (SODA)*, 2012.

[13] R. W. Hall. Adaptive least-expected time paths in stochastic, time-varying transportation and data networks. *Transportation Science*, 20(3):182–188, 1986.

[14] Harvard Humanitarian Initiative. *Disaster relief 2.0: The future of information sharing in humanitarian emergencies*. UN Foundation & Vodafone Foundation Technology Partnership, Washington, D.C. and Berkshire, UK, 2011.

[15] Humanitarian and Non-Profit Logistics Initiative. Available at http://hl.mccormick.northwestern.edu, February 2013.

[16] T. Ilhan, S. M. R. Iravani, and M. S. Daskin. TECHNICAL NOTE–The adaptive knapsack problem with stochastic rewards. *Operations Research*, 59(1):242–248, December 2010.

[17] T. Ilhan, S. M. R.Iravani, and M. S. Daskin. The orienteering problem with stochastic profits. *IIE Transactions*, 40(4):406–421, February 2008.

[18] P. Jaillet. A priori solution of a traveling salesman problem in which a random subset of the customers are visited. *Operations Research*, 36(6):929–936, 1988.

[19] E. Miller-Hooks. Adaptive least-expected time paths in stochastic, time-varying transportation and data networks. *Networks*, 37(1):35–52, January 2001.

[20] E. Miller-Hooks and B. Yang. Impact of travel time models on quality of real-time routing instructions. *Transportation Research Record*, 1857(1):21–29, January 2003.

[21] N. Morrow, N. Mock, A. Papendieck, and N. Kocmich. *Independent Evaluation of the Ushahidi Haiti Project*. Development Information Systems International, 2011.

[22] R. Munro. Crowdsourcing and the crisis-affected community: Lessons learned and looking forward from Mission 4636. 2012.

[23] R. Munro, L. Gunasekara, S. Nevins, L. Polepeddi, and E. Rosen. Tracking epidemics with natural language processing and crowdsourcing. In *Spring Symposium for Association for the Advancement of Artificial Intelligence (AAAI)*, Stanford, California, 2012.

[24] R. Munro and C. D. Manning. Short message communications : users, topics, and in-language processing. In *Proceedings of the 2nd ACM Symposium on Computing for Development*, Altanta, Georgia, 2012.

[25] W. B. Powell. *Approximate Dynamic Programming: Solving the Curses of Dimensionality (Wiley Series in Probability and Statistics).* Wiley-Interscience, Hoboken, New Jersey, first edition, 2007.

[26] N. Secomandi and F. Margot. Reoptimization approaches for the vehicle-routing problem with stochastic demands. *Operations Research*, 57(1):214–230, January 2009.

[27] I. Sungur, T. Ren, F. Ordonez, M. Dessouky, and H. Zhong. A model and algorithm for the courier delivery problem with uncertainty. *Transportation Science*, 44:193–205, 2010.

[28] H. Tang and E. Miller-Hooks. Algorithms for a stochastic selective travelling salesperson problem. *Journal of the Operatinal Research Society*, 56(4):439–452, 2005.

[29] UN Global Pulse. Big data for development : Challenges & opportunities. Technical Report May, United Nations Global Pulse, 2012.

[30] P. Vansteenwegen, W. Souffriau, and D. V. Oudheusden. The orienteering problem: A survey. *European Journal of Operational Research*, 209(1):1–10, February 2011.

[31] S. E. Vieweg. *Situational Awareness in Mass Emergency : A Behavioral and Linguistic Analysis of Microblogged Communications.* Ph.d., University of Colorado, 2012.

[32] G. Yen, F. Yang, T. Hickey, and M. Goldstein. Coordination of exploration and exploitation in a dynamic environment. In *Proc of International Joint Conference on Neural Networks*, volume 2, pages 1014–1018, 2001.