Tap-Tap and Pay (TTP): Preventing Man-In-The-Middle Attacks in NFC Payment Using Mobile Sensors

Maryam Mehrnezhad, Feng Hao and Siamak F. Shahandashti
Tap-Tap and Pay (TTP): Preventing Man-In-The-Middle Attacks in NFC Payment Using Mobile Sensors

M. Mehrnezhad, F. Hao and S. F. Shahandashti

Abstract

The reader-and-ghost attack is a real concern in mobile NFC payment applications. A malicious reader relays the user's NFC-enabled mobile phone to a remote legitimate reader to charge for a higher amount than what the user expects to pay. Using an NFC shield cannot prevent the attack, since the user consciously instantiates the NFC payment, though without realizing that the reader is controlled by an attacker. Recent solutions generally involve using ambient sensors to measure the ambient properties of the surrounding environment to ensure that the NFC-enabled phone and the reader are at nearby locations. Unfortunately, all these solutions fail completely once the attacker's reader and the legitimate reader are located in the same or similar physical environment.

In this paper, we propose the first and currently the only viable technical solution to defeat the reader-and-ghost attack even when the attacker' reader and the legitimate one are located in the same physical environment. Our solution is called "Tap-Tap and Pay" (TTP). It works by asking the user to physically tap the reader twice in succession to initiate an NFC payment. The physical tapping causes random but correlated vibrations at both devices, which are hard to forge (or reproduce) and can be reliably measured by accelerometers. Accordingly, we design the TTP protocol such that the NFC transaction will proceed only if the two vibration signals are found sufficiently similar. As compared with previous solutions, ours is fast, simple to use, easy to deploy, and above all, prevents attacks even if the attacker's reader and the legitimate one are located in the same environment.
Abstract

The reader-and-ghost attack is a real concern in mobile NFC payment applications. A malicious reader relays the user's NFC-enabled mobile phone to a remote legitimate reader to charge for a higher amount than what the user expects to pay. Using an NFC shield cannot prevent the attack, since the user consciously instantiates the NFC payment, though without realizing that the reader is controlled by an attacker. Recent solutions generally involve using ambient sensors to measure the ambient properties of the surrounding environment to ensure that the NFC-enabled phone and the reader are at nearby locations. Unfortunately, all these solutions fail completely once the attacker's reader and the legitimate one are located in the same or similar physical environment.

In this paper, we propose the first and currently the only viable technical solution to defeat the reader-and-ghost attack even when the attacker's reader and the legitimate one are located in the same physical environment. Our solution is called "Tap-Tap and Pay" (TTP). It works by asking the user to physically tap the reader twice in succession to initiate an NFC payment. The physical tapping causes random but correlated vibrations at both devices, which are hard to forge (or reproduce) and can be reliably measured by accelerometers. Accordingly, we design the TTP protocol such that the NFC transaction will proceed only if the two vibration signals are found sufficiently similar. As compared with previous solutions, ours is fast, simple to use, easy to deploy, and above all, prevents attacks even if the attacker's reader and the legitimate one are located in the same environment.

About the authors

Maryam Mehrnezhad: Maryam Mehrnezhad is a PhD student at Newcastle University, working under supervision of Dr Feng Hao. Maryam gained her BSc and MSc, both in Software Engineering at Ferdowsi University of Mashhad, Iran, in 2008 and 2011, respectively. There, she worked on different security related topics such as Soft Computing in Cryptography, Buffer Overflow Attacks, Security Engineering, and Usable Security. She joined Centre for Software Reliability (CSR), Newcastle University in 2012. Her current research focuses on Mobile Security.

Feng Hao: Feng Hao received a BEng (2001) and a MEng (2003) in electrical and electronic engineering from Nanyang Technological University, Singapore, and a Ph.D (2007) in computer science from the University of Cambridge, England. His research interests include biometrics, cryptography, fuzzy search algorithms, information coding, and error correction codes.

Siamak Fayyaz Shahandashti: Siamak F. Shahandashti is a research associate in Newcastle University, UK. His research interests include applied cryptography, privacy-preserving protocols, electronic voting, and biometric authentication. He has received his PhD in 2009 from Wollongong University, Australia, and has been with Victoria University, Sydney, and École Normale Supérieure, Paris, before joining Newcastle University.

Suggested keywords

NFC PAYMENT
MITM ATTACK
MOBILE SENSOR
ABSTRACT
The reader-and-ghost attack is a real concern in mobile NFC payment applications. A malicious reader relays the user’s NFC-enabled mobile phone to a remote legitimate reader to charge for a higher amount than what the user expects to pay. Using an NFC shield cannot prevent the attack, since the user consciously instantiates the NFC payment, though without realizing that the reader is controlled by an attacker. Recent solutions generally involve using ambient sensors to measure the ambient properties of the surrounding environment to ensure that the NFC-enabled phone and the reader are at nearby locations. Unfortunately, all these solutions fail completely once the attacker’s reader and the legitimate reader are located in the same physical environment.

In this paper, we propose the first and currently the only viable technical solution to defeat the reader-and-ghost attack even when the attacker’s reader and the legitimate one are located in the same physical environment. Our solution is called “Tap-Tap and Pay” (TTP). It works by asking the user to physically tap the reader twice in succession to initiate an NFC payment. The physical tapping causes random but correlated vibrations at both devices, which are hard to forge (or reproduce) and can be reliably measured by accelerometers. Accordingly, we design the TTP protocol such that the NFC transaction will proceed only if the two vibration signals are found sufficiently similar. As compared with previous solutions, ours is fast, simple to use, easy to deploy, and above all, prevents attacks even if the attacker’s reader and the legitimate one are located in the same environment.

