Trustworthy and Protected Data Collection for Event Detection Using Networked Sensing Systems

Authors:

Md Zakirul Alam Bhuiyan and Jie Wu

The 37th IEEE Sarnoff Symposium, September 19 -21, 2016, Newark, NJ, USA

TEMPLE UNIVERSITY

FORDHAM UNIVERSITY
THE JESUIT UNIVERSITY OF NEW YORK

Outline

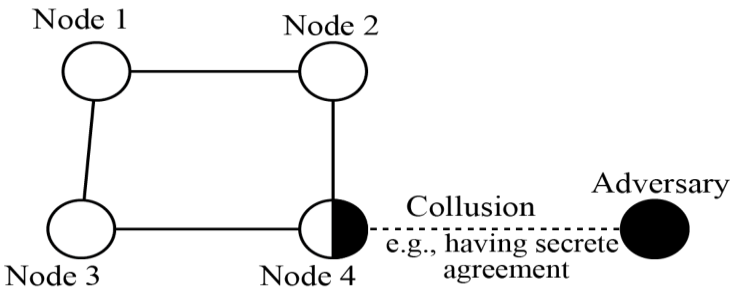
- Motivation
- Existing Work
- Proposed Framework
- Trustworthy Data Collection
- Protected Data for Aggregation
- Conclusion & Future Work

Motivation

- Wireless Networked Sensing Systems
  - Various applications
    - Crowd sensing, structural health monitoring (SHM) or damage event detection
  - Requirements
    - The quality of the data or the quality of the monitoring and timely detection of an event
      - E.g., Structural damage or fire

Motivation

- Challenges with the Quality of the Monitoring
  - Untrustworthy data
    - Security attack
      - Collusion attack and the malicious attack
      - Some sensors constantly provide truthful data while others may generate biased, compromised, or even fake data
    - Fault occurrence

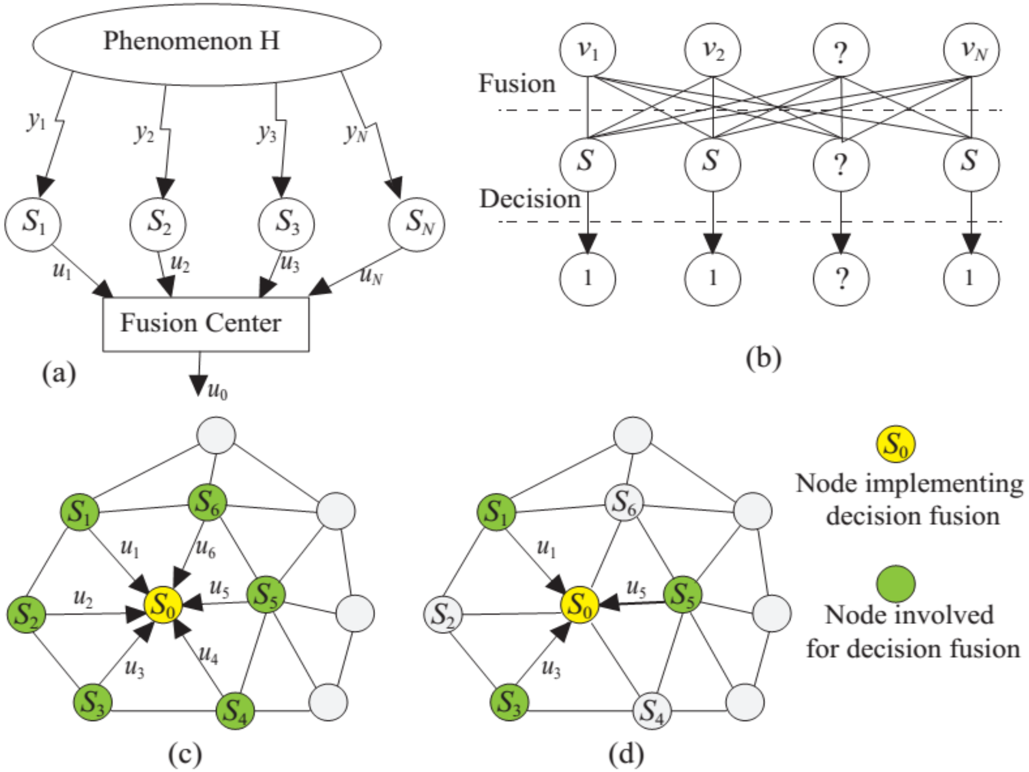


Motivation

- Challenges with the Quality of the Monitoring
  - **Unprotected data**
    - Data alteration
      - During transmission
      - After transmission, and
      - Before aggregation

