Contextual Proximity Detection in the Face of Context-Manipulating Adversaries

Babins Shrestha*, Nitesh Saxena*, Hien Thi Thu Truong[†], N.Asokan[‡]
*University of Alabama at Birmingham, USA, Email: {babins,saxena}@uab.edu

†University of Helsinki, Finland, Email: hien.truong@cs.helsinki.fi

‡Aalto University and University of Helsinki, Finland, Email: asokan@acm.org

Abstract—Contextual proximity detection (or, co-presence detection) is a promising approach to defend against relay attacks in many mobile authentication systems. We present a systematic assessment of co-presence detection in the presence of a context-manipulating attacker. 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. Second, based on these capabilities, we show that an attacker who can manipulate the context gains a significant advantage in defeating context-based co-presence detection. For systems that use multiple sensors, we investigate two sensor fusion approaches based on machine learning techniques – features-fusion and decisions-fusion, and show that both are vulnerable to contextual attacks but the latter approach can be more resistant in some cases.

I. 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 V). Traditional cryptographic authentication typically involves a challengeresponse protocol whereby \mathcal{P} proves the possession of the key K that it pre-shares with V by constructing a valid response to a random challenge sent by 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 [10] 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., [14], [15]), 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} [9]. Using this protocol, \mathcal{V} can verify whether \mathcal{P} is within a close proximity thereby detecting the presence of relay attacks [14], [15]. However, distance bounding systems may not be currently feasible on 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 [17], [27], [21], [16]. This defense leverages the notion of "context" derived from onboard device sensors based on which $\mathcal{P}\text{-}\mathcal{V}$ 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 *contextual co-presence detection*, including audio [17], radio (*WiFi* [27], *Bluetooth* [26] and *GPS* [16], and the physical environment (*temperature*, *humidity*, *gas* and *altitude/pressure*) [24]. Many single modalities, such as audio and WiFi, 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 [26], and combination of physical sensors [24], has been shown to further reduce false negatives and false positives.

Our Contributions: The focus of prior work on contextual copresence 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 [13]). In this paper, we are extending this model to the realm of a context-manipulating attacker. There are two main parts and contributions of the paper, as summarized below:

1. Novel Context Manipulation Attacks: 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. We demonstrate attacks against a variety of modalities studied in prior work including audio, radio (Bluetooth/WiFi), and physical (temperature, humidity, gas and altitude).

Our work is the first to consider context-manipulation attacks against audio and physical modalities. In particular, we demonstrate how an attacker in close proximity of the sensors can successfully manipulate the physical environment "seen" by these sensors, without the need to manipulate the global surrounding environment or compromise the devices/sensors themselves. Our attacks are described in Section III.

2. Co-Presence Detection with Context Manipulations: Based on the above manipulation capabilities, we comprehensively examine and quantify the advantage a multi-modality attacker, who can manipulate multiple sensor modalities simultaneously, can have in defeating co-presence detection over a zero-modality attacker (one studied in prior work). To accomplish this, we re-orchestrated the co-presence detection approaches based on machine learning techniques in audio-only [17], audio-radio [26], physical [24] 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 II quantifies the attacker success rates).

For systems that use multiple modalities, we investigate two different sensor fusion approaches – *features-fusion* (proposed in [26]) 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 IV.

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 as difficult as previously believed, and the security offered by contextual co-presence detection is therefore weaker. 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 V). 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 [20] with respect to context-manipulating adversaries.

II. BACKGROUND AND MODELS

A. Relay Attacks and 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 [13]: 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 $\mathcal V$.

The attacker could take the form of a "ghost-and-leech" [18] 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 V to conclude that P is in proximity. This is an attack applicable to zero-interaction authentication systems. A similar attack applies to proximity-based payment systems [12], [14].

