Resilient Disaster Communications in the Social-Media Era

JUNO-2 Meeting

October 11, 2019

K. K. Ramakrishnan
University of California, Riverside

Toru Hasegawa, Yuki Koizumi
Osaka University
Masakatsu Nishigaki, Tetsushi Ohki
Shizuoka University
Yoshinobu Kawabe
Aichi Institute of Technology

Importance of Communication for Disaster Management

- Communication is key to improving outcomes in the aftermath of a disaster
- Keys to an effective response to a catastrophic incident:
  - Effective communication within and among dynamically formed first responder teams
    - Public safety teams comprising: law enforcement, health, emergency, transport and other special services, depending on the nature and scale of the emergency
- First responders are not the only ones that can help. Increasingly, volunteers are playing a significant part in disaster management
- In the aftermath of a disaster, likely to face communication challenges
  - Infrastructure may be impacted
  - Lack of personnel to support emergency communications
- Complement with social media
- Security and Resiliency are major concerns

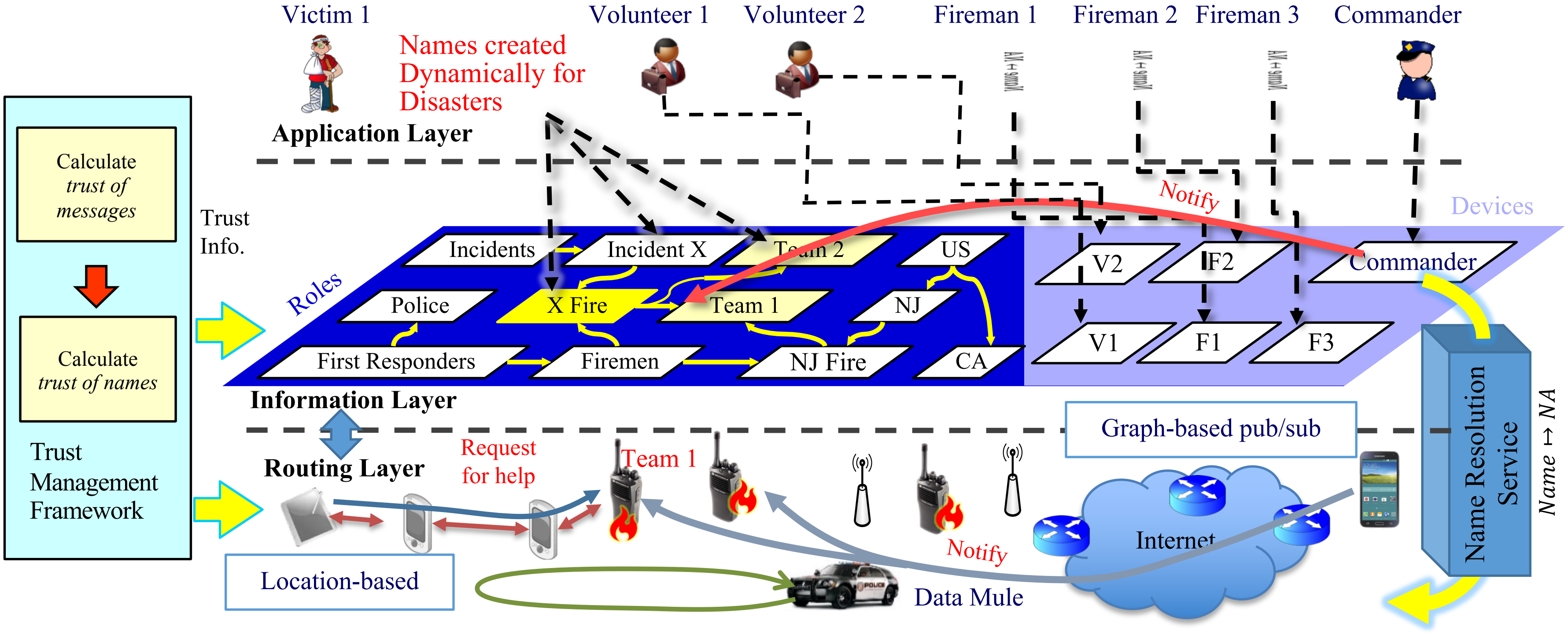
- **Project Objective: A network architecture for information and communication resilience in disaster management, that is also secure, integrates volunteers and social media seamlessly in disaster response**

Challenges

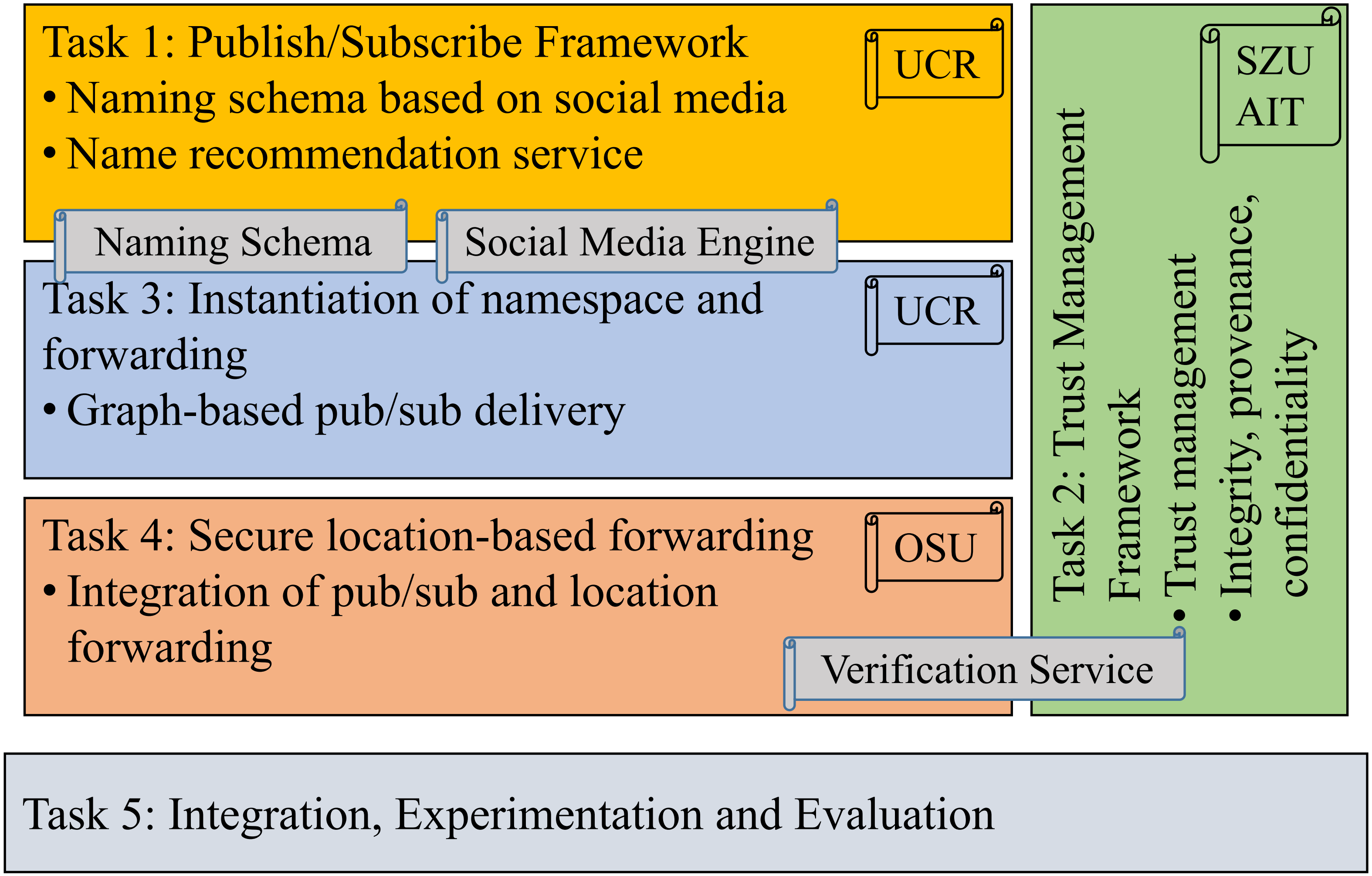
- Challenge 1: Designing a naming and forwarding framework in dynamic disaster environments, focusing on communications between and among honest first responders, while including trusted volunteers and victims
  - Task 1: Design, Creation and Instantiation of namespace
  - Task 3: Publish/Subscribe Framework; timely forwarding of relevant information
- Challenge 2: Security and resiliency against dishonest volunteers when the root of trust is lost
  - Task 2: Trust management
  - Task 4: Secure location-based forwarding

Proposed System Architecture

- **Information Layer -** (Role-Based) Communication
  - Facilitate communication: dynamically formed first-responder teams
    - Communication based on dynamically created roles, rather than locations
    - Include citizens (victims and volunteers) willing to help
- Secure and resilient: incorporate social media communications, based on Information Centric Networking (ICN)



Project Management



Task 1: Publish/Subscribe Framework

- Naming schema based on social media
- Name recommendation service

UCR

Naming Schema

Social Media Engine

Task 3: Instantiation of namespace and forwarding

- Graph-based pub/sub delivery

UCR

Task 4: Secure location-based forwarding

- Integration of pub/sub and location forwarding

OSU

Verification Service

Task 2: Trust Management Framework

- Trust management
- Integrity, provenance, confidentiality

SZU
AIT

Task 5: Integration, Experimentation and Evaluation

UCR: University California, Riverside OSU: Osaka University
SZU: Shizuoka University AIT: Aichi Institute of Technology

