Rethinking the Future of Wireless Emergency Alerts: A Comprehensive Study of Technical and Conceptual Improvements

SUMEET KUMAR, School of Computer Science, Carnegie Mellon University, USA
HAKAN ERDOGMUS* and BOB IANNUCCI, Department of Electrical and Computer Engineering, Carnegie Mellon University, USA
MARTIN GRISS*, Martin Griss Associates, USA
JOÃO DIOGO FALCÃO†, AiFi, USA

The Wireless Emergency Alerting (WEA) service is a standards-based transport and presentation channel used nationwide in the United States. The service can deliver short text warnings to wireless subscribers through a cell broadcast mechanism. For emergency situations in which a broadcast modality and a single, short text message are sufficient to convey information, the WEA service can be efficient and effective. However, the content to be delivered may necessitate more than a single, unchanging short message. In this research, we first examine the WEA service from the perspective of alert originators. We then use the insights gained to explore the efficacy of a range of potential extensions to the service. The extensions mainly address the importance of user context and the ability to create awareness through careful attention to the integrity of the vital information. We evaluated these extensions using a WEA emulation testbed in two public usability trials. We present an analysis of the broad range of improvements as a basis for further research into improving the service. We conclude that (1) precise geo-targeting augmented with location information and maps is an important aspect of capturing users' context, and (2) presenting information in a digested form can markedly improve the actionability and the accuracy of interpretation.

CCS Concepts: • **Information systems** → Location based services; Mobile information processing systems; • **Human-centered computing** → Ubiquitous and mobile computing design and evaluation methods;

Additional Key Words and Phrases: Wireless Emergency Alerts, Context awareness, Emergency alerting, Mobile emergency systems

ACM Reference Format:
Sumeet Kumar, Hakan Erdogmus, Bob Iannucci, Martin Griss, and João Diogo Falcão. 2018. Rethinking the Future of Wireless Emergency Alerts: A Comprehensive Study of Technical and Conceptual Improvements. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 2, 2, Article 71 (June 2018), 33 pages. https://doi.org/10.1145/3214274

*Corresponding author
†This work was done while the author was affiliated with Carnegie Mellon University.

Authors' addresses: Sumeet Kumar, sumeetku@cs.cmu.edu, School of Computer Science, Carnegie Mellon University, Pittsburgh, Pennsylvania, USA, 15213; Hakan Erdogmus, hakan.erdogmus@csv.cmu.edu; Bob Iannucci, bob@sv.cmu.edu, Department of Electrical and Computer Engineering, Carnegie Mellon University, NASA Ames Research Park, Building 23 (MS 23-11), Moffett Field, California, USA, 94035; Martin Griss, martin.l.griss@gmail.com, Martin Griss Associates, USA; João Diogo Falcão, joao@aifi.io, AiFi, USA.

2474-9567/2018/6-ART71 $15.00
https://doi.org/10.1145/3214274

1 INTRODUCTION

The Wireless Emergency Alerting (WEA) service [4], formerly known as the Commercial Mobile Alert System (CMAS), is part of the United States Federal Emergency Management Agency's Integrated Public Alert and Warning System (IPAWS) [2]. Using the Short Message Service Cell Broadcast (SMSCB) protocol [6], the WEA service provides a dissemination path for alert and warning messages to the public on WEA-capable mobile devices. When an alert is sent, it is targeted using cell towers, and is delivered to all WEA-capable devices covered by those cell towers. This form of targeting is coarse at best. The inability to provide fine-grained targeting, combined with the limitations of text-based short messages in delivering adequate information and/or actionable advice, is frequently cited as a disadvantage that increases the likelihood of citizens to opt out of this mode of alerting or ignore alert messages. This behavior in turn reduces the effectiveness of the WEA service.

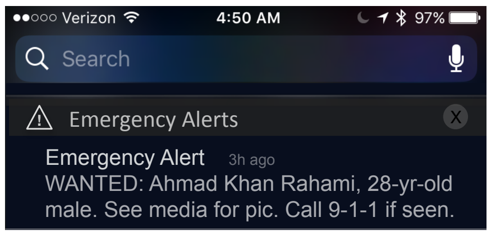


Fig. 1. Example WEA Message. This alert was sent to the citizens of New York City in an attempt to capture a bombing suspect on the morning of September 19, 2016.

Rapid improvements in mobile and network technologies enable a significant re-thinking of WEA to address these concerns. When SMSCB was created, phones were little more than devices for making voice calls and handling text messages. Now, smartphones have capabilities exceeding those of personal computers from a decade ago. Similarly, network capacities have improved dramatically when compared to the initial demonstration of SMSCB. New protocols such as Long Term Evolution (LTE) broadcast have emerged, and user expectations have shifted from text messaging to rich messaging. Considering how both network and phone technologies have changed dramatically since the inception of SMSCB, in this work, we look for ways to overcome the WEA service's shortcomings within the framework of the current WEA architecture.

The primary goals of this research are to (1) gain insight into WEA adoption and acceptance issues, in particular with respect to perceived poor public response to alert messages, (2) develop and test strategies for overcoming these issues, and (3) explore the technical feasibility of the available options. These goals are well aligned with the future research directions set forth for the emergency alerting and warning system.[1] With these goals in mind, we performed an extensive series of studies in two phases, with end-to-end implementation in a testbed of both simple and complex improvements to the WEA service. The timeline of the studies in Fig. 2 shows the work schedule and components, each of which is explained in more detail, with the exception of the Geo-Target Compression work, which is described elsewhere [20, 21] and is beyond the scope of this paper.

The WEA service is part of the United States government's emergency warning infrastructure, and is currently only in use nationwide in that country. Thus we studied it in a strictly national context. Some of results, ideas, and lessons learned, especially with respect to technical aspects, may be applicable to similar systems used in or planned for other jurisdictions, however we haven't investigated their generalizability and implications outside the United States.

The rest of this paper is organized as follows. We explain the overall design of the research in Section 2. We introduce the first phase, the Alert Originator Requirements phase, in Section 3 and summarize the insights

[1] https://www.nap.edu/resource/24935/RH-alerts.pdf

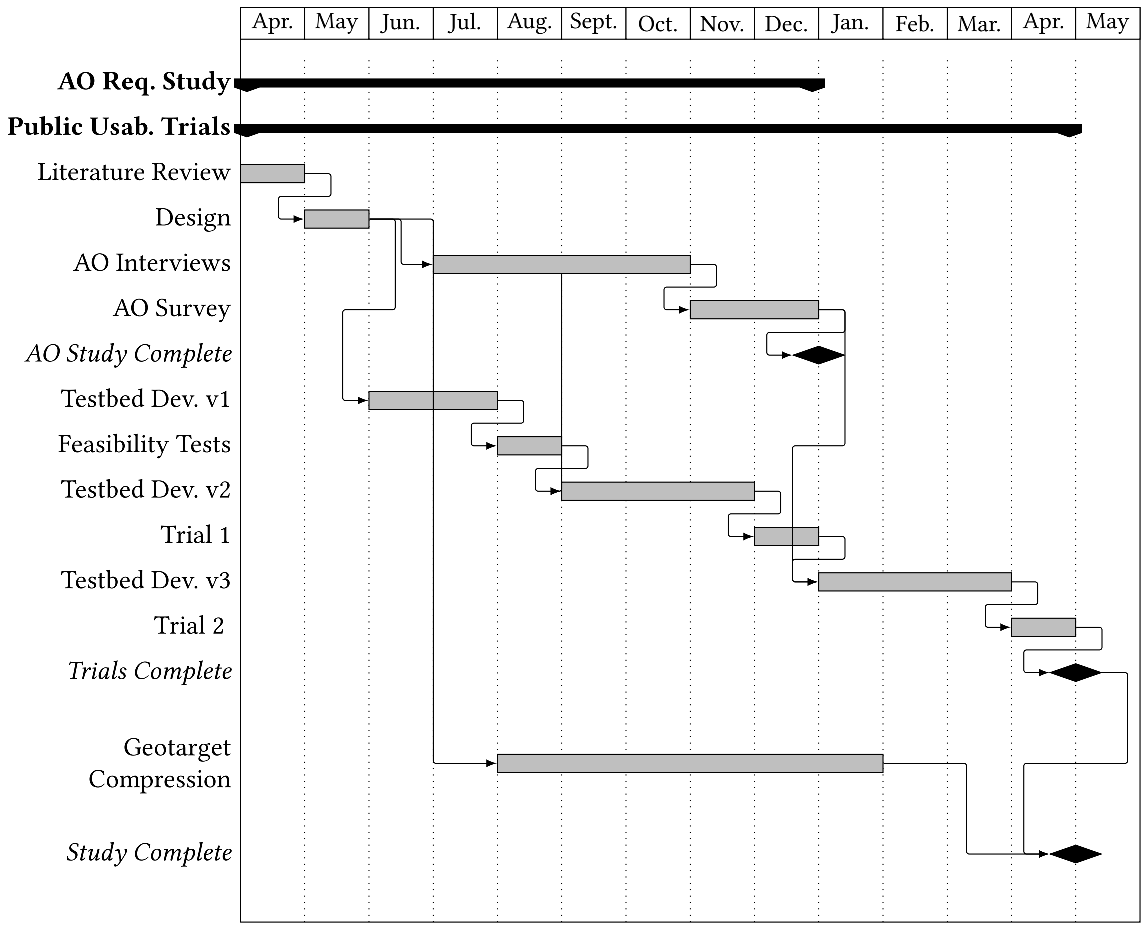


Fig. 2. Gantt chart of the CMU WEA Study showing major phases. The Alert Originator (AO) Requirements Study phase involved AO Interviews and the AO Survey. The Public Usability Trial phase involved the development of different versions of the WEA+ emulation testbed implementing various enhancement features and the execution of two real-time, live-user trials that took advantage of the testbed. Diamonds mark major milestones.

that were relevant to the next phase. In Section 4, we present the next phase, the Public Usability Trial phase, including the development of the WEA emulation testbed and experimental framework (WEA+), experiment designs used in this phase, and data analysis approach. The results of the second phase are presented in Section 5 together with the features and characteristics selected for testing. Study limitations are discussed in Section 6. Section 7 presents related work. This is followed in Section 8 by a summary of our recommendations. The paper concludes with Section 10, which discusses future work.

2 RESEARCH DESIGN

In this research we used a mixed methods approach divided into two distinct phases and addressing different but mutually re-enforcing research questions.

The first phase of the research–the Alert Originator Requirements Study–focused on deepening our understanding of the challenges confronting the authorities who create the alerts, or alert originators (AOs). We designed and performed two investigations consisting of an interview study conducted using a qualitative approach, mainly Grounded Theory [32], exploring a smaller number of deep insights with a small sample, and an online survey using a a mixed quantitative and qualitative approach, exploring a larger set of shallower insights with a much larger sample. The interview study was also used to inform the focus of the online survey, in terms of selecting findings to further validate with a large audience and filling in the gaps.

The Alert Originator Requirements Study served primarily to scope and ground our technical contribution and its validation in the second phase. We used the results, as they emerged, to identify and revise promising technical improvements to the WEA service that addressed pressing alert originator concerns. These improvements surfaced during multiple workshops conducted within a research team of eight to eleven participants over a period of one year during which the team continuously reviewed the results, suggested concrete new features for the WEA service, revised the features previously proposed, explored their feasibility, and discussed and devised strategies to validate their effectiveness. Concurrently, and incrementally, the team implemented a flexible testbed to support the implementation and testing these new features.

The second phase of the work–the Public Usability Trials–included all the design and implementation efforts pertaining to potential WEA enhancements, culminating in three experiments that evaluated the impact of these enhancements. This phase involved the development of a set of extensions to WEA that we call WEA+, our primary technical contribution.

WEA+ was implemented and tested with live subjects in three experiments. According to priorities identified in the Alert Originator Requirements Study, we focused on technical enhancements that could take advantage of advanced smartphone capabilities, leaving out most human-factors-related changes (*e.g.,* support for multiple languages and use of different fonts, font sizes, and sounds). The experiments involved 237 subjects who received simulated alerts in real time in simulated emergency situations. To our knowledge, this is the first study of its kind that evaluated WEA capabilities and extensions in a realistic setting with human users in real time. The experiments used a novel repeated-measures design to compare a dynamically created randomized control group that received regular WEA messages with a dynamically-created randomized treatment group that received enhanced alerts according to multiple effectiveness criteria. The criteria involved feedback provided by the subjects in real-time as well as automated measurements, both of which were incorporated to WEA+ as part of a set of mechanisms that allow for easy experimentation.

The data collected during the second phase experiments were analyzed using standard statistical techniques that are well suited for frequency data. They included Chi-square goodness of fit and independence tests accompanied by proper effect size measures, such as Cramer's V, Phi statistic, and Odds Ratio that match the data and statistical tests used.

Finally, the findings of the two phases were synthesized to a set of high-level recommendations to inform the evolution of the WEA service based on evidence from our work.

3 ALERT ORIGINATOR REQUIREMENTS STUDY

Following the Federal Emergency Management Agency's terminology, we use the term Alert Originator (AO) to refer to emergency services personnel who work in a role that involves assessing or managing emergency situations and crafting, approving and ultimately disseminating public alert messages. As a key part of our research, we conducted a comprehensive study to update our understanding of the AO's perspective given the WEA service's current use. We organized this preliminary study in two parts: (1) In-depth interviews with a small number of selected AOs and (2) A large-scale online survey of the AO community. The investigations helped us confirm or refute standing assumptions about the WEA service, identify its strengths and shortcomings, and solicit recommendations on how to improve its adoption, effectiveness, and relevance.

Formalized in Goal-Question-Metric (GQM) [8] terms, the purpose of the Alert Originator Requirements Study was:

- to *validate* previous findings and recommendations regarding the viability, limitations, advantages, and use of the WEA service (G1) *and identify* impediments and opportunities regarding the adoption of the WEA service by Alert Originators (G2) and acceptance of the WEA service by the public (G3)
- *for the purpose of* prioritizing, developing, and evaluating enhancement goals and options for future wireless broadcast services (G4)
- *with respect to* increased geo-targeting specificity (G5) of alert messages and improved relevance (G6) to their recipients *from the perspective of* Alert Originators (G7)
- *in the context of* emergency alert systems and services used by different jurisdictions, agencies, and response communities throughout the nation (G8).