Co-presence detection schemes aim to address such relay attacks. Figure 1(a) shows a typical system model of an authentication/authorization protocol using contextual co-presence, adapted from [26]. In this defense, \mathcal{P} (respectively \mathcal{V}) preshares a key K(K') with a "comparator" 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 C. 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, C concludes that P and Vare co-present. When \mathcal{C} is integrated with \mathcal{V} , K' is not used. Figure 1(b) 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 [17], radio context including WiFi – W and Bluetooth – B [26], and physical environmental attributes, temperature – T, humidity – H, concentration of gases – G and altitude – Al [24].

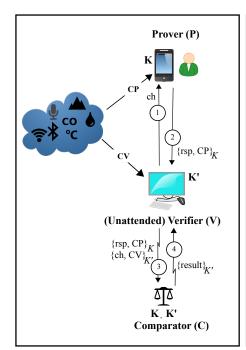
B. Threat Model for a Single-Modality Contextual Attacker

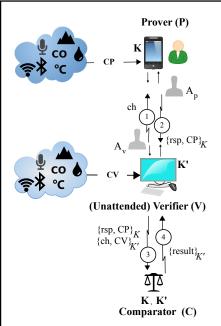
Our focus is on a context-manipulating attacker against co-presence detection (going beyond a Dolev-Yao attacker). Truong et al. [26] briefly explored the problem of characterizing such a contextual attacker. They only consider an attacker who is capable of manipulating a single sensor modality at a 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:

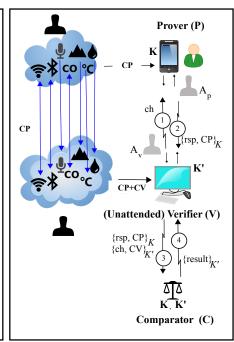
- A_p , A_v can measure the context information that \mathcal{P} , \mathcal{V} would sense, respectively.
- A_v can fool \mathcal{V} into sensing the context information A_v chooses. Specifically A_v can receive context information from A_p and reproduce it near \mathcal{V} .
- $A_v(A_p)$ cannot suppress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

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

We reiterate that this model assumes that V is **unattended**, and as such, the attacks we will present in this paper are







(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 [26] 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.

based on this assumption. Vehicles parked in underground parking lots/decks represent an apt example of unattended verifiers. Relay attacks against such vehicles have already been demonstrated in [15]. 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 vulnerable to our attacks.

III. CONTEXT MANIPULATION ATTACKS

A. 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 II that our threat model allows the attacker to add to the ambient audio at 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 V's side can go undetected is valid since V may be unattended in many scenarios (as our model in Section II 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 [1]. We use the features and classifier described in prior work [17]. Our results are presented in Section IV-B.

B. Manipulating Radio-Frequency Sensor Modalities

Prior work suggests that manipulating the radio context is possible in general. The work presented in [25] describes attacks on a public WiFi-based positioning system. They used a Linux laptop as an Access Point (AP) with the Scapy packet manipulation program [6] to spoof WiFi APs. Similarly, spoofing bluetooth device addresses has already been demonstrated in prior work [19], [26], 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 [26] (see Section II-B) is reasonable. Furthermore, in the case of 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 WiFi spoofing ourselves, and studied how it can be used to match the WiFi context at two ends. In our experiment, we used a Linksys router (WRT54G) to create a spoofed hotspot. We flashed DD-WRT firmware [11] 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 (TX) 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 dBm 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 WiFi context matching attack becomes rather straightforward. The attacker can even have advantage in environments where number of WiFi 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.

C. Manipulating Physical Environment Sensor Modalities

As discussed in [24], 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-theshelf 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 $(\mathcal{V} \text{ or } \mathcal{P})$, 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 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 contextual 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 nontrivial. 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 (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.

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.

Increasing the Temperature: In situations where \mathcal{P} (e.g., a car key indoors) is at a higher temperature than 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 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 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 [4] had a power of 1875 watt AC. A video demonstration of our attack has been uploaded to YouTube [5].

