Sitek and Kotulski *EURASIP Journal on Information Security* (2018) 2018:5
https://doi.org/10.1186/s13635-018-0076-9

EURASIP Journal on Information Security

RESEARCH Open Access

POS-originated transaction traces as a source of contextual information for risk management systems in EFT transactions

Albert Sitek* and Zbigniew Kotulski

Abstract

Transaction traces analysis is a key utility for marketing, trend monitoring, and fraud detection purposes. However, they can also be used for designing and verification of contextual risk management systems for card-present transactions. In this paper, we presented a novel approach to collect detailed transaction traces directly from payment terminal. Thanks to that, it is possible to analyze each transaction step precisely, including its frequency and timing. We also demonstrated our approach to analyze such data based on real-life experiment. Finally, we concluded this paper with important findings for designers of such a system.

Keywords: Payment transaction, Transaction traces, Traces analysis, Context, Contextual security, CVM

1 Introduction

The design and evaluation of new systems, algorithms, and techniques heavily rely on the existence of productive datasets. There are number of papers in the literature that focus on this issue. For example, in [1], one can find comprehensive overview of existing datasets to evaluation of cyclostationarity-based network Intrusion Detection Systems (IDSs), together with their own dataset with real traffic and up-to-date attacks. To the Authors' best knowledge, there is no such a dataset with transaction data from a payment terminal available that will allow to design and evaluate new solutions in payment ecosystems.

An acronym POS stands for Point-of-Sale: the device that represents the most critical function in a retailer's supply chain, the checkout process. The cash register is the moment of truth at which the consumer must commit to a purchase and offer tender in exchange for the goods [2]. In the interest of all is to make that process quick, safe, optimal, and comfortable. One of the possible ways to achieve that is to perform extended analysis of data gathered from a POS device and to introduce necessary improvements. This data is an effective source of information about shoppers, their purchases, and behaviors and is commonly used in retail environment for:

- Analysis of purchase trends
- Demand planning and forecasting
- Consumer marketing programs
- Strategic account development

Extensive studies have been conducted by various researchers in that field. For example, authors of [3] emphasized an importance of Point-of-Sale data sharing among suppliers in terms of demand forecasts. On the other hand, the paper [4] points out an impact of POS data inaccuracy and inventory data errors. An thought-provoking way of POS data utilization has been described in [5]. Authors proposed there the way how to track usage patterns of residential pesticides based on those data. As we can see, POS-originated data can be used for various purposes. One of them will be described later in this paper.

From the technical point of view, a POS device usually consists of two physical devices:

(1) A POS terminal: the main unit, responsible for tracking and recording customer orders, managing inventory, creating printouts, etc.

*Correspondence: asitek@tele.pw.edu.pl
Institute of Telecommunications of WUT, Nowowiejska 15/19, 00-665 Warsaw, Poland

(2) A payment terminal: the peripheral, autonomous device responsible for performing electronic transaction. Usually, it is a programmable Personal Identification Number Pad (PIN Pad, without a printer)

In this paper, we are focusing only on data gathered exclusively from a payment terminal. Those data will be presented as a source of contextual information for a dedicated risk management systems.

The rest of this paper is organized as follows: Section 2 provides the technical background to fully understand the consecutive sections, Section 3 describes the performed experiment in details, and Section 4 contains the experimental results, while Section 5 summarizes the paper and maps out future work.

2 Problem formulation and related works

This section formulates the problem and provides the crucial background to understand it in hand. Section 2.1 summarizes the information about card-present transaction and electronic card ecosystem. Section 2.2 presents the current utilization of transaction's logs collected in various points of the payment system. Section 2.3 describes the concept of context-aware systems and security. Finally, Section 2.4 discloses the usage of contextual risk managements in payment systems and explains the motivation and goal of this paper.

2.1 Card-present transaction overview

Payment card transactions are gaining more and more popularity in each country across the world [6]. Only in Poland, there were more than 1.1 billion transactions performed in Q2.2017 [7]. People are getting used to pay by card instead of paying in cash. They are commonly treating electronic payments as fast and secure way to finalize the transaction. It is because of increasing adoption of chip cards [8] compliant with EMV specification [9]. This standard has been firstly proposed by Europay, MasterCard, and Visa in 1993. Currently, it is promoted by EMVCo which associates all major Payment Card Schemes: Mastercard, Visa, American Express (AmEx), Discover, China UnionPay (CUP), and Japan Credit Bureau (JCB), and covers both contact and contactless payment cards. According to some statistics [9], transaction made with contactless card is 53% faster than a traditional magnetic stripe credit card transaction and 63% faster than using cash. There is also emerging trend observed on the market to emulate contactless payment card with a smartphone [10], thanks to services like Android Pay [11] or Samsung Pay [12] that uses Host Card Emulation technique (HCE) [13] and Near Field Communication (NFC) interface [14]. Thankfully, a smartphone emulating payment card is treated and read by the payment terminal as a physical card, so no changes are required in the payment infrastructure to handle those devices correctly.

