Sensor-Based Proximity Detection in the Face of Active Adversaries

Babins Shrestha, Nitesh Saxena, *Member, IEEE*,
Hien Thi Thu Truong, *Member, IEEE*, and N. Asokan, *Fellow, IEEE*

Abstract—Context-centric sensor-based proximity detection (or, contextual co-presence detection) is a promising approach to defend against *relay attacks* in many mobile authentication systems, especially against *unattended terminals* (such as cars parked in unmonitored parking lots, remote gas station pumps, or stolen laptops). Prior work demonstrated the effectiveness of a variety of contextual sensor modalities for this purpose, including audio-radio environment (ambient audio, Wi-Fi, Bluetooth, and GPS, and combinations thereof) and physical environment (temperature, humidity, gas, and altitude, and combinations thereof). In this paper, we present a systematic assessment of such co-presence detection in the presence of a strong, context-manipulating attacker against unattended terminals. *First*, we show that it is feasible to *manipulate*, *consistently control*, and *stabilize* the readings of different acoustic and physical environment sensors (and even multiple sensors simultaneously) using low-cost, off-the-shelf equipment. Specifically, we show that it is possible to control the temperature using a home-grade hair dryer, affect the gas readings using a smoking cigarette, impact the altitude/pressure with a simple air compressor, or relay audio signals recorded at one end to the other thereby causing both sides to perceive a very similar acoustic environment. *Second*, based on these capabilities and the strengthened threat model, we show that an attacker who can manipulate the context gains a significant advantage in defeating contextual co-presence detection. For systems that use multiple sensors, we investigate two sensor fusion approaches based on machine learning classification techniques—*features-fusion* and *decisions-fusion*, and show that both are vulnerable to context manipulation attacks but the latter approach can be more resistant in some cases. We further consider other defensive approaches that may be used to reduce the impact of even such a strong context-manipulating attacker. Our work represents the first concrete step towards analyzing, extending, and systematizing prior work on contextual co-presence detection under a stronger, but realistic adversarial model.

Index Terms—Sensors, environmental sensors, context manipulation, relay attack

1 Introduction

Authentication is critical to many mobile and wireless systems where one communicating device (prover $\mathcal{P}$) needs to validate its identity to the other (verifier $\mathcal{V}$). Traditional cryptographic authentication typically involves a challenge-response protocol whereby $\mathcal{P}$ proves the possession of the key K that it pre-shares with $\mathcal{V}$ by constructing a valid response to a random challenge sent by $\mathcal{V}$. Examples of systems where such authentication is deployed include payment transactions between NFC/RFID devices and point-of-sale systems, and zero-interaction authentication [12] scenarios between a token and a terminal (e.g., phone-laptop, or key-car). Unfortunately, the security and usability benefits provided by these authentication systems can be subverted by means of *relay attacks*, as demonstrated by prior research (e.g., [16], [18]), which involve two non co-present colluding attackers, one near $\mathcal{P}$ and one near $\mathcal{V}$, simply relaying protocol messages back and forth between $\mathcal{P}$ and $\mathcal{V}$.

A known defense to relay attacks is *distance bounding*, where a challenge-response authentication protocol allows $\mathcal{V}$ to measure an upper-bound of its distance from $\mathcal{P}$ [10]. Using this protocol, $\mathcal{V}$ can verify whether $\mathcal{P}$ is within a close proximity thereby detecting the presence of relay attacks [16], [18]. Although distance bounding systems are gradually becoming commercially available [1], they may not be feasible on all cost-sensitive commodity devices (such as smartphones or payment tokens) due to their sensitivity to measurement errors (of elapsed time).

The presence of ubiquitous and low-cost sensing capabilities on many modern mobile devices has facilitated a potentially more viable relay attack defense [19], [20], [25], [32]. This defense leverages the notion of "context" derived from on-board device sensors based on which $\mathcal{P}$-$\mathcal{V}$ proximity, or lack of it, could be determined. In other words, in a benign setting, where $\mathcal{P}$ and $\mathcal{V}$ are co-present, both would record a similar context with a high probability. In contrast, if the system is subject to a relay attack, and $\mathcal{P}$ and $\mathcal{V}$ are non co-present, devices' context should be different with a high probability.

Extensive recent prior work demonstrated the feasibility of using different types of sensor modalities for such

- *B. Shrestha is with VISA Inc., Austin, TX 78759. E-mail: babishre@visa.com.*
- *N. Saxena is with the University of Alabama at Birmingham, Birmingham, AL 35294. E-mail: saxena@uab.edu.*
- *H.T.T. Truong is with the NEC Laboratories Europe, Heidelberg 69115, Germany. E-mail: hien.truong@neclab.eu.*
- *N. Asokan is with Aalto University, Espoo 02150, Finland. E-mail: asokan@acm.org.*

Manuscript received 6 Dec. 2016; revised 20 Mar. 2018; accepted 7 May 2018. Date of publication 22 May 2018; date of current version 7 Jan. 2019. (Corresponding author: Babins Shrestha.)
For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org, and reference the Digital Object Identifier below.
Digital Object Identifier no. 10.1109/TMC.2018.2839604

contextual co-presence detection, including audio[1] [20], radio (*Wi-Fi* [32], *Bluetooth* [31] and *GPS* [19], and the physical environment (*temperature*, *humidity*, *gas* and *altitude/pressure*) [28]. Many single modalities, such as audio and Wi-Fi, were shown to be performing quite well for contextual co-presence detection resulting in low *false negatives* (i.e., rejecting a co-presence instance; a measure of usability) and low *false positives* (accepting a non co-presence instance; a measure of security). In addition, *fusion* of multiple modalities, including combination of audio-radio [31], and combination of physical sensors [28], has been shown to further reduce false negatives and false positives.

Our Work versus Related Work. The focus of prior work cited above on contextual co-presence detection largely centered on evaluating the system's security under the assumption that it is very hard to manipulate the contextual environment (i.e., it considered only a Dolev-Yao attacker [15]). In this paper, we are extending this model to the realm of a context-manipulating attacker, especially against unattended terminals. Vehicles parked in underground parking lots/decks represent an apt example of unattended verifiers.[2] Relay attacks against such vehicles have already been demonstrated in the literature [18] and reportedly being executed in the wild by the car criminals [17]. Other examples include stolen laptops in a zero-interaction authentication system. Payment scenarios, such as those involving parking meters or remote gas station pumps, also involve unattended payment terminals and are thus also subject to our study.

Our Contributions.The main focus of our work is on the assessment of existing contextual co-presence detection systems against active attackers. The primary contributions of this paper are two-fold:

1. *Building Simple Context Manipulation Attacks Against Unattended Terminals*. We show that it is feasible to manipulate the readings of different sensors (and combinations thereof) using low-cost, off-the-shelf equipment, representing a realistic attacker against unattended terminals. We demonstrate attacks against a variety of modalities studied in prior work including audio, radio (Bluetooth/Wi-Fi), and physical (temperature, humidity, gas and altitude).
   In particular, we demonstrate how an attacker in close proximity of the sensors can successfully control, manipulate and stabilize the physical environment "seen" by these sensors, without the need to manipulate the global surrounding environment or compromise the devices/sensors themselves. For instance, we show that it is possible to control the temperature using a home-grade hair dryer or ice cubes, affect the gas readings using a smoking cigarette, or impact the altitude/pressure with a simple air compressor made up of a plastic bag. We also observe that with practice, an attacker can increase his effectiveness surprisingly quickly, suggesting that our attacks should be taken as a lower bound—dedicated attackers are likely to fare significantly better than our results show. Our attacks are described in Section 3.
2. *Quantifying the Security of Co-Presence Detection in the Presence of Context Manipulations*. Based on the above manipulation capabilities, we comprehensively examine and *quantify* the advantage a *multi-modality attacker* can have in defeating co-presence detection over a zero-modality attacker (one studied in prior work). A *multi-modality attacker* can manipulate multiple sensor modalities simultaneously while a zero-modality attacker cannot manipulate any modality. To accomplish this, we re-orchestrated the co-presence detection approaches based on machine learning classification techniques in audio-only [20], audio-radio [31], physical [28] and (*a newly-proposed*) audio-radio-physical systems, in a way that non co-present data samples were manipulated for different modality combinations. Our results show that the attacker advantage increases *many-folds* in several cases (Table 2 quantifies the attacker success rates).
   For systems that use multiple modalities, we investigate two different sensor fusion approaches—*features-fusion* (proposed in [31]) and *decisions-fusion* based on majority voting, and show that both approaches are vulnerable to contextual attacks but the latter can be more resistant in some cases, at the cost of slight degradation in usability. Our detailed analysis is presented in Section 4.

Broader Impact and Lessons Learned. Our work represents the first concrete step towards analyzing, extending and systematizing prior work on contextual co-presence detection under a stronger, but realistic adversarial model. It suggests that tampering with context may not be very difficult, and the security offered by contextual co-presence detection therefore weakens.

Although a sophisticated attacker would likely fare better at manipulating the context (compared to our attacks), we also suggest potential strategies (including decisions-fusion) that may still be used to strengthen the security of co-presence detection against a multi-modality attacker (Section 5). At a broader level, our work calls the security of contextual co-presence detection into question, and motivates the need of re-evaluating the security of other context-centric systems in the face of context manipulation. For instance, our work may be extended to analyze the security of other promising context-based systems such as contextual access control [24] with respect to context-manipulating adversaries.