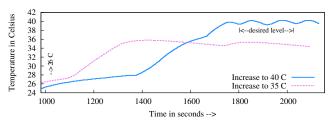


Fig. 2: Increasing T to desired level (35 °C and 40 °C)

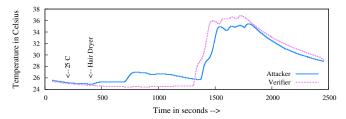


Fig. 3: VS and FS on same location; the attacker trying to increase temperature to 35 $^{\circ}C$.

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;

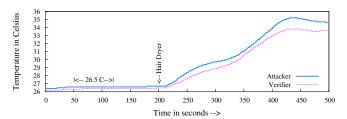


Fig. 4: Heating an area; VS and FS within a range of 15 cm; the attacker trying to increase temperature to 35 °C.

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 (Figure 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 (Figure 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.

Decreasing the Temperature: In some scenarios, it might be necessary for the attacker to reduce the temperature recorded by V (e.g., when P is indoors and 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 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 few trials within a variance of +/-0.3 °C as shown in Figure 5.

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

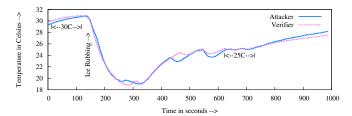


Fig. 5: Decreasing temperature with an ice cube; the attacker trying to decrease to 25 °C.

move the hot coffee cup nearer to, and farther away, from the sensors to control the humidity level. Using this strategy, we were able to increase the humidity by 10%, i.e., from normal humidity of 55% to 65% (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% 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.

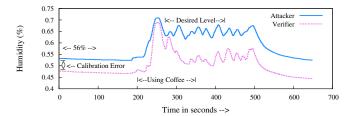


Fig. 6: Increasing humidity with hot coffee; the attacker trying to increase to 65%.

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).

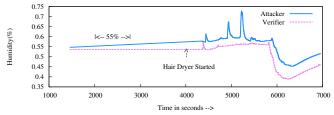


Fig. 7: Decreasing humidity with hair dryer such that VS and FS are at same location; the attacker trying to decrease to 50%.

3) Gas Manipulation: Following prior work [24], 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

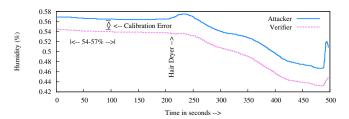


Fig. 8: Decreasing humidity with hair dryer such that VS and FS are within a range of 15 cm; the attacker trying to decrease to 50%.

II-B), 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 [8]. 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.

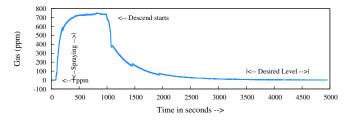


Fig. 9: Effect of aerosol spray in CO level; increasing the CO gas level to arbitrary value and wait to decrease to desired level.

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 ppm 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.

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 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 [3] 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 C).

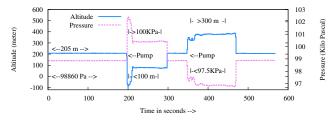


Fig. 10: Using an air pump to change pressure to the sensors wrapped inside a Ziploc bag by pumping air in and out.

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 [7].

D. Manipulating Multiple Sensor Modalities Simultaneously

As demonstrated by prior work [17], [26], [24], 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 driesup 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.

IV. PERFORMANCE OF CO-PRESENCE DETECTION SYSTEMS UNDER CONTEXT MANIPULATION

In light of the attacks presented in Section III, we first extend the rudimentary context attacker model from [26] as follows:

- We allow multi-modality attackers who can simultaneously control multiple sensor modalities, in addition to the singlemodality attacker of [26].
- We assume that a context attacker can manipulate radio contexts in both directions. The same assumption applies to gas sensors in light of our aerosol spray attack.

