Social Media Alert and Response to Threats to Citizens (SMART-C)

(Invited Paper)

Nabil Adam
US Department of Homeland Security
Science & Technology Directorate
Washington, DC, USA

Jayan Eledath*, Sharad Mehrotra+, Nalini Venkatasubramanian+
* SRI International, Princeton, NJ
+ University of California, Irvine

***Abstract*— Social media, such as blogs, Twitter, and information portals, have emerged as the dominant communication mechanism of society. Exploiting such input to gain awareness of an incident is a critical direction for research in effective emergency management. In this paper we present an overview of the SMART-C system, which is part of the social media initiative at the Department of Homeland Security. The system aims to enable robust bidirectional communication between emergency management and the public at large throughout the disaster life-cycle via a multitude of devices and modalities including cell phones, MMS messages, text messages, blogs, Twitter, etc. A discussion of the major components of SMART-C and related research challenges is included. These components include mechanisms to model event level semantic information, a platform for implementing multi-sensor fusion, mechanisms for estimating the veracity of information, data cleaning to reduce uncertainty and enhance accuracy of event detection and notification, and spatiotemporal analyses for pattern and trend analyses for higher level observations.**

Keywords - social media; emergency management; alerting; robust data analytics

I. INTRODUCTION

Social media, such as blogs, Twitter, and information portals, have emerged as the dominant communication mechanism of today's society. In the context of emergency management, exploiting such input to gain awareness of an incident is a critical direction for research. Dynamic real-time incident information collected from on-site human responders about the extent of damage, the evolution of the event, the needs of the community and the present ability of the responders to deal with the situation combined with information from the larger community could lead to more accurate and real time situational awareness that allows informed decisions, better resource allocation and thus a better response and outcome to the total crisis. Component technologies such as mobile devices, ubiquitous networking, location technologies, etc., which are integral components of a system that supports robust communication between response agencies and the public at large, have significantly matured to make such a system a reality. Social media and mobile devices are also changing how communities at large seek information in disasters. Reaching populations with customized and timely alerts through multiple channels (traditional media, Internet technologies, mobile services and social networks) can help inform those at risk and assure those not at risk with messages that accurately reflect the levels of vulnerability of the target population.

Many past and current projects (e.g., [11]), as well as several studies conducted by the National Academies [14, 19] have indeed established the need for a robust, seamless, and scalable bidirectional communication between the response agencies and the public at large. A system that provides such a capability should support robust and scalable capabilities for real-time collection and processing of incident-related data from citizens using diverse media. Information obtained from social media and human input may be structured (e.g., interactive websites, speech and text dialog systems) or unstructured (e.g., textual reports, voice, video inputs, MMS messages). Components of such a system include mechanisms to model event-level semantic information, a platform for implementing multi-sensor fusion, mechanisms for estimating the veracity of information, data cleaning to reduce uncertainty and enhance accuracy of event detection and notification, and spatio-temporal analyses for pattern and trend analyses for higher level observations.

In this paper we present such a system, Social Media Alert and Response to Threats to Citizens (SMART-C), which is part of the social media initiative at DHS [33]. SMART-C supports an extensible plug-n-play architecture into which new mechanisms, techniques, modalities, and systems can be further incorporated to both extend the set of supported functionalities and/or to leverage the functionalities provided in the context of specific systems. SMART-C's capabilities include:

- Enrichment of incident information, specifically, semantic enrichment through multi-modal analysis (text, speech, video) to create event level representation
- Integration with other data sources such as demographics, socio-economic, environmental data.

COLLABORATECOM 2012, October 14-17, Pittsburgh, United States

DOI 10.4108/icst.collaboratecom.2012.250713

- Enhanced response planning capabilities by utilizing enriched incident information and appropriate resource databases
- Allowing addition and/or request updates from various sensors (cell phones, surveillance cameras, etc.) in real-time for improved situational awareness.
- Querying targeted sensors for recent updates based on their location and capabilities.
- Targeted monitoring of social media feeds for event relevant information.
- Generation and customization of alerts for specific population groups based on contextual information such as location, physical disabilities, language impairments, and socio-economic factors.

The rest of this paper describes these capabilities and the research challenges to be addressed to achieve them.

II. SMART-C VISION AND GOALS

A. *Vision*

The vision of SMART-C is to design a flexible platform to support two-way responder/citizen collaboration that uses social media to enhance communications before, during, and after an emergency. The framework is envisioned for use throughout the disaster lifecycle, including warnings in anticipation of the event, alerting during the event, and post-event analysis and recommendations. This will be achieved through the integration of multiple data ingestion, enrichment and alerting technologies, through a design that is modular, scalable, and able to support both current and future needs.

SMART-C will support robust and scalable capabilities for real-time collection and processing of incident-related data (reports of conditions and locations, updates and assistance requests) from social media. Such data may be structured (e.g., interactive websites, speech, and text dialog systems) or unstructured (e.g., textual reports, voice, video inputs, MMS messages). The integrated SMART-C platform will include mechanisms to model event level semantic information, a platform for implementing multi-sensor fusion, mechanisms for estimating the veracity of information, data cleaning/de-noising to reduce uncertainty, and data enrichment utilizing both metadata as well as shallow content analysis.

The framework will also include a client sub-system through which alerts are generated, customized, and disseminated to citizens during disasters. One implementation of this application will operate on smartphone/tablet devices and desktops that can be used by first responders.