Summary of 1st Year

- We designed the architecture and submitted a joint paper to ICT-DM 2019
  - UC Riverside: used natural language processing pipelines/ML to map social media info' to map to the naming framework developed; analyzed data from 2 disasters in CA.
  - Osaka University designed Pub/Sub protocol for emergency calls (Globecom WS, 2019) and preliminarily evaluated performance of volunteers' crowdsourcing
  - Shizuoka University designed a biometric authentication of volunteers; did a questionnaire survey
  - Aichi Institute of Technology designed a trust model for volunteers (IFIPTM 2019)
- Publications
  - Joint paper is submitted to ICT-DM 2019 (Under review):Mohammad Jahanian, Toru Hasegawa, Yoshinobu Kawabe, Yuki Koizumi, Amr Magdy, Masakatsu Nishigaki,Tetsushi Ohki and K. K. Ramakrishnan, "DiReCT: Disaster Response Coordination with Trusted Volunteers"
  - Yoshinobu Kawabe, Yuki Koizumi, Tetsushi Ohki, Masakatsu Nishigaki, Toru Hasegawa and Tetsuhisa Oda, "On Trust Confusional, Trust Ignorant, and Trust Transitions," in Proceedings of 13th IFIP WG 11.11 International Conference on Trust Management (IFIPTM) 2019, July 2019.
  - Yuki Koizumi, Yoji Yamamoto and Toru Hasegawa, "Emergency Message Delivery in NDN Networks with Source Location Verification," to appear in Globecom 2019 Workshop, Dec. 2019.
- Meetings
  - US-JP: October 27-28, 2018 (Tokyo)
    Aug 2018 (UCR), March 28- 29, 2019 (UCR), August 25-26, 2019 (UCR)
    JP: November 15, December 16, 2018
    January 6, February, February 22, March 15, April 6, May 27,
    June 16, June 29, July 12, August 8, September 4

System Model

- Objective
  - Timely delivery of the right information to the right recipients in disasters
  - Disaster response coordination including trusted volunteers
- Players

First responder (FR)

- Perform tasks for disaster management

Volunteer

- Support tasks of first responders

Verifier

- Perform tasks to verify social media posts according to requests by the voting authority

Incident commander (IC)

- Send commands to FRs according to event reports from the VA

Voting authority (VA)

- Ask verifiers to check the credibility of social media posts
- Only send credible social media posts to the IC

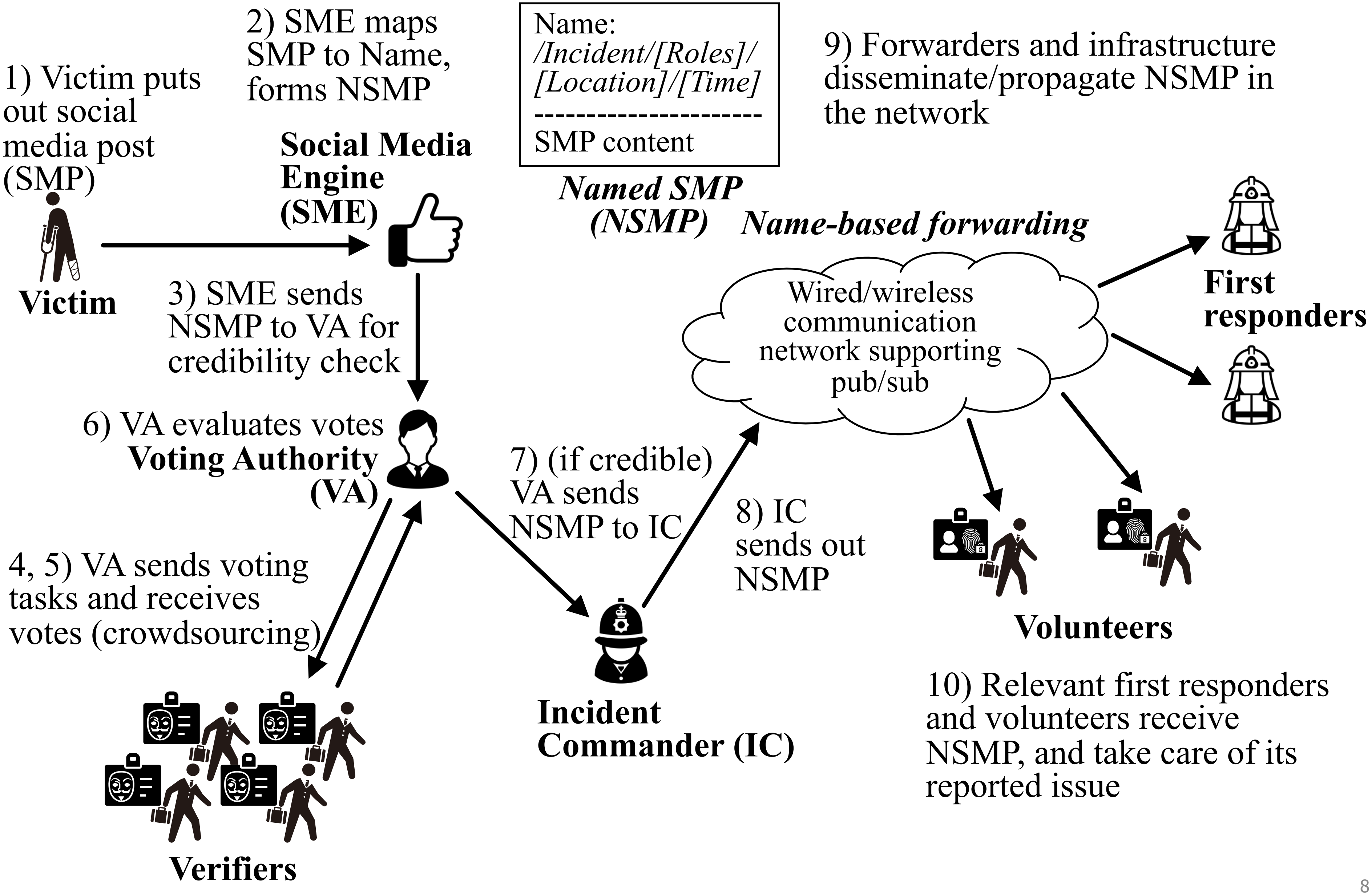
Social media engine (SME)

- Analyze social media posts to the social media and map them to the name space (named social media posts)

Victim

- Post messages to the social media

Scenario Walkthrough



1) Victim puts out social media post (SMP)

Victim

2) SME maps SMP to Name, forms NSMP

Social Media Engine (SME)

Name:
/Incident/[Roles]/[Location]/[Time]

SMP content

Named SMP (NSMP)

3) SME sends NSMP to VA for credibility check

4, 5) VA sends voting tasks and receives votes (crowdsourcing)

Verifiers

6) VA evaluates votes

Voting Authority (VA)

7) (if credible) VA sends NSMP to IC

Incident Commander (IC)

8) IC sends out NSMP

Name-based forwarding

Wired/wireless communication network supporting pub/sub

9) Forwarders and infrastructure disseminate/propagate NSMP in the network

First responders

Volunteers

10) Relevant first responders and volunteers receive NSMP, and take care of its reported issue

Architectural Components

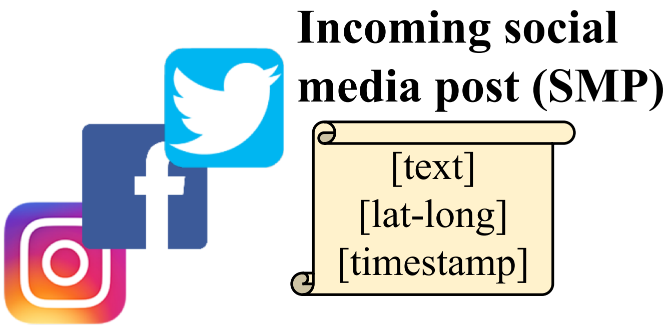
- Naming Schema
  - Unifies the interactions between all different actors (civilians, first responders, etc.) and guides the subscription and publication paths
  - Namespace represents entities related to and critical in incident management, and captures complex relations among them.
- Social Media Engine (SME)
  - Incoming social media posts (SMP), possibly including latitude/longitude, and timestamp, in addition to text, goes through a sequence of stages to be mapped to a (set of) name(s) of the namespace structure
  - Machine-learning based 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
- Verification Service
  - Set of crowdsourced voluntary verifiers check credibility of SMP. A majority vote used to bring only credible information into disaster response activity
  - Trust management system identifies trustworthy volunteers/verifiers based on biometric signature and a trust model

Social Media Engine

- Civilians use social media - free-form text
  - No knowledge of the namespace
  - Need: deliver relevant information to the right entities in the namespace
- Propose using a Social Media Engine (SME) to intelligently map social media posts (e.g., tweets) to the right name. Publishing to the appropriate first responders and volunteers
  - Using NLP/ML techniques
- SME pulls social media posts from Internet/server-based social media platforms, and analyzes and disseminates them

Social Media Engine

- Input: user-generated social media posts (e.g. Tweets) in an online mode
  - May contain text, geo info (e.g., lat-long), and timestamp