A. Analysis Methodology

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 audio-radio system [26] used a dataset to evaluate resistance against single-modality attackers. The physical system [24] 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 a set of new audio relaying experiments to collect a dataset to evaluate the performance of contextual co-presence detection based on audio. Furthermore, we collected a new dataset corresponding to the audio-radio-physical system (which was not considered in prior work).

When evaluating prior systems, we used the same classification techniques used in the original evaluations (Decision Tree and Random Forest), implemented in Scikit-learn [22]. 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 to "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. The decision of co-presence or non co-presence is made one-time only based on the output of the prediction model. This is the approach followed in prior work [26], [24].
- 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 based on majority voting to produce a final decision. This is a new approach we are exploring in this paper. 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.

B. Audio-Only System

Halevi et al. [17] proposed the use of (only) audio for copresence detection. Their work showed that audio is a good ambient context resulting in 100% accuracy and 0% False Positive Rate (FPR). To assess how an attacker can manipulate ambient audio via the streaming attack (Section III-A), 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: WiFi 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 WiFi channel). The attacker devices used a Skype connection as the audio relay channel.

The audio features used in [17] are based on audio frequency. Therefore, to evaluate the impact frequency on the attack feasibility, we tested three different levels 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 5000 Hz or more).

TABLE I: Relay attack success rate (FPR) for audio streaming via WiFi and Cellular networks

Acoustic relaying environments $(P \text{ freq} \rightarrow V \text{ freq})$	WiFi	Cellular
$High \rightarrow Medium$	100%	40%
$High \rightarrow Low$	100%	20%
$Medium \rightarrow Medium$	100%	0%
$Medium \rightarrow Low$	100%	60%
$Low \rightarrow Low$	20%	0%
Others	0%	0%

We used the dataset for ambient audio of previous work [26] which collected ambient acoustic data to build the classification model (F1 of 0.86 and FPR of 9.3%). The 100 samples we collected via audio streaming channels are fed to the classifier

for prediction. Table I presents the FPR of non co-presence detection under the streaming attacks over WiFi and cellular data channels. The results indicate that the attacker (1) has a higher chance of success using the WiFi 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 WiFi 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.

The audio features we used, i.e., the ones proposed in [17], 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 [23], 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).

C. Audio-Radio System

Truong et al. [26], evaluated the performance of an audioradio 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%), a context attacker could increase the FPR: from 0.18% to 65.8% (manipulating W), from 1.1% to 1.2% (B); from 1.62% to 3.01% (audio). Now, we will analyze the same system against a bi-directional (for radio), multi-modality attacker. To model the attack, in each run, the non co-presence samples in the test 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 [26], 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., WiFi access points or Bluetooth devices). The both-sides context adversary can be modeled by replacing S_a with $S_a \cup \{(m, s) \ \forall (m, s) \in S_b, m \not \in S_a^{(m)}\}$, and S_b with $S_b \cup \{(m, s) \ \forall (m, s) \in S_a, m \not \in S_b^{(m)}\}$.

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 context attackers.

TABLE II: FPRs with/without different contextual attacks in various audio/radio/physical systems. Notations: Sets of manipulated sensors are put inside curly braces $\{\}$. $\{X\}$ 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).