B. *Goals*

SMART-C supports several modules and capabilities in a flexible manner such that as new capabilities and technologies arise, they can be more easily incorporated into the system. The overall goal is to first develop a baseline system that can be fielded in a few select locations across the country; implement solutions to new requirements derived from these field experiments and then roll the system out to multiple locations. SMART-C has the following features:

- Multi-modal Data Ingestion: ingest rich, multi-modal information obtained from many different types of sources.
- Full Disaster Lifecycle Alerts: produces alert generation and delivery for full disaster lifecycle (e.g., pre-incident warnings, precautions/advisories, and post-incident actions for seeking assistance and disaster mitigation)
- Customization: alerts will be customized for specific population groups based on contextual information such as current location, physical disabilities, language impairments, etc. The system will support interfaces to import contextual information from diverse information sources/feeds. These could be GIS data obtained/collected as a prior operation and/or dynamically acquired contextual information about the recipient during real-time alerting (e.g., specific conditions/status of recipient).
- Analytical Capabilities: the underlying technologies provide enrichment of incident information through multimodal event recognition; enhanced response planning using enriched incident information and resource databases; and geo-location of incoming multimodal data. Each stage of processing/analytics will extract increasingly rich and high-level semantic information from the data, beginning with raw data parsing, continuing with semantic extraction of individual events, and culminating in higher-level semantics with grouping and removal of redundancy.
- Manual Override: supports manual override (Responders and emergency management officials decide the severity of the alert and specify the precautionary measures). Manual override offers opportunity to validate the reliability of the alerts prior to delivery.
- Plug-n-Play Technology: supports the ability to add new information sources (e.g., Twitter feeds, citizen-provided data through text/multimedia messaging and voice calls), add new recipients, add new analytics, or delivery/dissemination mechanisms. This key feature will allow the system to be enhanced as new research and technologies are developed.
- Scalability and Robustness: supports graceful degradation to failures and scalability to large numbers of users and diverse customizations.
- Reliability: integrates algorithmic techniques to determine reliability of alerts based on reliability of information sources and through manual override.

III. SMART-C ARCHITECTURE

The SMART-C architecture (Figure 1) is composed of servers - on which data are ingested, cleaned and analyzed to extract information that is customizable for first responders

and citizens – and client devices on which alerts are visualized.

dictate the kind of information to be provided in an alert are incorporated into a rule-base. This is used to generate

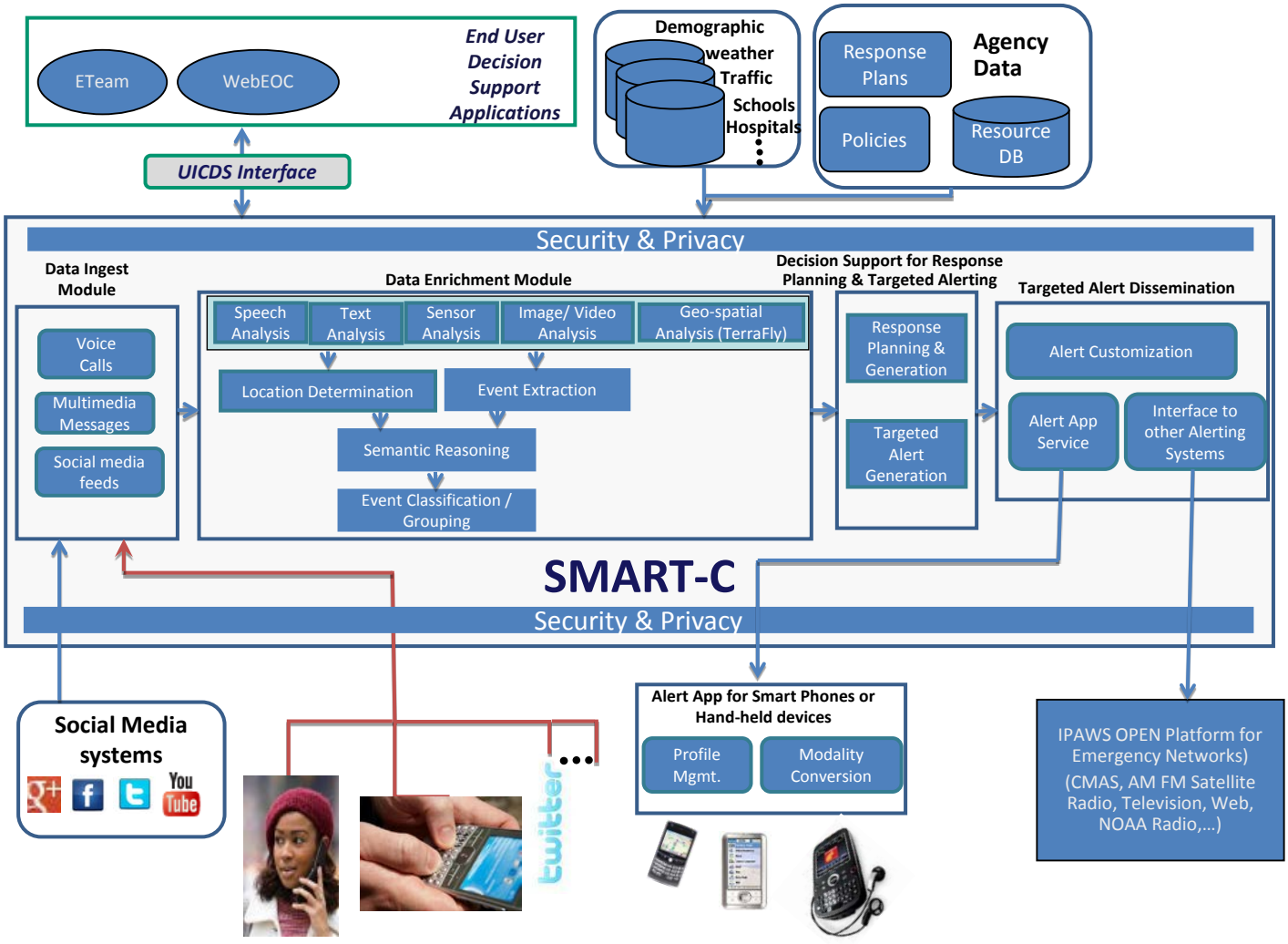

Figure 1. Figure 1: SMART-C consists of servers for data ingestion and enrichment, and alerting tasks, and client devices that visualize results and alerts.