Existing Work

- Existing Work
  - Security related work
  - Decision-making related work



Our Framework: TPDC

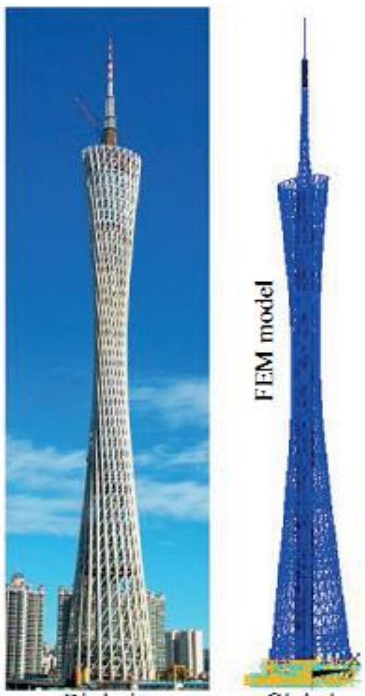
- Trustworthy and Protected Data Collection
  - Identify whether the acquired data is trustworthy or not, and finally transmit the trustworthy data.
  - Identify whether the received data is protected or not before aggregation
- Two Solutions
  - Trustworthy data acquisition
    - We use ‘mutual information independence (MII)’ as an indirect signal measurement, assuming that a prior correlation model presents
  - Protected data collection
    - We use a truth discovery approach

Our Framework: TPDC

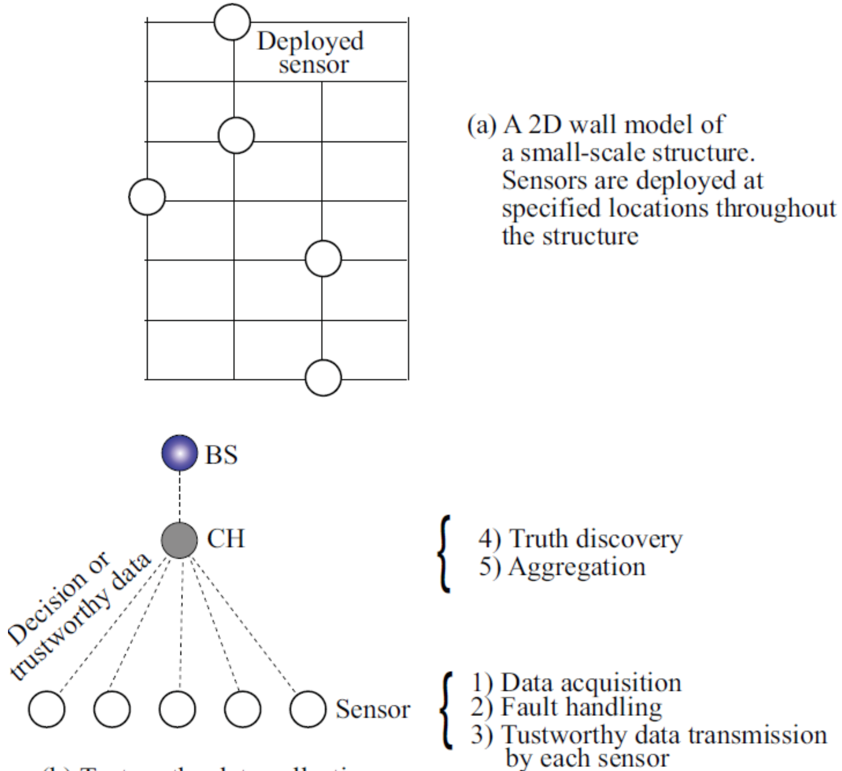
- A Hierarchical WNSS
  - A set of energy-constrained sensors
    - Organized into CHs connecting a BS
    - A CH forwards a final decision of an event or aggregated data to the BS
  - Target application: SHM, smart city applications
  - Event detection and attack/fault detection
    - A minimum communication range, sensors are allowed to share their signals with their neighbors

Our Framework: TPDC

- Monitoring the Health of Civil Structures
  - e.g., building, bridge, aircraft.