Keywords
NFC payment, MITM attack, mobile sensor

1. INTRODUCTION

NFC payment: Near Field Communication (NFC) payment is an upcoming technology that uses Radio Frequency Identification (RFID) to perform contactless payments. An RFID system has two parts: the RFID tag that can be attached to any physical object to be identified; and the RFID reader that can interrogate a card within physical proximity in radio frequency. An NFC-enabled payment card has an embedded RFID tag. To make an NFC payment, the user just needs to hold the card in front of an NFC reader and wait (a short while) for the confirmation. NFC payments are usually limited to small-value purchases (e.g., buying a coffee or a bus ticket). In the UK, the maximum amount in each NFC transaction is limited to £20 only [4].

A mobile phone can also be used as an NFC payment card. HSBC Hong Kong Mobile Payment¹ and Google Wallet² are examples of NFC payment applications on mobile phones. Using the mobile phone for NFC payment is considered convenient since people carry mobile phones all the time. The new generation of smart phones has commonly been equipped with NFC sensors. Using the NFC development API, one can turn an NFC-enabled phone into an NFC card or an NFC reader depending on the application. It is estimated that mobile payment using NFC will reach 670 billion US dollars by 2015 [17].

In this paper, we focus on mobile payment using NFC. Hence, by “the NFC card”, we refer to an NFC-enabled device functioning like a payment card, which in our context is an NFC-enabled mobile phone. By “the NFC reader”, we refer to a payment terminal that communicates with the card via NFC. A legitimate NFC reader is one that is authorised by the bank and is connected to the back-end banking network for payment processing.

It is known that NFC payment is vulnerable against Man-In-The-Middle (MITM) attacks [6]. Here, two types of the MITM attacks are of particular concern, known as the ghost-and-leech and reader-and-ghost attacks.

Ghost-and-leech attack: In the first attack, the attacker places a pair including an NFC card and a reader — known as the “ghost” and “leech” respectively — between the user’s NFC card and an honest NFC reader [10]. The “leech” secretly interrogates the user’s NFC card without the user’s awareness, relays the card response to the remote “ghost” which in turn pretends to be a legitimate card and makes an NFC payment. Such an attack can be carried out by using a pair of NFC-enabled mobile phones, as demonstrated in [6, 5].

Reader-and-ghost attack: A more severe type of the MITM attack is called the reader-and-ghost attack (see Fig. 1).

¹www.hsbc.com.hk
²wallet.google.com
Figure 1: The reader-and-ghost attack: The malicious reader colludes with the malicious card (ghost) and fools the honest card to pay for something more expensive to an honest reader.

In this attack, the user consciously initiates an NFC payment with a legitimate-looking terminal (reader) but the reader actually relays data to another NFC payment terminal to pay for something more expensive. As an example, a user may wish to pay for a coffee but unknowingly, her card is abused to pay for an expensive purchase in a jewellery shop. Since the charge displayed on the attacker’s NFC reader can be an arbitrary amount, users could be easily deceived without any awareness. The feasibility of this attack has been demonstrated in [3].

As compared with the earlier ghost-and-leech attack, the reader-and-ghost attack is more serious and harder to prevent. In the first attack, the user is not aware at all that her card is being used. However, the threat can be mitigated by keeping the card inside an NFC-blocking sleeve. The same countermeasure does not work in the second attack, because the user consciously initiates an NFC payment. For ordinary users, it is difficult, if not impossible, to tell if the NFC reader they deal with is honest or controlled by an attacker. (Here, as in all previous works [7, 13, 19], we do not consider the malicious reader directly communicating with the bank, as that would be identified by the bank.)

Overview of the existing solutions: Researchers have proposed several methods to address the reader-and-ghost attack. In general there are two main approaches. The first is based on distance-bounding protocols, which estimate the physical distance between the two devices by measuring the latency of communication. However, implementing a distance-bounding protocol is not easy and the performance is extremely sensitive to computational delays incurred at the end devices. Furthermore, they usually require shared secret keys between the card and reader, which is not an option in some applications [13]. The second approach is based on using ambient sensors to measure the characteristics of the surrounding environment, such as light, sound, temperature and GPS [20, 7, 13, 19]. The underlying assumption is that the malicious and honest readers should be in two different locations with different environments. However, all these solutions will not work if the two readers are located in the same environment (e.g., in two nearby stalls in the same mall).

One might argue that the reader-and-ghost attack can be easily stopped if we require the users to be vigilant. First of all, we should note that explicit user confirmation (pressing yes/no dialogue to confirm the communication) does not prevent the attack at all. In the reader-and-ghost attack, the user consciously unlocks the phone and makes a payment. Hence, explicit user confirmation (confirmation of the communication, not the amount) does not solve the problem in this scenario. However, if the user manually verifies the payment amount shown on the mobile phone screen, she may be able to detect the difference. However, as explained by other researchers [7, 11], it is imprudent to totally rely on the user’s vigilance. In reality, because the mobile NFC payment is mainly for quick and low-value purchases, many users do not have time or do not even bother to manually check the display on their phone screens. Therefore, we believe it is necessary and important to develop a secure and usable payment solution without having to rely on the user’s vigilance.

Basic idea: In this work, we aim to design a countermeasure that defends against the reader-and-ghost attack even when the malicious and honest readers are in the same environment. Similar to the other works, we assume that the NFC-enabled mobile phone and the reader are equipped with sensors; in particular, we will use accelerometers.