The data ingest module handles citizen input from multi-modal sources (voice, text, image, video). It pre-processes the received data based on its modality. For example, voice calls are transcribed and stored to support data sharing with incident management applications. The data enrichment module further analyzes the pre-processed data to extract semantic information. More specifically, this semantic enrichment is performed in the context of location determination and event representation. In addition, the semantic information extracted from multiple users' data is fused for event classification and disambiguation. Additional details are provided in Section IV B. This enriched event data is then processed to generate more targeted alerting by integrating the semi-structured event representation with other relevant local data such as demographics, socio-economics, weather, environmental and geo-spatial characteristics, and resource availability in different areas and jurisdictions. The GIS-enhanced information can enable decision makers (e.g., alert policy designers) to match the on-ground situation (using citizen inputs) and consequently determine alerting requirements in different areas. Policies that messages for specific population groups that are categorized based on location, physical disabilities, language impairments, and socio-economics (e.g., people who do not own a car and rely on public transportation). For example, in the case of an industrial fire, the alert sent to people living closer to the industrial facilities storing hazardous chemicals would be different from the alert message sent to people living at a safe distance from those facilities.

This architecture is designed to address two key challenges: (1) while there's a large amount of information available from these new media sources, it is often less resilient during disasters - cellular networks, for example, are subject to overload or damage, limiting their efficacy in severely impacted regions; (2) the sheer size of the incoming data requires solutions that can sift through and handle large-scale data, and extract relevant, meaningful and grouped information. SMART-C addresses these challenges by being modular, processing multiple sources of information, supporting multiple alerting mechanisms and leveraging technology advances in

cloud computing, distributed processing, text analytics, computer vision, machine learning and web services.

IV. SMART-C COMPONENTS

Each of the three components identified in Section III – data ingestion, enrichment and customized alerting – are now described in more detail.

A. Data Collection

Information useful for creating situational awareness in emergency response situations is available from multiple sources and in diverse formats. Data collection architecture must be able to support capture of data from heterogeneous information sources. Furthermore, it must seamlessly scale to large-scale emergencies such as hurricanes /earthquakes, and continue to be operational in operating conditions during emergencies where failures and infrastructure disruption are a norm, and information may be missing and/or partially available. We discuss three challenges in design of a robust data collection architecture for emergency conditions we are addressing in SMART-C.

Scalability: In geographically spread disasters, such as earthquakes, the number of information sources and the resulting data generated may be very voluminous. Real-time processing and analysis of such data could overwhelm any computing and communication infrastructure. While dynamically acquired cloud resources alleviate some of the overheads (e.g., by parallelizing the analysis), scaling data collection to such potentially large data requires additional techniques. In particular, Smart-C supports mechanisms to dynamically adapt spatial and temporal granularity of data acquisition in order to optimize the overall situational awareness under resource constraints (e.g., network bandwidth) and source restrictions (e.g., maximum allowed queries to web site for a given unit of time). Such mechanisms enable prioritization of the collection of diverse data most useful for analysis and event modeling.

Heterogeneity: The data collection component in SMART-C enables ingest of information from multiple sources that generate data at different levels of semantic abstraction. To support a flexible and effective data collection from such disparate sources, Smart-C supports: (a) Structured approaches for modeling the information sources and the data obtained from them; (b) Languages and abstractions to specify the information needs of the specific response related applications. (c) Techniques to translate sensing information needs into specific sensing tasks that obtain data from the underlying sensors at appropriate quality levels. (d) Strategies to ensure resilient sensor data collection despite failures in the sensing and communication infrastructure and limitations in the sensor data processing mechanisms. In Smart-C, programming abstractions hide application programmers from having to deal with low-level sensor specifics – applications specify their higher-level data needs, which are then automatically translated into lower-level data acquisition plans. Furthermore, the underlying runtime is designed to seamlessly overcome errors and failures to the extent possible.

Relevance – One of the goals in data acquisition is ensuring that data collected is relevant to the event under consideration. For example, an event hurricane may lead to acquiring tweets for related concepts. In Smart-C the design of the acquisition framework addresses four interrelated aspects of the problem: precision, recall, ranking, and clustering. Twitter can be viewed as a large stream of messages, only few of which are actually related to the event of interest. The precision aspect of the problem is to be able to accurately retrieve these relevant messages, avoiding retrieving too many irrelevant ones such that the analyst is not overwhelmed with a flood of irrelevant information. In turn, twitter messages might describe relevant sub-events using keywords and terms that are not present in the description of the event of interest. Hence, the retrieval challenge of the problem is to be able to get as many of the relevant messages as possible. In general, precision and recall are known to form a trade-off, as it is typically easy to achieve high precision but low recall and vice versa - therefore the challenge is to try to maximize both of them at the same time. The retrieved messages might describe just a few related sub-events of a larger event. Thus it could be useful to cluster the messages (perhaps according to multiple analyst-specified criteria) such that each cluster corresponds to one sub-event. Finally, the analyst might want to see clusters and tweets inside clusters be ranked such that the most important information is presented first. Furthermore, all such information should be collected and presented to the analyst in a real-time fashion.

To address the precision challenge, context-based filtering techniques could be applied that would remove messages that are unlikely to be related based on the context of the event and the message [8]. A simple example is filtering out a message with known GPS location that is far from the event of interest created by the user who is not friend/follower of any other poster on the event. This is since it is likely to be just an opinion of that user on the event -- at best, rather than real first hand observation from a person on the ground. To improve recall, boosting methods [3] can be used that collect initial set of message related to the event, and then from the set choose new discovered frequent/ important additional keywords as query terms. Such query terms could also be chosen using ontologies such as SWEET. To rank clusters features such as messages size of clusters, reputation of people whose messages ended-up in the cluster, how many times messages have been retweeted, etc are used. Furthermore, Smart-C is exploring the role of topic modeling to detect subevents in the set of relevant messages and cluster them.

