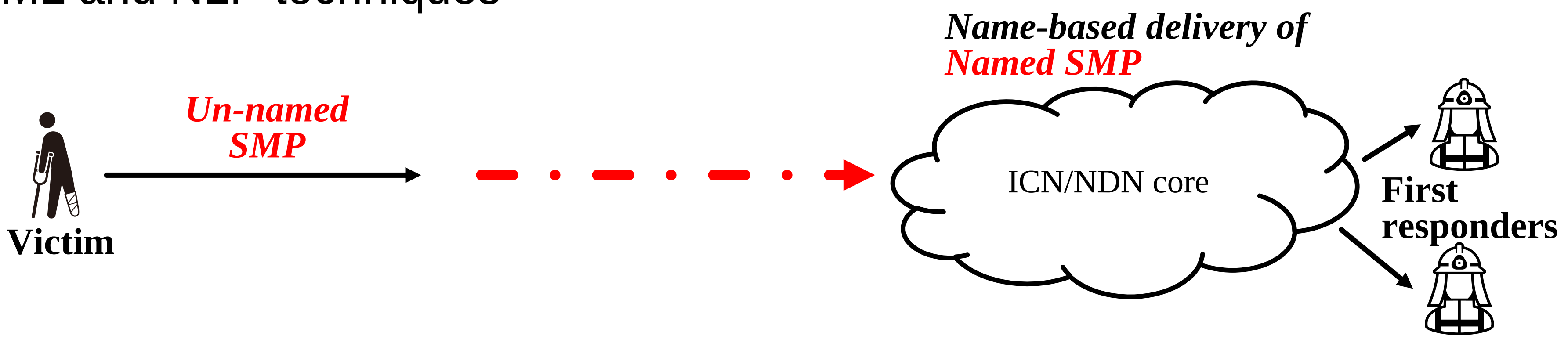
FLARE: Federated Active Learning Assisted by Naming for Responding to Emergencies

Viyom Mittal, Mohammad Jahanian, K. K. Ramakrishnan
Department of Computer Science and Engineering
University of California, Riverside

UC RIVERSIDE

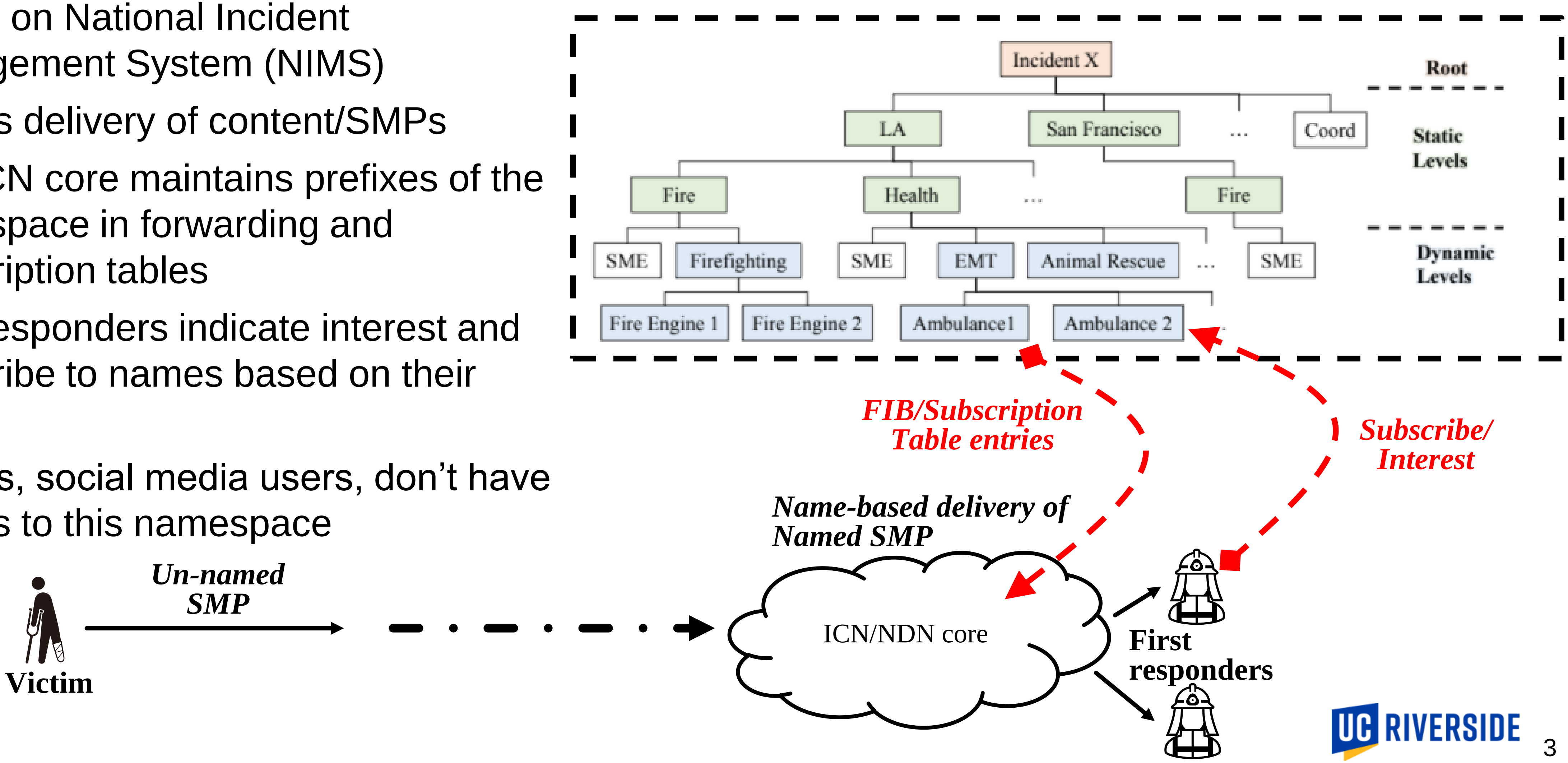
Name-based Delivery of Critical Information: Social Media

- Social media posts (e.g., tweets): relevant content sent to the right recipients; timely delivery
  - During disasters, “victims” post about incidents; need relevant first responders to receive & act
- Information-Centric Name-based delivery is very beneficial for such distribution
  - However, every content needs to be named by ‘publisher’ before “entering the network”
  - Both interest/data request-response and pub/sub (COPSS, CNS) depend on named publications
  - Challenge: what if content publishers do not have access to the namespace to assign names (INDENT THIS LESS)
  - Such as civilian social media users reporting, i.e., victims
- FLARE: Deliver free-form text content (social media posts, SMPs) to relevant users w/ICN
  - Using ML and NLP techniques