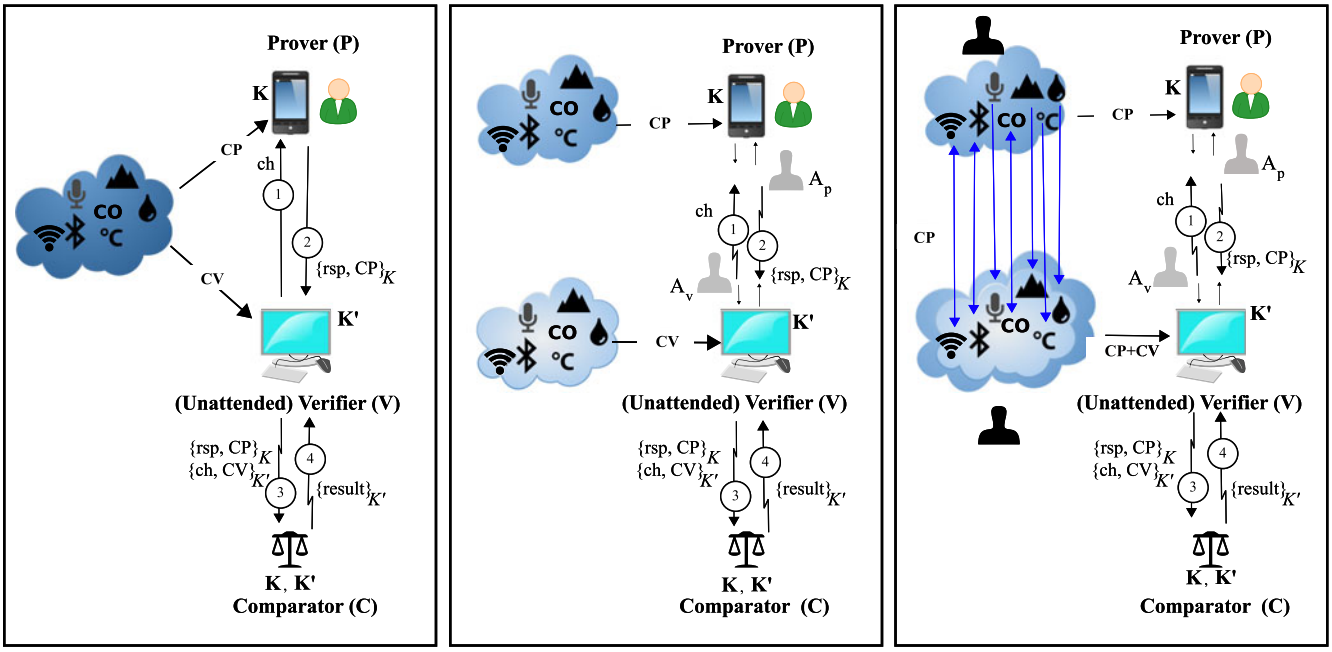
2 Background and Models

2.1 Relay Attacks & Contextual Co-Presence Detection

The goal of the adversary against a challenge-response authentication system is to fool $\mathcal{V}$ into concluding that $\mathcal{P}$ is nearby and thus needs access to $\mathcal{V}$ even when $\mathcal{P}$ is actually far away. The attacker possesses standard Dolev-Yao capabilities [15]: it has complete control of the communication channel over which the authentication protocol between $\mathcal{P}$ and $\mathcal{V}$ is run but does not have physical possession of $\mathcal{P}$ nor is able to compromise (e.g., through malware) either $\mathcal{P}$ or $\mathcal{V}$.

1. The work presented in [21] also makes use of audio to detect whether an authentication token is present near the browser, forming a two-factor authentication scheme. However, the focus of this work is not to defeat proximity or relay attacks but rather to defeat remotely located adversaries, and it is therefore not studied in this paper.

2. Clarke [11] reports that most of the theft happens at unattended places such as parking lots where there is rarely much surveillance.



(a) P and V see the same environment (b) Dolev-Yao attackers (c) (Uni/bi)-directional single-modality and multiple-modality attackers

Fig. 1. System model of proximity based authentication with contextual co-presence. In our work, the unidirectional single-modal attacker model [31] is extended to bidirectional and multiple-modality attackers (highlighted with blue arrows). Radio sensors and Gas are subject to bidirectional attacks in our new model.

The attacker could take the form of a "ghost-and-leech" [22] duo (A_p, A_v) such that A_p (respectively A_v) is physically close to $\mathcal{P}$ ($\mathcal{V}$), and A_p and A_v communicate over a high-speed connection. Such an adversary pair can compromise the security of traditional challenge-response authentication by simply initiating a protocol session between $\mathcal{P}$ and $\mathcal{V}$, relaying messages between them, leading $\mathcal{V}$ to conclude that $\mathcal{P}$ is in proximity. This is an attack applicable to zero-interaction authentication systems. A similar attack applies to proximity-based payment systems [14], [16].

Co-presence detection schemes aim to address such relay attacks. Fig. 1a shows a typical system model of an authentication/authorization protocol using contextual co-presence, adapted from [31]. In this defense, $\mathcal{P}$ (respectively $\mathcal{V}$) pre-shares a key K (K') with a "comparator" $\mathcal{C}$ (which may be part of $\mathcal{V}$ or a separate entity, depending on the scenario). When $\mathcal{P}$ sends a trigger to $\mathcal{V}$, it responds with a challenge ch. $\mathcal{P}$ and $\mathcal{V}$ then initiate context sensing for a fixed duration t. $\mathcal{P}$ computes a response rsp (using K), appends it to the sensed context information CP and sends both $\mathcal{V}$, protected by K. $\mathcal{V}$ forwards this to $\mathcal{C}$. In the meantime, $\mathcal{V}$ finishes sensing its own context and sends the resulting context data CV protected using K' to $\mathcal{C}$. $\mathcal{C}$ then recovers CP, CV, ch and rsp. It checks the validity of rsp and compares if CP is sufficiently similar to CV. If both checks succeed, $\mathcal{C}$ concludes that $\mathcal{P}$ and $\mathcal{V}$ are co-present. When $\mathcal{C}$ is integrated with $\mathcal{V}$, K' is not used. Fig. 1b shows how contextual co-presence can thwart a Dolev-Yao relay attacker.

Prior work has proposed the use of different sensor modalities for such co-presence detection: ambient audio – Au [20], radio context including Wi-Fi – W and Bluetooth – B [31], and physical environmental attributes, temperature – T, humidity – H, concentration of gases – G and altitude – Al [28].

2.2 Threat Model: Single-Modality Contextual Attacker

Our focus is on a context-manipulating attacker against co-presence detection (going beyond a Dolev-Yao attacker). Truong et al. [31] briefly explored the problem of characterizing such a contextual attacker. They only consider an attacker who is capable of manipulating a single sensor modality at a time ("single-modality attacker", in our parlance). Again, in this model, an attacker cannot compromise $\mathcal{P}$ and $\mathcal{V}$ devices. Based on the rationale that $\mathcal{V}$ is often *unattended*, whereas $\mathcal{P}$ is in the possession of a human user, they speculated that the context attacker can manipulate context without detection only in one direction. More precisely, they modeled a single-modality attacker as follows:

- A_p, A_v can measure the context information that $\mathcal{P}$, $\mathcal{V}$ would sense, respectively.
- A_v can fool $\mathcal{V}$ into sensing the context information A_v chooses. Specifically A_v can receive context information from A_p and reproduce it near $\mathcal{V}$.
- A_v (A_p) cannot suppress any contextual information from being sensed by $\mathcal{V}$ ($\mathcal{P}$).

Fig. 1c illustrates this threat model. Later in Section 4, based on our context-manipulation attacks presented in Section 3, this current model will be extended, to incorporate multi-modality attackers, who can perform the above (single-modality) tasks corresponding to multiple modalities simultaneously.

3 Context Manipulation Attacks and Attack Experiments

In this section, we present our context manipulation attacks against audio, radio and physical sensor modalities, and their various combinations. We explain the concepts

underlying the attack approaches and present attack experiments when needed. "Modality" refers to the *type* of sensor that generates the raw input data used for modeling the ambient context [29].

3.1 Manipulating Audio Sensor Modality

To manipulate ambient audio, an adversary must find a way to make ambient audio on one side similar to that on the other side. Recall from Section 2 that our threat model allows the attacker to add to the ambient audio at $\mathcal{V}$'s side without being noticed, allowing him to relay/stream the ambient audio in real-time from $\mathcal{P}$'s side to $\mathcal{V}$'s side thereby causing the features used for audio correlation almost match at both sides. The assumption that manipulating audio at $\mathcal{V}$'s side can go undetected is valid since $\mathcal{V}$ may be unattended in many scenarios (as our model in Section 2 assumed). The attacker duo can use any reliable audio streaming tool to stream the audio from $\mathcal{P}$'s side to $\mathcal{V}$'s side. They can execute this attack conveniently using mobile phones and wireless data connection. We evaluated how well such an attacker can succeed in fooling audio-based co-presence detection by streaming ambient audio using Skype [2]. We use the features and classifier described in prior work [20]. Our results are presented in Section 4.2.1.

3.2 Manipulating Radio-Frequency Sensor Modalities

Prior work suggests that manipulating the radio context is possible in general. The work presented in [30] describes attacks on a public Wi-Fi based positioning system. They used a Linux laptop as an Access Point (AP) with the Scapy packet manipulation program [7] to spoof Wi-Fi APs. Similarly, spoofing bluetooth device addresses has already been demonstrated in prior work [23], [31], both of which reported bluetooth-based relay attacks. An attacker can control the received signal strength by controlling the transmission power of his masquerading devices. Therefore, we conclude that the threat model assumed in [31] (see Section 2.2) is reasonable. Furthermore, in the case of Radio Frequency (RF) sensor modalities, it is reasonable to assume that an attacker can also manipulate the RF environment at $\mathcal{P}$'s end without being noticed (since radio waves are imperceptible to human users). Therefore, limiting the attacker to unidirectional manipulation only is too restrictive.

We tested the feasibility of Wi-Fi spoofing ourselves, and studied how it can be used to match the Wi-Fi context at two ends. In our experiment, we used a Linksys router (WRT54G) to create a spoofed hotspot. We flashed DD-WRT firmware [13] to the router since the default firmware did not allow us to spoof the Basic Service Set Identifier (BSSID). The router used in our experiment is portable, easily available in the market, and much cheaper than other devices which can also be used to spoof the hotspot such as laptops or smartphones.

The DD-WRT control panel also provides an option to change the transmission power with which we can increase/decrease the signal strength. The normal signal strength for the router detected by our target device (a MacBook Air laptop) was around -39 dBm. The router and the target device were located around 30 cm apart. Merely by adjusting router settings, we were able to vary the signal strength of the router, as sensed by the target device, between -25 and -48 dBm. By changing the distance between the target device and the spoofed router, we were able to further reduce the signal strength down to -87 dBm. This suggests that the adversary has a high degree of control in manipulating sensed signal strength. Based on this spoofing and Received Signal Strength Indicator (RSSI) manipulation capability, the Wi-Fi context matching attack becomes rather straightforward. The attacker can even have advantage in environments where number of Wi-Fi APs is low. For example, we observed that there are less than five APs in outdoors such as parking lot. In such cases, the attacker would only need to spoof $\mathcal{P}$'s side.

3.3 Manipulating Physical Environment Sensor Modalities

As discussed in [28], it may seem hard to manipulate physical modalities, Temperature T, Humidity H, Gas G and Altitude Al. For example, it appears that an adversary has to change the temperature or humidity of the entire environment surrounding the victim device which may be quite challenging or detected easily. However, in this section, we show that, by using off-the-shelf devices, manipulating physical context is not only feasible but also realistic and effective by tampering with the "local" environment close to one of the devices (e.g., an unattended $\mathcal{V}$). Our attacks do not require the compromise of the devices ($\mathcal{V}$ or $\mathcal{P}$), but rather only manipulation of environment close to their sensors. In order to monitor the current ambient readings as they are being changed, the attacker has to use his sensors. These ambient readings serve as a feedback for the attacker while he attempts to change the current $\mathcal{V}$'s ambience. The feedback sensor needs to be placed very close to the victim sensor so that the two provide similar readings.

