SurFi: Detecting Surveillance Camera Looping Attacks with Wi-Fi Channel State Information

Nitya Lakshmanan
National University of Singapore
nityalak@comp.nus.edu.sg

Inkyu Bang†
Agency for Defense Development
ikbang@add.re.kr

Min Suk Kang
National University of Singapore
kangms@comp.nus.edu.sg

Jun Han
National University of Singapore
junhan@comp.nus.edu.sg

Jong Taek Lee
ETRI
jongtaeklee@etri.re.kr

ABSTRACT

The proliferation of surveillance cameras has greatly improved the physical security of many security-critical properties including buildings, stores, and homes. However, recent *surveillance camera looping attacks* demonstrate new security threats — adversaries can replay a seemingly benign video feed of a place of interest while trespassing or stealing valuables without getting caught. Unfortunately, such attacks are extremely difficult to detect in real-time due to cost and implementation constraints. In this paper, we propose SurFi to detect these attacks in real-time by utilizing commonly available Wi-Fi signals. In particular, we leverage that channel state information (CSI) from Wi-Fi signals also *perceives* human activities in the place of interest in addition to surveillance cameras. SurFi processes and correlates the live video feeds and the Wi-Fi CSI signals to detect any mismatches that would identify the presence of the surveillance camera looping attacks. SurFi does not require the deployment of additional infrastructure because Wi-Fi transceivers are easily found in the urban indoor environment. We design and implement the SurFi system and evaluate its effectiveness in detecting surveillance camera looping attacks. Our evaluation demonstrates that SurFi effectively identifies attacks with up to an attack detection accuracy of 98.8% and 0.1% false positive rate.

CCS CONCEPTS

• **Security and privacy** → *Mobile and wireless security*.

KEYWORDS

CSI, surveillance video, Wi-Fi

ACM Reference Format:
Nitya Lakshmanan, Inkyu Bang†, Min Suk Kang, Jun Han, and Jong Taek Lee. 2019. SurFi: Detecting Surveillance Camera Looping Attacks with Wi-Fi Channel State Information. In *12th ACM Conference on Security and Privacy in Wireless and MobileNetworks (WiSec '19), May 15–17, 2019, Miami, FL, USA.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3317549.3324928

† Research done while working at National University of Singapore.

WiSec '19, May 15–17, 2019, Miami, FL, USA

ACM ISBN 978-1-4503-6726-4/19/05…$15.00
https://doi.org/10.1145/3317549.3324928

1 INTRODUCTION

Surveillance cameras are now everywhere. Big cities around the world heavily rely on video surveillance to protect themselves from various threats.Naturally, video surveillance systems become attractive targets of attacks. Recent attacks demonstrate surveillance camera looping techniques by launching a software attack on the camera or tapping their Ethernet cables [3, 7]. Such new avenue of attacks can potentially enable unauthorized access to a security-sensitive area or unauthorized activities (e.g., stealing valuables or breaking properties) by replaying seemingly legitimate video feeds.

Unfortunately, mitigation against such attacks is surprisingly challenging. First, many existing legacy surveillance cameras have no proper end-to-end integrity protection and their hardware replacement/upgrade would incur prohibitive costs. Second, detecting replayed video feeds purely based on video signal analysis is impractical because the surveillance cameras would frequently capture nearly identical but authentic video feeds (e.g., an empty corridor or jewelry shop), which would incur too many false positives. Third, the deployment of additional infrastructure for detection (e.g., LED lights blinking in a predefined pattern) will require additional efforts such as secret sharing which potentially introduces a new attack surface.

In this paper, we try to answer the following question: *Can we utilize any auxiliary information from devices that are present along with the surveillance cameras in the place of interest?* To answer this question, we investigate whether we can leverage universally deployed Wi-Fi transceivers to effectively detect replayed video feeds. This is based on the observation that co-located devices, though different in sensing modalities, may perceive the same events. In this work, we propose SurFi[1], which detects in real-time the video feed looping attacks by comparing the video feeds with channel state information (CSI) signals, that can be easily captured via commercial Wi-Fi transceivers. Figure 1 exemplifies a scenario where an adversary first loops a seemingly legitimate video feed of an untampered vault, while he is actually breaking into the vault. SurFi

[1] SurFi stands for 'Surveillance with Wi-Fi.'