Our solution is based on the following observation: As the result of physical tapping between a pair of sensor-rich devices (say two NFC-enabled phones acting as a card and a reader), the tapping event causes unpredictable but strongly correlated vibrations at both sides, which are hard to reproduce and can be reliably measured by accelerometer sensors.

At a high level, our solution works as follows. The user physically taps the mobile phone against a reader twice in succession to initiate an NFC payment. Both the phone
and the reader use the embedded accelerometer sensors to measure the vibrations caused by tapping and send the measurements to the bank. The bank subsequently compares the two measurements and authorises the transaction only if the two vibration samples are found sufficiently similar. In some sense, the vibrational features are similar to the concept of behavioral biometrics (e.g., an on-line handwritten signature). The main difference is that the vibration features are unique to a specific tapping event rather than to a person. Details of our solution will be explained in Section 3.

Contributions: Main contributions of this paper are summarised below.

1. We propose a “Tap-Tap and Pay” (TTP) as a new countermeasure to prevent reader-and-ghost attacks. To the best of our knowledge, our solution is the first that works even when the malicious and honest readers are physically close to each other or within the same environment.

2. We present a proof-of-concept implementation of TTP by using smartphone accelerometer sensors. Experiments confirm that vibrations induced from the same tapping are closely correlated, while they are not if originating from different tappings.

3. We conduct user studies to evaluate the usability and user perception about the security of our TTP prototype. Based on the feedback, users generally find our solution easy to use and fast. The physical tapping makes many users feel they are in control of the payment process and hence consider it as more secure than the entirely contactless payment method.

2. RELATED WORK

Security and privacy issues for RFID communication, and in particular NFC, have been studied intensively in the literature. However, there are only a few works addressing the MITM attack and in particular reader-and-ghost attacks against NFC-based protocols, and unfortunately, the solutions proposed in most of these works do not seem to meet sufficient efficiency, security, and usability levels. In this section, we review these solutions and discuss the limitations of each.

Note that a related but different security issue in NFC-enabled devices is vulnerability against unauthorised readings. Many of such devices, e.g., contactless cards, are always active, and hence, a malicious reader in the proximity of such a device is able to trigger a response from the device, without the user being aware of such a reading. A number of security and privacy violations have been reported in the literature exploiting such unauthorised readings. An example is the work of Emms et al. [4] which demonstrates attacks that successfully verify user PIN codes offline or extract sensitive private information such as card-holder’s birthday through such unauthorised readings. A simple solution to unauthorised readings (similar to ghost-and-leech attack) is to keep the card inside an NFC-blocking sleeve. A number of other solutions have been proposed in the literature to defend against such attacks, including Secret Handshakes [2], UWave [12], and Still and Silent [18]. The main focus of this research, however, is defending against reader-and-ghost attack.

There are two general approaches in the literature to defend against the reader-and-ghost attack: distance bounding and ambience sensing protocols. The idea behind distance bounding protocols is that a device in the proximity of a reader should be able to respond to unpredictable challenges by the reader within a tight timespan, whereas an attacker would have some delay since it has to communicate with the victim in order to get the correct response to each challenge. It has been argued in the literature (e.g. by Czeskis et al. [2] among others) that distance bounding protocols are highly sensitive to time delay and require fine grained timing mechanisms, and hence might not be fully consistent with the existing NFC hardware. Therefore, in the following we mainly focus on ambience sensing protocols and discuss them in detail.

Ambience sensing protocols are proposed as a countermeasure against reader and ghost attack. All the ambience sensing protocols in the literature work based on the underlying idea that a device and a reader are physically situated at the same environment, and hence can “sense” similar ambience, whereas in a reader and ghost attack scenario, the two pairs of communicating parties, i.e. the victim and the malicious reader as the first pair, and the malicious device and the honest reader as the second pair, are physically in different environments, and hence the ambience measurements by the victim and the honest reader will be different enough to make the mix-and-match between the communications belonging to the two transactions detectable.

An inherent shortcoming of the ambience sensing approach is that it is unable to differentiate between multiple transactions occurring at the same or similar enough environments. For instance, consider multiple point of sale terminals at a supermarket or a department store, which are situated beside one another. Sensors at different points of sales in this case can have very similar measurements of ambient factors such as audio or light.

A further observation on the ambience sensing in general is that this approach may merely prevent the reader and ghost attack, but is not effective against unauthorised readings. A number of different sensor measurements have been proposed for ambience sensing protocols. Some proposed protocols are based on sensors that are more frequently available on mobile devices such as GPS, audio, and light sensors, whereas others require specialised devices equipped with sensors such as temperature, humidity, precision gas, altitude, and pressure sensors. In the following we discuss these proposed schemes in more detail.

GPS: Ma et al. [13] propose using location data (GPS) to defend against MITM attacks. The general idea is to store a list of the locations of valid readers on read-only memory (e.g., an EEPROM) augmented to the tag. The location of the tag is continuously measured using the GPS sensor and compared to the list to turn the tag on only when in the proximity of legitimate reader locations.

Relying on GPS is challenging since GPS data is not considered reliable at indoor environments (where most payments are expected to take place). Indeed, the reported experiments were all carried out outdoors. Furthermore, to compensate for GPS sensor’s inherent inaccuracy, Ma et al. suggest averaging multiple readings to achieve a reliable measurement. Such multiple readings take around 10 seconds to complete, which seems to be too long for payment in practice, specially contactless payment. Apart from such
efficiency and usability issues, using user GPS data raises user location privacy concerns.