Our experiments demonstrate how different sensor modalities can be *manipulated, controlled* and *stabilized* to enable successful relay attacks. Arbitrarily changing a sensor's readings, at the verifier's side, based on a physical activity may be straightforward but consistently maintaining and controlling these readings to match those at the prover's side, is non-trivial. For example, it may be obvious that temperature can be increased using a hair dryer (a simple tool used in our temperature manipulation experiments), but how to maintain it at a desired level for a reasonable period of time (during which the attack can be launched) is not obvious. While we present several direct/explicit ways to manipulate many modalities, we also demonstrate some indirect/implicit techniques. For example, we show how altitude can be manipulated by changing pressure (i.e., without relocating the device to a different altitude). When performing the attacks, we need to consider that the attacker will not have access to the direct readings from the actual ($\mathcal{V}$) device and hence has to use his own sensors to monitor the current ambient readings during the attack. These ambient readings serve as a feedback for the attacker while he attempts to change the current $\mathcal{V}$'s ambience. The feedback sensor needs to be placed close to the victim sensor so that both provide similar readings.

3.3.1 Temperature Manipulation

We were able to successfully alter the temperature to a desired level using various household items, such as a hair dryer, a coffee mug, and ice cubes. All of our experiments were performed with Sensordrone devices serving as both $\mathcal{V}$ and the attacker's feedback sensor.

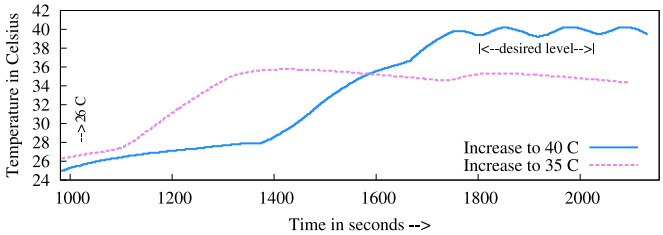


Fig. 2. Increasing T to desired level (35 °C and 40 °C).

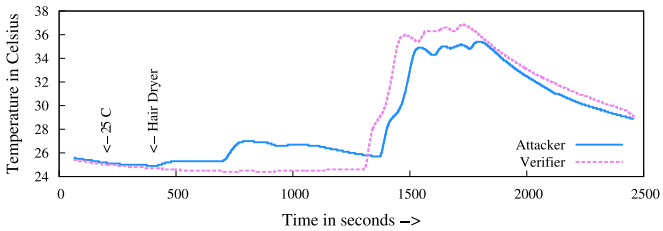


Fig. 3. *VS* and *FS* on same location; the attacker trying to increase temperature to 35 °C.

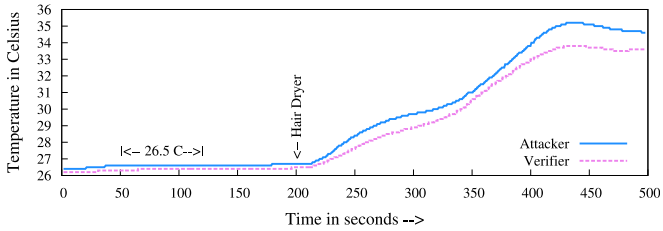


Fig. 4. Heating an area; *VS* and *FS* within a range of 15 cm; the attacker trying to increase temperature to 35 °C.

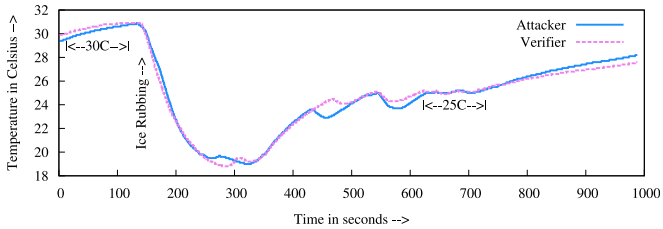


Fig. 5. Decreasing temperature with an ice cube; the attacker trying to decrease to 25 °C.

Increasing the Temperature. In situations where $\mathcal{P}$ (e.g., a car key indoors) is at a higher temperature than $\mathcal{V}$ (e.g., a car parked outside in winter), the attacker must increase the temperature. We first used a hair dryer to heat-up the area around the Sensordrone such that the temperature is increased to a desired level. To monitor how the temperature increases as we bring the hair dryer closer to $\mathcal{V}$, we first placed the hair dryer far enough and then brought the hair dryer closer to the sensors in a way that we can handle the increase in temperature gradient. In our experiment, we first tried to increase the temperature to 40 °C and then to 35 °C. After a few attempts, we could successfully increase the temperature to a desired level and stabilize for almost 2 minutes (Fig. 2). The lab temperature when the experiments were performed was around 26 to 27 °C. The hair dryer we used [5] had a power of 1,875 watt AC. A video demonstration of our attack has been uploaded to YouTube [6]. To perform this attack in a real world, the attacker can use a battery-operated device or a power outlet from his vehicle when $\mathcal{V}$ is located in a parking lot.

Our next set-up uses two sensors, $\mathcal{V}$ sensor (*VS*) and feedback sensor (*FS*), to change the temperature. Depending on whether or not the attacker knows where the sensor is precisely located on $\mathcal{V}$ device, he may place *FS* either exactly on top of *VS* or away from it. We performed the hair dryer test such that: (1) *FS* is placed at the same place as *VS*; (2) *FS* is placed such that *VS* is closer to hair dryer than *FS*; and (3) *FS* is placed such that *FS* is closer to hair dryer than *VS*.

For the first case, we were able to match the temperature on both sensors to a large extent when performing the heating activity (Fig. 3). However, if the attacker does not know the location of *VS* then the sensor device closer to the hair dryer ends up getting more heated. These attacks are described in Appendix A in detail. Hence, the attacker should heat up the whole area as he may not be able to place his *FS* exactly on top of *VS*. Subsequently, we tried to apply the heat not just focusing on one particular area but rather heating the entire area within a range of 15 cm. Using this approach, we could effectively change the temperature around *VS* with feedback from *FS* as the two temperature curves move side by side (Fig. 4). We were able to control the temperature to a desired level within a variance of $+/-0.3$ °C for more than one minute in *FS* device. It took us less than 4 minutes to reach the desired temperature level. We also observed that our ability to make the sensor reach the target temperature and sustain it for a significant duration (~ 1 minute) increased dramatically with experience. This suggests that determined and professional attacker will fare significantly better than our results.

Decreasing the Temperature. In some scenarios, it might be necessary for the attacker to reduce the temperature recorded by $\mathcal{V}$ (e.g., when $\mathcal{P}$ is indoors and $\mathcal{V}$ is outdoors during summer conditions). To decrease the temperature readings, we used an ice cube and rubbed it against the sensor. The environment on the other hand increased the temperature. By using the ice cube, we first tried to drop the temperature below 20 °C and then let the environment increase the temperature naturally. This natural increase of the temperature was very slow, and when the temperature started increasing beyond the desired temperature level, we gently rubbed the ice again to stabilize the temperature. We conducted experiment in a parking deck where the ambient temperature was around 30 °C. Our goal was to change the temperature down to 25 °C. We rubbed the ice cube on the sensors (both $\mathcal{V}$ and feedback sensors) until the temperature decreased to less than 20 °C. Afterwards, the temperature started rising slowly naturally. When it reached around 25.2 °C, the ice cube was rubbed gently again on the sensors such that the temperature drops slightly. We were able to decrease the temperature and stabilize it at 25 °C for more than a minute after a few trials within a variance of $+/-0.3$ °C as shown in Fig. 5.

3.3.2 Humidity Manipulation

To alter humidity, we used common household items such as hot coffee (for increasing humidity) and hair dryer (for decreasing humidity).

Increasing the Humidity. Coffee fumes when brought close to *VS* would increase the humidity level. An attacker has to move the hot coffee cup nearer to, and farther away, from the

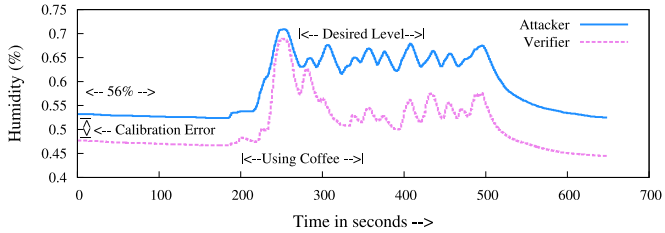


Fig. 6. Increasing humidity with hot coffee; the attacker trying to increase to 65 percent.

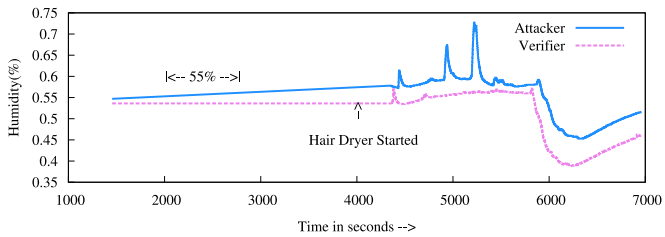


Fig. 7. Decreasing humidity with a hair dryer such that VS and FS are at same location; the attacker trying to decrease to 50 percent.

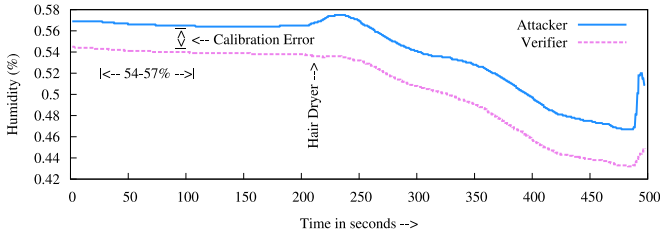


Fig. 8. Decreasing humidity with a hair dryer such that VS and FS are within a range of 15 cm; the attacker trying to decrease to 50 percent.

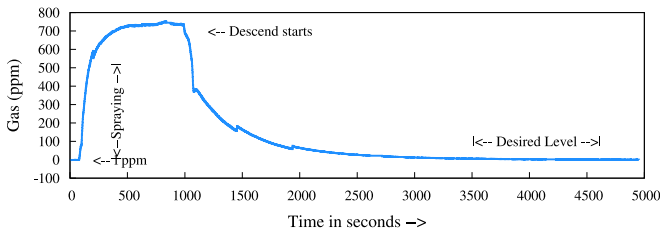


Fig. 9. Effect of aerosol spray in CO level; increasing the CO gas level to arbitrary value and wait to decrease to desired level.

sensors to control the humidity level. Using this strategy, we were able to increase the humidity by 10 percent, i.e., from normal humidity of 55 to 65 percent (Fig. 6). The attacker needs to use FS to control the humidity. On our first attempt, we were able to control the humidity with a variance of $+/-3$ percent for almost 30 seconds. In the second attempt, we could raise the humidity to the desired level for more than one minute (106 seconds) with the same threshold.

