DiReCT: Disaster Response Coordination with Trusted Volunteers

Mohammad Jahanian*, Toru Hasegawa†, Yoshinobu Kawabe‡, Yuki Koizumi†, Amr Magdy*,
Masakatsu Nishigaki§, Tetsushi Ohki§ and K. K. Ramakrishnan*
*University of California, Riverside †Graduate School of Information Science and Technology, Osaka University
‡Aichi Institute of Technology §Shizuoka University

***Abstract*—In disaster management, timely delivery of the right information to the right people can improve outcomes and save lives. With fewer first responders available, volunteers play an increasing role in disaster assistance. A communication framework that brings victims, volunteers, and first responders together for timely disaster assistance can be incredibly helpful. Recent observations, across multiple incidents, have shown that people extensively use social media for communicating in disaster situations. This paper proposes DiReCT, a framework for coordination of disaster response leveraging volunteers in a trustworthy fashion. DiReCT integrates social media with Information-Centric dissemination to help in disaster-related communications. To enable efficient dissemination, we exploit a dynamic namespace that captures the complex incident-management's organizational relationships among participants' roles, in addition to time and location. The DiReCT architecture facilitates both content retrieval and recipient-based publish/subscribe.**

Users, especially regular citizens and volunteers, may not have access to the full incident-management namespace. We propose a social media engine to intelligently map social media posts to the right name(s), using Natural Language Processing and Machine Learning, to steer social media posts towards appropriate first responder(s) that may be most appropriate in helping with the issue posted. Managing volunteers' trust is crucial to ensure only credible posts get to the right people. We use a reputation-based trust model, that is used to guide the authorization, management and evaluation of trustworthiness of content and users. Preliminary evaluation results show the proposed social media and trust management-based solution can be effective.

I. Introduction

Communication saves lives in disaster situations. Effective communication among citizens in need of help, first responders and others that are able to help during and in the aftermath of a disaster can affect outcomes dramatically. Timely and relevant information delivered to the right person(s) helps manage disasters better. Security and trust in this context are just as important, as these can be situations where people are vulnerable, which adversaries and malicious actors may seek to exploit. Additionally, the integration of social media in daily lives has also dramatically changed how victims, volunteers, and first responders exchange information and seek or provide help during and after a disaster. Informal ad-hoc volunteer groups are increasingly becoming an integral part of rescue efforts, with extensive use of social media for grassroots organization of effective rescue missions. We seek to develop a communication framework in disaster situations that brings multiple parties together, first responders from multiple organizations and volunteers, enabling critical information to be shared among relevant groups. As part of that, we seek to use multiple modes of communication, including the use of social networks for hyper-local assistance, in a safe and trustworthy manner. This paper proposes DiReCT, a framework to coordinate disaster response using trusted volunteers, social media, and crowd-sourced verification. Through this coordination, DiReCT enables a bridging between social media platforms (which victims interact with, to report events), and publish/subscribe frameworks [1] (to deliver reports to incident commanders, first responders and volunteers).

Naming is fundamental to networking [2]. Information-Centric Networks (ICNs) use names as the location-independent identity at the network layer (e.g., NDN [2], MobilityFirst [3]). Each application domain may also have its own namespace and hierarchy. This namespace needs to be carefully managed in a disaster scenario, so that only authorized entities may create and update the namespace that is used by all participants (victims, volunteers and first responders).

A key aspect of our work is in the integration of social media into managing the disaster response. People (especially common citizens) use the normal forms of social media communications (Twitter, Facebook, etc.) to send and get information during disaster situations, since that is what they are familiar with. This was observed across different parts of the world, e.g., in California Wildfires in 2018 [4] and more recently, Hurricane Dorian in 2019 [5]. Furthermore, people are likely to have little knowledge or understanding of the notions of a structured namespace to determine where to publish information or ask for information (e.g., generate an Interest in NDN). It would be ideal to allow the use of these social media platforms in the manner people are used to, with free-form text, possibly enhanced with pull-down menus to determine key meta-information to associate with the message. This then would require mapping individual messages to the namespace for publishing, expressing interest or creating subscriptions. The approach we explore is to process the natural language in the social media posts (Tweets, Facebook posts, etc.) and map it to appropriate names in the namespace. This allows messages to be delivered to the correct entity (such as first responder or incident commander) based on the derived names. Additional handling of the message may be based on the decisions by those individuals (e.g., handling false positives).

A number of security and trust issues arise when we accept volunteers as an integral part of disaster management. An

978-1-7281-4920-2/19/$31.00 ©2019 IEEE

even greater concern is the use of social media to disseminate information, especially given the recent concerns related to the spread of information that is not true (aka, fake news [6]) through social media. The need to allow only trustworthy entities to publish to a name and to subscribe and receive valuable information (e.g., prevent malicious actors from receiving information destined to first responders) becomes important.

How do we develop trust in this environment? While trust usually appears to be a higher layer issue, it has critical ties to how information is allowed to be published into a namespace and is propagated in the network. We do not have a long-term reputation framework to assist us. Instead, we have to establish the veracity of a social media post in a short time. A voting authority solicits verification of social media posts by crowdsourcing citizens to verify and vote on the social media posts. Based on the crowd-sourced votes, the voting authority deems some posts as "true" and others as "fake". While fake posts get discarded, the voting authority publishes the true posts to their mapped names in the namespace. It may also send them to the incident commander first before publication, depending on the nature of the posts' content. The publications are then fed to the pub/sub-based dissemination framework, "pushing" them towards the right recipients (i.e., subscribers) [1]. For the vote crowdsourcing component, we suggest the use of a short-term reputation system which uses coins to reward good behavior and reduces the coins disproportionately fast for bad behavior on the part of the verifiers. Coins are removed from a verifier for a false/erroneous vote/verification. The coin balance of a trusted verifier will be higher and this can be used to guide their selection as a trusted verifier.