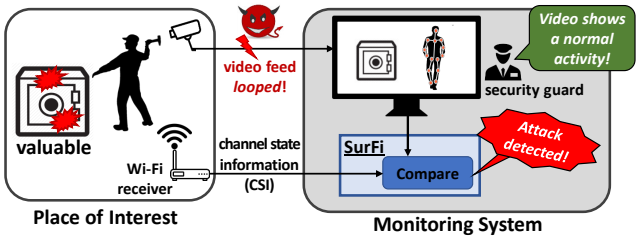

Figure 1: SurFi detects the surveillance camera looping attack by comparing live video feed with Wi-Fi channel state information (CSI) signal.

successfully detects this attack in real-time by comparing the looped video feed with the Wi-Fi CSI signal.

Designing SurFi comes with difficult challenges due to signal comparison across the two different sensing modalities. First, a direct signal comparison is impossible because signals from heterogeneous sensing modalities (i.e., video and CSI signals) are inherently *semantically different*. In particular, state-of-the-art video processing (such as OpenPose [5]) results in displacement of main body points (including head, elbow, waist, and knees) while CSI signals are time-series samples of received wireless symbols across different frequency bands. Second, we cannot simply rely on machine learning classification on CSI signals, unlike many research utilizing CSI signals for activity recognition [1, 8, 17, 18]. This is because we *cannot* expect to collect training data for all possible authorized and unauthorized activities of arbitrary visitors of a place of interest (e.g., a jewelry shop or a bank ATM office). To address these main challenges, we propose to extract common *attributes* from the time-frequency analysis of the signals of differing sensor types. Particularly, we observe that the two signals capture the *common timing information* of occurrence of activities and the corresponding *main frequency component*. Furthermore, we evaluate our proposal with real-world implementation and demonstrate that SurFi can detect potential attacks (i.e., non-matching video and CSI pair) with up to an attack detection accuracy (or true positive rate) of 98.8% and 0.1% false positive rate.

2 THREAT MODEL

We first define a *place of interest* as an indoor, open space under video surveillance. Some activities in the place of interest are defined unauthorized (e.g., breaking/opening jewelry boxes by unauthorized personnel, breaking/moving ATM machines) and the video surveillance system aims to identify such unauthorized activities. We also assume that the surveillance camera's field-of-view covers the entire place-of-interest. This is a reasonable assumption because one installs the video surveillance systems to protect valuable assets, hence purposely minimizing blind spots (i.e., area not in the camera's field-of-view).

In this paper, we consider the *surveillance camera looping attack*, where an adversary is capable of replaying a seemingly legitimate video feed (i.e., containing authorized activities only) to trick the targeted surveillance monitoring system. The ultimate goal of this attack is to evade the detection of the adversary's unauthorized activities by the authorities (e.g., security guards and personnel) as shown in Figure 1.

Recently, security researchers have demonstrated the feasibility of such attacks. In BlackHat 2013 [7], researchers demonstrated how a vulnerability in the web server interface of a surveillance camera can be leveraged to replace the live video feed with a legitimate-looking image such as an empty store or a lift compartment. In DefCon 2015 [3], another group of researchers demonstrated live video feed looping by hijacking an Ethernet cable connection between a camera and a surveillance system without breaking the physical connection.

Unlike the surveillance camera looping attack, there have been no known CSI measurement looping attacks to date. Hence, we assume that the adversaries cannot loop CSI measurements.

3 SYSTEM DESIGN

We now describe how SurFi utilizes the Wi-Fi signal to verify whether the surveillance camera system is under attack. Figure 2 depicts the flowchart diagram of SurFi's design consisting of four steps. SurFi continuously receives the real-time video feeds from the surveillance cameras as well as the CSI data from a Wi-Fi receiver. The two sets of signals are input to the *Data Pre-processing* module. This module extracts the displacement of body keypoints from the video feeds and removes noise from the CSI signals (§3.1). Subsequently, SurFi monitors the denoised CSI signals to detect the start of an event (§3.2). On detecting the start of an event, the *Attribute Extraction* module extracts three attributes (i.e., start time, end time, and prominent frequency of an event) from both the body keypoints and the denoised CSI signal (§3.3). Following this, the *Comparison* module compares the attributes and outputs the similarity score (§3.4). Finally, the *Decision* module takes the similarity scores of multiple events and outputs the final decision of whether the video is looped (§3.5).