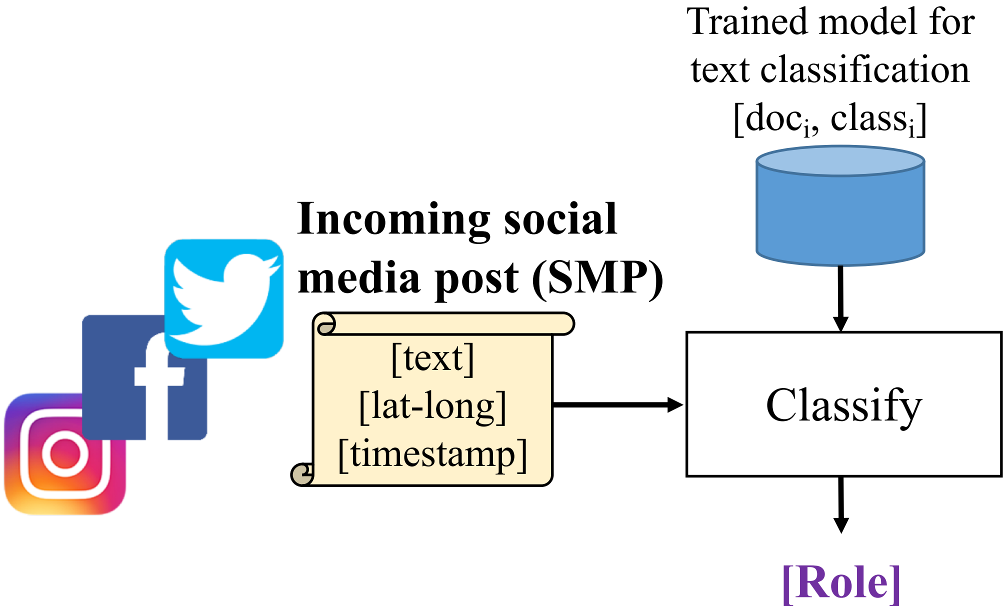


Incoming social media post (SMP)

[text]
[lat-long]
[timestamp]

Social Media Engine

- Extract incident role associated with the text of social media post (SMP)
  - Using supervised classification
  - Trained model from previous/similar disasters, labeled according to namespace template



Trained model for text classification [doc_i, $class_i$]

Incoming social media post (SMP)

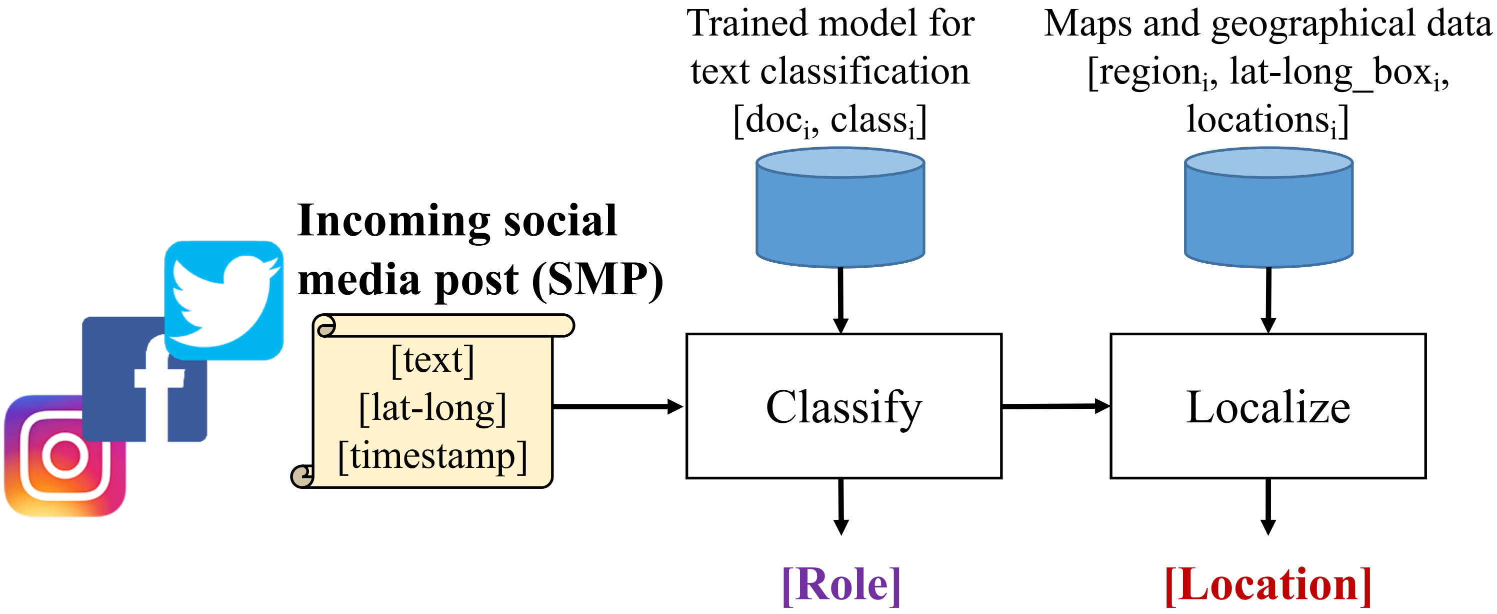
[text]
[lat-long]
[timestamp]

Classify

[Role]

Social Media Engine

- Extract location from the SMP
  - Can be geo-tag metadata, or location names in the text



Trained model for text classification $[doc_i, class_i]$

Maps and geographical data $[region_i$, lat-long_box_i, $locations_i]$

Incoming social media post (SMP)

[text]
[lat-long]
[timestamp]

Classify

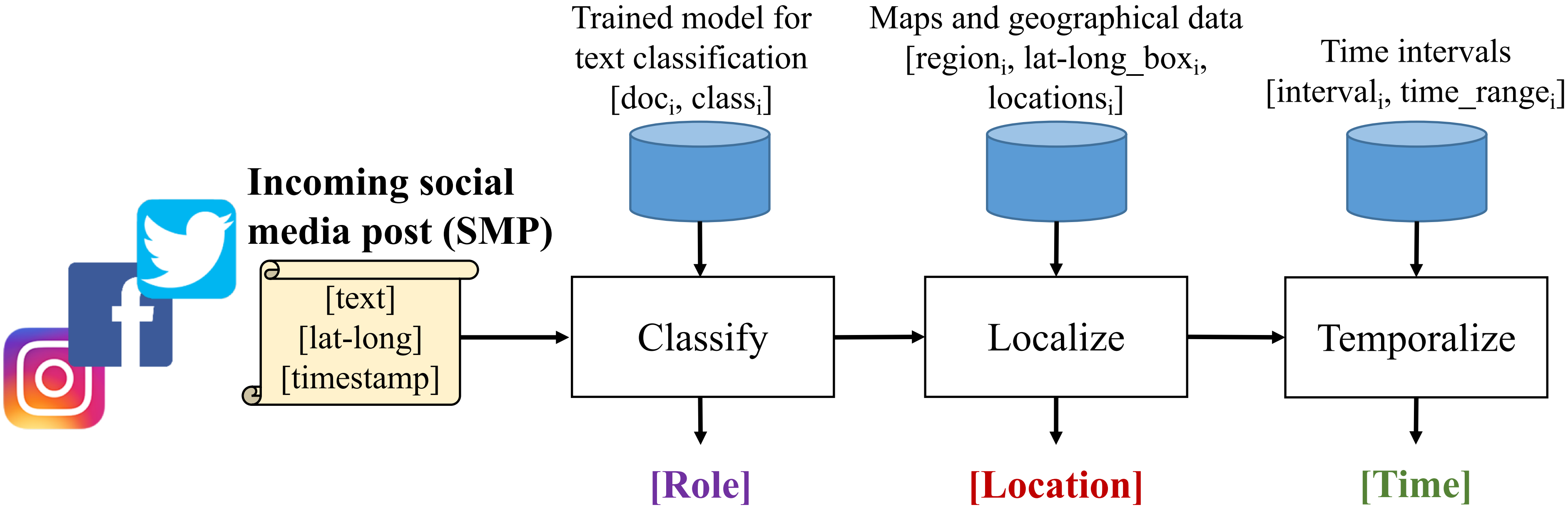
Localize

[Role]

[Location]

