Detecting Compromised Edge Smart Cameras using Lightweight Environmental Fingerprint Consensus

Deeraj Nagothu, Ronghua Xu, Yu Chen
Binghamton University
Binghamton, New York, USA
{dnagoth1,rxu22,ychen}@binghamton.edu

Erik Blasch, Alexander Aved
The U.S. Air Force Research Laboratory
Rome, New York, USA
{erik.blasch,alexander.aved}@us.af.mil

ABSTRACT

Rapid advances in the Internet of Video Things (IoVT) deployment in modern smart cities has enabled secure infrastructures with minimal human intervention. However, attacks on audio-video inputs affect the reliability of large-scale multimedia surveillance systems as attackers are able to manipulate the perception of live events. For example, Deepfake audio/video attacks and frame duplication attacks can cause significant security breaches. This paper proposes a Lightweight Environmental Fingerprint Consensus based detection of compromised smart cameras in edge surveillance systems (LEFC). LEFC is a partial decentralized authentication mechanism that leverages Electrical Network Frequency (ENF) as an environmental fingerprint and distributed ledger technology (DLT). An ENF signal carries randomly fluctuating spatio-temporal signatures, which enable digital media authentication. With the proposed DLT consensus mechanism named Proof-of-ENF (PoENF) as a backbone, LEFC can estimate and authenticate the media recording and detect byzantine nodes controlled by the perpetrator. The experimental evaluation shows feasibility and effectiveness of proposed LEFC scheme under a distributed byzantine network environment.

CCS CONCEPTS

• **Security and privacy → Authentication**.

KEYWORDS

Deepfake Detection, Environmental Fingerprint, Electrical Network Frequency (ENF) Signals, Proof-of-ENF (PoENF) Consensus

ACM Reference Format:
Deeraj Nagothu, Ronghua Xu, Yu Chen and Erik Blasch, Alexander Aved. 2021. Detecting Compromised Edge Smart Cameras using Lightweight Environmental Fingerprint Consensus. In *BlockSys '21: the Third ACM International Workshop on Blockchain-enabled Networked Sensor System, November 2021, Coimbra, Portugal.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3485730.3493684

BlockSys '21, November 2021, Coimbra, Portugal

ACM ISBN 978-1-4503-9097-2/21/11…$15.00
https://doi.org/10.1145/3485730.3493684

1 INTRODUCTION

The growing smart city technological adaptation demands many security and privacy practices put in place. With advancing large-scale infrastructures, there is a constant demand to maintain a secure channel for monitoring and decision-making. The exponential growth in the deployed number of intelligent Internet of Video Things (IoVT) devices for smart surveillance makes online video stream processing the most researched topic in Smart Cities. The edge computers have allowed a more conventional way to process the incoming audio-video streams for a broad spectrum of applications like behavioral recognition and suspicious event identification [21]. Advanced applications in smart surveillance framework lead to targeted networks, and visual layer attacks focus on compromising the device integrity. Visual layer attacks open new dimensions for perpetrators to manipulate the real-time scene without triggering alarms like replay attacks and frame modification attacks [18]. Recently, with advanced machine learning models in edge computers, DeepFakes [26] have become a common attack vector with minimal computing resources.

Generative deep learning models have become an enabling technology for manipulating the event perception through manipulated multimedia injections. Audio manipulations can be performed with pre-existing source recordings of the targeted individual. A person's voice can be generated in real-time using text to speech software with little training time [19]. Video manipulations involve real-time face swapping or facial reenactment attacks, where a source facial expressions can be projected over a targeted face to present a false perception of an individual's actions. Such attacks are quite popular for celebrities and politicians, and with the growth of edge computing resources, similar attacks can be launched in the IoVT environment and manipulate real-time streams framing an individual for false actions. These visual layer attacks have become a potent security risk for the IoVT environment, and detection of such attacks with minimal downtime is a high priority. This work proposes an environmental fingerprint-based detection and attack localization technique for secure online authentication of audio and video channels using a distributed consensus mechanism.