It is important to integrate volunteers into the disaster response. We use first responders to participate in establishing the trustworthiness of volunteers. Specifically, we envisage the use of biometrics of the volunteers to be provided and verified by first responders as a deterrence against malicious behavior, under an assumption that citizens who submit their biometrics would never behave maliciously [7]. First responders are responsible for managing volunteers, and a volunteer submits his/her biometric information, such as fingerprints, to one of the nearby first responders and get his/her key-pair, which is directly derived from the biometric information [8].

We evaluate the applicability of our use of social media information using Tweets collected during (and just after) two recent wildfires in the state of California in the United States. DiReCT uses Natural Language Processing (NLP) to map the tweets to the right first responder(s) automatically. We demonstrate that it does so accurately up to 96% of the time. Even for the ones that are inaccurate (4%), they can be recovered by the help of the first responder, manually forwarding it to the correct recipient. For trust management, our simulation results show that our short-term reputation-based trust framework with the use of coin incentives ensures the majority of messages are evaluated correctly (in terms of "true" or "fake"), even when a majority of verifying volunteers vote dishonestly.

The contributions of this paper are: 1) a system that integrates critical components and actors in disaster scenarios, i.e., first responders, volunteers and social media, in a name-based, trustworthy information dissemination model, 2) a trust management model for volunteers based on short-term reputation and incentives, to prevent misinformation from spreading in the network, 3) a social media engine that intelligently and automatically maps free-form social media posts to the right names for publication in a pub/sub framework, and 4) demonstration of the effectiveness of our trust management and social media engine through measurements from our evaluations.

II. Related Work

There have been many studies characterizing and designing network and communication frameworks for disaster management (surveyed in [9], [10]). Disasters can have major network-related impacts such as infrastructure damage and excessive congestion [10]. Different communication and network technologies have been proposed for disaster management, including Cellular, P2P and Satellite [9]. In addition to technological challenges, there are known social and organizational challenges when designing an effective and efficient network architecture for disasters [9]: a common language between organizations and citizens, as well as a structured, non-ad-hoc organization of disaster response is needed and desired [9].

In the past decade, the use of name-based and information-centric capabilities in disasters have gained interest as they offer important benefits, such as location-independence, content-oriented security, and content caching, enabling inherent support for timely delivery of content and mobility [11]. To enhance efficiency, Publish/Subscribe (pub/sub) methods in ICN have been proposed to support efficient, timely dissemination of information in a point-to-multipoint manner [12], [1], [13]. CNS [1] enables recipient-based pub/sub for disaster scenarios. However, there are additional challenges that need to be addressed to use name-based communications. There is a need for users to know the precise names to be used while publishing a critical disaster-related information, which they may not know. Also, security and trust of participants (especially with volunteers) needs to be managed.

Security and trust are important. Named Data Networking supports content-oriented security and name-based trust schema [14]. Building trust is key to ensure credible information dissemination and communication. Work in [15] describes a number of attacks and their corresponding countermeasures related to untrustworthy access and propagation of fake information. Work in [16] proposes trustworthy crowdsourcing systems to prevent malicious behaviors in such systems.

Social media has been increasingly used for information dissemination in incident response, which can be very beneficial, especially when traditional means of communications, e.g., 911, are down or overloaded [17], [4], [5]. Social media server-based extensions and plugins have been developed for help during disasters, providing users with useful information, e.g., updates, warnings, offers, maps, etc [17]. Our work integrates social media with name-based communication and content dissemination, to intelligently guide social media posts to the right recipients, as opposed to the currently unstructured ways for dissemination, e.g., in form of retweets.

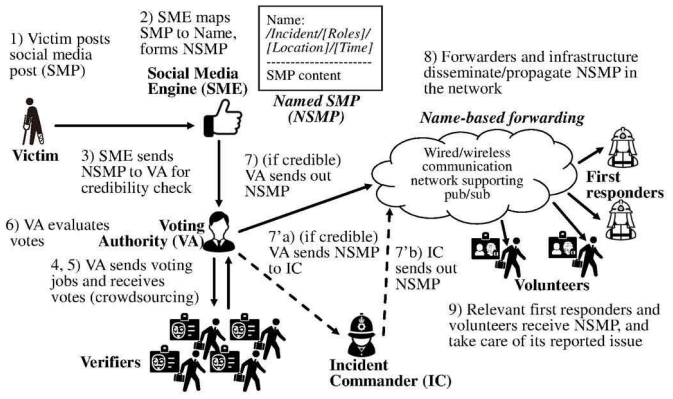


Fig. 1. System model

III. System Model

The system we consider is one where people involved in a disaster (first responders, victims and volunteers willing to help) can communicate easily using a name-based communication framework. Each participant may have an associated *role*, and the namespace captures the relationships among these roles. We expect that the *incident commander* creates, updates and manages the namespace, which may follow a structure that is similar to the specifications in the National Incident Management System (NIMS) [18], reflecting the organizational roles of participants. Volunteers, first responders and incident commanders publish into the namespace, with messages that are relevant to the disaster and intended to reach the relevant people. They also subscribe to names or name prefixes based on their role and the information they expect to receive. Information can be accessed by the name, using an Interest/Data framework as in NDN [2] as well as using a publish/subscribe framework such as the recipient-based pub/sub in [1]. An important consideration is that the namespace is likely to dynamically change as new incidents [1] occur or when new roles are added or assigned to specific first responders and volunteers. Generally, exchanged messages refer to *events*. An event is defined as a 3-tuple of $\langle$*Type*, *Time*, *Location*$\rangle$; an example for an event is $\langle$*Fire*, *11am April 1*, *Yamadaoka Suita-city*$\rangle$. The event type is a task or issue associated with a role.

