Card Fraud in a PSD2 World: A Few Examples

Prepared by Timur Yunusov

Head of Offensive Security Research

A Cyber R&D Lab Publication

July 21st 2020

https://www.cyberdlab.com

Contents

1. Summary of Findings

In January of 2018, the European Union adopted the Payment Service Directive v2.0 (PSD2), which, among other regulations, requires multi-factor authentication in order to prevent tap-and-go fraud with contactless cards.

However, the fast-changing threatscape in the payments industry and emerging attack tecnhiques are challenging even the best ones. In our recent research, we found several ways to bypass required multi-factor authentication for both MasterCard and Visa cards, including offline PIN attacks.

2. Introduction

With each passing day, more and more people around the globe choose tap-and-go payments via their contactless cards. For example in 2019 48% of in-person transaction[1] through Visa are contactless.

Alongside the popularity of new types of payments, new ways of fraud also became more and more popular. One of them was specifically with contactless cards limits for transactions that didn't require Cardholder Verification Methods (CVM) [1].[2]

This article represents similiar research that we performed in the middle of 2019 around Visa and Mastercard contactless cards. Since PSD2 [2] came into full force in September, after our original contactless research work was published, we decided to re-visit this topic and focus on some of the specific aspects of the new regulation.

One of the most interesting requirements of the PSD2 regulation was Strong Customer Authentication (SCA), which requires issuers to enforce multi-factor authentication for all types of payment transactions from time to time. It is important to clarify that the SCA requirement attempts to reduce the risks associated with in-person Tap-and-go [3] payments, which previously did not require payment verification.

Since our study was focused on contactless cards limits for No CVM [4] transactions, we decided to repeat our research again on several European banks with implemented SCA.

Most payment institutions were unable to make the original transition deadline of January 2020, so mandatory SCA has been pushed back until at least 2021.

[1] Contactless cards: 'Whoa, it's the future!' https://www.creditcards.com/credit-card-news/contactless-tap-and-go-cards-us-market

[2] All specialist's terminology used in this article is defined in the section Glossary of Terms and marked in the text as "[Number of reference]".

During our research, we found multiple ways to circumvent SCA requirements at four out of five banks that we focused on. Here, we present our findings.

3. State of Payment Security

3.1. Life Before PSD2: A Question of Limits & Liability

Before PSD2, issuing banks didn't have any mandatory limits on card payments. Instead, acquiring banks and merchants set Tap-and-go limits for contactless transactions. Tap-and-go limits were individually configured on each payment terminal and were optional. For example, the limit recommended in the U.K. until March 2020 was £30. With the 2020 pandemic, this number has risen to £45 per transaction. Because these limits are optional, each acquiring bank can set their own limits on the terminals they provide—even as high as, let's say, £1,000, if the merchant is a luxury store.

Each merchant (store) may also ask to set tap-and-go limits that are higher or lower than the officially recommended ones. But what does this mean for fraud? Because the cardholder is not responsible for contactless fraud, liability accrues to the merchants, or acquiring banks. In the U.K., for example, a robbed cardholder must reimburse the first £50 of stolen money.[3] The merchant and/or acquiring bank must make up the difference. Given this, it becomes obvious why few stores are willing to go beyond the standard limits: Scammers would target them to make payments with the higher limits, potentially sticking the store with liability.

3.2. Tap-and-Go & Contactless Fraud

Contactless fraud is as old as contactless technologies. Although it makes up just three percent of card fraud by value, contactless fraud causes annual losses in the U.K. in excess of £1 million.[4]

The most popular contactless fraud scheme for lost and stolen cards at the moment is to make as many Tap-and-go payments as possible, which do not require entering a PIN code for payments. For example, spliting the store total into multiple transactions, each within the no CVM limits. So when trying to buy a product for £135 with a limit of £45 at the terminal, the attacker asks a seller to charge three payments of £45 eachwith no need to enter the PIN. This can be continued until the card is blocked or there are no funds left on the card.

[3] Q&A: All you need to know about contactless payments https://www.moneysupermarket.com/money-made-easy/q-and-a-all-you-need-to-know-about-contactless-payments/

[4] Surge in contactless card fraud - stealing £1.18m in 10 months https://www.standard.co.uk/news/crime/surge-in-contactless-card-fraud-stealing-118m-in-10-months-a4030256.html

4. Putting Cards to the Test

The study involved six cards from five banks, which, as of the time of our research, used Strong Customer Authentication for their cards.

Goal of our testing scenario was: Using a "stolen card" and without knowledge of the PIN code, to make more than five payments for a total exceeding £225 (which is the current U.K. limit).