This paper proposes a Lightweight Environmental Fingerprint Consensus (LEFC) based detection of compromised smart cameras in edge surveillance systems. Electrical Network Frequency (ENF) is adopted as an environmental fingerprint due to its presence in audio-video recordings in an indoor environment where smart surveillance network devices are typically deployed. ENF provides ground truth evidence on multimedia manipulations by cross-referencing the source ENF signal, similar throughout the power grid interconnect. Any irregularities can be used to identify

potential manipulations. To enable a distributed approach to detect false media injection attacks, we adopted a distributed ledger technology (DLT). The DLT allows the network devices to provide security and trust without relying on a centralized architecture and third-party intervention. The DLT technology demonstrates a great potential in revolutionizing the edge computing paradigm [25]. Equipped with Spatio-temporal sensitive ENF signal, we propose a lightweight Proof-of-ENF (PoENF) consensus algorithm to provide efficiency and security to the participating smart surveillance nodes. In each PoENF consensus round, the participating nodes broadcast the estimated ENF from the recorded media (audio/video). The consensus round selects the ground truth ENF among the broadcasted signals. Assuming no more than f nodes controlled by the perpetrator in a network with $n \geq 2f + 3$ nodes, all honest nodes can make agreement on valid ENF and false ENF are identified using a measure of similarity with correlation coefficient.

This paper makes the following contributions.

- A distributed secure-by-design visual layer attack detection system is introduced with key components and workflow;
- An IoVT system is deployed with varying byzantine node rates, and system resilience is evaluated with an ENF score;
- A distributed consensus mechanism with sliding windows detects and localizes an attack; and
- Different attack vectors like false frame injection and DeepFakes representing byzantine nodes are tested, and the proposed LEFC system can detect with a high accuracy and display tolerance to network delays.

The remainder of the paper is organized as follows. Section 2 discusses the key components of LEFC, including ENF fingerprinting and blockchain consensus protocols. Section 3 introduces the LEFC system design rationale and basis for attack detection and localization. Section 4 presents the numerical analysis of LEFC system with varying byzantine nodes and discusses attack detection parameters and related discussions. Finally, a conclusive summary is provided in Section 5.

2 BACKGROUND AND RELATED WORK

2.1 ENF as an Environmental Fingerprint

Electrical Network Frequency (ENF) is a power network frequency provided in the form of infrastructure utility. The power supply frequency is 60 Hz in the US and 50 Hz in other Asian and European countries. The fluctuations in the power supply frequency are due to the power supply demand and the load balancing mechanism, and these fluctuations are referred to as the ENF signal. ENF fluctuations are similar throughout a power grid interconnect and are random due to the unpredictability of the power requirements. The frequency fluctuations in a power grid carry small latency due to the distance propagation and can tag a recording with its geographical location [10]. To represent the randomness of the ENF data, we used the correlation coefficient to measure the similarity between two signals. ENF data is collected for approximately 200 hours, and the resulting heatmap of the correlation coefficient is shown in Figure 1. The diagonal represents the highest correlation as expected for signals with similar fluctuations, whereas similar fluctuations are not repeated.

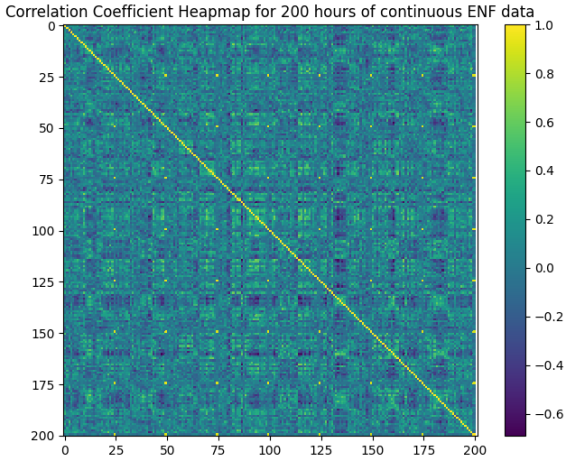

Figure 1: Heatmap of the Correlation coefficient of ENF data for 200 Hours of Power Recording

2.2 ENF in Multimedia Recordings

In multimedia recordings like audio and video, the ENF is embedded through multiple power-dependent sources. In the case of audio, the ENF is embedded through electromagnetic induction when the recorder is directly connected to the power grid [12]. For battery-powered devices, the ENF is embedded through the background hum [8]. The audio recordings consist of a higher sampling frequency allowing for a robust estimation of the ENF signal.

The source of ENF in video recordings depends on the type of imaging sensor used in the video recorders. The illumination frequency, i.e., 120 Hz, which is twice the nominal frequency, projects the ENF fluctuations in the power grid, and the recorded photons by the imaging sensor capture the fluctuations. The two common sensors used for general-purpose video applications are complementary metal-oxide-semiconductor (CMOS) and charge-coupled device (CCD) sensors. For each sensor, the photon capturing mechanism varies with the shutter mechanism used. Video recorders with CCD sensors deploy a global shutter mechanism. The sensor is exposed to light for a time instant, resulting in capturing samples equal to the number of frames per second (FPS) of the camera. With the low sampling rate of the CCD sensor, the estimation of 120 Hz illumination frequency from the samples captured is not possible. It thus relies on aliasing frequency for ENF estimation [11].