The goals G1-G8 were further refined into specific research questions, which the the AO Requirements Study attempted to answer (see Table 7 in the Appendix).

3.1 Design and Recruitment

The subjects for the Alert Originator Requirements Study were selected from a sample population of 599 alert originators assembled from a combination of sources, including a prior survey conducted by CMU's Software Engineering Institute as well as personal contacts and referrals.

From this pool, we selected 22 interview candidates, balancing geography and role/level. Thirteen candidates responded positively to our invitation (a response rate of 59%), representing a mix of employees from city, county, state and national organizations. The participants' titles spanned Senior Coordinator of the Office of Emergency Services (OES), Director of OES, Program Lead of Emergency Communications, Emergency Management Liaison, Assistant Director of Emergency Management, Preparedness Coordinator, Director of National Center, OES Coordinator, Emergency Management Specialist, Emergency Operations Center (EOC) Coordinator, and Head of Protective Services. Their affiliations spanned Santa Clara County, California, City of Palo Alto, California, National Weather Service, Harris County, Texas, Johnson County, Kansas, State of Florida, Commonwealth of Massachusetts, National Center for Missing and Exploited Children, the National Aeronautics and Space Administration (NASA) Ames Research Park, and Department of Homeland Security/FEMA. All interviewees had either significant experience or significant familiarity with WEA. Additional details of the interviewees are available in Table 8 of the Appendix.

The interviews helped identify major themes to address and specific points to further probe, refine, and validate in the online survey. A survey instrument was developed based on these themes and specific points. The survey instrument was first tested with interview participants, contacts in our initial sample with whom we had a previous working relationship, and a small group of Software Engineering Researchers, a partner in our project, who had previously worked on WEA. Survey questions were revised based on the feedback received for deployment to the larger sample of alert originators. We had to remove 144 contact from our initial list because we couldn't validate their contact information or email addresses. This left us with a sample of 455 to whom to send the online survey. The response rate to the online survey was 19%. Finally, the results of the interviews and online survey were synthesized into a small number of key observations.

3.2 Interview Methodology and Insights

The interviews addressed several central topics related to the research questions, but also had an open component to allow emerging themes. They were conducted using a combination of semi-structured and open interviewing techniques and lasted up to an hour for each interview. Specific topics addressed included: whether and how the AOs were using WEA, and what, if any, barriers they encountered in adopting WEA; the need and perceived feasibility of more precise geo-targeting; and issues related to the relevance of WEA to and its acceptance by the general public. We also sought to uncover new requirements based on current experience with the service and opportunities afforded by smartphone capabilities, connections to social media, and inclusion of links and rich media in alerts. Interview notes were transcribed, summarized individually, and thematically analyzed during a full-day workshop based on a Grounded Theory approach [32].

The major interview themes that pertained to the online survey were:

3.2.1 Longer Messages. Current WEA service has a limit of 90 characters per alert. Most interviewees stated that they were unable to craft meaningful messages to the general population within this constraint.

3.2.2 Better Geo-targeting. The majority of the interviewees stated that increasing the geo-targeting precision of the WEA service will rectify key adoption challenges by delivering alert messages only to those impacted by an emergency.

3.2.3 WEA as a Warning Alarm vs. a Rich Information Service. Some AOs perceived the WEA service as a "bell ringer" technology, akin to sounding the first alarm only. Others believed that the natural evolution of the WEA service should involve uncovering ways to directly embed additional information within the alert messages themselves–including maps, location information, and links–and to couple them with effective interpretation, incident follow-up, and closure mechanisms to improve the situational awareness of the recipients.

Insights from the AO interviews primarily influenced our decisions regarding the initial WEA enhancements to tackle during the field experiments and the first trial.

3.3 Online-Survey Methodology and Insights

To gather additional data and to refine the results from the interviews, we deployed a 30-question survey to the larger AO community. Survey questions covered these categories: basic demographic info; perceived WEA adoption rate; adoption and opt-out reasons for AOs; suggestions for improving WEA adoption; suggestions regarding features desired in alert generation tools, mobile devices, and the WEA service itself; and the role of WEA within the larger landscape of emergency information dissemination.

The majority of the survey questions were multiple choice. These were processed using simple frequency analysis, as shown in Table 1 for two sample questions.

The remaining few questions had free-text responses, requiring subsequent manual thematic analysis [9]. We used the University of Pittsburgh's Coding Analysis Toolkit [10] to process the free-text responses, cluster them and map the clusters to sets of emerging themes that capture their essence. This manual coding process of the free text responses reads each response and identifies words or phrases that seemed to indicate prevalent themes, iterating several times over the responses to develop meaningful clusters. For example, in one of the questions, we discovered themes that received multiple responses such as "smaller or more precise polygon," "opt out or customize message types received," "keep it short more as alarm bell," "provide enough information to cause action," "educate message creators," "educate message receivers or proactive outreach," and "do not send annoying, fatiguing irrelevant messages." Similar themes were discovered in the other free text responses.

The online survey was given to a list of 455 AOs. We received 88 responses, out of which 79 were usefully complete. The majority of the respondents had significant experience in emergency alert origination (over 10 years), but nearly half (44 %) did not have prior experience with WEA. 75% percent indicated their organizations regularly issued emergency alerts. The sample was diverse, with representation from city-level, county-level, state-level, and national bodies.

Here we list the main findings pertaining to the online survey. Raw survey results, thematic analysis of open-ended questions and comments, and further details can be found in our full report [16].

Table 1. Sample Survey Questions and Results. Full results are available in [16].

Question 8: How likely are you to adopt and use WEA if the minimum geographic area that you can define for the delivery of the alert is...

Number of Responses: 67 (76%) Number of Skips: 21 (24%)

	Very Unlikely	Unlikely	Likely	Very Likely
as small as 1 square city block	11	3	15	37
as small as 1 neighborhood (10 city blocks by 10 city blocks	8	4	23	31
1 square mile	6	5	27	26
as large as 10 square miles	11	10	23	19
as large as 100 square miles	27	13	9	17

Question 13: Recognizing the trade-off between message length and limits and costs of wireless broadcast technology, how effective do you believe the WEA service can be if the maximum message length is?

Number of Responses: 67 (76%) Number of Skips: 21 (24%)

	Very Ineffective	Ineffective	Effective	Very Effective
90 characters	3	21	36	5
280 characters	0	5	27	35
500 characters	5	16	18	27
1000 characters	20	18	12	16

3.3.1 Ways to Improve WEA Adoption. The AOs interpreted WEA adoption from two perspectives: their own propensity to use the WEA service as a viable emergency warning channel to issue alerts and the propensity of the public to rely on the WEA service to receive emergency information. 55 AOs provided suggestions for improving WEA adoption by the AOs. The top suggestions were to: permit smaller geo-targets (61%); increase allowable message length (12%); and provide a mechanism to test WEA messages (10%). 54 AOs provided suggestions for improving WEA adoption by the public. The top suggestions were to: allow recipients to customize which messages they receive (13%); avoid over-warning with too many alert messages (12%); ensure that WEA messages are actionable (11%); improve geo-targeting precision (9%); and allow longer messages (9%).

3.3.2 Importance of Geo-targeting. The ability to easily define a geo-target during alert creation was the top preferred front-end feature for the majority of AOs, followed by compliance with the Common Alerting Protocol (CAP) format, a standard for crafting warning messages with meta data [26]. With respect to desired geo-targeting precision, respondents were most likely to use WEA (at 81%) if minimum geo-targeting resolution were less than roughly the size of an urban neighborhood.

3.3.3 Message Length. Although a majority, 61%, of the surveyed AOs felt the current 90-character maximum length was sufficient, a larger majority felt that longer messages of up to 500 characters (93% for up to 280 characters and 67% for up to 500 characters) would be more effective *even if doing so would increase the cost and push the limits of wireless broadcast technology.* This result suggests an unqualified support for increased message length in future WEA service.

3.3.4 Role and Proper Use of WEA. A significant majority, 75%, of AOs thought that the AO community must rethink the vision of WEA. 99% agreed that a WEA message should give the recipient enough actionable information beyond simply sounding the alarm. 84% thought that a WEA message about an impending emergency should be followed up with regular status updates, including a closure message.

3.3.5 Links and Richer Media in Alerts. The surveyed AOs did not give unqualified support for enhancing WEA messages with external links and rich media. A majority of the respondents disagreed that WEA messages should be enriched with maps or images *if doing so would complicate alert generation or jeopardize the willingness of carriers to participate in the WEA service.* The respondents were divided (54% for vs. 46% against) between whether embedding links to websites and social media into alert messages would be worthwhile *if doing so has the risk of overloading the communication network.*

The results of the AO survey confirmed the majority of the findings from the interviews. They also triggered modifications to the enhancements tested during the first public usability trial (Experiment 1) and influenced the additions implemented and tested during the second trial (Experiments 2 and 3). These changes included measuring response rate and delay, removing the tracking of update alerts (which was not found useful), addition of two location-aware enhancements (Location Prediction and Location History), and addition of a situational awareness enhancement tested in a newly designed, separate experiment (Situation Digest view in Experiment 3).

3.4 Alert Originator Requirements Study Conclusions

Key observations from the requirements study pertain to the WEA service's wide reach as a major advantage, importance of geo-targeting specificity, leveraging smartphone capabilities, and strategies for increasing both the capacity of an alert to prompt action and the likelihood of an alert to be relevant to the recipient via longer messages, maps, rich media, and other enhancements, provided that these enhancements do not lead to network overloading or discourage carrier participation.

Most AOs reaffirmed they use or plan to use the WEA service because it is able to nearly instantaneously reach anyone in a designated area (similar to "bell ringer" or "sounding a siren") without the need for prior registration. The WEA service's ability to leverage the increasing prevalence of smartphones was touted as a major advantage, which we heavily leveraged in the Public Usability Trials phase when designing and testing the various enhancements for WEA+. The results of the study with alert originators thus informed the selection and design of enhancement features tested during the next phase of our work. Table 2 links these results to enhancement features tested.

4 PUBLIC USABILITY TRIALS

During the Public Usability trials, we explored both the usefulness and applicability of a range of hypothesized enhancements and extensions to the WEA service. We developed a prototype of an enhanced, emulated WEA service and experimental framework—the WEA+ system—which included an alert creation and monitoring subsystem (Fig. 3), a message delivery subsystem (Fig. 4), and native smartphone applications (Fig. 5 and Fig. 6) for processing delivered alerts and presenting them to recipients.

We then conducted two one-week-long usability trials involving three controlled experiments. The purpose of the public usability phase was to test selected WEA enhancement features with human subjects. This phase included, in addition to the trials, software development for three iterations of WEA+ and the associated smartphone applications and multi-location field feasibility tests with preliminary capabilities (mostly focused on implementing precise geo-targeting). During the trials, using simulated emergency scenarios, we assessed the effectiveness of new system features by measuring and comparing participant responses to alerts issued in real-time with and without any enhancements.

The high-level research questions and study designs of the experiments conducted in the trials are presented in the following subsections. Details and supporting materials can be found in the Appendix, Tables 9 to 22.

Table 2. Influences of the Alert Originator Requirements Study on the Public Usability Trials

Enhancement Feature Tested	**Relevant AO Requirements Study Findings**
Long Message	Sections 3.3.1, 3.2.1, 3.3.4 and 3.3.3: pertaining to the need to convey more information using longer text messages; also with regard to improving actionability through being able to improve message content
High-Information Map	Sections 3.3.1, 3.2.3, 3.3.4, and 3.3.5: pertaining to importance of actionable information and inclusion of rich media, including maps; although in the survey, the AOs did not give unqualified support for improving actionability with rich media embedded in alerts due to concerns regarding complex alert generation and resistance from wireless carriers, we believed these concerns to be easily overcome via technical solutions that take advantage of increasingly more pervasive capabilities built into smartphones and advancements in cellular networks
Geo-Targeting	Sections 3.3.1, 3.3.2 and 3.2.2: pertaining to the ability to precisely target alert messages to an effected area, which was prevalent in both the survey and interview components
External Link	Section 3.2.3 and 3.3.5: pertaining to actionable information and support for embedding of additional information in alerts
Situation Digest	Sections 3.2.3, 3.3.1 and 3.3.4: pertaining to the importance of providing actionable information and proper incidence follow-up, updates, and closure with effective interpretation
Location History and Location Prediction	Sections 3.3.1, 3.2.3, 3.3.4: pertaining to the importance of improving actionability; these features were designed to leverage geo-targeting and maps to provide context awareness, thus increasing the relevance of the messages to the recipients, and ultimately actionability

4.1 Research Questions

The following high-level research questions were tackled in the two public trials: (a) *Potential Enhancements*: Would certain improvements to the WEA service make it more effective for the public?; (b) *Alert Characteristics*: How do certain alert characteristics impact the effectiveness of the WEA service for the public?; and (c) *Overall Impressions*: After exposure to the WEA service, does the public appreciate the benefits of the WEA service, with or without improvements?

Regarding (b) above, in two experiments, we studied certain factors that pertain to the attributes or nature of the alerts themselves, such as the timing or purpose of an alert, rather than possible enhancements or changes to the WEA service. These are what we refer to as *alert characteristics*. Regarding (c) above, we evaluated the overall impressions about WEA via a final questionnaire sent to the subjects immediately after each experiment.

4.2 WEA+ Framework

The trials were conducted using our experimental platform WEA+. The WEA+ system comprised an Android mobile application (Fig. 5), an iOS application (Fig. 6) and a back-end subsystem (Fig. 4). The back-end included a web-based interface to create and schedule new alerts (with and without enhancements) as well as to monitor active alerts (Fig. 3), two web servers to persist and push the scheduled alerts, and a repository to store the data and the feedback sent by users. Figure 3 shows a screenshot from the interface of the WEA+ subsystem used for creating, issuing, and monitoring regular and enhanced WEA messages issued to the trial subjects.