Figure 1 presents standard payment card scheme. It involves four parties:

- Cardholder: a person that is buying goods or services at the Point-of-Sale and wishes to pay by card
- Merchant: a shop that accepts payment cards

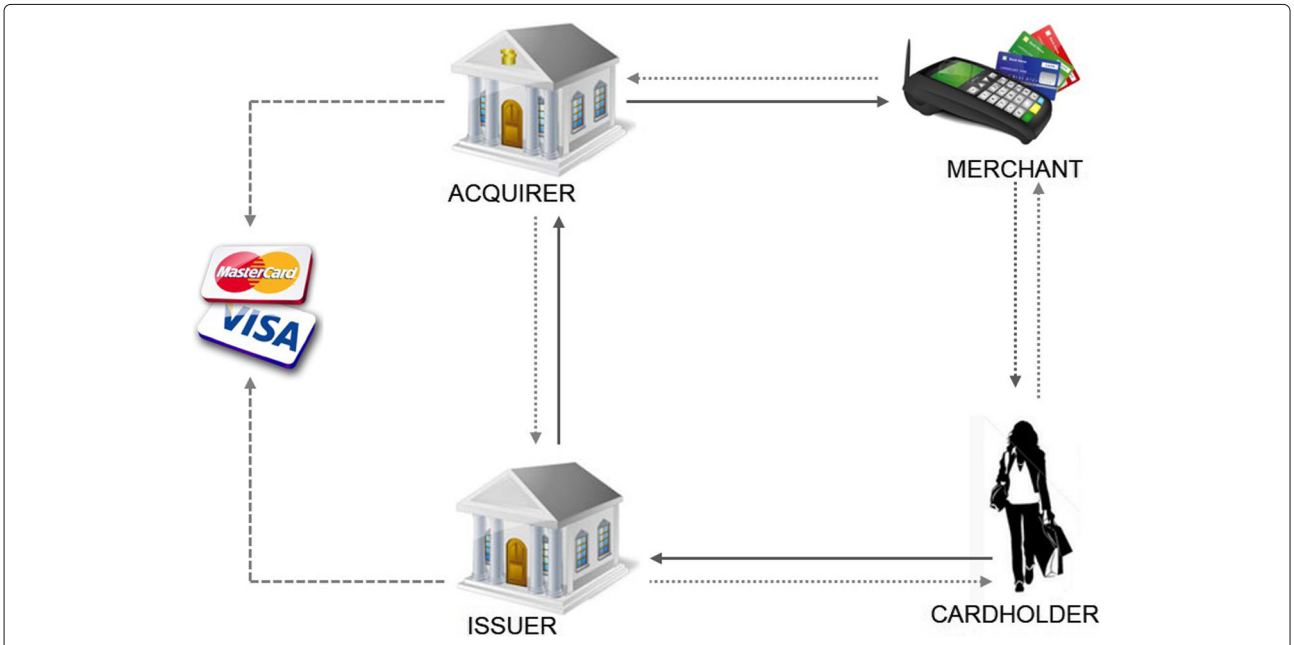

Fig. 1 Payment card scheme architecture

- Acquirer: an institution that is responsible for transaction processing and settlement
- Issuer: an institution, usually bank, that issued payment card to the Cardholder

It is worth to mention that payment transaction can be sent from the Acquirer to the Issuer directly (the so-called on-us) or via payment scheme network (Visa/Mastercard, etc.). So, payment organizations may not be aware of all transactions made by certain Cardholder.

Payment transaction compliant with the EMV specification consists of several steps [15]. Figure 2 depicts in details all possible transaction flows that can happen for both contact and contactless cards. The most remarkable steps that have a significant impact on the transaction processing time are Cardholder Verification and Transaction Authorization. The Cardholder can be verified by following Cardholder Verification Methods (CVMs):