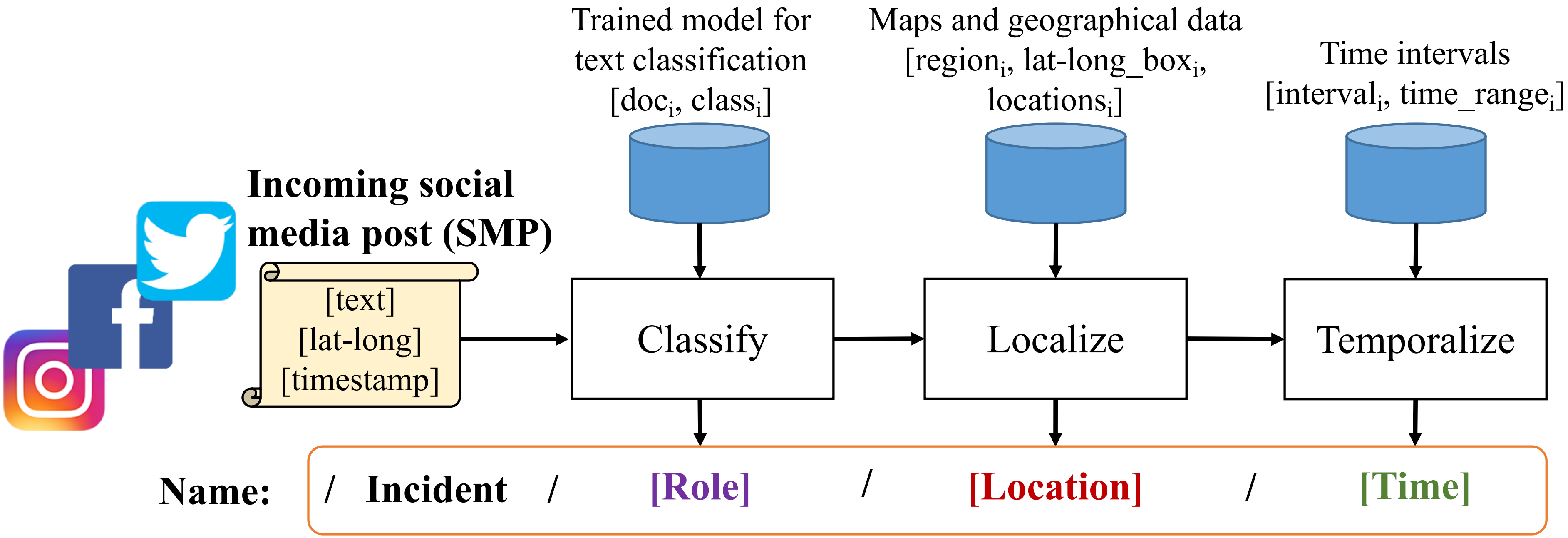
Social Media Engine

- Extract time stamp from SMP and map it to the right time interval


Trained model for text classification
[doc$_i$, class$_i$]
Maps and geographical data
[region$_i$, lat-long_box$_i$, locations$_i$]
Time intervals
[interval$_i$, time_range$_i$]
Incoming social media post (SMP)
[text]
[lat-long]
[timestamp]
Classify
Localize
Temporalize
[Role]
[Location]
[Time]

Social Media Engine

- Using these extracted elements, we can form the “name” (e.g., a hierarchical name), to use for publication into the network



Incoming social media post (SMP)

[text]
[lat-long]
[timestamp]

Trained model for text classification [doc$_i$, class$_i$]

Maps and geographical data [region$_i$, lat-long_box$_i$, locations$_i$]

Time intervals [interval$_i$, time_range$_i$]

Classify

Localize

Temporalize

Name: / **Incident** / **[Role]** / **[Location]** / **[Time]**

Social Media Engine

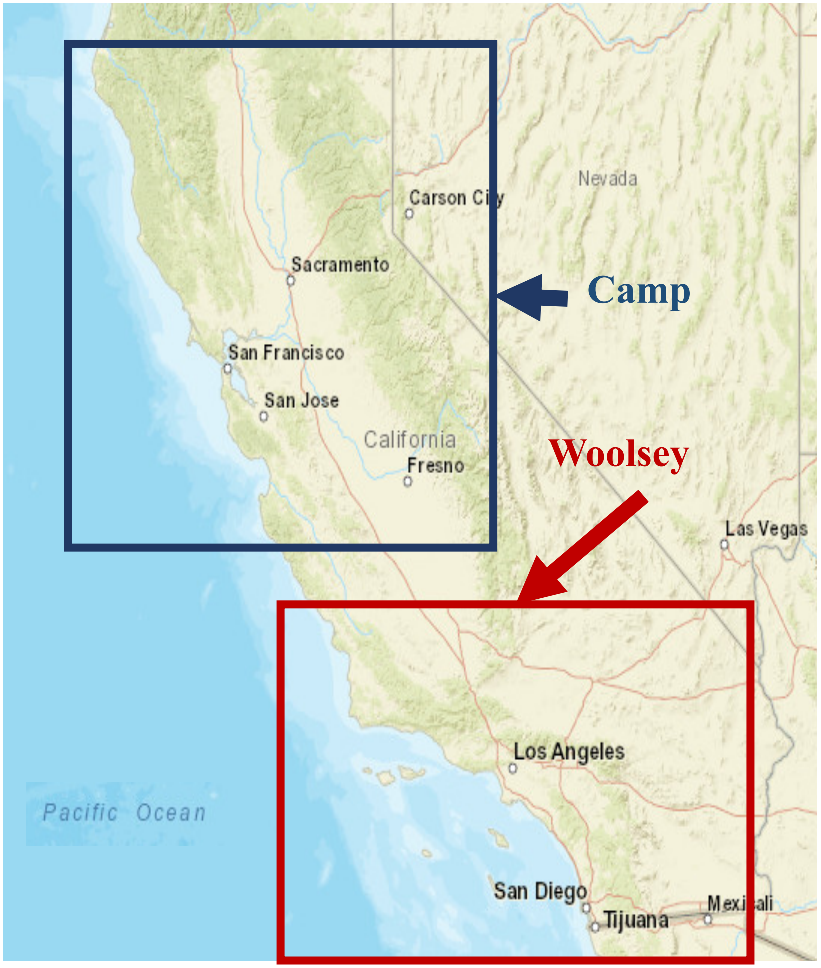
- There is a (small?) chance of inaccuracy in mapping
  - Example: a medical doctor receives a report regarding an urgent need for fighting a fire
  - In such a case, he/she can either: 1) re-publish the SMP to the network picking the right names; or 2) send it as a unicast message to his/her incident commander
  - This provides resiliency against inaccurate mapping and delivery

- SME does not determine veracity and importance of the SMP
  - Veracity is determined via crowd-sourced verification service
  - Importance is determined by the relevant first responder

- Future work: automatic veracity and importance prediction by the SME

Case Study: California Wildfires in 2018

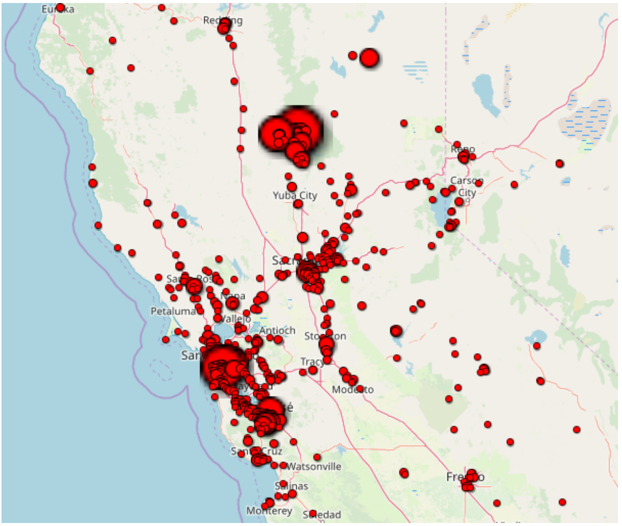
- To evaluate performance of SME:
- We collected tweets for two major wildfires in California, in 2018
  - From Nov. 7th to Nov. 26th, 2018
  - Geo bounding boxes shown on map
  - Camp Fire
    - 959,740 tweets
    - Northern California
  - Woolsey Fire
    - 1,961,131 tweets
    - Southern California



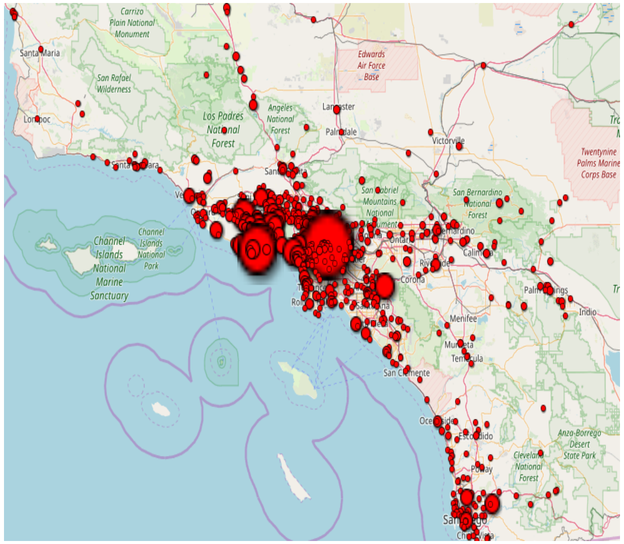
Spatial and Temporal Analysis

- Extract and analyze disaster-related tweets
- Spatial correlation: Higher density of disaster-related tweets near the more affected areas