Resilience to communication failures: End devices, the communication medium, and the application context are subject to constant changes or failures in dynamic environments. For example, communication infrastructure may be damaged, devices can be turned on and off, users may move from one place to another and lose connectivity, networks may be congested and packets are dropped. Our prior experience in resilient sensing in disasters (real world exercises in the context of SAFIRE, RESCUE Extreme Networking[34,35]) has indicated that communication failures will occur and are often aggravated by the presence of hazards leading to loss of fixed infrastructures such as the Internet. Existing network delivery mechanisms are also sensitive to interference, noise and this can result in degradation of information quality even when

infrastructure may be available. Mobility is an additional factor and the ability to perform resilient data collection from users on the move is challenging. The challenge is then to design end-to-end data collection systems that can adapt to the unreliability of the network and communication infrastructure. Adaptation to failures and surge capacity demands are hard to realize using approaches at lower levels of the networking infrastructure (e.g. medium access and network layers of the ISO stack). More recent efforts aim to use middleware driven approaches to manage communication by exploiting and merging multiple networking technologies such as mesh network deployments, delay tolerant networks, and mobile adhoc networks [36,37,38,39]. The ability to create spontaneous networks where nodes have the ability to make communication decisions locally, using available knowledge of network status and taking into account information needs and tolerance parameters (timing, accuracy, reliability) is critical to enabling communication resilience. What is new and challenging is using these ideas to manage communication over diverse network technologies, leveraging the components' network capabilities seamlessly in a quality sensitive manner. Examples of techniques to enhance communication reliability include (a) exploiting multiple access technologies to form connected networks where networks are partially damaged or (b) techniques to exploit node mobility to ferry data to points from where they can be reliably uploaded when there is more significant network damage/loss.

B. Data Enrichment

SMART-C is designed to ingest complex, multi-modal data from various sources. As such, there is a need for algorithms that are capable of analyzing the ingested data and extracting meaningful semantic information from it. With appropriate extraction, such modalities can provide a very rich source of situational information. While some existing systems support citizen input, SMART-C is designed to handle multi-modal data at a semantically richer level extracting meaningful situation information for alerting in real-time. Realization of such a system exploits advances in multimodal data processing including mechanisms for speech recognition (E.g., SRI's DynaSpeak and Decipher systemc) and natural language understanding for analysis of audio and textual inputs, as well as, image and video processing mechanisms to extract event-level information.

An example of such mechanisms embedded into SMART-C is our recent work in [24] that maps low-level features aggregated into Bag-of-Words (BoW) representations compiled over the length of the video clip. The approach used a Support Vector Machine (SVM) as the basic classifier for learning event models and used intersection kernels for all BoW features. The output of each classifier is the probability of event detection for each video clip. In [24] we showed that the proposed method performs well on the public TRECVID MED11 dataset, despite the complexity of the task and challenging nature of the data. We also showed that simple feature-fusion strategies (such as products of individual feature probabilities) outperform more sophisticated ones.

Another example analysis technology supported in SMART-C is the geo-localization of data from ad-hoc sensors like cell phones and social media sources, which is crucial for providing first responders with situational awareness about an incident. This can be done at several levels of complexity, including usage of direct meta-data, analysis of speech data for landmarks or natural language directions, or complex computer vision algorithms that can localize images based on skylines or building facades. Having information about where the reported event occurred allows for various geospatial analyses that can support alert customization to subscribers. For example, TerraFly-based database queries regarding healthcare facilities, critical infrastructure, and demographic information can be used to determine who to alert and the information content that an alert should have.

In addition to supporting multimodal analysis techniques for multimodal event extraction, SMART-C supports a wealth of techniques that systematically exploits domain semantics in the form of disaster event ontologies, relationships embedded in the data, variety of contextual information (e.g., GPS), and spatio-temporal reasoning to enhance the data quality in the context of events extracted from diverse multimdoal information sources. It leverages data cleaning techniques developed as prior work including a domain-independent graphical-based entity resolution framework [10,11]. Such data cleaning techniques are used to reduce and/or mask the inherent uncertainty in automated algorithms for extracting semantic information such as higher-level objects, relationships, events, etc. from multimodal data which, in turn, can improve the reliability of the inference made on the data. We illustrate the usage of data cleaning techniques in the context of supporting situational awareness over data generated using state-of-the-art automated speech recognition systems such as Dynaspeak. In such systems, data is represented and used at different levels – in the form of words transcribed from speech using ASR, as speech segments annotated with intent, incident, emotions etc. and as situational data synthesized from conversations. Given uncertainty associated with automated analysis, data quality challenges arise at all the different levels of information processing. For instance, in speech recognition, such uncertainty arises in the form of N-Best lists or word graphs outputted by ASR systems. In order to generate a deterministic transcription, such systems output the top-1 (most probable answer) that often is wrong (specially when the speech is noisy or emotive) resulting in very high word error rates. In [28], we explored MLE based technique to exploit contextual information (e.g., word co-occurrences, location, etc.) to improve quality of transcription. Given an N-best list, our method outputs a set of words along with associated probabilities conditioned based on available contextual information. Given complexity of the MLE approach, a more efficient heuristic based on branch and bound method that can support real-time speech input was also developed. Our results [27] shows that significant quality improvements can result from using such a technique – 20-25% quality improvement of F1-measure for annotating images using speech. What is more interesting is that, as the noise level in speech increases (and thus the word error rate in the input speech increases), with the carefully designed and tuned strategy the relative improvement increases. Even for noisy data, we obtained F-measure of over 0.5 compared to 0.2 by ASR.

