Resilient Disaster Communications in the Social-Media Era

JUNO-2 Kick-off Meeting

October 25, 2018

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

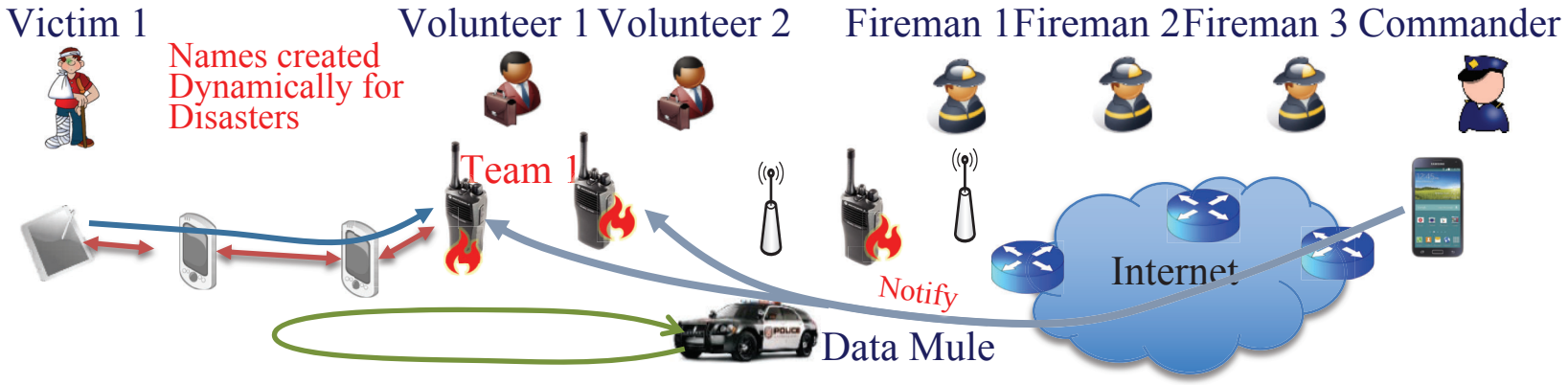
Importance of Communication for Disaster Management

- 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
- Lack of personnel to support emergency
- In the aftermath of a disaster, likely to face communication challenges
  - Infrastructure may be impacted
- Complement with social media with data communications: Security?

- Security and Resiliency are major concerns
- **<u>Project Objective:</u> A network architecture for information and communication resilience in disaster management that is also secure; integrate volunteers; include social media**

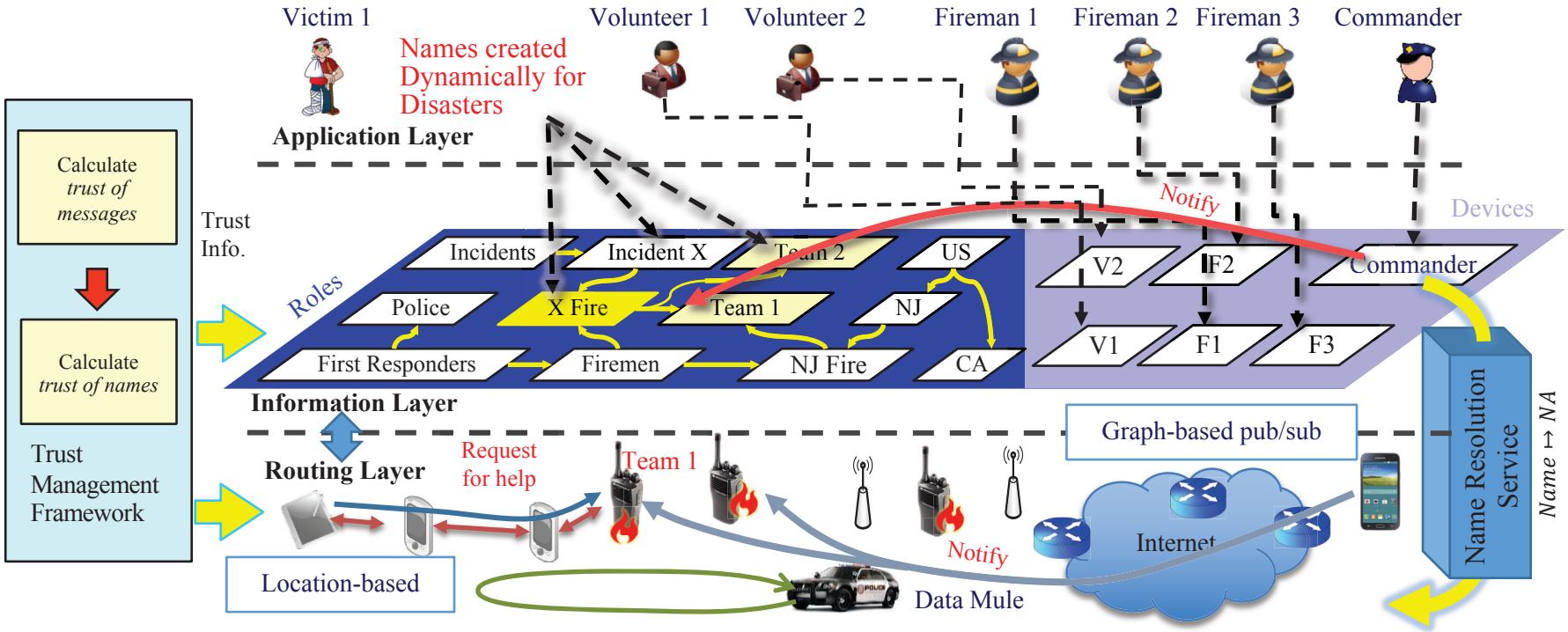
System Model and Assumptions

- Network Model
  - Enhanced Information Layer building on Information Centric Networking (ICN) concepts, with a Publish/Subscribe service
  - Multi-hop communication to allow communication even in fragmented networks, and disruptions
- Security Model
  - Honest Players: First responders and incident commander
  - Potentially Dishonest Players: Volunteers and Victims
    - No long term history/reputation available for use as basis of trust
- Safely manage information flow and support rescue efforts



Proposed System Architecture

- **Information Layer -** (Role-Based) Communication
  - Facilitate communication: dynamically formed first-responder teams
    - Communication based on dynamically created roles, not network locations
    - Include citizens, including victims and volunteers willing to help
- Secure and resilient; integrate social media communication into an Information Centric Networking (ICN) framework



Challenges

- Challenge 1: Designing a naming and forwarding framework in dynamic disaster environments, focusing on communication among honest first responders, and including trusted volunteers and victims

- Relationships among participants are dynamic and, often, transient

  - Task 1: Instantiation of namespace
    - Naming scheme for players in disasters
    - Graph-based naming scheme: Multi-rooted tree structure for representing multiple organizations

  - Task 2: Publish/Subscribe Framework and forwarding
    - Publish/Subscribe forwarding mechanism for graph-based name prefixes
    - Name prefix distribution to routers and participants

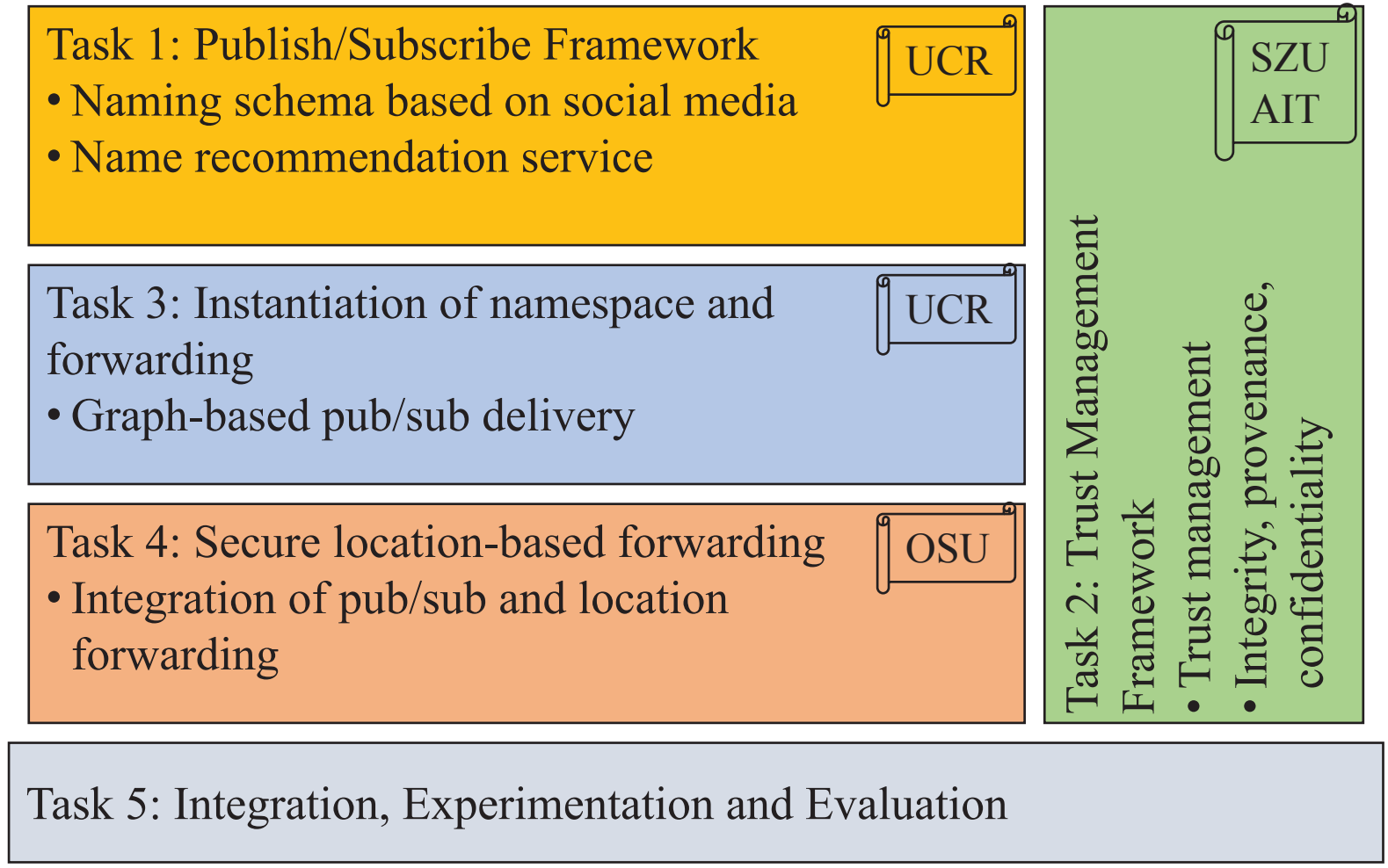
Challenges - continued

- Challenge 2: Security and resiliency against dishonest volunteers when the root of trust is lost

  - Task 3: Trust management
    - Managing trust of volunteers and victims without using any certification authority
    - Providing trust information, with first-responders choosing the method for secure communication

  - Task 4: Secure location-based forwarding
    - Protecting privacy (names) of first responders from volunteers
    - Extending the forwarding framework against malicious forwarders (volunteers) in ad-hoc and disruptive environments

Project Management



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

Collaboration and Joint Research Efforts

- Build on long (6-7 year) history of collaboration among several members of the team
- Expect to have bi-weekly or monthly calls between PIs
  - Include students as and when progress is being shared and ensure all are in synch.
- Relatively frequent face-face meetings
  - Already had one face-face kick-off meeting in UC Riverside in Aug. 2018
  - Will meet here in Tokyo during this visit
- Where possible, have students from participating institutions visit for an extended period of time to enable closer sharing of artifacts, development of prototype.

Publish/Subscribe Framework

- Overview and Motivation
  - Delivering the messages to the right people (i.e., first responders dealing with the particular incident, and/or volunteers nearby that can help)

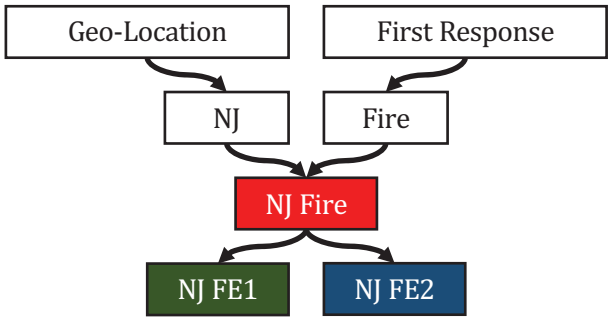
- Objective and Intellectual Merit
  - An information layer for disaster management and clearly integrating social media information so we can automate the dynamic matching among victims, first responders and volunteers in a secure and timely manner

- Research Challenge
  - A graph-based naming framework, so that the relationships come automatically from the social media data generated in calls for help in a real disaster
    - An acyclic directed graph that has multiple roots

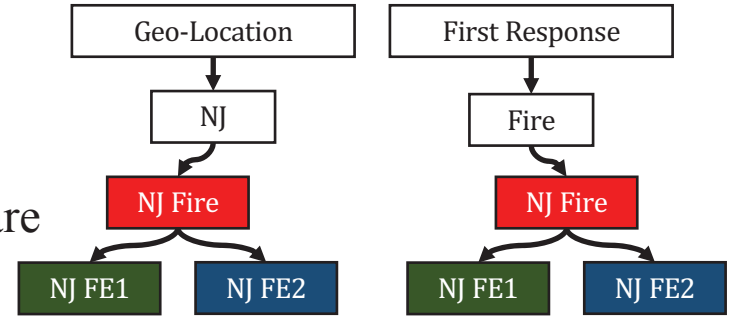
Namespace Design

- **Multi-dimensional**
  - E.g. `FireEngine1` has Time, Location and Department attributes (dimensions)
- **Graph structure**
  - More efficient than NDN-style strict hierarchy
- **Dynamic**
  - Edges (relations) pop in and out of existence
- **Publish/Subscribe service interface**
  - Support a publish/subscribe capability for users to share information
  - Multiple entities can publish to a name
  - Uses a shared multicast structure in network, using rendezvous points (RPs)

Graph Structure



Hierarchical Structure



Hierarchical names:

```
/Geo-Location/NJ/NJ Fire
/First Response/Fire/NJ Fire
/Geo-Location/NJ/NJ Fire/NJ FE1
/First Response/Fire/NJ Fire/ NJ FE1
/Geo-Location/NJ/NJ Fire/NJ FE2
/First Response/Fire/NJ Fire/ NJ FE2
```

Improve Efficiency of Disaster Management by Graph-based Namespace

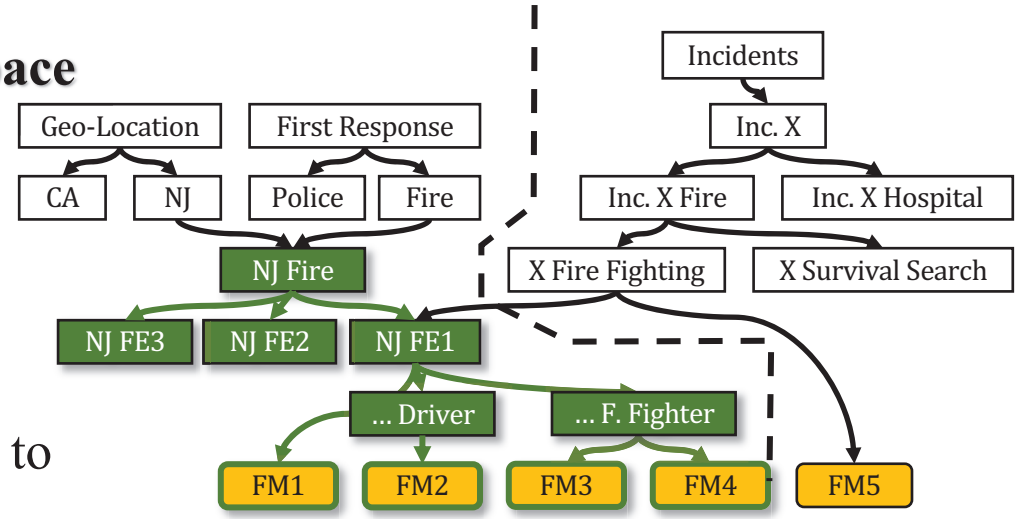
- Example namespace
  - Organizational structure: need information flow to members
    - Graph enables multiple dimensions (geo-location & functionality)
  - Incident place holder
- First responders instantiate roles
- Instantiate a disaster management template: preplanned namespaces
- Dispatch units to deal with functions in an incident
- Send messages to a role, e.g., "NJ Fire"

Need: Support a graph-based namespace in the network

Dynamic Nature of Namespace

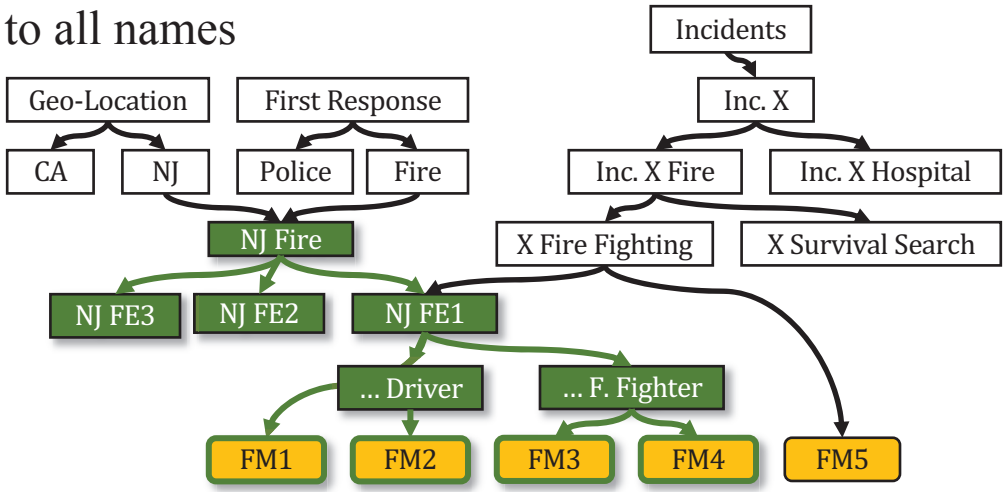
Dynamic installations of disaster namespaces

The namespace can evolve according to the situation



Supporting Graph-Pub/Sub in The Network

- Alternatives:
  - Perform BFS/DFS at each router
    - Benefit: fewer messages to be delivered
    - Issue: computation and storage cost at each router, infeasible, inefficient
  - Network only deals with flat names/ids (e.g., IP multicast, MF multicast & COPSS with flat names)
    - Subscribers subscribe to all names & publish to one
    - Subscribers subscribe to one & publish to all names



Solution: information layer to do the name expansion

Trust Management

- **Overview and Motivation**
  It is essential to introduce trust value in both the pub/sub forwarding framework (network layer) and the location-based forwarding framework (routing layer) to achieve completely trustful networking. Thus, the outcome of this task constitutes the basis of all the other tasks.