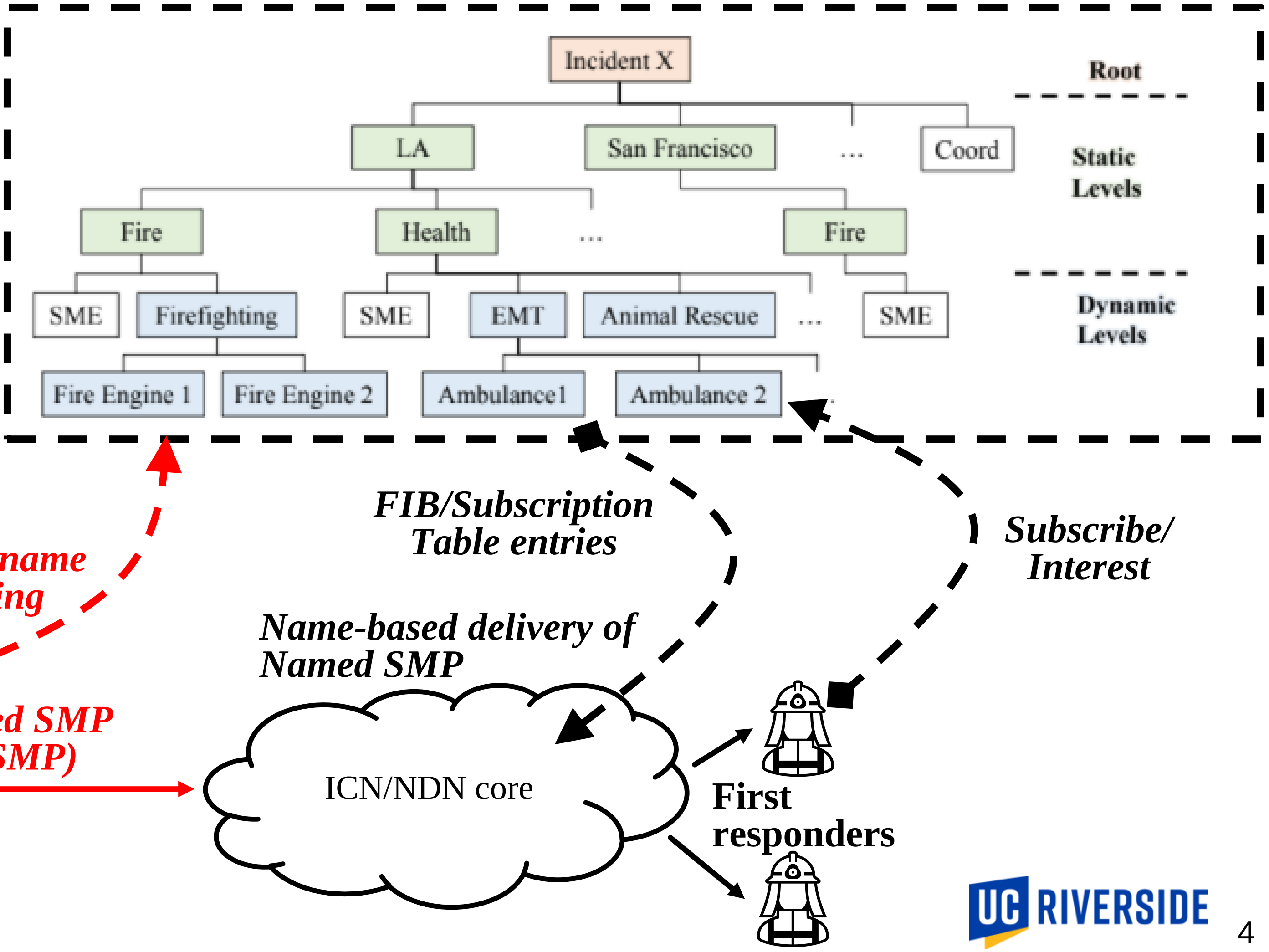
Namespaces: Emergency Response information Dissemination

- Incident Namespace
  - Based on National Incident Management System (NIMS)
  - Guides delivery of content/SMPs
  - The ICN core maintains prefixes of the namespace in forwarding and subscription tables
  - First responders indicate interest and subscribe to names based on their roles
  - Victims, social media users, don’t have access to this namespace



From free-form content to name-based dissemination

- Social Media Engine (SME)
  - Maps free-form text (e.g., social media posts) to the right name components, and assigns names, e.g., *"/IncidentX/LA/Fire/Firefighting/FireEngine1/LARegionA"*, and publishes to right first responders
  - SME also uses the namespace for its operations
  - Incident Commander(s), Dispatchers: manage namespace