Decreasing the Humidity.A hair dryer can be used to dry-up the air around the sensor to reduce the humidity. The setup of this experiment is similar to the hair dryer temperature increase experiment. We tried to decrease the humidity of VS by monitoring the humidity change on FS. When two devices are placed exactly at the same location, the humidity decreases and matches consistently between the two devices (Fig. 7). Even when the two devices are placed 15 cm apart, the drop in the humidity readings coincides (Fig. 8).

3.3.3 Gas Manipulation

Following prior work [28], we study Carbon Monoxide (CO) level as a modality for co-presence detection. While manipulating this modality, an attacker may not be detected even when he alters the gas content near either $\mathcal{V}$ or $\mathcal{P}$ (unlike the rudimentary model of Section 2.2), unless there is a significantly large change, or gas monitors are installed. This provides flexibility to the attacker to increase/decrease the CO level at both sides such that both readings match.

Increasing the Gas (CO). We performed several activities such as using a smoking cigarette to exhale a high amount of CO gas to the sensor, and using a car exhaust to increase the CO level. We also found out that room heaters emit gases which increase CO readings when we placed the sensor device on top the gas vent while the heater was turned on. The aerosol spray also increased the CO level when it was sprayed around on top of the sensor. The effect of different propane gas heaters as well as aerosols air fresheners on gas content has been mentioned in [9]. All these activities, though, increased the CO level abruptly, it takes a long time for sensor reading to descend back to normal, which provides the attacker with a sufficiently long attack window as shown in Fig. 9. The effects of cigarette and car exhaust on CO level are described in Appendix B in detail. We observed these activities for more than five times, and noticed that it took more than thirty seconds to decrease by 1 ppm when gas level decreased below 10 ppm which is already above average of normal gas level.

*Decreasing the Gas (CO).*To reduce the gas level, an attacker needs to "purify" the air from the CO content around the sensors. We implemented this strategy using a kitchen exhaust fan which is used to remove pollutants. We found that when sensor was placed near the exhaust fan, it decreased the CO gas content.

The gas reading heavily depends upon the location of $\mathcal{P}$ and $\mathcal{V}$. In a heavy traffic or polluted area, this may be higher than 10 ppm while in a normal workplace, it may be around 0 to 5 ppm. If $\mathcal{P}$ is located in low CO area while $\mathcal{V}$ is located in high CO area, the attacker may use the kitchen exhaust fan activity to decrease the CO level in $\mathcal{V}$'s location. However, if the attacker cannot reduce the CO level by significant amount, he can always collude with the attacker at $\mathcal{P}$'s side to increase the CO level using an aerosol spray. This can increase the CO level by significant amount and then it only takes a while to fall back to the normal gas level. This effect can be confirmed from Fig. 9.

3.3.4 Altitude Manipulation

The altitude of a location is inversely correlated to the pressure at that location. The Sensordrone device detects the pressure, and uses it to calculate the altitude based on a standard conversion method.

Manipulating sensors so as to increase or decrease altitude directly seems very difficult. In order to manipulate the altitude readings, one may physically carry the verifier device to a higher or lower altitude as needed. If the verifier device is portable (such as a stolen laptop), doing so is easy. However, there are many scenarios where directly changing the altitude is not feasible (e.g., when $\mathcal{V}$ is a car and $\mathcal{P}$ is a car key carried in victim's pocket). We show that it is still

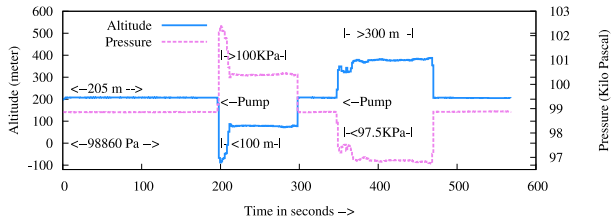


Fig. 10. Using an air pump to change pressure to the sensors wrapped inside a Ziploc bag by pumping air in and out.

possible to manipulate altitude readings *indirectly* by manipulating the pressure readings.

*Increasing the Altitude.*To increase the altitude indirectly, an attacker must decrease the pressure near the sensors. To achieve this functionality, we created a low-cost air compressor. We placed the sensor inside a Ziploc bag and then used an electric air pump [4] to suck-up the air from the bag. When $\mathcal{V}$ is large in size or shape (such as a car), an attacker just needs to create an enclosure around its sensor, while if it is a portable/small device (e.g., a laptop), the device itself can be placed inside a bag. When the air pump sucks up the air around the sensors enclosed inside the Ziploc bag, the weight of air exerted on the sensor is reduced. This reduces the pressure around the sensor and hence increases the altitude level. In our experiment, we effectively altered the altitude by more than 60 meters (Fig. 10). By using an air pump with a higher power, the attacker can further increase the altitude level. A vacuum cleaner may also be used in place of an air pump (as described in Appendix 9).

Decreasing the Altitude. To decrease the altitude (i.e., increase the pressure), we placed the sensor inside a polythene bag and applied high pressure by squeezing the bag, blowing air into the bag, and finally using the air pump device to blow the air inside. First, we wrapped the sensor inside a polythene bag to see if there is any change in altitude when we blow air into the bag by mouth, or squeeze the air tight polythene bag. This increased the pressure by very high amount and decreased the altitude correspondingly. However, it was not doable in a controlled way, i.e., sometimes the altitude decreased by 5 meters while on other occasions, it decreased by 50 meters. Ideally, an attacker would want to have a relatively long time window where the desired altitude remains constant for him to perform the relay attack. To address this issue, we used the air pump mentioned above. Filling up the air into the bag increased the pressure and decreased the altitude such that it remained constant for almost 14 seconds. A video demo of this experiment has been uploaded to YouTube [8].

3.4 Manipulating Multiple Sensor Modalities Simultaneously

As demonstrated by prior work [20], [28], [31], a contextual co-presence detection system can use combinations of several sensor modalities. In such cases, the attacker needs to manipulate multiple modalities at the same time (multi-modality attacker). However, performing one activity may be altering not only the target modality but also one or more other modalities that a system might be using for context detection, such as (T and H) or (Al and Au) even though they are not directly correlated.

For example, hair dryer increases temperature but also dries-up the air (i.e., potentially reduces the humidity) around the sensor where it is applied. It also changes the ambient noise. An attacker needs to manipulate in such a way that if the multiple modalities are involved in the system he should change the target modality without altering other modalities by effective amount. We also found that hair dryer activity results in a huge momentary change in gas level. However, the reading comes back to normal when hair dryer is applied for a long period of time. Altitude and pressure did not change with the hair dryer activity. Hair dryer activity also does not impact on RF signals. Hence, hair dryer activity can be used to manipulate the system which uses either temperature or humidity along with gas, altitude and RF signals.

Using aerosol spray to increase the gas content does not have effective change on any other modalities besides humidity. Similarly, updating RF signals does not seem to have any effect on physical modalities. Therefore, an attacker can simultaneously manipulate radio, temperature and gas while he hopes that audio, altitude and humidity either match the minimum criteria from both sides or is not used by the system.

Using an ice cube to decrease the temperature does not affect other modalities effectively. However, if the ice melts then it may affect the humidity of the space near the sensors. In our experiment, we saw that humidity fluctuates when we tried to decrease the temperature using an ice cube. Hence, using an ice cube to decrease temperature activity can be used with all other modalities except altitude and humidity.

Hot coffee cup changes the humidity along with the temperature, while other modalities remain unchanged. In this case, an attacker can manipulate humidity along with radio, audio and gas while he cannot control temperature and humidity together.

When an attacker has to use an air pump or vacuum cleaner to increase or decrease the altitude, it affects ambient noise. Also, an air pump was used in conjunction with a Ziploc bag where the sensors were wrapped to create an enclosed space. When an attacker performs such activity with an enclosed space, it will be very difficult for him to change gas, temperature or humidity. We thus may only claim that the attacker can manipulate altitude along with radio modalities.

To summarize, our attacks support the following combinations of multi-modality manipulations: (1) Al, B, W; (2) Au, B, G, (increase for H), W; (3) Au, B, G, (decrease for T), W; (4) Au, B, G, W; (5) B, G, H, W; (6) B, G, T, W. However, a more sophisticated attacker (than the one we considered) may use different techniques to possibly attack other combinations too.

4 Security of Co-Presence Detection Systems under Context Manipulation

In light of the attacks presented in Section 3, we first extend the rudimentary contextual attacker model from [31] as follows:

- We allow multi-modality attackers who can simultaneously control multiple sensor modalities, in addition to the single-modality attacker of [31].
- We assume that a contextual attacker can manipulate radio contexts in *both directions*. The same assumption applies to Gas sensors in light of our aerosol spray attack.

4.1 Analysis Methodology

4.1.1 Datasets

To fairly evaluate the resilience of co-presence detection systems in the presence of our contextual attacker, we used the same datasets and the same set of features originally used to evaluate the systems in question. The previous audio-radio system [31] used a dataset to evaluate resistance against single-modality attackers. The previous physical system [28] used a dataset to model a zero-modality attacker. We use these datasets to evaluate the resistance of the respective systems against multi-modality attackers. In addition, we conducted new audio relaying experiments to collect data and evaluate audio-based co-presence detection performance. Furthermore, we collected a new dataset corresponding to the audio-radio-physical system (which was not considered in prior works).

Data-PerCom. The dataset from Truong et al. [31] contains 2,303 samples, of which 1,140 samples (49.5 percent) are from co-present devices and 1,163 (50.5 percent) from non co-present devices. Each sample contains data from sensor modalities available at the time on the respective devices (2,117 with audio, 1,600 with Bluetooth, 782 with GPS and 2,269 with Wi-Fi). Recording time varies from sensor to sensor: 2 minutes for GPS scanning, 10 scans for Wi-Fi (about 30 seconds), 10 seconds for ambient audio, and 10 scans for Bluetooth (up to 12 seconds for each scan).

Data-FC. The data from Shrestha et al. [28] contains sensor data for ambient temperature, precision gas, humidity, and altitude. Data was recorded and labeled according to the location and time of the place. The data was also marked how the device was held, i.e., either in hand or in pocket. The experiment was conducted in a variety of places, not just confined to labs and typical university offices. The locations included: parking lots, office premises, restaurants, chemistry labs, libraries as well as halls with live performance and driving on interstate highways. We collected a total of 207 samples at 21 different locations. The different samples collected from the same place are "paired" to generate co-presence data instances whereas those from different places are paired to generate non co-presence data instances. We ended up with 21,320 instances of which 20,134 instances belonging to non co-presence class and 1,186 instances belonging to co-presence class.