While the above discussion illustrates benefit of exploiting semantics at the word level representation, data cleaning approaches can be expected to be even more effective at higher levels of data representation (i.e., event and situational level). The reason is that at the higher level, additional event level semantics become available that can be used to improve quality. For instance, if the underlying system has uncertainty whether a speech segment mentions the location of an individual being on the 14th floor or the 40th floor (a result of an ASR error percolating all the way through information extraction pipeline), we could eliminate the alternative of 40th floor with very high confidence if we knew that the building in question contains only 25 floors. While the above illustrates a simple instance of exploitation of domain knowledge, in general, we could exploit a variety of semantics in the form of context (e.g., who is the speaker, who is he speaking to, topic of conversation, emotional state, speaker location, time of speech), domain knowledge (rules, constraints), spatial and temporal reasoning, external knowledge (e.g., ontologies, and external information sources), as well as attributes of entities and events and the relationships among the events and entities. Such an exploration, based on our previous work on data cleaning in the context of relational and textual data [30, 31, 32] is currently ongoing. We are further exploring how data quality techniques can be used for semantic interpretation of multimodal including video data. Some initial work in this area can be found in [29].

C. *Customized Alerting*

With the advent of mobile computing and the Internet, it is now possible to collect granular information and provide specific targeted messages – this is the cornerstone of SMART-C. For dissemination of alerts by the emergency response agencies to citizens, the SMART-C system leverages the recent growth in the number of citizens who are equipped with smartphones and/or are connected to the Internet. SMART-C alerting has two components – a user side app on a mobile phone and server side modules that generate and customize alerts for individual subscribers based on current context. The client module is an Alert Service Application that can be downloaded on smartphones using which users can directly receive customized alerts, which can be converted into the appropriate modality based on the user preference. SMART-C will be designed using an open architecture that will allow the integration of other alert dissemination modality platforms such as Integrated Public Alert and Warning System (IPAWS). Specifically, SMART-C will encapsulate a given alert in a CAP message in the Standard EDXL-DE envelope for delivery to the general public. This will enable us to use different channels including Commercial Mobile Alert Service (CMAS) for cellular broadcast, AM/FM broadcast, satellite broadcast, TV broadcast, Web, and NOAA radio service.

On the server side, the alert dissemination module will support customized delivery of alerts to specific sub-populations based on location and other contextual parameters. Based on the current status of the event and its expected propagation, status of various infrastructures, and information about the individuals (e.g., location, connectivity, special needs, etc.), an alerting component will generate and deliver appropriate messages to the affected individuals. First, an audience or channel-mapping step will use prior knowledge about the alerting context to customize the audience set and channels based on dissemination needs and channel accessibility. A predefined, but modifiable policy determines which recipients should receive notification and the type of message to be sent. A flexible policy definition mechanism will allow customizations based on the type of event, the location, or geographical information, leveraging our previous work [10]. The policy-based architecture will also determine whether and how often an individual user may be reached as the event evolves and the required protective actions change. This message will be transmitted, after human verification via existing underlying communication technology infrastructure (e.g., IPAWS).

Enabling reliable alerting in the presence of communications failures: Existing alerting and messaging mechanisms prespecify a communication strategy (e.g.broadcast, multicast, unicast) to reach a given set of recipients using specific communication modalities (e.g. email, SMS-based cellphone alerts). Pervasive networking environments of today consist of multiple heterogeneous wired and wireless networks that co-exist and overlap spatially and/or temporally; that if combined, can potentially render opportunities for enhanced and resilient dissemination performance[40]. Future systems must include the ability to adapt on-the-fly both the dissemination strategy (broadcast/multicast/unicast) and the networks over which content is delivered based on current recipient context (location, device availability) and network context (what networks are deployed and their current availability) and application level information (user priorities/preferences, crisis context).For instance, broadcast might be appropriate to reach a large pool of recipients with stringent time constraints (despite the cost of data delivery to non-desired recipients); multicast may be preferred for dissemination of customized information to a small pool of recipients (despite of the cost of network overlay maintenance). Another key challenge is to develop protocols that can exploit geographical and social correlations in the information dissemination process for more efficient and targeted alerting. In our prior efforts on personalized alerting in disasters we have observed that information needs, societal relationships and evolution of disaster events are all geocorrelated. For example, the set of recipients of a message are often geocorrelated (e.g. the community associated with an elementary school including teachers, staff, parents and students). Societal context is also relevant in adapting the alerting content for broader reach e.g. language translation for non-English speaking communities in a neighborhood. Disasters and their spread are dynamic and often geo-correlated (e.g. tornadoes) – it is as important to warn citizens that the tornado has moved away as it is to warn those who are likely to be impacted. Typical failure models assume uniformly random failure in the overall network[41,42]; further research is required to understand how a continuous series of geographical failures in ongoing events or “moving events” (e.g. tornadoes) can impact dissemination.

V. Related Work

Smart-C embodies a large number of technologies (including smart-phones, cloud computing, social media, data collection, information enrichment, multimodal data analysis, GIS reasoning and alert generation) that work in synchrony to provide a platform for bidirectional communication between response agencies/personnel and the citizens with the objective of creating improved awareness and alerting. Research and development in such technologies is progressing at a rapid pace (in the emergency management domain as well as other domains). We focus our discussion on related work on two such relevant technologies (a) analyzing social media for event awareness, and (b) systems for alert generation.

A. Analyzing Social Media for Situational Awareness

Social media, such as Twitter, contains large amounts of user-contributed content for a wide variety of real world events, concepts and entities. Recent studies have explored diverse ways in which individuals have used Twitter during disasters and identified opportunities such information offers for disaster response. [4] studied statistical differences in ways Twitter was used during different types of mass convergence events -- political conventions versus crisis situations such as Hurricane Ike, and Gustav. In [23], authors analyzed tweets related to Oklahoma grass fires and Red River floods to identify features/aspects of tweets most important with the objective to guide IT researchers on the nature of extractors to build for situational awareness. Their work emphasized the need for robust extractors for location, location references, and situational updates. The ongoing work on Smart-C addresses exactly these challenges. Another related study [1] explored the tweets generated by on the ground medical teams during Haiti disaster. The work established the major role a tweet-based communication system could play in coordinating an effective response by providing an open channel through which responders/ citizens on the ground could request appropriate assistance/resources.