The overall procedure is shown in Fig. 1. A *victim* (or any *data initiator*) posts a disaster-related message (report, update, etc.) on social media, e.g., tweeting, in free-form text. A *social media engine* (SME) collects, analyzes and maps *social media posts* (SMPs) to a name, of form "/Incident/[Role]/[Location]/[Time]" according to the incident namespace (described in §IV-A), using NLP/ML procedures (described in §IV-B).

After the name-mapping procedure, and for the purpose of fact-checking social media posts, the SME sends the "named SMP" (NSMP) to the *verification service* which is a crowdsourcing service involving a set of volunteers that vote on the veracity of posts, referred to as *verifiers*, who help in establishing the trustworthiness (credibility) of the NSMP. The *voting authority* is a trusted third party which manages the verification service. The voting authority eventually sends the verified NSMP to the incident commander or to those first responders that are appropriate (with the choice depending on the real-time-criticality of the matter) who are part of the namespace. The *trust management* evaluates the trustworthiness of verifiers in performing the verification task, which helps with ensuring the spread of only credible information in the network. A *first responder* receives a command from the incident commander as well as a verified NSMP from the voting authority.

In addition, we also consider *volunteers* who seek to actively participate in the disaster response. A volunteer registers herself/himself with a volunteer service by providing credentials, including potentially biometric information, under an assumption that volunteers who provide their biometrics would never behave maliciously. Since these volunteers are trusted, they are allowed to publish messages and subscribe to names in the incident namespace.

Our design assumes the initial interaction between users and social media servers (e.g., Twitter) over the Internet. However, the communications in the pub/sub framework, e.g., between first responders and volunteers, can be a combination of infrastructure-based (e.g., through fixed routers) or infrastructure-less (e.g., device-to-device, through data mules and/or drones) communication links. The use of information-centric dissemination allows for information delivery over such diverse links [11], especially to deal with situations such as mobility and network link failures. We assume that the incident commander and first responders are honest, while volunteers may not all be honest. First responders and volunteers are assumed to have identifiers and credentials. The incident commander and first responders have certificates issued by an incident authority, that is a trusted third party and works as the root certification authority. Volunteers are assumed to not have certificates issued by the trusted third party and their credentials are self-certified. Moreover, volunteers may have multiple pairs of identifiers and credentials.

IV. Architectural Components

To facilitate the desired bridging between social media platforms and publish/subscribe frameworks, DiReCT uses a namespace that unifies and organizes the information exchanges, a social media engine that maps free-form social media posts to the right names through NLP/ML techniques, and a verification service that assesses the credibility of social media posts through crowdsourcing verifiers and reputation systems. This section explains these key components.

A. Naming Schema

Naming is a key component of DiReCT, as it unifies the interactions between all different actors (civilians, first responders, etc.) and guides the subscription and publication paths. DiReCT's namespace follows an NDN-style, hierarchical structure [2] (it can be extended to a graph-based namespace as well, as proposed in [13]). The namespace represents entities related to and critical in incident management, and captures complex relations among them. An example namespace is (partially) depicted in Fig. 2. The namespace in DiReCT has the structure of "/Incident/[Role]/[Location]

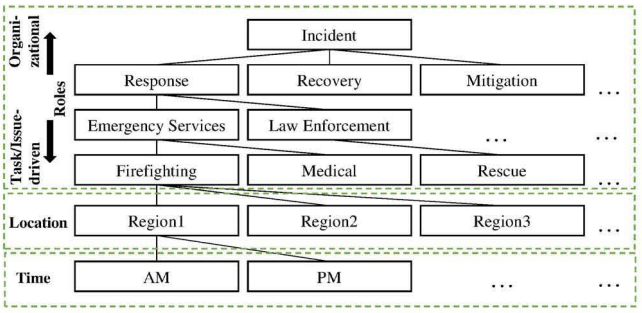


Fig. 2. Incident management namespace

/[Time]", where elements within brackets "[...]" can be any number of name components. Each dimension may correspond to some aspect of an incident and can contain any number of name levels in the namespace hierarchy. For example, the 'Role' dimension can consist of organizational roles (e.g., NIMS [18]) towards the root, and task- or issue-driven (i.e., incident-specific) roles towards the leaves of the tree, to cover the critical components for managing the incident. This design is suitable to model the *what*, *where*, and *when* aspects of content, which are critical aspects of incident information.

The namespace follows a recipient hierarchy for dissemination, as proposed in CNS [1] for a recipient-based pub/sub. The paths followed for publication/subscription for recipient-based pub/sub is the reverse of topic-based pub/sub [12]. In DiReCT, subscribing to a prefix, implicitly means also subscribing to all ancestors of that prefix. For example, a subscriber of "/Incident/Response/EmergencyServices/Firefighting", would receive all publications corresponding to that name, as well as publications to names above it, i.e., "/Incident/Response/EmergencyServices", etc. Conversely, a content published to a prefix, implicitly means it will also be disseminated towards all subscribers of the descendants of the prefix. For example, a content published to "/Incident/Response/EmergencyServices/Firefighting/Region1" will also be received by subscribers of "/Incident/Response/EmergencyServices/Firefighting/Region1/AM", etc.