(a) Bird view of GNTVT (b) Global structure



(a) A 2D wall model of a small-scale structure. Sensors are deployed at specified locations throughout the structure

4) Truth discovery
5) Aggregation

1) Data acquisition
2) Fault handling
3) Tustworthy data transmission by each sensor

(b) Tustworthy data collection

Our Framework: TPDC

- **Models**
    - **Security attacks**
        - Sensors may produce abnormal signals from security attacks
            - Collusion attack
            - Malicious attack
    - **Sensor faults**
        - **Debonding fault**— sensor node may slightly or completely detach from the host structure, which affects in vibration capturing
        - **Signal fault**—this is caused by precision degradation, breakage, etc., especially in vibration capturing,
        - Faults in offset, bias, or amplification gain factors
        - Noise faults

Trustworthy Data Collection

- Signal Comparison
  - A sensor compares its acquired signal with a reference signal set and get a correlation value
  - The sensor exchanges its correlation value with its neighbors in each sampling instant so that any discrepancy in signals can be discovered
- Signal Correlation Analysis
  - Given signals, MII is a function, defined by the quantify how much the measurement correlation between the sensor signals, and then between sensor nodes

Trustworthy Data Collection

- Decision-making
  - If the MII value of a signal is larger than a given correlation value (calculated from a set of reference signals), the signal is considered to be compromised by an attack or there is a fault
  - Otherwise, the signal is considered trustworthy

Protected Data for Aggregation

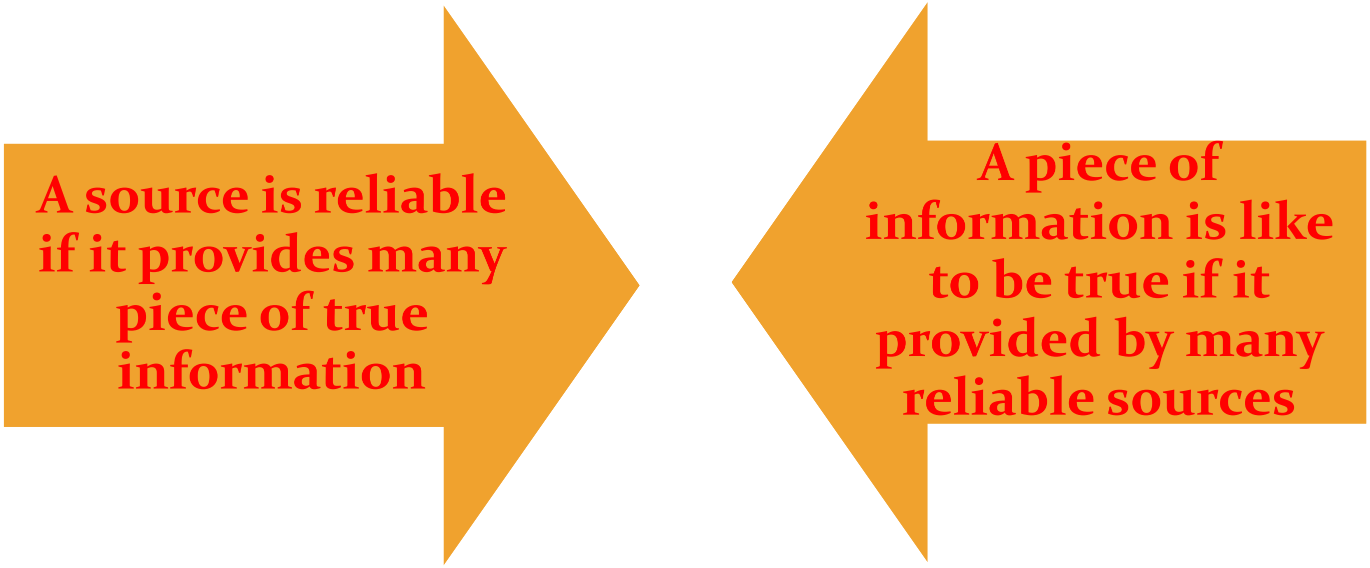
- Once a sensor has **trustworthy data**, it may be altered at the sensor or intermediate sensor before/after transmission
  - i.e., a CH may receive unprotected (or altered) data for aggregation.
- To discover a unreliable sensor or unprotected data at the CH, we use the **truth discovery** approach

Truth Discovery

- **It is used in many domains in order to resolve conflicts with multiple noisy data sensors**
  - **The medias provide billions of pieces of information, unfortunately, not all are reliable, relevant accurate, unbiased, or up-to-date**
  - **Before being used, the information are evaluated for truth.**

Truth Discovery

- Key idea
  - Evaluating 'true information' and its 'source reliability'
- Principle
  - Infer both truth and source reliability from the data



A source is reliable if it provides many piece of true information

A piece of information is like to be true if it provided by many reliable sources