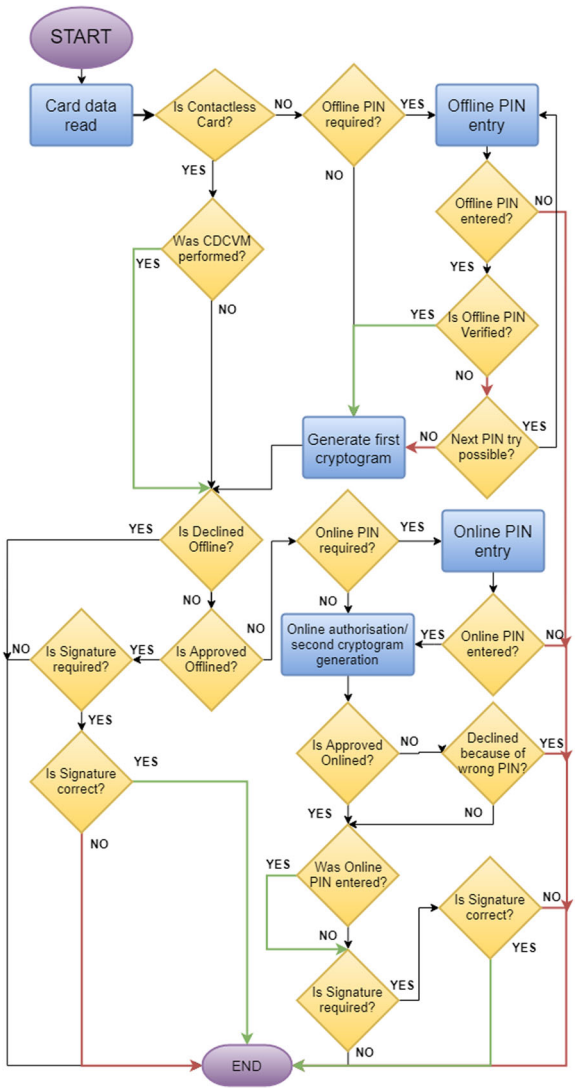

Fig. 2 Transaction's flow diagram

- Offline PIN: verified by the card, only for contact EMV
- Consumer Device CVM (CDCVM): verified by the device, only for HCE transactions
- Online PIN: verified by the Issuer
- Signature: verified by the Merchant
- No CVM: no verification at all

On the other hand, transaction can be authorized in following way:

- Online: by sending an authorization request to the Issuer
- Offline: locally, by the card

The decision which Cardholder Verification Method and authorization method should be used is being made based on terminal's configuration and data retrieved from the card (encoded on the card by the Issuer during its personalization phase). Those parameters are for example:

- Cardholder Verification Limit (CVL), only for contactless transactions, the amount above which the Cardholder must be verified: currently 50 PLN in Poland
- Floor Limit, the amount above which the transaction must be authorized online

2.2 Payment transaction traces analysis

As one can easily see in Fig. 1, there are three points, where transaction traces can be collected and analyzed: by Merchant, Acquirer, and Issuer. In Section 1, we described what can be done with transaction data collected by the Merchant. This information is rich; it contains payment credentials, content of the basket for particular transaction, some loyalty information, and so on. In this section, we are focusing on transaction traces that can be collected by Acquirer and Issuer. Usually, those traces contain only pure payment-related information without any details about Cardholder or content of the basket. Transaction traces at the level of Acquirer can be used only for statistical purposes. We must note that an Acquirer is not aware of all transaction made by some Cardholder, but only of those performed by the Acquirer's terminal. On the other hand, transaction traces from the Issuer contain information about all transactions made by the Cardholder. Therefore, this data is profitable and worthwhile for various purposes [16]:

- Law enforcement: Law enforcement agencies can subpoena records from the credit card Issuer to find out the time, date, and place of a credit card purchase,

in other words information that may be helpful in determining the last known location of a crime victim or suspect. National security agencies also track terrorist activity by monitoring certain purchases.
- Marketing: Issuers use past purchasing patterns as a basis for offering additional products. Someone purchasing airline tickets with their credit card may get offers of airline rewards credit cards or travel-related services from the Issuer or an affiliate.
- Risk management: Cardholders who continually go over their credit limits or exhibit unusual spending habits (such as charging large amounts of merchandize on a card they had previously rarely used) may be at greater risk of not paying their bills or filing for bankruptcy.
- Fraud detection: Credit card companies and Issuers monitor spending to detect unusual purchasing habits that could be red flags for fraud. This topic is quite popular across researchers. There are plenty of fraud detection systems proposed in the literature that incorporate various data mining techniques like neural network [17, 18], outlier detection [19], data clustering [20], support vector machine [21], artificial immune system [22], self-organizing maps [23–25], hidden Markov model [26], Bayesian classifier [27], fuzzy systems [28], and genetic algorithm [29]. Some of them has been tested using real transaction traces gathered, e.g., from a bank from UK [30, 31].

2.3 Context-aware systems and security

According to the Dictionary [32], context can be defined as follows: *context is the set of circumstances or facts that surround a particular event, situation, etc.* In other words, it can be described as a set of information about the entity in particular moment in time. An author of [33] divided contextual information into two parts:

- Internal, which describes user state such as emotions, look, and history of life
- External, which refers to environment states such as time, location, and temperature

The utilization of contextual information for various solutions in a very popular topic among researchers. They started to propose the so-called context-aware systems. Those systems can adapt their operations to the current context without explicit user intervention and thus aim at increasing usability and effectiveness by taking environmental context into account [34]. An example of such a system can be a solution proposed in [35]. Authors extended the instant messaging paradigm by adding context awareness to support information management within a hospital. All users (in this case nurses, physicians, etc.) are equipped with mobile devices to write messages that are sent when a desired set of conditions are met. For example, a user can formulate a message that should be dispatched to the first doctor that enters room number 110 after 6 am. Such a system is aware of following contextual elements: time, roles, location, and device state.