4.1. How Banks Implement Cumulative Limits

The main purpose of cumulative limits [5] is to fight Tap-and-go fraud. With cumulative limits in place, stolen unblocked cards in the U.K. are limited to a maximum of five consecutive contactless transactions worth a total of £225. After meeting the £225 cumulative limit, the cardholder is asked to insert the card, use the chip and provide the PIN.

To understand how to circumvent these limits, we need to examine in detail how banks implement their checks.

Important to clarify, that as offensive security researchers, we usually do not have complete information about the specifics of procedure in banks. Thus, our work is based on a series of experiments, communications with banks and other partners, and publicly available information.

The most common step for banks is to conduct checks during payment authorization. The issuing bank receives the authorization request message specified in ISO-8583,[5] which contains fields with the following information:

- Amount of payment
- Type of payment
- Card info
- Transaction attributes for card authorization on the issuer's side
- Issuer Application Data (IAD) [6]
- Terminal Verification Results (TVR) [7]
- Cardholder Verification Methods (CVM) Results [8]

The integrity of some of these fields is checked using a cryptogram on the hardware security module (HSM) [9]. However, some of them are not checked, depending on the cryptogram

[5] ISO 8583
https://en.wikipedia.org/wiki/ISO_8583

version, card brand and sometimes even the payment authorization software. Sometimes important checks can be accidentally disabled, as we will describe later. This is not new and, in fact, happens with some frequency, such as in the so-called "Brazilian attack" described by Brian Krebs.[6]

The issuer checks the fields at least three times:

- The integrity of the cryptogram is checked on the HSM. This is to be sure that none of the essential fields (such as amount and currency) from the terminal were tampered with after the cryptogram left the card.
- During risk management: Because the information from the payment terminal is considered to be secure and unmodified, the bank uses this information to decide whether the transaction is legitimate. For example, the bank checks which types of cardholder verification and offline authentication were performed.
- Based on the transaction type, the bank will increment or reset the counter. If it's a chip transaction confirmed by PIN, the counter is reset. If the cumulative amount exceeds £225, payment is declined, and a message prompting to "insert card" is displayed.

Hence, the issuer has enough data to make a decision: Update the cumulative limit counter, reset it, proceed with the transaction or decline the transaction due to exceeded limit.

One of the five tested banks has introduced an additional measure - a transaction counter on the card itself. Which means that after five contactless transactions, contactless payments on the card will no longer work, requiring to use the chip. This measure led to mass reissuance of cards. (We will come back to this later)

4.2. History of Offline PIN & CDA Downgrade Attacks

Offline PIN attacks (PIN-OK) against Chip and PIN cards [10] have been known since 2009, when Omar Choudary, Steven Murdoch, et. al. from the University of Cambridge pointed out the possibility of such attacks. In 2010, an organized crime group was able to pull off such an attack to the tune of more than half a million euros in France.[7]

The PIN code is checked during the cardholder verification process based on the CVM list. The EMV specification allows checking the PIN not only on the bank side, but also on the card itself using so-called offline modes. In this case, the terminal can pass the PIN to the chip for verification in plaintext or encrypted form. But in either case, as long as the PIN is correct, the terminal receives the answer "9000" (meaning "everything is OK") from the card. And here is

[6] Revolution' Crimeware & EMV Replay Attacks
https://krebsonsecurity.com/2015/04/revolution-crimeware-emv-replay-attacks/
[7] When Organized Crime Applies Academic Results
https://eprint.iacr.org/2015/963.pdf

where a man-in-the-middle attack can be effective. A special device for this attack was created by the original research team, and we have used an equivalent one in our own work.

Figure 6.2: **Forward Commands** application tested on Natwest CAP reader. The SCD has blocked the transaction after the PIN has been entered and is waiting for the user to select if the transaction should continue (yes) or not (no)

[A wedge device for man-in-the-middle attacks]

A lot would seem to have changed since 2009. The offline PIN now is rarely used as a priority cardholder verification method. Cards and terminals now support the modern CDA [11] secure offline authentication scheme, which specifically thwarts PIN-OK attacks by preventing changes to the CVM list and CVM results fields. As for the old DDA [12] authentication scheme, which does not protect the integrity of card fields, banks have implemented additional mitigation features: The card now can send information to the issuer if the offline PIN has failed verification. This field, too, cannot be tampered with, because it's part of the payment cryptogram that is checked during authorization.

However, results of our research indicated that banks did not put enough attention to these indicators, which should affect the state of the cumulative limit counter.