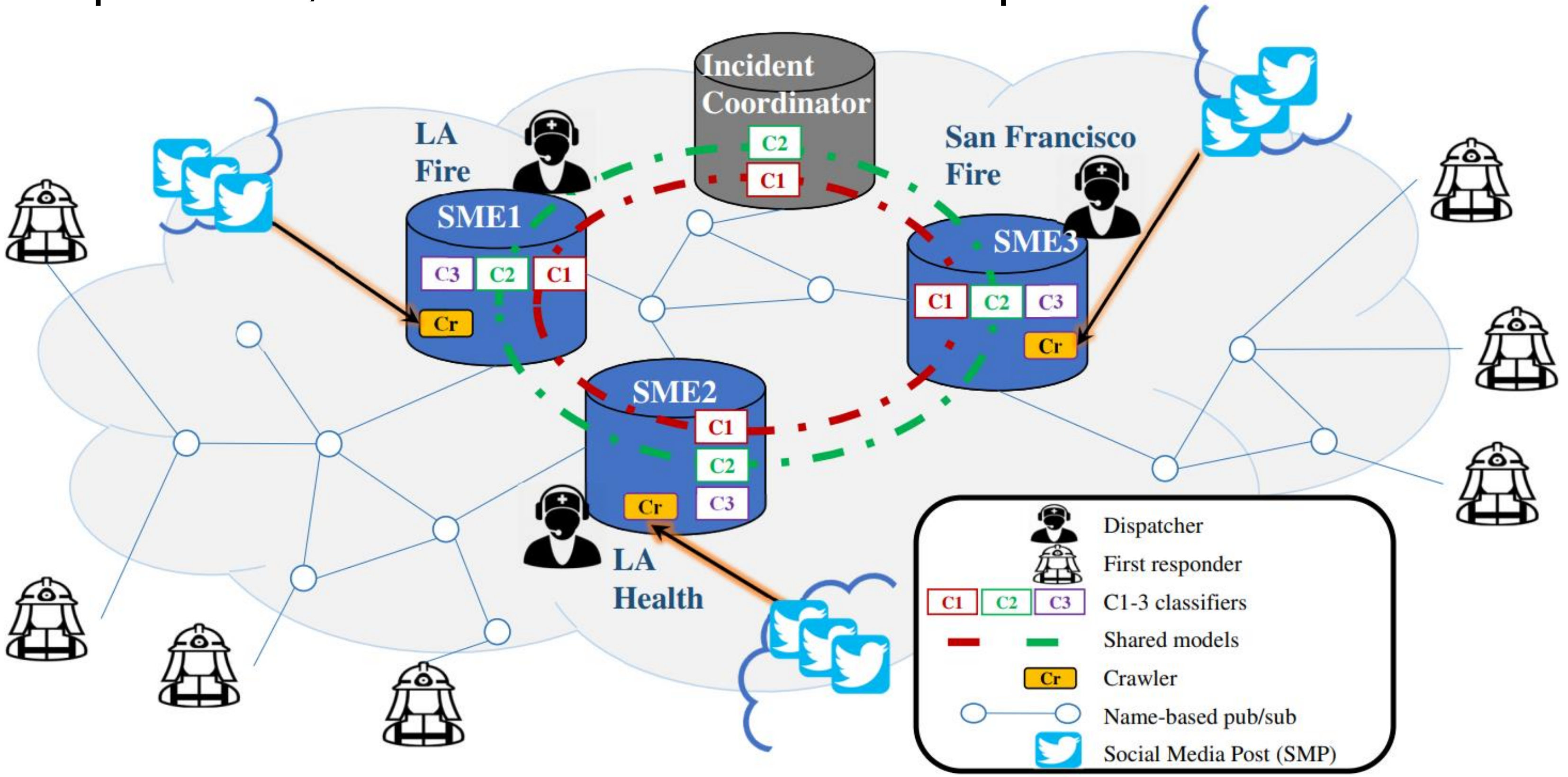


Challenges with traditional classifiers

- Difficulties with traditional learning approach for inference of text messages
  - May not have enough training data to begin with
  - Too much labeling (typically manual) effort required
  - Static classification techniques are not suitable for classification w.r.t. dynamic namespace
  - Unable to utilize specialized knowledge of various participating entities

FLARE: Architecture overview

- **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
- Media Engines associated with departments, equipped with multiple classifiers (C1-C3), and dispatchers, and the full incident namespace



Mapping content to names - components of MEs

- Each Media Engine (ME) has its own (decentralized) **“Crawler”**:
    - Collects text-based content and/or crawls SMPs in real-time during or in the aftermath of disaster.
    - Collects data with specific keywords based on the department’s specialization.

Mapping content to names - components of MEs

- Each Media Engine (ME) has its own (decentralized) **“Crawler”**:
  - Collects text-based content and/or crawls SMPs 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):** predicts if a text or 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

UC RIVERSIDE

Mapping content to names - components of MEs

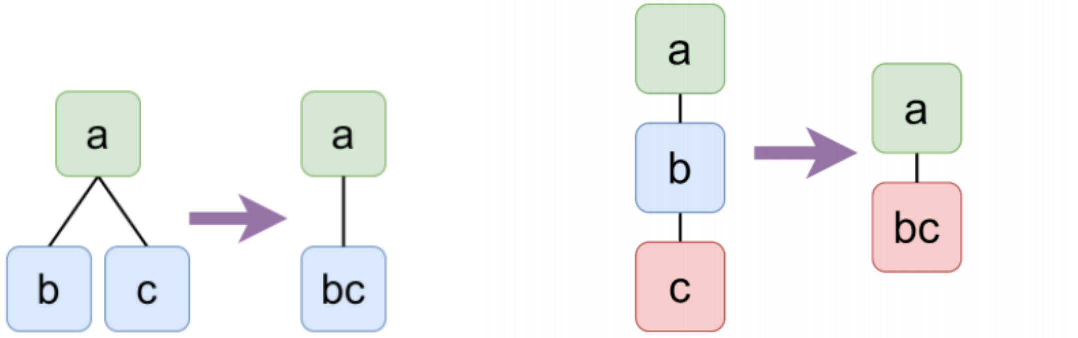
- Each Media Engine (ME) has its own (decentralized) **“Crawler”**:
  - Collects text-based content and/or crawls SMPs in real-time during or in the aftermath of disaster.
  - Collects data with specific keywords based on the department’s specialization.
- Multiple classifiers
  - **C1 (incident relevance predictor):** predicts if a text or 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
- Finally the **“Named MediaPost Generator”** componet takes the output of C3 as input and forms and publishes the MP with full hierarchical name, e.g.: *“/IncidentX/LA/Fire/Firefighting/FireEngine1”*

Media Engine Features (To address the challenges)

- **DNN classifier with Universal Sentence Encoder (USE):**
  - USE is pre-trained for sentence embedding over huge data corpus allowing it to capture rich semantic information. This helps in faster learning without the requirement of large initial dataset.
- **Active learning:**
  - Reduces the manual labeling effort of the dispatcher by selecting only crucial messages required for training of the classifier.
  - Helps in providing support for dynamic namespace changes
- **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 the free form text messages across different Media Engines for their finer-grained classification by specialized knowledge of the dispatcher.
  - Leverages organizational expertise in labelling more efficiently to eventually achieve better performance.

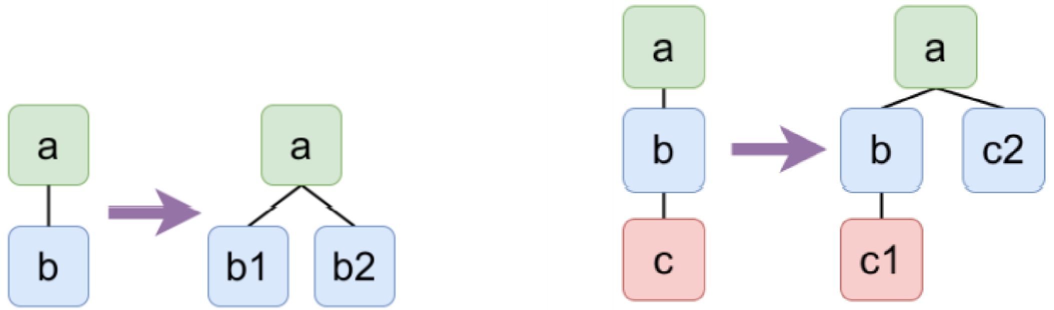
Dynamic Namespace Support

- During the course of a disaster, the incident namespace may have to be updated as the situation evolves.
- These updates may be because of addition or deletion of roles, as well as changes in the command chain, according to the real-time needs of the emergency response tasks. These updates can be applied at various levels of the namespace.
- These updates can be categorized into two types:
   1. **Updates not requiring classifier update:** Eg: Name deletion, name merging, etc.
   2. **Updates requiring classifier update:** Eg: Name addition, name splitting, etc.