- **Objective and Intellectual Merit**
  An objective of this task is to add **"ephemeral trust"** to untrusted parties such as victims and volunteers and establish a trust chain originating from a first-responder.

- **Research Challenges**
  Our research challenge is to ensure the ephemeral trust of messages/participants in disaster situations.
  A key idea is to employ two types of information sources:
  - Verify consistency of multiple-messages in social media communications from volunteers/victims, and
  - A deterrence capability emphasized by their biometric information.

Solution

- **(i1) Assignment of ephemeral trust:**
  Social media messages reported from volunteers/victims may have some degree of uncertainty. Some volunteers/victims may even send false messages or non-urgent messages deliberately. We develop a scheme to evaluate the veracity of messages in a disaster situation.

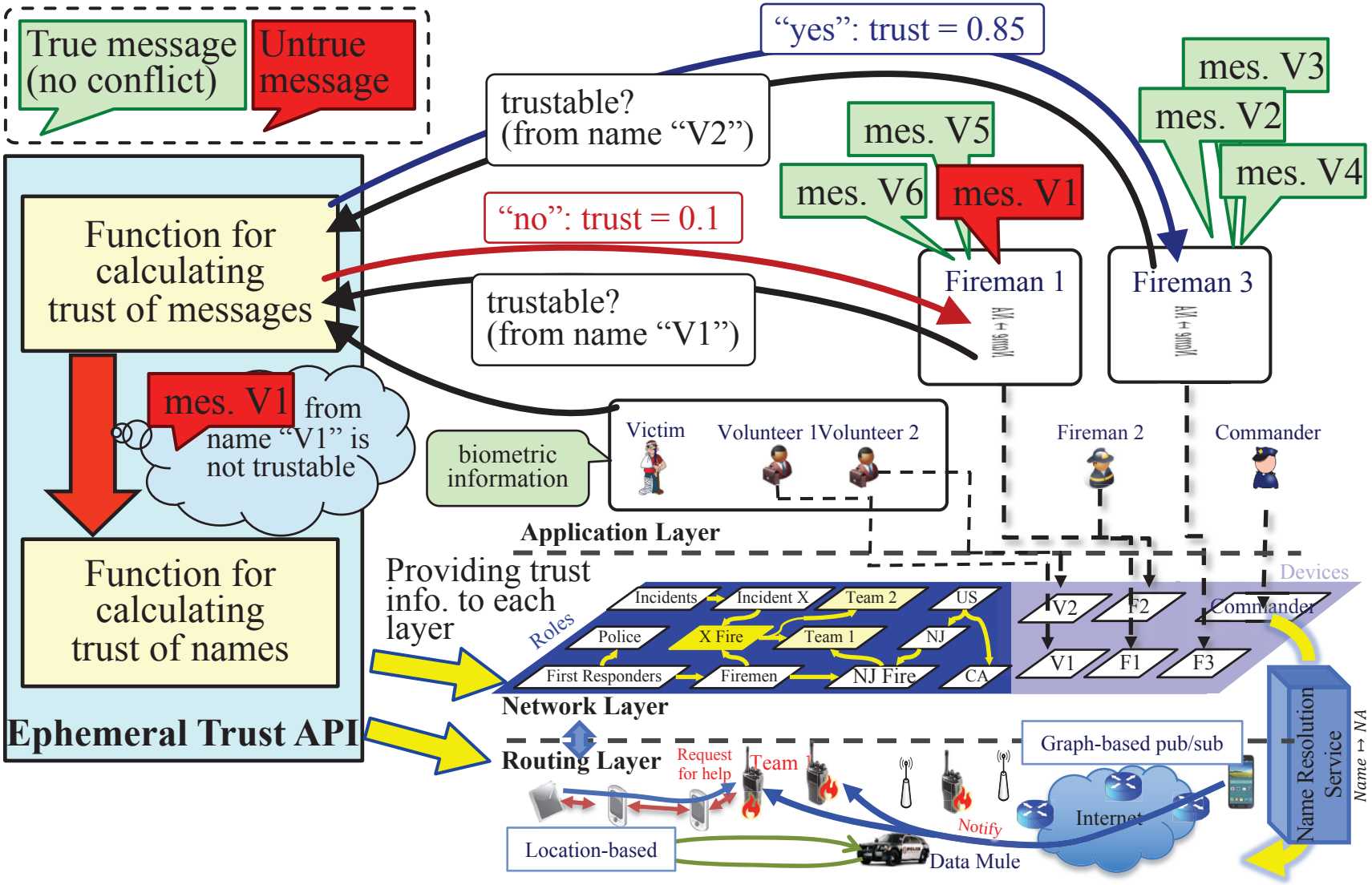
- **(i2) Trust value management of ephemeral trust:**
  A possible way to formulate the veracity of messages is a game-theoretic approach. The trust value of a participant/message is formalized by a utility function, and the sender of true messages is assigned a trust value as a reward.

- **(i3) Deterrence provided by biometrics:**
  We introduce a concept of **"deterrence-based trust"** in our ephemeral trust model. By using biometric signature, even when participants are completely alone and isolated, they can leave evidence (biometric signature) linked to their actions.

- **(i4) Development of trust API:**
  We develop a **"trust API"** which can be used not only for trustful/effective name resolution in pub/sub forwarding framework but also for trustful/effective route determiniation in location-based forwarding framework.

Trust Management - continued

- API

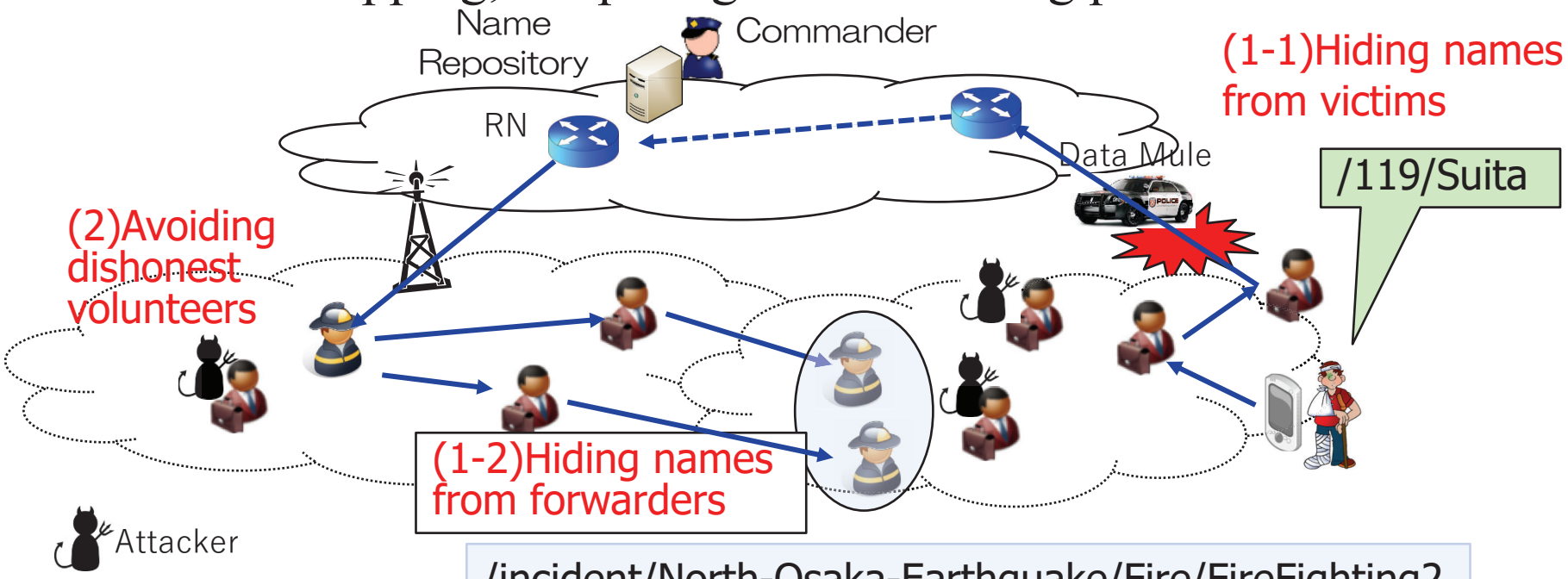


Secure Location Based Forwarding

- Overview and Motivation
  - Secure and resilient forwarding between honest first responders and between honest first responders and a volunteer/victim in multi-hop environments, assuming that
    - Security
      - No certificate authority is reachable from first responders
      - Some volunteers as forwarders may be dishonest
    - Reachability
      - Network is fragmented, thus the Internet backbone may not be available
- Objective and Intellectual Merit
  - Securely deliver urgent messages from a victim to one of the nearest first-responders when central emergency offices are not reachable

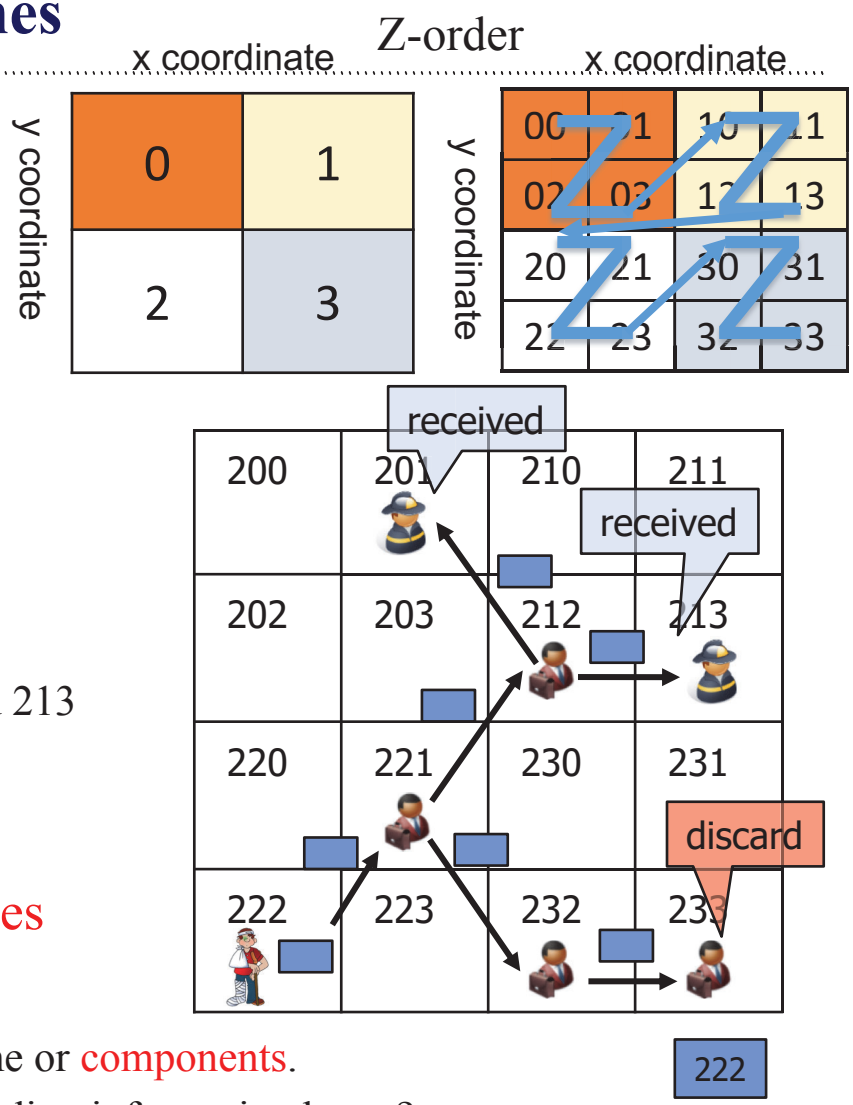
Secure Location Based Forwarding - continued

- Research Challenges
  - (1)Privacy: Hiding first responders' names from victims and volunteers
  - (2)Security: Resiliency against dishonest volunteers: eavesdropping, tampering and discarding packets



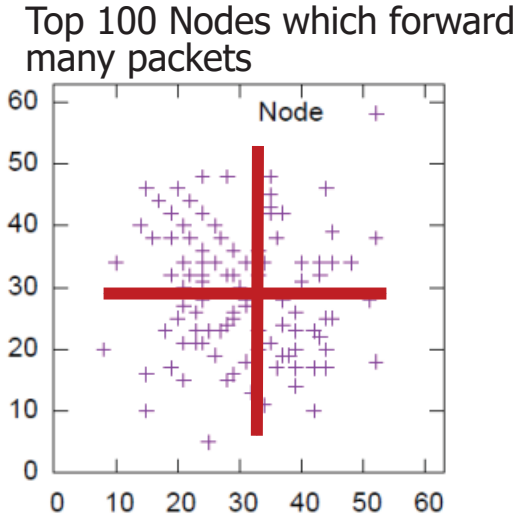
(1) Hiding First-Responders' Names

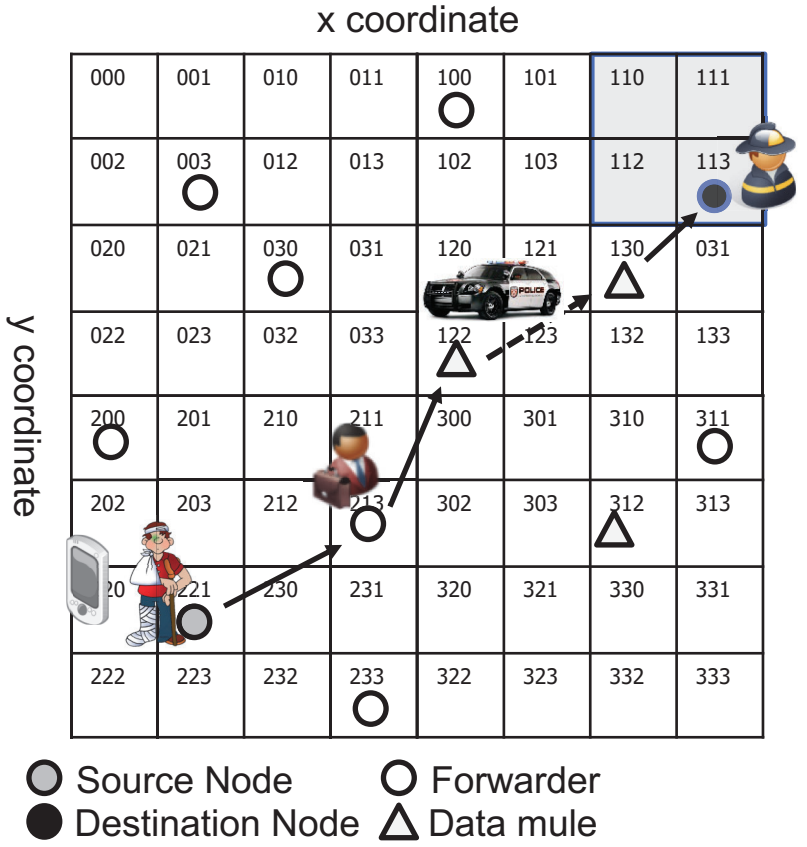
- Against victims: Location based forwarding
  - Hide first-responders' names by using location names according to Z-order
  - Walking scenario
    - A victim sends a message to its location name 222
    - It is delivered to first-responders who subscribe to nearby locations like 201 and 213
- Against forwarders: Obfuscating first-responder names
  - Names are obfuscated as keyed hashes
  - Issues
    - How to hash a name? Either a whole name or components.
    - How to aggregate hashed names in forwarding information bases?
    - How to share keys among first-responders and a commander?