Recent work on analyzing tweets for information has focused on event/entity detection and resolution in tweets, techniques to group/cluster tweets about the same event/ same topic, classify tweets based on relevance to given events, and analyzing tweets for trend detection. Challenges in tweet analysis arise due to poor quality of the tweets, ungrammatical construction of the messages, extensive usage of acronyms, as well as extensive dependence on the appropriate context to properly interpret the tweets. In [4], authors explore various ways to select a set of most important tweets from a collection of tweets about an event that have high textual quality, are related to the associated event, and are useful (viz., are informative) for the event. In [4] authors address the challenge of distinguishing between real-world events and non-event messages and identifying events in real time on Twitter. In [2] they first group events that are topically similar together using an online clustering technique. Four kinds of Twitter message's features help detecting clusters that are associated with events. Temporal features are mainly about the volume of messages for an event during its associated time. Social features including retweets and mentions are using for capturing the interaction of users in a cluster's messages. Topical features describe the topical coherence of a cluster, based on a hypothesis that event clusters tend to revolve around a central topic, whereas non-event clusters do not. Also non-event clusters often center on a few terms (e.g. "sleep", "eat") that do not reflect a single theme. Using the above features an event classifier (that classifies tweets into event or non-event) is trained by applying standard machine learning techniques.

The work in [4] explores mechanisms to detect emergent topics in Twitter stream using user-assigned hashtags. A sliding time window (about one hour for Twitter) is used to detect unusual shifts in correlations of tag pairs in a window that is considered to be an indicator of an emergent topic. Another similar system is TwitterMonitor [9], which also detects trends in twitter streams.

While the above work is general and not specific to emergency management domain, in [22] authors have explored applying machine learning techniques (e.g., SVM) to classify tweets during mass disruptions to identify local/on the ground tweets from others, separate between original information and derived information (e.g., in the form of repost, re-tweet, etc.). The classification is done based on meta information such as number of followers, follower growth, initial follower count, friend count, etc. and not does not exploit content features. The authors are able to achieve 70% accuracy in labeling tweets as being generated on the ground or remote

B. Alerting Systems and Technologies

Alerting systems aim to provide information to the public at large specifically to encourage self-protective actions, such as evacuation from endangered areas, sheltering-in-place, and other actions designed to reduce exposure to natural and human-induced threats. Effective information dissemination during crises consists of conveying accurate and timely information to those who are actually at risk (or likely to be), while providing reassuring information to those who are not at risk and therefore do not need to take self-protective action. Alerting systems and technologies have been long studied by social scientists [12, 20, 25, 6, 21, 10] – such studies indicate that the key factors that introduce challenges to effective information dissemination in crisis situations include time constraints (i.e. the amount of warning time), accurate targeting of warning and delivery of personalized alerts that are relevant to an individual's specific context.

For example, warning time can range from no warning (a sudden terrorist bombing) to a minute or less (real-time seismic alerts), to several minutes (tornadoes) to days (hurricanes). Effective warnings provide clear information both on who is at risk—i.e., who must act on the warning—and who is outside the zone of danger. To motivate action, alert information must specify what geographic areas and what populations are at risk, while reassuring those not at risk that they need not take self-protective action. The strategies employed must thus incorporate place-specific data to ensure that the information that is communicated accurately reflects differential levels of vulnerability. The final challenge is in the customization of the delivery process that must accommodate the public's use of multiple communications media and the fact that on any given day the public is highly mobile. To add to this fact, the

infrastructure over which dissemination is conducted may have been destroyed, partially failed and overloaded.

The eventual goal of alerting system design and implementation is to provide timely, valid, and accurate information to those at risk in appropriate formats and communication channels such that the messages elicit appropriate self-protective behavior on the part of recipients. Prior efforts such as the RESCUE and SIGNAL projects (http://www.itr-rescue.org), have developed an understanding of the key factors in effective dissemination to the public. Infrastructures such as the NOAA Weather Radio All Hazards (NWR) and the Emergency Alert System (EAS) [13], along with standards like SAME and CAP can provide the basic mechanisms for the dissemination of the warnings over dedicated radio frequencies and public media. Furthermore early warning systems, like the Local Tsunami Warning [5] and the USGS Shakecast have a similar approach for the dissemination of warnings. Push-based technologies such as reverse 911, SMS based messaging, and auto-dialers provide limited opportunities for targeted messaging based on geographical and user specific context, but have experienced problems with accuracy, scalability and reliability and do not support the customization and prioritization required for effective response. Recent studies on communication behavior in disasters indicate that the public is becoming more proactive in using new technology advances that provide anytime, anywhere connectivity. Several groups in industry and academia are also investigating analysis of social media, such as analysis of information diffusion in social media [17] machine learning to determine high-value data in social media [23] or sentiment analysis [18]. What is lacking, however, is a system utilizing these trends and citizen provided information in the design of an alerting system.

A plethora of mobile applications have been developed (often to address specific events or specific platforms) – imapWeather (for the Iphone platform) and WeatherBugElite are example alerting systems that focus on weather advisories based on reports from the National Weather Service (NWS). Such systems monitor the current location/geography of the subscribers and alert them to severe weather conditions that they may encounter; apps may also allow users to subscribe for multiple locations simultaneously (to monitor safety of friends and family). Weather Bug Elite includes the ability to provide enhanced map based interfaces, extended forecasts and other user level features (summaries, GPS tracking etc.).