For video recorders with CMOS sensors, the rolling shutter mechanism captures the photon sample in each row, allowing for a high sampling rate compared to the CCD sensors. Depending on the manufacturer, each camera model has its idle period where few samples are not captured for each frame. Using the Filter-Bank model, the idle period along with the ENF can be estimated [24]. CMOS is the most commonly deployed camera sensor among both imaging sensors due to its cost efficiency and broad applicability.

The nominal frequency ENF signal is estimated using spectrogram estimation techniques with the samples captured from the multimedia recordings. ENF is computed from each harmonic and combined using their respective signal to noise (SNR) ratio as the weights for robust estimation. A detailed discussion of ENF estimation techniques from audio and video recordings are presented in our previous work [16, 17].

2.3 Blockchain and Consensus Protocols

Blockchain initially was implemented as an enabling technology of Bitcoin [20], which aims to provide a cryptocurrency by recording and verifying commercial transactions among trustless entities in a decentralized manner. Thanks to the decentralized Peer-to-Peer (P2P) network architecture and cryptographic security mechanisms, miners and validators in a blockchain network utilizes a consensus protocol to guarantee auditability and integrity of data on the distributed ledger instead of relying on any third party trust authority [29]. As one of the most important component in a blockchain system, *consensus* states that the processes have to reach agreement on of a value (called decision value) under fault-tolerant distributed network environment. Given diverse consensus protocols, blockchains can be categorized as permissionless or permissioned chains.

As the first practical BFT consensus, PBFT [7] uses the State Machine Replication (SMR) scheme to address the Byzantine General Problem [14] in distributed networks. It has been widely adopted as a basic consensus solution in the *permissioned* blockchains, like Hyperledger Fabric [3]. Given assumption that at most $\lfloor \frac{n-1}{3} \rfloor$ out of total of n nodes in a blockchain network are Byzantine faults, the PBFT algorithm can guarantee both liveness and safety in synchronous permissioned network environments. PBFT consensus demonstrates deterministic finality on distributed ledger with energy efficiency and high transactions throughput. However, it inevitably incurs high latency and communication overhead to synchronously execute a consensus protocol in large scale networks.

To jointly address the critical issues, such as pseudonymity and scalability in a asynchronous open-access network environment, the Nakamoto protocol is widely used cryptocurrency blockchains, like Bitcoin [20] and Ethereum [6]. The Nakamoto protocol relies on a Proof-of-Work consensus algorithm, which uses a computation intensive cryptographic hash value searching game to reward the winner of block generation. For a Proof of Work (PoW) consensus network, the probability of mining valid blocks is proportion to computing power percentage that a participant can have compared with the whole network. Given assumption that an adversary cannot control majority (51% attack) of computing resources of the consensus network, Nakamoto protocol can guarantee security and scalability in a *permissionless* blockchain network.

To improve the performance and resource efficiency in PoW, a number of alternative Proof of X-concept (PoX) schemes have been proposed. To reduce unnecessary wastage of computational resources in PoW, Permacoin [15] adopts a Proofs of retrievability (PoRs) [13], which requires miners to invest their storage capacity rather than solo computational power. Like Permacoin, a Proof-of-Useful-Work (PoUW) [31] consensus protocol relies on the partially decentralized trust model inherent in Intel Software Guard Extensions (SGX) to achieve security and resource efficiency. The above mentioned consensus algorithms requires large storage or specific hardware features, however, they are not suitable for heterogeneous IoVT devices with limited computation and storage capability.

Unlike PoW and its variants, our PoENF consensus neither requires high demand of computation and storage resource for mining, nor depends on security guarantees supported by a trusted hardware platform [30]. Thus, using an ENF-based environmental fingerprint consensus is a promising solution to detect compromised smart cameras connected to the power grid.