Location based Forwarding

- Distance vector routing protocol
  - Routing information: Aggregation based on Plaxton
- Preliminary evaluation
  - Packets are forwarded between randomly chosen nodes on a 64 x 64 mesh network
  - Some nodes forward more packets than the others





	0	1	2	3
1st digit	030	122 via 213	-	312
2nd digit	200	213	-	233

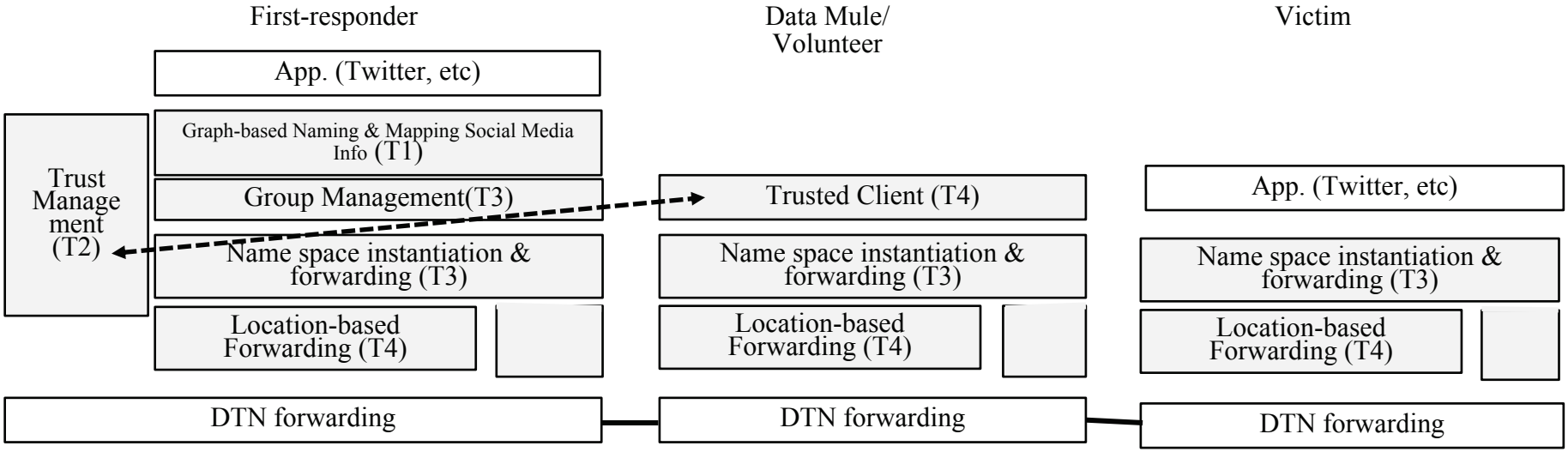
Forwarding table at 211

Resiliency against dishonest volunteers' behavior

- Approaches
  - Selecting volunteers with high trust values to avoid dishonest forwarders on forwarding paths
  - Reliably recording/sharing information (trust values and authentication information) of volunteers to prevent dishonest volunteers among volunteers by leveraging blockchain
    - Assuming that forwarding paths among them are not either secure or resilient to failures
- Research issues
  - Energy efficient byzantine fault tolerant algorithm rather than proof of work
    - Multi-hop environments where broadcasting is not available

Integration and Experiment

- Validation based on simulation and prototyping
  - Integrate graph-based namespace, pub/sub, forwarding and security functionality for design and performance evaluation
  - Integrate graph-based forwarding (T1 and T3) and location-based forwarding (T4) on a prototype
  - Integrate trust management (T2) with above (T1, T3 & T4)



Conclusion

- Secure and resilient disaster communication in the Social Media era
  - Protocol design and evaluation based on analysis, simulation and prototype experiments over testbeds like Cutei and NDN testbed
- Dissemination
  - Open source software
    - Software on open source ICN software: Cefore, NFD(NDN Forwarding Daemon)
      - Example: An emergency message delivery service like 119/911 calls
  - Publications and Standardization
    - ACM ICN, IEEE Transactions, et. al.
    - IRTF ICNRG

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

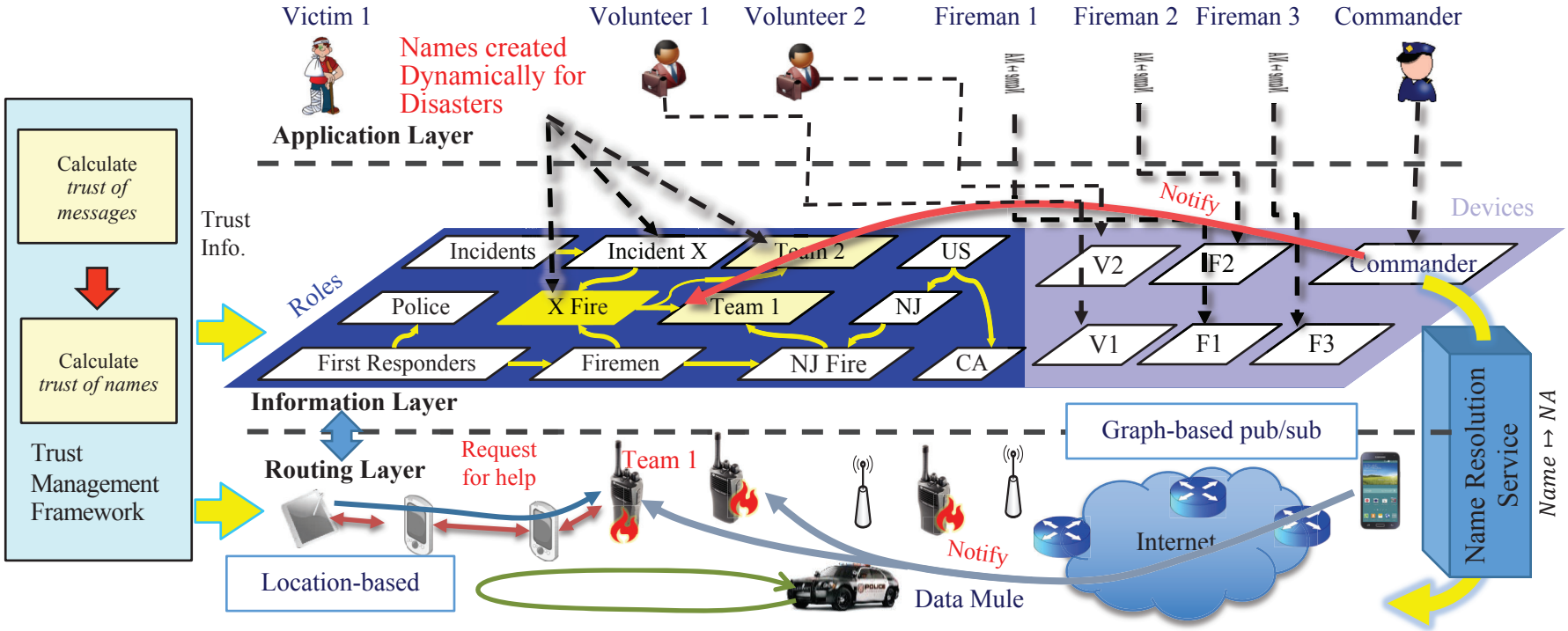
- **<u>Project Objective:</u> A network architecture for information and communication resilience in disaster management, that is also secure, integrates volunteers and social media seamlessly in disaster response**

Challenges

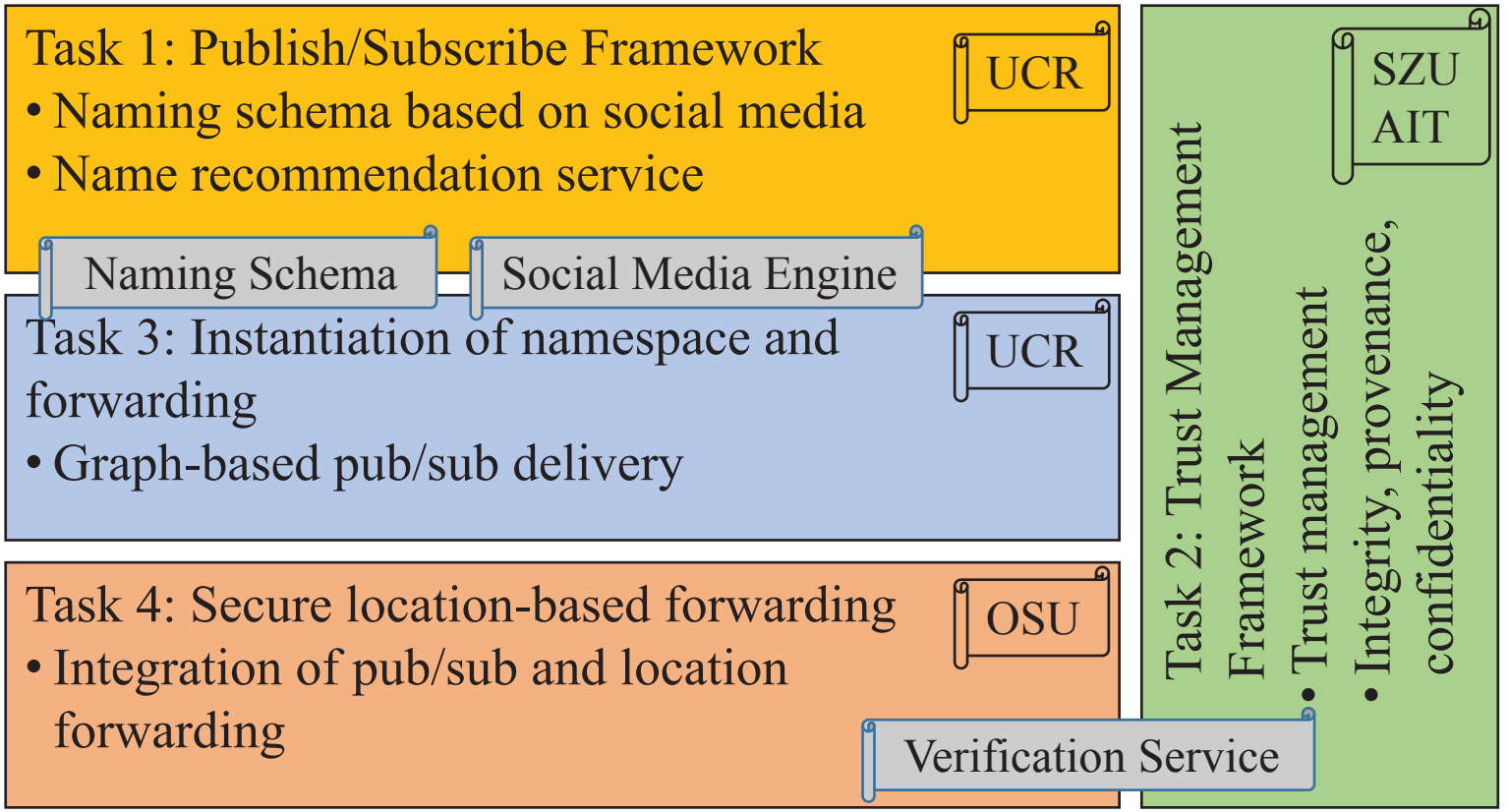
- Challenge 1: Designing a naming and forwarding framework in dynamic disaster environments, focusing on communications between and among honest first responders, while including trusted volunteers and victims
  - Task 1: Design, Creation and Instantiation of namespace
  - Task 3: Publish/Subscribe Framework; 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 5: Integration, Experimentation and Evaluation

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

5

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

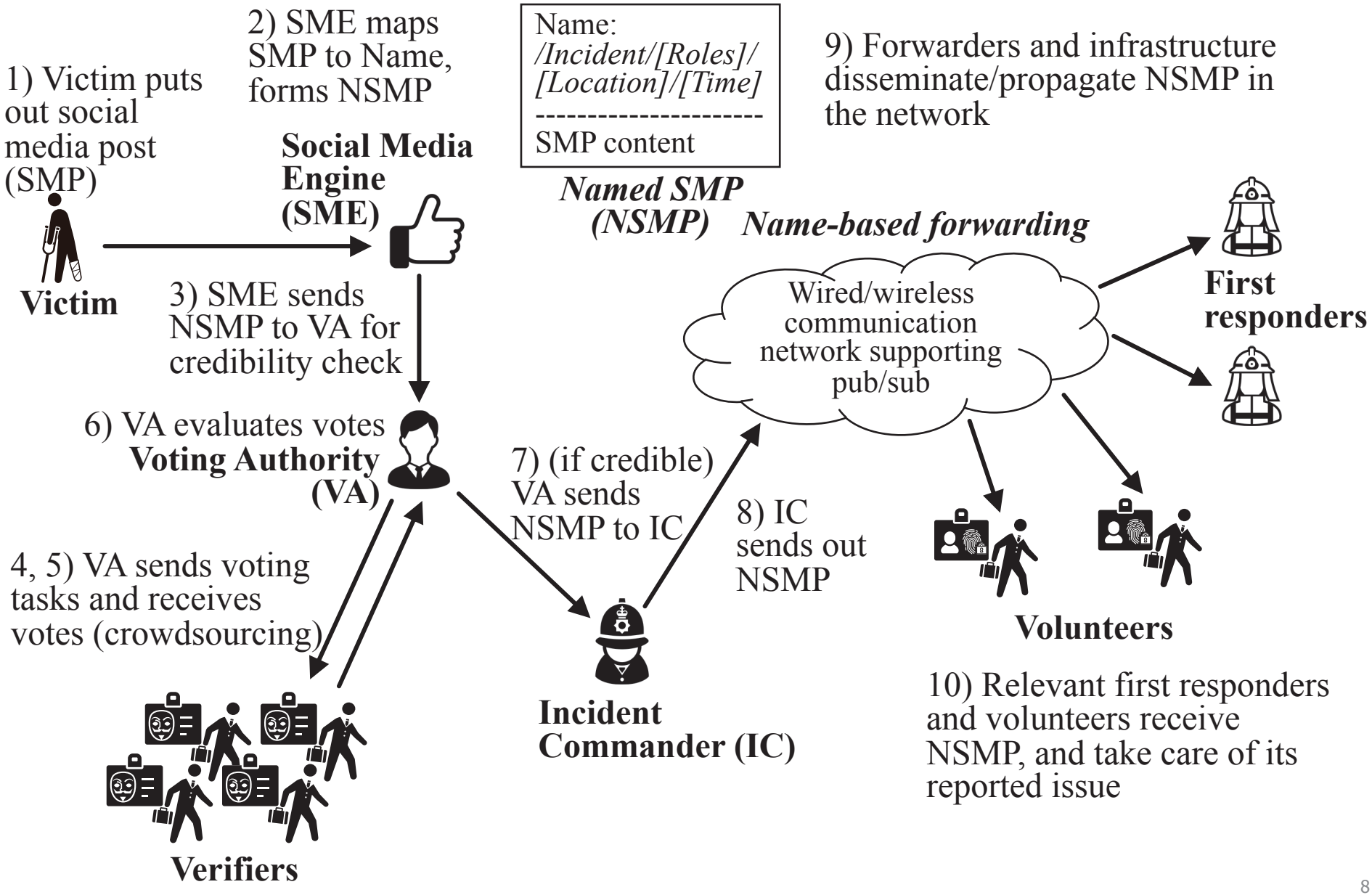
Social media engine (SME)

- 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



Architectural Components

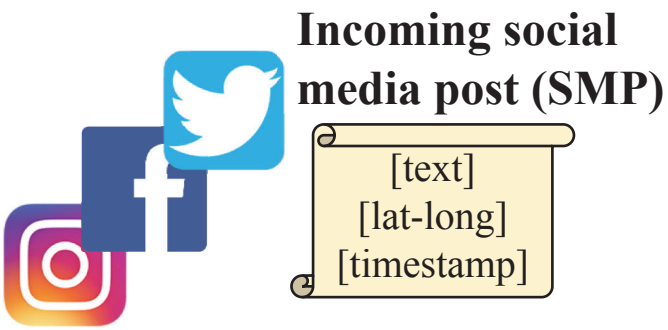
- Naming Schema
  - Unifies the interactions between all different actors (civilians, first responders, etc.) and guides the subscription and publication paths
  - Namespace represents entities related to and critical in incident management, and captures complex relations among them.
- Social Media Engine (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