Another group of context-aware systems concerns various security services, like access control, encryption (information confidentiality), and authentication. They are known in the literature as context-aware security systems [36]. For example, authors of [37] proposed the solution for mobile ad hoc networks that analyzes various contextual information, such as communication channel status, battery status, and weather condition, and then uses them to determine whether the misbehavior is likely a result of malicious activity or not. Other examples of context-aware security (in terms of access control) can be found in [38–40].

Context-aware systems are usually capable to deal with special types of context and are tailored and optimized to operate in specific conditions, e.g., in hospital scenarios or ad hoc networks. Unfortunately, they do not have to be flexible and extensible. To simplify the development of context-aware applications, researchers started to propose the so-called context-aware frameworks [41]. An abundant overview of such frameworks can be found in [34].

In the next section, we will present how context-aware systems can be utilized during card-present transactions.

2.4 Contextual risk management in EFT systems

As mentioned in Section 2.1, card-present transactions can have plenty possible scenarios. The way how the transaction is being processed depends on various factors: transaction amount, card used, data stored on the card, and terminal's configuration. We can say that transaction processing rules are constant for every transaction: it means that each Cardholder is treated equally, no matter what is his history and the context of current transaction. There are also clear rules regarding risk related to the transaction. If a disputed transaction has been authorized:

- With signature verification, then it would be charged to the merchant.
- With PIN verification, then it would be charged to the customer.
- By the card (Offline Authorization), it would be charged to issuing bank.

Such an approach is effortless, but it causes that a lot of transactions are processed "time and user experience-ineffectively" [42]. One can imagine that the transaction flow could be tailored to the Cardholder and to the particular transaction, based on various contextual factors.

It may give a lot of profits, e.g., better user experience, greater Cardholder's loyalty, and shorter transaction processing time. It should also assure acceptable level of transaction security. This is the main motivation why contextual risk management systems started to appear. They enable merchants to take some risk by allowing some payment transactions being authorized, for example, without any verification in exchange for the abovementioned profits. Such systems could be very useful on the markets, where the level of fraudulent transactions is low. For instance, such an information can be found in the European Central Bank's report [43], which says that the level of deceptions is very low in certain countries.

The topic of contextual risk management was already considered in our earlier papers. It was firstly discussed in [44], where we proposed a new Cardholder Verification Method which is One-time PIN verification. This method assumes that each transaction is authorized online and the decision if PIN verification should be performed is being made by the Issuer based on various contextual factors (like Cardholder's reputation, place, and time of the transaction). In the case of positive decision, encrypted PIN (or One-time PIN) is being sent to the terminal and a payment application verifies, if the encrypted PIN entered by the Cardholder is the same as the one received from the Issuer. This paper proposed only a modification of current payment system architecture but did not suggest any decision making algorithm.

Another approach has been discussed in [45], where we proposed the dedicated solution for great merchants (such as Carrefour and Auchan), where historical transaction data are kept on merchants' servers. The proposed architecture (depicted in Fig. 3) assumed that payment terminal sends contextual information (the tokenized card's number, transaction's amount, merchant's location, etc.) to the merchant's server and it receives back the decision whether the transaction should be authorized "offline" or "online". We also proposed a simple example of the algorithm that calculates floor limit for current transaction based on the transactions' periodicity factor, transactions' amount stability factor, and Cardholder's reputation. However, the quoted reputation system has a few flaws; for example, it is not possible to detect the situations where:

- The Cardholder enters the Offline PIN with success in the second attempt. Such a transaction is treated the same as that with PIN verified on the first try.
- He cancels the transaction on PIN entry screen. Such a behavior is suspicious, but it is not taken into account at all.
- A transaction with Online PIN that has been declined because of lack of funds (but PIN has been verified correctly). Such a situation should increase Cardholder's reputation, because the key information for the system is that the Cardholder knew the PIN code, but it works against it.

To extend and improve such a situation, in the paper [42], we proposed a new Cardholder's reputation system that can be utilized in contextual risk management solutions for payment transactions. This reputation system covers all possible transaction flows. Each transaction flow has constant rating assigned to it. The set of ratings for all possible transaction flows are parameters of the reputation system. The Cardholder's reputation for a forthcoming transaction n can be calculated as weighted average of the last N transactions limited to the range $< R_{\mathrm{MIN}}, R_{\mathrm{MAX}} >$, see Eq. (1). N, R_{MIN}, and R_{MAX} are parameters of the reputation system.

$$R_n = \begin{cases} R_{\mathrm{MIN}} & \text{if } \overline{R}_{n-i} < R_{\mathrm{MIN}} \\ \overline{R}_{n-i} & \text{if } \overline{R}_{n-i} \in \langle R_{\mathrm{MIN}}, R_{\mathrm{MAX}} \rangle \\ R_{\mathrm{MAX}} & \text{if } \overline{R}_{n-i} > R_{\mathrm{MAX}} \end{cases}, \tag{1}$$