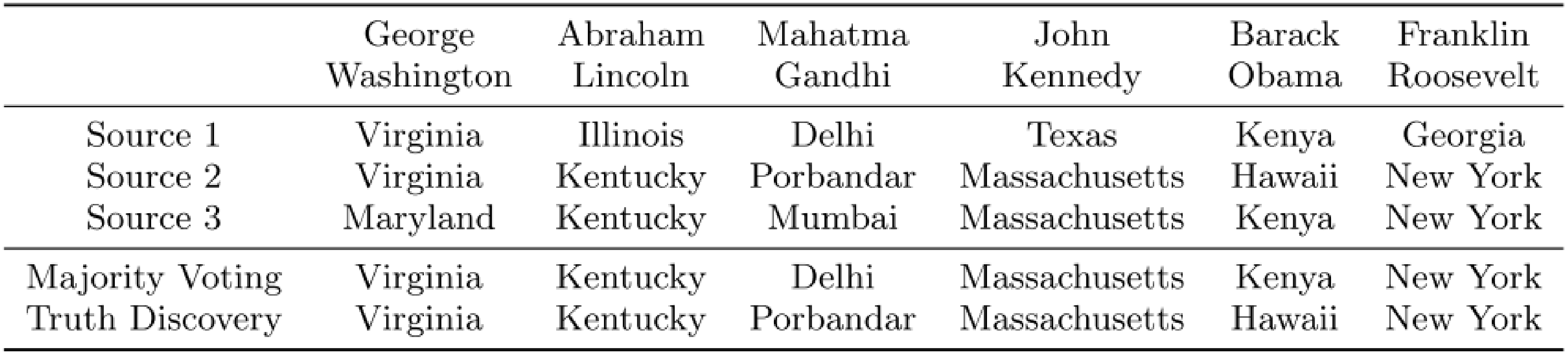
Truth Discovery

- Example 1
  - The top search results returned by Google for the query-- the height of Mount Everest

Source	Height	Vote
Source 1	29.035	5
Source 2	29.002	6
Source 3	29.129	3
Wikipedia	29.029	5