(a) Name merging at same level (b) Name merging across different levels

Fig: Namespace updates that do not require classifier update



(a) Name splitting at same level (b) Name splitting across different levels

Fig: Namespace updates that require classifier update

Dynamic Namespace Support - Evaluation

- We evaluate the instantiation of a new classifier with increased number of classes due to a change in the namespace that requires a change in the classifier against a static classifier.
- Dataset:
  - **Total tweets:** 1000
  - **Number of classes till first 400 tweets:** 2 (“Structure and Shelter”, “Other”)
  - **Number of classes after first 400 tweets:** 3 (“Structure and Shelter” split into “Structure, Building & Road Damages” and “Shelter, Shortage & Outage”)
- The dynamic classifier, is able to rectify the drop in accuracy. Thorough,there is a reasonable increase in the load on the dispatcher, but that is alleviated over time as the new classifier gets trained.

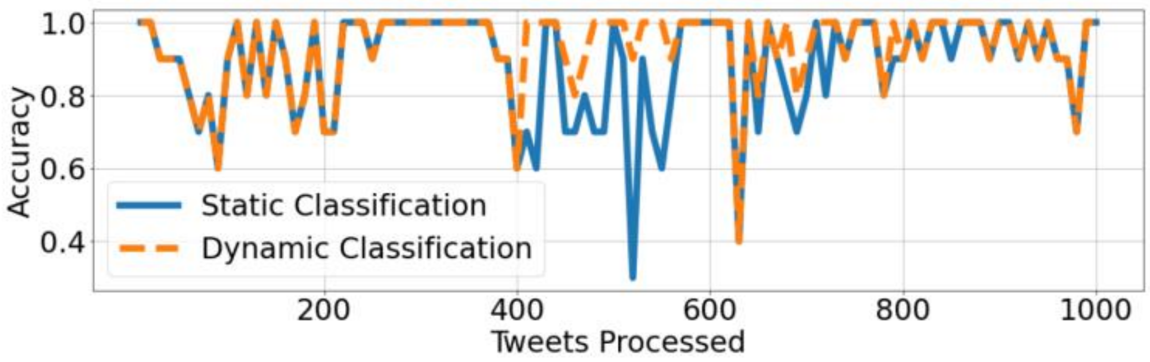


Fig: C3 accuracy while performing class split (split at 400 tweets)

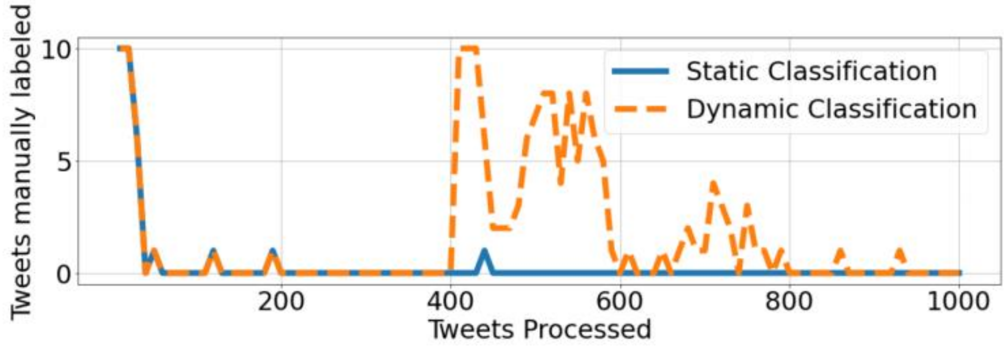
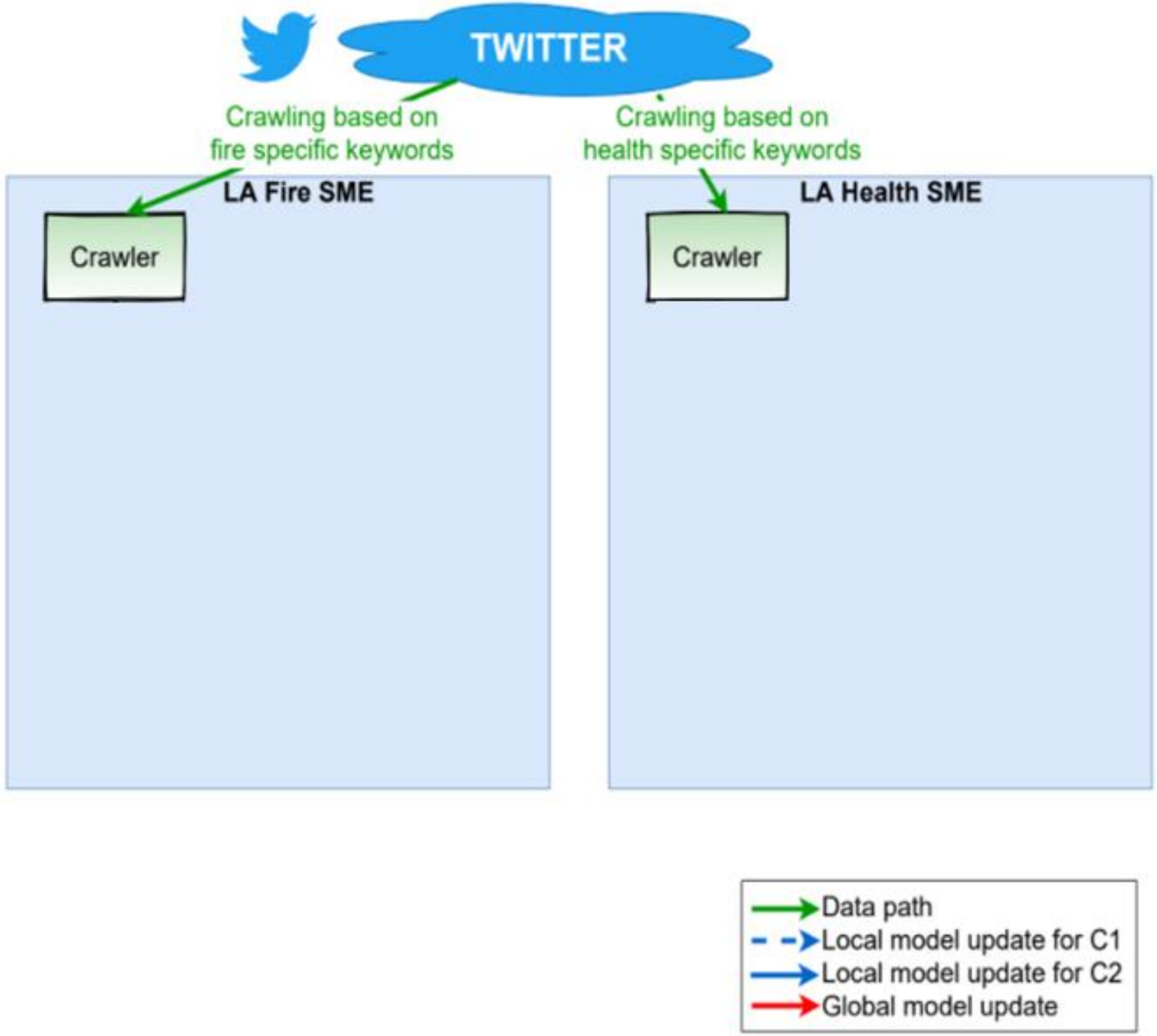