The proposed reputation system assumed that there must be at least N historical transaction stored in the system's database to calculate proper reputation value; otherwise, Cardholder's reputation is set to 0. Equation (2) shows the proposed formula how to calculate weights for the weighted average computation.

$$w_{Rni} = \frac{1}{2} e^{-\frac{t_n - t_i}{\tau_{RT} * AvgT}} * \mathrm{erfc}\left(\frac{(n-i-1)*2}{x_d} + x_m \right), \tag{2}$$

where n is the index of current transaction, i is the index of ith transaction, t_n is the time of current transaction, t_i is the time of ith transaction, $AvgT$ is the average distance (in time, measured in days) between transactions, τ_{RT} is the reputation system parameter (the decay factor), erfc is the complementary error function, x_d is the reputation system parameter (a dispersion parameter of the erfc function), and x_m is the reputation system parameter (a concentration parameter of the erfc function [46]).

Every presented approach is focusing only on internal part of contextual information. Comprehensive contextual risk management system should be able to take into account external part of contextual information. This is the subject of our future research.

All of mentioned papers presented various enhancements for current card payment ecosystem; however, all of them were tested using synthetic sets of data (prepared based on experts' knowledge), because of the lack of realistic production data. To our best knowledge, there is no solution known from the literature that describes a way how the traces directly from a payment terminal can

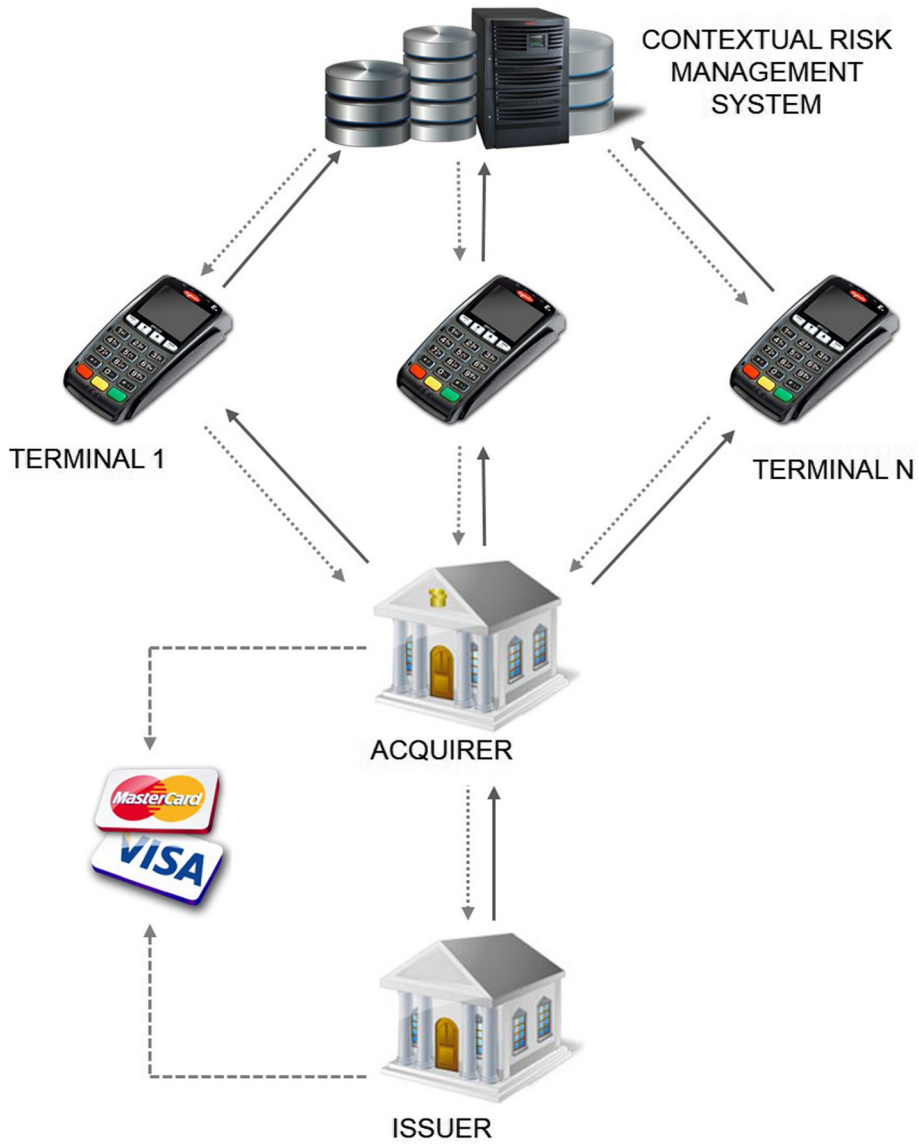

Fig. 3 Contextual risk management system architecture

be created, collected, and analyzed for risk management system's improvement.

In this paper, we present our approach to gather and analyze transaction traces collected directly from a payment terminal. Thus, in Section 3, we describe our experiment performed on large-scale production data. In Section 4, we present key results of the experiments and we show what useful information can be extracted in terms of design and configuration of the contextual risk management system.