First responders and volunteers subscribe to prefixes; e.g., a fireman dispatched to fight a fire in Region1 during AM hours, subscribes to "/Incident /Response/EmergencyServices/Firefighting/Region1/AM". The use of recipient-based pub/sub is very beneficial in our architecture since it allows the most relevant first responders, at the finest granularity possible, to receive incident-related publications, without causing distraction and information overload for 'non-relevant' first responders that may be busy dealing with other tasks. Also, leveraging the hierarchical structure of the namespace, and allowing subscription/publication to any desired granularity (and not necessarily each individual below/above a prefix) greatly decreases the number of pub/sub messages and state that has to be maintained in the network, as shown in [12], [1].

An incident commander manages, creates and updates the namespace. At the beginning of an incident, the initial namespace is derived from an a priori template. This template is pre-defined and follows the incident command chains designed specifically for that particular type of incident, e.g. [18]. A benefit of having this template is that it is also accompanied by a trained data set to enable supervised classification performed by the Social Media Engine (SME) during the incident, as we explain in §IV-B. As the incident progresses and new issues and tasks arise, either from monitoring the incident or suggestions/offers from volunteers, the incident commander dynamically modifies the (more fine-grained parts of the) namespace and notifies the nodes that have it or use it (such as rendezvous points [13]) for synchronization. For example, upon receiving an aid offer from a volunteer team suggesting that they want to help with firefighting in Region4 during PM hours, the incident commander adds the sub-tree "Region4 $\rightarrow$ PM" as a child of "Firefighting" in the namespace. For encryption and authentication, we envisage the use of Attribute-Based Encryption (ABE) [19] with rules and attributes following our namespace. However, we are unable to focus on those issues due to limited space.

B. Social Media Engine

DiReCT's recipient-based pub/sub allows social media posts (SMP), e.g., Tweets, to be sent as publications and disseminated in the network. However, the correct delivery of each publication depends on the name it has been published to. A civilian (e.g., a victim reporting an emergency and seeking help) may not have knowledge of or access to the namespace to pick the right prefix. Having users download the whole namespace (or have it pushed to them proactively) is costly both in terms of network usage and storage on a user device. Users would also have to manually go through a potentially large namespace to pick the right name for their publication which can be time-consuming and error-prone. DiReCT solves these problems by employing a social media engine (SME) that intelligently maps a SMP to the right part(s) of the namespace using natural language processing and machine learning.

Fig. 3 shows the overall architecture of SME: The incoming SMP, possibly including latitude/longitude, and timestamp, in addition to the text, goes through a sequence of stages to get mapped to a (set of) name(s) of the namespace structure shown in Fig. 2. This pipeline processes SMPs in an online fashion, i.e., as SMPs arrive and are captured from each social media platform. Using trained models for text (data from previous and/or similar incidents), the classification procedure maps the textual part of the SMP to the right roles, depending on what tasks and/or issues the SMP is referring to. The classes (labels) of this classification are the leaves of the 'Roles' part of the namespace only; i.e., tasks and issues only. Using maps and other geo-related databases, the localization step maps lat/long (and possibly location names mentioned within the text) to the right location names under the previously derived role names. Finally, the temporalizing step maps the timestamp in the SMP to the right time-related name, i.e., which time interval this particular timestamp belongs to. Having formed a complete name, the SME appends the name to the SMP and sends it forward, to the voting authority to be further processed and disseminated (explained in §IV-C).

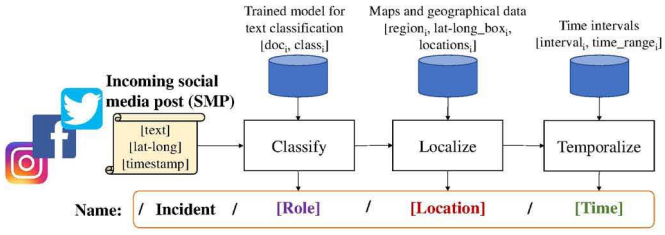


Fig. 3. Social media engine

It is worth noting that SME does not determine the importance of an SMP. It merely takes care of the mapping to a name for directing the delivery. Despite this mapping, there is chance for inaccuracy, i.e., an SMP is mapped to the wrong name, and thus delivered to the wrong first responder(s). If that happens, e.g., if a medical doctor receives a report regarding an urgent need for firefighting, he/she can either: 1) re-publish the SMP to the network picking the right names (as he/she has access to the namespace as a first responder); or 2) send it as a unicast message to his/her incident commander, who then can forward it as appropriate. This step recovers DiReCT from inaccurate deliveries. However, using a good classification/learning approach will greatly reduce these inaccuracies, so that only a very small percentage require subsequent correct forwarding.

C. Verification Service

1) Overview and Requirements: We develop a verification service involving a set of voluntary verifiers in a crowdsourced manner as a means of ensuring that we only bring credible information into the disaster response activity since disaster-related SMPs may contain a lot of fake and useless information [20]. When the voting authority, who is responsible for managing the verification service, receives a SMP from the SME, it assigns a verification task for an event reported in the SMP to some of the verifiers. The verifiers check and vote on whether the SMP is credible or not. If the majority of the verifiers judges the SMP to be credible, the SMP is sent to either first responders or the incident commander.

Important requirements of the verification service is twofold: One is *to mitigate malicious behavior of verifiers*, as they are potentially dishonest. Dishonest verifiers, for instance, may intend to bring fake information to the disaster response activity by voting maliciously. That is, they may vote "credible" for a fake SMP and "fake" for a credible one. The other is *to recruit as many verifiers as possible*, because the number of disaster-related SMPs is huge.