4.3. Threat Models & Materials Used for Testing

In our study we employed two threat models. In the first model, attackers possess their own POS terminal, which they can configure. The second threat model applies to unattended terminals, such as self-service kiosks in food chain restaurants, gas stations and so on.

For testing, we used six cards (two Visa, four MasterCard) from five different banks, and a terminal identical to those that can be found in stores in the U.K., Europe and the U.S.

This terminal was chosen to provide convenience for the research, as it contains a remote code execution vulnerability, which gives us a technological opportunity to intercept all data and monitor all risk management fields detailed in ISO-8583.

Figure 1. Terminal used for testing

4.4. Visa & MasterCard: Similar Idea, Different Implementations

Visa and MasterCard differ in their implementation of contactless payments, including different approaches on Tap-and-go limits. That's why we will explain the necessary steps for each of them separately. You can read more of our materials on about Visa and MasterCard limits in a separate paper.[8]

In order to conduct a testing attack on MasterCard cards, we had to use our own terminal. But for implementing attacks on Visa cards, we could have used unattended terminals, such as at a fast food chain restaurant or parking lot, since they did not require a terminal of our own.

In general, attacks on MasterCard cards are implemented in four steps:

1. An attacker needs to choose a terminal on which CDA is not used or where it can be downgraded to DDA.[9] Another option is to use an attacker-controlled terminal to reset limits in order to make purchases at other merchants. In our previous research, we

[8] First Contact: New vulnerabilities in Contactless Payments
https://drive.google.com/file/d/1KMvrdTgpw22Hvdgy4D_-ks_fW3WDEwT7/view?usp=sharing
[9] Practical EMV PIN interception and fraud detection
https://dev.inversepath.com/download/emv/emv_2014.pdf

have discussed how to obtain such terminals.[10] On our terminal, we disabled CDA, leaving only DDA supported. Contactless transactions on such terminals should be declined, but chip transactions may still be approved, potentially resetting the card's cumulative limit.

2. An attacker makes a chip payment for £1. The card transmits its CVM list to the terminal. By forging the CVM with a man-in-the-middle attack, we set offline PIN as the priority cardholder verification method. To the DDA scheme, this attack remains invisible.
3. The terminal chooses an offline PIN (encrypted or plaintext) and awaits the response of the PIN verification. When a wrong PIN was presented, we change the response from the card from “63C2” (verification failed, two tries left) to “9000” (PIN was checked correctly) and the terminal requests an online cryptogram. The card gets the cryptogram request, but has not received any PIN yet, so it puts a special flag in the IAD field. Different versions of cards use different bits. For some cards, the flag is "Offline PIN Verification Not Performed," and for others it’s, "Terminal Erroneously Considers Offline PIN OK." This data in the IAD field is transmitted to the issuer and cannot be modified because IAD is a part of the payment cryptogram.
4. The terminal sets all the values of the payment request and sends the data to the acquiring bank, from which it is then sent on to the issuer.

As mentioned, the issuer checks the transaction three times:

- During risk management, and at this stage, due to "Offline PIN Verification Not Performed" and "CDA / DDA failed" being set, the bank could have declined the transaction—and not reset the cumulative limit counter. However, we have encountered this specific event at three of the banks tested.
- Next, the integrity of the cryptogram is checked on the HSM. Since we did not replace any fields from the cryptogram input, this step will be successful.
- With the cumulative limit, the bank sees that the transaction was done with the chip and PIN. And even though there was no evidence that PIN was checked offline, the bank mistakenly resets the cumulative limit counter.

Now the hackers can go to any other store and make as many purchases as they want until they hit the £225 limit again.

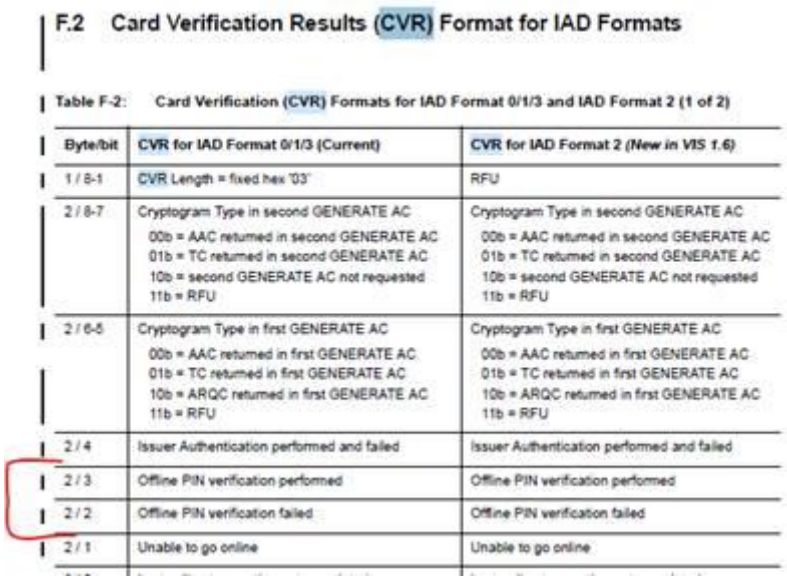
Attacks on Visa cards also take four steps:

1. An attacker with a stolen card who does not know the PIN can use a so-called wedge device in a man-in-the-middle attack and any unattended payment terminal.
2. An attacker makes a chip payment for £1. The card sends its CVM list to the terminal. By forging this list, an attacker sets the offline PIN as the priority CVM. This is possible

[10] For the Love of Money: Finding and Exploiting Vulnerabilities in Mobile Point of Sales Systems https://drive.google.com/file/d/1ADqAhqjcaVRr6jLA_GAuSngF5i-NsCxo/view

because the test Visa cards only used DDA mode during chip transactions for its offline authentication.

3. If PIN verification has failed, Visa cards will set up bits 2 and 3 of the byte 2 to clearly indicate a PIN OK attack.



F.2 Card Verification Results (CVR) Format for IAD Formats

Table F-2: Card Verification (CVR) Formats for IAD Format 0/1/3 and IAD Format 2 (1 of 2)

Byte/bit	CVR for IAD Format 0/1/3 (Current)	CVR for IAD Format 2 *(New in VIS 1.6)*
1 / 8-1	CVR Length = fixed hex '03'	RFU
2 / 8-7	Cryptogram Type in second GENERATE AC 00b = AAC returned in second GENERATE AC 01b = TC returned in second GENERATE AC 10b = second GENERATE AC not requested 11b = RFU	Cryptogram Type in second GENERATE AC 00b = AAC returned in second GENERATE AC 01b = TC returned in second GENERATE AC 10b = second GENERATE AC not requested 11b = RFU
2 / 6-5	Cryptogram Type in first GENERATE AC 00b = AAC returned in first GENERATE AC 01b = TC returned in first GENERATE AC 10b = ARQC returned in first GENERATE AC 11b = RFU	Cryptogram Type in first GENERATE AC 00b = AAC returned in first GENERATE AC 01b = TC returned in first GENERATE AC 10b = ARQC returned in first GENERATE AC 11b = RFU
2 / 4	Issuer Authentication performed and failed	Issuer Authentication performed and failed
2 / 3	Offline PIN verification performed	Offline PIN verification performed
2 / 2	Offline PIN verification failed	Offline PIN verification failed
2 / 1	Unable to go online	Unable to go online

Description of the attack indicators

4. The cryptogram is correct, and the issuing bank sees that the transaction was made with chip and PIN even though an offline PIN wasn't checked. The bank mistakenly resets the cumulative limit counter.

4.5. Attack Variations

There are also other ways to succeed in the attack. Cards use the same private keys and cryptogram versions for both contact and contactless chip transactions. That means that hackers can steal information needed to perform operations on their own terminals by means of mobile phones or special readers in a technique known as NFC relay or eavesdropping.

Another variation is to perform chip and signature transactions [13]. For this, in step one, an attacker would need to set the first cardholder verification method to "signature" and then sign the receipt.

4.6. Two Additional Methods Successful Against Individual Banks

A few additional checks and bypass mechanisms came up during the project, and we would like to share them too:

1. As mentioned already, one of the banks that we had tested implemented an additional check on the card itself. After five contactless transactions, the card stops responding via NFC and requires the chip instead. However, a cryptogram for a valid payment can still be obtained in two different ways:
- Attackers can request a cryptogram using a contact chip without providing an offline PIN.
- Attackers can reuse a cryptogram that already has been requested and authorized in the past. This is known as a replay attack,[11] which has been described previously.
2. One of the MasterCard issuing banks correctly detected our PIN OK attack and didn't allow resetting the limit via chip and signature. After several trials we put "Online PIN" in the CVM Results field. In the result the encrypted PIN would be attached to the payment request, and would have been decrypted and checked on the HSM, just like in a regular chip and pin transaction. However, we were not able to present the PIN code, because we didn't know it. Due to an issue with the card authorization software, the transaction was authorized and the cumulative limit was reset to 0. If not for a bug in the software, such a situation should never occur.