·	Audio-Radio		Physical		Audio	Audio-Radio-Physical		
Zero-modality	Fuse-F	Fuse-D-S (2)	Fuse-F (3)	Fuse-D-S (4)	Fuse-F (5)	Fuse-D-S (6)	Fuse-D-M	
ero-r	2.0%	2.0%	7.5%	13.0%	3.0%	27.1%	6.9%	
Ž	(FNR: 1.4%)	(FNR: 12.0%)	(FNR: 3.9%)	(FNR: 14.5%)	(FNR: 0.0%)	(FNR: 0.3%)	(FNR: 0.0%)	
	(F1: 0.977)	(F1: 0.925)	(F1: 0.928)	(F1: 0.861)	(F1:0.990)	(F1: 0.923)	(F1: 0.980)	
	{Au}: 3.0%	{Au}: 3.0%	{T}: 8.3%	{T}: 17.0%	{Au}: 87.7%	{Au}: 45.3%	{Au}: 36.9%	
lity	{B}: 2.7%	{B}: 9.0%	{G}: 11.9%	{G}: 20.0%	{B}: 100%	{B}: 45.8%	{B}: 36.9%	
Single-modality	{ W }: 99.8%	{ W }: 8.0%	{H}: 15.3%	{H}: 24.4%	{W}: 12.3%	{ W }: 44.8%	{ W }: 35.0%	
e-in			{AI}: 55.1%	{AI}: 33.1%	{AI}: 5.4%	{AI}: 37.9%	{AI}: 6.9%	
ngu					{G}: 5.9%	{G}: 29.6%	{G}: 6.9%	
S.					{H}: 3.4%	{H}: 29.1%	{H}: 6.9%	
					- {T}: 3.4%	{T}: 31.5%	{T}: 6.9%	
	{Au, B}: 3.6%	{Au, B}: 96.0%	{G, T}: 13.9%	{G,T}: 40.1%	$\{B\}\cup\{\widetilde{X}\}$: 100%	{2 sensors}:	$\{Au,B\}\cup\{\widetilde{X}\}$:	
>	{B, W}: 99.8%	{Au, W}: 96.0%	{G, H}: 15.7%	{H, T}: 41.9%		32.0-75.4%	> 97.5%	
la	{Au, W}: 100%	{B, W}: 100%	{H, T}: 29.6%	{Al, T}: 50.6%	$\{Au\}\cup\{\widetilde{X}\}{>}74.9\%$	{3 sensors}:	$\{Au, W\} \cup \{\widetilde{X}\}:$	
Jog Jog	{Au, B, W}: 100%	{Au, B, W}: 100%	{G, H, T}: 31.1%	{G, H}: 57.5%	~	37.4-97.5%	> 88.2%	
Multi-modality			$\{AI\} \cup \{X\}:$ 64.7-100%	{Al, H}: 61.2% {Al, G}: 65.5%	$\{\widetilde{X}\}\backslash\{Au,B\}{:}<12.3\%$	{4 sensors}: 97.5-100%	{Al, G, H, T}: 9.9%	
Ĭ			0.11/10076	rest: 100%		rest: 10%	{B, W}: 36.9%	
							rest: 6.9-87.7%	

Table II (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 [26].

In single-modality attack, manipulating WiFi, the dominant feature, results in a very high success rate with features-fusion. The results change when decisions-fusion was applied. Here, each sensor contributed equally to the voting process. 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% (featuresfusion) down to 8% (decisions-fusion). We recall that the performance difference of audio and radio sensors is not large (as reported in [26], 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. When manipulating any modality along with WiFi, the FPR is above 95%. We earlier argued that audio and radio modalities can be manipulated simultaneously.

D. Physical System

In [24], four physical modalities (Al, H, G, and T) were introduced 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%) against a zero-modality adversary.

Based on our attacks against physical modalities (Section III-C), we consider an adversary 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, all non co-presence samples in the test set were transformed to the "attack" value (distance 0). The distance is set to 0 as data collection in [24] 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 undersampling as in [24]: we divided the non co-presence samples into 19 subsets, and ran several rounds of cross validation taking 10 subsets in each round and aggregating the results in the end. In addition to the features-fusion employed in [24], we tested the decisions fusion similar to our audio-radio system analysis in the previous section.

Table II (columns 3 and 4) shows our analysis results. The system performance in zero-modality attack is well-aligned with the one reported in [24]. As in [24], among four physical modalities, Al performs the best. Consequently, manipulating only Al degrades the security vastly with features-fusion (FPR increases to over 50%). 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.

E. Audio-Radio-Physical System

We extended the data collector used in [26] to record physical sensor data using an attached Sensordrone device (as used in [24]). 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.