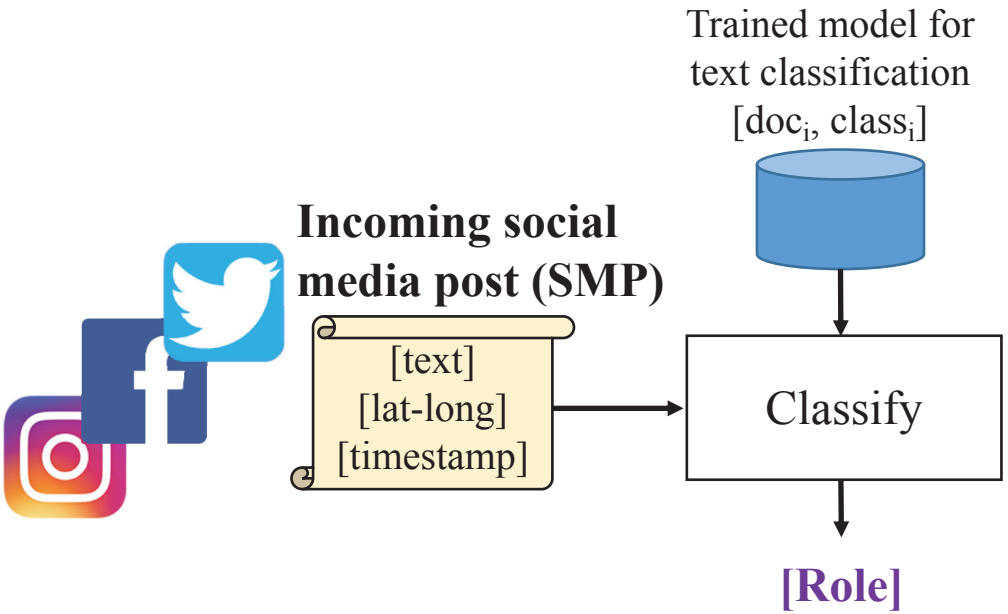
Social Media Engine

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



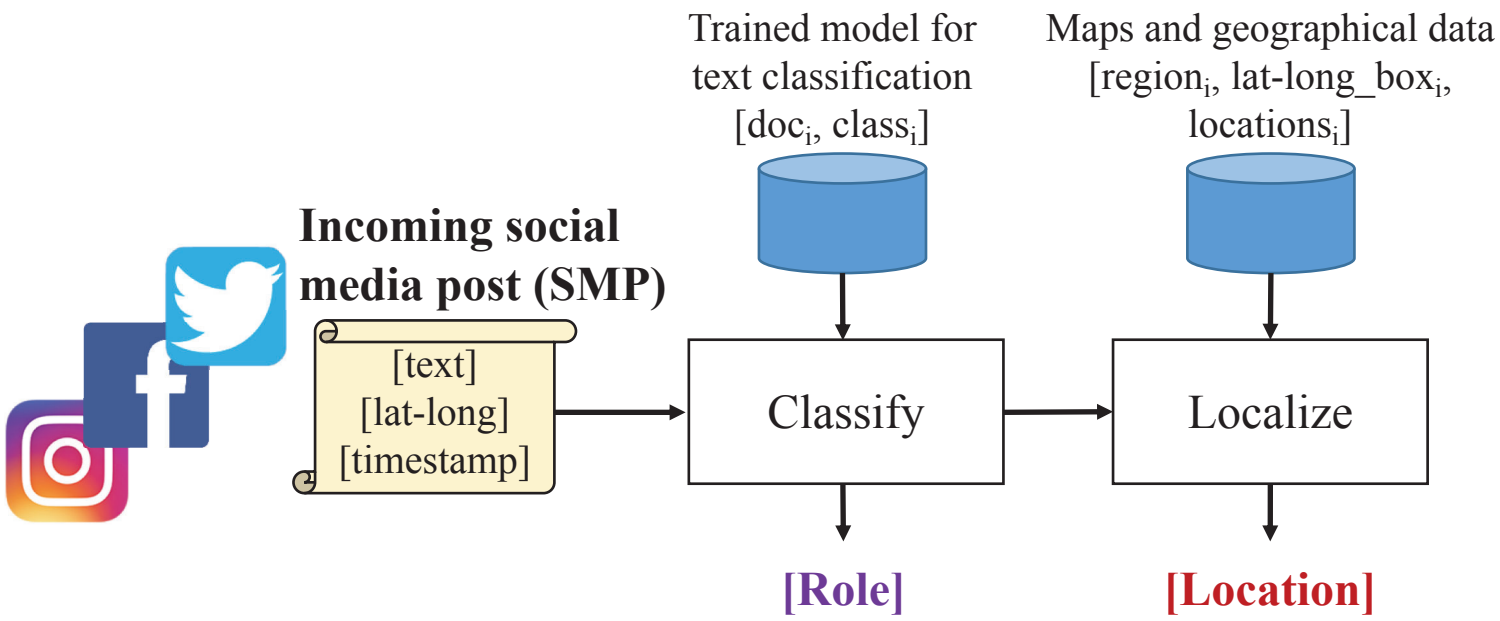
Social Media Engine

- 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



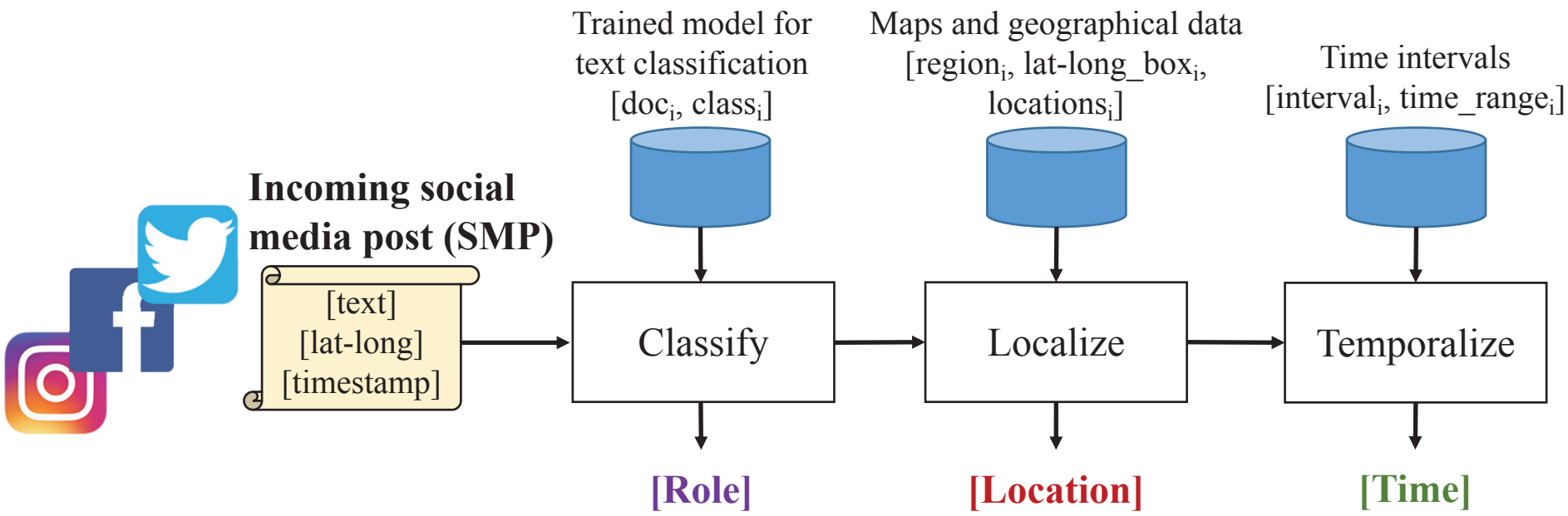
Social Media Engine

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



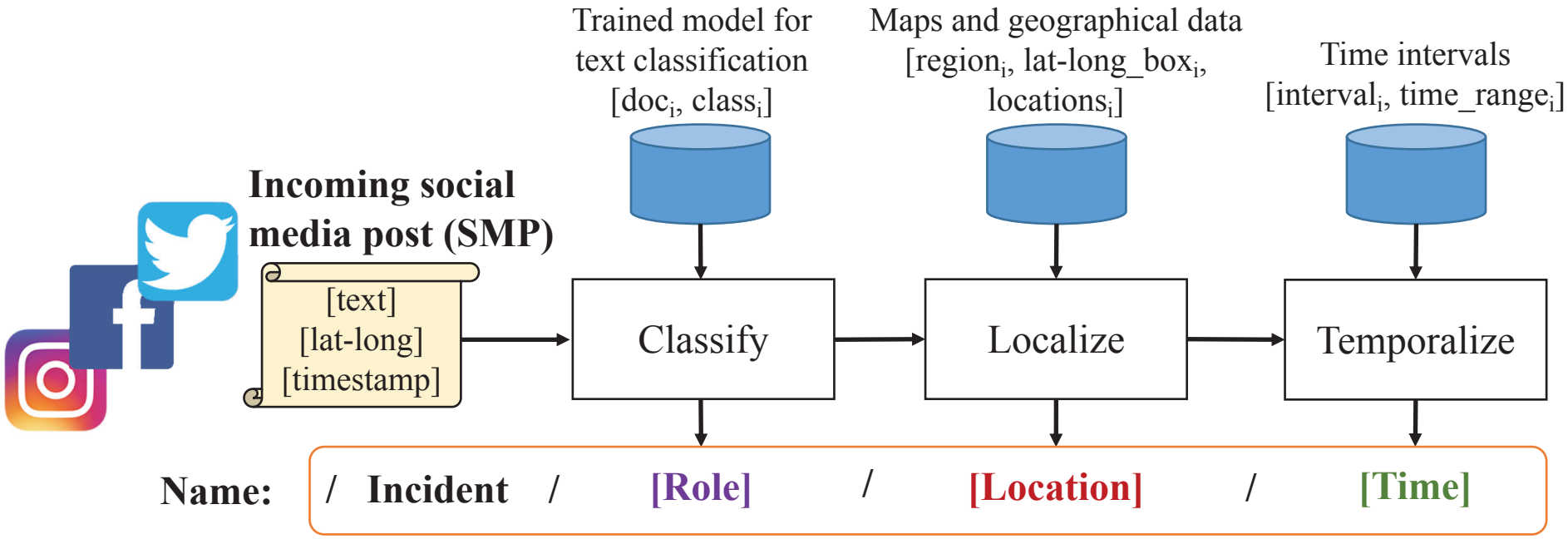
Social Media Engine

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



Social Media Engine

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



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

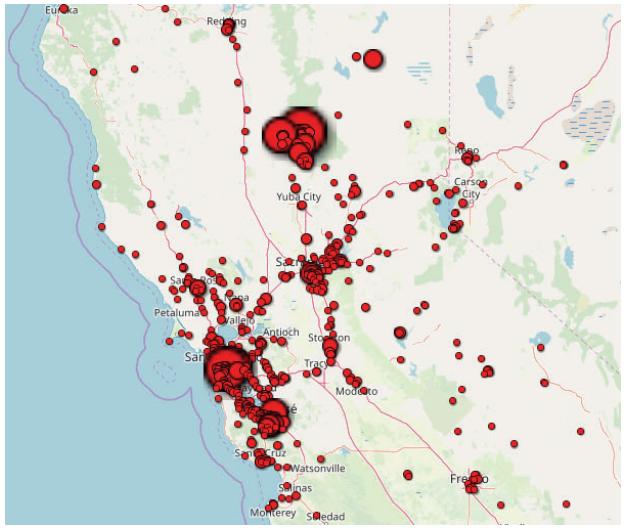
Case Study: California Wildfires in 2018

- 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

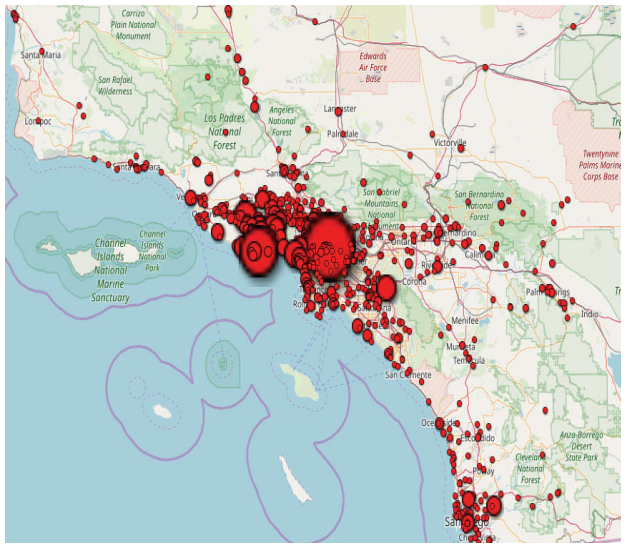
Spatial and Temporal Analysis

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

Camp

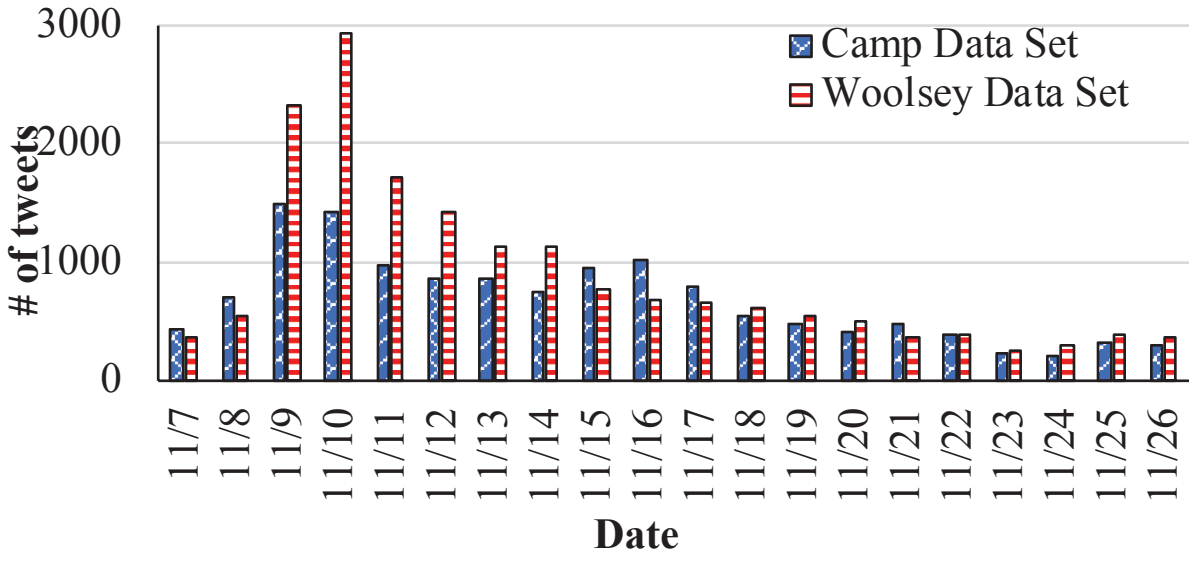


Woolsey

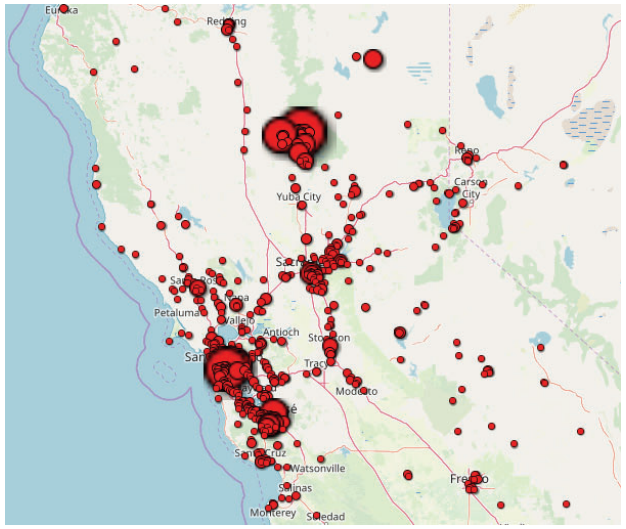


Spatial and Temporal Analysis

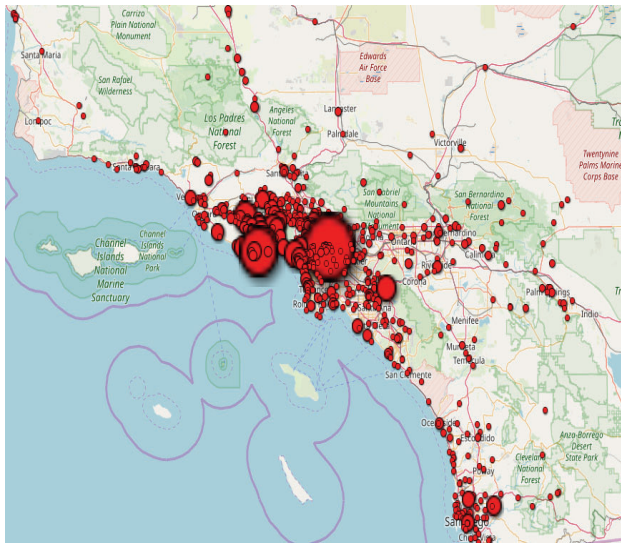
- Extract and analyze disaster-related tweets
- Spatial correlation: Higher density of disaster-related tweets near the more affected areas
- Temporal correlation: Higher density of disaster-related tweets during more intense days of the wildfires



Camp



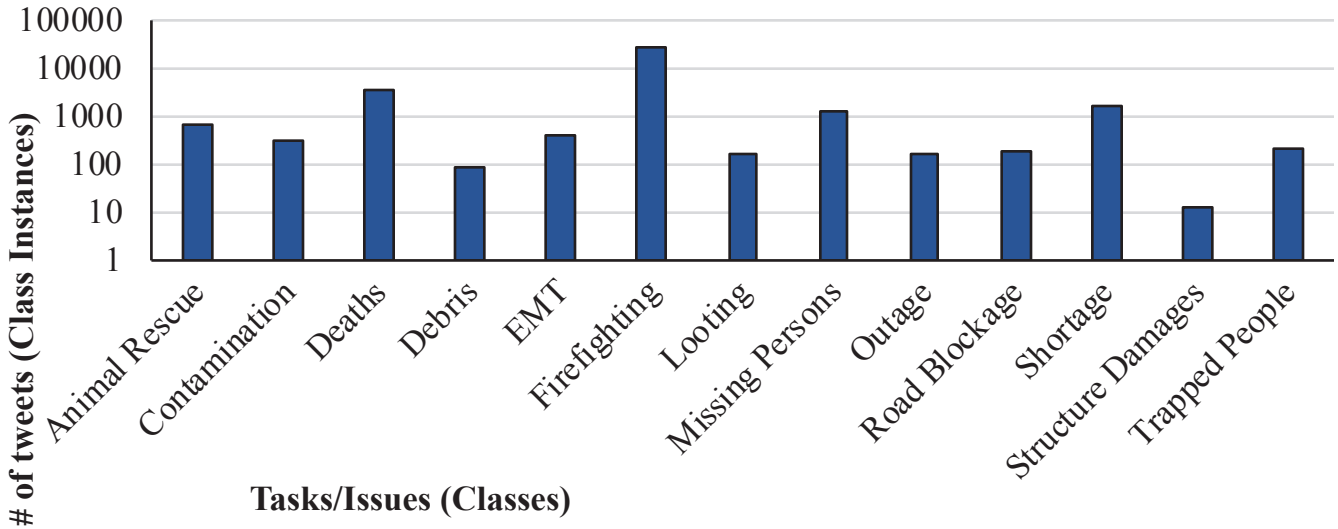
Woolsey



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

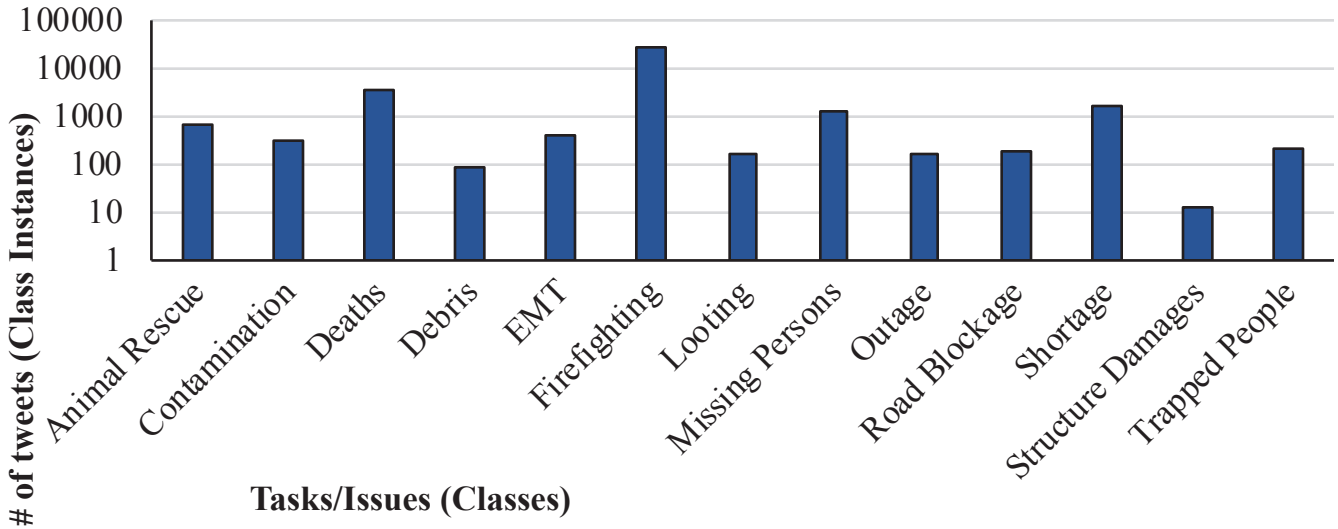
Tweets per label distribution ('Firefighting' is majority)