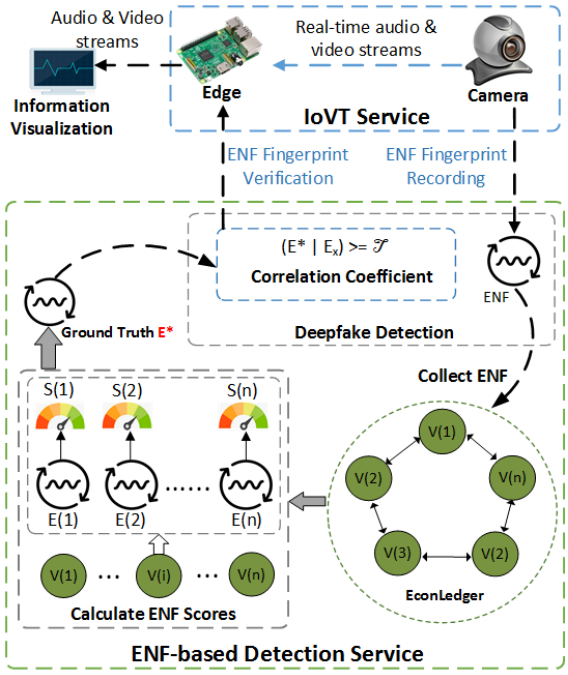

Figure 2: LEFC system architecture consisting of a upper-level IoVT service layer and a secure-by-design deepfake detection service layer.

3 LEFC SYSTEM ARCHITECTURE DESIGN

Our deepfake detection mechanism leverages ENF environmental fingerprint and blockchain technology to achieve a partial decentralized audio and video (A&V) authentication for smart surveillance systems. We assume a small scale IoVT network including about 100 nodes, like cameras and edge servers, and all devices are connected to the same regional power grid. Figure 2 represents the LEFC system architecture consisting of: 1) IoVT service layer that provides multimedia steams for smart surveillance systems; and 2) ENF-based detection service layer that support a security-by-design networking foundation for deepfake detection.

The IoVT service layer is deployed at the network of edge to enable A&V analytic tasks and information visualization for upper-level smart surveillance applications. All users and devices must be registered to join the permissioned network. Therefore, basic security primitives are guaranteed, such as public key infrastructure (PKI), data integrity verification [22], identity authentication [28] and access control [27]. Cameras send real-time A&V streams to on-site/near-site edge devices for lower level analytic tasks, like object-detection and situational contextual features extraction [9]. Then, edge devices transferred raw multimedia data along with extracted contextual information to information visualization unit, which provides on-line or off-line A&V recordings for authorized users. The following sub-sections provide details of components and workflow in ENF-based detection service.

3.1 PoENF Consensus Mechanism

The PoENF mechanism is mainly responsible to provide verifiable and traceable ENF data and select ground truth ENF benchmarks for further deepfake detections.

3.1.1 ENF Data Recording. Our solution adopts EconLedger as a "trust" and partial decentralized security infrastructure for cross-devices networking IoVT systems at the edge [30]. In a EconLedger network, a random PoENF consensus committee election strategy chooses a subset of valid nodes as committee members (validators). Then, validators of the current consensus committee collect ENF proofs from each other, and the cooperatively execute a lightweight PoENF consensus protocol to record all ENF data on a private distributed ledger that can be accessed by nodes within the network. Thanks to PoENF consensus protocol and immutable distributed ledger, deepfake detection can identify compromised cameras by auditing historical ENF data of nodes without relying on a centralized third party trust authority, which provides ground truth ENF benchmarks directly extracted from power grid [17].

3.1.2 ENF Scores Calculation. As ENF data from individual nodes are stored on distributed ledger, a validator v_i in committee D can extracts ENF proofs submitted by other validators in a specific time spot. Each ENF proof is a vector $E = \{e_1, e_2, ..., e_d\}$, where $e_i \in \mathbb{R}$ is the ENF frequency sample and d is the vector size defined by the sliding window of ENF estimation. Thus, each validator maintains a global view of collected ENF proof vectors $G = \{E_1, E_2, ..., E_K\}$, and calculates ENF scores block as Fig. 2 shows. For a $E_i \in G$, an ENF score S_i can be calculated by using sum of its relative distances between other ENF vectors, that is computed with the Euclidean norm. Finally, a ENF vector that deviates the least from all ENF vectors has the minimal ENF score, and it will be selected as a benchmark E^*.

However, an adversary can compromise validators, and a byzantine node sends poisoned E_b that is too far away from valid ENF vectors. As a result, they can force the PoENF algorithm to choose any arbitrary ENF vector. Our ENF score calculation algorithm adopts *Krum aggregation rule* [5] to guarantee an (α, f)-Byzantine resilience property, where $0 \leq \alpha \leq \pi/2$ is any angular value and f is a non-negative integer smaller than or equal to n. We require that each honest validator only maintain a $G = \{E_1, E_2, ..., E_n\}$ including $n \geq 2f + 3$ observed ENF proof vectors from PoENF committee members, and only at most f are sent by Byzantine nodes. For any $i \neq j$, let $i \rightarrow j$ denote the fact the that E_j belongs to the $n - f - 2$ closest vectors to E_i. Then we define the ENF score for v_i:

$$s(v_i) = \sum_{i \rightarrow j} \|E_i - E_j\|^2. \tag{1}$$

Therefore, each validator can use Eq. (1) to calculate ENF scores $(s(1), ..., s(n))$ associated with validators v_1 to v_n, respectively.