3.1 Data Pre-processing Module

In this step, we pre-process the two raw signals obtained from a Wi-Fi receiver and a surveillance camera. The raw CSI signal is composed of multiple time-series CSI values in each subcarrier (e.g., 30 subcarriers per antenna pair in IEEE 802.11n). We can reliably collect these signals by using various open source tools [6, 15]. We denoise these raw CSI values in order to reliably extract the attributes [18]. We reduce noise from a wide frequency range of the received CSI values because we do *not* have a pre-specified set of authorized and unauthorized human activities that may occur in a place of interest; i.e., an open space with public access such as ATMs and offices. Hence, we utilize Discrete Wavelet Transform (DWT) filter, which reduces the noise from all frequency bands available in the raw CSI data (e.g., 1–500 Hz in our experiments) [1].

For the live video feeds, we use OpenPose [5], a state-of-the-art real-time video processing tool, to detect the body keypoints in the video feeds. OpenPose returns the X-Y coordinates of the 25 body keypoints per video frame. Figure 3 shows an example of the body keypoints detected by OpenPose.

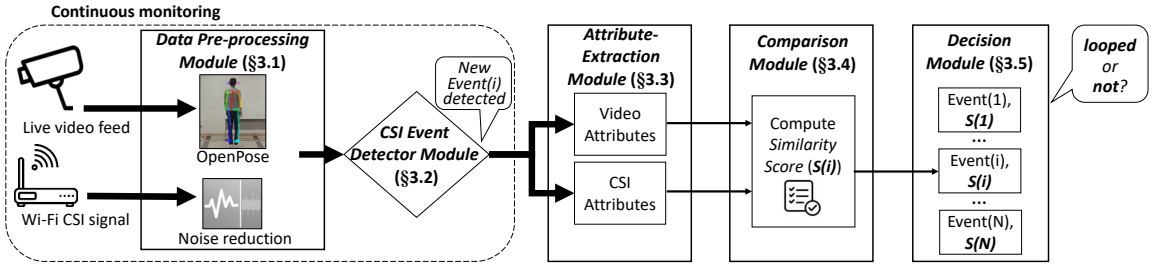

Figure 2: Figure depicts a flowchart illustrating the five modules of SurFi. The *Attribute Extraction Module* extracts common attributes from different sensing modalities (i.e., video and CSI). The computed scores are then compared in the *Comparison Module*, which in turn is input to the *Decision Module* to output the final attack detection decision.

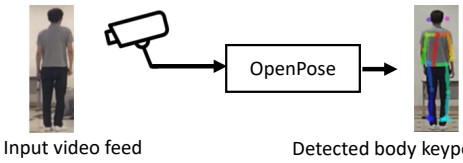

Figure 3: OpenPose extracts 25 body keypoints (including eyes, hand, feet, joints) and outputs their X-Y coordinate in real time.

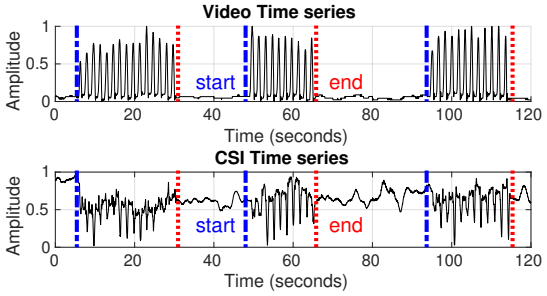

Figure 4: Time series of the left wrist body keypoint and the selected CSI subcarrier when a person moves his arm up and down.