3 The experiment

To conduct our experiment, we performed several steps, which are presented in the following sections.

3.1 Design and implementation of application's extension

We designed and implemented an extension to the existing payment application that allows it to record detailed information about each transaction. During the transaction, payment application records all important data and events related to the given transaction and stores it as an entry in transactions' traces batch file.

Table 1 presents the structure of one entry in the mentioned batch, while Table 2 lists all possible events that can be included in the trace. Figure 4 shows an example transactions' traces batch. Such a structure of a batch's entry has a few advantages:

Table 1 Structure of single transaction's trace

Pos.	Meaning	Details
1	Card's token	
2	Card read type	2: magstripe
		3: contact EMV
		5: contactless magstripe
		6: contactless EMV
3	Transaction's date	Format: YYYYMMDD
4	Transaction's time	Format: HHMMSS
5	Transaction's type	1: sale
		6: refund
6	Transaction's amount	
7	Transaction's trace	Sequence of events from Table 2

(1) There is a possibility to measure how long it took the Cardholder to enter PIN code, place a signature, etc.
(2) It is possible to detect the situation when the transaction was interrupted by the user, for example, by canceling the PIN entry.

Such an information would not be possible to retrieve from other viewpoint, e.g., from the Acquirer or the Issuer perspective.

The essential thing in the batch's entry is that we are using card's token to unambiguously identify certain payment card without revealing its Primary Account Number (PAN). The way how a token is calculated is fully compliant with Payment Card Industry (PCI) requirements [47], and it looks as follows:

$$\text{token} = \text{SHA256}\,(\text{PAN}|\text{EXP\_DATE}|\text{SSS}) \tag{3}$$

Table 2 Possible events to be recorded in transaction's trace

Event	Meaning	Timestamp	Value
crs	Card read started	X	
cr	Card read	X	
cp	CDCVM performed		
pofs	Offline PIN started	X	
pofc	Offline PIN canceled	X	
poff	Offline PIN failed	X	
pofv	Offline PIN entered	X	
ofd	Offline declined		
ofa	Offline approved		
pons	Online PIN started	X	
ponc	Online PIN canceled	X	
onr	Online result	X	X
2ar	2nd AC rejected		
ss	Signature started	X	
sf	Signature failed	X	
sv	Signature verified	X	

where:

- EXP_DATE: expiry date of the payment card
- SSS (Strong Secret Salt): securely distributed across all terminals set of constant 32 bytes

Thanks for that, the batch of transactions' traces:

(1) Can be processed in systems that are not PCI DSS (Payment Card Industry Data Security Standard) certified, because the batches do not contain sensitive data
(2) Contains card's token, which is only the pseudonym in terms of General Data Protection Regulation (GDPR) [48].

The GDPR is the European Union (EU) regulation that will take effect on May 25, 2018. It offers a new framework for data protection with increased obligations for organizations, and its reach is far and wide. The GDPR is applicable to any organization that intentionally offers goods or services to the European Union or that monitors the behavior of individuals within the EU. The GDPR defines pseudonymization as "the processing of personal data in such a way that the data can no longer be attributed to a specific data subject without the use of additional information" [49]. Moreover, the Regulation says that if it is possible to demonstrate that the System is not in a position to identify the data subject (presented case), Articles 15 to 20 shall not apply. Those Articles contain regulations about:

- Right of access by the data subject
- Right to rectification
- Right to erasure ("right to be forgotten")
- Right to restriction of processing
- Notification obligation regarding rectification or erasure of personal data, or restriction of processing
- Right to data portability

In other words, it would be much more easier to meet GDPR requirements after productive rollout of the system, thanks to the utilization of pseudonymization. Additionally, the GDPR provides an exception to the purpose limitation principle for data processing for scientific, historical, and statistical research; pseudonymization is sufficient to meet its requirements, so similar experiments in the future are possible.

3.2 Data collection

Each payment terminal in the field is managed remotely by the central system called terminal management system (TMS). It is responsible for the management of terminal's configuration, monitoring its internal state, etc. Most of TMSs have an ability to receives some files from the terminal. Usually, a payment terminal is connecting to its TMS

19755E51E42F043AFDE2592DFCDD9B45183E22C64DE4C664176F7605F59F52EB;6;20170319;203012;1;100;[{"evt":"crs","ts":"1489955413"},{"evt":"cr","ts":"1489955417"},{"evt":"cp"},{"evt":"ofa"}]

6427FB9766A0DC0E5721BE8C20CC140D26841B8397BBD38338EEF6160938954D;6;20170319;203027;1;200;[{"evt":"crs","ts":"1489955434"},{"evt":"cr","ts":"1489955441"},{"evt":"cp"},{"evt":"ofd"}]