Fig: C3 tweets labelled while performing class split (split at 400 tweets)

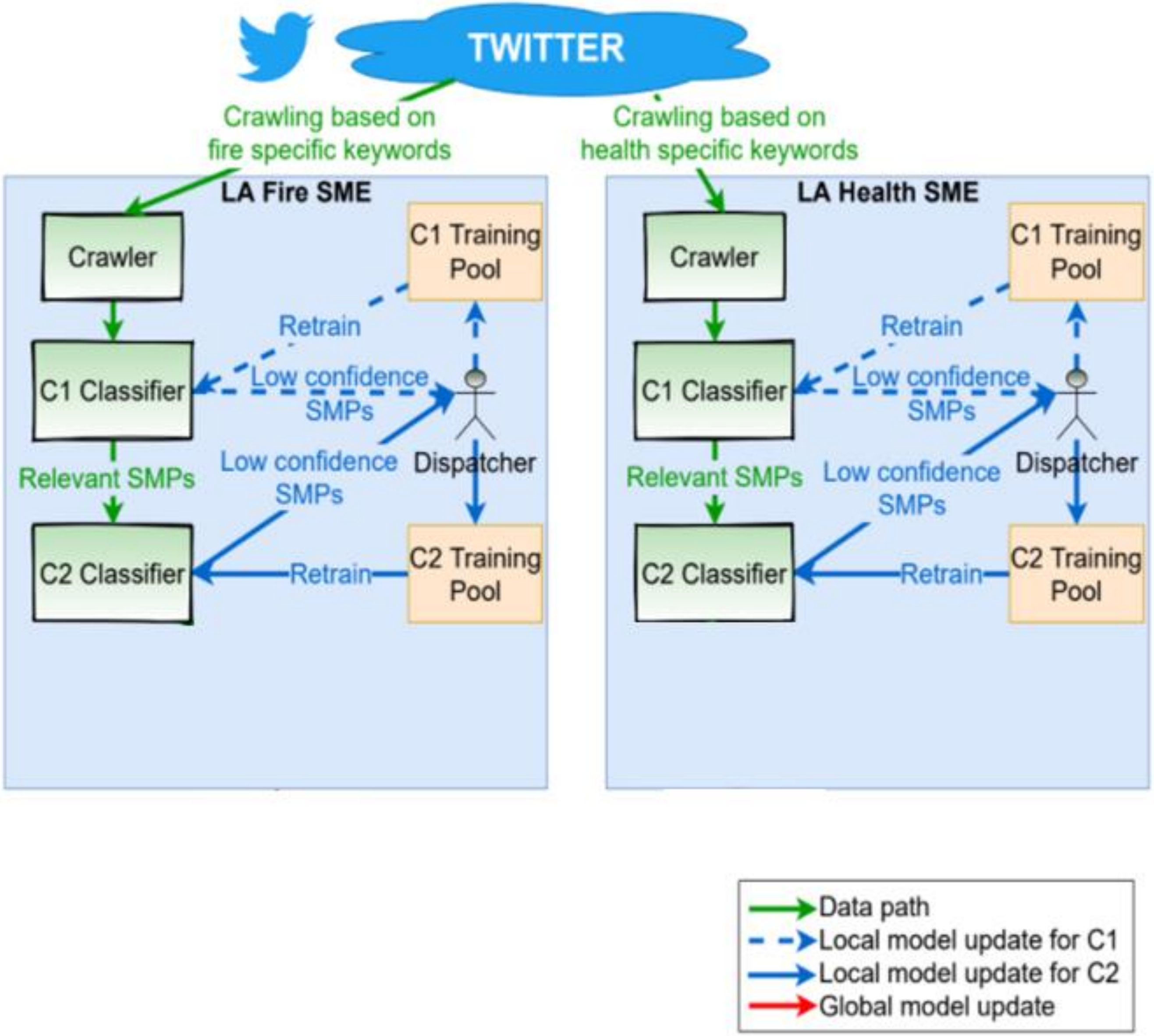
Media Engine Pipeline



Media Engine Pipeline

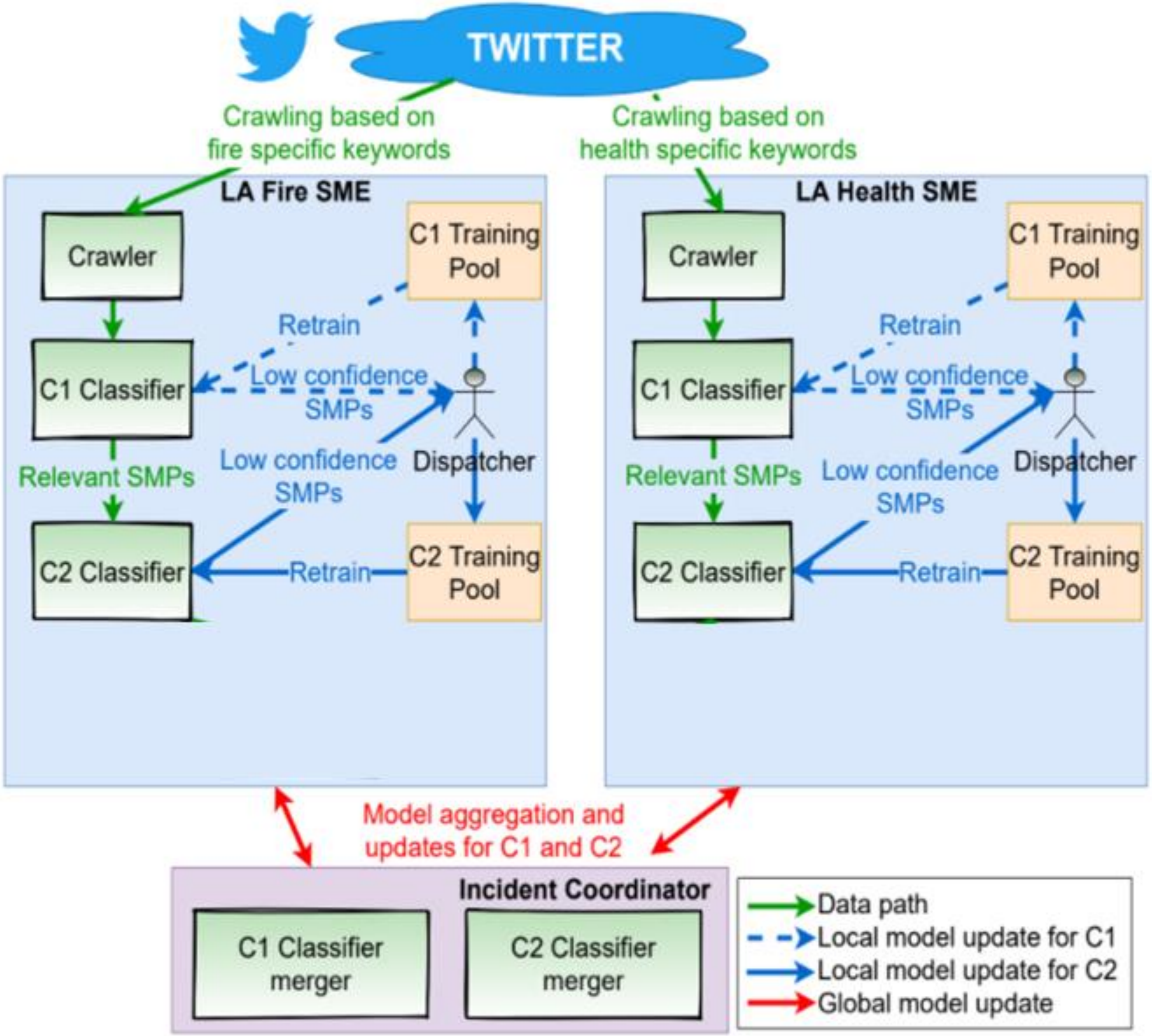
- **DNN classifier with Universal Sentence Encoder (USE):** Used in all the classifiers.

- **Active learning:** Used in all **C1**, **C2** and **C3** classifiers.