The high-level architecture and the key technologies used in the creation of WEA+ are illustrated in Fig. 4. In the figure, two Node.js®-based web servers expose services to create and schedule alerts and to collect feedback from mobile phones. The alert originator user interface (previously shown in Fig. 3) communicated with these web servers for creating, modifying, deleting, and monitoring alerts. The web servers persisted the alert data and user feedback in MySQL databases. An Amazon-based push notification service was used to deliver alerts to

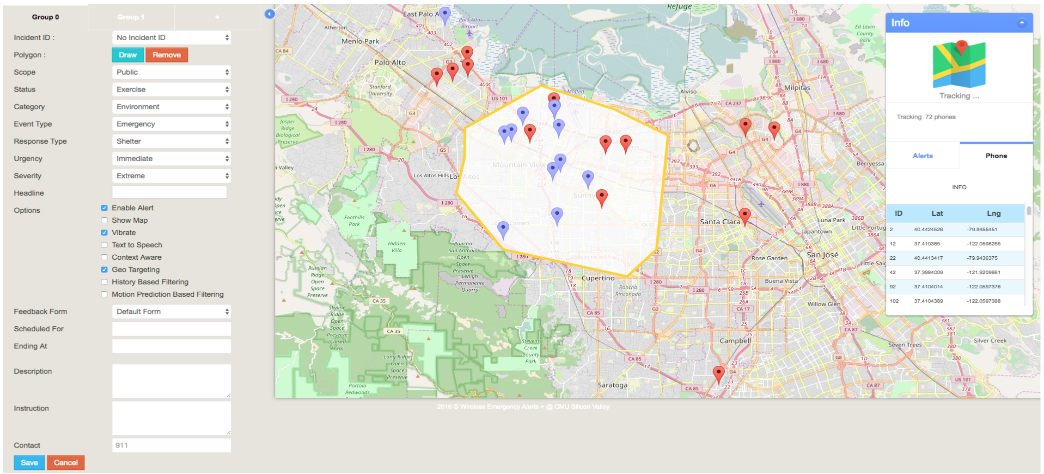


Fig. 3. The User Interface of the Alert Creation and Monitoring Subsystem of WEA+. The target region of an active alert, specified as a polygon, is overlaid on a map. The alert is delivered both in plain and enhanced form in a randomized way to a group of experiment subjects. The markers represent the positions and statuses of the subjects. The left pane shows attributes for the active alert.

mobile phones. Two native mobile applications—one for Android-based users (Fig. 5a) and one for iOS-based users (Fig. 6a)—were created to test different WEA features and characteristics.

In WEA+, the alert creation and monitoring subsystem (Fig. 3) plays a central role. It provides a web-based user interface, akin to a control center, to specify, schedule, and manage alert messages. It also allows A/B-type testing by randomly dividing the experiment subjects into two equally-sized groups. Each group receives the same alert, but the alert is presented with a distinct set of features to each group. In the experiments, to control the evaluated factor, we varied alert features one at a time between the two groups. For example, while testing the effect of text-to-speech (the factor being studied), subjects in group A received a spoken alert, whereas subjects in group B received the exact same alert in the regular way. All other factors were kept constant between the groups to isolate the factor being studied.

Figure 5a shows a snapshot of an enhanced alert delivered to an Android user. The alert is augmented with a high-information map centered around the alert's region. The alert shows a map overlaid with a red-framed polygon indicating the affected region. Also shown to the user are the the time period during which the alert is active, a text describing the alert content, and the alert type. If the precision geo-targeting feature is enabled for an alert, the mobile application running on the phone determines the location of the user, and shows the alert only if the user is inside the alert polygon or enters the alert polygon while the alert is still active. Fig. 5b shows a sample WEA+ feedback question for Android users. This feedback has a question and three options to select. Using 'Back' and 'Next' buttons, users can navigate to other feedback questions.

Consistent with the finding that AOs desire to make WEA compliant with CAP format [1] (Section 3.3.2) and with our desire to move WEA from a focus on *alerting* to true *situational awareness*, we created a new WEA protocol and supporting mechanisms to compress CAP to fit the transmission limitations of WEA and to then interpret CAP on smartphones for clear visual presentation [18]. This required solving three problems: (1) reducing the size of CAP messages without loss-of-information, (2) designing a mechanism to allow reduced-size CAP to be transported over un-modified WEA, and (3) creating a user interface for smartphones to present consolidated information from a *stream* of related WECAP messages in a concise form.

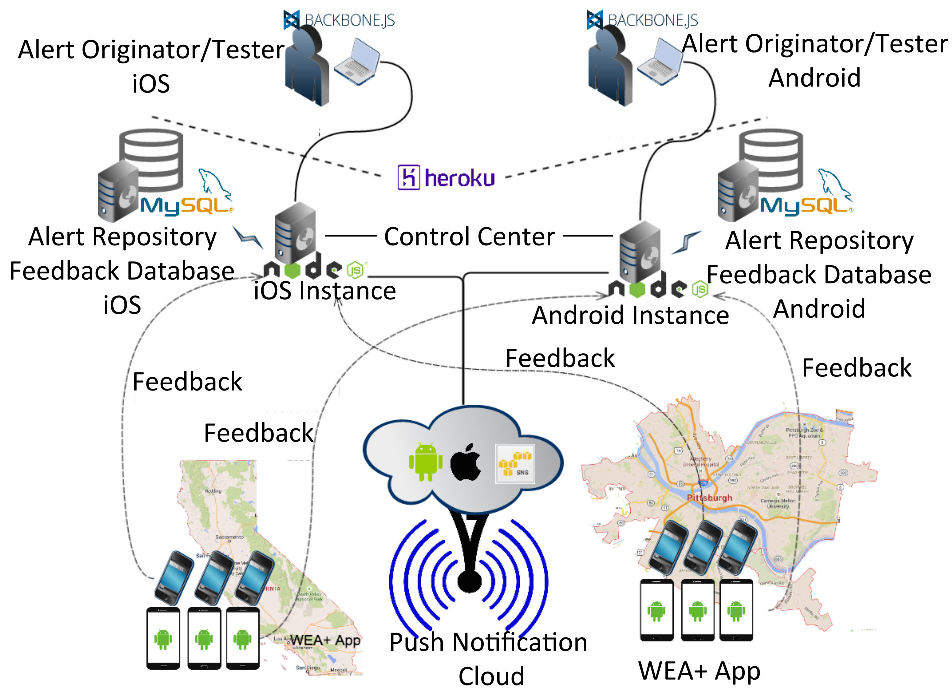


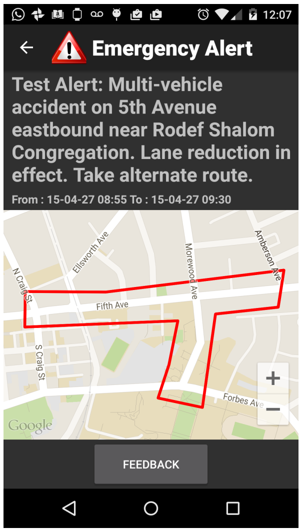
Fig. 4. Architecture of the WEA+ Testbed and Framework. WEA+ supports both Android and iOS clients. Alert originators create alerts on a Node.js®-based server. All data is stored in MySQL databases. Phones receive alerts via push notification managed by Amazon SNS, Google Cloud messaging service, and the iOS Push Notifications Service. Feedback is collected via question/answer screens on the phone.

Our solution to the first two problems was the creation of a domain-specific, compressed representation of WEA called WECAP, supported by a pagination coding scheme for WECAP messages that may exceed the WEA single-message size limit. Our solution to the user interface is the situation digest view depicted in Fig. 6a. A *headline* (in this example, **Earthquake Alert**) is prominent at the top of the display followed by the authoring entity and a time stamp. In the center of the display, citizens are presented with the immediacy of the recommended action as well the type of emergency. In smaller font, additional detail and instructions are presented. The icons on the left represent the immediacy, emergency type, and nature of instructions. On the top left of each icon is a colored circle that signifies the degree of immediacy and severity. This presentation supports situational awareness through an overlay mechanism that aggregates information from an ordered stream of related alerts (Fig 6b). Changes to individual fields such as severity and recommended actions are possible without resending redundant information.

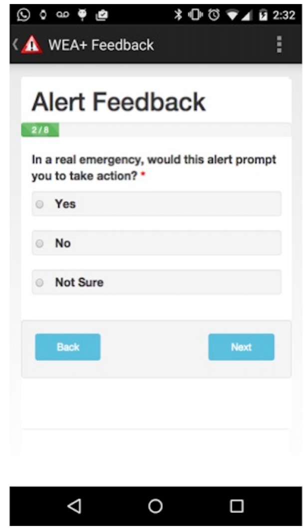
In our experiment, we evaluated the effectiveness of this approach over existing WEA.

4.3 Experiment Design, Subject Recruitment, and Analysis

The two public usability trials involved three experiments, during which we tested these features: inclusion of high-information maps, precise geographical targeting using on-phone geo-filtering, location-history-based targeting, use of a situation digest view related to a stream of connected alert messages, use of text-to-speech, targeting with location prediction, and inclusion of external links. We also evaluated certain alert characteristics,



(a) A Sample WEA+ Alert for Android Users. This alert is divided into (top) alert text and (bottom) map with impacted area. The goal was to test the High-Information Map feature combined with on-phone geo-filtering.



(b) A Sample WEA+ Feedback Question for Android Users. This screen has a feedback question and three options to select. Using 'Back' and 'Next' buttons, users can navigate to other feedback questions deployed right after an alert has been received.

Fig. 5. WEA+ Alert and Feedback for Android Users: Example Screen shots

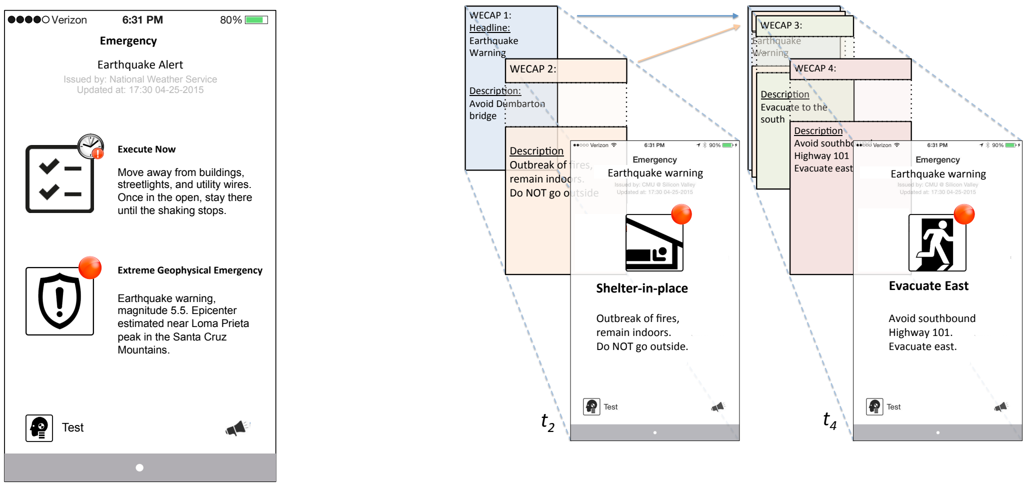
such as message length and timing, as well as overall impressions about WEA after exposure to the emulated WEA service. Table 3 summarizes the experiments.

The experiments were conducted with 140 subjects with Android devices and 93 subjects with iOS devices. Android subjects participated in Experiments 1 and 2, and the iOS subjects participated in Experiment 3. Experiment 1 was conducted only in Silicon Valley and Experiments 2 and 3 were conducted in two locations: Silicon Valley and Pittsburgh. Experiments 1 and 2 were designed to evaluate WEA+ enhancements (except situation digest view) and used the same single-factor, randomized repeated-measures design. The single factor was binary, representing either the control (Group A) or a tested enhancement feature (Group B). Experiment 3, which evaluated situational awareness, was conducted separately as it required significant changes to the operation and user interface of the mobile app as well as the experiment design. Experiment 3 assessed the users' ability to follow information change. This meant designing an experiment that could accommodate streams of related alerts, but the same randomization approach was retained. The experimental designs used are contrasted in Fig. 7.

4.3.1 Subject Recruitment and Participation for Experiment 1. Experiment 1 was conducted with volunteers from Carnegie Mellon University's Silicon Valley campus. Subjects included students, faculty and staff. It was only targeted for Android phone users since our testbed at that time only supported this platform. Ten days before the experiment start, the research team conducted an online pre-trial survey of the campus population to assess smart phone usage and determine eligible users with Android phones. 56 responses were received, of which 50% were Android users and 38% were iOS users. Since the experiment needed the subjects to use data and location services, the eligibility survey also inquired about data plans and usage patterns. 30% of the respondents

Table 3. Summary of Experiments: Experiment Locations, Mobile Platforms Used by Subjects, Distribution of Subjects, Distribution of Alerts, Features Tested, Characteristics Tested, and Outcomes Evaluated. The cross ("X") indicates that the enhancement feature, alert characteristic, or outcome in the corresponding row was evaluated during the experiment in the corresponding column. If the cross missing, the feature, characteristic, or outcome was omitted in the corresponding experiment. Certain outcomes were only applicable to certain tested features: these are indicated in brackets next the cross, where applicable.

	Trial 1	**Trial 2**	
Locations	Silicon Valley (SV)	Silicon Valley (SV) and Pittsburgh (Pgh)	Silicon Valley (SV) Pittsburgh (Pgh)
Experiments	Experiment 1	Experiment 2	Experiment 3
Platforms	Android	Android	iOS
# Subjects	52	SV | Pgh 42 | 46	SV | Pgh 54 | 43
# Alerts Issued per Subject	24	27	30
Enhancement Features Tested			
Long Message	X	X	
High-Information Map	X	X	
Geo-Targeting	X	X	
External Link	X	X	
Text-to-Speech	X	X	
Location History		X	
Location Prediction		X	
Situation Digest			X
Alert Characteristics Tested			
Timing	X	X	
Update Alert	X		
Alert- or Scenario-Based Outcomes			
Understanding	X	X	X (via Situational Awareness)
Relevance	X	X	
Hindsight Relevance	X (Geo-Targeting)	X (Geo-Targeting)	
Annoyance	X	X	
Actionability	X	X	X
Milling Behavior	X	X	
Adequacy	X	X	
Usefulness	X (High-Information Map and External Link)	X (High-Information Map and External Link)	
Situational Awareness			X
Response Delay		X (Text-to-Speech)	X
Response Rate		X (Text-to-Speech)	
Outcomes Related to Overall Perceptions			
WEA Benefits	X	X	
Feature Preference	X (High-Information Map and External Link)	X (High-Information Map, External Link, Location History, Location Prediction)	X



(a) Situation Digest View Display for iOS Users. The display supports situational awareness using structured alert information that aggregates content and meta data from a stream of related messages for an active emergency. Key elements are highlighted, e.g. immediacy, alert nature, and headline. Icons highlighting alert severity and nature are meant to support understanding and actionability. Other active emergency incidents are accessible by swiping left/right.