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

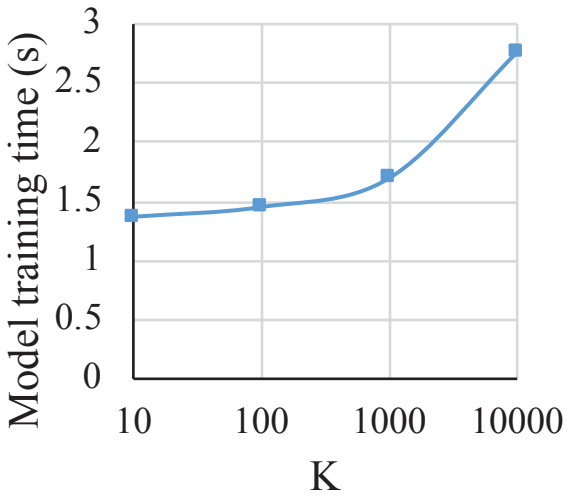
Tweets per label distribution ('Firefighting' is majority)

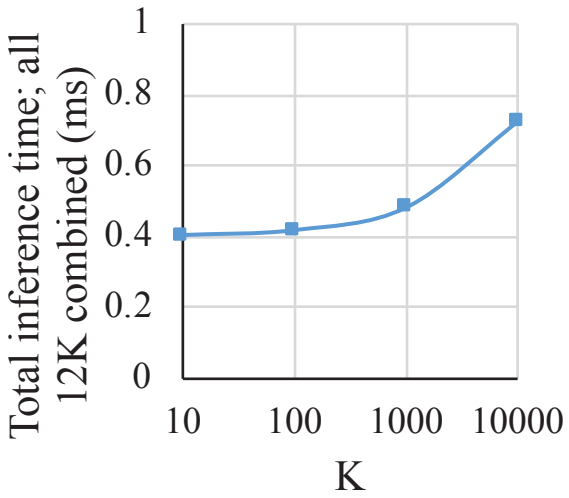


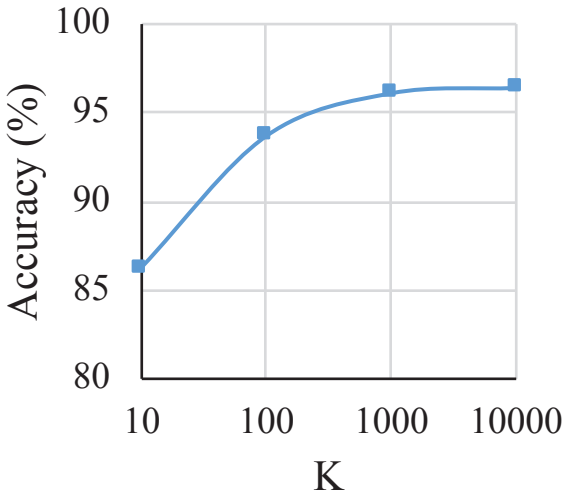
tf–idf: term frequency–inverse document frequency

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')

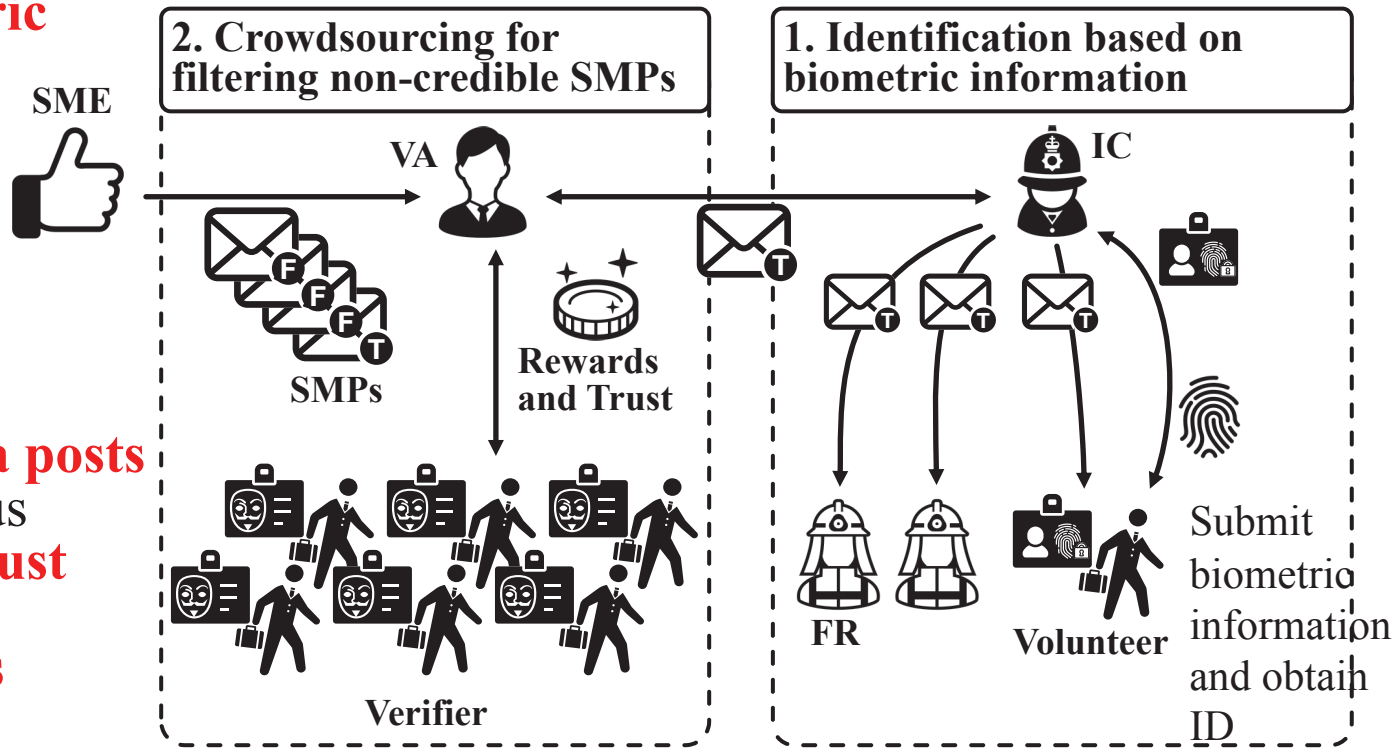
Metric	Precision	Recall	F1-score
Micro average	0.96	0.96	0.96
Macro average	0.88	0.81	0.84

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**



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

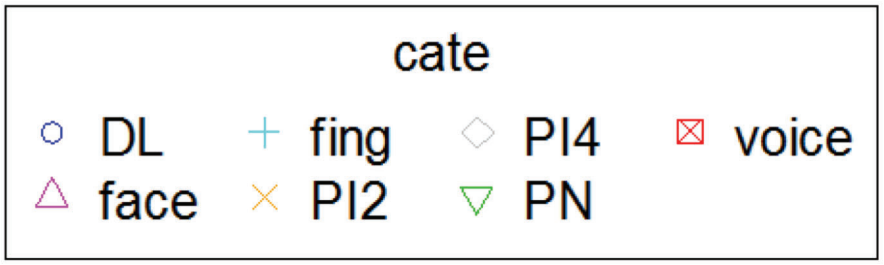
Survey Results: Which personal information pieces contribute to deterrence?

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

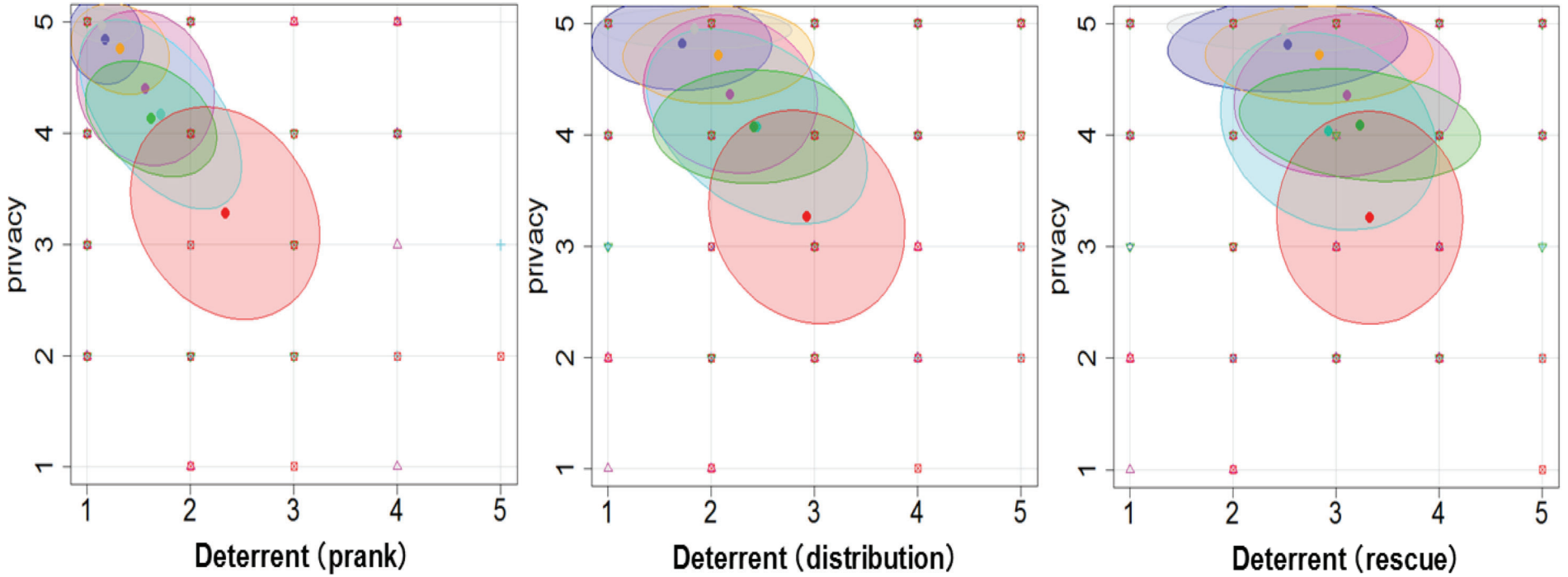
item / info	Privacy score	Deterrence score			Effect score		
		Prank	Distribution	Rescue	Prank	Distribution	Rescue
Non	-	1.76 (1.14)	1.98 (1.25)	1.78 (1.21)	-	-	-
Face	4.17 (1.14)	4.24 (1.02)	3.68 (1.17)	3.05 (1.40)	0.07 (1.42)	-0.49 (1.65)	-1.12 (1.93)
Fing	3.93 (1.15)	4.10 (1.11)	3.44 (1.34)	3.07 (1.31)	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



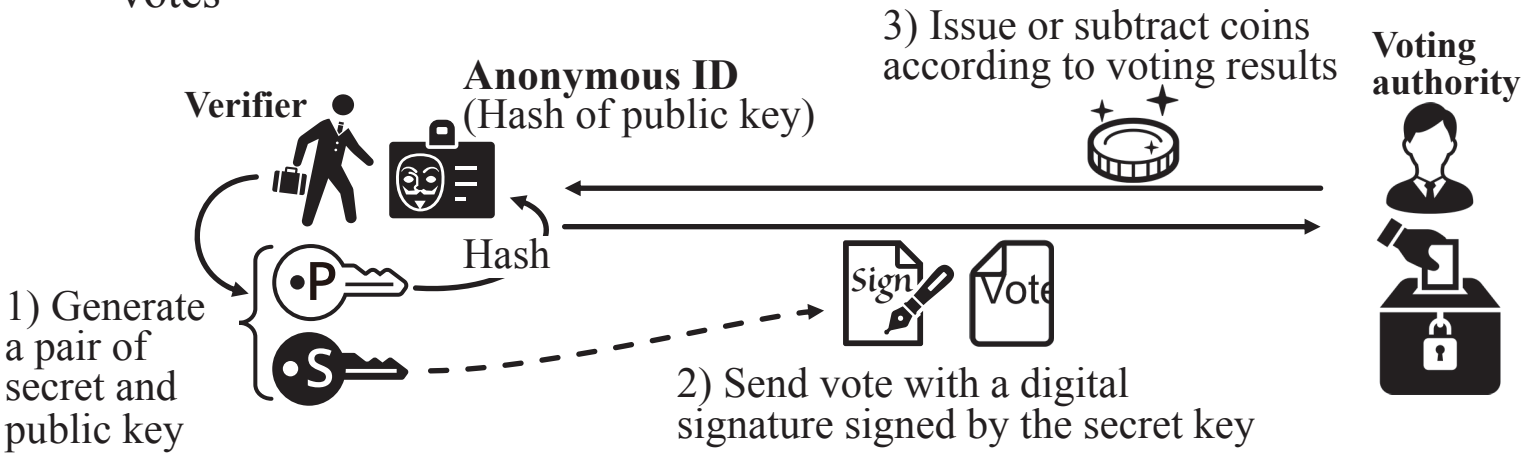
Observations

- 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"

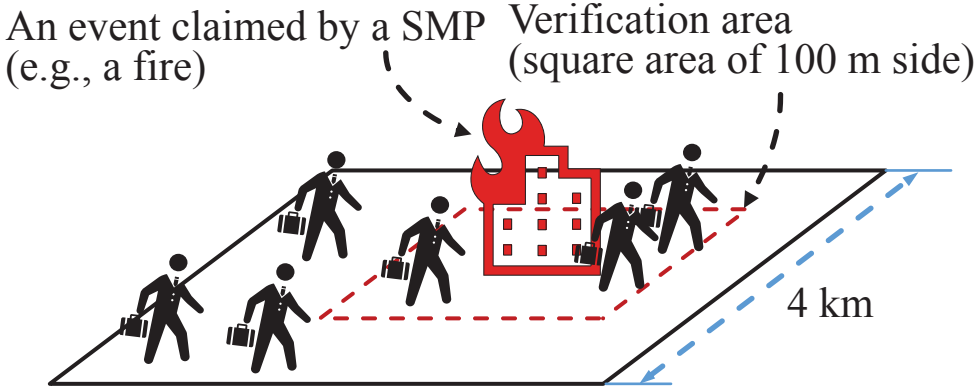
Crowdsourced Social Media Post Verification

- 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

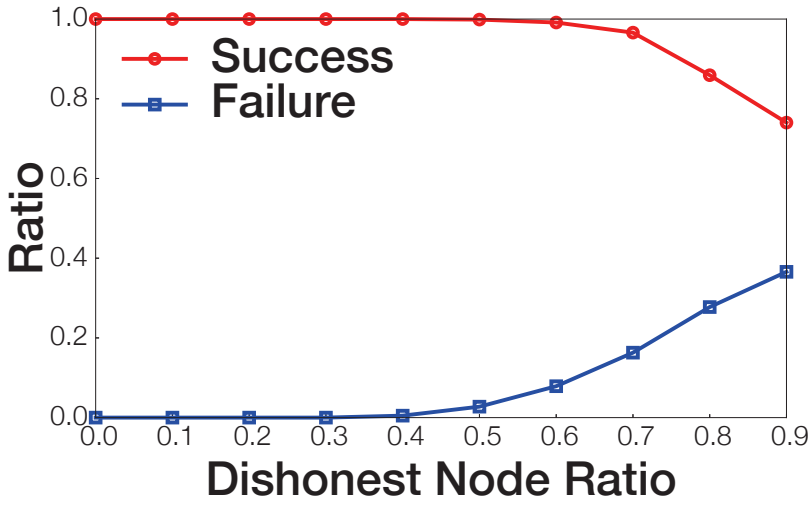


Evaluation of Crowdsourced Verification

- 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

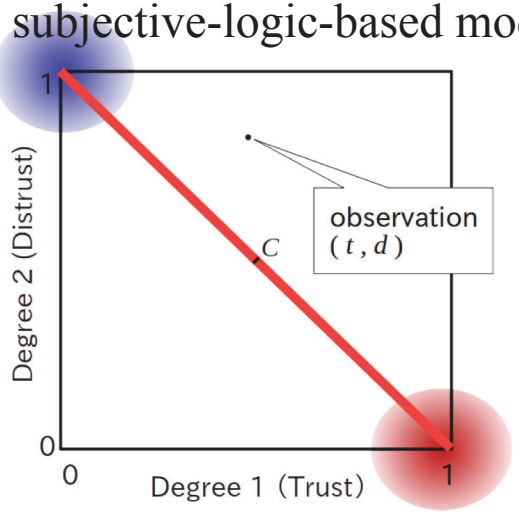


10^4 verifiers deployed randomly



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

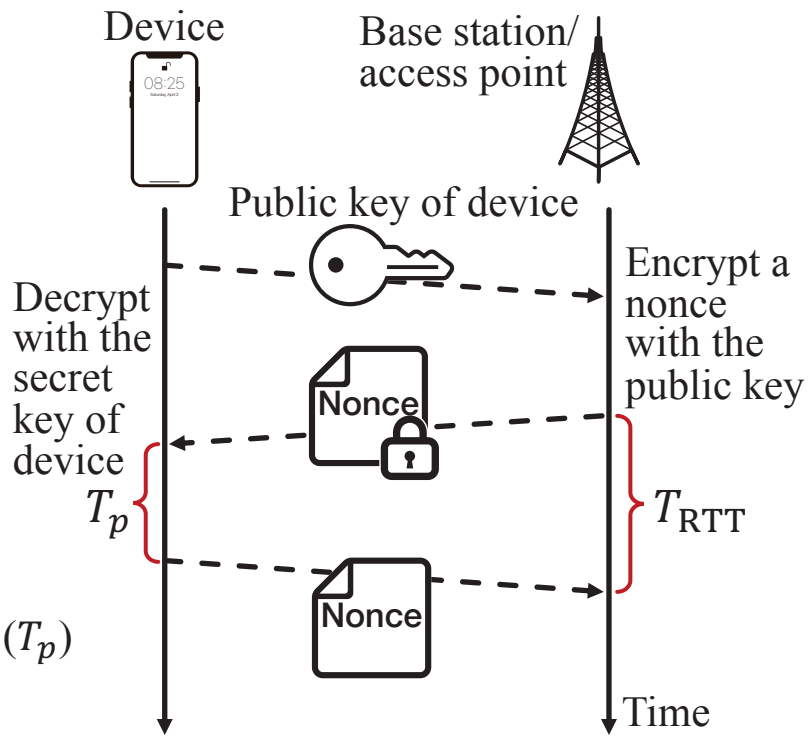
$$\frac{T_{\mathrm{RTT}}}{2s^{-1}} - d(l_d, l_b) \leq \varepsilon$$