74CF25636B04D9B598299D7E49108E1CCD5C3BC87130CA755CC44F65AEF981C1;3;20170319;203054;1;20000;[{"evt":"crs","ts":"1489955458"},{"evt":"cr","ts":"1489955474"},{"evt":"onr","ts":"1489955476","val":"0"},{"evt":"ss","ts":"1489955482"},{"evt":"sv","ts":"1489955483"}]

F2D927AC74CA911036DCC011A9687748AB9EC12A2FC342CB82C8A7C219225052;2;20170319;203143;1;70000;[{"evt":"crs","ts":"1489955496"},{"evt":"crs","ts":"1489955503"},{"evt":"cr","ts":"1489955503"},{"evt":"pons","ts":"1489955503"},{"evt":"pone","ts":"1489955506"},{"evt":"onr","ts":"1489955508","val":"0"}]

Fig. 4 Example transactions' traces batch

once a day. Because transactions' traces batch is a single text file, we made a change in the payment application, so that the whole file was uploaded to the TMS each day.

3.3 Data analysis

Once transactions' traces batches has been collected, they must be analyzed somehow. In [50], one can find an extensive overview of currently available tools and utilities that are widely used by researched to data analysis and statistics.

We started our analysis from designing the database to store transactions' traces (see Fig. 5). Then, we created dedicated Python's script that can parse collected traces and that inserts data to the database. Finally, we created a set of Python's scripts to analyze collected data. We utilized the following libraries:

- Pandas: the library providing high-performance, easy-to-use data structures and data analysis tools for the Python [51]
- NumPy: fundamental package for scientific computing with Python [52]
- Matplotlib: plotting library for Python [53]

To write our scripts and to test them effectively, we used IPython [54], which is the system for interactive scientific computing allowing, for example, to execute certain parts of Python's code independently, without loss of the whole context (currently read data, variables, etc.). We also used Jupyter [55], which is a comfortable IDE (Integrated Development Environment) to write and execute IPython's code easily.

3.4 Experiment's details

Our experiment lasted 5 months. We uploaded our modified payment application onto 68 terminals located in 18 shops belonging to one of the retail chain. All those shops are in North-West region of Poland, near the border with Germany. During this period, we collected 1,048,382 transactions' traces.

4 Experiment's results

Results of our experiment can be divided into the following groups.

4.1 Analyses of transactions' amount and time

We performed several analyses concerning transactions' amount and time, considering precise transaction time, card used, etc. Such an information can be useful for proper configuration of contextual risk management system, for example, by focusing mostly on frequently used cards or by adjusting risk management parameters in time, according to the present number of transactions.

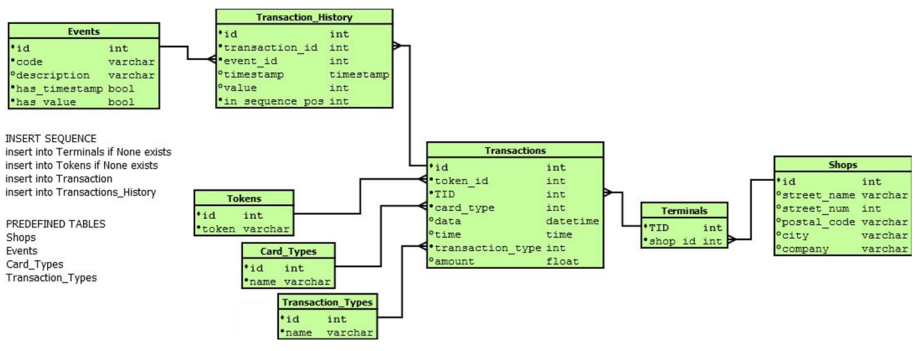

Fig. 5 Structure of database to store collected traces

Because of limited space in this paper, we are presenting the most noteworthy findings from the results.

Figures 6 and 7 illustrate our analysis for weekdays. The first one presents the quantity of transactions performed in time with a breakdown per card used, while the second one shows average transaction amount for all cards with a breakdown per card used. From those pictures, we can see that:

- There are around two to three times more transactions made with contactless EMV cards than those made with contact EMV cards.
- The level of other card types (contactless magstripe and magstripe) is insignificantly low.
- Customers are using contact EMV cards for high-value transactions. The average transaction amount for contact EMV cards is more than two times higher than that for contactless EMV cards.
- The highest frequency of transactions is between 15:30 and 18:30. There is around three times more transactions in this period than in the mornings.

Then, we did analogous analysis for weekends. Based on Figs. 8 and 9, we can see that:

- The ratio of transactions made with contactless EMV cards to that made with Contact EMV cards is comparable to weekdays.
- The level of other card types (contactless magstripe and magstripe) is still insignificantly low.
- The average transaction amount in time looks similar for weekends and for weekdays.
- Most transactions during weekends are made around midday and in the evenings.