3.2 CSI Event Detector Module

The event detector module receives the denoised CSI values and the X-Y coordinates of the 25 body keypoints, and checks if a start of a new event is detected. This module monitors the CSI values *only* because the video feeds can be manipulated by the looping attacks. The start of an event is detected using a metric referred to as *motion energy* [10], which is the energy contained in the frequency bands of the CSI signals. When the start of an event is detected, SurFi triggers the next attribute-extraction module and provide it the denoised CSI values and video keypoints until the end of the event is detected.

3.3 Attribute-Extraction Module

The two goals of the attribute-extraction module are (1) to select a set of attributes that enable *reliable* comparison of the two signals, and (2) to compute the attributes in *real-time*.

Requirements for *reliable* comparison. The first requirement is that the attributes need to be captured consistently from *both* the video and CSI signals. Fine-grained information retrievable from video signals (e.g., precise height or gait of a person) may not be adequately extracted from CSI [18], and, thus, not suitable for our purpose. The second requirement is that the attributes should capture the distinctive characteristics (e.g., the frequency of repeated actions) of *activities* with acceptable accuracy. One may argue that the accurate identification of individuals will increase the detection of looped video feed (e.g., the live video feed shows Alice whereas the CSI analysis identifies Bob). Although it may be useful, accurate identification of human beings in a large population (more than 8-10 people) based on the CSI signals remains an open problem [18, 19].

After extensive experiments under various environments, activities, and human participants, we have found the three attributes that satisfy the above requirements. The first two are time-domain components and the last is a frequency-domain component. The time-domain attributes are the *start and end time* of an event whereas the frequency-domain attribute is the *prominent frequency*. When extracting these attributes from the pre-processed video and CSI signals, we choose a single time-series component from each signal, as shown in Figure 4. That is, we choose one keypoint with the maximum signal magnitude in the frequency domain from the processed video data, which represents the largest displacement of a single body component during the measurement period. Also, we choose one subcarrier signal with the largest energy component from the denoised CSI data. This simplification reduces the attribute extraction and comparison to a single-dimension problem for real-time analysis while effectively capturing major activities performed in the place of interest.

Start and end time. We analyze the video and time-domain CSI signals to investigate whether the start and the end time of an activity or an event can be reliably captured. Figure 4 depicts an activity where a person moves his left arm up and down for a random time duration, repeating it three times.

Note that similar video signals can be obtained from the other body parts (e.g., the left elbow) but our algorithm picks the left wrist as it contains the largest magnitude. Both signals exhibit a similar start and end time, and this trend is consistently observed over multiple trials, people, and events. This allows us to conclude that they can be the reliable attributes for comparison between the video and CSI signals.

To extract the start and the end time of an event, we use a metric referred as *motion energy* [10] which captures the energy

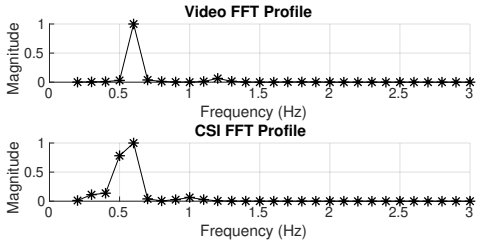

Figure 5: FFT profile of the left wrist body keypoint and the selected CSI subcarrier when a person moves his arm up and down. The prominent frequency of both the signals is 0.6 Hz, which is the frequency at which the event is performed.

in the different frequency bands of the CSI signal. The motion energy (E) can be calculated as $E = \sum_{i=1}^{L} FFT_{half}(i)^2$, where the $FFT_{half}(i)$ is the FFT coefficient magnitude calculated over a time window, L. Note that we only consider the first half of the FFT coefficients since other half is redundant and we ignore the DC component. We divide the total event time window into L of 0.1 seconds for both video and CSI signals since this gives sufficient time granularity for comparison of attributes.

We detect the start time of an event when its motion energy E increases and crosses a threshold. For the video feeds, we monitor the instantaneous E values and determine the start of an event when E becomes ten times larger than its moving average value. For CSI signals, which is still much noisier than the video signals, we determine the start of an event when the variance of E becomes larger than its moving average. The end times can be calculated similarly. We denote the start and the end times measured for an event from video and CSI signals as follows: τ^{V}_{start}, τ^{V}_{end}, τ^{C}_{start}, and τ^{C}_{end}.