Truth Discovery

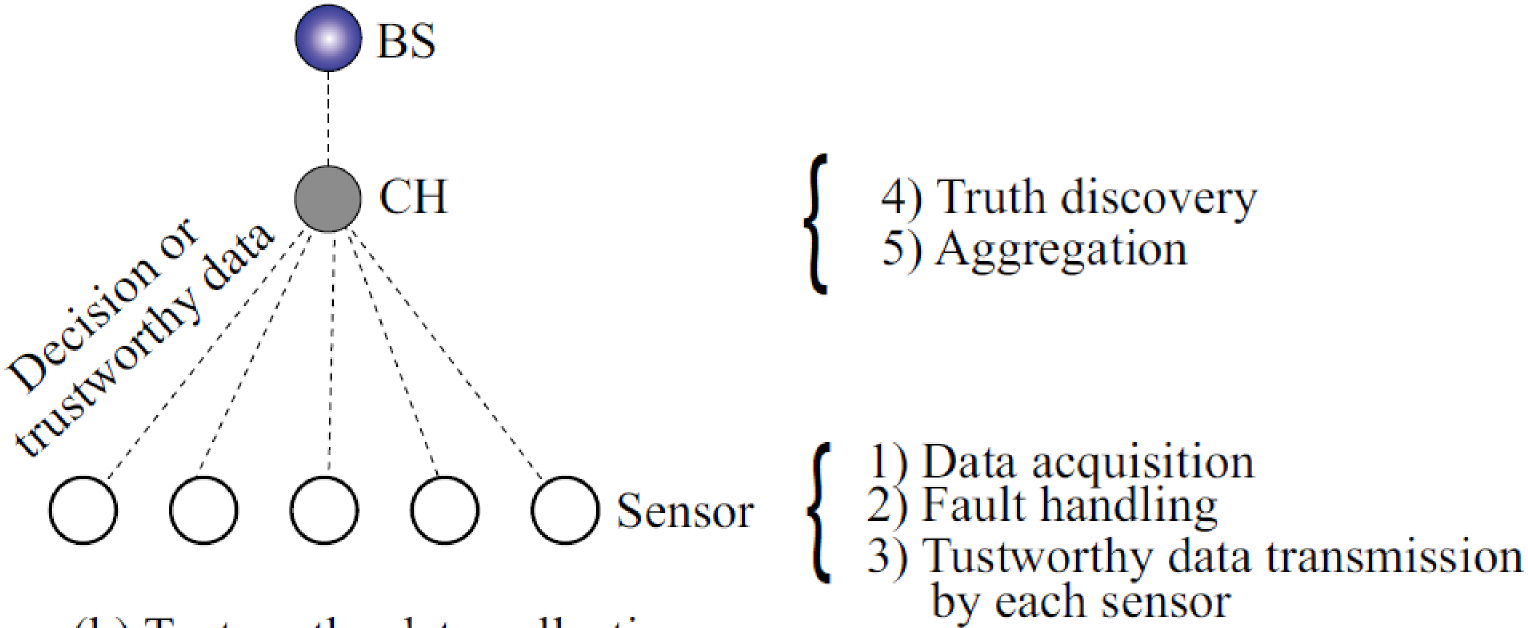
- Example 2
  - The birth place



	George Washington	Abraham Lincoln	Mahatma Gandhi	John Kennedy	Barack Obama	Franklin Roosevelt
Source 1	Virginia	Illinois	Delhi	Texas	Kenya	Georgia
Source 2	Virginia	Kentucky	Porbandar	Massachusetts	Hawaii	New York
Source 3	Maryland	Kentucky	Mumbai	Massachusetts	Kenya	New York
Majority Voting	Virginia	Kentucky	Delhi	Massachusetts	Kenya	New York
Truth Discovery	Virginia	Kentucky	Porbandar	Massachusetts	Hawaii	New York

Conflicting multi-source information

Truth Discovery Instead of Voting Scheme



4) Truth discovery
5) Aggregation

1) Data acquisition
2) Fault handling
3) Tustworthy data transmission by each sensor

(b) Tustworthy data collection

+The amount of truth value provided by sensor
+ the reliability of the sensor

Truth Discovery Instead of Voting Scheme

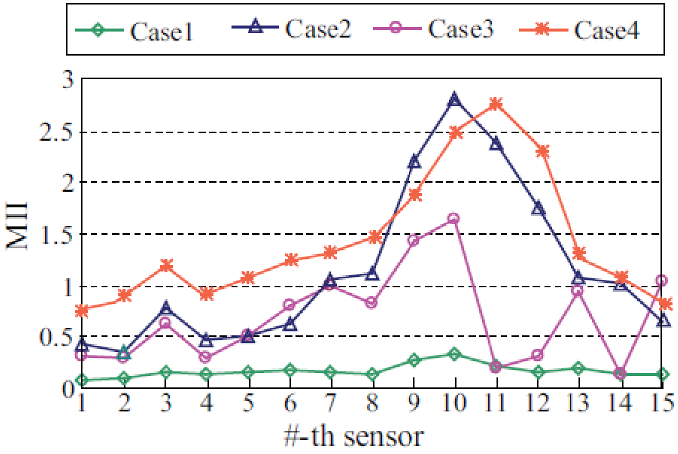
- A sensor's status value is given a high value if the sensor transmitted trustworthy data is close to the estimated ground truths (or given MII values).

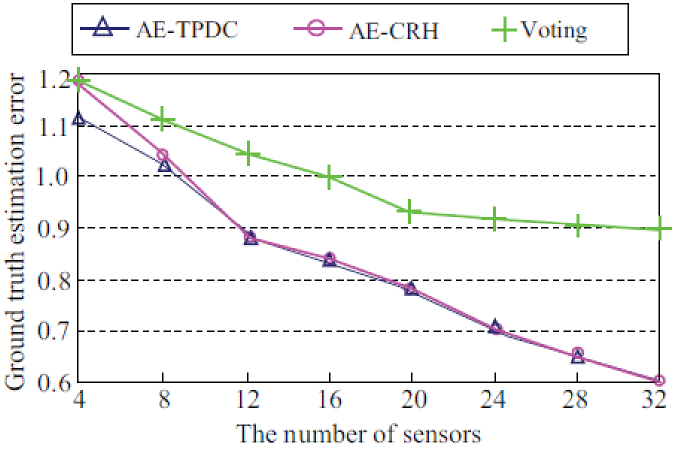
- A truth discovery algorithm
  - Begins with a random guess of ground truths
  - Iteratively conducts status value updates and truth updates until convergence.

Performance Evaluation

- **MATLAB**
- **Real data of from 800 sensors collected from GNTVT**
- **We use the data sets for the 100-sensor case in our simulations.**
- **A SHM toolsuite**
- Attack and fault injection:
  - Added additional noise
  - Change some sensor data

Performance Results





Conclusions and Future Work

- Conclusion:
  - 0/1 based decision-making or fault-tolerant approaches are not suitable for detecting security attacks and faults
- Future work
  - Noise vs. security attack
  - Noise vs. sensor fault
  - Noise vs. event occurrence
  - Security attack vs. sensor fault

Q & A

Contact

Email: mbhuiyan3@fordham.edu, zakirulalam@gmail.com
https://sites.google.com/site/zakirulalam/