Data-TMC-Audio. To assess how an attacker can manipulate ambient audio via the streaming attack (Section 3.1), we conducted a set of experiments to collect about 100 audio samples for the non co-presence case. The audio streaming was done over two different channels: Wi-Fi and cellular data. $\mathcal{P}$ was a Galaxy Nexus device while $\mathcal{V}$ was a Galaxy S3 device. Unidirectional streaming of the audio from $\mathcal{P}$'s side to $\mathcal{V}$'s side was done between a pair of devices (from a Galaxy S4 to an iPhone 5 in the case of the cellular data channel, and from a MacBook Air to a ThinkPad Carbon X1 in the Wi-Fi channel). The attacker devices used a Skype connection as the audio relay channel.

Data-TMC. We extended the data collector used in [31] to record physical sensor data using an attached Sensordrone device (as used in [28]). Different device models were used to record sensor data. Each device, in a pair of devices, was connected to its own Sensordrone device. Two users were involved in the data collection. Data was collected at different locations in two countries for ten days. The resulting dataset has 203 non co-presence samples and 335 co-presence samples.

4.1.2 Features for Co-Presence Detection

We summarize features used for co-presence detection in prior works.

Audio Features. Halevi et al. [20] proposed the use of (only) audio for co-presence detection. Audio features include max cross correlation and time frequency distance.

- Max cross correlation:
  $M_{corr}(a,b) = Max(cross\ correlation(X_a, X_b))$
- Time frequency distance:
  $D(a,b) = \sqrt{(D_{c,time}(a,b))^2 + (D_{d,freq}(a,b))^2}$ where, $D_{c,time}(a,b) = 1 - M_{corr}$, $D_{d,freq}(a,b) = ||FFT(X_a) - FFT(X_b)||$ is the euclidean norm of the distance.

Here X_a and X_b denote the raw (16-bit PCM) audio signals recorded by A and B and FFT(X_a), FFT(X_a) denotes the Fast Fourier Transforms of the corresponding signals.

Radio (Bluetooth, Wi-Fi, GPS) Features. Truong et al. [31] used a set of features with small variance for radio sensor modalities. They let a sample from an RF sensor modality be of the form (m, s) where m is an identifier of a sensed device and s is the associated signal strength. Also, they let S_a and S_b denote the set of records sensed by a pair of bound devices A and B, and let n_a and n_b denote the number of different beacons (i.e., Wi-Fi access points, satellites or Bluetooth devices) observed by devices a and b. We define the following sets:

$S_a = \{(m_i^{(a)}, s_i^{(a)}) \mid i \in \mathbb{Z}_{n_a - 1}\}$.
$S_b = \{(m_i^{(b)}, s_i^{(b)}) \mid i \in \mathbb{Z}_{n_b - 1}\}$.
$S_a^{(m)} = \{m\ \forall (m,s) \in S_a\}$, $S_b^{(m)} = \{m\ \forall (m,s) \in S_b\}$.
$S_\cap = \{(m, s^{(a)}, s^{(b)})\ \forall m | (m, s^{(a)}) \in S_a, (m, s^{(b)}) \in S_b\}$.
$S_\cup = S_\cap \cup \{(m, s^{(a)}, \theta)\ \forall m | (m, s^{(a)}) \in S_a, m \notin S_b^{(m)}\}$
$\cup \{(m, \theta, s^{(b)})\ \forall m | (m, s^{(b)}) \in S_b, m \notin S_a^{(m)}\}$,
θ is modality-specific (see below).
$S_\cap^{(m)} = \{m\ \forall m | (m, s^{(a)}, s^{(b)}) \in S_\cap\}$.
$S_\cup^{(m)} = \{m\ \forall m | (m, s^{(a)}, s^{(b)}) \in S_\cup\}$.
$L_a^{(s)} = \{s^a | (m, s^{(a)}, s^{(b)}) \in S_\cap\}$.
$L_b^{(s)} = \{s^b | (m, s^{(a)}, s^{(b)}) \in S_\cap\}$.

$S_\cap$ consists of devices seen by both A and B; $S_\cup$ represents all devices seen by A or B with θ filled in as the "signal strength" for devices that are *not* seen by either device.

1) Jaccard distance: $1 - \frac{|S_\cap^{(m)}|}{|S_\cup^{(m)}|}$
2) Mean of Hamming distance: $\frac{\sum_{k=1}^{|S_\cup|} |s_k^{(a)} - s_k^{(b)}|}{|S_\cup|}$
3) Euclidean distance: $\sqrt{\sum_{k=1}^{|S_\cup|} (s_k^{(a)} - s_k^{(b)})^2}$
4) Mean exponential of difference: $\frac{\sum_{k=1}^{|S_\cup|} e^{|s_k^{(a)} - s_k^{(b)}|}}{|S_\cup|}$
5) Sum of squared of ranks: $\sum_{k=1}^{|S_\cap|} (r_k^{(a)} - r_k^{(b)})^2$ where, $r_k^{(a)}$ (respectively $r_k^{(b)}$) is the rank of $s_k^{(a)}$ ($s_k^{(b)}$) in the set L_a (L_b) sorted in ascending order.
6) Subset count: $\sum_{i=1}^{T} f_i$. Here T is the scanning time (seconds)
   $f_i = 1$ if $S_{a_i}^{(m)} \neq \emptyset$, $S_{b_i}^{(m)} \neq \emptyset$,
   $(S_{a_i}^{(m)} \subseteq S_{b_i}^{(m)}$ *or* $S_{a_i}^{(m)} \supseteq S_{b_i}^{(m)})$
   $f_i = 0$ otherwise. S_{a_i}, S_{b_i} are the set of records by A and B respectively at the ith second

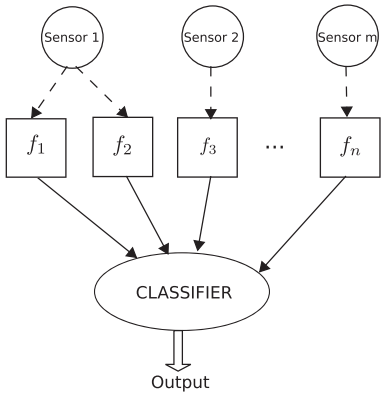


Fig. 11. Features-fusion technique.

Features 1-5 are used for Wi-Fi (θ is -100.). Features 1-3 are used with BDADDR as identifier (m) and average RSSI as signal strength (s) for Bluetooth(θ is -100).

Physical Sensor Features. Shrestha et al. [28] used Hamming distance as a feature on physical sensors for co-presence. Let L_i and L_j be a sensor reading captured by two devices at locations i and j. The Hamming distance is calculated as follows:

$$D(i,j) = |L_i - L_j|.$$

Given n different sensor modalities and the input data for the kth modality $L_i^{(k)}$ and $L_j^{(k)}$ from two samples, we have $D^{(k)}(i,j) = |L_i^{(k)} - L_j^{(k)}|$. With the data corresponding to n modalities, we obtain a feature vector of n elements of $D^{(k)}(i,j) \mid 1 \leq k \leq n$.

4.1.3 Classification Techniques

In evaluating prior systems, we used the same classification techniques as in the original evaluations (Decision Tree and Random Forest), implemented in Scikit-learn [26]. The results are reported after running ten-fold cross validation. We use *False Positive Rate (FPR) as a metric to represent the attacker's success probability*. FPR corresponds to "non co-presence" samples which are mislabeled as "co-presence", reflecting the security of the system (higher the FPR, lower the security). We use False Negative Rate (FNR) as a metric to represent the usability of the system. FNR represents "co-presence" samples that are mislabeled as "non co-presence" (lower the FNR, better the usability). F1 score is reported only for the overall performance of the classification model under zero-modality attack.

Whenever multiple sensor modalities are used, we fuse the data from these modalities before feeding it to the classifier. We considered the following fusion approaches.

Features-Fusion. The features of all sensor modalities are together fed to the classifier (see Fig. 11). The decision of co-presence or non co-presence is made one-time only based on the output of the prediction model. Prior work [28], [31] implemented this fusion technique.

Decisions-Fusion. Each of the n sensors (with all its features) is used separately by the classifier. As result there are n decisions made. All decisions are then combined to produce a final decision. This is an approach that has not been used for co-presence detection in previous works. Decisions-fusion can aggregate decisions from single sensor modalities or from subsets of sensor modalities, for example, three subsets can be built on top of seven sensors: acoustic = {Au }, radio = {B, W }, physical = {Al, G, H, T }. In the latter fusion approach, classifiers of subsets are built using features-fusion. Fig. 12 illustrates decision-fusion technique.

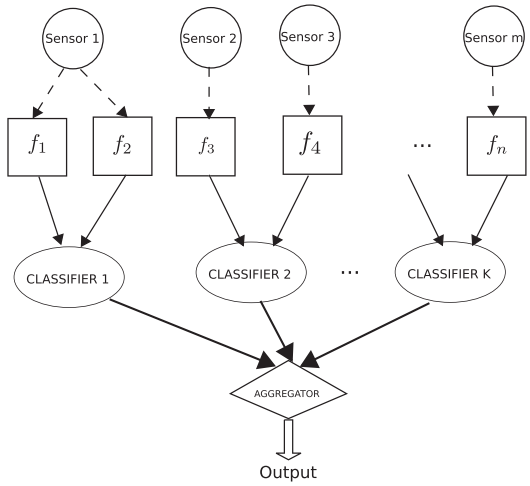


Fig. 12. Decisions-fusion technique.

4.2 Analysis Results

4.2.1 Audio-Only System

Halevi et al. [20] proposed the use of (only) audio for co-presence detection. Their work showed that audio is a good ambient context resulting in 100 percent accuracy and 0 percent False Positive Rate.

The audio features used in [20] are based on audio frequency. Therefore, to evaluate the impact of frequency on the attack feasibility, we tested three different ranges of ambient audio frequencies collected by controlled experiments where we set up the ambient noise surrounding recording devices falling into different categories. *Low ambient audio* (frequency less than 100 Hz); *Medium ambient audio* (frequency in the human audible range, at around 500 Hz); *High ambient audio* (frequency 5,000 Hz or more).

We used the dataset *Data-PerCom* for ambient audio to build the classification model (F1 of 0.86 and FPR of 9.3 percent). The 100 samples we collected via audio streaming channels in dataset *Data-TMC-audio* are fed to the classifier for prediction. Table 1 presents the FPR of non co-presence detection under the streaming attacks over Wi-Fi and cellular data channels. The results indicate that the attacker (1) has a higher chance of success using the Wi-Fi channel and (2) could be thwarted when either the ambient audio at $\mathcal{P}$ is low frequency or if the ambient audio at $\mathcal{V}$ is high frequency.