Prominent frequency. The third attribute we extract is the prominent frequency component of an event. We observe that it is possible to extract a single frequency component that exhibits the maximum signal magnitude in the frequency domain. Figure 5 shows the FFT plot of the selected video and CSI signals of the aforementioned activity discussed in Figure 4. We observe that both the video and CSI signals exhibit approximately 0.6 Hz main frequency component. Hence, we conclude that a repeated event is captured as the prominent frequency component in the frequency domain.

To calculate the prominent frequency of an event, we first apply FFT on the video signal and consider the frequency with the maximum magnitude as the prominent frequency, f^{V}. Then, we apply a bandpass filter on the CSI signal with the frequency we obtain from the video signal f^{V} to remove unrelated frequency components (e.g., less than 0.3 Hz which mainly captures slower movements such as posture changes [18]) and extract the maximum magnitude frequency, f^{C}.

3.4 Comparison Module

Given the two sets of attributes $\{\tau^{V}_{start}, \tau^{V}_{end}, f^{V}\}$ and $\{\tau^{C}_{start}, \tau^{C}_{end}, f^{C}\}$ from the video and CSI signals, we compute the similarity score $S(i)$ for a single observed event $Event(i)$. We first determine the *per-attribute* similarity scores between the two signals as follows,

$$AS_j = \begin{cases} 1 & \text{if } \triangle_j \leq T_j \\ 0 & \text{otherwise,} \end{cases}, \; j \in \{1, 2, 3\}, \tag{1}$$

where T_j is the per-attribute threshold and $\triangle_j$ is the difference between the three attributes of the two signals as follows: $\triangle_1 = |\tau^{V}_{start} - \tau^{C}_{start}|$, $\triangle_2 = |\tau^{V}_{end} - \tau^{C}_{end}|$, and $\triangle_3 = |f^{V} - f^{C}|$. The per-attribute thresholds are chosen empirically to obtain an acceptable detection accuracy with a low false positive rate per attribute. To achieve this, we compute the accuracy with which each attribute detects a legitimate video as legitimate and a looped video as looped. A low threshold value may result in misidentifying the legitimate video as looped, whereas a high threshold value will cause the looped video to be incorrectly detected as legitimate, thereby decreasing the accuracy. Thus, the threshold at which the accuracy is maximized is selected as the per-attribute threshold. Based on our experiments, we observe the start and end time attribute thresholds of 2.5 seconds and 2 seconds, respectively, and a frequency attribute threshold of 0.25 Hz to be optimal. Subsequently, we compute the *per-event similarity score*, which ranges from 0 to 3 as the following: $S(i) = \sum_{j=1}^{3} AS_j$.

3.5 Decision Module

The decision module takes as input the similarity scores of one or more individually observed events and outputs a final decision of whether the video feed is looped. When multiple events are considered, detection accuracy (i.e., how effectively SurFi detects when the video feed is looped) improves and false positive rate (i.e., how frequently SurFi misidentifies a legitimate case as an attack) decreases. This is because as more similarity scores from independently observed events are considered, SurFi achieves higher confidence level for the final decision. We first take the average of the similarity score of multiple events and compare it against a decision threshold. The choice of this threshold affects the detection accuracy and the false positive rate. We decide to set a low false positive rate for SurFi and then evaluate the detection accuracy.

4 EVALUATION

We evaluate the feasibility of SurFi by conducting preliminary experiments in a controlled environment. We describe our experiment setup and present the performance evaluation.

4.1 Experiment Setup

Figure 6 shows the setup of our place of interest. We set a controlled experiment in a small indoor office room (4.9-meter wide × 2.6-meter long). We set up a transmitter-receiver pair using two Thinkpad W500 laptops equipped with Intel NIC 5300, which are spaced 1.6-meter apart on the table. We use Linux 802.11n CSI Tool [6] on Ubuntu 14.04 to extract CSI values from Wi-Fi packets at 5 GHz frequency range. The transmitter sends one ping packet every 1 millisecond and the receiver collects the CSI values. Overall, we obtain the CSI values of 90 subcarriers as the two laptops are equipped with one transmitter and three receiver antennas.