3.1.3 Ground Truth ENF Selection. By executing ENF score calculation defined in Eq.1 for each $E_i \in G$, each validator can get a global ENF score list $S = \{s(1), s(2), ...S(n)\}$. Then, each validator can sort S to select the minimum ENF score as follows:

$$s^* = \min_{i \in \{1,...,n\}} (s(i)). \tag{2}$$

Finally, PoENF consensus requires that a honest validator always uses s^* as the ground truth ENF benchmark in current round. Thus, all honest validators can make an agreement on s^* if an adversary can control at most f nodes in PoENF committee. The ENF vector E_i that satisfies the condition $s(i) \leq s^*$ will be selected as a benchmark ENF E^* for deepfake detection.

3.2 ENF-based Multimedia Attack Detection

Each node generates a local ENF signal and the consensus committee agrees on a ground truth ENF for the current round. To measure the similarity between the ground truth ENF and the locally generated ENF in each node, we utilize cross-correlation coefficient (ρ). The correlation varies in the range $[-1, 1]$, where 1 represents highest measure of similarity. Based on the experimental evaluations, a threshold of 0.8 is used to detect the drop in correlation and to allow each consensus around to tolerate a few seconds delay in ENF braodcast. Equation 3 represents the correlation coefficient used to detect ENF signal variations in each node (E_l).

$$\rho(l) = \frac{\sum_{t=1}^{d} [E_l(t) - \mu_{E_l}][E^*(t-l) - \mu_{E^*}]}{var(E_l) * var(E^*)} \tag{3}$$

where l is the lag between the two signals, d is the vector size, μ is the mean of the signal, and var is the variance of the signal.

4 EXPERIMENTAL RESULTS

4.1 Experimental Setup

A proof-of-concept prototype is implemented in Python to verify feasibility of the proposed solution. The LEFC prototype emulates a small scale video surveillance system based on a local area network (LAN). A Dell Optiplex 7010 simulates a monitor server that manages permissioned network and aggregates audio/video streams from cameras installed at different location. While 20 Raspberry Pi-4 (Rpi) devices act as validators to collect ENF proofs and execute PoENF consensus algorithm.

To generate compromised multimedia recordings, *Descript* platform [1] is used to synthesize text-to-speak audio deepfakes, and *DeepFaceLive* is used [23] to create video deepfakes based on target users' faces along with frame replay forgeries [18]. We adopt a light micro-framework called Flask [2] to develop networking and web service Application Programming Interface (API) for deepfake detection services. All cryptographic functions, like hash functions and digital signature, are developed on the foundation of the standard python lib *cryptography* [4].

4.2 Numerical Results

To evaluate the performance of the running prototype under an IoVT-based edge network environment, a set of experiments is conducted by executing multiple complete round of PoENF consensus among committee members. We calculate an average of results for PoENF latency and use statistical analysis for PoENF effectiveness. The computation cost by message encryption and decryption are not considered during the test.

4.2.1 ENF Sliding Window Size. Each consensus round consists of generating ENF signal data from a certain period of multimedia

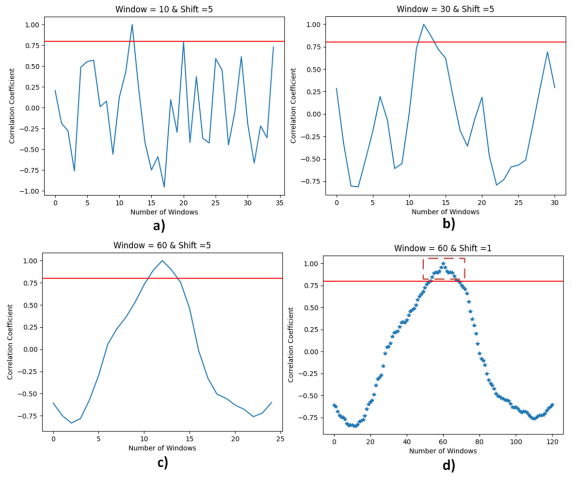

Figure 3: ENF analysis on sliding window and margin.

Table 1: PoENF latency with varying committee size.

