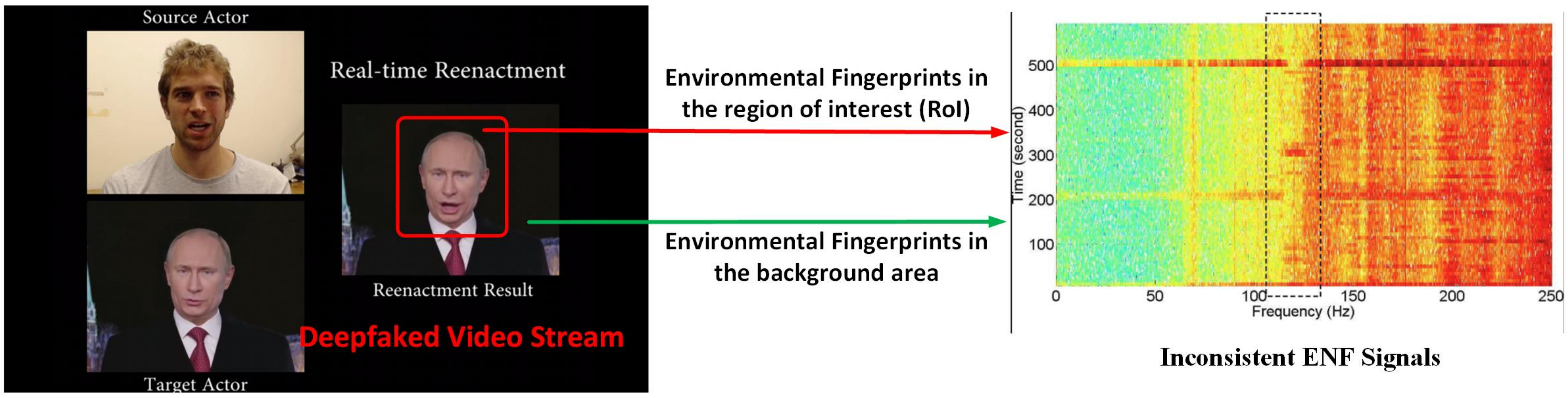
EAGER:
SAVED: Secure Audio and Video Data from Deepfake Attacks Leveraging Environmental Fingerprints

Yu Chen, Binghamton University, ychen@binghamton.edu



Inconsistent ENF Signals

Challenge:

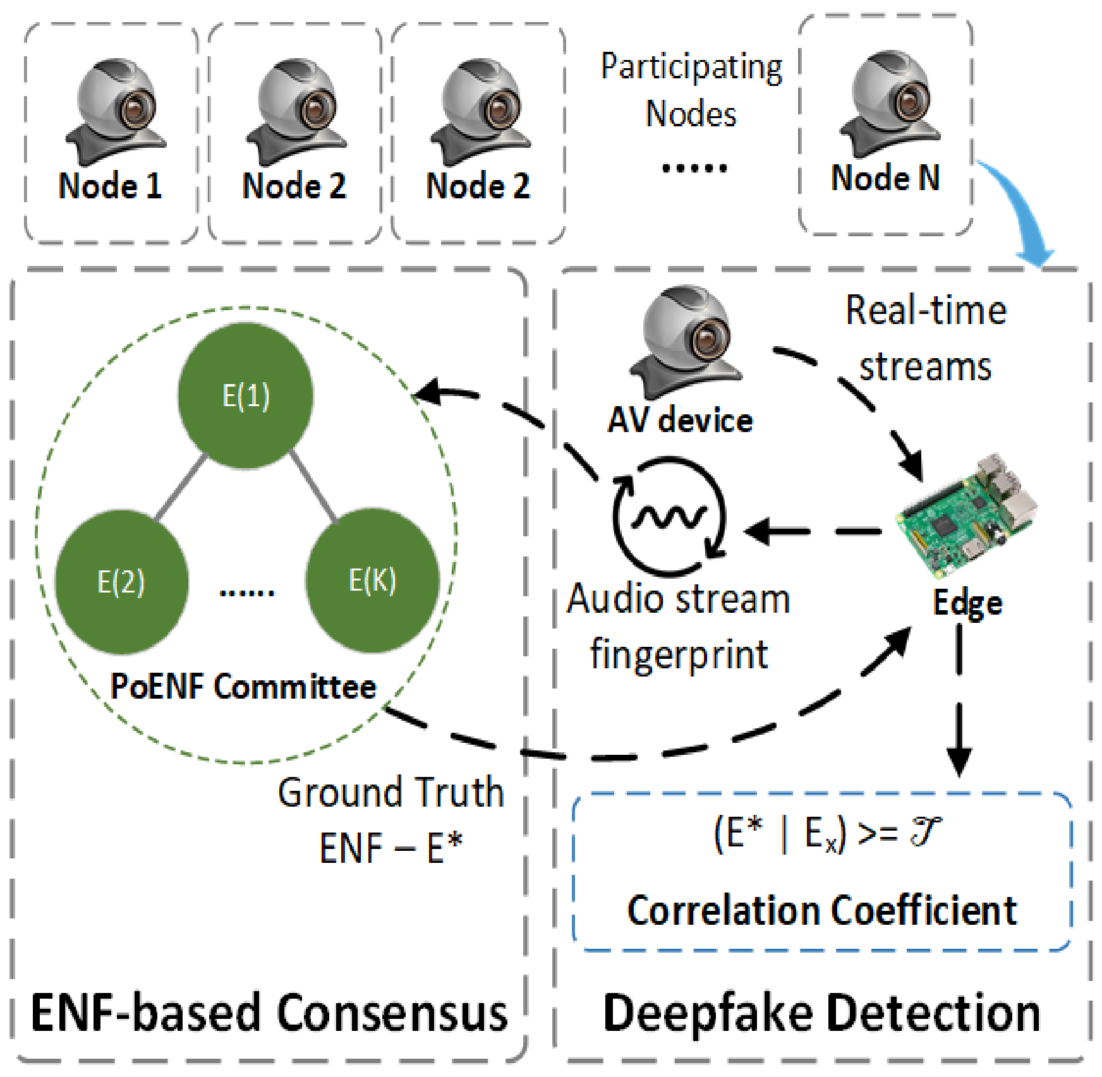
- Visual layer attacks using AI create entirely misclassification
- Deepfaked video, audio or photos are highly disturbing and able to mislead the public, raising further challenges in policy, technology, social, and legal aspects.
- Developing “better” ML-based detector is “Fighting fire with fire”, which leads to an endless AI arms race.

Scientific Impact:

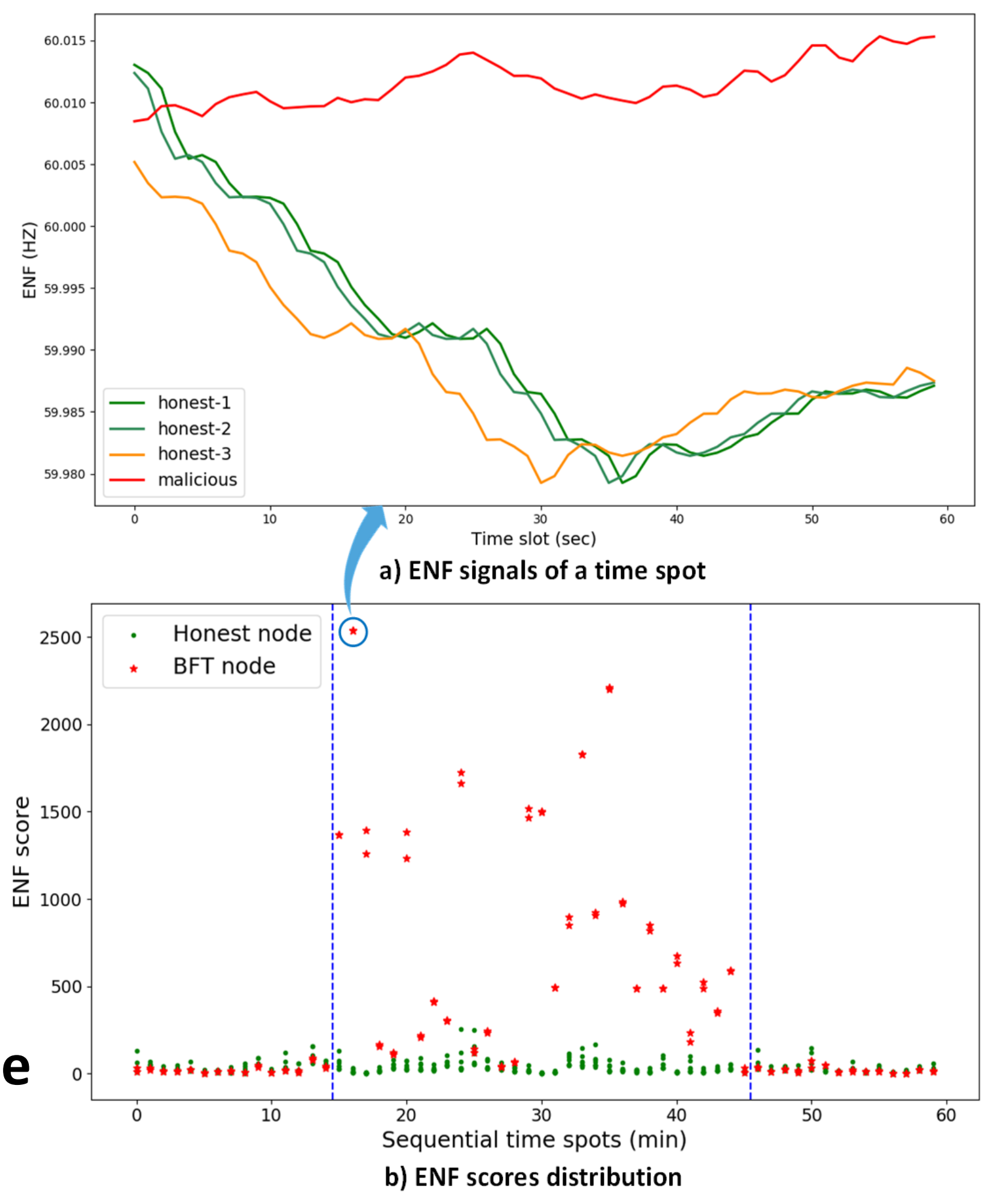
- A deeper understanding of deepfaked audio/video streaming data in terms of embedded, invisible electromagnetic signals.
- A comprehensive exploration of Electrical Network Frequency (ENF) signals as a spatial-temporal correlated environmental fingerprints.
- Robust edge computing paradigm being capable of online, real-time detection.

Decentralized ENF-Consensus based Deepfake Detection

- Determine ground truth ENF from all valid ENF transactions, each transaction is verified for source, timestamp verification and participation
- ENF score is calculated using Krum Aggregation rule to provide Byzantine resilience property



Experimental study: 10 nodes with BFT rate 0.2, attack ranges from 15 – 45 minutes.



a) ENF signals of a time spot

b) ENF scores distribution

Broader Impact:

- This project advances the research frontier of data security toward more reliable, secure, real-time AVS applications.
- It enhances mission-critical, delay-sensitive applications where fake or false inputs will cause disastrous consequences.

Broader Impact:

- Underrepresented and minority students are recruited and trained.
  - ✓ 2 Female
  - ✓ 1 AA student
- Outreach to build safe community collaborating with startups (IFT, 1854Cycling) and public safety practitioners.

Broader Impact and Broader Participation:

- Publications
  - ✓ 3 journal papers
  - ✓ 2 magazine articles
  - ✓ 2 Conference papers
- 2 Invited Talks
- Outreach Activities
  - ✓ Brochure/interview at a high school event

Award ID: CNS-2039342

The 5th NSF Secure and Trustworthy Cyberspace Principal Investigator Meeting (2022 SaTC PI Meeting)
June 1-2, 2022 | Arlington, Virginia