**Audio and Light:** Halevi et al. [7] suggested using ambient audio and light to defend against the reader and ghost attack. In this approach, both the card and the reader record ambient sound or light. The card encrypts its recording under an embedded key shared with the bank and sends the encrypted data to the reader. The reader then relays the received ciphertext along with its own recording to the bank. The bank decrypts the card data, compares it with the reader data and proceeds with processing the payment if the two recordings are sufficiently similar. To study the security of the suggested approach, Halevi et al. also developed a classifier which is able to efficiently distinguish different business type locations from the recordings.

It is not clear how effective this approach is in preventing the reader and ghost attack, since especially ambient light and to a great extent ambient sound at the honest reader environment can be easily reproduced at the malicious reader environment. Furthermore, user voice might be captured by the sensors, which raises privacy concerns. Finally, similar to the GPS-based solution, there are legitimate concerns about location privacy with this approach.

**Temperature, Humidity, Precision Gas, and Atmospheric Pressure:** Shrestha et al. [19] have explored the idea of using ambient meteorological information to determine if two communicating devices are located at the exact same place. They focus on a combination of four sensors: temperature, humidity (i.e. air moisture), precision gas (i.e. carbon monoxide (CO)), and atmospheric pressure (a.k.a. altitude).

Meteorological sensors are mostly found on highly specialised devices. Indeed, the experiments by Shrestha et al. are carried out on a Sensordone [4] which is an external device able to communicate with a smartphone. Hence, such a solution has little chance of practical deployment unless such sensors become readily available on everyday devices. Even in that case, such meteorological ambience data are public and fairly static, and can be easily recreated at the malicious reader environment (say, as long the two transactions occur at the same altitude). Therefore, it is unclear exactly what level of security such an approach would provide against reader and ghost attacks.

**Device Surface Temperature:** Uriena and Piramuthu [20] propose using independent measurements of the surface temperature of the tag by the tag and the reader to defend against relay attacks. They use the measured temperatures first as a shared secret between the tag and the reader to carry out mutual authentication and later as a seed to generate the randomness required for a distance bounding protocol.

Besides suffering from the mentioned limitations of the distance bounding protocols, the use of the low entropy temperature measurements for mutual authentication and randomness generation is questionable. Besides, accurate measurement of the tag surface temperature relies on specialised sensors that are not readily available (and indeed, the authors do not report any experiment results). Even if such sensors are available, it is not clear under what conditions the reader is able to measure the tag surface temperature with sufficient accuracy.

In summary, there are inherent limitations with both of the aforementioned approaches in the literature. On one hand, the distance bounding approach is extremely sensitive to computational delays, and on the other hand, the ambient sensing approach assumes, rather too strongly in our point of view, that the honest and malicious readers are physically in two easily distinguishable environments. In light of these issues, we propose a more effective solution to tackle the reader-and-ghost attack in the next section.

### 3. OUR SOLUTION: TAP-TAP AND PAY

In this section, we propose a solution called “Tap-Tap and Pay” (TTP) to defend against reader-and-ghost attacks.

The key idea in our solution is a notion of what we call “tap-signature”. To the best of our knowledge, this is the first time that this notion is introduced in this paper. A tap-signature refers to a type of the unique physical behaviour — more specifically, the naturally incurred vibrations between two physical devices once they are tapped to each other. This notion is similar to classic biometric approaches such as an on-line handwritten signature. A tap-signature is dynamic, naturally generated without requiring any special training and hard to forge. However, different from an on-line handwritten signature, the tap-signature is unique to a specific tapping event rather than to a user. As we will show, the unique tap-signature can be reliably measured by the two devices involved in the same tapping event, which underpins the working of our solution.

This work addresses the limitations of the previous works for the following reasons. First, since we do not require tight timing restrictions as in the distance bounding protocols, our approach is not too sensitive to timing issues, and hence is easier to implement. Besides, we only require the commonplace accelerometer which can be found on almost any mobile device. And second, since the one-time-use tap-signature in our proposed approach is transaction-specific, rather than ambience-specific, we do away with the strong assumption of the readers being physically in distinguishable environments.

#### 3.1 Overview

An overview of our solution is shown in Figure 2. First, to start an NFC payment, the user holds her mobile phone to tap the NFC reader twice in succession. The physical tapping causes unpredictable but correlated vibrations at the two devices, which are then measured by the embedded accelerometers. At this point, the reader detects the presence of an NFC card within physical proximity and starts a standard challenge-and-response process for the NFC payment. At a high level, this involves the reader sending a challenge to the card, and the card replying with a response signed by a symmetric key (via MAC) which is pre-shared with the issuing bank. Our solution does not alter this existing data flow; but within the card response, we propose to add an additional item acc_card, which is a sample of the tap-signature measured by the card. When the reader forwards the card's response to the issuing bank through a secure back-end network, it appends acc_reader, which is its own measurement of the tap-signature. The bank compares the two samples and approves the transaction only if it finds the two sufficiently similar. (Recall that in the reader-and-ghost attack in Figure 1, the user's NFC card and the jewellery shop's NFC reader are honest devices and...
In our solution, the user taps the reader twice in succession. There is no strict limit on the number of tappings. In fact, the more the user taps, the more features we can extract from the tappings, and hence the more secure the protocol is. This is similar to handwritten signatures — the more lines and shapes in a signature, the more difficult for others to forge it. With a single tap, we found that the features of the tap-signature are not rich enough to allow reliable verification. With double-tap, we were able to extract features that were sufficiently unique for our purpose. Of course, with more taps, we will be able to get more features, but at the expense of less convenience for the user. Hence, we chose double-tap as a suitable trade-off between security and usability.