Figure 5 contains a table summarizing the results of our testing. One bank in particular secured its cards against all methods that we attempted. A "+" indicates that the attack was successful; a "-" indicates that our attack attempt failed.

Bank	Card	PIN OK	Chip&Signature	Additional methods
#1	Visa #1	+	-	Replay
#2	Visa #2	+	+	N/A
	MC #3	+	+	N/A
#3	MC #4	-	-	N/A
#4	MC #5	+	-	N/A
#5	MC #6	-	-	Online PIN

Figure 4. Table of results

5. Conclusion

While regulations in the financial industry are usually one of the main drivers toward more secure financial operations, our research shows that this alone is not enough. Adopting PSD2 and SCA complacency measures significantly decrease opportunities of fraud, and yet, there are still ways for malicious actors to commit fraud. Banks and financial institutions need to take proactive measures in addition to those required by regulations to reduce possible fraud surface and increse entry barriers for high-profile fraud.

[11] First Contact: New vulnerabilities in Contactless Payments
https://drive.google.com/file/d/1KMvrdTgpw22Hvdgy4D_-ks_fW3WDEwT7/view?usp=sharing

Most common steps toward a proactive stance are communications to the hacker community through bug bounty programs, inboarding security research partners, or leveraging internal capabilities.

See the following section for some recommended mitigation measures to consider.

6. Recommended Mitigation Measures

We would like to discuss a few ways to address issues that were revealed in the research. Here are key points for protecting EMV and NFC cards against known attacks:

- If the offline PIN is not checked on the card, the card should set the corresponding bits in the IAD field, which later should be checked on the HSM; such transactions should be declined.
- An offline PIN correctly verified on the card, or an online PIN correctly verified on the HSM, should be the only reason for resetting cumulative limits. Other options include resetting via apps, which many banks allow.
- We recommend against using the same cryptogram versions or symmetric keys for NFC and chip transactions. When this is combined with checking that the right keys have been used on the HSM, fraudsters cannot steal data via NFC and reuse it later for chip transactions (or vice versa).
- Keep an eye on the types of offline data authentication that have been attempted and failed: CDA/DDA from the TVR and IAD fields in particular. This information by itself is not enough to block cards but could be a sign of a malfunctioning terminal or stolen card.

7. Glossary of Terms

[1] Cardholder Verification Methods (CVM) – Payment verification methods used to validate (authenticate) that the person presenting the card is the valid cardholder. Common methods are chip and PIN or chip and signature.

[2] PSD2 – A directive regulating the payment market in Europe.

[3] Tap-and-go – A payment scheme for contactless transactions when no cardholder verification is required under a certain limit, such as £45 in the U.K.

[4] No CVM – A payment where no cardholder verification methods are used. This method will be chosen during tap-and-go payments.

[5] Cumulative limits – The maximum amount of money that a cardholder can spend using contactless payments before being forced to use chip and PIN to enter PIN. The current agreement across EU/U.K. banks is to allow five transactions, which may also be represented by £150 until March 2020 or £225 after March 2020 (maximum amount of one payment in the U.K. has been increased up to £45 due to COVID-19 epidemy).

[6] Issuer Application Data (IAD) – Contains proprietary application data for transmission to the issuer in an online transaction.

[7] Terminal Verification Results (TVR) – The TVR is a series of bits set by the terminal reading an EMV card, based on logical tests (for example, has the card expired). This data object is used in the terminal's decision whether to accept, decline, or go online for a payment transaction.

[8] CVM results – Indicates the results of the cardholder verification performed during the transaction

[9] Hardware Security Module (HSM) – A physical computing device that safeguards and manages digital keys, performs encryption and decryption functions for digital signatures, strong authentication, and other cryptographic functions.

[10] Chip and PIN – One of the two verification methods that EMV-enabled cards can employ. Rather than physically signing a receipt for identification purposes, the user just enters a personal identification number (PIN), typically of four to six digits. This number must correspond to the information stored on the chip. Chip and PIN technology makes it much harder for fraudsters to use a found card, so if someone steals a card, they can't make fraudulent purchases unless they know the PIN.

[11] CDA – An offline data authentication scheme that protects some risk management fields passed from the card such as CVM List. This is implemented to tackle the PIN OK attack.

[12] DDA – An offline data authentication scheme. When the terminal generates a random number, the card signs it with a private key and returns to the terminal along with a public

key. To ensure that the card is genuine, the terminal checks that the unencrypted UN is exactly the same, which has been passed to the card. This scheme doesn't protect the integrity of any fields from the card.

[13] Chip and signature – Differentiates itself from chip and PIN by verifying a consumer's identity with a signature.