While existing systems offer relevant functionality (e.g., alert dissemination based on weather reports, location-based alerting, mobile app based alert delivery), there are several limitations of these systems when used in the context of disaster alerting, First, these systems primarily focus on weather-related alerts and their utility in the context of other hazards (e.g., wild-land fire) is limited. It is unclear how well these applications and their associated frameworks can operate robustly when the scale of the deployment is large or when there is significant damage to the infrastructure. In contrast, SMART-C incorporates input from multiple sources such as social media feeds, multimedia messaging, and citizen reports into alert generation. Data from such sources could significantly improve reliability, accuracy, and customization. SMART-C additionally offers the ability for geo-spatial analysis for fine-grained customization of alerts (e.g. geo-context such as vicinity to chemical factories). Finally, the policy-based alerting platform in SMART-C enables alerting functionality during all phases of the disaster lifecycle.

VI. Conclusions And Future Work

SMART-C is unique since it incorporates an integrated system capable of ingesting multi-modal data from many different sources and performing analytics to create a coherent picture of the event. Technologies that aim to derive input from individual sources (e.g., NWS) do not face challenges related to dealing with multiple sources of rich information, including disambiguation and removal of redundancy related to one event. The additional modalities could include rich information such as images or videos – the analysis of such data has not been the focus of existing social media studies, and is a key strength of the SMART-C approach and framework. In addition, since the system is designed using a plug-and-play, modular approach, such existing work is complementary; and additional sources and analytics can be easily plugged in through the service architecture.

Acknowledgements

Smart-C is being developed by a team of colleagues from Florida International University (Naphthali Rishe), Rutgers University (Vijay Atluri, Basit Shafiq, Soon Chun, Jaideep Vaidya), SRI (Zsolt Kira, Dalton Pont), and University of California, Irvine (Dmitri Kalashnikov, Stylianos Doudalis, Mehdi Sadri, Ye Zhao). Their contributions to the concept, design, and architecture of Smart-C are gratefully acknowledged.

References

[1] Sarcevic, Aleskandra, Leysia Palen, Joanne White, Mossaab Bagdouri, Kate Starbird, Kenneth M. Anderson, (2012). “Beacons of Hope” in Decentralized Coordination: Learning from On-the-Ground Medical Twitterers During the 2010 Haiti Earthquake 2012 ACM Conference on Computer Supported Cooperative Work, Bellevue, WA.

[2] Foteini Alvanaki, Sebastian Michel, Krithi Ramamritham, Gerhard Weikum: See what's enBlogue: real-time emergent topic identification in social media. EDBT 2012: 336-347

[3] H. Becker, M. Naaman, and L. Gravano Selecting Quality Twitter Content for Events, in Proc. of the Fifth Intl. AAAI Conference on Weblogs and Social Media (ICWSM), 2011

[4] H. Becker, M. Naaman, and L. Gravano, Beyond Trending Topics: Real-World Event Identification on Twitter, Hin Proc. of the Fifth Intl. AAAI Conference on Weblogs and Social Media (ICWSM), 2011.

[5] G. L. Crawford, “Noaa weather radio (nwr) - a costal solution to tsunami alert and notification,” Natural Hazards, vol. 35, no. 1, May 2005.

[6] K. Dow and S. L. Cutter, Emerging Evacuation Issues: Hurricane Floyd and South Carolina, Natural Hazards Review 3:12-18, 2002.

[7] Hughes, Amanda and Leysia Palen (2009). Twitter Adoption and Use in Mass Convergence and Emergency Events. Proceedings of the 2009 Information Systems for Crisis Response and Management Conference (ISCRAM 2009), Gothenberg, Sweden.

[8] Hojjat Jafarpour, Jay Lickfett, Kyungbaek Kim, Bo Xing, Sharad Mehrotra, Nalini Venkat, (2009), “A Policy Driven Meta-Alert System for Crisis Communications”, DHS Workshop on Emergency Management: Incident, Resource, & Supply Chain Management 2009

[9] Michael Mathioudakis, Nick Koudas: TwitterMonitor: trend detection over the twitter stream. SIGMOD Conference 2010: 1155-1158

[10] M. McGinley, D. Bennet, and A. Turk, "Design criteria for public emergency warning systems," Proceedings of the 3rd Intl. ISCRAM Conference, USA, May 2006

[11] S. Mehrotra, C. Butt , D. Kalashnikov, N. Venkatasubramanian, R. Rao, G. Chockalingam, R. Eguchi, B. Adams C. Huyck, (2004) "Project RESCUE: Challenges in Responding to the Unexpected", SPIE Internet Imaging Conference.

[12] D. S. Mileti, and L. Peek, The Social Psychology of Public Response to Warnings of a Nuclear Plant Accident, Journal of Hazardous Materials 75: 181-194, 2000

[13] L. K.Moore, "Emergency communications: The emergency alert system (eas) and all-hazard warnings," CRS Report for the Congress – Received through the CRS Web, July 2006

[14] National Academies Public Meeting, (2011), "Increasing National Resilience to Hazards and Disasters." organized by NACS, Engineering, and Public Policy. http://sites.nationalacademies.org/PGA/COSEPUP/nationalresilience/index.htm

[15] Rabia Nuray-Turan, Dmitri V. Kalashnikov, and Sharad Mehrotra Exploiting web querying for web people search. In ACM Trans. on Database Systems (ACM TODS), 37(1), Feb 2012

[16] Rabia Nuray-Turan, Dmitri V. Kalashnikov, and Sharad Mehrotra Adaptive connection strength models for relationship-based entity resolution. In *ACM Journal of Data and Information Quality*, 2012

[17] Palen, Leysia, Sarah Vieweg, Sophia Liu, Amanda Hughes (2009). Crisis in a Networked World: Features of Computer-Mediated Communication in the April 16, 2007 Virginia Tech Event. Social Science Computing Review, Sage, (pp 467-480).