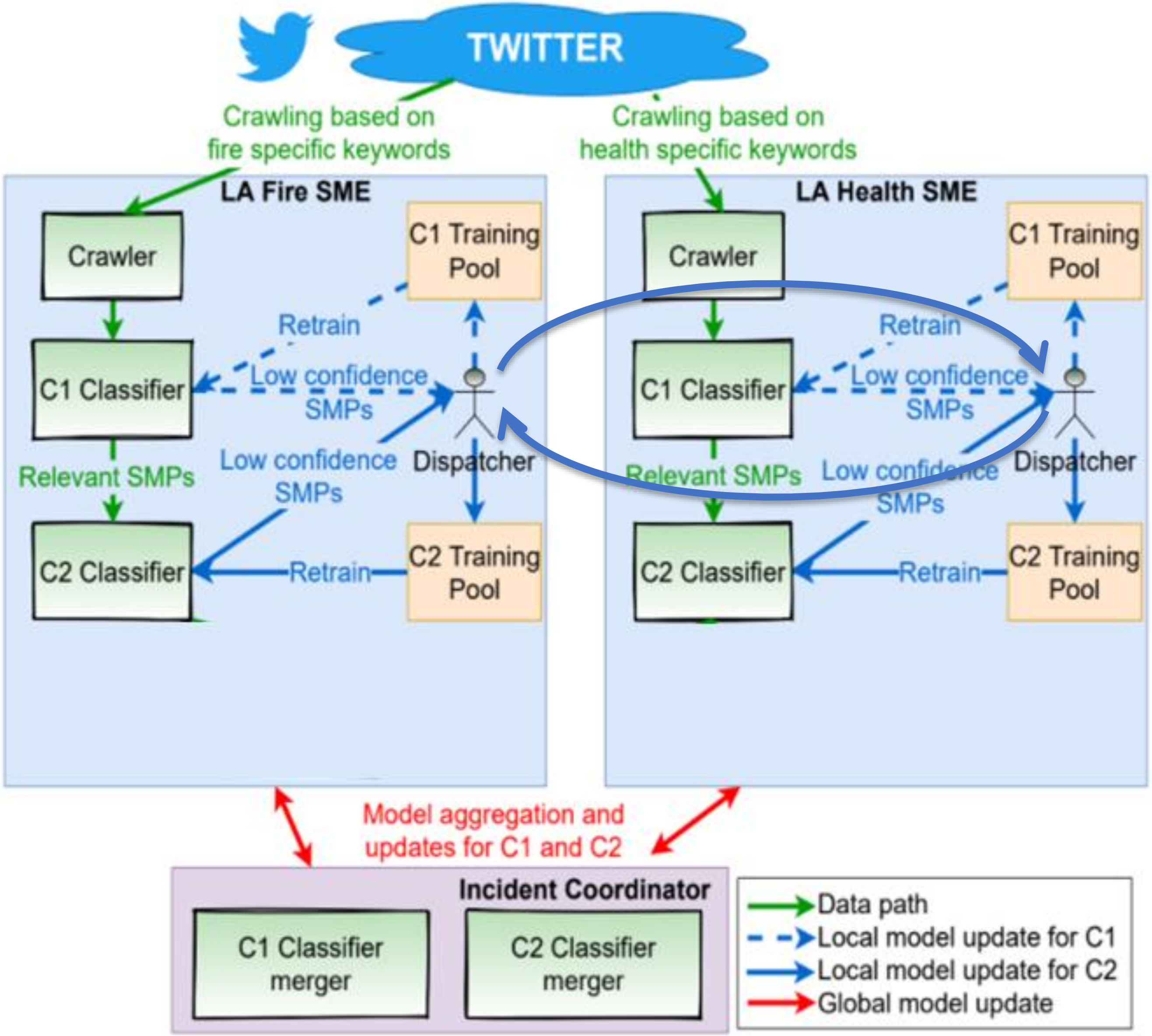
Media Engine Pipeline

- **DNN classifier with Universal Sentence Encoder (USE):** Used in all the classifiers.
- **Active learning:** Used in all **C1**, **C2** and **C3** classifiers.
- **Federated learning:** Used in **C1** and **C2** classifiers only -- both are common across all MEs.