To meet the first requirement, it is useful to use a trust management system, which is a mechanism to identify "trustworthy" people, for mitigating effects of malicious votes [16]. An important and unique challenge here is that the trustworthiness of verifiers must be built in a short-term process although human trustworthiness is typically built in a long-term process. It would be helpful if the trustworthiness of verifiers is built before a disaster, but clearly this is not feasible. To address this issue, we adopt a reputation system where the trustworthiness of verifiers is represented as their reputation based on their past voting results. Each verifier has his/her trust value, which is associated with his/her ID. If the verification service detects that a verifier votes maliciously, the degree of the trustworthiness of the verifier is reduced. In contrast, the degree of the trustworthiness of legitimate verifiers get increased. The degree of the trustworthiness reflects the past voting activities of each verifier, thereby avoiding the effect of malicious verifiers without tracking each vote activity.

The second requirement comes from the fact that the number of disaster-related SMPs is huge but only a small portion of the SMPs contains credible information. Gupta and Kumaraguru [20] analyzed disaster-related tweets regarding several high impact events of 2011 and they have revealed that only 17% of the total tweets about the events contain credible situational-aware information. Furthermore, we collected disaster-related tweets with the Twitter search API during the 2018 Japan floods [21] and observed that the maximum number of disaster-related tweets per day is 789,696 (July 7th, 2018). To extract credible information from a large amount of SMPs, we need help from a high number of verifiers. To meet this requirement, we adopt an easy-to-use registration mechanism, where each verifier uses a self-certified public key as her/his ID. IDs are used to identify verifiers in the trust management mechanism. In addition, the corresponding secret key is used to sign her/his voting results.

2) Trust Management System: The trust management system, which is inspired by crowdsourcing [16], identifies trustworthy people. The voting authority plays the role of a coordinator of both the crowdsourcing and the trust management system. Specifically, the voting authority recruits verifiers and manages the degree of trustworthiness, which is represented through *coins*, to the verifiers according to the quality of their completed voting results. Coins are issued to and removed from verifiers when they do proper and malicious voting, respectively. An increase in coins means that a verifier obtains a high reputation. In this way, a good or bad reputation is built in a short time scale.

When the verification service receives a SMP from the SME, it asks the verifiers near the location claimed by the SMP to vote on whether the SMP is credible or not. These requests are forwarded to verifiers near the location with location-based forwarding [22], [23], which is realized on top of a recipient-based pub/sub communication framework. Verifiers subscribe to names of their locations, which we call *location names*. Please note that the location name corresponds to the location of the 3-tuple for the event, as described in Section IV-A. A location is defined by one of several coordinate systems, like the road numbering system, the x-y coordinate system and a space-filling curve system like Z-ordering [24]. Next, the voting authority publishes the request to the location names. Specifically, the location names specify that verifiers at the locations should participate in the vote. Since a location does not specify a single point, but instead an area like a square of the predefined size, verifiers in that area receive the request.

Verifiers who receive the request check the credibility of the SMP and send their voting result with digital signatures signed with their secret keys to the voting authority. The voting authority first verifies the integrity of the voting results with the attached digital signatures and the corresponding public keys, i.e., the IDs of the verifies. It next judges the credibility of the SMP by computing the sum of the collected votes weighted by

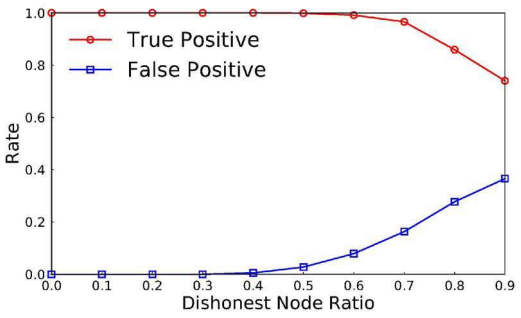


Fig. 4. True positive and false positive rates of votes

the coins of the verifiers. More precisely, the voting result c is computed as $c = \sum_{i \in \mathcal{V}} t_i v_i$, where $\mathcal{V}$, t_i and v_i are the set of verifiers who join this voting, the number of coins of verifier i and the vote of i, respectively. The vote v_i for/against the truthfulness of the SMP equals $1/-1$. The SMP is credible if the voting result c is greater than 0, otherwise it is fake.

Coins are issued to/removed from data initiators and verifiers as the reward/punishment. The number of coins issued to and removed from verifiers are predefined as r. Each verifier i earns or loses the coins $rcv_i / \left(\sum_{j=\mathcal{V}} t_j\right)$.

To compromise voting held in a certain area, a malicious verifier needs to have more than half of the total coins of all the verifiers in the area, because of the nature of majority voting weighted by coins. This makes it difficult for malicious verifiers to tamper with a voting result, thus helping to filtering out fake information. We also expect that the issuance and removal of coins motivate verifiers to work correctly. The design thus gradually excludes malicious verifiers.

V. Preliminary Evaluation

In this section, we provide preliminary results of our evaluation of our trust management and social media engine.

A. Verification Service

We evaluate how the verification service filters fake SMPs via a simulation experiment. We deploy 10,000 verifiers randomly in a square area of 4 km side, assuming that it is a city area heavily damaged by a significant earthquake. Many people stay at their offices or shelters, and hence their initial locations are randomly chosen. They move according to the Lévy walk model [25].