Estimated distance from the measured RTT

Distance between the BS location (l_b) and the location claimed by the device (l_d)

Estimation error caused by the processing delay (T_p)

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



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)

Resilient Disaster Communications in the Social-Media Era

JUNO-2 Meeting

August 18/19, 2021

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

UC RIVERSIDE UNIVERSITY OF CALIFORNIA

大阪大学 OSAKA UNIVERSITY

国立大学法人 静岡大学 National University Corporation Shizuoka University

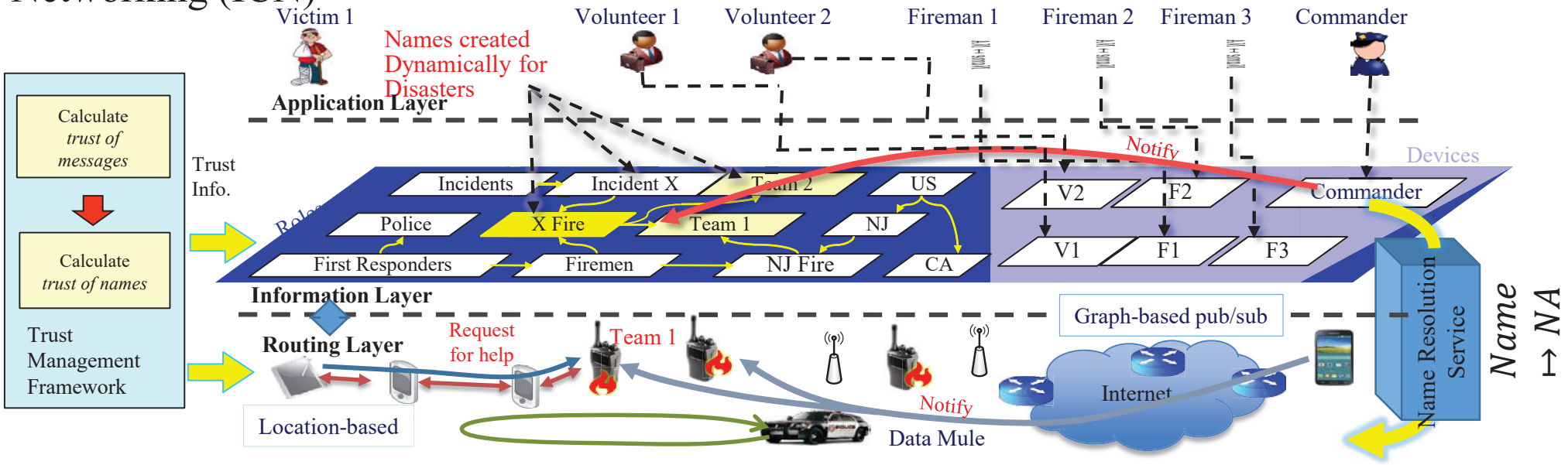
愛知工業大学

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**

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)



Objectives and achievements

- Task 1: Naming and Publish/Subscribe framework
  - We developed a dynamic graph-based namespace (initially defined by templates) to more appropriately reflect the organization of personnel involved in emergency response
  - Exploit recipient hierarchies for information dissemination, including social media posts
- Task 2: Trust Management
  - Developed the 2-level crowd sourcing system to which verifies credibility of SMPs by leveraging remote/on-site volunteers
- Task 3: Mapping free-form text (social media posts) to namespace
  - Social Media Engine performs online classification and inference using Natural Language Processing (NLP): exploit specialized knowledge of individual public-safety departments with active learning for accurate delivery of social media posts
- Task 4: Secure location-based forwarding
  - Developed the failure-resilient pub/sub protocol and the emergency service based on the protocol
- Task 5: Integration, Experimentation and Evaluation
  - Performed experiments for SMPs posted at large scale disasters and verified the trust management system is able to verify credibility of SMPs

Dissemination of Project Achievements: Summary

- We designed the architecture and published a joint paper
  - 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/implemented the Pub/Sub protocol and the emergency service, and implemented and evaluated the rescue classifier, a part of SME
  - Shizuoka University designed a biometric authentication of volunteers; did a questionnaire survey
  - Aichi Institute of Technology designed a trust model for volunteers
  - 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," in Proceedings of ICT-DM 2019, Dec. 2019. **(Best paper award)**
- Publications (Japan)
  - Kazuhiko Ohkubo, Tetsuhisa Oda, Yuki Koizumi, Tetsushi Ohki, Masakatsu Nishigaki, Toru Hasegawa, Yoshinobu Kawabe, "Trust representation under confusion and ignorance," in Proceedings of International Workshop on Informatics (IWIN) 2018, pp.195-202, Sept. 2018.
  - Sugimoto Genki , Fujita Masahiro , Mano Yuto , Ohki Tetsushi , Nishigaki Masakatsu, "Micro Disposable Biometric Authentication - An Application Using Fingernail 15 Minute Textures for Nonsensitive Services -," in Proceedings of Proceedings of the 2018 3rd International Conference on Biomedical Imaging, Signal Processing, pp. 68-73, Aug. 2018.
  - 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," in Globecom 2019 Workshop, Dec. 2019.
  - Yoshinobu Kawabe, Yuki Koizumi, Tetsushi Ohki, Masakatsu Nishigaki and Toru Hasegawa, "Toward Mathematical Analysis for Quantity of Trusts," in Proceedings of ITC-CSCC 2020, pp. 111-115, July 2020.
  - Yuki Koizumi, Yoji Yamamoto, Toru Hasegawa, "NDN-based Publish/Subscribe Communication in Disasters," in Proceedings of International Conference on Emerging Technologies for Communications, pp.1-4, Dec. 2020.
  - Takumi Kitagawa, Tetsushi Ohki, Yuki Koizumi, Yoshinobu Kawabe, Toru Hasegawa and Masakatsu Nishigaki "Deterrence-Based Trust: A Study on Improving the Credibility of Social Media Messages in Disaster Using Registered Volunteers," in Proceedings of International Conference on Network-Based Information Systems, pp. 188-201, Sept. 2021.

Dissemination of Project Achievements: Summary (cont'd)

- Publications (US)
  - Mohammad Jahanian, K. K. Ramakrishnan, "Name Space Analysis: Verification of Named Data Network Data Planes", The 6th ACM conference on Information Centric Networking (ICN), September 2019. (Best Student Paper Award)
  - Mohammad Jahanian, Jiachen Chen, K. K. Ramakrishnan, "Graph-based Namespaces and Load Sharing for Efficient Information Dissemination in Disasters", The 27th IEEE International Conference on Network Protocols (ICNP), October 2019.
  - Mohammad Jahanian, Jiachen Chen, K. K. Ramakrishnan, "Formal Verification of Interoperability Between Future Network Architectures Using Alloy", The 7th International Conference on Rigorous State-Based Methods (ABZ), May 2020.
  - Mohammad Jahanian, Jiachen Chen, K. K. Ramakrishnan, "Managing the Evolution to Future Internet Architectures and Seamless Interoperation", The 29th International Conference on Computer Communications and Networks (ICCCN), August 2020.
  - Mohammad Jahanian, K. K. Ramakrishnan, "CoNICE: Consensus in Intermittently-Connected Environments by Exploiting Naming with Application to Emergency Response", The 28th IEEE International Conference on Network Protocols (ICNP), October 2020.
  - Viyom Mittal, Mohammad Jahanian, K. K. Ramakrishnan, "Online Delivery of Social Media Posts to Appropriate First Responders for Disaster Response", The 3rd International Workshop on Emergency Response Technologies and Services (EmeRTeS @ ICDCN'21), January 2021.
  - Mohammad Jahanian, K. K. Ramakrishnan, "Name Space Analysis: Verification of Named Data Network Data Planes", The IEEE/ACM Transactions on Networking, Volume 29, Issue 2, April 2021.
  - Mohammad Jahanian, Jiachen Chen, K. K. Ramakrishnan, "Graph-based Namespaces and Load Sharing for Efficient Information Dissemination" The IEEE/ACM Transactions on Networking, 2021 (Early Access).
  - Viyom Mittal, Mohammad Jahanian, K. K. Ramakrishnan, "FLARE: Federated Active Learning Assisted by Naming for Responding to Emergencies", The 8th ACM conference on Information Centric Networking (ICN), September 2021. (Accepted)

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

On-site volunteer
- Perform tasks to verify credibility social media posts according to requests by the volunteer authorities
- Support tasks of first responders

Remote volunteer
- Perform tasks to verify social media posts according to requests by the volunteer authorities

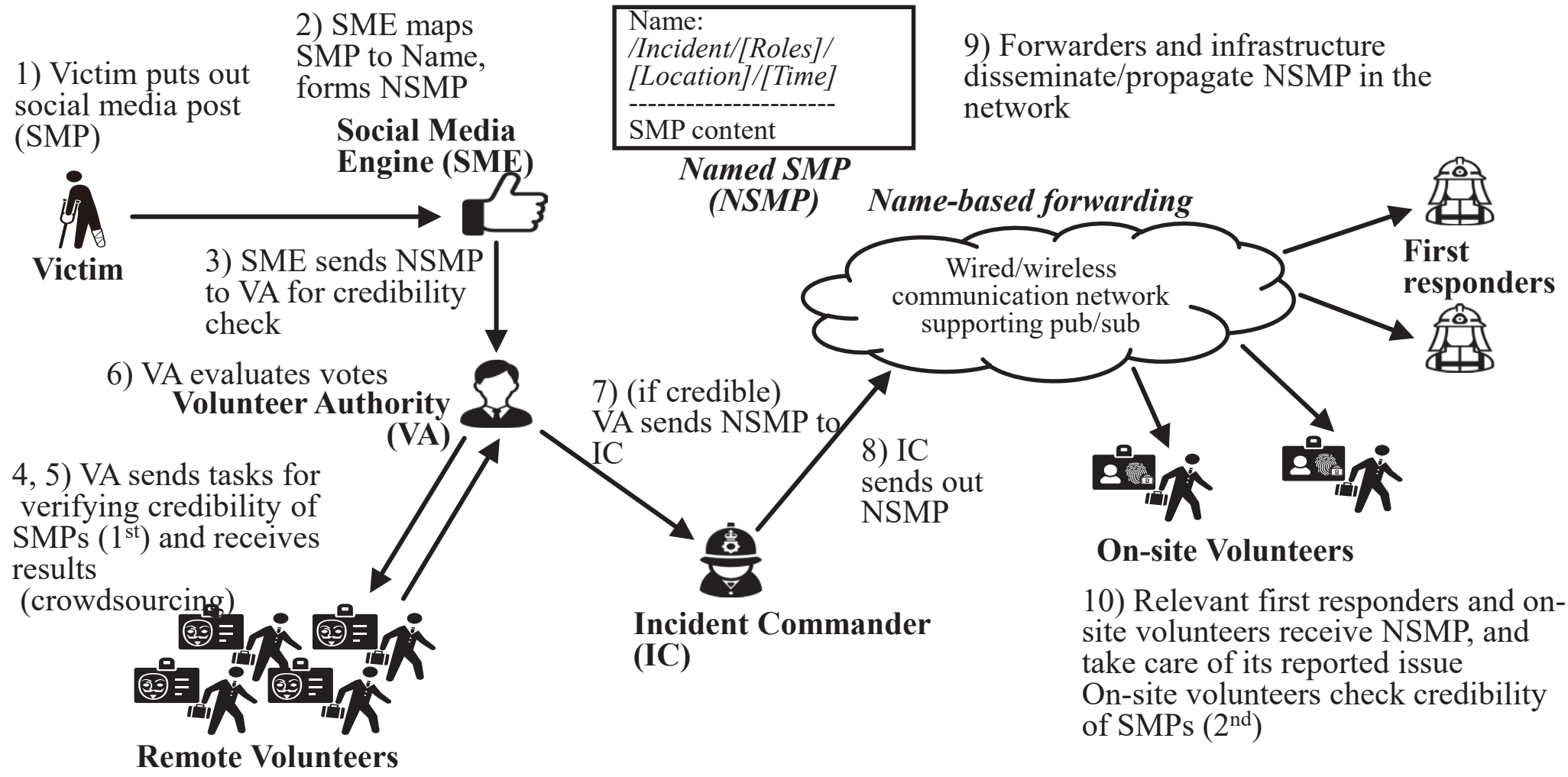
Incident commander (IC)
- Send commands to FRs according to event reports from the VA

Volunteer authority (VA)
- Ask remote/on-site volunteers 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



Architectural Components

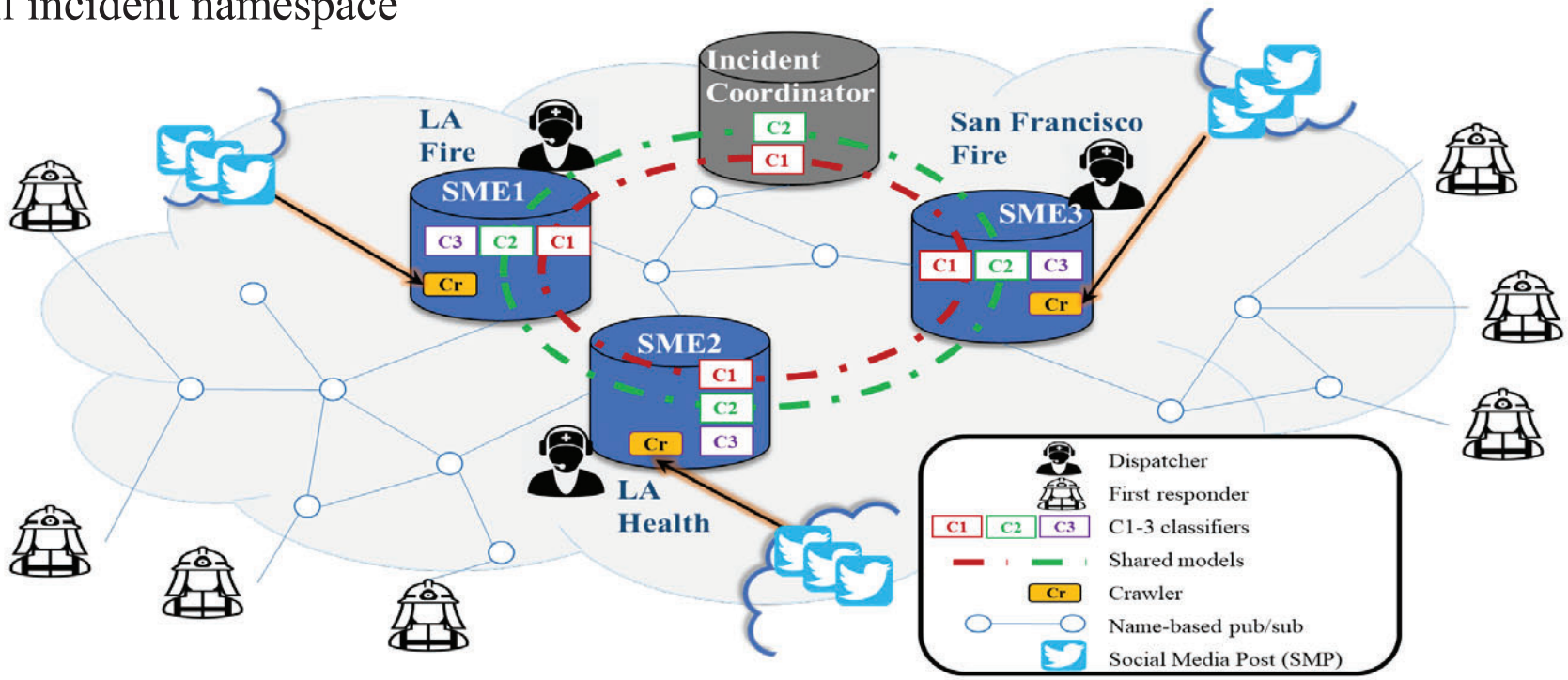
- Naming Schema
  - Unifies the interactions between all different actors (civilians, first responders, etc.) and guides the subscription and publication paths
  - Namespace represents entities related to and critical in incident management, and captures complex relations among them.
- Social Media Engine (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
  - 2-layer crowd sourcing verifies credibility of SMPs extracted by SME, wherein remote and on-site volunteers work for checking the credibility
  - Trust management system identifies trustworthy volunteers/verifiers based on biometric signature and a trust model

