Integration of Real Time Data in Urban Search and Rescue Technical Report *

Alex Huang¹, Alex Ma¹, Sara Schmidt¹, Nancy Xu¹, Brandon Zhang¹, Lewis Meineke¹, Zhenyu (Edwin) Shi¹, Jennifer Chan², and Irina Dolinskaya¹

¹Department of Industrial Engineering and Management Sciences, Northwestern University ²Department of Emergency Medicine, Northwestern University

May 31, 2013

Abstract

In this pilot project, we demonstrate a preliminary workflow for incorporating real time, "crowdsourced" data into disaster relief routing, using search and rescue (SAR) operations in the aftermath of the 2010 Haiti earthquake as a test setting. We develop two simple models of relief routing: one with real time data and one without. We use road and building data from OpenStreetMap, a collaborative mapping project that assumed a prominent role in the postearthquake data dissemination process, to define road networks for routing analysis. The real time model incorporates daily updates to this road network contributed by volunteers around the world. For the real time model, demand is predicted by filtering text messages sent to the standardized shortcode 4636 for actionable information on collapsed structures and trapped persons. The non-real time model predicts demand using historical demand indicators such as building footprints.

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 post-disaster 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 prior research with recent

^{*}This work was supported by the Center for Commercialization of Innovative Transportation Technology, a Google Faculty Research Award, and the Walter P. Murphy Society.

developments in the use of new technologies in disaster relief to assess the viability of real-time data for urban search and rescue (USAR) 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.

Humanitarian logistics since the Haitian earthquake

This work situates itself according to an increasingly organized humanitarian logistics research agenda that has emerged in the last few years, much of which uses the Haitian earthquake as its major framing device. This event is one of the most-studied disasters in recent years in part because it displays extreme forms of many of the characteristics separating humanitarian logistics from commercial logistics; some authors consider it a prime example of a "catastrophic" disaster, the distinction between catastrophic and non-catastrophic disasters being based in part on the inability of the impacted society to respond to catastrophe on its own [2, 10]. One of the primary reasons for using the insufficiency of local resources as a defining characteristic of catastrophe is the vastly increased difficulty of obtaining information when outside organizations must become involved in crisis response. In Haiti, this information gap meant that international response agencies faced especially daunting challenges in navigating the area and understanding which areas to prioritize for assistance.

While great progress has been made in recent years in the creation of technologies for sharing useful information in the aftermath of disasters, it is still unclear which sources of information will be most useful for organized crisis response. Large response agencies operate according to established protocols and best practices, and it is often perceived as both risky and costly to change these protocols to accommodate information streams of uncertain value. There is a growing consensus that new data sources offer extremely valuable information whose reliability compares favorably to established data sources if they are understood and used appropriately [7, 20], but developing frameworks for appropriate use in real-world settings is an ongoing project.

This work is one contribution toward the research agenda outlined in [4], which reviews the realworld problems related to vehicle routing for delivery of goods and services in disaster-affected regions, analyzes the representation of these problems in current operations research models, and identifies 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. There is a distinct need 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 [1, 19].

In this report, rather than focusing on the development of these models, we describe how data sources of the type that appeared in the aftermath of the Haitian earthquake could be incorporated into future models designed to accommodate them, using a simplified preliminary routing algorithm as a starting point. Section 2 introduces our data sources and processing workflows. Section 3 details two parallel solution approaches for using these data sources to generate USAR routes, one based in a traditional static framework and one that incorporates real time information. Section 4 discusses some results of the exercise and charts a course for future extensions.

2 Innovation: Integrating new data streams 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.

2.1 Study setting

We take as our motivating case a single urban search and rescue team planning a route through the Port-au-Prince urban area during a five-day-long period starting five days after the earthquake. The scale of need over a wide area, the state of confusion, and the condition of local infrastructure would probably have rendered it impractical to programmatically design search and rescue team routes in response to specific requests in the immediate aftermath of the earthquake even if the information processing framework had existed [10], but rescue operations were still needed long after the state of information improved and organized international response was fully underway. Although the first two days after an event are critical, survival under rubble frequently occurs well past the two-day mark [13]. In Haiti, text messages continued to arrive up to two weeks after the event with information on locations of trapped survivors, sent by witnesses, family members, and even the survivors themselves.