The initial number of coins for verifiers are set to two and the rewards to verifiers r is set to 0.2. Every 10 minutes, a SMP regarding a randomly chosen location is posted, and the credibility of the SMP, credible or fake, is determined randomly with probability of 0.5. The simulation duration is one day, which means we have 144 SMPs totally. Verifiers in a square area of 100 m side centered at the location claimed in the SMP join the voting on the SMP. For each posted message, verifiers who are within 100 square meters of the location vote on the message. Dishonest verifiers vote maliciously with probability of 0.5. We evaluate true positive and false positive rates by changing the ratio of dishonest verifiers to the entire verifiers. The true positive rate is the ratio of the number of voting trials which correctly identifies the credibility of SMPs and the false positive rate is the ratio of the number of voting trials which erroneously decide whether the SMP is credible.

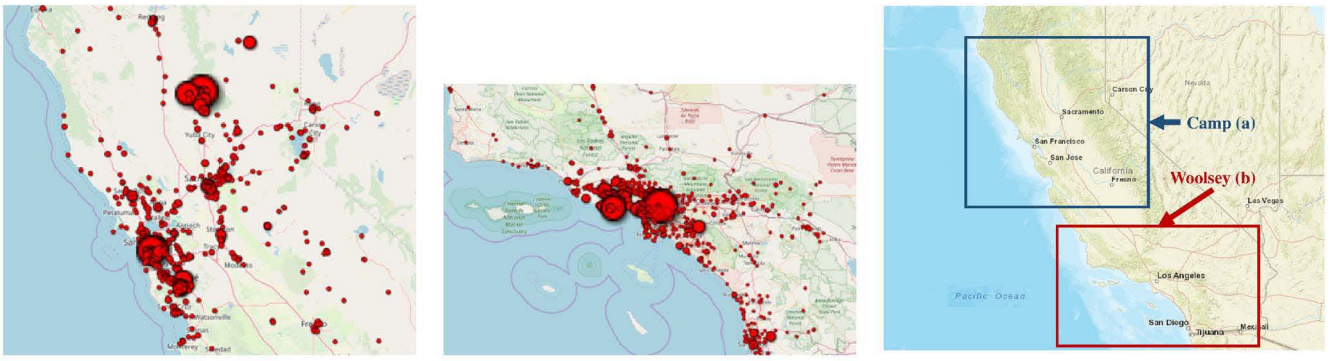
Fig. 4 shows the true positive and false positive rates for the ratio of dishonest verifiers. We have the following observations: First, the trust management system identifies fake SMPs even if the majority of verifiers is malicious. Even when 90 percent of the verifiers are dishonest, the credibility of about 80 percent of SMPs is identified. Coins work as a short-term reputation measure. The main benefit of the trust management system is to reduce the probability that first responders go to the locations claimed by fake SMPs. Second, this success is partly because verifiers are densely distributed in the disaster area. In the simulation, about 19 verifiers join a single voting trial on average. How to cope with the condition where verifiers are sparsely distributed needs to be studied further.

B. Social Media Engine

The SME processes social media posts using NLP & ML methods to map incoming posts to the right names. To evaluate the performance of SME in our context, we use Tweets collected from two disasters in California in 2018, namely the Camp [26] and Woolsey [27] wildfires. Crawling Twitter, we captured many Tweets (with no restrictions on keywords) sent from Nov. 7th to Nov. 26th, 2018 (several million) for each fire into the "Camp" and "Woolsey" tweet *pools* (un-processed collections). The geo-location bounding boxes for the two pools are shown in Fig. 5c. The geo-location restrictions were chosen according to the facts of the two wildfires [26], [27].

The Camp and Woolsey pools have 959,740 and 1,961,131 tweets respectively. We show the spatial and temporal distributions of fire-related tweets, doing a keyword-based text mining for tweets that include combinations of the words "fire" or "Camp/Woolsey", using the Apache Lucene API [28]. The analysis results are shown in Fig. 5 and Fig. 6, indicating strong correlation of tweet patterns with the actual progression of the events during the two wildfires: peaking around Nov. 9-10 and getting contained around Nov. 21-25 [26], [27]. Fig. 5 shows that the density of fire-related tweets is higher at areas most affected by the wildfires, i.e., Paradise, CA (Camp fire), and Thousand Oaks and Malibu, CA (Woolsey fire) [26], [27].

SME's NLP and ML procedures use the NLTK [29] and Scikit-learn [30] toolkits. Our implementation is in Python, and we evaluate the performance using a machine with Ubuntu 14.04.6 LTS using Intel(R) Xeon(R) CPU E5-2650 v4 @ 2.20GHz dual-socket with 14 cores each with hyper-threading enabled with 252GB RAM. To evaluate the performance, we use a subset of our Camp and Woolsey pools; to produce our Camp and Woolsey *data sets*. We identify a total of ~35K tweets across the two pools related to the wildfire incidents. We identified 13 classes (tasks/issues in namespace) and annotated them based on keywords related to each class. The classes picked are based on what we felt were the most important issues and roles during wildfires, in accordance with FEMA reports [31]. Fig. 7 shows these classes and the number of instances for each class in the two data sets, combined. 'Firefighting' has the most tweets associated with



(a) Camp (Northern California) (b) Woolsey (Southern California) (c) Bounding boxes for Camp and Woolsey

Fig. 5. Spatial distribution of fire-related tweets for Camp Fire (a) and Woolsey Fire (b) – (size of circles correlates with the number of tweets)