Federated Active Learning Assisted by Naming for Responding to Emergencies

- First responders and other users (victims seeking help) would prefer using free-form text for communication during disasters
  - Social media is being increasingly used for communicating urgent information
- Our name-based pub/sub dissemination can be very helpful to deliver relevant information to the right first responder in a timely manner
- Each publication: Guided by namespace to the most appropriate first responder(s)
  - Challenge: How to assign names to free-form text in **real-time**? (e.g., senders without access to namespace)
- FLARE: Framework to provide efficient, timely dissemination of relevant content to first responder teams assigned to different incident response roles using the specialized knowledge of different first responder departments in a real-time online manner in disasters
  - Classification and learning procedures to map textual social media posts (SMPs) to the right name
  - Multi-level classification - provides increasing level of detail for mapping to namespace
  - Active Learning (reduce labeling effort) and Federated Learning (cooperation of specializations)
    - Better overall delivery accuracy respecting flexibility in managing AL and FL for each classifier

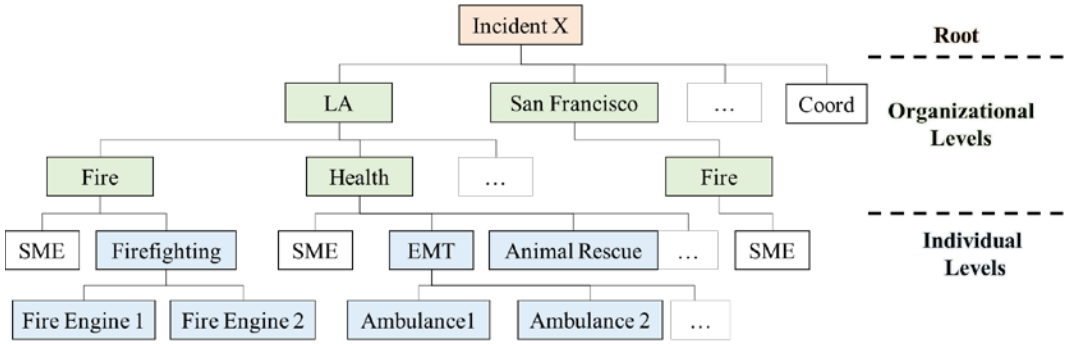
Overview of FLARE

- **Goal:** Provide efficient, timely dissemination of relevant content to first responder teams assigned to different incident response roles using specialized knowledge of first responder (and assisting) departments
- SMEs associated with departments, equipped with multiple classifiers (C1-C3), dispatchers, and the full incident namespace



Mapping Content to Names - Integrated with Learning

- **Map SMP to Names**
  - Map text messages and social media posts
  - Deliver to the correct individuals
  - Information-centric pub/sub (and request/response)



- **Participants**
  - Victims: Request for help using free-form text (tweets, etc.)
  - Incident Commander(s), Dispatchers: manage namespace
  - First responders: Subscribe to the appropriate level of the namespace - respond to messages
- **Social Media Engine (SME)**: Maps free-form text (e.g., social media posts) to the right name components, assigns names, e.g., *"/IncidentX/LA/Fire/Firefighting/FireEngine1/LARegionA"*, and publishes to correct first responder
- SME performs online classification and inference using Natural Language Processing (NLP)
  - Using a trained model to facilitate accurate inference

Mapping Content to Names - Components of SMEs

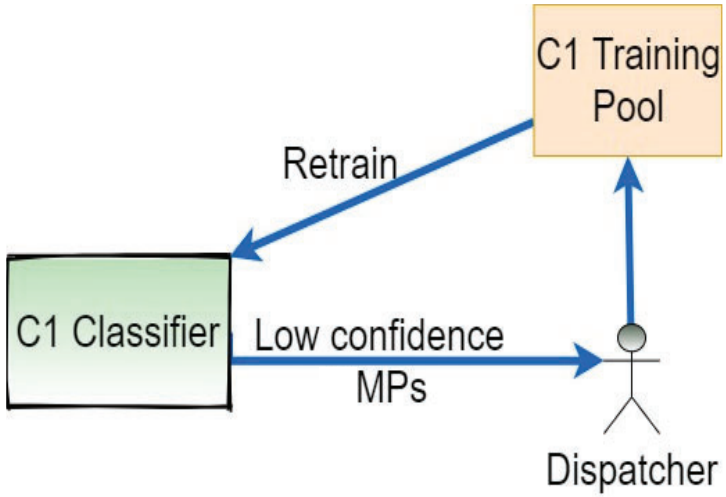
- Each Social Media Engine (SME) has its own (decentralized) **"Crawler"**:
  - Collects text-based content and/or crawls texts in real-time during or in the aftermath of disaster
  - Collects data with specific keywords based on the department's specialization
- Multiple classifiers
  - **C1 (incident relevance predictor):** Determines if a an SMP is **relevant** to the incident
  - **C2 (organization predictor):** Provides classification corresponding to a coarse organizational-level granularity in the incident namespace
  - **C3 (fine-grained role predictor):** Provides classification corresponding to the finer granularity of individual roles in the incident namespace
- Each SME has a **"Named SMP Generator":** takes the output of C3 as input and forms and publishes the SMP with full hierarchical name, e.g.,: *"/IncidentX/LA/Fire/Firefighting/FireEngine1"*

Social Media Engine Features

- **DNN Classifier with Universal Sentence Encoder (USE):**
  - Supports incremental learning allowing model to train on new dataset as it is made available
  - USE is pre-trained for sentence embedding over huge data corpus allowing it to capture rich semantic information
- **Active Learning:**
  - Reduces the manual labeling effort of the dispatcher by selecting only crucial messages required for training of the classifier
- **Federated Learning:**
  - Enables learning across various public-safety departments with specialized knowledge to handle notifications related to their roles, in a cooperative manner
- **Message Passing:** A technique to pass SMPs across different SMEs for their finer-grained classification by specialized knowledge of the dispatcher
  - Leverages organizational expertise in labeling more efficiently to eventually achieve better performance

Active Learning

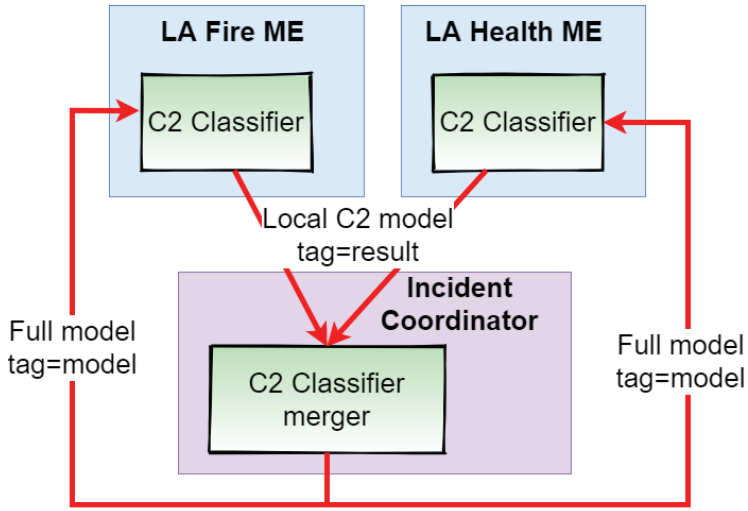
- Each classifier in the Media Engine is assisted by a dispatcher to label a small set of incoming SMPs
- **Specifications:**
  - Determine per-class prediction probability for each incoming SMP
  - Only label low confidence SMPs: Maximum prediction probability below a certain threshold
  - Configurable parameters:
    - Threshold: 0.8 for C1 and 0.7 for C2 (tunable)
    - Batch size: 50 (tunable)



- **Advantages:**
  - Reduced labeling effort for the dispatcher with no loss in accuracy
    - Labelling selective “informative” SMPs rather than all
  - Stream-based active learning enables FLARE to perform real time processing
    - Labelling as data comes in, rather than rely on an offline a priori trained data set

Federated Learning

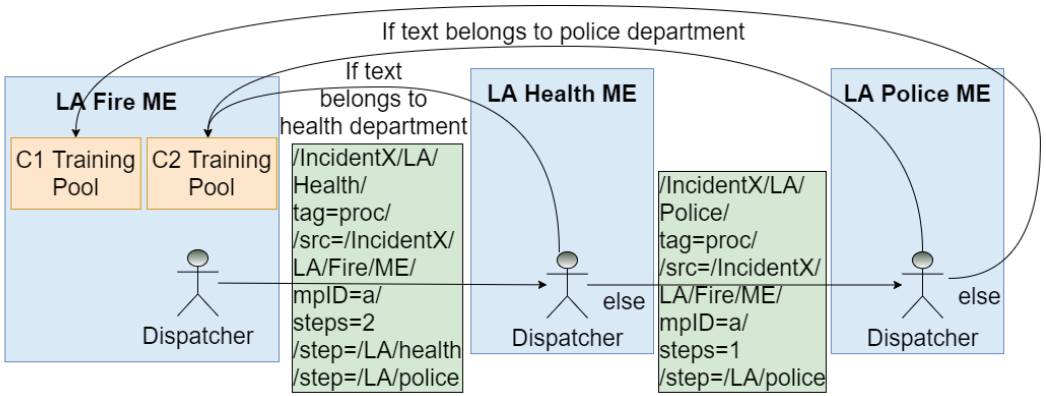
- Allows learning of model across multiple "workers" (SMEs), to assimilate and use the specialized knowledge of different entities (e.g., department-specific knowhow in disaster management) and collectively train the classifiers



- **Specifications:**
  - Algorithm: Vanilla Federated Learning
  - Iterations: Number of times models are aggregated across clients (tunable: based on stopping criteria)
  - Framework: **Flower**
  - Communications based on name-based pub/sub, with special "tags" (hierarchical name components)
    - SME-to-Coordinator, to pass local models "/…/tag=result/…"
    - Coordinator-to-SMEs, to pass aggregated models "/…/tag=model/…"
- **Advantage:** Since each SME works on a different set of data based on the keywords it uses for obtaining its dataset, its classifier is trained uniquely. FL (e.g., at incident coordinator) assimilates knowledge of the individual classifiers to create an aggregate model.

Message Passing

- Even with keyword-based crawling, it's possible that a message may not belong to the SME (department) that first receives it. FLARE uses message passing among SMEs for specialized labeling of text messages.
- If a dispatcher of a particular SME (e.g., Fire ME) is unable to classify the message, it is forwarded to the dispatcher for the SME that has the highest prediction probability predicted by the classifier.
- The process continues until the message is classified by one of the dispatchers - then it is added to the C2 training pool of the origin SME. If none of the dispatchers are able to classify it, the message is marked as irrelevant and added to the origin SME's C1 training pool.
- For message passing, we use the "proc" tag in the messages with required parameters (e.g., message ID, text, list of SMEs sorted by most to least likelihood of affinity to message).



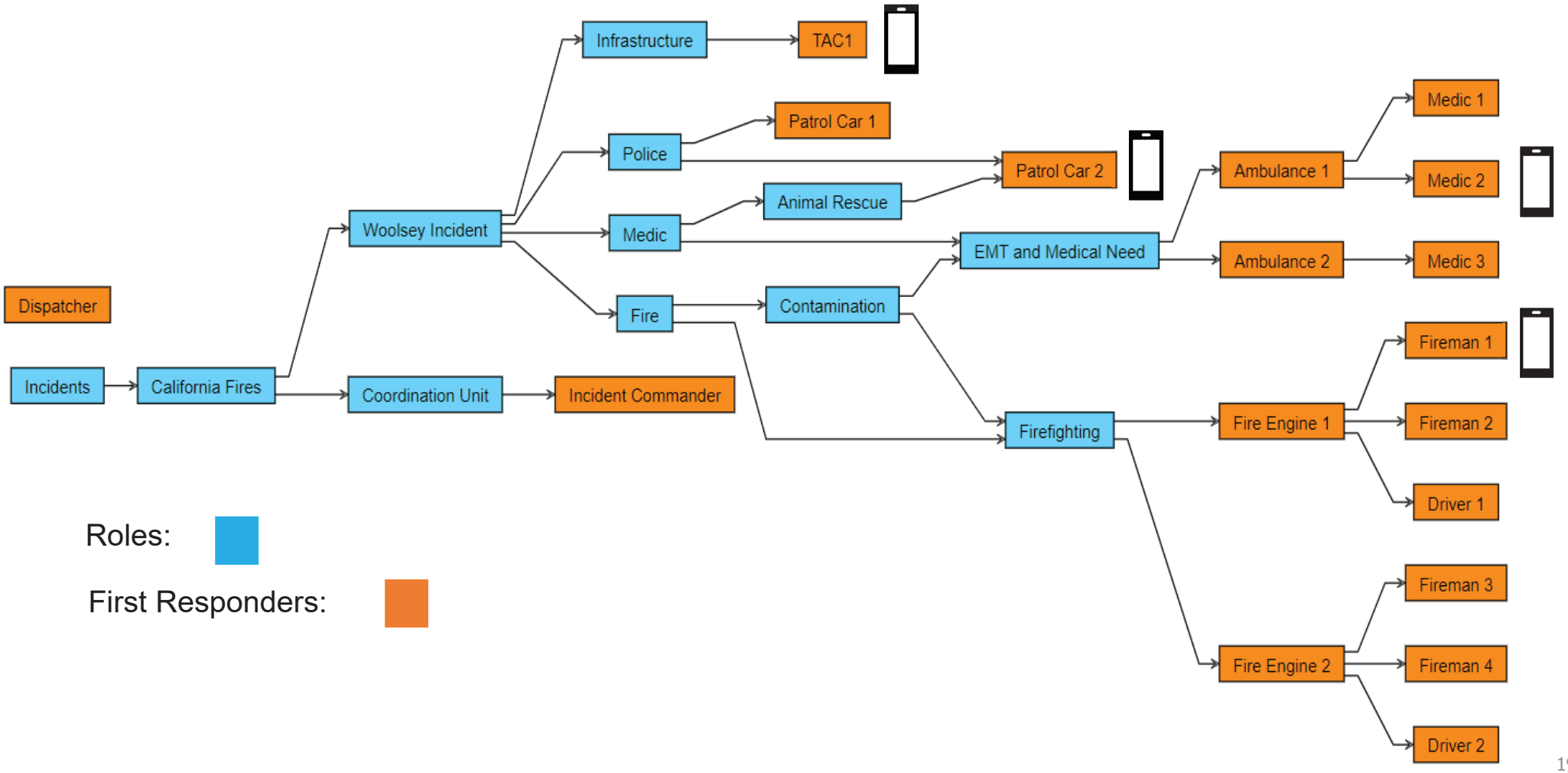
Results

- **Overall results with streaming twitter data collected in California Wildfires**
  - **98.38%** of all disaster-relevant tweets get published and delivered to "some" first responder(s), some may be to the incorrect organization/role
  - **1.62%** of tweets classified "irrelevant" by C1
    - Not examined further
    - But, these tweets appear to be borderline & non-actionable, e.g., opinion

	C1	C2	C3 (avg)
Accuracy (initial)	0.8262	0.6847	0.8553
Accuracy (dispatcher-assisted)	0.9091	0.8963	0.9291
Recall/F1 (initial)	0.9462	0.6183	0.8238
Recall/F1 (dispatcher-assisted)	0.9838	0.8589	0.9034
# of input tweets	3521 (of 3521)	2613 (of 2656)	2342 (of 2656)
# of correctly classified tweets	3201	2342	2176
# of tweets labelled	908	1223	441
Overall accuracy	0.9091	0.8818	0.8193

  - **88.18%** of all disaster-relevant tweets get published to first responder(s) in the right organization, whether or not it is to the right fine-grained role
    - Remaining **10.2%** delivered to incorrect organization - but can be delivered correctly based on the feedback from first-responders
  - Overall, **81.93%** of all disaster-relevant tweets get published to the first responder with correct role in right organization, at the finest granularity possible