While we use USAR as a study scenario, our model of USAR operations is highly stylized; the emphasis here is on the approach to new data sources as a sufficient support for optimization, rather than on the current optimization algorithm's direct applicability to the search and rescue setting. This closely focused emphasis is in response to an identified need for adaptation to new data sources and analytic techniques in the humanitarian sector [20], and in particular to the need to more deeply integrate these data sources into decision-making structures.

2.2 Data sources

Although a variety of data sources, such as aerial imagery and social media services, played significant roles in the aftermath of the earthquake and hold the potential to be integrated into routing workflows during future crises [22], we restrict the present focus to two specific, 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 [16, 8]. This workflow produced information that could be shared with response teams in near-real time–the average lapsed time from receipt of a message to translation, categorization, and submission to the data consumption stream was ten minutes, and over 80,000 messages were processed in total [14, 16].

This report is not the first to consider methods of estimating demand for first response from geolocated 4636 messages. [3] compares the spatial distribution of 4636 messages geolocated by Ushahidi Haiti to a point dataset of damaged buildings generated by volunteers assessing high-resolution aerial imagery over a two-month period following the earthquake [5]. These two spatial point patterns are highly correlated, indicating that quickly-arriving text messages can be used as an early indicator of the spatial distribution of damage, although it remains to be seen whether this relationship is generalizable to other catastrophic disaster settings. A comparison of the aerial damage assessment with a ground survey conducted by experts suggests the imperfect reliability of the former, even for marking highly damaged buildings [12], so more work must be done to understand the relationship between both of these measures and ground truth.

The procedure detailed in this report is considered to be a complement to the type of analysis described in [3]. The essential components of our procedure occur "downstream" of (that is, farther along in the workflow connecting data collection to organizational implementation than) surface estimation. Our demand estimation involves filtering and processing information in the text messages, rather than using the spatial distribution of all geocoded messages to estimate a surface; while it would be possible to combine these techniques, a mechanism would have to be designed for doing so.

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 supplemented or 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.

The OSM network in Port-au-Prince did not become sufficiently detailed to permit reliable routing until three to four days after the earthquake. As explained above, we restrict the analysis to the period beginning 1/17/2010, the fifth day following the earthquake. By this time, the OSM network had become dense enough for almost all georeferenced text messages in Port-au-Prince to be attached to a nearby network node with connectivity to the rest of the street network. It should be noted that the algorithm employed below would fail if demand nodes were mapped to network nodes without connectivity to the rest of the network.

While the software environment for routing on OSM networks has improved greatly in recent years, many organizations, small and large, use standard commercial routing sources such as Google Maps for basic routing needs. There are projects designed to combine the strong geographic coverage and data integrity of Google Maps with rapidly updated information on road outages [18]. Unfortunately, these commercial sources cannot fully address the routing needs of relief agencies. Their data structures are not designed to be updated very rapidly with crowdsourced information on road conditions except in highly specific proprietary systems such as smartphone-based traffic trackers; although this is changing with the increasing presence of Google Map Maker, the data contributed to the system can



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

only be accessed through Google interfaces. Furthermore, there are generally sharp restrictions on using results from commercial routing engines, and the routing engine itself is a black box; it is not possible, for example, to design a custom travel cost variable. For these and other reasons, we focus on OSM as the network data source of choice.

2.3 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, by a few days afterward, OSM 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 postearthquake 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 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. Figure 3 and Table 1 show a portion of the building dataset in map and tabular form.