3.2 Sensor data Preprocessing

Since TTP is the first work relying on device behaviour measured by mobile sensors, we performed some initial experiments to explore which sensor is suitable for measuring the physical tapping. To enable data collection, we developed two Android applications: Card app and Reader app. We installed them on two NFC-enabled smartphones, Nexus 5, which are also equipped with a plethora of sensors. Through experiments, we found the measurement of tapping-induced vibrations using the accelerometer was more reliable than that using the gyroscope. Hence, in this work, we only use the accelerometer data. It is likely that a combination of different sensors leads to better results, which will be worthwhile to explore in future research.

3.2.1 Accelerometer data

We use the embedded accelerometer sensor on the mobile phone to capture the vibration changes during the physical tapping. The accelerometer sensor returns acceleration data in three dimensions, obtained by measuring forces (including the force of gravity) applied to the sensor along the x, y and z axes (see Fig. 4). We tested data in all three dimensions and observed that the z-axis data work better than the x and y axes. This can be intuitively explained. When the user moves the card towards the reader, the movement is perpendicular to the surface of the reader, which gives the z-axis data the richest features. Similarly, at the NFC reader side, because the reader is fixed to a table, it is restricted to move sideways. The tapping-induced vibrations are mostly along the z axis. Hence, we only use the z-axis data in the paper. Note that most of pin-based readers which are NFC readers are horizontally placed. Furthermore, the angle is less relevant because we only used the z-axis data, which is perpendicular to the reader surface (Fig. 4). Using only z-axis has the further advantage that it gives us a very small data sequence and consequently faster data transfer. In the following sections, by "accelerometer data", we refer to the measurement along the z axis only.

---

5Since the prototyping of our TTP protocol requires the facility of bidirectional NFC using Host-based Card Emulation (HCE), we chose Nexus 5. At the time of experiments, the Nexus 5 was the only device allowing that facility.
3.2.2 Derivatives

As shown in the left two diagrams in Fig. 3, the accelerometer measurement by the card is more vibrant than that by the reader, since the card is moving in the hand of the user. They are also different in scale, depending on the start status of accelerometers. In order to smooth out the noise of accelerometer measurements specially in the card side and bringing them in the same scale, we apply the following equation (based on [9]) to approximate the first derivatives of the sequences. The results are displayed in Fig. 3 (right hand side).

\[ D_i = \frac{(acc_i - acc_{i-1}) + ((acc_{i+1} - acc_{i-1})/2)}{2} \]  

(1)

3.2.3 Sequence alignment

After obtaining the derivatives, we align the two sequences by first identifying the peaks. This can be simply achieved by searching for the maximum values with at least 0.1 seconds separation. The two sequences are then aligned based on the first peak (with a few linear shifts to get the best matching by trial-and-error). This alignment process does not depend on time synchronization, which is usually required in other solutions (e.g., based on audio measurements [7]). Experiments show that this simple alignment algorithm is 100% accurate and is fast.

After the alignment of two sequences, we take a segment of each sequence, starting from 0.2 second before the first peak until 0.2 seconds after the second peak. This covers the whole significant variation of the accelerometer data. Our analysis shows that the duration between two peaks is normally between 0.2 sec and 1.1 sec. With the 0.2 second before the first peak and 0.2 sec after the second peak, the whole recording time is in the range of 0.6 and 1.5 seconds.

3.3 Similarity comparison

Given the two sequences of data in the time domain, there are various standard methods to evaluate the similarity between them. Commonly suggested measures include correlation coefficients, covariance, cross covariance and cross correlation. Some of the related papers suggest using methods such as cross covariance [1] and cross correlation [2] to compare accelerometer data. All these methods usually work on data in the time domain. In [7], cross correlation is applied in both the time and frequency domains to compare audio data. In the frequency domain, coherence is a common measure for comparison, which is adopted in [14] [15] to determine the similarity between two sequences of accelerometer data.

Among various methods to compare similarity, we found
the correlation coefficients in the time domain and the coherence in the frequency domain, as the two most effective based on the collected user data. In this paper we use them as a part of our final decision maker to show that available mobile accelerometers are sufficiently sensitive to measure the tap-induced vibrations and hence produce unique tap-signatures.

3.3.1 Correlation coefficient (Time domain)

The correlation coefficient is commonly used to compare the similarity of the shapes of two signals. The intuition is that if the two measurements originate from the same double-tap, their signal shapes, especially their tap shapes would be highly correlated, and otherwise they would not be correlated. Given two vectors $i$ and $j$ and $\text{Cov}$ denoting covariance, the correlation coefficient is computed as below:

$$R_{i,j} = \frac{\text{Cov}_{i,j}}{\sqrt{\text{Cov}_{i,i} \cdot \text{Cov}_{j,j}}}$$

(2)

3.3.2 Coherence (Frequency domain)

Since the user taps the reader twice in succession, the time gap between the two taps is the dominant feature in the frequency domain. To obtain a similarity measure in the frequency domain, we apply the coherence method, which indicates the level of matching of features in the frequency domain between two time series. Given two sequences, $i$ and $j$, we compute the magnitude squared coherence based on the following equation, where $P_{ii}(f)$ and $P_{jj}(f)$ are power spectral densities of $i$ and $j$, and $P_{ij}(f)$ the cross power spectral density between $i$ and $j$.

$$C_{ij}(f) = \frac{|P_{ij}(f)|^2}{P_{ii}(f) \cdot P_{jj}(f)}$$

(3)