(b) Implementation of the WEA+ Digest View. The digest view supports situational awareness through an overlay mechanism to aggregate information from a stream of related alerts. Multiple incoming alerts in CAP format are received with updates to only certain fields (represented by transparency). At any point in time, the digest view shows a composition of the most recent and still valid fields from all messages in the stream. Only digested views (bottom images) are shown.

Fig. 6. WEA+ Alert and Digest View for iOS Users

indicated they would volunteer for a research experiment with smart phones and 52% said they might volunteer. After the eligibility survey, posters, on campus email lists and word of mouth officially advertised the experiment.

A kickoff event was planned to launch the experiment. A week before the kickoff even, an online signup form was deployed to formally solicit volunteers. 43 positive responses were received, with two volunteers indicating they would need loaner Android phones. To increase participation, and anticipating further requests for loaner phones at the kickoff, the research team decided to prepare 25 Android phones to loan to prospective volunteers. On the evening of day 1 of the experiment, the kickoff event was held during which the research team explained the project goals and the experiment to the attendees; explained how privacy, security and confidentiality would be handled; demonstrated the WEA+ Android app; completed the sign-up procedures including signing informed consent forms; gave instructions; and helped the volunteers install the WEA+ app and set up their phones. Test alerts were sent to test the WEA+ app and allowed the participants to become familiar with the app. About 60 volunteers attended the kickoff event and 43 signed up on the spot. The research team loaned 15 Android phones with SIM cards and data plans to volunteers who did not have Android phones. An additional nine volunteers who could not attend the kickoff event signed up on day 2, bringing the total number of subjects to 52 (six faculty/staff and 46 graduate students). One of these volunteers' phones failed to register with the WEA+ server so we collected data from a total of 51 phones over six days. This corresponds to a 16 percent participation rate from the sampled population of 324 students, faculty, scientists and administrative staff. Based on active phones detected from day to day, actual participation peaked at 48 during the trial. Most users were active every day, but some dropped out temporarily on certain days.

The participation instructions asked the subjects to keep their phones on and data and location services enabled throughout the experiment. Subjects were also asked to charge their phones each night. Installation instructions, the informed consent form and other trial-related information were posted at the project website (weacmu.org). During the trial, the project website was updated with fresh posts to keep the participants informed and motivated. During Experiment 1, subject participation levels stayed relatively steady, peaking around the second and third day. We tracked participation starting the day after the kick-off event and stopped on the last day of the experiment (day 8).

4.3.2 Subject Recruitment and Participation for Experiments 2 and 3. Experiments 2 and 3 followed a similar preparations schedule to Experiment 1, but was conducted simultaneously in two separate locations on two different smart phone platforms to test two disjoint sets of features. Experiment 2 was a partial, but larger replication of Experiment 1 and used a new version of the WEA+ Android app. Experiment 3 was new and tested a different paradigm via a new mobile WEA+ app deployed on iOS for iPhone users. Like in Experiment 1, Silicon Valley subjects were solicited from CMU's local campus and targeted students, staff, scientists and faculty. We did not deem this to be sufficient because of repeat targeting of the same population, however. To extend the sample, we mobilized our local emergency services and NASA Ames Research Center contacts to recruit extra subjects through an email, flyer and word-of-mouth campaign. Two kickoff/information meetings were held, one at the CMU campus the evening of the trial start and another at a Community Emergency Response Team (CERT) training event in neighboring Sunnyvale the week before. We reached out to Amateur Radio Emergency Services (ARES) volunteers as well. We changed our signup protocol to allow online signup with approval from CMU's Institutional Review Board.

The project website was updated to include an online signup form, a complete set of written instructions, all information packages (consent information and privacy/security policy), and instructional videos for both versions of the WEA+ app. A research team member also attended a Red Cross event the week before to distribute flyers and advertise the trial. In Pittsburgh, subjects were recruited from CMU main campus population, including the Software Engineering Institute, through an email and flyer campaign. A separate kickoff/information meeting was held on campus in Pittsburgh the evening of the experiments' start. One research team member was based in Pittsburgh and another traveled to Pittsburgh to help with logistics, recruitment and the kickoff event.

These additional recruitment activities, efforts and methods resulted in an increase in sample size compared to Experiment 1. A total of 185 volunteers signed up to be subjects in Experiments 2 and 3, of which 96 were in Silicon Valley and 89 were in Pittsburgh. Of the 96 Silicon Valley subjects, 42 were Android users and assigned to Experiment 2 and 54 were iOS users and assigned to Experiment 3. 50% were CMU students, faculty or staff and 50% had non-CMU affiliations (NASA, CERT volunteers, ARES volunteers, employees of various local emergency services organizations, and city employees including the cities of Mountain View, Palo Alto and Sunnyvale). Of the 89 Pittsburgh subjects, 46 were Android users and assigned to Experiment 2 and 43 were iOS users and assigned to Experiment 3. All Pittsburgh subjects had CMU affiliations. Since our Pittsburgh email campaigns specifically targeted Electrical and Computer Engineering and Computer Science populations, we estimate that most subjects belonged to these departments. In both locations, CMU subjects were predominantly students as opposed to staff or faculty. Across both locations, we were able to recruit 88 subjects for Experiment 2 and 98 for Experiment 3. Despite this higher than expected signup rate, actual participation varied from day to day, and not all users were active each day. Participation peaked at 66 for Experiment 2 and 65 for Experiment 3. The number of actual registered phones from which we were able to collect data was 72 for Experiment 2 and 98 for Experiment 3. Since we could support two major smart phone platforms, no loaner phones were used.

A detailed Frequently Asked Questions page was created and posted at the project website. All trial-related information, including participation and installation instructions, instructional videos, forms, confidentiality, privacy and security information were posted at the website. The website was updated regularly with fresh posts to keep the subjects engaged and informed. As in Experiment 1, subject participation during Experiments 2 and 3 were steady, again peaking around the second and third day.

4.3.3 Design of Experiments 1 and 2. For these experiments, we designed fictitious, but realistic emergency scenarios. Within a span of eight days, 24 alerts were issued to a total of 52 subjects in Silicon Valley in Experiment 1. In Experiment 2, 54 alerts were issued to a total of 88 subjects in both Silicon Valley and Pittsburgh, again within a span of eight days. We assessed the effectiveness of the WEA features tested by measuring and comparing the responses of the subjects to alerts issued in real-time. We used randomized A/B testing to evaluate each feature as explained above.

The alerts were modeled after real WEA messages based on examining a small corpus of about 120 non-weather alerts from various sources and sampling a large corpus of about 10,000 weather-related alerts from the National Weather Service. We mimicked the type of emergencies, message composition, and content of the samples examined, sometimes slightly perturbing the messages to fit the geography and context of the experiment locations (Pittsburgh and Silicon Valley). We also paid attention to established guidelines [23]. To avoid confusion with real WEA alerts, we prefixed each message with the text "THIS IS A TEST ALERT". The experimental protocol was approved by CMU's Institutional Research Board. The alerts' realism was evaluated in the final questionnaire and found to be reasonable by the subjects. A sample of the alert messages used is provided in the Appendix (see Table 9 for Experiment 1, 10 for Experiment 2, and 11 for Experiment 3).

To quantify public response outcomes, we asked the subjects to send feedback immediately after each alert. Feedback was given by clicking a button on the alert screen (Fig. 5a), and directing the user to a small set of multiple choice questions that they could quickly answer on their mobile device. To compare the tested feature relative to the control feature, a number of outcomes were studied: (a) *Understanding*: Was the alert easy to understand? (b) *Relevance*: Was the alert relevant to the recipient given the recipient's context? (c) *Annoyance*: Would the alert annoy the recipient in a similar real emergency situation? (d) *Actionability*: Would the alert prompt the recipient to take protective action in a similar real emergency situation? (e) *Milling Behavior*: Would the alert prompt the recipient to seek confirmation from alternative sources? (f) *Adequacy*: Does the alert contain sufficient information for the recipient to assess the situation? (g) *Usefulness*: Would the recipient find the alert useful in a similar real emergency situation? At the end of each experiment, the WEA+ app also solicited the subjects' overall impressions by a final questionnaire.

Experiment 2 is essentially a multi-location replication of experiment 1, with a few extra features and measurements. Experiment 1 was deployed in Silicon Valley, where Experiment 2 was deployed in both Silicon Valley and Pittsburgh with new subjects about three months after the conclusion of Experiment 1. Experiment 2 also used different alerts specialized to the new location and to add more variation, although the alerts were similar to those of Experiment 1 in that they were modeled based on real alerts from our corpus and composed using the same guidelines. We conducted Experiment 2 to increase the sample size and reconfirm the results of Experiment 1. We also added two extra features that were not tested in Experiment 1, namely Location History and Location Prediction, and measured response delay and response rate for the Text-to-Speech feature to better understand this feature. Both Experiments 1 and 2 were deployed on Android phones.

4.3.4 Design of Experiment 3. Experiment 3 had a slightly different design to test a completely different, more complex feature, the Situation Digest View, that was not tested in Experiments 1 and 2. It was deployed on an iOS phones to a different group of subjects than Experiments 1 and 2, but was conducted simultaneously with Experiment 2.

In this experiment, alerts were issued to 93 participants, again in a span of eight days. However this experiment was intended to test the subjects' understanding facing a stream of inter-related messages using a specific overlay-based presentation mechanism called the Situation Digest View. This meant that the users were exposed to a set of evolving emergency scenarios each of which ran like a story, with a series of alerts that overrode,

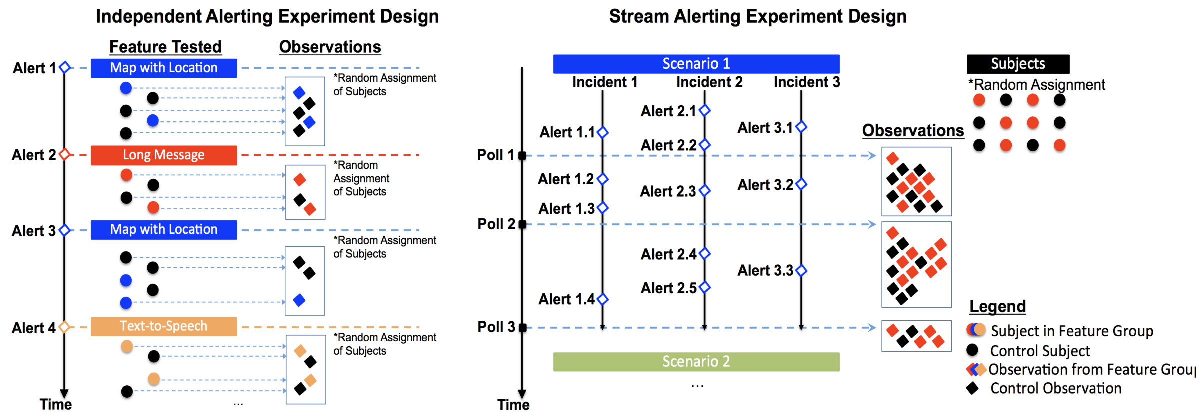


Fig. 7. Experimental Design for Public Usability Trial. Left: In experiments requiring independent alerting (Experiments 1 and 2), alerts were issued to a population randomly divided on-the-fly into two groups: the feature group received an enhanced alert and the control group a regular alert. Each subject provided feedback for every alert received. Each alert tested a single feature. Right: In the stream alerting experiment (Experiment 3), complex scenarios were created where multiple incidents would occur simultaneously and get updated in an interleaved fashion. Each alert was attached to a particular incident. At pre-determined times a *poll* would be issued, assessing the subjects' understanding of the scenario at hand. The control (situation digest view) and feature (ordinary WEA view) groups were randomly populated at the beginning of each scenario in a similar way to Experiments 1 and 2, but the groups were maintained until the next scenario.

added to, or updated previous alerts. During each of the scenarios we tested the users' understanding by sending a short poll at pre-determined times. For each scenario, the subjects were evenly and randomly divided into two groups similar to the other experiments: the treatment group received the alerts via the situation digest view, and the control group received them as an ordinary WEA message. However, in this experiment, understanding was measured based on the correctness of the subjects' answers to specific questions pertaining to an ongoing incident. The proportions of correct and partially correct answers were then compared. The single high-level outcome measure of this experiment was *situational awareness*, which objectively captured how well were the subjects were able to assimilate the evolving information given to them over time.

As in Experiments 1 and 2, the alerts were modeled after real WEA messages used in real emergencies and message composition was based on established guidelines [23]. However in one scenario used during the final days of the experiment (the alien invasion scenario), we relied on a completely fictitious, fantasy-like story line. This choice was deliberate because the scenario relied on total suspension of disbelief, unlike the other scenarios that involved plausible events consistent with the geography and context of the experiment locations. We used total suspension of disbelief to increase motivation and prevent too much focus on local details. But even in these cases, the form and language of the alerts mimicked real examples from our corpus. Again, to avoid confusion with real WEA alerts, we prefixed each message with the text "THIS IS A TEST ALERT". The experimental protocol was approved by CMU's Institutional Research Board.

5 RESULTS

We discuss the experimental results and the main findings of the Public Usability Trial phase in this section. Responses of subjects who participated in usability trials is summarized in Table: 4 and Table: 5. The important findings are summarized in Table: 6. The Appendix provides tabulated results for all experiments (see Tables 12 to 22), including frequencies, statistical tests conducted, whether the null hypotheses were rejected at the adopted significance level, as well as effect sizes. The results are pooled for Experiments 1 and 2, which tested the same set of enhancements.