2.4 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



Figure 2: Additional data processing workflow for real time model



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

Table 1: Portion of building table

ID	NAME	TYPE	AREA
115901	OAVCT Petionville	public building	0.13644
116161	Ecole Sainte Louise de Marilac	school	0.20323
117571	Port-au-Prince Hospital Bernard Mevs	hospital	0.89818
137063	Ecole Francklin	Bois	1.24675
138129	hopital cubano-haitiano	tent	1.39718
139379	OAVCT	public building	0.25438
139381	Sean hotel	hotel	0.20306
139408	Complex Oasis	hotel	1.25601
139503	Orphelinet Croix Glorieuse	orphanage	0.23783
139887	avsi	hospital	5.99784
140095	DGI annex	public building	0.18737
140139	Ecole Communautaire de Croix des Missions	school	0.09867

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.

Automated processing of text streams using natural language processing for tasks such as message categorization and full-text geocoding is a major area of current research, and there is a growing body of literature applying these techniques to disaster data [11, 6, 21, 17]. We instead rely on human-powered information processing. Even though our entire population of messages was interpreted manually, we still lose useful information by relying on this interpretation exclusively, since the categorization employed has been found to have been unreliable. This unreliability largely came in the form of categorizing general distress calls or other vague messages with too much specificity [15, 9].

2.5 Road processing

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. These datasets are then converted into Esri geodatabase format using the ArcGIS Editor for OpenStreetMap.

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 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.

While this project implemented the road processing component of the workflow in ArcGIS via Python, it is also possible to do the same work using open source tools, including various combinations of PostgreSQL, pgRouting, osm2po, and the Python library networkx. Implementing the full procedure, combining data processing and our custom algorithm in a fully open source framework, is the subject of current work.

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 4. 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 5).

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.

3.1 Optimization model

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.



Figure 4: Visual representation of the solution procedure

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 \boldsymbol{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				

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:

4

$$\sum_{j} \sum_{i} X_{ij} c_{ij} + Y_i S_i \le T_{max} \tag{2}$$

$$\sum_{j} X_{ij} = \sum_{j} X_{ji} \tag{3}$$

$$\sum_{j} X_{ij} = Y_i \tag{4}$$

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

• 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 NSet 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

3.2 Implementation

To address these needs, we develop a heuristic in MATLAB. Figure 4 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



Figure 5: An example of a recommended route. Map source: OpenStreetMap.

as a subroutine. Data preprocessing is accomplished with a variety of tools, including several small 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.

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 2. 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 6–9.

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



Figure 6: Simulation 1

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 the two models with time.

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

Table 2: Four sets of simulations

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



Figure 7: Simulation 2



Figure 8: Simulation 3



Figure 9: Simulation 4

sending it to end users (that is, into a routing algorithm such as this one) is a critical element of any usable tool.

- We also do not address the potential for bias in crowdsourced data. The types of bias introduced by crowdsourced data and the effect of any bias on the validity of model results is its own important research topic.
- 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.
- 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.