We define the similarity rate between the two signals based on magnitude squared coherence as the sum of the magnitude squared coherence values at all frequencies as follows:

$$F_{ij} = \sum_f C_{ij}(f)$$

(4)

3.3.3 Energy Difference

Our analysis shows that different users tap devices with different strengths; some taps are very gentle, some are medium, and some are very strong. We found the total energy levels of the card and reader sequences obtained from the tap are strongly correlated, while they are distinctive if obtained from different taps. Hence, we use the following measure to capture the distance of two signals $i$ and $j$ in term of the total signal energy levels.

$$D_{ij} = |\sum_i^{} - \sum_j^{}|$$

(5)

3.3.4 Combined method

We use a simple linear fusion method to combine all the three similarity measures in making a final decision. In this method, the bank decides to authorize or decline the transaction based on the weighted sum of the three measures: correlation coefficient, coherence, and the energy similarity. The decision is made based on comparing the weighted sum against a threshold.

We use a simple linear normalisation that maps the above values to the interval $[0,1]$. Let us denote these normalised versions by $\bar{R}_{ij}$, $\bar{F}_{ij}$, and $\bar{D}_{ij}$, respectively. Since unlike the other two measures, $\bar{D}_{ij}$ decreases with similarity, we define $\bar{E}_{ij}$ as below. Note that $\bar{E}_{ij}$ is also a normalised value belonging to the interval $[0,1]$.

$$\bar{E}_{ij} = 1 - \bar{D}_{ij}$$

(6)

Given $\bar{R}_{ij}$, $\bar{F}_{ij}$ and $\bar{E}_{ij}$, $T_{ij}$ calculates the total similarity rate of two signals $i$ and $j$ as below, where $a$, $b$ and $c$ are the weights of each method:

$$T_{ij} = a \cdot \bar{R}_{ij} + b \cdot \bar{F}_{ij} + c \cdot \bar{E}_{ij}$$

(7)
The weight parameters are determined through experiments based on the collected user data by testing all possible weights up to two decimal places for a, b, and c — under the condition that the sum of them is equal to 1 — and observing the equal error rate. The values which gave us the best error rate have been fixed as $a = 0.07$, $b = 0.30$, and $c = 0.63$. In the next section, we will explain the data collection process and user experiments in detail.

4. EXPERIMENTS AND RESULTS

We performed two different experiments: 1) to collect accelerometer data from double-tapping; 2) to conduct user studies to evaluate the usability and security aspects of the TTP solution from the users’ perspective.

4.1 Experiment 1: Data collection

4.1.1 Proof-of-concept prototype

We implemented a proof-of-concept prototype for the TTP system. Two Android mobile phone applications were developed, one for the card and the other for the reader. When the user taps the reader, the two applications independently record the accelerometer data and when the NFC card is identified by the reader, they start a two way NFC communication and simulate an NFC payment. In order to evaluate the performance of the prototype based on real-world data, we recruited 23 volunteers (students and staff in a university, 13 females) to participate in the data collection, each performing five tapping actions.

We made a short video guide to demonstrate how to do the double-tap and showed it to the user before the experiment. The video guide served as self-explanatory training, which was played for all users. Users generally found the video guide useful in helping them quickly grasp the instruction of “Tap-Tap and Pay” within a few seconds.

4.1.2 Experiment setup

We fixed the reader phone to the table by using a doublesided tape (Fig. 5,a). The front of the phone faced downwards and the back was labelled “Reader”. We used an application called MyMobiler\(^6\) to operate the reader through a USB connection. The GUI of the card application was very simple (Fig. 5, b). After launching the card application, the user just tapped the phone to the reader twice in succession and kept it close to complete an NFC payment. Once he was notified with a successful payment, he could repeat the experiment. The recorded accelerometer data at both the card and the reader were saved into a file for further analysis in Matlab.

4.1.3 Results

The False Negative Rate (FNR) is the rate that two measurements from the same tap event are determined as not matching. The False Positive Rate (FPR) is the rate that two measurements from two different tap events are determined as matching. Like in a biometric system, FNR and FPR vary according to a threshold. The Equal Error Rate (EER) is the rate where the FNR and the FPR curves intersect. The EER is commonly used as a measure to evaluate the overall performance of a biometric System. Based on the collected data, we computed the EERs in Table 1 based on the similarity comparison methods as described in Section 3.3.

<table>
<thead>
<tr>
<th>Method</th>
<th>Equal error rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Correlation coefficients</td>
<td>30.22%</td>
</tr>
<tr>
<td>Coherence</td>
<td>25.39%</td>
</tr>
<tr>
<td>Total energy</td>
<td>25.60%</td>
</tr>
<tr>
<td><strong>Combined method</strong></td>
<td><strong>17.65%</strong></td>
</tr>
</tbody>
</table>

Table 1: Equal error rates for different suggested methods

Overall, the Equal Error Rate of our prototype system is 17.65% by using the combined method (Table 1). This error rate may look relatively high, but it is worth noting that under the same test condition, all other solutions have 100% error rates (see section 4.3). From the 100% error rate to 17.65%, we consider this a significant improvement. We believe the error rate can be further reduced when more

\(^{6}\)www.mymobiler.com
sensitive sensors are available in future mobile phones. It is also worth comparing this error rate with the typical 25% EER reported in on-line handwritten signature verification systems [8]. We find that the dynamic “behaviour” of a mobile device caused by tapping is actually more consistent than the human “behaviour” of writing signatures.

4.2 Experiment 2: User study

We performed a second experiment to evaluate the usability and the user’s feeling about the security of the TTP prototype. We asked 22 different users (who partially overlapped the last users set, students and staff in a university, 7 females) to make two types of NFC payments using a mobile phone. The first one was a contactless payment, and the second one was an NFC payment using TTP.