5.1 Data Analysis Approach and Conventions

The majority of the trial data collected were responses to multiple-choice feedback questions and polls issued after alerts, in-between alerts, or at the end of an experiment. These questions and polls were intended to measure a set of outcome constructs, or dependent variables. The outcome construct levels represented in the questions and polls were of categorical, making it unsound to employ an analysis approach based on differences in mean values. The categorical, survey-like nature of the data made it most amenable to an analysis based on frequencies. The independent variables pertaining to response rate and response delay were the only exceptions to this.

We therefore used the standard Chi-square independence test [29] for all evaluated enhancements with the null hypothesis that the evaluated enhancement was independent of the response distribution. Thus the null hypothesis stated that the resulting outcome construct, or response variable, levels were independent of the value of the input variable, or group affiliation determined by the presence or absence of a tested factor. The Chi-square independence test is a non-parametric test. We compared the response variable's expected conditional frequency distribution inferred from the data to the actual, observed conditional frequency distribution. In such data, deviations from the expected behavior should follow a Chi-square distribution. If according to the underlying Chi-square distribution, the differences between expected and observed response level frequencies could not be simply due to chance with respect a selected confidence level, the null hypothesis was rejected and the alternative hypothesis that the response depends on the input was accepted.

To analyze responses to questions directly asking which option a subject preferred among a given set of options, the standard Chi-square goodness-of-fit test was used. This helped us determine whether one option significantly dominated others. The goodness-of-fit test is statistically similar to the independence test, except that the observed frequencies are not conditional on any input variable, and therefore the expected frequencies of the response levels are explicitly defined rather than inferred from the data. We tested the null the hypothesis that the observed unconditional frequencies matched the expected unconditional frequencies, and whether any differences were likely to be due to chance with respect to a selected confidence level. We adopted a uniform distribution for the expected frequencies of the response levels, rejecting the null hypothesis when the differences could not be due to chance according to the underlying Chi-square distribution at the selected significance level. The alternative hypothesis was that the responses were biased toward a specific subset of response levels. The uniform distribution was a reasonable choice given that we did not have any prior knowledge or expectations regarding the frequencies of the response levels; we assumed that if the subjects were on average indifferent about a probed issue, their responses would be distributed evenly among the presented answer choices.

In all Chi-square tests, the selected alpha level was 0.05. We selected this alpha level because the response variables (with the exception of subject's response time, which is not discussed in this paper, and the Situation Digest View in Experiment 3) capture perceived effectiveness according to a given criterion: they depend on human factors. We considered a test result to be significant when the p-value was below 0.05, rejecting the underlying null hypothesis. We considered a test result to be highly significant when the p-value was below 0.01. In discussing the findings below for each tested factor, we stick to this convention.

As per the Chi-square test, statistical significance was evaluated by comparing the Chi-square statistic of an analyzed outcome to the corresponding threshold value at the desired alpha level of the underlying Chi-square distribution. The Chi-square statistic is the sum of the squares of the relative deviations from the expected frequencies. In Chi-square independence tests, the tested factor (the input variable) always had two levels represented by the presence or absence of an enhancement feature or alert characteristic. There could be two or more categorical response levels (such as "Yes," "Unsure" and "No," or "Correct," "Partially Correct", and "Incorrect").

The number of levels of the input and response variables determined the dimensions of the underlying contingency table and the degree of freedom (df) of the underlying Chi-square distribution. In Chi-square goodness of fit tests, the number of response levels (categories) alone determined the degree of freedom (df) of the underlying Chi-square distribution since the contingency table always had a single column. In all independence tests in the analysis, the df values were either 1 or 2, as determined the number of rows and columns of the underlying contingency tables. In all goodness-of-fit tests, the df value was 3 corresponding to three response levels. The critical values of the Chi-square statistic at an alpha level of 0.05 are 3.84, 5.99, and 7.82 for df = 1, 2, and 3, respectively; and at an alpha level of 0.01, the critical values are 6.64, 9.21, and 11.3 for df = 1, 2, and 3, respectively.

Statistical significance is only one component of hypothesis testing, related to the odds of an effect being real. Equally important is the magnitude of the effect, or effect size (Cohen 1992). Effect size answers the question: how pronounced is the effect in the response? For the independence test, effect size quantifies the magnitude of the differences in responses when a tested factor is present and absent. For the goodness of fit test, it quantifies the deviation from the expected distribution. In both cases, the larger the effect size, the higher the tested construct's potential impact in practice, provided that the effect is also statistical significant.

In all Chi-square tests, we measured effect size in two different ways: (1) using Cramer's V statistic (C.V) for independence tests and the Phi statistic for goodness-of-fit tests [7]; and (2) using odds ratio (O.R.) for both kinds of tests. C.V and Phi represent theoretical measures, with widely accepted rules for interpreting them, whereas the O.R. is a practical measure, whose interpretation is more context-dependent. O.R. was calculated only for results that were statistically significant at the chosen alpha level.

Given the underlying degrees of freedom, the theoretical effect sizes were interpreted as follows [11]: (a) Very small: C.V smaller than 0.1; (b) Small: C.V larger than or equal to 0.1 and less than 0.3; (c) Medium: C.V larger than or equal to 0.3 and less than 0.5; and (d) Large: C.V larger than or equal to 0.5. An effect size between 0.2 and 0.3 as measured by C.V is considered normal in studies dealing with human behavior where outcomes might be affected by multiple uncontrolled factors [11]. This was the case in our experiments.

Phi and C.V were calculated as follows:

$$\text{Phi} = \sqrt{\chi^2/n} \tag{1}$$

$$\text{C.V} = \sqrt{\chi^2/n(k-1)} \tag{2}$$

where χ^2 is the Chi-square statistic, n is the number of observations, and k is smaller of the number of levels of the input variable (number of columns in the contingency table) and the number of levels of the response variables (number of row in the contingency table). The value of k was 2 in all independence tests; therefore, Phi and C.V. coincided in all cases.

As for the O.R., the odds of a positive response for a given group was defined as the ratio of the frequency of a positive response (for example, corresponding to the level "Yes") in that group to the frequency of a negative response (for example, corresponding to the levels "No" and "Unsure" together) in that group. In all tested factors except the Situation Digest view, the mid-level responses were combined with the most negative level in calculating O.R. In evaluating situational awareness with the Situation Digest view, the middle response level (Partially Incorrect) was ignored in the calculation of O.R. and only the absolute positive level (Correct) and absolute negative level (Wrong) were considered. Based on the scheme adopted, the O.R. for a tested factor (an alert feature or characteristic) was calculated as the odds of a positive response when the tested factor was present (the feature group or characteristic group) to the odds of a positive response when the tested factor was absent (the control group). Thus the closer the O.R. to 1, the smaller the effect size was. The interpretation of the magnitude of O.R. depended on the tested factor's context.

Table 4. Summary of Results of Public Usability Trial Experiments 1 and 2. For each tested feature, subjects' responses pertaining to a set of outcomes (Understandability, Relevance, Annoyance, Actionability, Milling Behavior, and Adequacy) were solicited immediately after each alert using a set of questions displayed on their phones asking the subjects whether the alert aligned with an outcome. The responses ('Yes','No', or an intermediate or neutral response) to each outcome question were aggregated across all alerts. The table shows the total counts selected by the subjects in the contol and treatment groups of each tested factor. Factors correspond to WEA enhancement features, which are represented by the different rows.

<table>
<tr><th>Outcome / Tested Factor</th><th>Understanding (No|Partially|Yes)</th><th>Relevance (No|NotSure|Yes)</th><th>Annoyance (No|Somewhat|Yes)</th><th>Actionability (No|NotSure|Yes)</th><th>Milling Behavior (No|NotSure|Yes)</th><th>Adequacy (No|NotSure|Yes)</th></tr>
<tr><td colspan="7">WEA Enhancement Features</td></tr>
<tr><td>Without Long Message</td><td>(27|24|381)</td><td>(129|69|231)</td><td>(273|59|87)</td><td>(144|70|207)</td><td>(129|58|243)</td><td>(117|39|252)</td></tr>
<tr><td>With Long Message</td><td>(9|20|223)</td><td>(57|35|156)</td><td>(185|27|33)</td><td>(54|49|137)</td><td>(53|31|165)</td><td>(49|23|162)</td></tr>
<tr><td>Without High-Information Map</td><td>(26|29|425)</td><td>(134|74|269)</td><td>(310|67|92)</td><td>(152|88|230)</td><td>(127|71|279)</td><td>(125|44|285)</td></tr>
<tr><td>With High-Information Map</td><td>(12|16|355)</td><td>(97|29|243)</td><td>(252|52|73)</td><td>(124|63|185)</td><td>(77|43|153)</td><td>(39|32|195)</td></tr>
<tr><td>Without Geo-Targeting</td><td>(20|16|300)</td><td>(126|44|163)</td><td>(212|50|66)</td><td>(127|55|147)</td><td>(122|47|164)</td><td>(85|28|206)</td></tr>
<tr><td>With Geo-Targeting</td><td>(3|2|96)</td><td>(11|9|79)</td><td>(74|8|16)</td><td>(25|16|58)</td><td>(28|12|47)</td><td>(13|8|62)</td></tr>
<tr><td>Without External Link</td><td>(22|17|260)</td><td>(83|39|174)</td><td>(209|26|58)</td><td>(101|31|163)</td><td>(85|30|167)</td><td>(86|28|163)</td></tr>
<tr><td>With External Link</td><td>(12|9|148)</td><td>(50|19|100)</td><td>(107|22|39)</td><td>(48|27|92)</td><td>(25|10|66)</td><td>(16|9|75)</td></tr>
<tr><td>Without Text-to-Speech</td><td>(29|31|314)</td><td>(106|56|193)</td><td>(232|39|77)</td><td>(118|57|171)</td><td>(92|43|172)</td><td>(70|31|194)</td></tr>
<tr><td>With Text-to-Speech</td><td>(16|24|375)</td><td>(105|63|242)</td><td>(281|55|65)</td><td>(103|80|219)</td><td>(97|53|262)</td><td>(102|37|245)</td></tr>
<tr><td>Without Location History</td><td>(7|6|256)</td><td>(92|24|153)</td><td>(178|37|53)</td><td>(98|39|131)</td><td>(89|34|147)</td><td>(32|20|209)</td></tr>
<tr><td>With Location History</td><td>(5|3|123)</td><td>(16|8|106)</td><td>(92|13|26)</td><td>(25|17|88)</td><td>(30|14|86)</td><td>(31|8|88)</td></tr>
<tr><td>Without Location Prediction</td><td>(3|0|70)</td><td>(26|5|43)</td><td>(48|6|20)</td><td>(26|7|41)</td><td>(31|2|41)</td><td>(7|2|64)</td></tr>
<tr><td>With Location Prediction</td><td>(0|1|46)</td><td>(13|1|34)</td><td>(31|4|13)</td><td>(17|1|30)</td><td>(13|2|33)</td><td>(2|4|41)</td></tr>
</table>

5.2 Analysis of Subjects' Responses

Public Usability Trial experiments collected perceptions and preferences of users via real-time feedback questions. Subjects' responses are summarized in Tables 4 and 5. We analyzed the responses to accept or reject an underlying null hypothesis to see whether a WEA enhancement affected the responses with respect to several criteria, and to gauge the magnitude of the effect the enhancement had via C.V, Phi, and O.R. In this section, we describe these calculations in detail using one tested enhancement feature as an example. The results of all statistical tests are available in the Appendix (see Tables 12 to 22.)

Let's take the test of short vs. long messages as an example tested feature. For this feature, a subset of subjects was selected as the control group and received short messages (let's call this Group 0). The rest of the subjects constituted the treatment group and received long messages (let's call this Group 1). After receiving an alert in either format, subjects were immediately asked to answer a feedback question on their phones. Suppose the question was 'Did you understand this alert message?' For this question, users could respond by choosing 'No'(n), 'Partially'(p), or 'Yes' (y). Suppose the number of subjects in Group 0 who responded 'No' were N_{0n}, 'Yes' were N_{0y}, and 'Partially' were N_{0p}. Similarly, for subjects in Group 1, suppose the respective numbers were N_{1n}, N_{1y}, and N_{1p}. Thus, the total number of responses in Group 0 were $N_{0n} + N_{0y} + N_{0p}$, and in Group 1 were $N_{1n} + N_{1y} + N_{1p}$.

Table 5. Summary of Users' Feedback for Situation Awareness Experiment (Experiment 3). For each scenario, the subjects' responses to a series of polls were objectively evaluated as 'Wrong', 'P.Correct' (Partially Correct), or 'Correct'. The rows shows the total number of subjects that fell into each response category for the treatment group (With Situation Digest View) and the control group (Without Situation Digest). The columns indicate the outcome evaluated: understanding of nature of emergency, action to take, and identification of the emergency's immediacy.

Scenario \ Outcome	**Understanding of Type of Emergency** (Wrong|P. Correct|Correct)	**Understanding of the Action to Perform** (Wrong|P. Correct|Correct)	**Identification of the Immediacy of Emergency** (Wrong|P. Correct|Correct)
Earthquake with Plume			
Without Situation Digest view	(50|0|51)	(45|0|56)	(70|0|31)
With Situation Digest view	(20|0|79)	(26|0|73)	(41|0|58)
Random Alerts			
Without Situation Digest view	(25|0|42)	(41|0|26)	(37|0|30)
With Situation Digest view	(11|0|55)	(18|0|48)	(35|0|31)
Severe Weather			
Without Situation Digest view	(46|73|14)	(66|54|15)	(89|19|81)
With Situation Digest view	(29|77|14)	(27|67|29)	(70|24|81)
Alien Catastrophe			
Without Situation Digest view	(118|101|1)	(141|79|3)	(132|100|59)
With Situation Digest view	(78|160|18)	(127|109|15)	(154|98|89)
Bad Weather			
Without Situation Digest view	(6|0|48)	(30|0|265)	(30|0|255)
With Situation Digest view	(4|0|48)	(34|0|257)	(34|0|257)
All Scenarios			
Without Situation Digest view	(245|174|156)	(316|133|131)	(358|119|466)
With Situation Digest view	(142|237|214)	(220|176|195)	(334|122|516)

We used these frequencies in the Chi-square goodness-of-fit test. We first calculated the underlying χ^2 statistic. We then compared this value with the critical χ^2 value for the corresponding alpha level to maintain or reject the null hypothesis.