References

- A. M. Anaya-Arenas, A. Ruiz, and J. Renaud. Relief distribution networks : A systematic review. Technical report, Interuniversity Research Centre on Enterprise Networks, Logistics and Transportation, 2012.
- [2] J. Barnshaw, L. Letukas, and E. Quarantelli. The characteristics of catastrophes and their social evolution: An exploratory analysis of implications for crisis policies and emergency management procedures. 2008.
- [3] C. Corbane, G. Lemoine, and 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, Feb. 2012.
- [4] L. E. de la Torre, I. S. Dolinskaya, and K. R. Smilowitz. Disaster relief routing: Integrating research and practice. *Socio-Economic Planning Sciences*, 46(1):88–97, Mar. 2012.
- [5] R. T. Eguchi, S. P. Gill, S. Ghosh, W. Svekla, B. J. Adams, G. Evans, J. Toro, K. Saito, and R. Spence. The january 12, 2010 Haiti earthquake: A comprehensive damage assessment using very high resolution areal imagery. In 8th International Workshop on Remote Sensing for Disaster Management, Tokyo Institute of Technology, pages 1–8, Tokyo, Japan, 2010.
- [6] J. Gelernter and N. Mushegian. Geo-parsing messages from microtext. Transactions in GIS, 15(6):753-773, Dec. 2011.
- [7] Harvard Humanitarian Initiative. Disaster relief 2.0: The future of information sharing in humanitarian emergencies. Technical report, UN Foundation & Vodafone Foundation Technology Partnership, Washington, D.C. and Berkshire, UK, 2011.
- [8] J. Heinzelman and C. Waters. Crowdsourcing crisis information in disaster-affected Haiti. Technical report, United States Institute of Peace Special Report, 2010.
- [9] V. Hester, A. Shaw, and L. Biewald. Scalable crisis relief: Crowdsourced SMS translation and categorization with Mission 4636. In *Proceedings of the First ACM Symposium on Computing* for Development, London, 2010.
- [10] J. Holguín-Veras, M. Jaller, L. N. Van Wassenhove, N. Pérez, and T. Wachtendorf. On the unique features of post-disaster humanitarian logistics. *Journal of Operations Management*, 30(7-8):494– 506, Nov. 2012.
- [11] M. Imran, S. Elbassuoni, C. Castillo, F. Diaz, and P. Meier. Extracting information nuggets from disaster-related messages in social media. In T. Comes, F. Fiedrich, S. Fortier, J. Geldermann, and L. Yang, editors, *Proceedings of the 10th International ISCRAM Conference*, number May, Baden-Baden, Germany, 2013.
- [12] N. Kerle. Remote sensing based post-disaster damage mapping with collaborative methods. In S. Zlatanova, R. Peters, A. Dilo, and H. Scholten, editors, *Intelligent Systems for Crisis Management: Lecture Notes in Geoinformation and Cartography*, Lecture Notes in Geoinformation and Cartography, pages 121–133. Springer, Berlin, Heidelberg, 2013.

- [13] A. G. Macintyre, J. A. Barbera, and E. R. Smith. Surviving collapsed structure entrapment after earthquakes: a "time-to-rescue" analysis. *Prehospital and Disaster Medicine*, 21(1):4–17, 2006.
- [14] P. Meier and R. Munro. The unprecedented role of SMS in disaster response: Learning from Haiti. SAIS Review, 30(2):91–103, 2010.
- [15] N. Morrow, N. Mock, A. Papendieck, and N. Kocmich. Independent evaluation of the ushahidi Haiti project. Technical report, Development Information Systems International, 2011.
- [16] R. Munro. Crowdsourcing and the crisis-affected community: Lessons learned and looking forward from mission 4636. 2012.
- [17] 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.
- [18] S. Nedkov and S. Zlatanova. Google maps for crowdsourced emergency routing. In *The XXII Congress of the International Society for Photogrammetry and Remote Sensing*, volume XXXIX, Melbourne, Australia, 2012.
- [19] P. Roy, C. Brewster, and P. Albores. Logistical framework for last mile relief distribution in humanitarian supply chains: Considerations from the field. In *Proceedings of the International Conference on Manufacturing Research 2012*, Birmingham, UK, 2012.
- [20] United Nations Office for the Coordination of Humanitarian Affairs. Humanitarianism in the network age. Technical report, United Nations, 2013.
- [21] S. Verma, S. Vieweg, W. J. Corvey, L. Palen, J. H. Martin, M. Palmer, A. Schram, and K. M. Anderson. Natural language processing to the rescue? Extracting "situational awareness" tweets during mass emergency. In *Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media*, pages 385–392, Barcelona, 2011. Association for the Advancement of Artificial Intelligence.
- [22] M. Zook, M. Graham, T. Shelton, and S. Gorman. Volunteered geographic information and crowdsourcing disaster relief: A case study of the Haitian earthquake. World Medical & Health Policy, 2(2):6–32, Jan. 2010.