Camp



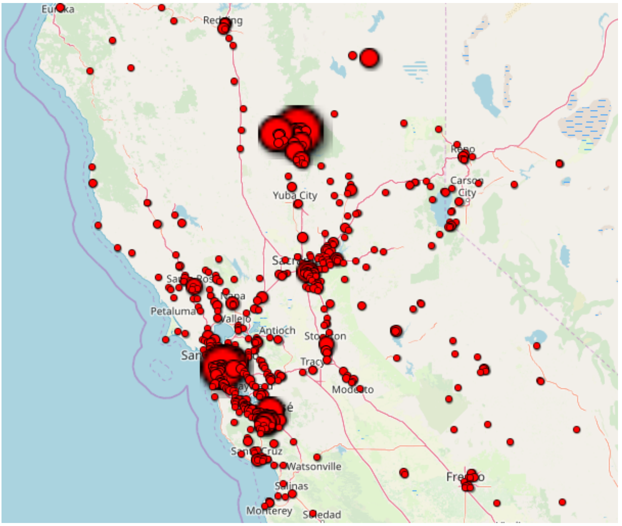
Woolsey

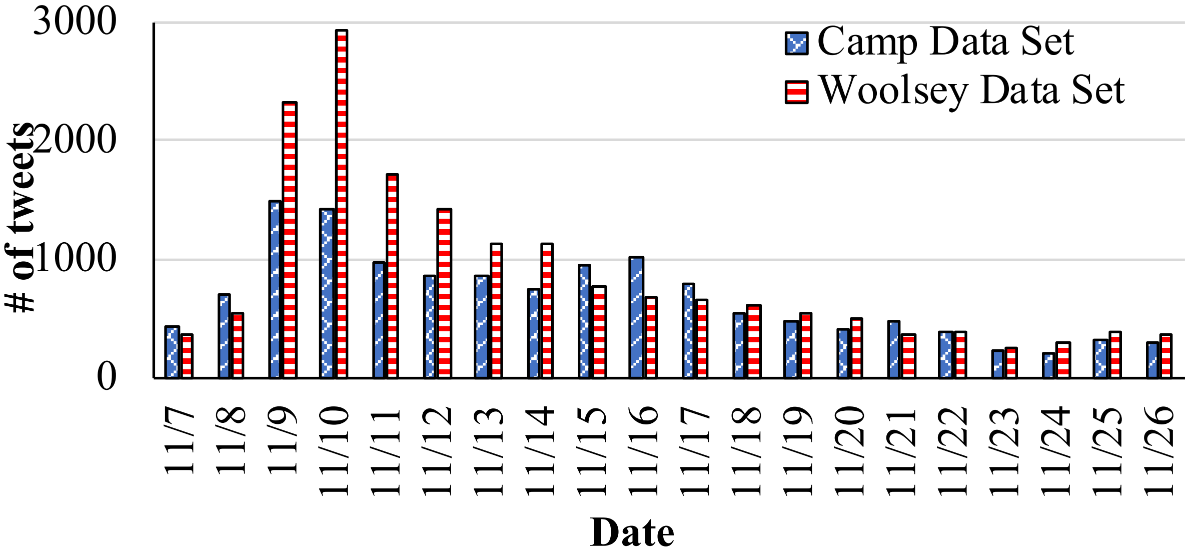

Carrizo Plain National Monument
Santa Maria
San Rafael Wilderness
Lompoc
Los Padres National Forest
Angeles National Forest
Lancaster
Palmdale
Edwards Air Force Base
Victorville
Twentynine Palms Marine Corps Base
San Gabriel Mountains National Monument
San Bernardino National Forest
Channel Islands National Marine Sanctuary
Channel Islands National Park
Corona
Menifee
Murrieta
Temecula
San Clemente
Indio
Escondido
Poway
Cleveland National Forest
Anza-Borrego Desert State Park

Spatial and Temporal Analysis

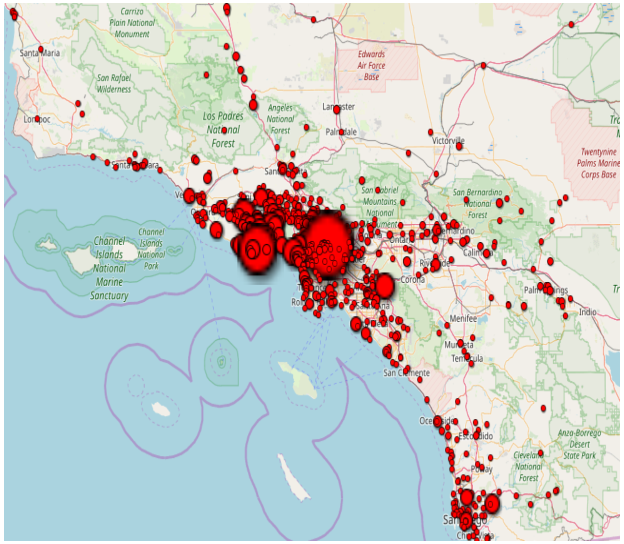
- Extract and analyze disaster-related tweets
- Spatial correlation: Higher density of disaster-related tweets near the more affected areas
- Temporal correlation: Higher density of disaster-related tweets during more intense days of the wildfires

Camp





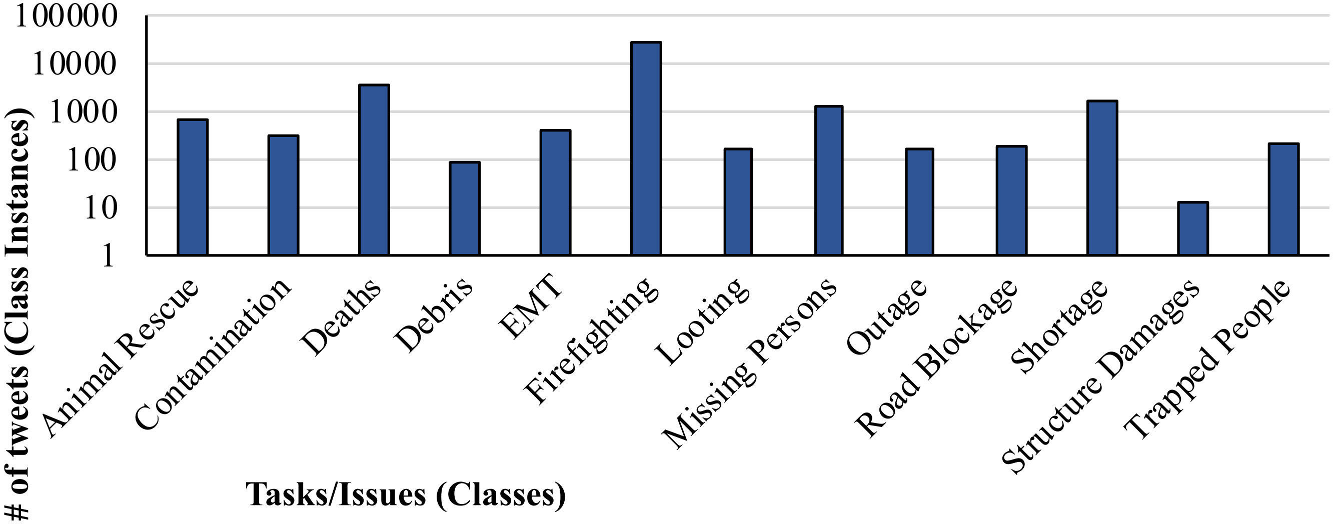
Woolsey


Carrizo Plain National Monument
Santa Maria
San Rafael Wilderness
Lompoc
Los Padres National Forest
Angeles National Forest
Lancaster
Palmdale
Edwards Air Force Base
Victorville
Twentynine Palms Marine Corps Base
Channel Islands National Marine Sanctuary
Channel Islands National Park
San Gabriel Mountains National Monument
San Bernardino National Forest
Corona
Menifee
Murrieta
Temecula
San Clemente
Oceanside
Escondido
Poway
Cleveland National Forest
Anza-Borrego Desert State Park
San Diego

Social Media Engine Procedure

- Data sets: we filter out tweets irrelevant to disaster
  - Woolsey: ~23K tweets; Camp: ~12K tweets
- Training data: Woolsey data set; Test data: Camp data set
- Annotate data with incident-related labels (task/issue-driven roles in the namespace) based on keywords

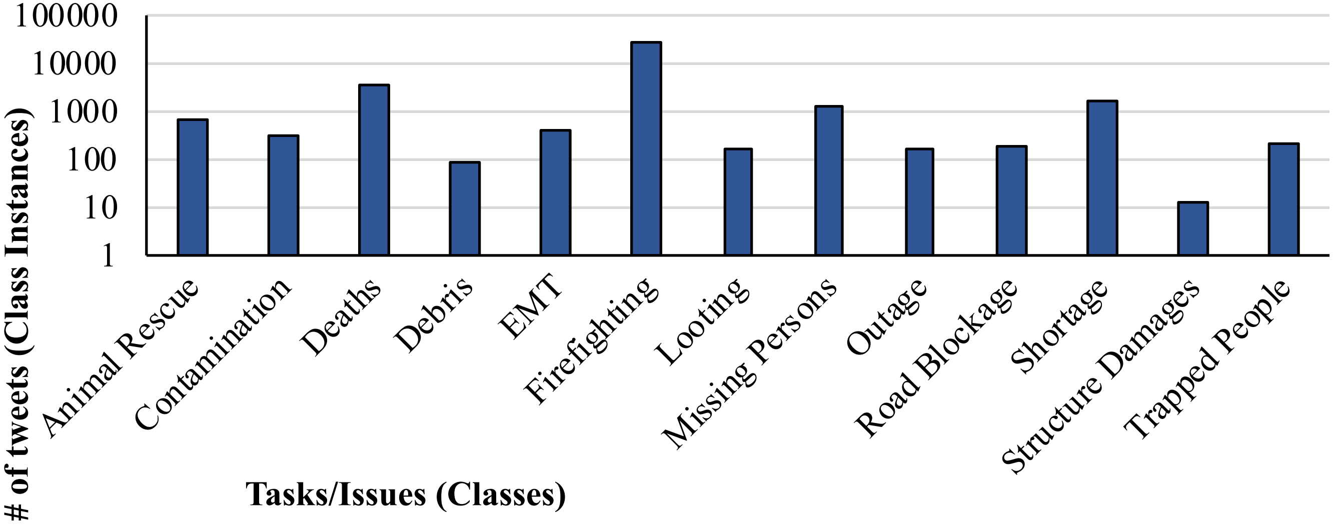
Tweets per label distribution (‘Firefighting’ is majority)



SME Procedure

- Data sets: filter out tweets irrelevant to disaster
  - Woolsey: ~23K tweets; Camp: ~12K tweets
- Train data: Woolsey data set; Test data: Camp data set
- Annotate data with incident-related labels (task/issue-driven roles in the namespace) based on keywords
- Learning procedure: used tf-idf vectorization (how important a term (1-2 words) is to a document in a collection): ‘feature-value’ vector
- Classification using Random Forest for mapping from tweets to names