4.2.1 Experiment setup

In this experiment, we asked each of the 22 users to carry out two tasks: making an NFC payment first using the contactless method and then using TTP. We developed two prototype Android mobile applications, one for the card and the other one for the reader, to simulate the same paying experience as using the contactless NFC payment. Before the experiment, we presented users a study description, including a short introduction of mobile contactless payment using NFC, followed by a general description of mobile payment using TTP. In the first task, the user was asked to hold the phone near the reader and wait for the confirmation message. In the second task, the user was asked to double-tap the reader, keep the phone near the reader and wait for the confirmation. Figure 7 shows the GUIs of the two tasks in this experiment.

4.2.2 Findings

After completing the two tasks, the users were asked to fill in a questionnaire and rate the level of convenience, speed and feeling of the security of each payment method in a likert scale from level 5 to 1 (corresponding to “strongly agree”, “agree”, “neutral”, “disagree”, and “strongly disagree”). They were also asked to write free comments about the usability and the security aspects in this experiment. Figure 8 shows the average user rating of the contactless payment and TTP.

As shown in Fig. 8, users generally found the contactless payment very convenient, while TTP has a lower rating in this factor. Including a physical action makes it less convenient for users. As one user commented: “... the fact that I need to keep the device close to the reader after tapping make the experience less convenient”.

![Figure 7: GUIs of user study Android applications](image)

![Figure 8: User study: average user rating of contactless payment and TTP](image)
<table>
<thead>
<tr>
<th>Sensor/ Solution</th>
<th>Prevents attacker at same place</th>
<th>False negative (%)</th>
<th>False positive (%)</th>
<th>Recording duration (sec)</th>
<th>On mobile sensors</th>
<th>Based on ambience or device</th>
</tr>
</thead>
<tbody>
<tr>
<td>Audio [7]</td>
<td>x</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>✓</td>
<td>Ambience</td>
</tr>
<tr>
<td>Light [7]</td>
<td>x</td>
<td>5</td>
<td>6.5</td>
<td>2</td>
<td>✓</td>
<td>Ambience</td>
</tr>
<tr>
<td>GPS [13]</td>
<td>x</td>
<td>0–100</td>
<td>N/A</td>
<td>10</td>
<td>✓</td>
<td>Ambience</td>
</tr>
<tr>
<td>Temperature (T) [19]</td>
<td>x</td>
<td>23.74</td>
<td>32.40</td>
<td>instant</td>
<td>x</td>
<td>Ambience</td>
</tr>
<tr>
<td>Precision Gas (G) [19]</td>
<td>x</td>
<td>15.26</td>
<td>30.36</td>
<td>instant</td>
<td>x</td>
<td>Ambience</td>
</tr>
<tr>
<td>Humidity (H) [19]</td>
<td>x</td>
<td>16.25</td>
<td>29.81</td>
<td>instant</td>
<td>x</td>
<td>Ambience</td>
</tr>
<tr>
<td>Altitude (A) [19]</td>
<td>x</td>
<td>8.57</td>
<td>16.25</td>
<td>instant</td>
<td>x</td>
<td>Ambience</td>
</tr>
<tr>
<td>THGA [19]</td>
<td>x</td>
<td>2.96</td>
<td>5.81</td>
<td>instant</td>
<td>x</td>
<td>Ambience</td>
</tr>
<tr>
<td>Accelerometer (TTP)</td>
<td>✓</td>
<td>17.65</td>
<td>17.65</td>
<td>0.6–1.5</td>
<td>✓</td>
<td>Device</td>
</tr>
</tbody>
</table>

Table 2: Comparing solutions against the reader and ghost attack. Note that in all ambient sensor-based solutions [7, 13, 19], the reported error rates are based on testing under distinct environments. If the attacker's reader is in the same environment as the honest reader, the False Positives in all these solutions will be nearly 100%.

comparing to the first one. I feel after tapping I automatically bring the phone close enough to the reader, but in first task, my phone was not close for a while and it took longer. The security rating on TTP from the user's point of view about TTP is higher than the contactless payment. By a physical tapping action, users felt in control of the transaction and would not worry about accidental payments. One user commented about the security of the two payment methods: “As before [task 1] payment is very easy. I like the action of tapping the reader as this made me feel more in control of when the transaction took place. I felt this method [TTP] was more secure due to the action of tapping to start the transaction. This meant I knew when the transaction took place”. Another user wrote a similar comment: “The payment [in task 1] is very easy, but I don’t know when the connection between wallet and reader is made; range or time, so I would keep my payment device away from the reader to be sure until I want to pay.”

4.3 Comparison with previous works

In this section, we compare our results with those of previous solutions against the reader-and-ghost attack. For distance-bounding solutions, a meaningful comparison of different characteristics seems to be hard since such protocols are highly sensitive to timing delays, whereas our approach is reasonably indifferent to minuscule timing variations. Hence in our comparison, we mainly focus on ambience sensing based solutions. We consider a number of security, efficiency, and usability factors. Table 4.2.2 summarises the comparison.

In terms of security, our proposed approach is the first solution able to prevent the reader-and-ghost attack when the honest and malicious readers are located at the same place. In contrast, ambience sensing solutions are inherently incapable of distinguishing an honest party and a malicious party located at the same environment, and hence cannot prevent attacks by adversaries that are physically at the same place.

