All Your Payment Tokens Are Mine: Vulnerabilities of Mobile Payment Systems

Speaker: Zhe Zhou, zhouzhe@fudan.edu.cn
Pre-Tenure Associate Professor,
School of Computer Science, Fudan University, China

This white paper describes a co-lead project with Xiaolong Bai. Please see the USENIX Security'17 paper "Picking Up My Tab: Understanding and Mitigating Synchronized Token Lifting and Spending in Mobile Payment"[1] co-authored by Xiaolong Bai, Zhe Zhou, Xiaofeng Wang, Zhou Li, Xianghang Mi, Nan Zhang, Tongxin Li, Shi-Min Hu, Kehuan Zhang, for detailed and complete information.

Mobile payment service makes our life far more convenient. However, according to our latest findings, even some extreme mobile popular payment products have vulnerabilities that may result to their users' financial losses. The threatens are mainly a result of a problematic off-line payment scheme.

---

1 Background

Mobile payment transactions should be able to be completed even without mobile-phone's connectivity, which provides users good experience even in poor cellular network area. Still, the security of such an off-line payment should be guaranteed.

There are a lot of popular payment schemes, like Wechat pay, Alipay and

Samsung pay, and they all support payment in the off-line mode. Besides, their schemes mainly base security on payment tokens that are generated every time users try to pay. Figure. 1 illustrates the process of payment token generation.

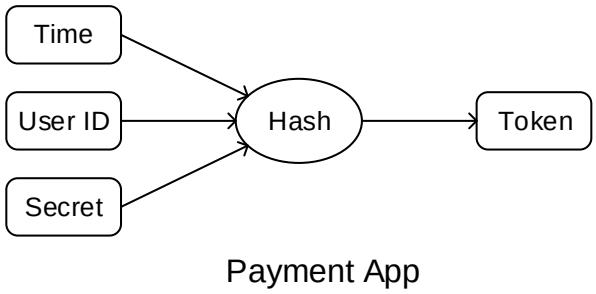

Figure 1: Payment token generation process.

The token, which is the key of the security of their schemes, will be modulated to a broadcast channel (not network) for sending to the vendor. E.g., 1) encode the token to a piece of QR code and display it on screen, 2) encode it to a clip of audio and play the audio or 3) encode it to magnetic signal via a coil antenna, as showing by Figure. 2

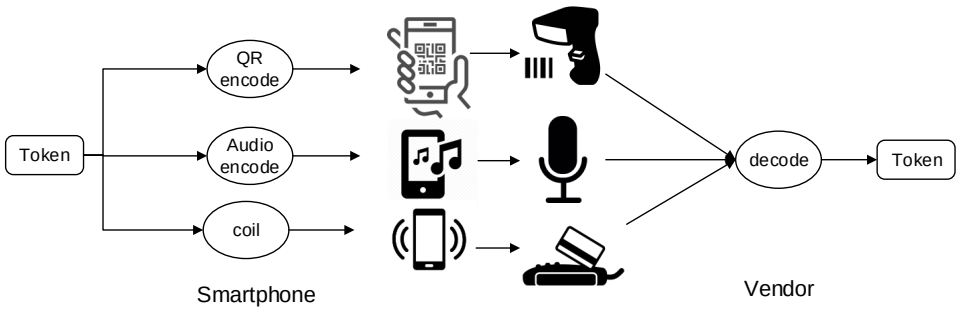

Figure 2: Payment token transmission process.

The vendor, once receives the signal from the channel by 1) scanning the QR code, 2) recording the audio played by the phone, 3) or receiving the magnetic signal through POS (will be explained later), will decode the token and relay it to the back-end server for verification and complete the transaction.

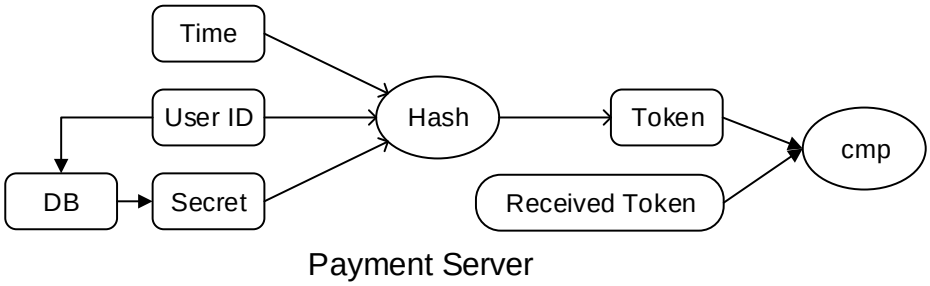

Figure 3: Payment token verification process.

The server will check the validity of the token and approve or deny the transaction. The server checks 1) if the token is already used by looking up the token in the history, 2) if the token is already expired, by checking if the claimed generation time is within a valid period, 3) if the token is indeed generated by the claimed user by generating a token using the user's secret key and compare the generated one with the received one. After having checked the validity of the token, the server is convinced that the token is indeed just generated by the claimed user for the transaction the vendor claims, so can proceed to make decisions according to other security unrelated issues, like the balance.

The scheme has at least two security weaknesses but cannot be exploited by attackers easily: 1) The token is transmitted in a broadcast fashion, so nearby attackers are potentially able to acquire the token. 2) The token is not restricted for a specific transaction use, so a token, once generated, can be used by a vendor to authorize a transaction with any amount of money under the daily limit. Attackers cannot easily exploit the two weaknesses because 1) when an attacker successfully

sniffs a token, the token is simultaneously legally received by the vendor, who will immediately send the token to server for verification, making the sniffed one a used one. 2) Vendors are assumed to be trusted, so they will not overcharge their customers with the token. If they do so, customers will later notice the overcharge and ask for reimbursement.