No. of Validators	10	20	50	100	200
Latency (sec)	0.02	0.08	0.5	1.9	7.7

recording time. We adopt the sliding window technique to compensate for the signal capture and estimation to enable real-time detection. Each sliding window is treated as a single ENF consensus round, and the shift in the window can reduce the computational resources required for ENF signal estimation. Figure 3 represents the comparison of two ENF signals using different sliding window sizes and shift sizes, where the signal similarity is represented using a correlation coefficient. With a smaller window size like 10 seconds in (3a) and 30 seconds in (3b), the signal length is smaller, and the fluctuations in correlation are higher. Whereas, with 60 seconds sliding window in (3c), the correlation follows the bell curve where the peak represents the highest signal similarity and no lag. A sliding window size of 60 seconds is used to reduce the false positives in consensus rounds, and a shift size of 5 seconds is used to compensate for ENF estimation times. The sliding window can tolerate ±6 seconds of delay as shown in (3d) to account for signal broadcast delay caused by network congestion. The threshold of 0.8 detects the drop in correlation and reduces the false positives.

4.2.2 PoENF Consensus Latency. Given the above mentioned analysis on sliding windows used in ENF signal extraction, the size of an ENF vector d used by following test results is 60. Then LEFC evaluates time latency for PoENF when committee size K changes. Table 1 presents the cumulative time taken for a round of PoENF consensus including ENF proof collection, ENF score calculation and ground truth ENF selection. Time latency is dominated by the ENF score calculation, which has the complexity of $O(K^2d)$. Thus, processing time dramatically increases as the number of validators in committee scales up. Assuming a small IoVT network including less than 100 nodes, the time latency of PoENF consensus procedures in LEFC is no more than 2 seconds. Therefore, it can satisfy required sliding window and margin in ENF extraction.

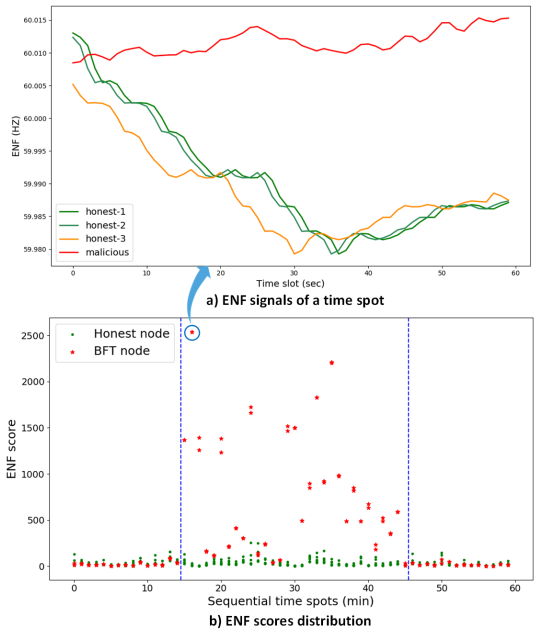

Figure 4: ENF scores distribution with sequential time spots. (total nodes: 10, BFT rate: 0.2, attack range: 15-45 mins)

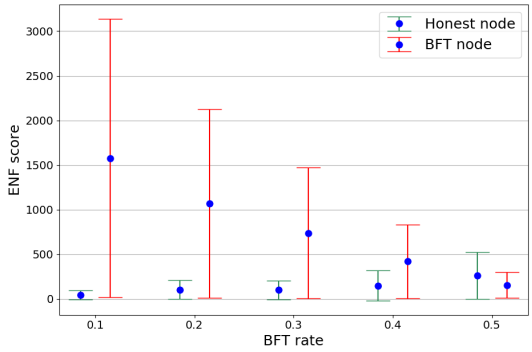

Figure 5: Statistical comparison between honest and BFT nodes. (total node: 20, BFT rate: 0.1-0.5)

4.2.3 ENF-based Detection Effectiveness. We also verify how effective ENF-based detection can identify compromised cameras under deepfake attack scenarios. Figure 4 shows ENF scores distribution based on a synchronous ENF recording period (60 minutes). There are total 10 nodes that participant PoENF consensus. Two byzantine nodes start deekfake attacks at the 15 minute mark, and such a attack ends at 45 minute mark. In a certain sliding window period of attack points, ENF proof vectors from malicious nodes have distinct trends compared with these generated by honest nodes, as Fig. 4a shows. As a result, ENF scores associated with BFT nodes are much larger than ENF scores of BFT nodes given an attack range, as Fig. 4b shows points between blue lines. Because a ground truth E^* has the minimal ENF score at a time spot, it can be used to classify malicious nodes according to ENF scores distribution.