[18] Bo Pang and Lillian Lee (2008). "Opinion Mining and Sentiment Analysis", Foundations and Trends in Information Retrieval: Vol. 2: No 1–2.

[19] Ramesh R. Rao, Jon Eisenberg, and Ted Schmitt, (2007), "Improving Disaster Management: The Role of IT in Mitigation, Preparedness, Response, and Recovery", The National Academies Press. ISBN-13: 978-0-309-10396-1.

[20] J. H. Sorensen, Hazard Warning Systems: Review of 20 Years of Progress, Natural Hazards Review 1: 119-125., 2000.

[21] J. H. Sorensen, and B. Vogt, Community Processes: Warning and Evacuation, pp. 183-199 in H. Rodriguez, E. L. Quarantelli, and R. R. Dynes (eds.) Handbook of Disaster Research. New York:Springer, 2006.

[22] Starbird, Kate, Grace Muzny and Leysia Palen (to appear 2012). Learning from the Crowd: Collaborative Filtering Techniques for Identifying On-the-Ground Twitters during Mass Disruptions. To appear in the Proceedings of the Conference on Information Systems for Crisis Response and Management (ISCRAM 2012), Vancouver, BC

[23] Vieweg, Sarah, Amanda Hughes, Kate Starbird, Leysia Palen (2010). Microblogging During Two Natural Hazards Events: What Twitter May Contribute to Situational Awareness. The Proceedings of the ACM 2010 Conf. on Computer Human Interaction (CHI 2010).

[24] Amir Tamrakar, Saad Ali, Qian Yu, Jingen Liu, Omar Javed, Ajay Divakaran, Hui Cheng, Harpreet Sawhney, "Evaluation of Low-Level Features and their Combinations for Complex Event Detection in Open Source Videos", CVPR 2012.

[25] K. Tierney, M. K. Lindell, and R. W. Perry, Facing the Unexpected: Disaster Preparedness and Response in the United States, Washington, DC: Joseph Henry Press, 2001.

[26] Sudha Verma et.al . NLP to the Rescue? Fifth Intl. AAAI Conference on Weblogs and Social Media, July 2011, Barcelona, Spain.

[27] C. Desai, D. Kalashnikov, S. Mehrotra, and N. Venkatasubramaniam, "Using Semantics for Speech Annotation of Images", ICDE 2009

[28] D. Kalashnikov, S. Mehrotra, Jie Xu, N. Venkatasubramanian, "A Semantics-Based Approach for Speech Annotation of Images", IEEE Transactions on Knowledge and Data Engineering, 23(9), September 2011

[29] Liyan Zhang, Ronen Vaisenberg, Sharad Mehrotra, and Dmitri V. Kalashnikov, "Video Entity Resolution: Applying ER Techniques for Smart Video Surveillance", In Workshop on Information Quality and Quality of Service for Pervasive Computing (IQ2S 2011) in Conjunction with IEEE PERCOM 2011, invited paper, Mar 21-25, 2011.

[30] Rabia Nuray-Turan, Dmitri. V. Kalashnikov, Sharad Mehrotra, Yaming Yu, "Attribute and object selection queries on objects with probabilistic attributes", ACM Trans. Database Syst. 37(1): 3 (2012)

[31] Rabia Nuray-Turan, Dmitri V. Kalashnikov, Sharad Mehrotra, "Exploiting Web querying for Web people search", ACM Trans. Database Syst. 37(1): 7 (2012)

[32] Zhaoqi Chen, Dmitri V. Kalashnikov, Sharad Mehrotra, "Exploiting context analysis for combining multiple entity resolution systems", SIGMOD Conference 2009.

[33] N. Adam and B. Shafiq, "Spatial Computing and Social Media in the Context of Disaster Management," IEEE Intelligent Systems, December, 2012

[34] C. Davison, D. Massaguer, L. Paradi, M. Reza Rahimi, B. Xing, Q. Han, S. Mehrotra, N. Venkatasubramanian. Practical Experiences in Enabling and Ensuring Quality Sensing In Emergency Response Applications, Wkshp on Pervasive Networks for Emergency Management (PerNEM), 2010.

[35] B. Xing, M. Deshpande, S. Mehrotra and N. Venkatasubramanian. Gateway Designation for Timely Communications in Instant Mesh Networks, IEEE PerCom on Pervasive Wireless Networks, 2010.

[36] B. Xing, S. Mehrotra and N. Venkatasubramanian, RADcast: Enabling Reliability Guarantees for Content Dissemination in Ad Hoc Networks, INFOCOM 2009.

[37] M. Stehr and C. Talcott. Planning and learning algorithms for routing in disruption-tolerant networks. MILCOM 2008.

[38] Ngoc Do, Cheng-hsin Hsu, Nalini Venkatasubramanian. HybCast: Efficient Rich Content Dissemination over Hybrid Cellular and Ad Hoc Networks, IEEE SRDS 2012.

[39] O. Wolfson and B. Xu, "Mobile Peer-to-peer Data Dissemination with Resource Constraints", Proc. of the 8th International Conference on Mobile Data Management, Mannheim, Germany, May 2007.

[40] [GG03] M. Gruteser, and D. Grunwald, Anonymous Usage of Location-based Services through Spatial and Temporal Cloaking, In proceeding of MobiSys 2003.

[41] [DKH12] Mayur Deshpande, Kyungbaek Kim, Bijit Hore, Sharad Mehrotra, Nalini Venkatasubramanian. ReCREW: A Reliable Flash Dissemination System. IEEE Transactions on Computers, 2012.

[42] [JHM09] Hojjat Jafarpour, Bijit Hore, Sharad Mehrotra and Nalini Venkatasubramanian:CCD: Efficient Customized Content Dissemination in Distributed Publish/Subscribe, ACM/IFIP/USENIX Middleware 2009, Urbana-Champaign, IL,USA, November 2009.