This simple streaming attack with commodity devices shows that the audio-only system is highly vulnerable to relay attacks, especially via the Wi-Fi channel. The attack has very high success rate regardless of hardware variations and network delays inherent to streaming. However, an attacker can succeed only when relaying ambient audio from a higher frequency acoustic environment to a similar or lower frequency acoustic environment, such that, the higher frequency dominates the lower frequency, and makes $\mathcal{V}$ falsely record $\mathcal{P}$'s ambient noise instead of the real "localized" ambient noise.

TABLE 1
Relay Attack Success Rate (FPR) for Audio Streaming via Wi-Fi and Cellular Networks

Acoustic relaying environments ($\mathcal{P}$ freq $\rightarrow$ $\mathcal{V}$ freq)	Wi-Fi	Cellular
High → Medium	100%	40%
High → Low	100%	20%
Medium → Medium	100%	0%
Medium → Low	100%	60%
Low → Low	20%	0%
Others	0%	0%

The audio features we used, i.e., the ones proposed in [20], are not sensitive to time synchronization. This is effective in terms of co-presence detection (i.e., results in very low FNR). However, as we can see from our experiments, these features also enable the attacker to succeed in the relay attack with a very high chance. Other audio features, such as the ones proposed in [27], require tight synchronization and could be more resistant to relaying. Unfortunately, because of their high sensitivity to synchronization, these features did not perform well in the benign (co-presence) case based on our experiments (i.e., resulted in high FNR).

4.2.2 Audio-Radio System

Truong et al. [31], evaluated the performance of an audio-radio system against a unidirectional, single-modality attacker. They showed that while the system achieves good performance (F1 of 0.98) and high security (FPR of 2.0 percent), a contextual attacker could increase the FPR: from 0.18 to 65.8 percent (manipulating W), from 1.1 to 1.2 percent (B); from 1.62 to 3.01 percent (audio). Now, we will analyze the same system and same dataset *Data-PerCom* against a bi-directional (for radio), multi-modality attacker. To model the attack, in each run, the non co-presence samples in the test data were transformed as below.

Audio. Because raw audio data is additive, and one-side context manipulation for audio is tested, an adversary can be modelled by replacing $\mathcal{V}$ side audio (X_a) to be the sum of its own ambient audio and $\mathcal{P}$ side audio ($X_a + X_b$).

Radio (B and W). In [31], the set of radio records from two devices A and B are defined as: $S_a = \{(m_i^{(a)}, s_i^{(a)}) \mid i \in \mathbb{Z}_{n_a-1}\}$, and $S_b = \{(m_i^{(b)}, s_i^{(b)}) \mid i \in \mathbb{Z}_{n_b-1}\}$, where (m, s) with m is an identifier and s is associated signal strength of a beacon; n_a and n_b denote the number of different beacons (i.e., Wi-Fi access points or Bluetooth devices). The both-sides contextual adversary can be modeled by replacing S_a with $S_a \cup \{(m, s)\ \forall (m, s) \in S_b, m \notin S_a^{(m)}\}$, and S_b with $S_b \cup \{(m, s)\ \forall (m, s) \in S_a, m \notin S_b^{(m)}\}$.

We considered two approaches of fusing sensor data against bi-directional relay attacks and showed which of them is more suitable for resisting against the presence of contextual attackers.

Table 2 (columns 1 and 2) presents the analysis results of training model combining all three audio-radio modalities (Au, B and W) and testing with different attacks. Zero-modality attack shows the very low FPR with both fusion methods. The FNR for decisions-fusion is higher compared to that for features-fusion. For features-fusion, the results are aligned with the ones reported in [31].

In single-modality attack, manipulating Wi-Fi, the dominant feature, results in a very high success rate with features-fusion. The results change when decisions-fusion was

TABLE 2
This Table Shows How Well Different Types of Context-Manipulating Attackers Perform in Scenarios Employing Different Types of Fusion

	Audio-Radio		Physical		Audio-Radio-Physical		
	Fuse-F (1)	*Fuse-D-S* (2)	*Fuse-F* (3)	*Fuse-D-S* (4)	*Fuse-F* (5)	*Fuse-D-S* (6)	*Fuse-D-M* (7)
Zero-modality	2.0% (FNR: 1.4%) (F1: 0.977)	2.0% (FNR: 12.0%) (F1: 0.925)	7.5% (FNR: 3.9%) (F1: 0.928)	13.0% (FNR: 14.5%) (F1: 0.861)	3.0% (FNR: 0.0%) (F1:0.990)	27.1% (FNR: 0.3%) (F1: 0.923)	6.9% (FNR: 0.0%) (F1: 0.980)
Single-modality	{Au }: 3.0% {B }: 2.7% {W }: 99.8%	{Au }: 3.0% {B }: 9.0% {W }: 8.0%	{T }: 8.3% {G }: 11.9% {H }: 15.3% {Al }: 55.1%	{T }: 17.0% {G }: 20.0% {H }: 24.4% {Al }: 33.1%	{Au }: 87.7% {B }: 100% {W }: 12.3% {Al }: 5.4% {G }: 5.9% {H }: 3.4% {T }: 3.4%	{Au }: 45.3% {B }: 45.8% {W }: 44.8% {Al }: 37.9% {G }: 29.6% {H }: 29.1% {T }: 31.5%	{Au }: 36.9% {B }: 36.9% {W }: 35.0% {Al }: 6.9% {G }: 6.9% {H }: 6.9% {T }: 6.9%
Multi-modality	{Au, B }: 3.6% {B, W }: 99.8% {Au, W }: 100% {Au, B, W }: 100%	{Au, B }: 96.0% {Au, W }: 96.0% {B, W }: 100% {Au, B, W }: 100%	{G, T }: 13.9% {G, H }: 15.7% {H, T }: 29.6% {G, H, T }: 31.1% {Al }$\cup\{\widetilde{X}\}$: 64.7-100%	{G,T }: 40.1% {H, T }: 41.9% {Al, T }: 50.6% {G, H }: 57.5% {Al, H }: 61.2% {Al, G }: 65.5% rest: 100%	{B }$\cup\{\widetilde{X}\}$: 100% {Au }$\cup\{\widetilde{X}\} > 74.9\%$ $\{\widetilde{X}\}\backslash$ {Au, B }: $< 12.3\%$	{2 sensors}: 32.0-75.4% {3 sensors}: 37.4-97.5% {4 sensors}: 97.5-100% rest: 10%	{Au, B }$\cup\{\widetilde{X}\}$: $> 97.5\%$ {Au, W }$\cup\{\widetilde{X}\}$: $> 88.2\%$ {Al, G, H, T }: 9.9% {B, W }: 36.9% rest: 6.9-87.7%

The horizontal blocks refer to increasingly powerful attackers with the ability to manipulate zero, one, or two context sensor modalities. Values in the table are FPRs with/without different contextual attacks in various audio/radio/physical systems. Notations: *Sets of manipulated sensors are put inside curly braces {}. $\{\widetilde{X}\}$ denotes an arbitrary set of sensor modalities. Fuse-F: features-fusion, Fuse-D-S: decisions-fusion from single modalities, Fuse-D-M: decisions-fusion from subsets of modalities.* Result highlights*: Manipulation of sensor modalities, especially multiple of them, can significantly reduce security (increase FPR) in most cases. Decisions-fusion can help improve security when dominant sensors are manipulated, but it may reduce usability (increase FNR).*

applied. In such case, manipulating any single sensor, even the most powerful one, does not significantly degrade the overall security. The FPR in case W was manipulated decreases from 99.8 percent (features-fusion) down to 8 percent (decisions-fusion). We recall that the performance difference of audio and radio sensors is not large (as reported in [31], F1 ranges from 0.857 for Au to 0.989 for W). This explains why decisions-fusion reduces the overall performance slightly (F1 reduces from 0.977 to 0.925) in case of zero-modality attack but significantly improves the security under a single-modality attack. The security is very low in multi-modality attack, and neither of the fusion approaches could restore the security level when majority of the sensors are under attacker's control. We earlier argued that audio and radio modalities can be manipulated simultaneously.

4.2.3 Physical System

Shrestha et al. [28] introduced four physical modalities (Al, H, G, and T) for co-presence detection. The performance of the features-fusion based classifier trained with their dataset is good (F1 of 0.957, FPR of 5.81 percent) against a zero-modality adversary.

Based on our attacks against physical modalities (Section 3.3), we consider an adversarial model where an attacker can manipulate the physical context on one side (unattended verifier) to match the sensor readings at the other side (prover). To model this attack, using the dataset *Data-FC*, we transformed all non co-presence samples in the test set to the "attack" value (distance 0). The distance is set to 0 as data collection in [28] was done by a single device at a given point of time, hence, no hardware effect or calibration error was taken into account. The non co-presence class in the dataset is about 18 times larger than co-presence class. To correct this imbalance, we applied the same under-sampling as in [28]: we divided the non co-presence samples into 19 subsets, ran several rounds of cross validation taking 10 subsets in each round and aggregated the results in the end. In addition to the features-fusion employed in [28], we tested the decisions-fusion similar to our audio-radio system analysis in the previous section.

Table 2 (columns 3 and 4) shows our analysis results. The system performance in zero-modality attack is well-aligned with the one reported in [28]. As in [28], among four physical modalities, Al performs the best. Consequently, manipulating only Al degrades the security vastly with features-fusion (FPR increases to over 50 percent). Decisions-fusion in general brings lower security and lower performance/usability in zero-modality attack and single-modality attack. However, it avoids the dominance of sole sensor in case the attacker can control such sensor (Al in this case). Decisions-fusion can also help improve security against a multi-modality attacker who manipulates Al along with other sensors. Compared to audio-radio system, in physical system, attacking each single modality results in higher success rate.

4.2.4 Audio-Radio-Physical System

Unlike the dataset for physical sensors [28] which was collected from one device at a time only, the dataset *Data-TMC* that we collected for this work contains data from pairs of devices, and therefore hardware variance and calibration errors between co-presence device sensors need to be taken into account. When we try to model the contextual attack on given sensor(s), distance 0 does not ensure that the attack will succeed. As the classifier is trained with data which may contains noise, we compute the mode of the histogram for distance values for the co-presence samples. As the data aggregated is from two participants, histograms of distance values are not uninomial but multinomial. Multinomial distribution implies several modes. For each physical sensor, we choose a mode value and assign it as the distance value. The mode values for Al, G, H and T are 13.54, 0.3, 6.61 and 0.153, respectively. As the manipulation by replacing the radio data at both sides has to be identical, the distance features for radio sensors are set to 0.