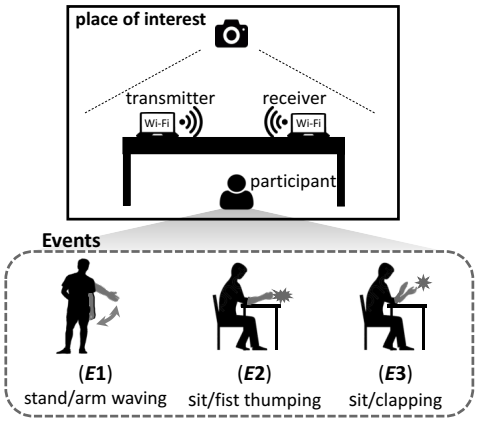

Figure 6: Experiment setup depicting the location of the camera recording the events performed, and the two Wi-Fi transceivers used for collecting CSI, and the three events we test.

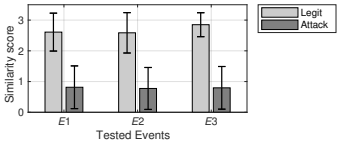

Figure 7: The distribution of similarity scores of the three events for the legitimate and attack cases. In the *legitimate* cases, we compare a *live* video feed and the CSI signal of the same event. In the *attack* cases, we instead use a *looped* video feed with a different event. The bars represent the average scores whereas the error bars denote the standard deviations.

For video feed collection, we record video clips (at 13-Megapixel resolution at the 30 frames/second sampling rate) using a camera on a mobile phone installed in the middle of the wall, with the field-of-view that faces a participant.

We ask the participants to perform the following three events:

(*E*1) standing and moving the left arm up and down repeatedly at 0.6 Hz frequency;
(*E*2) sitting and thumping right fist on the table repeatedly at 1 Hz frequency; and
(*E*3) sitting and clapping his/her hands repeatedly at 1.6 Hz frequency.

In our experiments, we test the three events with four participants and each event is tested 21 trials per participant. We instruct the participants to perform the selected activity at a certain frequency. To represent more realistic event occurrences, we randomize the starting and ending time of each trial (chosen uniformly at random between 5 to 15 seconds range, and 25 to 35 seconds range, respectively) such that each trial length ranges from 10 to 30 seconds.

4.2 Attribute-based Similarity Comparison

We now evaluate how SurFi utilizes the three attributes to compare similarities between the video and CSI signals, ultimately to detect for any potential attacks. Figure 7 depicts

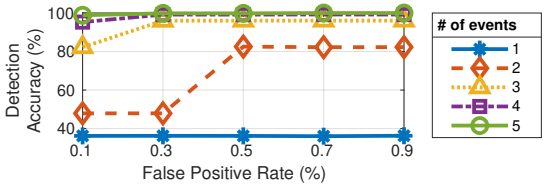

Figure 8: The attack detection accuracy improves as SurFi uses more events for attack detection.

the distribution of the per-event similarity scores $S(i)$ of an observed $Event(i)$ for the legitimate and attack cases across the three event types. For the legitimate cases, we compute $S(i)$ by comparing the attributes extracted from the live (thus authentic) video feed and the CSI signal. Subsequently, we treat the remaining cases as attack cases. Hence, for the attack cases, we compute $S(i)$ by comparing the attributes extracted from the looped video feeds, which contain different event types, and the CSI signal; e.g., $E1$ is performed in a looped video feed while $E2$ is performed in the actual experiment. Overall, we have 84 trials per event type (e.g., $E1$, $E2$, $E3$) representing the legitimate cases and 14,112 trials per event type representing the attack cases. The legitimate cases show much higher average per-event similarity scores (i.e., 2.6, 2.5, and 2.8) than the attack cases (i.e., 0.81, 0.77 and 0.79). The clear difference in the per-event similarity scores of the two cases shows that the selected attributes can effectively differentiate the cases independently of event types and subjects. Yet, their high variance (e.g., often close to 1) may cause some incorrect attack detection and false alarms.