The Chi-square statistic was calculated using equation 3, where $i \in 0, 1$ represents the group identifier (Group 0 or 1) and the j are possible responses n, p, or y. The E_i are the expected number of observations and the O_i are the actual number of observations.

$$\chi^2_{i,j} = \left(\frac{(O_{i,j} - E_{i,j})^2}{E_{i,j}}\right) \tag{3}$$

Since we did not have the expected number of observations from a benchmark, we estimated them by distributing the total number of each answer option proportionally to the two groups, using equation 4:

$$E_{ij} = \sum_i N_{ij} * \frac{\sum_j N_{ij}}{\sum_{ij} N_{ij}} \tag{4}$$

For the example question assessing the understanding outcome of long vs. short alerts, the cell at the intersection of the row 'Without Long Message / With Long Message' and column 'Better Understanding' of Table 4 gave the

Table 6. Summary of Public Usability Trial Findings and Implications.

WEA enhancement feature	Evidence in favor / Potential	Positively impacted outcome constructs	Negatively impacted outcome constructs	Implications
On-phone Geo-filtering	Strong / Significant	1. **Relevance** 2. Actionability 3. Adequacy	Hindsight Relevance	Requires embedding geo-target into alert at origin and location services to be enabled on phone for filtering. Straightforward to implement on phone with no UI changes. Invisible to alert originator and recipient.
Situation Digest	Strong / Significant	1. **Situational Awareness**		Requires new alert creation process/tools and use of CAP meta-data. Meta-data encoded and bundled with alert content for transmission, and unbundled or decoded on phone. Visible to both alert originator and recipient. Changes phone functionality, including UI. No changes to the WEA network architecture are necessary.
Location History	Strong / Significant	1. **Relevance** 2. Usefulness 3. Actionability 4. Adequacy	Hindsight Relevance	Used with geo-targeting. Straightforward implementation on phone possible with no UI changes. Can be invisible to alert originator and recipient. Effectiveness depends on prediction algorithm.
High-Information Map	Moderate / Medium	1. **Usefulness** 2. Relevance 3. Adequacy		Used with geo-targeting. Implementation on phone is straightforward, but requires new UI. Pre-cached maps on phones alleviate network congestion concerns.
External Link	Weak / Low	1. **Usefulness** 2. Adequacy		Requires only minor changes to phone app to make links clickable. Policy change is necessary to allow (restricted) links in alerts. Network congestion may be a current concern.
Location Prediction	Weak / Low	1. **Usefulness**	Hindsight Relevance	Used with geo-targeting. Implementation on phone possible with no UI changes. Prediction algorithm can be complex. Can be invisible to alert originator and recipient. Effectiveness heavily depends on prediction algorithm.
Long Message	Weak / Low	1. Actionability 2. Annoyance 3. Milling Behavior		No changes in alert creation side beyond alert originator training and modifications to existing tools to enable longer message construction. Requires policy change and Cooperation from wireless carriers is required.
Text-to-Speech	Weak / Unknown	1. Response Rate 2. Understanding 3. Actionability	Response time (short-term)	No changes in alert creation beyond avoidance in alert text of content not easily converted to speech. Simple changes to phone app.

number of observations for each possible answer option. We thus had:

$$E_{0\mathrm{n}} = (N_{0\mathrm{n}} + N_{1\mathrm{n}}) * \frac{N_{0\mathrm{n}} + N_{0\mathrm{y}} + N_{0\mathrm{p}}}{N_{0\mathrm{n}} + N_{0\mathrm{y}} + N_{0\mathrm{p}} + N_{1\mathrm{n}} + N_{1\mathrm{y}} + N_{1\mathrm{p}}}$$
$$= \frac{(27 + 9) * (27 + 24 + 381)}{27 + 9 + 24 + 20 + 381 + 223} = \frac{36 * 432}{684} = 22.74$$

Therefore, $\chi^2_{0\mathrm{n}}$ was $(E_{0\mathrm{n}} - N_{0\mathrm{n}})^2/E_{0\mathrm{n}} = 0.80$ for Group 0 and response level n. We calculated the remaining χ^2_{ij} similarly, and aggregated them to obtain the Chi-square statistic:

$$\chi^2 = \sum_{j=\mathrm{n,p,y}} \sum_{i=0,1} \left(\chi^2_{ij}\right) = 3.57$$

Finally, to accept or reject the null hypothesis, we compared the value of χ^2, as calculated above, to its critical value at an alpha level of 0.05 (0.01, respectively) for df = 2. The results for each tested feature/factor are provided in the Appendix. In the Appendix, tests that are significant at an alpha level of 0.05 (significance) and 0.01 (strong significance) are differentiated in Tables 12 to 22.

We summarize important findings next.

5.3 Important Findings

5.3.1 Alerts with High-Information Maps. To verify the utility of including high-information maps, we implemented map-based alerts in WEA+ (Fig. 5a). In addition to displaying the alert text and alert type, the WEA+ mobile app showed the recipient a local map of the area, the affected region (using a polygon), the location of the user, and the time-period of the alert. We compared alerts displaying this visual to those that didn't. Based on the analysis of the subjects' responses, we found that alerts with high-information maps increased the relevance of the alerts to the recipients. Maps also appeared to affect the information content of alert messages, as measured by adequacy. Both improvements were significant at an alpha level of 0.01. When explicitly asked whether the map in the alert just shown was useful, subjects overwhelmingly responded that they were. This result was also highly significant at an alpha level of 0.01. We did not find any evidence of a positive or negative effect of high-information maps on actionability, annoyance, and milling behavior.

5.3.2 Alerts with On-Phone Geo-Filtering . Since most modern phones are equipped with capable GPS receivers, it is possible to use client-side filtering to discard alerts that are not in the vicinity of a user's current location. We compared responses to geo-targeted alerts with on-phone geo-filtering to responses to generally targeted alerts. We found that precise, fine-grained geo-filtering on the phone improved alert relevance to recipients. The improvement was highly significant at an alpha level of 0.01 with a near-medium theoretical effect size as measured by C.V and considerably large practical effect size as measured by O.R. (over three times improvement in the odds). Actionability was also better with geo-filtering, but this effect was not as strong (significant only at an alpha level of 0.05). We conclude that precision geo-filtering has a significantly positive impact on alert relevance and a small to moderate impact on actionability.

5.3.3 Alerts Filtered with Location History. Because users are likely to be interested in areas that they visit often, we wanted to evaluate the impact of using a user's past locations to filter alerts. When a geo-targeted alert was sent with the location-history feature turned on, the mobile application used the geo-target polygon embedded in the alert and location history stored on the phone to determine if the recipient had ever visited the alert region recently. If the recipient had visited the alert area in the recent past, the alert was shown; otherwise it was discarded. We compared geo-filtered alerts with and without the location history feature turned on. We found that location history improved an alert's relevance to recipients. The improvement was highly significant at an alpha level of 0.01 with a near-medium theoretical effect size as measured by C.V and considerably large practical effect size as measured by O.R. Actionability and adequacy also improved when geo-filtering was combined with location-history-based filtering. Both were significant at an alpha level of 0.01. We conclude that location history is a highly desirable enhancement that complements on-phone geo-filtering.

5.3.4 Update Alerts. In Experiment 1, using a small sample, we tested whether alerts that updated the status of an ongoing emergency (update alerts) caused a different reaction than alerts that first announced an emergency (initial alerts). We found that initial alerts were more likely to trigger information-seeking behavior as compared to update alerts (as measured by milling behavior outcome) and were more actionable. But update alerts performed better in terms of adequacy, which is a component of information content. The results were significant at an alpha level of 0.05, but the effect sizes were small-to-medium as measured by C.V. The superior adequacy of update alerts is positive, but this result is conditional on whether the improvement leads to correct situational awareness. The standard WEA service treats update alerts in an ad-hoc manner, and this may lead to confusion in complex, rapidly changing situations.

5.3.5 Alerts Using Situation Digest View. Our hypothesis was that alerting with explicit situational awareness is important for complex and evolving emergency situations [18]. We tested this hypothesis using the situation digest view shown in Fig. 6b. This mode of alerting was expected to result in a higher-level of awareness faster. In order to measure the efficacy of this concept, in Experiment 3, we compared the subjects' responses to changing information in multiple emergency scenarios. We assessed three distinct elements of situational awareness: *e.g.,* action to take, alert/emergency nature, and immediacy.

We found, as expected, that situational awareness improvement via a digest view over the current mode used in the WEA service was dependent on the complexity of the scenario presented. Overall (when data from all scenarios were pooled), the digest view significantly outperformed the regular view in all awareness outcomes (action, nature, and immediacy). In complex scenarios where multiple concurrent alerts needed to be sent in order to update the subjects, the digest view performed significantly better in improving awareness outcomes (except in one scenario with respect to immediacy), with a medium theoretical effect size as measured by C.V and a very large practical effect size as measured by O.R. In simple scenarios where a small number of alerts were sent, the digest view did not show significant improvements, but it didn't have a negative impact either. At the end of the Experiment 3, when we asked the subjects which view they preferred, a near-majority of the subjects responded with a preference for the digest view: only 14 percent preferred the regular WEA view and the remaining 38 percent were either indifferent or unsure. The results were significant with a large effect size. All statistical tests that were significant for the situational view were significant at an alpha level of 0.01. We conclude that supporting situational awareness through a mechanism like the digest view would be a worthwhile improvement for the WEA service.

5.3.6 Alerts Using Location Prediction. On-phone geo-filtering and location history use recipients' current and past location information to filter alerts, however, they do not consider their future location, *e.g.,* in the case of a citizen moving towards a fire region covered by an active alert. Location prediction uses a person's movement information (speed and direction) to predict their future presence in the alert region. We implemented and tested this feature in WEA+. We expected this feature to improve alert outcomes, however our results did not support this expectation. Beyond the subjects' perception of the usefulness of this feature, there was no significant effect on any of the outcomes evaluated.

5.3.7 Inclusion of External Link in Alerts. Currently, inclusion of clickable links to websites and references to social media tags are disallowed in WEA messages due to fear of possible network congestion. However, this concern will become less and less valid as cellular technology advances. External links may improve the information content of alert messages by allowing easy access to additional resources. We therefore implemented link inclusion as a WEA+ feature and tested it. Although the majority of the subjects found it a desirable addition (this result was highly significant at an alpha level of 0.01) after receiving alerts with links, other alert outcomes were not significantly affected when compared to alerts without links. External links improved the information content of alert messages as measured by adequacy (this result was significant at an alpha level of 0.01 in pooled observations, but not systematic across the two experiments), however the observed effect was small.

5.3.8 Alerts with Text-to-Speech (TTS). Recipients alerted using different tonal modes might react to alert messages differently and at different speeds and rates. We wanted to find out if there was a difference between alerts notifying the recipient using vibration and a ringtone and those notifying the recipient with vibration and the spoken alert text instead of a ringtone. There were no significant differences in any of the alert outcomes, except a mild positive impact on understandability and actionability, with a small effect size. We also measured response rate for this feature, and TTS-enhanced alerts had a significantly higher response rate than regular alerts. This result may be interpreted as a precursor to real actionability as opposed to perceived actionability. In conclusion, we think that TTS could be beneficial for actionability in proper contexts, *e.g.,* while driving, but this needs to be investigated further.

5.3.9 Long-Message Alerts. Current WEA service limits alert messages to 90 characters. We evaluated the potential benefits of longer messages by comparing alerts significantly longer than 90 characters to alerts that obey the current character limit. The results show that actionability and milling behavior were most affected by message length, in favor of long messages and the results were significant at an alpha level of 0.05, not particularly strong. The significance and size of the improvements were alert-specific. The message length did not affect understanding and adequacy outcomes.

5.4 Alert Timing

We compared responses to alerts sent at different times (early morning, working hours, late evening, and weekend) to see whether the timing affected any of the alert outcomes. We didn't find any differences. However, this alert characteristic is both demographic-specific and difficult to isolate without a large alert sample. It is possible that the sample size was too small to assess this characteristic.

5.4.1 Overall Impressions. At the end of experiment, a final questionnaire was sent to gauge the overall perceptions of the subjects. The questionnaire included two questions on the WEA benefits: (1) Do you believe wireless emergency alerts are useful? (2) Do you believe wireless emergency alerts could save lives? Subjects nearly unanimously felt that the WEA service was useful and could save lives.

6 LIMITATIONS

Convenience sampling poses the main threat to the external validity of the Public Usability Trial results. The subjects who participated in the experiments were volunteers who were largely recruited from two University campuses. They were predominantly technology savvy and comfortable using their smartphones' advanced capabilities. Therefore there is a risk that their responses to technology-based features may differ from those of the average citizen. Also self-selected volunteers tend to be more motivated than the general population. As an alleviating factor, selection bias applies equally to the two compared groups in an internally randomized design, which was the case in both trials.

In addition, the experiments relied on suspension of disbelief and alerts send under artificial emergency scenarios. People may behave differently in real-world situations when faced with real dangers to their safety and to their property. Short of staging actual emergencies or deceiving the subjects—neither of which would be safe—we cannot entirely eliminate this threat in a controlled study. In the post-trial questionnaires, over two-thirds of the subjects found the level of realism in the test alerts acceptable, which moderates this threat.

In Experiments 1 and 2, the outcomes were evaluated via self-assessment based on the subjects' responses to a set of questions posed after receiving each alert. Therefore we evaluated the perceived value of the tested enhancements with respect to a set of pre-determined constructs, raising an internal validity threat. For these experiments, although the outcomes were not validated by other, more objective means, we believe perceived value to be a relevant and important consideration for the public adoption and acceptance of the WEA service. In Experiment 3, the outcomes were evaluated by objectively measuring the correctness of the subjects answers, eliminating this threat.