To evaluate how malicious nodes' ratio influences detection results, we make comparative experiments on ENF scores when BFT rate changes given a fixed committee size. Figure 5 shows ENF score statistics between honest nodes and BFT nodes as BFT rate

increases from 10% to 50%. Each bar indicates standard deviation (std) with a mean represented by a blue dot. The difference of means by honest and BFT nodes decreases as more nodes are compromised. Given assumption that $n \geq 2f+3$ in PoENF consensus, a committee with size $n = 20$ can tolerate up to 8 BFT nodes, which means that BFT rate can be no more than 0.4. As Fig. 5 shows, both mean and std by BFT nodes are smaller than honest nodes when BFT rate is 0.5. Thus, ENF-based deepfake detection will fail owing to fact that an adversary has controlled committee to disturb PoENF consensus process. However, our ENF-based detection can still identify deepfake attack nodes if honest nodes are no less than 60%.

4.3 Discussions

ENF can be embedded in audio/video in both power grid-connected and battery-powered devices in indoor environments. But for outdoor environments which lack artificially powered light with illumination frequency, ENF in video recordings is absent. However, most smart cameras are deployed in an indoor environment with power grid-connected, our LEFC is suitable to detect compromised devices based on embedded ENF recordings. Moreover, PoENF consensus relies on a small committee to improve efficiency but at the cost of partial decentralization. A scalable random committee strategy is promising to enhance security and scalability. We leave above open questions in future work.

5 CONCLUSIONS

By integrating ENF as environmental fingerprint with DLT, this paper proposes LEFC to detect compromised or malfunctioned smart cameras deployed in edge surveillance systems. Thanks to the unique spatio-temporal fluctuation property of ENF signals, experimental study based on a proof-of-concept prototype demonstrates that proposed LEFC is effective to identify deepfake attack under a distributed byzantine network environment. However, there are still open issues to solve before bringing LEFC into practice. Our ongoing efforts include validating LEFC in a real-world smart surveillance system and evaluating overall performance, robustness and security given different attack scenarios.

ACKNOWLEDGMENTS

This work is supported by the U.S. National Science Foundation (NSF) via grant CNS-2039342 and the U.S. Air Force Office of Scientific Research (AFOSR) Dynamic Data and Information Processing Program (DDIP) via grant FA9550-21-1-0229. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of the U. S. Air Force.