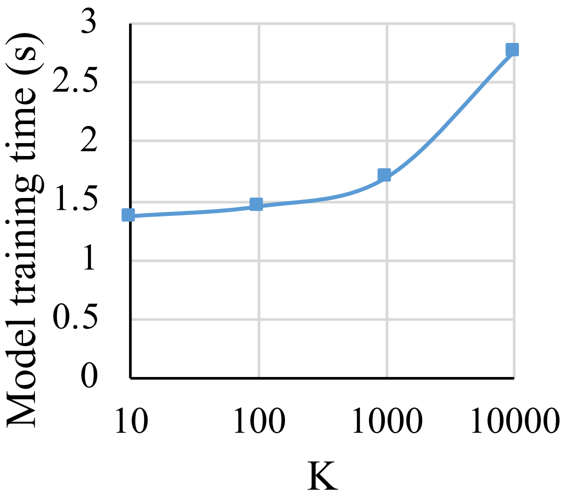
Tweets per label distribution (‘Firefighting’ is majority)

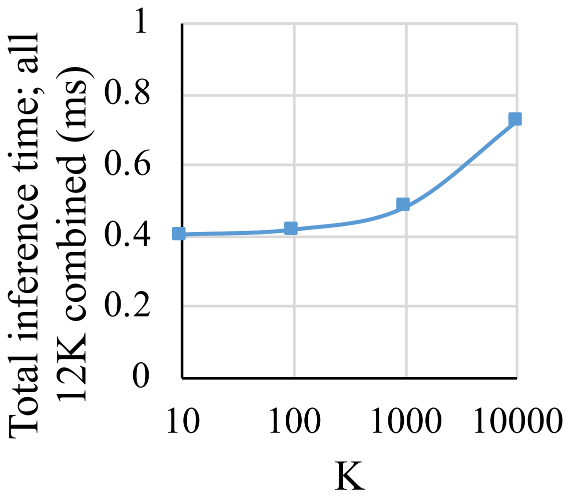


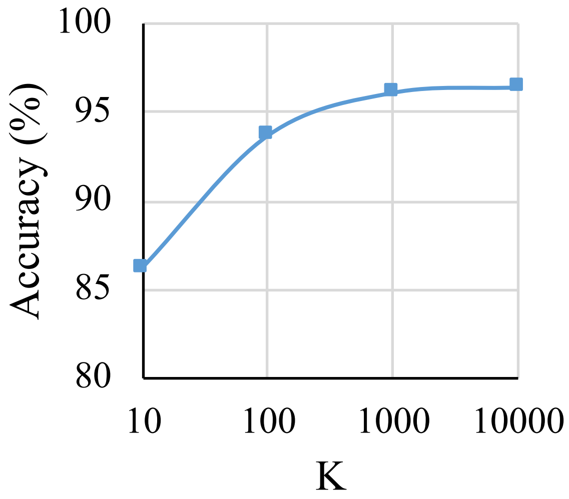
tf–idf: term frequency–inverse document frequency

Social Media Engine Results

- Feature selection to prevent overfitting, & reduce processing overhead
  - K-best feature selection using chi^2 test measure (intelligently picks the top K, most relevant features)







- Selection of K has impact
  - For K>1000 training and inference time exponentially higher, but almost no accuracy increase
  - Picking K=1000 is reasonable, based on accuracy & performance
- Server machine: Ubuntu machine with Intel(R) Xeon(R) CPU E5-2650 v4 and @2.20GHz dual-socket with 14 cores 252GB RAM

SME Results

- Achieve 96% accuracy (w/ K=1000)
  - Implies 96% of tweets would be mapped correctly and reach the right first responders to deal with the issue
    - Contrast: current unstructured way of re-tweeting - may be too difficult to rely on ad-hoc re-tweets
  - Only 4% would be mapped incorrectly
    - Our system can recover: through re-publishing or sending to incident commander from incorrectly reached first responders
- Other important metrics
  - Micro average is a better metric for our experiment, since our data set is imbalanced (~%70 of tweets belong to one class, namely ‘Firefighting’)

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

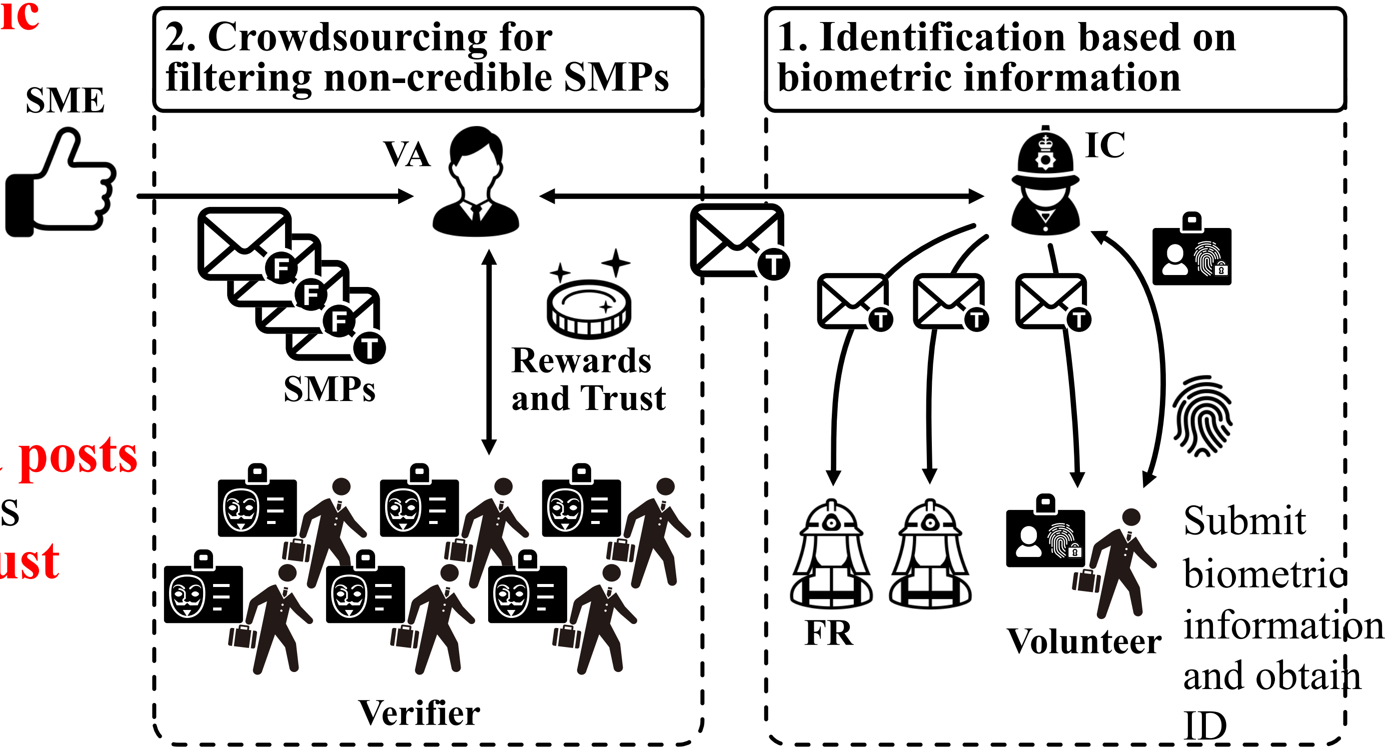
Macro average: sum of metric values (e.g., Precision) for all classes, divided by the number of classes (i.e., class-by-class averaging).

Micro average: average of metric values taking into account the number of per-class instances as well (i.e., item-by-item averaging).

Verification Service

- Objectives
  1. To identify the **trustworthiness of volunteers** to incorporate only **trustworthy volunteers** to disaster management
  2. To identify the **credibility of social media posts** among a vast amount of disaster-related social media posts to deliver only **credible social media posts** to the incident commander

- Approaches
  1. To use **biometric information** for creating the IDs of volunteers
  2. To develop **crowdsourced verification of social media posts** with anonymous verifiers and **trust management of the verifiers**


SME
2. Crowdsourcing for filtering non-credible SMPs
VA
F
F
F
T
SMPs
Rewards and Trust
Verifier
1. Identification based on biometric information
IC
T
T
T
T
FR
Volunteer
Submit biometric information and obtain ID

Questionnaire: Survey for Biometric Information Exposure

- Understand how exposing personal information including biometric information prevents volunteers from doing malicious behavior

Requirements

RQ1. Does personal information act as a deterrent to false information transmission?

- Biometric information: face, fingerprint, voiceprint
- Device information: driver's license, mobile phone number
- Census-register information:
  name + address, name + address + date of birth + gender

RQ2. Is the degree of exposure to privacy equal to the deterrence by personal information?

RQ3. Does the effect of deterrence change according to the degree of lie?

- Prank, Distribution, Rescue

Survey

Survey1. Question for assessing privacy score (*Sp*)

- When you present your personal information to other, how much do you think you are exposing your privacy?

Survey2. Question for assessing deterrence score (*Sd*)

- When a person has to register his personal information, do you think that he will send a malicious message?