4.3 Attack Detection Accuracy

We now evaluate the overall performance of SurFi when multiple events are observed within a short time duration (e.g., a few minutes) to increase the attack detection accuracy. We define two performance metrics as follows:

(1) *Attack detection accuracy* (or true positive rate): a ratio that the looped video feeds are correctly identified as looped;
(2) *False positive rate*: a ratio that the legitimate video feeds is incorrectly detected as looped.

We randomly append events together to emulate multiple event sequence. We then take the average of per-event similarity scores of the sequence. As discussed earlier, the more events we observe and use to detect the attacks, the higher confidence we have for our final decision, thus improving the attack detection accuracy. Figure 8 shows the attack detection accuracy of the SurFi's attack detection for a different number of events varying from 1 to 5 and a different target false positive rates (0.1% to 0.9%). We can observe that SurFi achieves 36% detection accuracy with only a single event when we target the false positive rate no higher than 0.1%. It can achieve up to 98.8% detection accuracy when it uses five events for the same target false positive rate. As the target false positive rate increases, the detection accuracy further increases as depicted in the figure. We can achieve higher detection accuracy as we use more events for attack detection because the variance of

the similarity score distributions decreases as the number of events increases [9].

5 DISCUSSION

In this section, we discuss the potential improvements for future work and some of the deployment considerations.

Stronger adversary model. One of the ways that an attacker could defeat SurFi is to mimic a subject's activities played in the compromised video. However, the level of sophistication required for such an attack would be significant, as small differences will cause differences in the *attributes*. Furthermore, SurFi's *Decision Module* requires the attacker to consistently mimic multiple events in series, making circumventing the SurFi defense even harder. We leave the study of more sophisticated attacks for future work.

Deployment considerations. When the SurFi system is deployed in practice, one may need to perform some *calibrations* because some thresholds used in our attack detection (e.g., thresholds for per-attribute similarity scores in Section 3.4 or multiple-event based decision in Section 3.5) should be adjusted to the new environment. SurFi also require to strategically locate the receiver with a certain distance away from the wall to minimize false positives that may occur due to the activities behind-the-wall [9].

6 RELATED WORK

This section presents related work on wireless sensing as well as cross-modality sensing. Refer to Section 2 for the recent work on video camera looping attacks.

Wireless Sensing. Many researchers demonstrate the feasibility of leveraging wireless signal (including Wi-Fi) for recognizing various human movements including human activity and gesture recognition [1, 2, 10, 13] as well as person identification using gait patterns [16, 18, 19]. All of these research utilize machine learning classification that requires the collection of training data. While SurFi also utilizes the analysis of CSI signals, we address an inherently more challenging problem as we cannot expect to collect training data. Bagci *et al*. [4] consider the problem of physical tampering of the Wi-Fi enabled cameras (e.g., moving or rotating the surveillance camera), which is different from our problem.

Cross-modality Sensing. Recent body of work studies how to utilize heterogeneous sensing modalities for added benefits. Researchers correlate video camera image with other sensing modalities such as RSSI value or IMU data to provide additional verification in different applications [11, 12, 14]. We are inspired by these techniques but SurFi addresses an inherently more difficult challenge of performing cross-modality correlation without the need of deploying specific sensors on users or objects, but rather utilizing the sensed information from existing infrastructure.

7 CONCLUSION

In this paper, we propose SurFi, a system that detects *surveillance camera looping attack* in real-time. SurFi leverages existing Wi-Fi infrastructure (thus requiring no additional hardware or deployment costs) to extract channel state information (CSI) to process and correlate video and CSI signals to detect any mismatches. SurFi increases its detection confidence as more events are perceived and correlated by the two heterogeneous sensing modalities. Our proof of concept of SurFi achieves up to an attack detection accuracy of 98.8% and false positive rate of 0.1%. As our future work, we plan to conduct more comprehensive experiments with diverse events and environments against a more sophisticated adversary model.

ACKNOWLEDGMENTS

We thank the anonymous reviewers of this paper for their helpful feedback. This research was partially supported by the grants from Singapore Ministry of Education Academic Research Fund Tier-1 (R-252-000-690-114, R-252-000-A26-133).