REFERENCES

[1] [n. d.]. Descript | Create podcasts, videos, and transcripts. https://www.descript.com/ accessed: Oct. 18, 2021..
[2] [n. d.]. Flask: A Pyhon Microframework. http://flask.pocoo.org/. accessed: Oct. 18, 2021..
[3] [n. d.]. Hyperledger Fabric. https://www.hyperledger.org/use/fabric. accessed: Oct. 14, 2021..
[4] [n. d.]. pyca/cryptography documentation. https://cryptography.io/. accessed: Oct. 18, 2021..
[5] Peva Blanchard, Rachid Guerraoui, Julien Stainer, et al. 2017. Machine learning with adversaries: Byzantine tolerant gradient descent. In *Advances in Neural Information Processing Systems*. 119–129.
[6] Vitalik Buterin et al. 2014. A next-generation smart contract and decentralized application platform. *white paper* (2014).
[7] Miguel Castro, Barbara Liskov, et al. 1999. Practical Byzantine fault tolerance. In *OSDI*, Vol. 99. 173–186.
[8] Jidong Chai, Fan Liu, Zhiyong Yuan, Richard W Conners, and Yilu Liu. 2013. Source of ENF in battery-powered digital recordings. In *Audio Engineering Society Convention 135*. Audio Engineering Society.
[9] Ning Chen, Yu Chen, Erik Blasch, Haibin Ling, Yang You, and Xinyue Ye. 2017. Enabling smart urban surveillance at the edge. In *2017 IEEE International Conference on Smart Cloud (SmartCloud)*. IEEE, 109–119.
[10] Ravi Garg, Adi Hajj-Ahmad, and Min Wu. 2021. Feasibility Study on Intra-Grid Location Estimation Using Power ENF Signals. arXiv:2105.00668 [eess.SP]
[11] Ravi Garg, Avinash L Varna, Adi Hajj-Ahmad, and Min Wu. 2013. "Seeing" ENF: power-signature-based timestamp for digital multimedia via optical sensing and signal processing. *IEEE Transactions on Information Forensics and Security* 8, 9 (2013), 1417–1432.
[12] Catalin Grigoras. 2005. Digital audio recording analysis–the electric network frequency criterion. *International Journal of Speech Language and the Law* 12, 1 (2005), 63–76.
[13] Ari Juels and Burton S Kaliski Jr. 2007. PORs: Proofs of retrievability for large files. In *Proceedings of the 14th ACM conference on Computer and communications security*. 584–597.
[14] Leslie Lamport, Robert Shostak, and Marshall Pease. 1982. The Byzantine generals problem. *ACM Transactions on Programming Languages and Systems (TOPLAS)* 4, 3 (1982), 382–401.
[15] Andrew Miller, Ari Juels, Elaine Shi, Bryan Parno, and Jonathan Katz. 2014. Permacoin: Repurposing bitcoin work for data preservation. In *2014 IEEE Symposium on Security and Privacy*. IEEE, 475–490.
[16] Deeraj Nagothu, Yu Chen, Alexander Aved, and Erik Blasch. 2021. Authenticating video feeds using electric network frequency estimation at the edge. *EAI Endorsed Transactions on Security and Safety* 7, 24 (2021), e4.
[17] Deeraj Nagothu, Yu Chen, Erik Blasch, Alexander Aved, and Sencun Zhu. 2019. Detecting Malicious False Frame Injection Attacks on Surveillance Systems at the Edge Using Electrical Network Frequency Signals. *Sensors* 19, 11 (2019), 2424.
[18] Deeraj Nagothu, Jacob Schwell, Yu Chen, Erik Blasch, and Sencun Zhu. 2019. A study on smart online frame forging attacks against video surveillance system. In *Sensors and Systems for Space Applications XII*, Vol. 11017. International Society for Optics and Photonics, 110170L.
[19] Deeraj Nagothu, Ronghua Xu, Yu Chen, Erik Blasch, and Alexander Aved. 2021. DeFake: Decentralized ENF-Consensus Based DeepFake Detection in Video Conferencing. In *IEEE 23rd International Workshop on Multimedia Signal Processing*. Tampere, Finland.
[20] Satoshi Nakamoto. 2019. *Bitcoin: A peer-to-peer electronic cash system.* Technical Report. Manubot.
[21] Seyed Yahya Nikouei, Yu Chen, Alexander Aved, Erik Blasch, and Timothy R Faughnan. 2019. I-safe: Instant suspicious activity identification at the edge using fuzzy decision making. In *Proceedings of the 4th ACM/IEEE Symposium on Edge Computing*. 101–112.
[22] Seyed Yahya Nikouei, Ronghua Xu, Deeraj Nagothu, Yu Chen, Alexander Aved, and Erik Blasch. 2018. Real-time index authentication for event-oriented surveillance video query using blockchain. In *2018 IEEE International Smart Cities Conference (ISC2)*. IEEE, 1–8.
[23] Ivan Perov, Daiheng Gao, Nikolay Chervoniy, Kunlin Liu, Sugasa Marangonda, Chris Umé, Mr Dpfks, Carl Shift Facenheim, Luis RP, Jian Jiang, et al. 2020. Deepfacelab: A simple, flexible and extensible face swapping framework. *arXiv preprint arXiv:2005.05535* (2020).
[24] Hui Su, Adi Hajj-Ahmad, Ravi Garg, and Min Wu. 2014. Exploiting rolling shutter for ENF signal extraction from video. In *Image Processing (ICIP), 2014 IEEE International Conference on*. Citeseer, 5367–5371.
[25] Melanie Swan. 2015. *Blockchain: Blueprint for a new economy.* " O'Reilly Media, Inc.".
[26] Luisa Verdoliva. 2020. Media forensics and deepfakes: an overview. *IEEE Journal of Selected Topics in Signal Processing* 14, 5 (2020), 910–932.
[27] Ronghua Xu, Yu Chen, Erik Blasch, and Genshe Chen. 2018. Blendcac: A smart contract enabled decentralized capability-based access control mechanism for the iot. *Computers* 7, 3 (2018), 39.
[28] Ronghua Xu., Yu Chen, Erik Blasch, and Genshe Chen. 2019. Exploration of blockchain-enabled decentralized capability-based access control strategy for space situation awareness. *Optical Engineering* 58, 4 (2019), 041609.
[29] Ronghua Xu, Deeraj Nagothu, and Yu Chen. 2021. Decentralized video input authentication as an edge service for smart cities. *IEEE Consumer Electronics Magazine* 10, 6 (2021), 76–82.
[30] Ronghua Xu, Deeraj Nagothu, and Yu Chen. 2021. EconLedger: A Proof-of-ENF Consensus Based Lightweight Distributed Ledger for IoVT Networks. *Future Internet* 13, 10 (2021), 248.
[31] Fan Zhang, Ittay Eyal, Robert Escriva, Ari Juels, and Robbert Van Renesse. 2017. {REM}: Resource-Efficient Mining for Blockchains. In *26th {USENIX} Security Symposium ({USENIX} Security 17)*. 1427–1444.