Implementation Example: Incident Name Space



Implementation Example: Processing Tweets

Demo

Text Messages

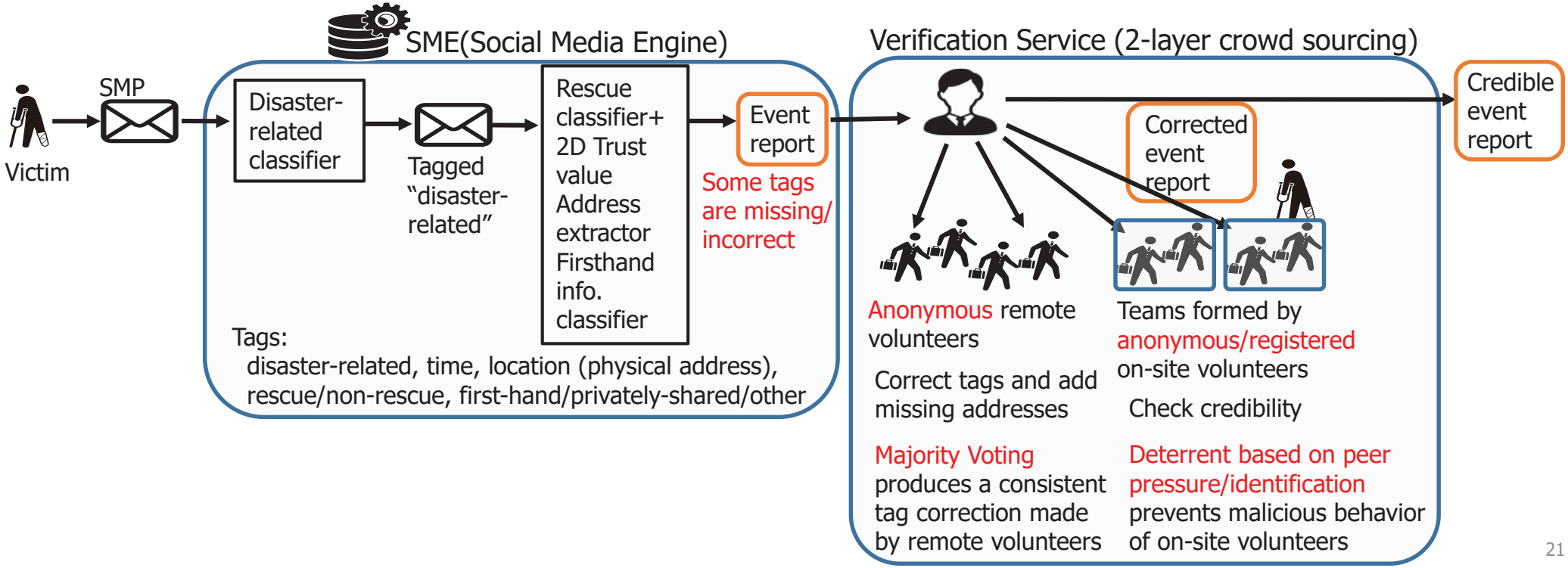
- Stuck in traffic due to bridge collapse.
- Case of looting witnessed on abc street.
- They're doing fire works when I get off wow.
- Severly injured, in need of medical assistance.
- Fire from above Sylmar Ca at around 4:30 pm.
- I am here for the animations. They're on fire.
- 5 people found dead in cars in Paradise wildfire.
- Animals waiting to be rescued in Malibu.
- Smokes and flames all over the place, unable to breathe.
- Foot cramps will be the death of me.

```
They're doing fire works when I get off wow. Irrelevant
Enter custom text message: Severly injured, in need of med
ical assistance.
Severly injured, in need of medical assistance. EMT and Me
dical Need
Enter custom text message: Fire from above Sylmar Ca at ar
ound 4:30 pm.
Fire from above Sylmar Ca at around 4:30 pm. Fire
Enter custom text message: I am here for the animations. T
hey're on fire.
I am here for the animations. They're on fire. Irrelevant
Enter custom text message: 5 people found dead in cars in
Paradise wildfire.
5 people found dead in cars in Paradise wildfire. Law
Enter custom text message: Animals waiting to be rescued i
n Malibu.
Animals waiting to be rescued in Malibu. Animal Rescue
Enter custom text message:
```

TAC 1 | Medic 2 | Patrol Car 2 | Fireman 1

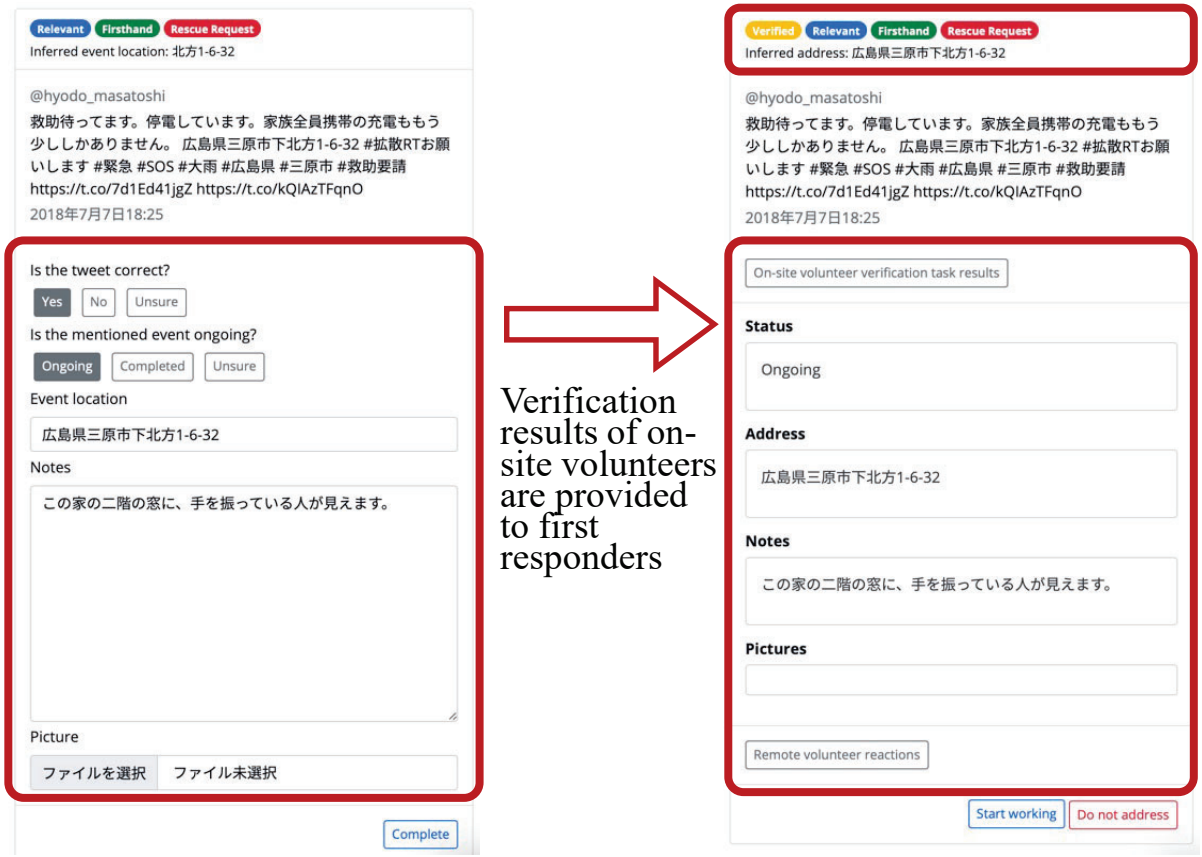
Experimentation of SME and verification service

- Data Set: $7*10^7$ SMPs (tweets) at Western Japan Heavy Rain, July 2018
- SME successfully extracts rescue SMPs from the data set
- 2-layer crowd sourcing successfully verifies credibility of event reports (NSMPs) extracted by SME: Emulation and Agent-based simulation



Implementation of SME and Collaboration Tool

- SME (classifiers)
  - Implement the three classifiers using BERT fine-tuned with tweets of Western Japan Heavy Rain
    1. Disaster related
    2. First-hand information
    3. Rescue request
- Collaboration tools for remote, on-site volunteers and first responders
  - Remote volunteers correct classification results
  - On-site volunteers verify tweets and input the verification results
  - First responders only see tweets verified by on-site volunteers and tweets of high-confident classification to rescue requests



(a) On-site volunteers' view

(b) First responders' view

Screenshots of the collaboration tool

Extracting Rescue SMPs

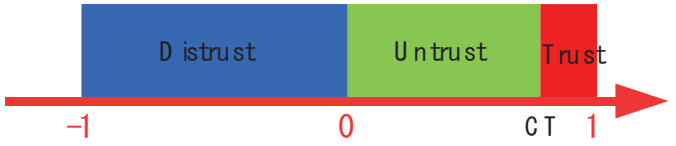
- Two classifiers for disaster-related (UCR) and rescue (Osaka) are used
- Leaning-based classifier and correction based on 2D trust model
- Evaluation summary:
  - Extraction results of learning-based filter
    - Accuracy 0.96, Precision 0.84, Recall 0.70
  - Correction based on 2D trust model improves precision to 0.90 (explained by next 2 slides)
- Analysis based on attention values
  - Attention: How much a particular word will contribute to classification results
  - Words representing location names, flooding situations, and victims have a high impact on the rescue request classification

Label / Prediction	Rescue	Others
Rescue	42	8
Others	18	605

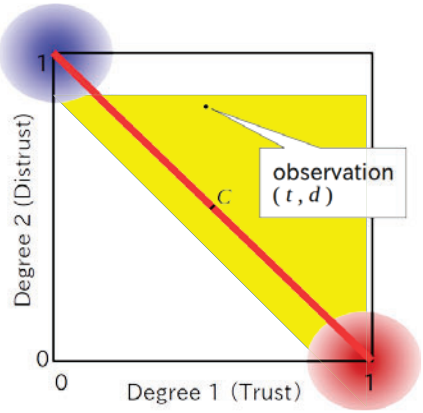
Classification Results	Words of higher attention values
Rescue request (True Positive)	town, - (hyphen), floor, people, left behind, flood, water, city, isolated, Chome (Town number in Japanese)
Rescue request (False Positive)	town, #, floor, city, San (title in Japanese), Chome
Non Rescue Request (True Negative)	#, @, town, http, please, city, people, dissemination
Non Rescue Request (False Negative)	#, evacuation, number, help, city

Correction by 2D Trust values

- Conventional (1D) trust model assumes that a human evaluator can calculate the degree (d) of distrust if the degree (t) of trust is given. (i.e. d=1-t)
  - 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. iTrust'05 (2005)



- However, such an assumption is sometimes too strong (e.g., a FR may judge that a message is basically credible while also having some doubtful aspects), and we introduced a 2D trust model where trust and distrust degrees are independent.
  - Fuzzy logic proves our model's correctness, and we found that our model is more powerful than subjective logic (SL), since only our model can deal with **contradictory situations** (top right region filled yellow).



- To exclude non-rescue requests, we assign 2D trust values to messages in disasters.
  - A high trust degree is assigned if a unique postal code can be derived from a message (i.e., a complete address is provided in the message).
  - A high distrust degree is assigned if there are "non-matching" words (Notice that words and phrases such as "Dear all," "your concern," and "Thank you" are unlikely to be used in an urgent call for help).

Correction by 2D Trust values (Cont'd)

- Example 1:
  (Trust: high, Distrust: low)
  A unique postal code (710-1304) is derived and it does not contain "non-matching" words.
- Example 2:
  (Trust: low, Distrust: not calculated)
  63 postal codes are derived.
  - 716-0009, 716-0039, 716-0018, 716-0015, ...
- Example 3:
  (Trust: high, Distrust: high)
  A unique code (709-0631) is derived, but it contains a non-matching word (物資, **goods**). In this case, a net trust value (t-d) is low.

拡散希望 友人のお母さんを救助ねがいます。倉敷市 真備町 尾崎■■-■■ 女性 1 名と犬 1 匹 2 階まで浸水 脚立にのり、水から逃れています。どうか拡散をお願いします。
＃倉敷市 ＃真備 ＃高梁川 ＃豪雨 ＃救助要請 ＃ SOS ＃自衛隊 ＃救助

To everyone: My friend's mother needs help. One woman and one dog are flooded up to the 2nd floor at ■■-■■ Osaki, Mabi-cho, Kurashiki-shi They're having to use a stepladder to escape from the water. Please spread the word. #Kurashiki-shi #Mabi #Koryo river #flooding #rescue request #SOS #Self-Defense Forces #rescue

倉敷真備町は無人家屋で田畑は放置地だったが、総社の秦がまずい。葡萄畑の水をポンプでぬいてるが水浸し。高梁市の叔母が床上浸水。倉敷の叔母以外はみんな避難所。ハトコは総社アルミ工場爆発の消火中。実家借主の息子さんが、コンビニで玄関ごと吹き飛ばされるも無事

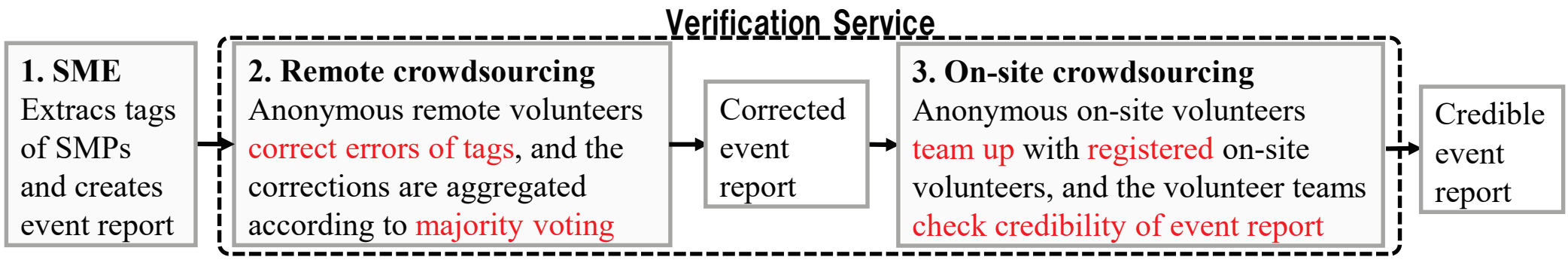
Kurashiki Mabi-cho was uninhabited and the fields were abandoned, but Hata in Soja-shi was in bad shape. They are pumping water out of the vineyard, but it is flooded. My aunt in Takahashi-shi was flooded above floor level. Everyone except my aunt in Kurashiki-shi is in a shelter …. <snip>

岡山市東区東平島の避難場所が小学校らしいんだけど、そこ昨日の夜まで浸水してたから別の場所に避難してるのかな？避難場所へ物資届けたくても遠すぎるしな。。他県からだと何ができるのか。

I heard that the evacuation site for Higashi-hirajima, Higashi-ku, Okayama-shi is the elementary school, but it was flooded until last night, so are they perhaps evacuating to someplace else? Maybe it's too far to deliver goods to the evacuation site? What can you do if you're from another prefecture?

Effectiveness of Verification Service

- **Mechanism :** Verification Service is composed of 2-layer crowd sourcing



- **Goal of evaluation:** The effectiveness of Verification Service is confirmed by
  - Conducting an emulation of "Remote crowd sourcing":
    - Evaluating whether anonymous remote volunteers can correct errors in "tags assigned by SME"
    - Evaluating whether anonymous remote volunteers can extract "detailed physical addresses of disaster locations" from tweets
  - Conducting an agent-based simulation of "On-site crowd sourcing":
    - Evaluating whether a team of registered/anonymous on-site volunteers can check the credibility of each Event Report
    - Evaluating whether "deterrent based on identification/peer pressure" can enhance trustworthiness of registered/anonymous on-site volunteers in conducting credibility check

Effectiveness of Remote crowdsourcing

- **Objective:** Evaluating how remote crowdsourcing correct errors of "tags in SMPs" assigned by SME
- **Approach:**
  - Chooses randomly 1000 disaster-related SMPs tagged "rescue" and "non-rescue" and evaluates how many corrected SMPs meet the condition of event report
  - Evaluates whether physical address is extracted (corrected) from the content of the "SMP where physical location is not identified by SME"
- **Observation:** Confirmed that the credibility of tagged SMPs can enhanced by a large number of remote anonymous volunteers

	Number of Extracted Event Reports (original)	Number of Extracted Event Reports (corrected)
Rescue	28	12
Non-Rescue	84	47

Wrong tags were corrected by remote volunteers, which extracted trustworthy 59 event reports out of 1000 tagged tweets.

	Corrected Non-rescue to Rescue	Corrected Rescue to Non-Rescue
Rescue	N/A	6
Non-Rescue	0	N/A

Six wrong Rescue tag were corrected by remote volunteers.

	Exact home address level	Street/town level with landmark	Street/town level without landmark
Rescue	4	0	2
Non-Rescue	7	17	13

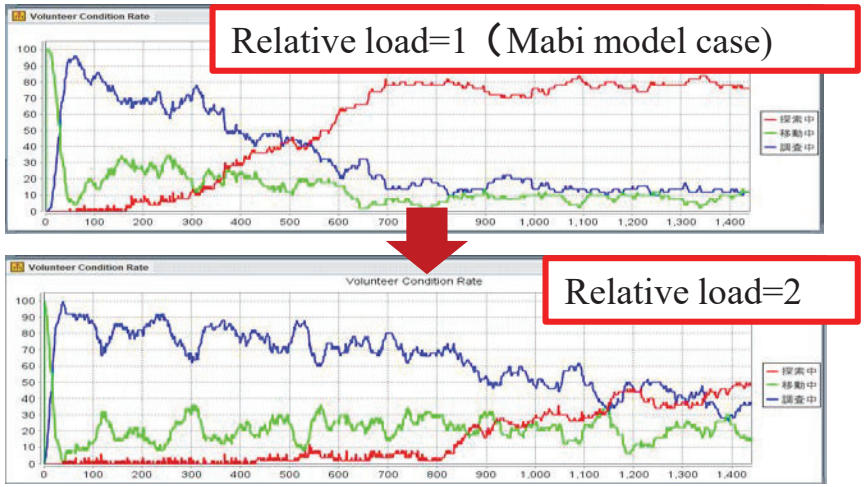
A tweet with a detailed address is likely to be an event report.

Effectiveness of On-site crowdsourcing

- **Objective:** Evaluating how on-site crowdsourcing checks credibility of event reports
- **Approach:**
  - Simulates the behavior of registered/anonymous on-site volunteer teams in disaster areas through agent-based simulation, based on Western Japan Heavy Rain, July 2018
  - Conducts simulations that take into account "deterrent based on identification" and "deterrent based on peer pressure".
- **Observation:** Clarified the effect of the ratio of registered on-site volunteers on the accuracy of credibility check and the effect of the relative load (depending on amount of Event Reports, number of teams, area of disaster site) on the speed of information cleansing.

Ratio of registered on-site volunteers	True Rate	False Rate
0 %	75.0%	25.0%
12.5%	77.3%	22.7%
25%	86.4%	13.6%
50%	95.8%	4.2%

Accuracy of credibility check is greatly improved when the volunteers includes more than 1/4 of registered volunteers.



The time required for credibility check increases as the relative load per volunteer team increases. The result can contribute to estimate the number of people required according to the load in each affected area.

Conclusion and Plan of the Future

- 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 integrated them
  - Evaluated the architecture based on data from SMPs
- Plan of the Future
  - Write the paper based on the integrated architecture and further evaluation

Acknowledgement

- This work was supported by the US National Science Foundation grant CNS-1818971, US Department of Commerce and NICT Contract No. 193.