2 Threat Analysis

An attacker however can make use of the weaknesses of off-line payment schemes. As mentioned before, an attacker can easily acquire a live token from the broadcast channel, but the token will be immediately legally spent by the vendor. Nonetheless, the token can be kept alive, if the attacker can somehow interrupt the transaction routine. If this happens, the attacker then can send the token to somewhere else for spending with amount up to the daily limit.

Accordingly, we devised a series of attacks against different payment services. Those attacks follow the same attack routine: 1) sniff a token; 2) interrupt the ongoing transaction, 3)spend the token at somewhere else.

3 Audio Pay

Our first target is audio pay, which is operated by the largest mobile payment service provider in China.

Audio pay is deployed in vending machines. Vending machines are connected to the service provider via cellular network and use microphones to receive audio tokens. Next, we illustrate how we attacked audio pay in real world.

Token Sniffing Attack places a running recorder next to a vending machine.

Transaction Interruption Once an user pays at the machine, the attacker turns on his cellular network jammer immediately.

Spending The recording is sent to another attacker who can purchase from a vending machine by replay the recording.

4 MST

Samsung employs a technique named MST to transmit tokens between a phone and a POS. POS machines were designed to read magnetic stripe cards only, with a magnetic sensor embedded to sense the magnetic field variation resulted from card swiping. Samsung creatively added a coil to inside their phones, which can broadcast magnetic signal carrying a token. The POS machine can sense the signal and feel as if someone is swiping a card, so it can normally decode the token encapsulated in a card track format.

Token Sniffing We designed a coil antenna (Figure .4) to sniff the magnetic signal produced by the phones. The token inside a piece of sniffed signal can be decoded with our implemented decoder.

Transaction Interruption Again, the POS can be jammed with a cellular network jammer.

Spending The token can be written to a magnetic strip card with a commercial magnetic strip writer, with which the token can be spent at a POS.

5 QR Code

QR code is the most popular mobile payment method now. Vendors use a scanner to scan users' phone screens to acquire the QR code carrying payment tokens.

Token Sniffing We assume the user's phone is infected with a malicious app. The app can turn on front camera to capture images during the scanning. We found that the QR code shows in the glass of the scanner lens, because of reflection. So

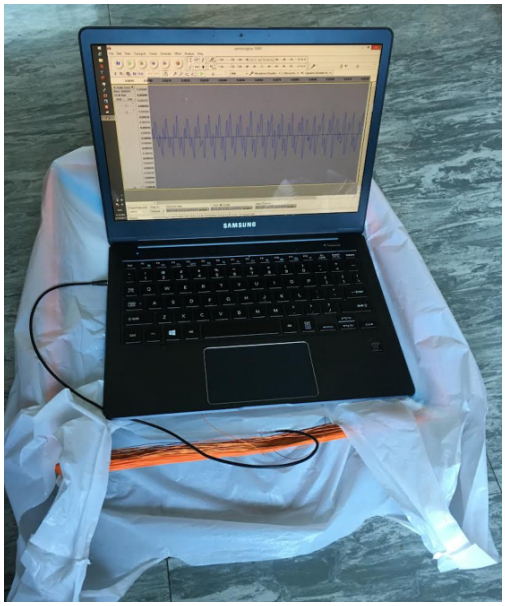

Figure 4: MST token sniffing antenna.

the attacker can decode the token from the image captured by the front camera. See Figure. 5.

Transaction Interruption The app draws a white box over one of the positioning marks of the QR code displayed on screen. The QR code can no longer be recognized by the scanner. But users can barely notice this.

Spending The token can be re-encoded to a QR code, which can be spent at any POS.

6 Bonus Attack

A QR code carrying a payment token, which is assumed to be sensitive, not only can be used to make payment, but also can be used as a kind of name card. When someone wants to receive money transfer, he can show the QR code and then the payer scans the QR code and complete the rest procedures without entering the

Figure 5: An image reflecting a payment QR code captured by the front camera of a phone. The right lower positioning mark is masked by the malware.

payee's account number. To be noticed is that the token inside the QR code is useless for the transfer and only the user ID is extracted and used.

This function is exploited by us to steal users' money even when the victim is not infected by any malware.

Token Sniffing We assume that the payer is infected with a malicious app. When he scans the victim's QR code, the malicious app jump to front with an identical UI and directly take a photo to rob a QR code.

Transaction Interruption Interruption is important here, because the attack can be noticed by the payer, if it wastes too much time. The malicious app immediate launch a bluetooth pairing request to the victim's phone, and immediately cancel the pairing, and quit immediately.

The rational behind the interruption is a bit complex. The pairing request will incur the paring confirmation in the victim's phone, which shows a window asking confirmation from the user. While the pairing is immediately canceled, so

the windows will immediately disappear. The life time of the windows can be as short as some hundreds of millisecond but the payment app showing the QR code will regenerate a token and a QR code, resulted from the window overlay.

The sniffed token is kept alive, because the transaction will go on with the newly generated QR code.

Spending The robbed QR code can be spent at any POS.

Follow up We demonstrated the attack to the largest payment service provider in China. They revoked the function of QR code as a name card for transfer.

References

[1] BAI, X., ZHOU, Z., WANG, X., LI, Z., MI, X., ZHANG, N., LI, T., HU, S.-M., AND ZHANG, K. Picking up my tab: Understanding and mitigating synchronized token lifting and spending in mobile payment. In *26th USENIX Security Symposium (USENIX Security 17)* (2017), USENIX Association, pp. 593–608.