Unlike the dataset for physical sensors ([24]) which was collected from one device at a time only, we collected 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 II (columns 5, 6, 7) reports our analysis results with different fusion methods. Under zero-modality attack, featuresfusion 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 decisionsfusion would not differ much. Decisions-fusion from subsets of sensors has a moderate performance, worse than featuresfusion 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% to 36.9%. Recall that manipulating single sensor is not difficult as we demonstrated in Section III.

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% 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 decision fusion from single modalities.

V. DISCUSSION

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, to 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 audio. 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.

Limitations and Future Work: There are certain limitations of our work. 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 copresence 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.

VI. 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) that may be used to strengthen the security of co-presence detection against a multi-modality attacker. Some of these countermeasure may require a thorough future investigation, which we plan to pursue.

REFERENCES

- [1] About Skype What is Skype. Skype. http://www.skype.com/en/about/.
- [2] Car Vacuum Activity. Youtube. https://www.youtube.com/watch?v=vN_ZBi_rmJc.
- [3] Electric Air Pump. Walmart. http://www.walmart.com/ip/33563196.
- [4] Hair Dryer. Conair. http://infiniti.conair.com/catalog.php?pcID= 47&products id=232.
- [5] Hair Dryer Activity. Youtube. https://www.youtube.com/watch?v= 3QG79NH0qjA.
- [6] Scapy. Online. http://www.secdev.org/projects/scapy/.
- [7] Ziploc and Air Pump Attack Video. Youtube. https://www.youtube.com/ watch?v=Fv2F8rY6bzw.
- [8] M. R. Bloomberg and S. Cassano. Fire safety education. http://www.nyc.gov/html/fdny/pdf/safety/fire_safety_education/2010_ 02/08_smoke_and_carbon_monoxide_alarms_english.pdf.
- [9] S. Brands and D. Chaum. Distance-bounding protocols. In Advances in Cryptology - EUROCRYPT, International Conference on the Theory and Applications of Cryptographic Techniques, 1993.
- [10] M. D. Corner and B. D. Noble. Zero-interaction authentication. In Proc. 8th annual international conference on Mobile computing and networking, MobiCom '02, 2002.
- [11] DD-WRT.com. www.dd-wrt.com Unleash Your Router. Available online at http://www.dd-wrt.com/site/index.
- [12] Y. Desmedt, C. Goutier, and S. Bengio. Special Uses and Abuses of the Fiat-Shamir Passport Protocol. In A Conference on the Theory and Applications of Cryptographic Techniques on Advances in Cryptology (CRYPTO), 1988.
- [13] D. Dolev and A. C.-C. Yao. On the security of public key protocols. IEEE Transactions on Information Theory, 29(2), 1983.
- [14] S. Drimer and S. J. Murdoch. Keep Your Enemies Close: Distance Bounding Against Smartcard Relay Attacks. In *USENIX Security*, 2007.
- [15] A. Francillon, B. Danev, S. Capkun, S. Capkun, and S. Capkun. Relay Attacks on Passive Keyless Entry and Start Systems in Modern Cars. In NDSS, 2011.
- [16] L. Francis, G. P. Hancke, K. Mayes, and K. Markantonakis. Practical NFC Peer-to-Peer Relay Attack using Mobile Phones. In Workshop on RFID Security – RFIDSec'10, 2010.