The response rates within the Public Usability Trials (Experiments 1 to 3), in a great majority of the cases, were reasonable and within or above acceptable ranges of 30%-40% for internal surveys. For the surveys sent to all subjects at the end of each experiment, response rates ranged from 38% to 87%. It was 38% for Experiment 1 (20 out of 52 subjects), 87% for Experiment 2 (47 out of 54 subjects), and 61% for Experiment 3 (37 out of 60 subjects). For the just-in-time, on-phone feedback received after each alert, response rates depended on the feature tested, the number of feedback questions deployed for an alert, the time an alert was sent, and whether the subject's phone was active at the time an alert was issued (turned on with the appropriate permissions enabled to register the phone and receive the alerts). Number of active subject phones varied from day to day during the course of each experiment, averaging 87% (44 out of 52) for Experiment 1, 69% for Experiment 2 (61 out of 88), and 58% (57 out of 97) for Experiment 3. Maximum participation as measured by active phones on any given time was 94%, 91%, and 66% for Experiments 1, 2, and 3, respectively. Daily average response rates to alert feedback questions ranged from 31% to 76% for Experiment 1 and 65% to 77% for Experiment 3. However when average response rates to alert feedback questions are calculated for individually tested features across alerts that were used in testing a specific feature, they showed more variability. For example, for the High Information Map feature, it was 42% (343 actual responses vs. 810 expected for a set of feedback questions); for the Long Message feature, it was 70% (2444 actual responses vs. 350 expected); and the the External Link feature, it was 45% (148 actual responses vs. 328 expected). The lowest response rates recorded among all tested alert features were for location-aware ones—Geo-Targeting, Location History, and Location Prediction. This was because, by design, only a fraction of the subjects received the alerts depending on the area geo-targeted, their movement, and their location histories. The response rate in those cases could fall below 25% (e.g., it was 21% for Geo-Targeting, 99 actual responses vs. 472 expected, because many subjects were outside the targeted area and did not receive the alerts). This reduced the effective sample size when analyzing certain outcomes measures for these features.

Our focus in this work has been the WEA service and addressing its specific limitations. The WEA service is part of a larger landscape of emergency warning network and constitutes only one channel within the Integrated Public Alert and Warning System (IPAWS) [2]. We do not claim generalizability of all of our results to other pieces of IPAWS, which are subject to different constraints. It is conceivable that some of the results, especially those pertaining to geo-targeting and filtering of streamed messages, can be ported to other warning services as devices that are capable of receiving warnings via those services increase in their sophistication, for example by becoming location aware.

The WEA service is currently in use in the United States. Our results were derived specifically in this context with respect to national needs and capabilities, and relying on the opinions and experiences of US-based alert originators and behavioral patterns of US-based users. Thus the results' generalizability to other countries and contexts may be limited.

7 RELATED WORK

Many researchers have studied the role of information technology, particularly via social media, in improving response to crisis events, for example focusing on increased participation [28], effective messaging [30], predictive sociobehavioral analysis to improve response outcomes [27], and impact on situational awareness [33]. The proliferation of mobile phones have amplified the importance of information technology in such contexts. The WEA service occupies a special place in the landscape of available platforms reachable via mobile phones. This service is distinct from social media because it is uni-directional, has a much wider and uniform reach, messaging and use are carefully controlled by authorized personnel, and the technology behind it, while ubiquitous, is rather old, has been static for years, and is subject to many more constraints than what is possible with modern social media platforms. Our motiviation in this work was to improve the WEA service by pushing its boundaries and demonstrating the feasibility and effectiveness of several enhancements to better position it in the modern landscape of disaster communications.

Several previous works have reported on the effectiveness and adoption of the WEA service and similar alerting mechanisms and suggested strategies for possible improvements for WEA delivery. We summarize the most relevant ones here and explain their relationship to our work.

Research conducted at the Software Engineering Institute (SEI) [31, 34] identifies trust, both on the part of the alert originators as well as the public, as key in the success of the WEA service. Based on surveys of the public as well as alert originators, SEI researchers constructed a formal trust model, which was analyzed via simulation. They determined that maximizing the alert originators' use of the WEA service requires improving three key outcomes: appropriateness, availability, and effectiveness. These outcomes in turn depend on trust factors such as sourcing, frequency, and messaging. We did not set out to re-identify such factors that influence adoption, since the SEI work on this topic is pretty comprehensive. Instead we set out to identify and investigate a set of technical enhancements that could modernize the service. As a starting point of our study, we targeted alert originators because as experts, they have a deep understanding of the technical needs that could address the underlying factors. Such enhancements are difficult to evaluate in the hypothethical: they are best studied in a context in which the pubic could actually see and use them. So instead of presenting them in the abstract to a user base, we took a different approach. We implemented them in a testbed, and deployed them to live subjects.

SEI reports also suggest that the specificity of how alerts are targeted to an affected geographical region, or *geo-targeting*, is a critical component for building trust. The more precise this geo-targeting is, the more relevant the alerts are to the recipients, increasing their trust in the system. Other work on possible WEA improvements include a comprehensive study of WEA integration considerations, again by SEI, [5], which posited that the ubiquity of smartphones enable novel technical solutions both for improved geo-targeting and for addressing various other current WEA limitations. Although such improvements, including better geo-targeting, were posited, they were not tested in any realistic context. Thus we explored many of these suggestions in our research after validating them with alert originals in a pre-study.

The importance of accurate geo-targeting was also highlighted in several other reports. The Department of Homeland Security's (DHS) WEA service recommendations [4] and SEI's WEA best practice recommendations [23] conjecture that AOs will use the WEA service more extensively if alert messages can be better targeted to the size and location of the geographic region impacted by the emergency event. With respect to geo-targeting granularity, Nagele and Trainor [24] stated that being able to set an appropriate polygon size could be an important factor in improving public response to alerts. However, this approach would be useful only if the actual delivery mechanism respects the finer resolution of smaller targets, which we guaranteed in our work through client-side filtering. The precise, fine-grained implementation that we propose adds client-side filtering to that achieved by SMSCB at the level of cell towers: it is not meant to replace the standard base-station-based targeting, but rather to augment it.

The WEA service currently supports only text messages. The 2013 DHS report [4] recommended that WEA should also support richer media content in alerts, including both maps and other rich media forms such as images and audio. Inclusion of such artifacts could convey more information to the public about the situation and the required action. Location-based services are often most appealing when combined with maps that allow the users to visualize pertinent information in the proper context and in real time. Such maps could be called *high-information*. Though maps have been frequently used in mobile applications, their use in emergency alerts delivered to mobile devices is still not very common. A study by University of Maryland's National Consortium for the Study of Terrorism and Responses to Terrorism (UMD-START) [17] concluded that the inclusion of a high-information map specifying the alert region and the recipient's location could have a significant and positive effect on public response outcomes including interpretation and personalization, with a potential to improve protective action-taking. We leverage this result in our prototypical implementation of the enhanced WEA service.

The WEA service limitation of only supporting text messages was highlighted in other studies as well. Two National Research Council (NRC) reports [15], [14] and the 2013 DHS recommendations [4] pointed to the 90-character limitation of current WEA implementations as a hindrance, with the proposition that a modest increase in message length might be beneficial. The 2013 DHS report [4] recommended that WEA also support richer media content in alerts, including both maps and other rich media forms such as images and audio. Inclusion of such artifacts could convey more information to the public about the situation and the required action. We validated these assumptions in the pre-study with alert originators and later deployed related improvements in a realistic context to test them.

Some of the above-mentioned improvements were evaluated via human studies involving focus groups and surveys [17], where subjects are presented with different alternatives on paper in the context of an emergency scenario and express a preference. But these studies were not conducted in real time with using simulated emergency scenarios, which we undertook in our work.

The UMD-START report [17] concluded that visual stimuli including color, size, shape, bold font, iconography, sound, and the character of audible tones that indicate the arrival of a message could influence WEA message interpretation and subsequent message response, although the precise effects of such enhancements were unknown. In our work, we do not study visual elements and user-experience related factors because we scoped our study to focus technical enhancements.

Additionally, the 2013 DHS recommendations [4] as well as the 2011 NRC report [14] suggest enhancing WEA delivery by allowing the public to be notified when a WEA message is issued to their home area or to a recipient-specified area of interest. Although this technical assessment was suggested, it was never prototyped and tested. We investigate the effects of similar improvements in our study in the form of a set of context-aware WEA features which were implemented, deployed, and tested with live subjects.

The work reported here is based on the research detailed in a comprehensive technical report available from the Department of Homeland Security [16]. A summary of central location-based features (including different geo-targeting options—such as precision targeting and targeting assisted by location history–but excluding situational awareness and other alert characteristics) and the results pertaining to them were presented at the 2016 IEEE Symposium on Technologies for Homeland Security [22]. Iannuccu et al. [18] developed the original concept of *situational awareness* in the context of WEA. This concept was implemented by the Situation Digest View feature in the WEA+ testbed and tested in Experiment 3 of our study. Iannuccu et al. also explain on how to adapt and extend the Commercial Alerting Protocol (CAP) to support this paradigm in a real cell-broadcast environment. This publication provides a unified view of WEA enhancements tested and the results which includes design and results of the Alert Originator Requirements Study as a motivator and pre-cursor to the technical WEA work, a consolidated experimental design section and secondary results and implications of WEA experiments not discussed in earlier papers. We have also included the recommendations and future work that were re-framed after the recent Federal Communications Commission (FCC) ruling on WEA.

8 RECOMMENDATIONS

Our studies have led to insights about the perceived current value of WEA and the potential future value that could come from specific improvements. We summarize the central insights next along with associated recommendations.

8.0.1 Deep Integration of Location-based Context will Improve the WEA Service's Value. AOs believe that increased geo-targeting resolution and precision will significantly improve the effectiveness of the WEA service. Our experiments demonstrate that this is both feasible and beneficial. Precise geo-targeting through on-phone geo-filtering can be combined with other location-aware enhancements—such as location history—to add more value, as long the geo-target can be embedded in the alert. Using compression algorithms for boundary polygon encoding [20], such embedding is both feasible and efficient, even with current length limitations. The Public Usability Trial results provide strong evidence for acceptance and value for these new features. We recommend that WEA alert creation tools be modified to allow embedding boundary polygons used for geo-targeting into WEA messages. We further recommend that smartphone manufacturers build flexible mechanisms into future phones to take advantage of boundary polygon information and other user context information available within the phone in making the alert delivery decision locally.

8.0.2 Rich-Media Integration into WEA is a Question of How, not If. The Alert Originator Requirements Study provided support for the integration of rich media (such use of photos, maps, and carefully typeset and laid-out text as one would find in a well-crafted web page) into the WEA service. The usability trial results indicated that integrating maps enhanced with the alert region polygon and the recipient's location into WEA messages were perceived as highly desirable and has the potential improve alert outcomes significantly. The call to include rich media content in alert messages arises from the fundamentals of (a) widespread use of smartphones and (b) the pervasiveness of the World Wide Web and the ways in which the Internet sets the standard for how information is conveyed. When SMSCB was selected as the WEA information transport mechanism, cellular networks did not support broadcast of rich media. Since that time, advances in cellular network architecture (such as LTE broadcast and in-network content caching) have made rich media broadcast possible. Similarly, inclusion of maps will be facilitated by the availability of pre-cached, built-in maps in future generations of smartphones. In light of these facts and supporting study findings, we recommend a re-consideration of WEA at the level of network standards bodies and developers of WEA smartphone software to support different content forms.

8.0.3 WEA Will Benefit from a Transition From a Focus on Alerting to a Focus on Awareness. The WEA service's fundamental nature as a 90-character text message broadcasting service may work well in situations where the descriptions of the emergency events and the recommended preparedness actions are simple and relatively unchanging. However, for situations that unfold over time and in which instructions to the public may be revised in the span of minutes to hours with many different, possibly conflicting alerts and updates issued (*e.g.,* in the case of an earthquake that causes bridges to collapse, triggers fires or the release of hazardous materials, or requires management of changing evacuation protocols), the current service's means of presenting information to citizens may not be well suited. Errors from interpreting individual text-based alert messages and updates out of their original order may lead to serious consequences. We demonstrated that providing software on smartphones to digest sequences and sets of related WEA messages and to present the digested information as a situational awareness view resulted in significantly better understandability compared to the standard WEA presentation of alerts on the phone. This new way of structuring and viewing streams of alert messages represents a change in the role of WEA from a focus on alerting (sending many messages) to a focus on awareness (assisting the user by digesting what has been sent into a comprehensive, up-to-date view). The WEA service's implementation today does not support such digesting. To address this, we also developed the means by which this digesting capability can be retro-fitted to the current WEA architecture.

9 SUMMARY OF CONTRIBUTIONS

This paper presented a multi-methods study of possible enhancements to the WEA service. Its contributions are three-fold:

- Non-technical contributions: We analyzed the requirements of potential enhancements through a comprehensive survey of alert originators in the United States and in-depth interviews with select alert originators. Based on insights gathered from these two components, we identified a subset of technical improvements we could test within an emulation testbed to demonstrate both the feasibility and the effectiveness of the enhancements.
- Technical contributions: We built an emulation testbed with extensive instrumentation and automated user-feedback capabilities to test possible enhancements to the WEA service. We implemented the selected WEA enhancements within this testbed and deployed them to live users in three experiments. The creation of the testbed was a substantial software and system building effort, involving over ten graduate students and undergraduate interns. We analyzed quantitatively the data collected from these experiments. The data captured users' responses with respect to several pre-defined dimensions of effectiveness.
- Synthesis of technical and non-technical contributions: Results of the non-technical contributions and technical contributions were combined and triangulated to generate a set of recommendations, subsequently influencing the Federal Communication Commission's perspective about the future of the WEA service within the United States.

To our knowledge, our work constitutes the first work of its kind that investigated improvements to the WEA service using a mixed-methods research design involving both conceptual and technical components, including a fully functional WEA testbed in which to deploy the improvements, and combining both qualitative and quantitative approaches. Our work yielded significant results that impacted emergency warnings policy in a national context.