Table 2 (columns 5, 6 and 7) reports our analysis results with different fusion methods. Under zero-modality attack, features-fusion performs the best while decisions-fusion from single modalities performs the worst. Features-fusion uses all possible features for training so that the classifier can be built based on the best features or best combination of features (B and Au with our current dataset). Thus, it returns the best results (in the absence of context manipulation) compared to any other ways of fusing sensor data. Decisions-fusion based on single modalities lets the worst sensors being able to contribute to the voting scheme, thus bringing down the overall performance. This is the case in our dataset where radio sensors and audio sensor perform better than physical sensors. Note that if all sensors perform equally well, features-fusion and decisions-fusion would not differ much. Decisions-fusion from subsets of sensors has a moderate performance, worse than features-fusion but better than decisions-fusion from single modalities. This hybrid approach avoids mis-learning as in the case of using a single modality only.

Let us now assess the security of this co-presence detection system when any single modality is controlled by the attacker. Depending on how sensors are fused, the impact of manipulated sensor varies. In features-fusion, as the classifier decision relies on the best features of dominant sensors, the FPR increases drastically when such sensors are manipulated (i.e., Au or B in our dataset). In contrast, when weaker sensors (physical or W) are manipulated, it has a relatively small impact on the security as the resulting FPR increases a bit compared to a zero-modal attack (especially for W). Decisions-fusion reduces attacker success rate when single sensor is manipulated, for example, FPR of manipulating B decreases from 100 percent (features-fusion) to 36.9 percent (decisions-fusion). Recall that manipulating single sensor is not difficult as we demonstrated in Section 3.

An attacker has the highest chance to succeed if he can control the dominant sensors or a subset of sensors that contain the dominant sensors. In such case, the success rate could reach 100 percent with only one single dominant sensor (i.e. B in our dataset) if the system uses features-fusion or with majority dominant sensors (i.e., Au and B). In most cases, attacking the set of weak sensors (e.g., {Al, G, H, T }) does not impact the security much, except when system uses decisions-fusion from single modalities.

5 Discussion, Potential Mitigations, and Future Work

Classifier Analysis. In our attack analysis, we used classification techniques as implemented by the original contextual

systems [28], [31]. We used different machine learning techniques (Decision Tree and Random Forest) with ten-fold cross validation implemented in Scikit-learn [26] as mentioned in Section 4.1. When we train/build the classifier model using benign training datasets, the classifier chooses the thresholds such that it yields the best result with high F1 score and low FNRs/FPRs. This model works best in case there is no context manipulation attack. Our results under this setting are inline with that reported in original systems [28], [31]. For evaluating the performance of our context manipulation attacks, we used the exact same classifier models since they demonstrated the best performance in the setting with no context manipulation attacks. The classifier model may have performed better if the classification thresholds were modified such that the model is more robust towards False Positive Rates, i.e., result in lower FPR. However, it is important to note that changing the thresholdization setting to yield low FPR may result in high FNR and thereby decrease the overall F1 score which leads to decreasing the usability of the system. Since usability is an important attribute of the contextual co-presence detection systems, our analysis represents a valid attack setting that serves to demonstrate the weaknesses of, and potential fixes to, the existing usable systems.

Reducing Attack Success with Decisions-Fusion. In the previous section on analysis of an audio-radio-physical system, we showed that decisions-fusion reduces attack success rates in cases where the minority of the sensors are manipulated. However, this may come at the cost of higher FNR which represents the usability of co-presence systems. Decisions-fusion from single sensors improves security when individual sensors perform well. However, it increases the attack success rate for weak sensors as they equally contribute to the voting. For example, in the context of the audio-radio-physical system, attacking weak sensors such as H or G brings relatively high success rate compared to features-fusion. Decisions-fusion from subsets of sensors reduces the FPR in general especially when dominant sensors are controlled by the attacker.

*Other Potential Countermeasures.*Typically, during the authentication/deauthentication process, the prover moves nearer to/farther away from the verifier. In this case, the radio signals changes gradually, i.e., if prover and verifier move towards APs, then new APs will be shown, or their signal strengths will continuously grow, while if they move further away from APs, their strengths will decrease or the APs will not be visible at all. If the verifier or prover device detects much more APs (or Bluetooth devices) nearby all of a sudden, it probably indicates a radio manipulation attack. The system can be made aware of such situations.

We noticed that when the verifier is in an environment which has high frequency noise, an attacker tends to fail with audio streaming. This can be used to design an active defense mechanism such that whenever audio contextual information is requested, the verifier can emit a high frequency, potentially random audio sounds. This audio signal can be for a short duration, and does not need to be loud (not high amplitude). As a result, the chances of attacker succeeding in a relay attack could be reduced.

When an authentication request has been initiated or finished, the user can be passively notified at both devices. Passive notification can be a flashing of LED light or beep on the prover device (key or phone). Hence, even if the verifier device is left unattended, user may notice at the end of the prover device that someone is trying to authenticate to the verifier, or has authenticated on user's behalf. Whether or not users would actually pay attention to such notifications should be subject to further scrutiny. It may help reduce the risk of context-manipulation relay attacks.

*Limitations.*There are certain limitations of our work. Our attack experiments were done in lab settings under normal environmental conditions. It may require more effort or sophistication from the adversary in real-life settings, under an arduous environment, to make these attacks work. For example, using ice cubes to decrease temperature below freezing point may not work, or using a hair dryer in a very cold outdoor environment may make it difficult to increase the temperature to a desired value, accessing a sensor when the sensor placement is unknown or access to sensor is difficult, and so on. Nevertheless, our work demonstrates that the designers of contextual security systems should consider such attacks seriously while developing such solutions. The dataset we used for analyzing the attacks in audio-radio-physical system is relatively small. It was collected from limited number of devices. It might not represent all possible scenarios and environments. However, it was sufficient to demonstrate the impact of attacks and defensive solutions. It gave insights for better understanding of the contextual co-presence detection system and possible defenses to improve security against different contextual attacks. Further work may be needed to collect and analyze a larger scale dataset to evaluate this system. The decisions-fusion from subsets of sensors seems to be the most appropriate solution for improving security against context manipulation attacks. However, we have analyzed it only with three subsets: acoustic (Au), radio (B, W) and physical subsets (Al, G, H, T). In design of a real system in the future, we would like to test different subsets combinations to find the best candidate for fusion.

6 Conclusions

Contextual co-presence detection has been shown to be a very promising relay attack defense in many mobile authentication settings suitable for off-the-shelf, sensor-equipped devices. We presented a systematic assessment of co-presence detection in the presence of a context-manipulating attacker. Our work suggests that tampering with the context can be achieved with simple yet effective strategies, and the security offered by co-presence detection is therefore weaker than previously believed. We also suggested potential countermeasures (e.g., decisions-fusion based machine learning classification technique, that may be used to strengthen the security of co-presence detection against a multi-modality attacker. Some of these countermeasures may require a thorough future investigation, which we plan to pursue.

Appendix A
Increasing the Temperature When the Attacker does not Know VS's Location

An attacker who does not know the location of VS will try to keep the FS as close as possible and perform the attack activity. We placed the FS 10 cm apart from the VS and performed

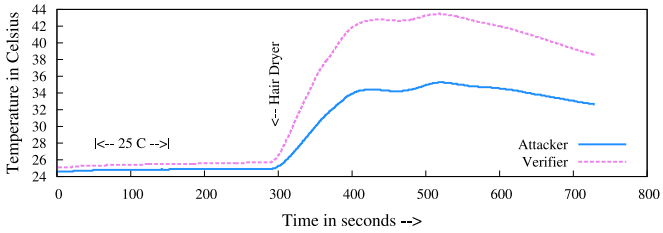


Fig. 13. Increasing the temperature; location of VS unknown to the attacker; VS is 10 cm closer to the hair dryer than FS; the attacker trying to increase temperature to 35 °C.

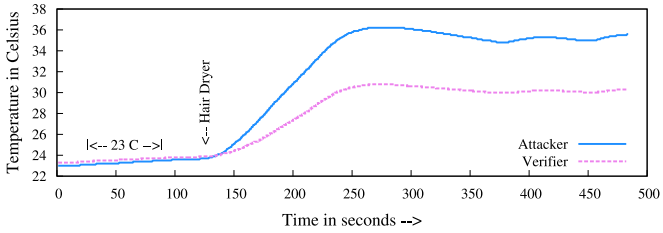


Fig. 14. Increasing the temperature; location of VS unknown to the attacker; FS is 10 cm closer to the hair dryer than VS; the attacker trying to increase temperature to 35 °C.

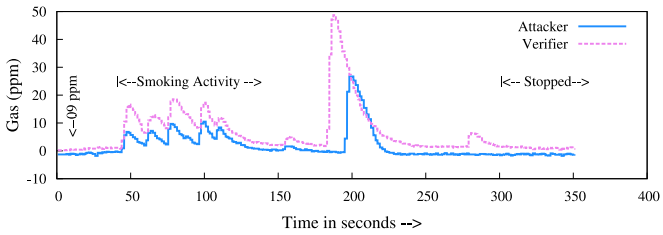


Fig. 15. Effect of cigarette in CO level; increasing the gas content to an arbitrary value and waiting to decrease to desired level.

experiment in two settings. In the first setting, the hair dryer is closer to VS as shown in Fig. 13, and in the latter setting, the hair dryer is closer to FS as shown in Fig. 14.

Appendix B
Increasing the CO Gas Level

We effectively manipulated the CO gas sensor using cigarette and car exhaust. The increase in the gas level due to the activity is abrupt when CO is blown onto the sensors, however, it takes a while for the sensors to fall back to normal readings. This provides an enough time window for the attacker as depicted in Figs. 15 and 16.

Appendix C
Increasing the Altitude Using a Car Vacuum

As an alternative to air pump, we tried a portable car vacuum cleaner for inducing an altitude increase. When we hovered the vacuum cleaner pipe around the sensors, it did not have any effect. However, when we put the pipe just on top of the sensor, it increased the altitude by 10-11 meters as shown in Fig. 17. An attacker can adjust the altitude to a desired level by changing the power level of the vacuum cleaner, similar to the air pump manipulation. The earlier part of the Fig. 17 shows a little fluctuation in altitude when we hovered the pipe around the sensors while the later

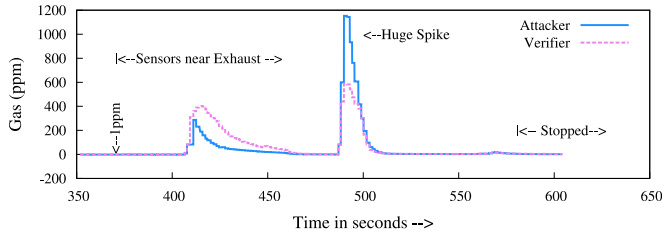


Fig. 16. Effect of car exhaust in CO level; increasing the CO gas level to arbitrary value and wait to decrease to desired level.