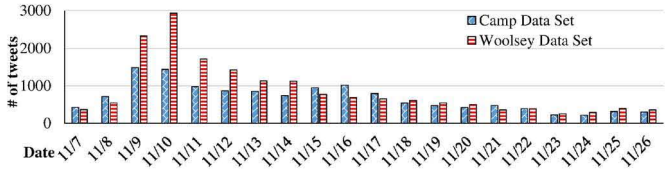


Fig. 6. Temporal distribution of fire-related tweets

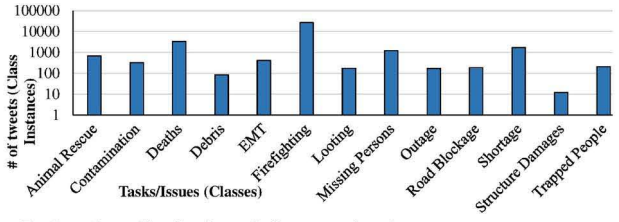
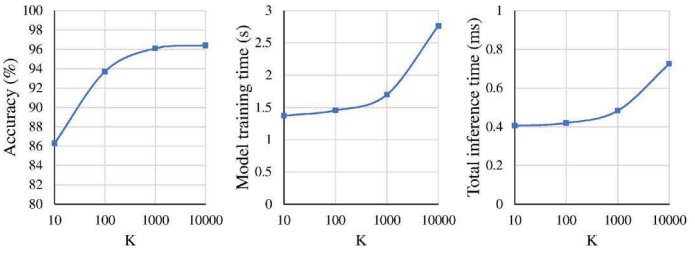


Fig. 7. Per-class distribution of disaster-related tweets



(a) Accuracy (b) Model training (c) Total inference

Fig. 8. Performance of the SME procedures

TABLE I
INFERENCE METRICS FOR K=1000

Metric	Precision	Recall	F1-score
Micro average	0.96	0.96	0.96
Macro average	0.88	0.81	0.84

it (around 75%), which is reasonable since the incident is fire-related. Inference, i.e., assignment of a tweet to a class, is followed by mapping it to a name. Each class corresponds to a name in the namespace; e.g., a tweet classified as a 'Firefighting' instance, would be mapped to "/Incident/Response/EmergencyServices/Firefighting/..." (according to the namespace, e.g., Fig. 2). The time and location of the name can be derived from geo-location and timestamp of the tweets. Then the SMPs are sent to verifiers for accuracy and credibility. Additionally, methods similar to content-based fake news detection [6] can be leveraged for faster, more thorough, and automated fact checking of SMPs by the SME.

Our learning procedure consists of tokenization, filtering out unwanted tokens (stopwords, etc.), stemming, and vectorization. For vectorization, we use tf-idf [30], allowing n-grams of size 1 and 2. For inference, we use Random Forest classification. An important feature of DiReCT is its processing of tweets in an online way; thus, we use training data from a previous and/or similar incident, to enable an accurate and fast classification of new tweets. In our experiment, we use the Woolsey data set (with ∼23K tweets) for training, and the Camp data set (with ∼12K tweets) for testing. Tweets from the Camp data set (i.e., test set) are processed one by one, classified into one of the classes using the trained model based on the other similar data set, namely the Woolsey data set, to be mapped to an appropriate name.

Feature selection is an important step in learning, as a good feature selection prevents overfitting and reduces processing time. We use K-Best feature selection using the chi2 method [30], which intelligently picks the top K most relevant features, and perform training and classification based on them. There are ∼20K features in our classification We use values of 10, 100, 1000, and 10000 for the value of K in the 'K-best' feature selection process and compare the results, as displayed in Fig. 8. The tf-idf vectorization takes 2.68 seconds. Fig. 8a shows the accuracy (calculated according to [30]) for different values of K: it shows that for values of K at or above 100, we reach accuracies of above 93%, which we believe is very good. For K from 100 to 10000, the accuracy does not change much, while the model training time (shown in Fig. 8b, excluding tf-idf vectorization step) keeps increasing for increasing K. Fig. 8c shows the total inference (classification) time for all (12,697) test tweets; per-tweet average inference time would be the values shown divided by 12,697. The results show the increasing growth rate and also reasonably low execution times. These latency values (both training and inference times) are important in online settings for the server-based SME. The small inference times demonstrate that using DiReCT, we can quickly (in the order of microseconds) classify a tweet and map it to the right names (and therefore to the right first responders who

can help), with high accuracy. Generally, the training data can be either from another incident having high similarity with the current incident (as we do here), or new tweets arriving from the current incident. In the former case, training needs to be done only once. In the latter case, which uses a more relevant training set (as every incident may have its unique characteristics), the training needs to be periodically done with the new data added, i.e., re-training. For faster model re-training, using incremental learning methods such as [32] can be leveraged, albeit sacrificing some accuracy.

Taking into account accuracy, training and classification times (Fig. 8), we pick K=1000 as a reasonable value for the feature set. It achieves 96% classification (and thus mapping to the correct name) accuracy which is very good: it means that out of 12,697 tweets, 12,189 of them get to the correct first responders with the publisher. Note that we are assuming that the civilian user does *not* know anything about the namespace, and we are performing the mapping automatically. Only 508 tweets would be inaccurately delivered, which can be appropriately forwarded manually afterwards; this shows the significant benefit of using DiReCT. Table. I shows other metrics of our algorithm which are important as they answer different questions about the inference performance and its practical usefulness in name mapping. These metrics are Precision (e.g., of all messages sent to firefighters, how many were actually about fire?), Recall (e.g., of all the messages actually about fire, how many did we send to firefighters?) and F1-score (weighted average of Precision and Recall), both as micro and macro averages (calculated according to [30]). Macro average metric values are a sum of metrics for all classes, divided by the number of classes. Micro average metrics, on the other hand, take into account the number of per-class instances, thus giving a more fine-grained averaging. As seen in the table, macro average values are less than micro average values. Note that micro average values (all 96%) are a better metric for our data set since our data set is not a balanced one (75% of instances belong to a single class, namely 'firefighting'). These results show the good performance of our learning/inference at the social media engine, indicating the effectiveness of DiReCT in mapping social media posts to the right names leading to the relevant first responders and volunteers.