10 DISCUSSION AND FUTURE WORK

In 2016, the Federal Communications Commission (FCC) proposed several improvements to the WEA service as part of its Notice for Proposed Rulemaking [13], which featured multiple threads of our research. The accompanying press release [12] distilled six recommendations out of FCC's proposed strategy. These recommendations overlap with four of our recommendations regarding an increase in the maximum length of WEA messages, support for inclusion of links in WEA alerts, need to deliver the alerts to more granular geographic areas, and provisions to make it easier for authorities to test WEA services, train personnel, and raise awareness. A subsequent National Academics of Sciences (NAS) report [25] further discussed and cited our preliminary findings as a basis for defining the future of the WEA service, strengthening the impact of our research. While our research addressed a number of issues of central importance to WEA effectiveness and adoption and provided evidence-based answers, many open questions remain and further improvements are possible. Some of our findings were sensitive to the specific schemes and algorithms used in the implementation of the features, so we recommend subsequent re-testing. For example, the location prediction algorithm could be made more robust by taking advantage of activity recognition using machine learning techniques. Map visualizations could be extended with additional information such as the recipient's distance from the alert zone, estimated time of entry into the affected area, or the nearest route for evacuation.

Newer smartphones equipped with a multiplicity of sensors have a growing awareness of their users' context. In the second trial, we had the ability to infer and categorize user activity, but did not take advantage of this information in any of the features tested. Activity information could be used to avoid distracting the recipient in dangerous situations, for instance, when one is driving at a high speed on a motorway, or changing the notification mode, for example by switching to text-to-speech. Other contextual and preference information could be explicitly set or learned. Integration with social networking apps, call histories, and/or chat histories is possible, subject to privacy concerns.

The WEA service represents a unique type of communication channel. The content is carefully controlled, the traffic is labeled, a broadcast mechanism is available, and phones are designed to process the content specially. But even with these positive attributes, the service has diverged from other communication channels on smartphones as these phones have evolved. A next-generation WEA service might seek to:

10.0.1 Cast WEA as an App. On-phone WEA handling software could be in the form of a smartphone *app* that can be securely updated and evolved without upgrading the phone itself. This would afford continual improvements as user expectations, network capabilities, and current alerting research evolve. Legacy, built-in WEA could be maintained in parallel for some years.

10.0.2 Extend communications for Resilience and Local Access. As part of a survivable communications strategy [19], wireless alerting *even when the recipient's carrier network is down* could be facilitated by giving smartphones the ability to recognize (digitally signed) WEA messages from local AOs over WiFi, Bluetooth, FM radio (RDS) [3] or other wireless bearers.

10.0.3 Enrich Alert Creation. Enable AOs to author rich (HTML-based) content, augment CAP to carry it, and expand WEA to WECAP [18]. Rendering should use the browser mechanisms built into phones. Take advantage of LTE broadcast and remove the short-text limit. Digitally sign all alerts.

10.0.4 Close the Alerting Loop. Today, the WEA channel is open-loop in the sense that alerts go out and alert originators only see the results in terms of the collective actions of the served population. Examine the possibility of closing the loop: provide for recipient responses in future mobile alerting applications (*e.g.,* a button saying "this alert was not relevant for me") to enable deeper studies of alert targeting. The responses need not come back during the emergency but can be trickled back over days following an alert so as to not create inappropriate network load. Our testbed and experimental framework (WEA+) demonstrated the feasibility and usefulness of such a back channel.

As proposed, our WEA+ extensions can be encoded densely to fit within the bounds of the existing WEA format, possibly with pagination–using multiple, chained alerts if a given message will not fit in a single one. We studied the question of how our format might result in additional network load based on re-coding a corpus of existing messages [20]. Considering current practice as represented by a corpus of approximately 10,000 previously-issued National Weather Service alerts, and using our proposed encoding [18], we observed that only 5% of today's alerts would require more than two such alerts. Furthermore, no alert in the corpus would require more than five paginated transmissions. In the legacy 2G/GSM networks that formed the basis for WEA, even this would not be a substantial additional load. In today's LTE networks increasingly dominated by smartphones with offline or built-in maps, the concern further diminishes. As such, we do not believe the network load under the coding and compression schemes proposed in [18] and [20] would be a problem or serve as an impediment to adoption, contrary to what many alert originators cautioned with respect to various WEA enhancements. Quite the opposite, we believe that modern networks have capacities that far exceed those assumed when WEA was designed and, as such, we strongly recommend re-consideration of the balance between the network traffic used for emergency alerting and the public good of conveying richer, clearer, and more targeted information.

Furthermore, services can be downgraded to a lowest common denominator selectively to match network and recipient device limitations to respect the fact that, while smartphones and faster networks are becoming pervasive, older networks and so-called feature phones will persist for some time. It is both possible and practical for cellular networks to send both traditional WEA alerts and our proposed WEA+ alerts. Feature phones and non-upgraded smartphones can display traditional WEA and ignore enhanced alerts. Smartphones with WEA+ capabilities can ignore WEA and decode an alert using the advanced features. More broadly, we advocate the use of all possible means to get alerts to citizens, including piggybacking alerts on top of FM radio broadcasts, as others [3] have advocated. It is also conceivable to use smartphones' WiFi and Bluetooth radios as alternate means for receiving alerts–possibly over local networks in the even that cellular networks are impaired.

ACKNOWLEDGMENTS

This research was sponsored by the Department of Homeland Security, Science and Technology Directorate, First Responders Group (Award HSHQDC-14-C-B0016), to advance the WEA Research, Development, Test and Evaluation Program. We thank our sponsor for their support and input. The work presented in this paper and other related publications represent the position of the authors and not necessarily that of DHS. We thank Dr. Céecile Péraire, Dr. Mirjana Spasojevic, and Rachel Hinman, who contributed to the AORS portion of this work. Besides the report authors, the WEA+ team at Carnegie Mellon University included at various points in time graduate students Maxim Kovalev, Abhinav Jauhri, Joel Krebs, Behrouz Rabiee, Harsh Alkutkar, Harsh Vinayak, Pallavi Mathane, and Ervin Teng, undergraduate intern Michael Wong, and former SEI technical staff member Joseph Elm. We would also like to acknowledge the involvement of volunteers who participated in the trials and interviews, without whom this research would not have been possible.

REFERENCES

[1] 2010. Common Alerting Protocol. http://docs.oasis-open.org/emergency/cap/v1.2/CAP-v1.2-os.html. (2010). Accessed: 2016-03-29.

[2] 2010. Integrated Public Alert and Warning System (IPAWS). *Federal Emergency Management Agency* (2010). http://www.fema.gov/emergency/ipaws

[3] 2013. FEMA and National Public Radio Work Together to Increase Emergency Alert Preparedness for People Who Are Deaf or Hard-of-Hearing. *Federal Emergency Management Agency* (2013). http://www.fema.gov/news-release/2013/10/22/fema-and-national-public-radio-work-together-increase-emergency-alert

[4] 2013. *Wireless Emergency Alerts System Enhancement Recommendations.* Technical Report. U.S. Department of Homeland Security, Science and Technology Directorate. http://www.firstresponder.gov/TechnologyDocuments/Wireless%20Emergency%20Alerts%20System%20Enhancements.pdf

[5] 2014. *Study of Integration Considerations for Wireless Emergency Alerts.* Special Report CMU/SEI-2013-SR-016. Software Engineering Institute. http://repository.cmu.edu/cgi/viewcontent.cgi?article=1781&context=sei

[6] 3GPP. TS 44.012, 2001. Short Message Service Cell (SMSCB) Support on the Mobile Radio Interface. (3GPP. TS 44.012, 2001).

[7] E. Agresti. 1996. *Introduction to Categorical Data Analysis.* John Wiley and Sons.

[8] V.R. Basili, H.D. Rombach, and G. Caldiera. 1994. Encyclopedia of Software Engineering. In *The Goal Question Metric Approach.* Wiley.

[9] V. Braun and V. Clarke. 2006. Using Thematic Analysis in Psychology. *Qualitative Research in Psychology* 3, 2 (2006), 77–101. https://doi.org/10.1191/1478088706qp063oa

[10] CAT. [n. d.]. Coding-Analysis Toolkit. ([n. d.]). http://cat.ucsur.pitt.edu/

[11] J. Cohen. 1992. A power primer. *Psychological Bulletin* 112, 1 (1992), 155. https://doi.org/10.1037/0033-2909.112.1.155 Publisher: American Psychological Association.

[12] Federal Communications Commission. September, 2016. FCC Strengthens Wireless Emergency Alerts as a Public Safety Tool—PS Docket No. 15-91. (September, 2016). https://apps.fcc.gov/edocs_public/attachmatch/DOC-341504A1.pdf

[13] Federal Communications Commission. September, 2016. Report and Order and Further Notice of Proposed Rulemaking—Wireless Emergency Alerts—Amendments to Part 11 of the Commission's Rules Regarding the Emergency Alert System. (September, 2016). https://apps.fcc.gov/edocs_public/attachmatch/FCC-16-127A1.pdf

[14] National Research Council. 2011. *Public Response to Alerts and Warnings on Mobile Devices: Summary of a Workshop on Current Knowledge and Research Gaps.* The National Academies Press, Washington, DC. https://doi.org/10.17226/13076

[15] National Research Council. 2013. *Geotargeted Alerts and Warnings: Report of a Workshop on Current Knowledge and Research Gaps.* The National Academies Press, Washington, DC. https://doi.org/10.17226/18414

[16] H. Erdogmus, M. Griss, B. Iannucci, S. Kumar, J.D. Falcão, A. Jauhri, and M. Kovalev. 2015. *Opportunities, Options, and Enhancements for the Wireless Emergency Alerting Service.* Technical Report CMU-SV-15-001. Carnegie Mellon University.

[17] B. Hamilton, B. Liu, D. Mileti, J. Sutton, S. Madden, and M. Wood. 2014. *Comprehensive Testing of Imminent Threat Public Messages for Mobile Devices.* Technical Report. National Consortium for the Study of Terrorism and Responses to Terrorism (START), College Park, MD. http://www.start.umd.edu/research-projects/comprehensive-testing-imminent-threat-public-messages-mobile-devices-mdp

Prepared for the Department of Homeland Security Science and Technology Directorate, Support to the Homeland Security Enterprise and First Responders Group.

[18] B. Iannucci, J. D. Falc ao, S. Kumar, H. Erdogmus, and M. Griss. 2016. From Alerting to Awareness. In *2016 IEEE Symposium on Technologies for Homeland Security (HST)*. 1–6. https://doi.org/10.1109/THS.2016.7801354

[19] B. Iannucci, J. Cali, R. Caney, and S. Kennedy. 2013. A Survivable Social Network. In *2013 IEEE International Conference on Technologies for Homeland Security (HST)*. 252–258. https://doi.org/10.1109/THS.2013.6699009

[20] A. Jauhri, M. Griss, and H. Erdogmus. 2015. Small Polygon Compression for Integer Coordinates. In *Third Conference on Weather Warnings and Communication*. Raleigh, North Carolina.

[21] A. Jauhri, M. Griss, and H. Erdogmus. 2016. Small Polygon Compression. In *2016 Data Compression Conference (DCC)*. 606–606. https://doi.org/10.1109/DCC.2016.31

[22] S. Kumar, H. Erdogmus, J. Diogo FalcÃčo, M. Griss, and B. Iannucci. 2016. Location-aware wireless emergency alerts. In *2016 IEEE Symposium on Technologies for Homeland Security (HST)*. 1–6. https://doi.org/10.1109/THS.2016.7568902

[23] J.D. McGregor, J.P. Elm, E.T. Stark, J. Lavan, R. Creel, C. Alberts, C. Woody, R. Ellison, and T.M. Keim. 2014. *Best Practices in Wireless Emergency Alerts*. Special Report CMU/SEI-2013-SR-015. Software Engineering Institute. http://repository.cmu.edu/cgi/viewcontent.cgi?article=1782&context=sei

[24] Danielle E. Nagele and Joseph E. Trainor. 2012. Geographic Specificity, Tornadoes, and Protective Action. *Weather, Climate, and Society* 4, 2 (2012), 145–155. https://doi.org/10.1175/WCAS-D-11-00047.1

[25] Engineering National Academies of Sciences and Medicine. 2017. *Emergency Alert and Warning Systems: Current Knowledge and Future Research Directions*. The National Academies Press, Washington, DC. https://doi.org/10.17226/24935

[26] OASIS. 2010. *Common Alerting Protocol, v. 1. 2.* OASIS Standard. OASIS Emergency Management TC. http://docs.oasis-open.org/emergency/cap/v1.2/CAP-v1.2-os.pdf Technical Committee: OASIS Emergency Management TC.

[27] Leysia Palen and Kenneth M. Anderson. 2016. Crisis informatics—New data for extraordinary times. *Science* 353, 6296 (2016), 224–225. https://doi.org/10.1126/science.aag2579

[28] Leysia Palen and Sophia B. Liu. 2007. Citizen Communications in Crisis: Anticipating a Future of ICT-supported Public Participation. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '07)*. ACM, New York, NY, USA, 727–736. https://doi.org/10.1145/1240624.1240736

[29] A.F. Siegel. 1994. Practical Business Statistics. (1994).

[30] Kate Starbird and Jeannie Stamberger. 2010. Tweak the tweet: Leveraging microblogging proliferation with a prescriptive syntax to support citizen reporting. In *Proceedings of the 7th International ISCRAM Conference*.

[31] R.W. II Stoddard, J.P. Elm, S. McCurley, T. Sheard, and T.M. Keim. 2013. *Wireless Emergency Alerts: Trust Model Technical Report*. Special Report CMU/SEI-2013-SR-021. Software Engineering Institute. http://repository.cmu.edu/cgi/viewcontent.cgi?article=1783&context=sei

[32] A. Strauss and J.M. Corbin. 1998. *Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory*. SAGE Publications. https://books.google.ca/books?id=wTwYUnHYsmMC

[33] Sarah Vieweg, Amanda L. Hughes, Kate Starbird, and Leysia Palen. 2010. Microblogging During Two Natural Hazards Events: What Twitter May Contribute to Situational Awareness. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '10)*. ACM, New York, NY, USA, 1079–1088. https://doi.org/10.1145/1753326.1753486

[34] C. Woody and R. Ellison. 2013. *Maximizing Trust in the Wireless Emergency Alerts*. Special Report CMU/SEI-2013-SR-027. Software Engineering Institute. https://resources.sei.cmu.edu/asset_files/SpecialReport/2014_003_001_84093.pdf

Received November 2017; revised April 2018; accepted April 2018