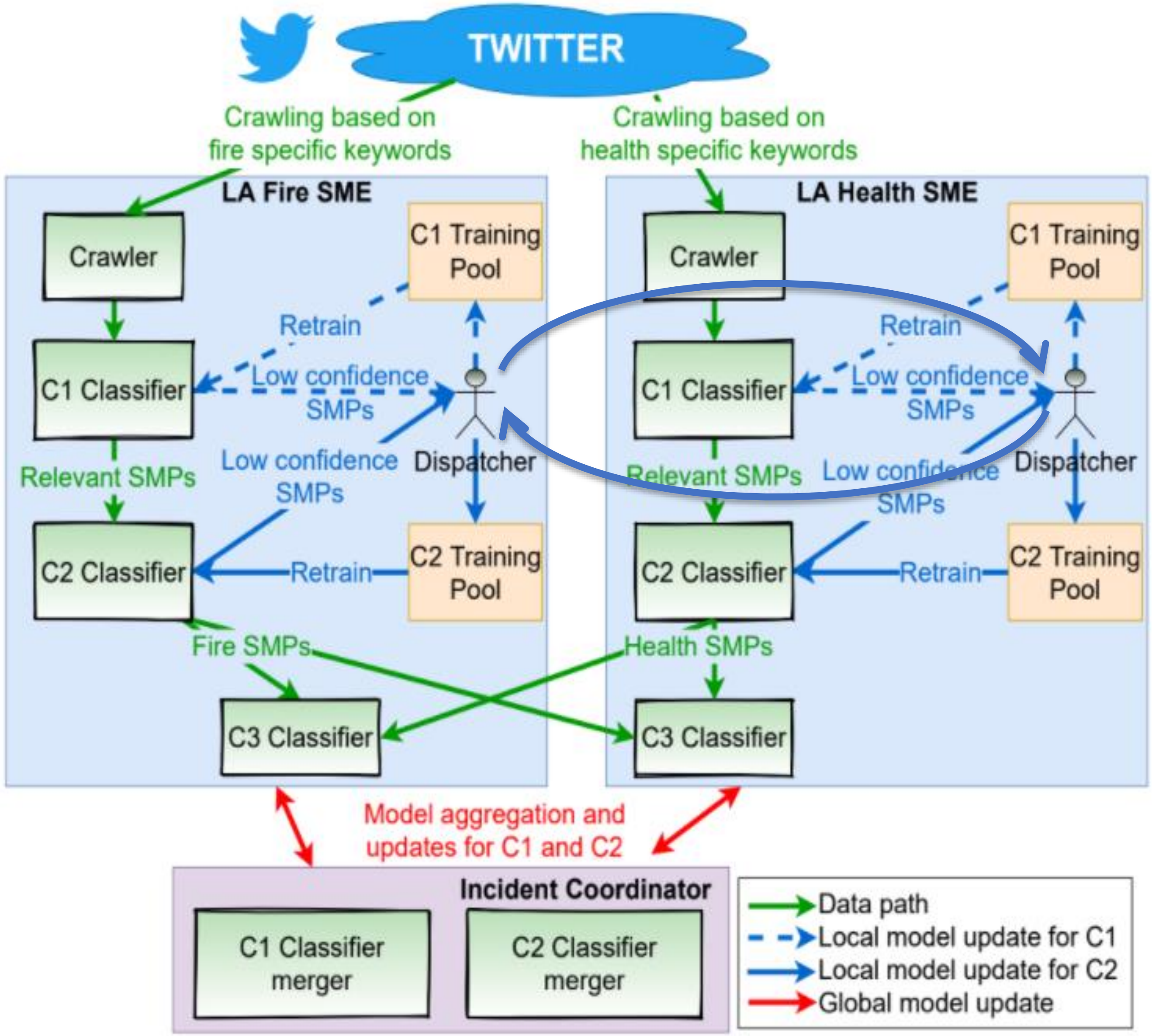
Media Engine Pipeline

- **DNN classifier with Universal Sentence Encoder (USE):** Used in all the classifiers.
- **Active learning:** Used in all **C1**, **C2** and **C3** classifiers.
- **Federated learning:** Used in **C1** and **C2** classifiers only -- both are common across all MEs.
- **Message passing:** Used **only for C2** classification when dispatcher is unable to classify the text message appropriately.



Media Engine Pipeline

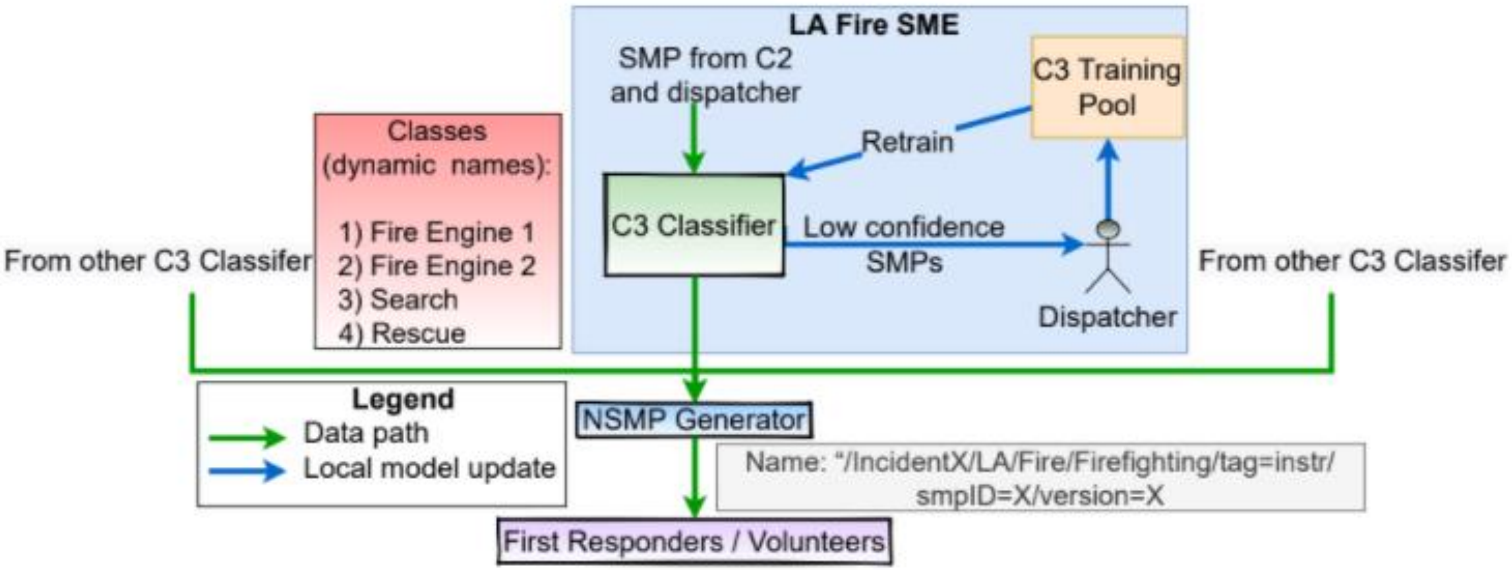
- **DNN classifier with Universal Sentence Encoder (USE):** Used in all the classifiers.
- **Active learning:** Used in all **C1**, **C2** and **C3** classifiers.
- **Federated learning:** Used in **C1** and **C2** classifiers only -- both are common across all MEs.
- **Message passing:** Used **only for C2** classification when dispatcher is unable to classify the text message appropriately.