Our proposed approach achieves comparable results in terms of false negative and false positive rates. Halevi et al. [7] report false negative and false positive rates for distinguishing different business types (such as library, concert hall, restaurant, etc.), which assumes that the adversary is at a different business location. Under such a rather strong assumption, they report false positive and false negative rates of 0% for audio sensor, and around 5% for light sensor. Ma et al. [13] report a 0% false negative rate under the rather strong condition that the attacker is located 20 meters or farther, 67.5% when the distance is more than 5 meters, and 100% when the distance is less than one meter. They do not report false positive rates. Shrestha et al. [19] report false negative rates approximately in the range of 10%–25% and false positive rates approximately in the range of 15%–30% for individual sensors. By combining the sensor readings, they achieve a false negative rate of about 3% and a false positive rate of about 6%. Our scheme achieves an equal error rate of 17.65% which is comparable to most of the above except for results combining multiple sensor readings or those based on strong assumptions.

In terms of usability, our scheme compares favourably against the proposed protocols based on GPS and meteorological sensors, and is comparable with those based on audio and light sensors. Our protocol needs a sensor recording of only 0.6 to 1.5 seconds which is sufficiently fast for contactless payment. Schemes based on audio and light sensors achieve similar timings. However, the GPS-based protocol of Ma et al. requires 10 seconds of sensor recording which makes the system not suitable for contactless payment. Our scheme is based on accelerometer which is readily available on most mobile devices, as are microphones (audio), light sensors, and GPS. However, meteorological sensors are only available on specialised devices. This imposes a significant barrier in adopting such protocols in practice.

Some participants in our study had the concern of physical damage to the device and tapped very gently. Our analysis shows that in general, as expected, gentle taps cause small vibrations that are less correlated at the two sides. We expect the error rates in TTP to improve with the equipment of more advanced accelerometers on future mobile phones.

5. DISCUSSION

5.1 Security against reader-and-ghost attack

Consider a reader-and-ghost attack on the TTP system. The user performs a double-tap on a malicious reader. Since the user’s mobile phone cryptographically protects the integrity (via MAC) of the accelerometer measurement using a key known only to the bank, the malicious reader cannot
alter the measurement. The malicious reader should be able to measure the same vibrations, so the accelerometer data is not secret. However, the difficulty for the attacker is to reproduce the same or sufficiently close physical vibrations by tapping with an honest reader. This is hard, if not impossible, for a human being to precisely control the timing, strength and location of the tapping. We might imagine the attacker uses a robot arm to precisely control the tapping action. While this may be theoretically possible, the attack could be detected by staff at the NFC payment terminal.

5.2 Security against unauthorised readings

We believe that the double-tap pattern generation requirement captures the user’s intention in essence, and hence we may assume, barring exceptional circumstances, that the user “intends” to make the payment. This provides a level of protection against unauthorised readings and the ghost-and-leech attacks, since the protocol can require the two tapping actions in quick succession to start the communication. Li et al. have studied the application of tapping as an intuitive human gesture to initiate an NFC communication in [11]. It is possible that an attacker may try to activate the user’s NFC payment application by “bumping” the user’s pocket or bag where the phone is kept. Such an attack is at the risk of being detected by the user.

5.3 Usability of TTP

Although we slightly change the original usage model of the contactless payment protocol by requiring the user to “Tap-Tap and Pay” instead of contactless payment. We consider this a minor modification. The concept of proximity, i.e. keeping the device “close” (up to 4 cm [16]) to the reader for a short while is still unchanged, and we do not require the user to carry any extra hardware, press any extra button, or perform any extra action other than double-tapping.

Our user study data suggest that many users rated TTP as even faster than the contactless payment protocol. Our observations show that users normally keep the mobile phone in a very close distance with the reader after double tapping, whereas in the contactless payment protocol, it takes the user a while to come up with the proper distance. Hence the user is more likely to feel more time is passing by. Furthermore, a number of our users expressed concerns while passively waiting for the contactless payment protocol to initiate, such as if they are holding the card close enough to the reader. Our users tended to feel better when they physically tapped their cards to the reader which seems to have given them a feeling of being in control of starting the process.

5.4 Privacy of the card owner

Some proposed protocols in the literature, namely those based on recordings of ambient sound or GPS, pose privacy concerns to the user. For example, the user may not want her voice or location to be recorded. This is not an issue in our protocol, since double-taps are generated on the fly per transaction.

5.5 Extension to other sensors and signatures

Although we have used the recordings from the accelerometer for double-taps, it is conceivable to use a combination of sensors along with accelerometer, such as gyroscope, for the recordings of the double-tap. It is also possible to devise more complex physical patterns, such as triple-tap, for applications that require higher security assurance. Hence, we view our work, not as a definitive solution, but as a first step in exploiting unique “behavioural” features from a physical object (instead of a human being) for security applications.

6. CONCLUSION

In this paper, we propose “Tap-Tap and Pay” to address an unsolved problem in the mobile NFC payment: namely, the reader-and-ghost attack. All previous solutions fail once the attacker’s reader is physically close to the real payment terminal or if they are located in the same environment. Our solution is the first that works regardless of the attacker’s location or environment. The novelty in our solution is based on extracting a unique “tap-signature”, which is dynamic, unpredictable, naturally occurring and difficult for a human to forge (or reproduce). Experiments show that the “tap-signature” can be reliably measured by using an accelerometer sensor embedded with a mobile phone. Although the TTP solution requires the user to physically tap the reader twice in succession to initiate an NFC payment, feedback from our user studies indicates that users generally find it as convenient as the contactless payment method; many of them feel using TTP is more secure as the users are in control of the payment process.

7. REFERENCES