REFERENCES

[1] H. Abdelnasser, M. Youssef, and K. A. Harras. 2015. WiGest: A ubiquitous WiFi-based gesture recognition system. In *Proc. of INFOCOM*.
[2] Fadel Adib and Dina Katabi. 2013. See through walls with WiFi!. In *Proc. of SIGCOMM*.
[3] Van Albert and Banks. 2015. Looping Surveillance Cameras through Live Editing. https://www.youtube.com/watch?v=RoOqznZUClI
[4] Ibrahim Ethem Bagci, Utz Roedig, Ivan Martinovic, Matthias Schulz, and Matthias Hollick. 2015. Using Channel State Information for Tamper Detection in the Internet of Things. In *Proc. of ACSAC*.
[5] Zhe Cao, Gines Hidalgo, Tomas Simon, Shih-En Wei, and Yaser Sheikh. 2018. OpenPose: realtime multi-person 2D pose estimation using Part Affinity Fields. In *arXiv preprint arXiv:1812.08008*.
[6] D Halperin, W. Hu, A. Sheth, and D. Wetherall. 2011. Tool release: Gathering 802.11n traces with channel state information. In *ACM SIGCOMM CCR*.
[7] Craig Heffners. 2013. Exploiting Network Surveillance Cameras Like a Hollywood Hacker. https://www.youtube.com/watch?v=B8DjTcANBx0
[8] Wenjun Jiang, Chenglin Miao, Fenglong Ma, Shuochao Yao, Yaqing Wang, Ye Yuan, Hongfei Xue, Chen Song, Xin Ma, Dimitrios Koutsonikolas, et al. 2018. Towards Environment Independent Device Free Human Activity Recognition. In *Proc. of MobiCom*. 289–304.
[9] Nitya Lakshmanan, Inkyu Bang, Min Suk Kang, Jun Han, and Jong Taek Lee. 2019. SurFi: Detecting Surveillance Camera Looping Attacks with Wi-Fi Channel State Information (Extended Version). https://arxiv.org/abs/1904.01350
[10] Tapia W. Munguia. 2008. Using machine learning for real-time activity recognition and estimation of energy expenditure. In *Ph.D. dissertation*.
[11] Le T. Nguyen, Yu Seung Kim, Patrick Tague, and Joy Zhang. 2014. IdentityLink: User-device Linking Through Visual and RF-signal Cues. In *Proc. of UbiComp*.
[12] Shijia Pan, Carlos Ruiz, Jun Han, Adeola Bannis, Patrick Tague, Hae Young Noh, and Pei Zhang. 2018. UniverSense: IoT Device Pairing Through Heterogeneous Sensing Signals. In *Proc. of HotMobile*.
[13] Qifan Pu, Sidhant Gupta, Shyamnath Gollakota, and Shwetak Patel. 2013. Whole-home Gesture Recognition Using Wireless Signals. In *Proc. of MobiCom*.
[14] Carlos Ruiz, Shijia Pan, Adeola Bannis, Xinlei Chen, Carlee Joe-Wong, Hae Young Noh, and Pei Zhang. 2018. IDrone: Robust Drone Identification Through Motion Actuation Feedback. *Proc. of IMWUT*.
[15] Matthias Schulz, Jakob Link, Francesco Gringoli, and Matthias Hollick. 2018. Shadow Wi-Fi: Teaching Smartphones to Transmit Raw Signals and to Extract Channel State Information to Implement Practical Covert Channels over Wi-Fi. In *Proc. of MobiSys*. ACM.
[16] Wei Wang, Alex X. Liu, and Muhammad Shahzad. 2016. Gait Recognition Using Wifi Signals. In *Proc. of UbiComp*.
[17] Wei Wang, Alex X Liu, Muhammad Shahzad, Kang Ling, and Sanglu Lu. 2015. Understanding and modeling of WiFi signal based human activity recognition. In *Proc. of MobiCom*.
[18] Yunze Zeng, Parth H. Pathak, and Prasant Mohapatra. 2016. WiWho: WiFi-based Person Identification in Smart Spaces. In *Proc. of IPSN*.
[19] J. Zhang, B. Wei, W. Hu, and S. S. Kanhere. 2016. WiFi-ID: Human Identification Using WiFi Signal. In *Proc. of DCOSS*.