VI. Conclusion

We proposed DiReCT, a framework to coordinate disaster response with first responders that receive timely relevant information and trusted volunteers. DiReCT bridges civilian-oriented social media platforms with a pub/sub information dissemination architecture for first responders and volunteers. It uses a hierarchical naming schema, NLP/ML-based social media analysis, and crowd-sourced reputation-based verification. Results from our preliminary evaluation show that DiReCT is effective and efficient in providing the correct mapping between free-form text and pub/sub-based names, and ensures that volunteers are trustworthy and information disseminated is first verified.

VII. Acknowledgements

This work was supported by the US National Science Foundation grant CNS-1818971, US Department of Commerce, National Institute of Standards and Technology (award 70NANB17H188), and the National Institute of Information and Communications Technology in Japan (Contract No. 193).

References

[1] J. Chen *et al.*, "CNS: Content-oriented notification service for managing disasters," in *ICN*, 2016.
[2] L. Zhang *et al.*, "Named data networking," *SIGCOMM CCR*, 2014.
[3] D. Raychaudhuri *et al.*, "Mobilityfirst: a robust and trustworthy mobility-centric architecture for the future internet," *SIGMOBILE MCCR*, 2012.
[4] M. Saunders, "Social media: California wildfires force thousands to evacuate," https://tinyurl.com/y2pfolyd, Nov. 2018.
[5] E. Cranley, "'Our government failed us': Bahamians were left to coordinate rescue efforts on social media after Hurricane Dorian," https://tinyurl.com/y3kvvmce, Sep. 2019.
[6] G. B. Guacho *et al.*, "Semi-supervised content-based detection of misinformation via tensor embeddings," in *ASONAM*, 2018.
[7] K. Nandakumar and A. K. Jain, "Biometric authentication: System security and user privacy," *IEEE Computer*, vol. 45, pp. 87–92, Nov. 2012.
[8] K. Takahashi *et al.*, "Signature schemes with a fuzzy private key," *International Journal of Information Security*, vol. 18, no. 5, pp. 581–617, Oct. 2019. [Online]. Available: https://doi.org/10.1007/s10207-019-00428-z
[9] L. Yan, "A survey on communication networks in emergency warning systems," *Sci. Comput*, 2011.
[10] Z. EL Khaled and H. Mcheick, "Case studies of communications systems during harsh environments: A review of approaches, weaknesses, and limitations to improve quality of service," *IJDSN*, 2019.
[11] J. Seedorf *et al.*, "The benefit of information centric networking for enabling communications in disaster scenarios," in *GC Wkshps*, 2015.
[12] J. Chen *et al.*, "COPSS: An efficient content oriented publish/subscribe system," in *ANCS*, 2011.
[13] M. Jahanian *et al.*, "Graph-based namespaces and load sharing for efficient information dissemination in disasters," in *ICNP*, 2019.
[14] Y. Yu *et al.*, "Schematizing trust in named data networking," in *ICN*, 2015.
[15] R. Tourani *et al.*, "Security, privacy, and access control in information-centric networking: A survey," *IEEE Communications Surveys and Tutorials*, 2018.
[16] H. Yu *et al.*, "Challenges and opportunities for trust management in crowdsourcing," in *WI-IAT*, 2012.
[17] M. Jahanian *et al.*, "The evolving nature of disaster management in the internet and social media era," in *LANMAN*, 2018.
[18] FEMA, "National Incident Management System," https://www.fema.gov/national-incident-management-system, 2019.
[19] J. Bethencourt *et al.*, "Ciphertext-policy attribute-based encryption," in *SP*, 2007.
[20] A. Gupta and P. Kumaraguru, "Credibility ranking of tweets during high impact events," in *PSOSM*, 2012.
[21] Japan Meteorological Agency, "The heavy rain event of July 2018," https://tinyurl.com/yygoags6, Jul. 2018.
[22] G. Grassi *et al.*, "Navigo: Interest forwarding by geolocations in vehicular named data networking," in *WoWMoM*, 2015.
[23] Y. Kurihara *et al.*, "Compact data structures for location-based forwarding in NDN networks," in *ICN-SRA*, 2018.
[24] G. M. Morton, "A computer oriented geodetic data base; and a new technique in file sequencing," IBM, Tech. Rep., 1966.
[25] I. Rhee *et al.*, "On the levy-walk nature of human mobility," *TON*, 2011.
[26] Wikipedia, "Camp fire (2018)," https://en.wikipedia.org/wiki/Camp_Fire_(2018).
[27] ——, "Woolsey fire," https://en.wikipedia.org/wiki/Woolsey_Fire.
[28] "Apache Lucene," https://lucene.apache.org/.
[29] S. Bird *et al.*, *Natural Language Processing with Python*, 1st ed. O'Reilly Media, Inc., 2009.
[30] F. Pedregosa *et al.*, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, 2011.
[31] FEMA, "State and federal partners respond to the california wildfires," https://www.fema.gov/news-release/2018/11/17/4407/state-and-federal-partners-respond-california-wildfires, Nov. 2018.
[32] A. Saffari *et al.*, "On-line random forests," in *ICCV Workshops*, 2009.