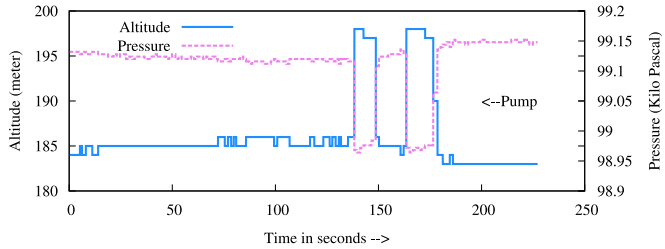


Fig. 17. Using a car vacuum cleaner to reduce pressure around the sensor and increase the altitude.

spikes clearly show that there was an increase of almost 10 meters when the pipe was touched to the sensors. A video demo of our attack has been uploaded to YouTube [3] to show the effect of portable car vacuum cleaner on the pressure/altitude sensors.

Acknowledgments

The authors would like to thank *TMC* reviewers for their useful feedback on the paper. This work was supported in part by US National Science Foundation grants (CNS-1526524 and CNS-1547350), the Academy of Finland (projects 274951 and 309994), and Intel (ICRS-SC).

References

[1] 3db Technologies GmbH, "Online." [Online]. Available: https://www.3db-access.com/, Last accessed on: June 4, 2018.
[2] About Skype - What is Skype, "Skype." [Online]. Available: http://www.skype.com/en/about/, Last accessed on: June 4, 2018.
[3] Car Vacuum Activity, "Youtube." (2014, July 29). [Online]. Available: https://www.youtube.com/watch?v=vN_ZBi_rmJc
[4] Electric Air Pump, "Walmart." [Online]. Available: http://www.walmart.com/ip/33563196, Last accessed on: June 4, 2018.
[5] Hair Dryer, "Conair." [Online]. Available: http://infiniti.conair.com/catalog.php?pcID=47&products_id=232, Last accessed on: June 4, 2018.
[6] Hair Dryer Activity, "Youtube." (2014, July 29). [Online]. Available: https://www.youtube.com/watch?v=3QG79NH0qjA
[7] Scapy, "Online." [Online]. Available: http://www.secdev.org/projects/scapy/, Last accessed on: June 4, 2018.
[8] Ziploc and Air Pump Attack Video, "Youtube." (2014, July 29). [Online]. Available: https://www.youtube.com/watch?v=Fy2F8rY6bzw
[9] M. R. Bloomberg and S. Cassano, "Fire safety education." [Online]. Available: http://www.nyc.gov/html/fdny/pdf/safety/fire_safety_education/2010_02/08_smoke_and_carbon_monoxide_alarms_english.pdf, Last accessed on: June 4, 2018.
[10] S. Brands and D. Chaum, "Distance-bounding protocols," in *Proc. Int. Conf. Theory Appl. Cryptographic Techn.*, 1993, pp. 344–359.
[11] R. Clarke, *Thefts of and from Cars in Parking Facilities*. US Department of Justice, Office of Community Oriented Policing Services, Washington, DC, 2002, https://ric-zai-inc.com/Publications/cops-w0691-pub.pdf
[12] M. D. Corner and B. D. Noble, "Zero-interaction authentication," in *Proc. 8th Annu. Int. Conf. Mobile Comput. Netw.*, 2002, pp. 1–11.

[13] DD-WRT.com, "www.dd-wrt.com | Unleash Your Router." [Online]. Available: http://www.dd-wrt.com/site/index, Last accessed on: June 4, 2018.
[14] Y. Desmedt, C. Goutier, and S. Bengio, "Special uses and abuses of the Fiat-Shamir passport protocol," in *Proc. Conf. Theory Appl. Cryptographic Techn.*, 1987, pp. 21–39.
[15] D. Dolev and A. C.-C. Yao, "On the security of public key protocols," *IEEE Trans. Inf. Theory*, vol. 29, no. 2, pp. 198–208, Mar. 1983.
[16] S. Drimer and S. J. Murdoch, "Keep your enemies close: Distance bounding against smartcard relay attacks," in *Proc. 16th USENIX Secur. Symp.*, 2007, Art. no. 7.
[17] Forbes, "Why the most expensive cars are the easiest to steal." (2014, Nov. 17). [Online]. Available: http://www.forbes.com/sites/marcwebertobias/2015/11/17/why-the-most-expensive-cars-are-the-easiest-to-steal/#35a3ae9c205d
[18] A. Francillon, B. Danev, S. Capkun, "Relay attacks on passive keyless entry and start systems in modern cars," in *Proc. Netw. Distrib. Syst. Secur. Symp.*, 2011.
[19] L. Francis, G. P. Hancke, K. Mayes, and K. Markantonakis, "Practical NFC peer-to-peer relay attack using mobile phones," in *Proc. Int. Workshop Radio Freq. Identification: Secur. Privacy Issues*, 2010, pp. 35–49.
[20] T. Halevi, D. Ma, N. Saxena, and T. Xiang, "Secure proximity detection for NFC devices based on ambient sensor data," in *Proc. Eur. Symp. Res. Comput. Secur.*, 2012, pp. 379–396.
[21] N. Karapanos, C. Marforio, C. Soriente, and S. Capkun, "Soundproof: Usable two-factor authentication based on ambient sound," in *Proc. USENIX Conf. Secur. Symp.*, 2015, pp. 483–498.
[22] Z. Kfir and A. Wool, "Picking virtual pockets using relay attacks on contactless smartcard," in *Proc. 1st Int. Conf. Secur. Privacy Emerging Areas Commun. Netw.*, 2005, pp. 47–58.
[23] A. Levi, E. Cetintas, M. Aydos, C. K. Koc, and M. Caglayan, "Relay attacks on bluetooth authentication and solutions," in *Proc. Int. Symp. Comput. Inf. Sci.*, 2004, pp. 278–288.
[24] M. Miettinen, S. Heuser, W. Kronz, A.-R. Sadeghi, and N. Asokan, "ConXsense: Automated context classification for context-aware access control," in *Proc. 9th ACM Symp. Inf. Comput. Commun. Secur.*, 2014, pp. 293–304.
[25] A. Narayanan, N. Thiagarajan, M. Lakhani, M. Hamburg, D. Boneh, et al., "Location privacy via private proximity testing," in *Proc. Netw. Distrib. Syst. Secur. Symp.*, vol. 11, 2011.
[26] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine learning in python," *J. Mach. Learn. Res.*, vol. 12, pp. 2825–2830, Nov. 2011.
[27] D. Schurmann and S. Sigg, "Secure communication based on ambient audio," *IEEE Trans. Mobile Comput.*, vol. 12, no. 2, pp. 358–370, Feb. 2013.
[28] B. Shrestha, N. Saxena, H. T. T. Truong, and N. Asokan, "Drone to the rescue: Relay-resilient authentication using ambient multi-sensing," in *Proc. 18th Int. Conf. Financial Cryptography Data Secur.*, 2014, pp. 349–364.
[29] R. Siegwart, I. R. Nourbakhsh, and D. Scaramuzza, *Introduction to Autonomous Mobile Robots*. Cambridge, MA, USA: MIT Press, 2011.
[30] N. O. Tippenhauer, K. B. Rasmussen, C. Pöpper, and S. Čapkun, "Attacks on public WLAN-based positioning systems," in *Proc. 7th Int. Conf. Mobile Syst. Appl. Serv.*, 2009, pp. 29–40.
[31] H. T. T. Truong, X. Gao, B. Shrestha, N. Saxena, N. Asokan, and P. Nurmi, "Comparing and fusing different sensor modalities for relay attack resistance in Zero-Interaction Authentication," in *Proc. IEEE Int. Conf. Pervasive Comput. Commun.*, 2014, pp. 163–171.
[32] A. Varshavsky, A. Scannell, A. LaMarca, and E. De Lara, "Amigo: Proximity-based authentication of mobile devices," in *Proc. Int. Conf. Ubiquitous Comput.*, 2007, pp. 253–270.

Babins Shrestha received the PhD degree in computer and information sciences from the University of Alabama at Birmingham (UAB), in 2016. He is a senior information security analyst with VISA Inc., Austin, Texas. He works in the Cyber Defense Team at VISA where he facilitates the detection and prevention of advanced emerging cyber threats through the development of new tools, techniques, automation, and sensor enrichment. He worked with Dr. Nitesh Saxena as a member of the SPIES (Security and Privacy In Emerging computing and networking Systems) Lab at UAB. He has several journal and conference papers published corresponding to his work. He also has four years of work experience in web application development at Verisk Information Technologies, previously known as D2Hawkeye Services where he worked as a senior software engineer.

Nitesh Saxena is an associate professor of computer and information sciences with the University of Alabama at Birmingham (UAB), and the founding director of the Security and Privacy in Emerging Systems (SPIES) Group/Lab. He works in the broad areas of computer and network security, and applied cryptography, with a keen interest in wireless and mobile device security, and the emerging field of usable security. His current research has been externally supported by multiple grants from the US NSF and NIJ, and by gifts/awards/donations from industry, including Google (two Google Faculty Research awards), Cisco, Comcast, Intel, Nokia, and Research in Motion. He has published more than 110 journal, conference, and workshop papers, many at top-tier venues in computer science, including: the IEEE Transactions, ISOC NDSS, ACM CCS, ACM WWW, ACM WiSec, ACM ACSAC, ACM CHI, ACM Ubicomp, IEEE Percom, IEEE ICME, and IEEE S&P. On the educational/service front, he currently serves as the director and principal investigator for the UAB's Scholarship for Service (SFS) program and a co-director for UAB's MS program in computer forensics and security management. He serves as an associate editor of the *Flagship Security Journals*, the *IEEE Transactions on Information Forensics and Security* (*TIFS*), and the *Springer's International Journal of Information Security* (*IJIS*). His work has received extensive media coverage, for example, at NBC, MSN, Fox, Discovery, ABC, Bloomberg, MIT Tech Review, ZDNet, ACM TechNews, Yahoo! Finance, *Communications of ACM*, Yahoo News, CNBC, Slashdot, Computer World, Science Daily, and Motherboard. He is a member of the IEEE.

Hien Thi Thu Truong received the PhD degree in computer science from INRIA and the Université de Lorraine, France, in 2012. From 2013 to 2016, she conducted her postdoctoral research with the University of Helsinki, Finland. Since 2016, she has been a research scientist in security at NEC Laboratories Europe in Germany. She has more than 10 years of R&D experience gained while working in Vietnam, Japan, France, Finland, and Germany. Her research interests cover the design and analysis of system security; privacy and trust for distributed systems; and mobile security. Recently, she focuses on applying machine learning and data analysis techniques to solve usable security problems. She is a member of the IEEE.

N. Asokan is a professor of computer science with Aalto University where he directs the Helsinki-Aalto Center for Information Security – HAIC. He is a fellow of the IEEE and an ACM distinguished scientist.