66 valid answers out of 200 subjects recruited on crowdsourcing platform

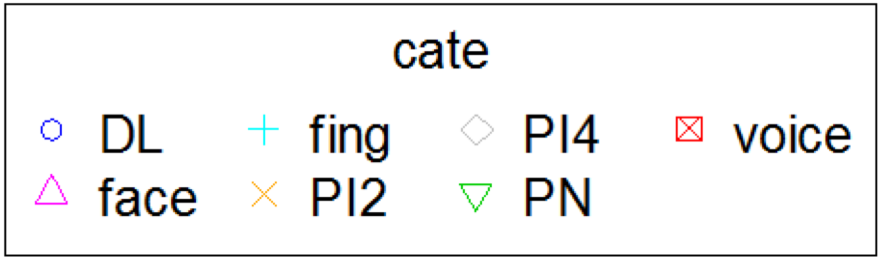
Survey Results: Which personal information pieces contribute to deterrence?

Privacy score *Sp*, Deterrence score *Sd*, Effect score *Sd-Sp*: avg(SD)

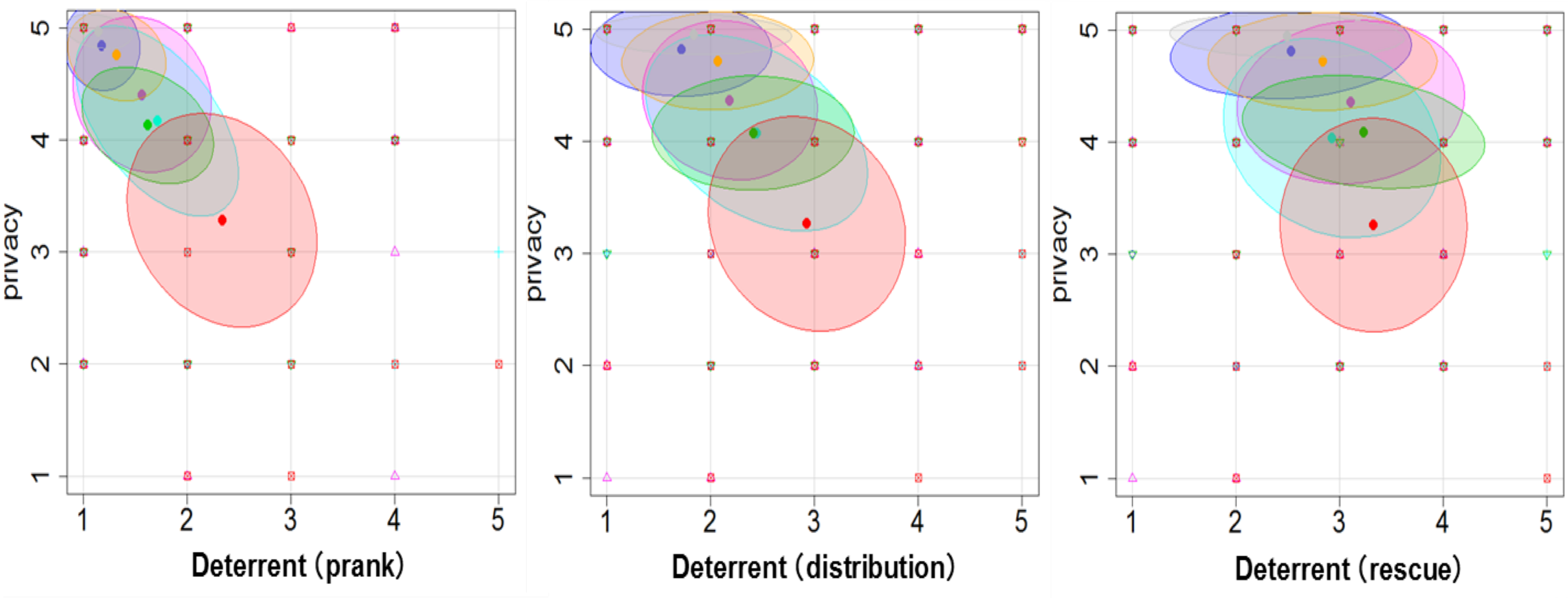
item info	Privacy score	Deterrence score			Effect score		
		Prank	Distribution	Rescue	Prank	Distribution	Rescue
Non	-	1.76 (1.14)	1.98 (1.25)	1.78 (1.21)	-	-	-
Face	4.17 (1.14)	4.24 (1.02)	3.68 (1.17)	3.05 (1.40)	0.07 (1.42)	-0.49 (1.65)	-1.12 (1.93)
Fing	3.93 (1.15)	4.10 (1.11)	3.44 (1.34)	3.07 (1.31)	0.17 (1.16)	-4.89 (1.38)	-0.85 (1.56)
voice	3.27 (1.23)	3.61 (1.18)	3.05 (1.22)	2.73 (1.18)	0.34 (1.54)	-0.22 (1.62)	-0.54 (1.67)
PI4	4.80 (0.459)	4.63 (0.662)	3.90 (1.32)	3.39 (1.38)	-0.17 (0.738)	-0.90 (1.37)	-1.41 (1.40)
PI2	4.61 (0.628)	4.51 (0.746)	3.80 (1.21)	3.20 (1.36)	-0.10 (0.86)	-0.80 (1.38)	-1.41 (1.53)
DL	4.61 (0.771)	4.63 (0.733)	4.10 (1.20)	3.51 (1.43)	-0.02 (1.07)	-0.51 (1.47)	-1.10 (1.73)
PN	4.02 (0.790)	4.27 (0.922)	3.54 (1.27)	2.88 (1.47)	0.24 (1.04)	-0.49 (1.52)	-1.15 (1.51)

non: nothing   face: face   fing: fingerprint   voice: voiceprint
PI2: name + address   PI4: name + address + date of birth + gender
DL: driver's license   PN: mobile phone number

Scatter diagram of Privacy score vs Deterrence score



Vertical axis: privacy score
Horizontal axis: deterrence score
drawn a 30% probability ellipse



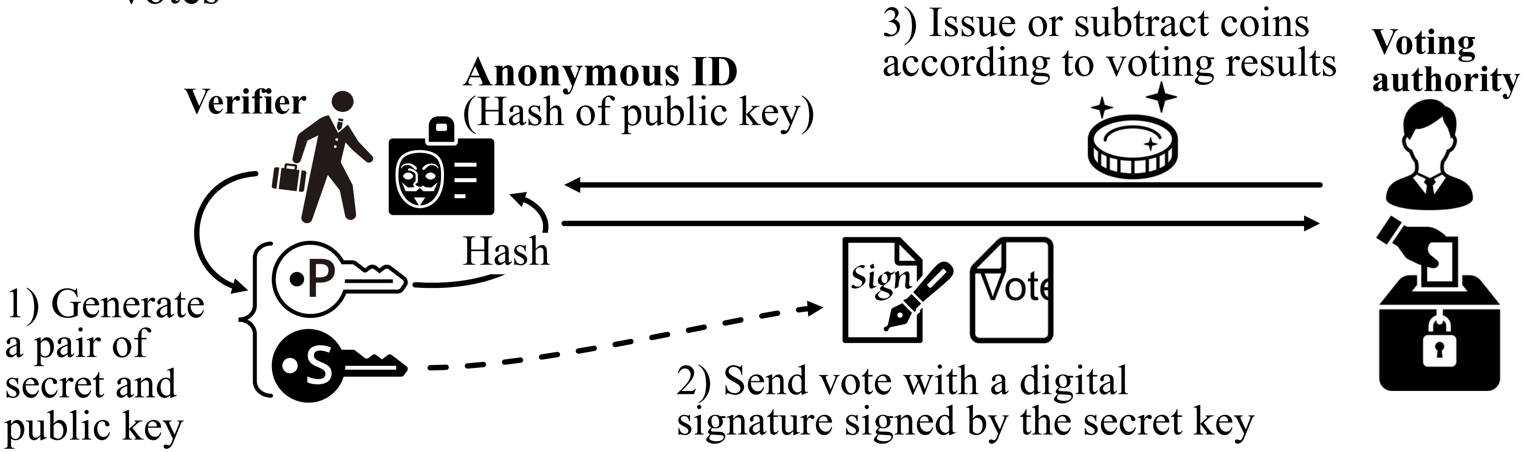
Observations

- Privacy score roughly equals to deterrence score
  - In the case that we need to prevent malicious behavior as much as possible
    - Privacy concern is of secondary importance
    - Solution is to ask volunteers to enroll their Driver's licenses
  - In the case that we need to protect privacy as much as possible
    - Deterrence to malicious behavior is not strong
    - Solution is to ask volunteers to enroll their Voiceprints

- Voiceprint is the most effective modality, because
  - Biometric information, has the biggest value of "deterrence score - privacy score"
  - Degree of "deterrence to malicious acts" is bigger than the degree of "privacy concern"

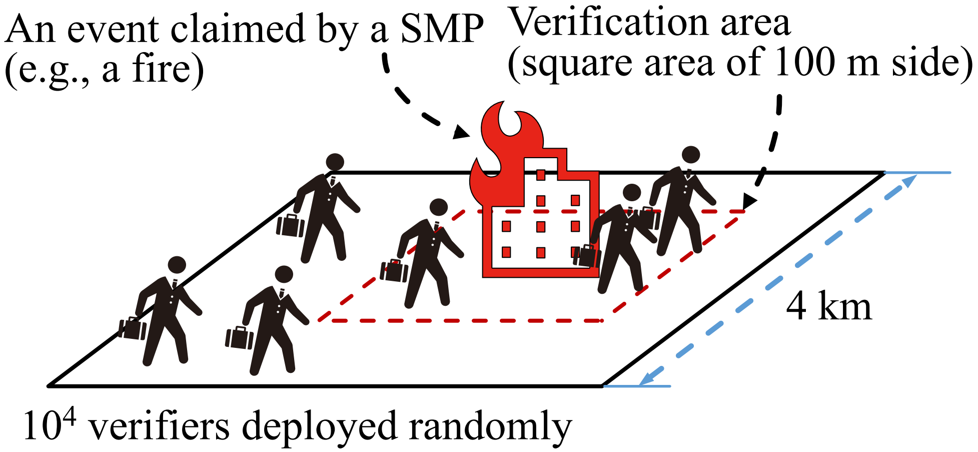
Crowdsourced Social Media Post Verification