- [17] T. Halevi, D. Ma, N. Saxena, and T. Xiang. Secure Proximity Detection for NFC Devices Based on Ambient Sensor Data. In European Symposium on Research in Computer Security (ESORICS), 2012.
- [18] Z. Kfir and A. Wool. Picking virtual pockets using relay attacks on contactless smartcard. In First International Conference on Security and Privacy for Emerging Areas in Communications Networks, SE-CURECOMM, 2005.
- [19] A. Levi, E. Cetintas, M. Aydos, C. K. Koc, and M. Caglayan. Relay Attacks on Bluetooth Authentication and Solutions. In *Computer and Information Sciences - ISCIS*. Springer Berlin Heidelberg, 2004.
- [20] M. Miettinen, S. Heuser, W. Kronz, A.-R. Sadeghi, and N. Asokan. ConXsense: Automated Context Classification for Context-aware Access Control. In 9th ACM Symposium on Information, Computer and Communications Security, ASIA CCS '14, 2014.
- [21] A. Narayanan, N. Thiagarajan, M. Lakhani, M. Hamburg, and D. Boneh. Location Privacy via Private Proximity Testing. In *Proc. Network and Distributed System Security Symposium (NDSS)*, 2011.
- [22] 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.*, 12:2825–2830, Nov. 2011.
- [23] D. Schurmann and S. Sigg. Secure communication based on ambient audio. *IEEE Transactions on Mobile Computing*, 2013.
- [24] B. Shrestha, N. Saxena, H. T. T. Truong, and N. Asokan. Drone to the Rescue: Relay-Resilient Authentication using Ambient Multi-Sensing. In Proc. Eighteenth International Conference on Financial Cryptography and Data Security, 2014.
- [25] N. O. Tippenhauer, K. B. Rasmussen, C. Pöpper, and S. Čapkun. Attacks on Public WLAN-based Positioning Systems. In 7th International Conference on Mobile Systems, Applications, and Services, MobiSys '09, 2009.
- [26] 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 *IEEE International* Conference on Pervasive Computing and Communications, PerCom 2014, 2014.
- [27] A. Varshavsky, A. Scannell, A. LaMarca, and E. De Lara. Amigo: Proximity-Based Authentication of Mobile Devices. In *International Conference on Ubiquitous Computing (UbiComp)*, 2007.

APPENDIX

A. Increasing the temperature when the attacker does not know VS's location

An attacker who doesn't know the location of VS will try to keep the FS as close as possible and perform the attack activity. In our experiment, we placed the FS 10 cm apart from the VS and performed 1) when the hair dryer is closer to VS as shown in Fig. 11, and 2) when the hair dryer is closer to FS as shown in Fig. 12.

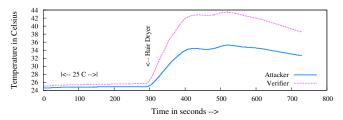


Fig. 11: Increasing the temperature; location of VS unknown to the attacker; VS is 10 cm closer to hair dryer than FS; the attacker trying to increase temperature to 35 °C.

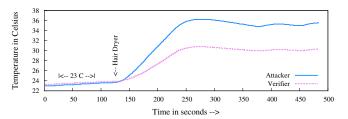


Fig. 12: Increasing the temperature; location of VS unknown to the attacker; FS is 10 cm closer to hair dryer than VS; the attacker trying to increase temperature to 35 °C.

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 13 and 14.

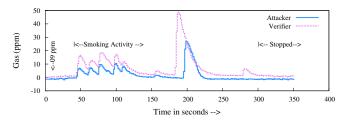


Fig. 13: Effect of cigarette in CO level; increasing the gas content to an arbitrary value and waiting to decrease to desired level.

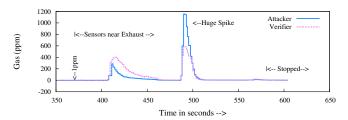


Fig. 14: Effect of car exhaust in CO level; increasing the CO gas level to arbitrary value and wait to decrease to desired level.

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. 15. 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. 15 shows a little fluctuation in altitude when we hovered the pipe around the sensors while the later 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 [2] to show the effect of portable car vacuum cleaner on the pressure/altitude sensors.

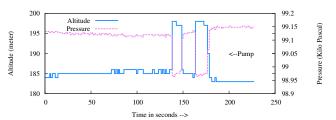


Fig. 15: Using a car vacuum cleaner to reduce pressure around the sensor and increase the altitude.