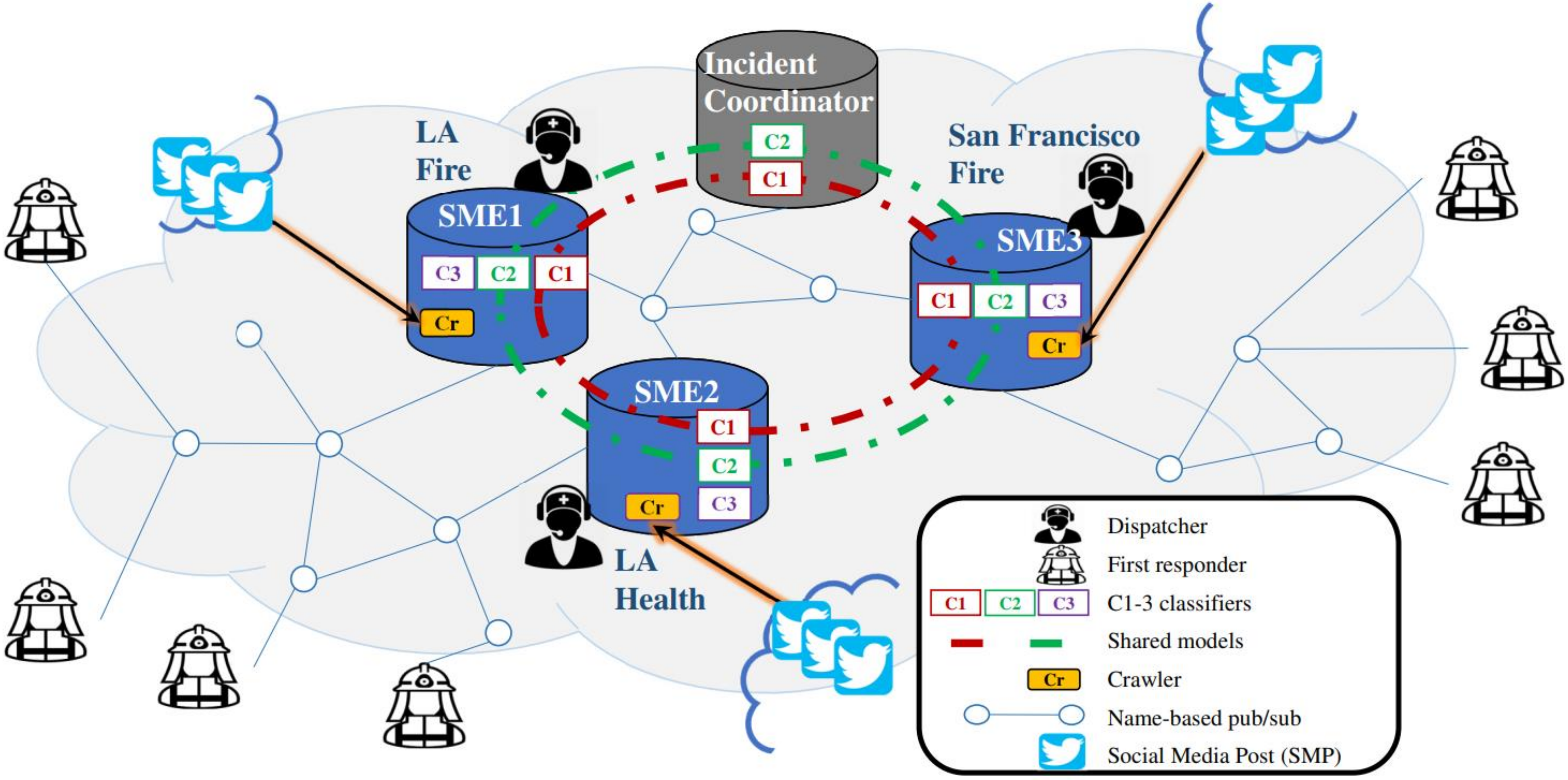
UC RIVERSIDE

Media Engine Pipeline

- **DNN classifier with Universal Sentence Encoder (USE):** Used in **C3** as well
- **Active learning:** Used in C3 as well
- **Dynamic namespace adaptation:** Specific to C3 only



Overall architecture



Results

- **Overall results with streaming data**
  - **98.38%** of all disaster-relevant tweets get published and delivered to “some” first responder(s),
    - some may be to incorrect organization/role.
  - **1.62%** of tweets classified “irrelevant” by C1
    - 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.

Summary of FLARE

- FLARE is a framework to map text messages and social media posts to the correct first responders in an online real-time manner, during a disaster, through NLP/ML-based Media Engines (MEs)
- FLARE uses namespace-assisted multi-classifier pipeline. Enables even some inaccurately labelled tweets reach “somewhat relevant” first responders , e.g., having the right department
- Utilizes active learning to reduce manual labelling effort
- Uses federated learning for collaborative learning among the MEs and dispatchers of different organizations
- Message passing among MEs to take advantage of specialized knowledge of different dispatchers and for improving classifier learning
- With fully streaming real-world Twitter data on California Wildfires, FLARE maps and delivers 81.93% of all disaster-relevant tweets to the right first responders, at the finest level of granularity

Q/A

Thank you!