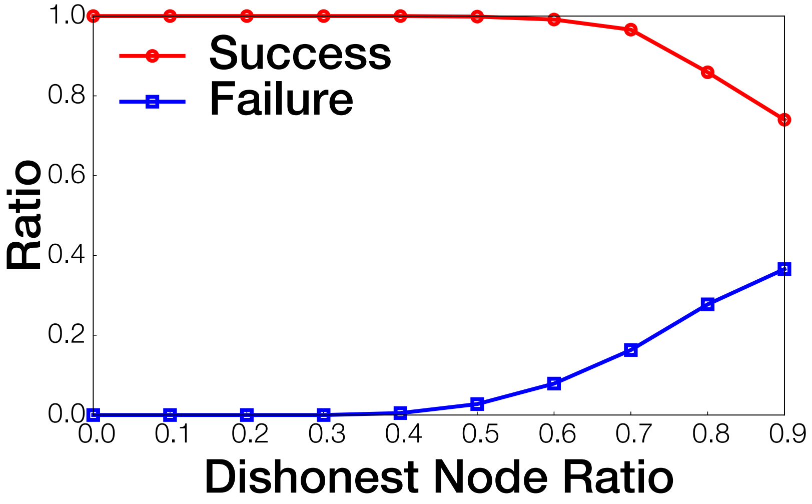
- Challenges
  1. To recruit as many verifiers as possible
  2. To prevent wrong votes from dominating the total votes because **verifiers are potentially dishonest**
- Solution
  1. Easy registration mechanism
     - Use self-generated public and secret key and use the hash of the public key as an ID
  2. Trust management system
     - Each verifier has her/his trust value, which is managed as virtual coins named trust coins
     - A vote of each verifier is weighted by her/his trust value
     - Trust values of verifiers are decreased/increased depending on having made wrong/correct votes


3) Issue or subtract coins according to voting results
Voting authority
Anonymous ID
(Hash of public key)
Verifier
Hash
P
S
Sign
Vote
1) Generate a pair of secret and public key
2) Send vote with a digital signature signed by the secret key

Evaluation of Crowdsourced Verification

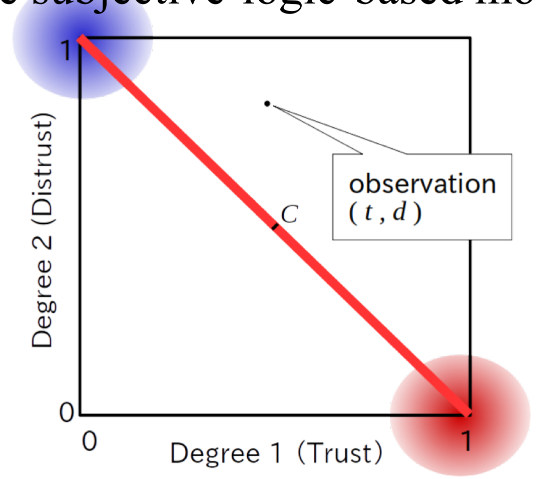
- Results
  - Most of fake SMPs are successfully filtered by the crowdsourced verification even if the majority of verifiers behave dishonestly
  - Comparison of this approach with other approaches, such as subjective logic, are under way
    - Extending the trust management so that the trust is defined in a two-dimensional space based on fuzzy logic (the next slide)
- Conditions
  - Overview
    - Ask verifiers around an event reported by a SMP and identify if the SMP is credible
  - Details
    - Verifiers: Dishonest verifiers make wrong votes with probability of 0.5
    - SMPs: A SMP is posted every 10 minutes reporting an event of randomly chosen point


An event claimed by a SMP (e.g., a fire)
Verification area (square area of 100 m side)
4 km
10^4 verifiers deployed randomly



Trust Model for Volunteers

- Conventional trust theory assumed that human evaluators could calculate the degree of "distrust" if the degree of "trust" on a trustee was given
  - Marsh, S., Dibben, M.R.: Trust, untrust, distrust and mistrust – an exploration of the dark(er) side. In: Proceedings of the Third International Conference on Trust Management. pp. 17-33. iTrust'05 (2005)
- However, such assumption is too strong, and we introduced a 2D trust model where trust and distrust degrees are independent.
- By applying Fuzzy logic, trust notions (trust, distrust, and untrust) developed in the conventional 1D trust theory are naturally extended
- We compared our 2D model and subjective-logic-based 2D model
  - Our 2D model can deal with contradictory situations (ex. a consistent message from an unknown sender), while the subjective-logic-based model cannot do.

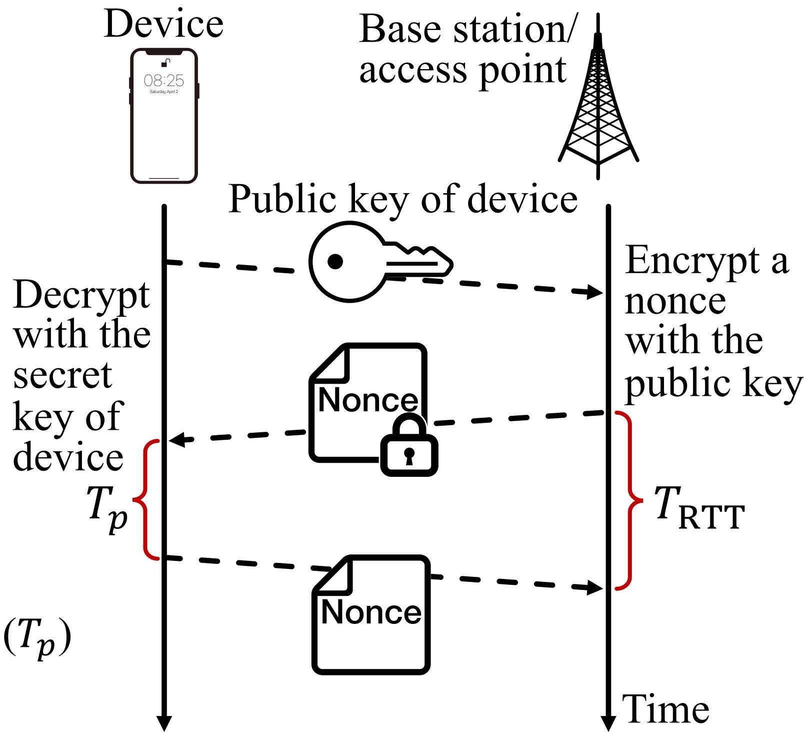


Emergency Communication with Location Verification

- Design principles
  1. Emergency message delivery with specifying a well-known name on top of pub/sub networks
  2. Secure emergency message delivery with attribute-based encryption (ABE)
  3. Location verification based on measurements of signal propagation delay
- Location verification
  - A base station (or access point) measures RTT to a device by sending a nonce
  - The claimed location of the device (l_d) is correct if the claimed difference and the estimated distance is sufficiently small

$$\underbrace{\frac{T_{\mathrm{RTT}}}{2s^{-1}}}_{\substack{\text{Estimated distance from}\\ \text{the measured RTT}}} - \underbrace{d(l_d, l_b)}_{\substack{\text{Distance between the BS location } (l_b) \text{ and}\\ \text{the location claimed by the device } (l_d)}} \le \underbrace{\varepsilon}_{\substack{\text{Estimation error caused by the}\\ \text{processing delay } (T_p)}}$$

S: the speed of signal
d: the distance between two locations



Device
08:25
Base station/ access point
Public key of device
Encrypt a nonce with the public key
Decrypt with the secret key of device
Nonce
T_p
T_{RTT}
Nonce
Time

Publications

- Joint paper
  - Submitted to ICT-DM 2019
    - Mohammad Jahanian, Toru Hasegawa, Yoshinobu Kawabe, Yuki Koizumi, Amr Magdy, Masakatsu Nishigaki,Tetsushi Ohki and K. K. Ramakrishnan, “DiReCT: Disaster Response Coordination with Trusted Volunteers”
- Paper
  - Yoshinobu Kawabe, Yuki Koizumi, Tetsushi Ohki, Masakatsu Nishigaki, Toru Hasegawa and Tetsuhisa Oda, “On Trust Confusional, Trust Ignorant, and Trust Transitions,” in Proceedings of 13th IFIP WG 11.11 International Conference on Trust Management (IFIPTM) 2019, July 2019.
  - Yuki Koizumi, Yoji Yamamoto and Toru Hasegawa, “Emergency Message Delivery in NDN Networks with Source Location Verification,” to appear in Globecom 2019 Workshop, Dec. 2019.

Conclusion and Plan of the 2nd Year

- Summary
  - Designed the architecture, DiReCt, so that timely delivery of the right information to the right people can improve outcomes and save lives
  - Designed architectural components and preliminarily evaluated them
- Plan of the 2nd year
  - Revise the architectural components so as to capture how people behave in disasters
    - Integrate the verification service and the trust model
  - Empirically evaluate the architecture based on data which is derived from SMP and auxiliary information at a disaster (e.g., Typhoon in 2018)