4.2 Analysis of transactions' processing time

Figure 10 presents the transaction processing time of all examined transactions. As we can see, majority of transaction lasts for about 5–6 s, but still, there are plenty transactions with processing time around 10 s. Of course, it depends on various factors: card used, authorization method, Cardholder Verification Method, etc. On the other hand, Fig. 11 shows transaction processing time of

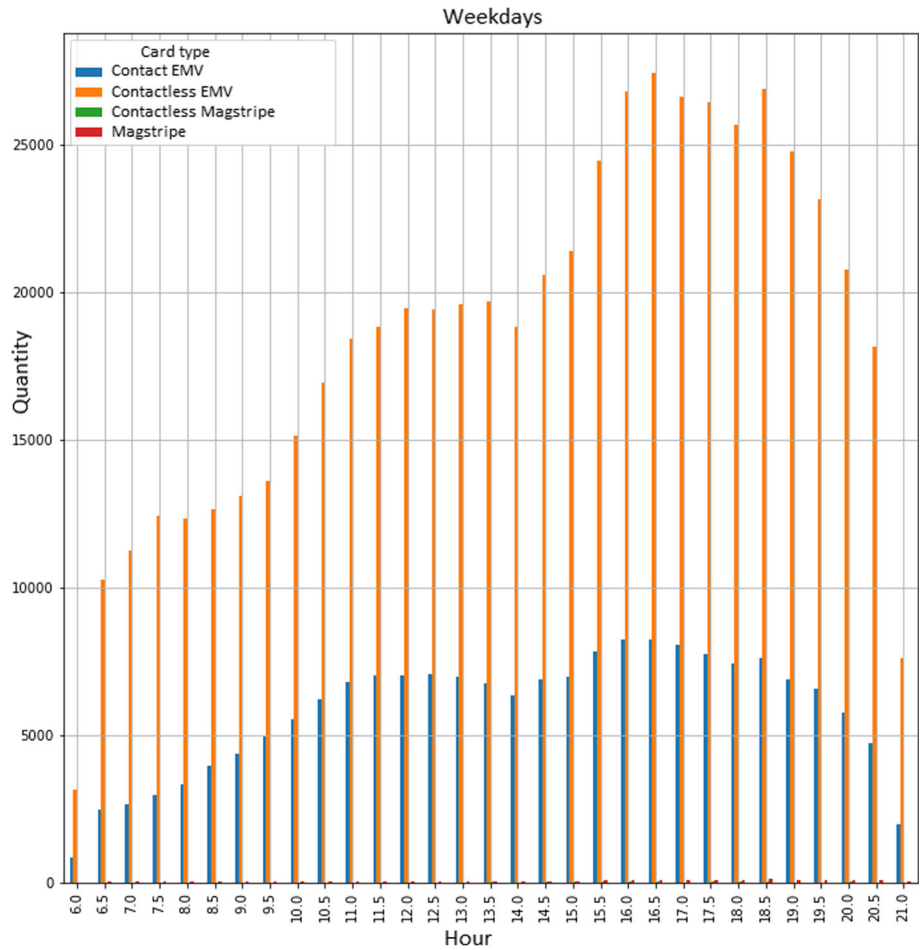

Fig. 6 Quantity of transactions in time with a breakdown per card type, during weekdays

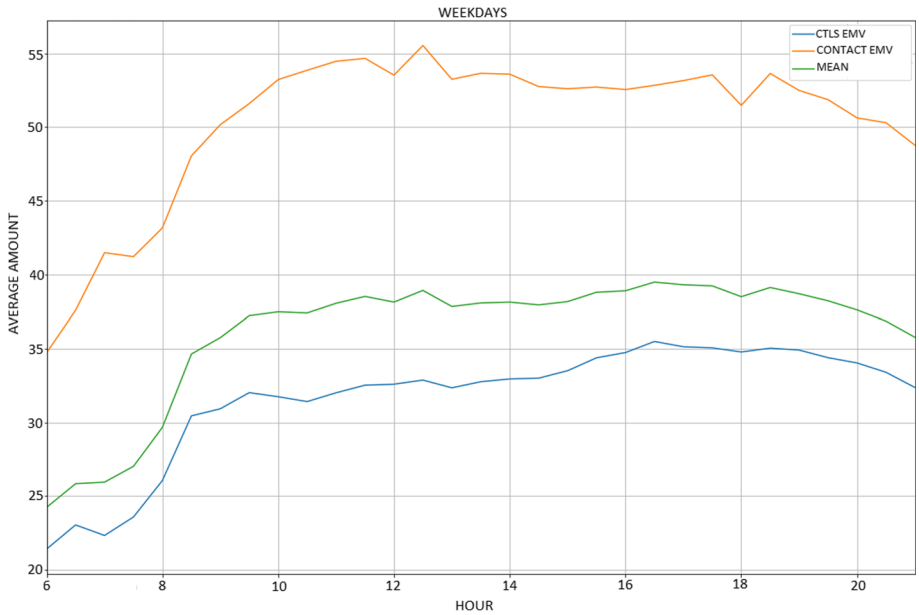

Fig. 7 Average transaction amount in time with a breakdown per card type, during weekends

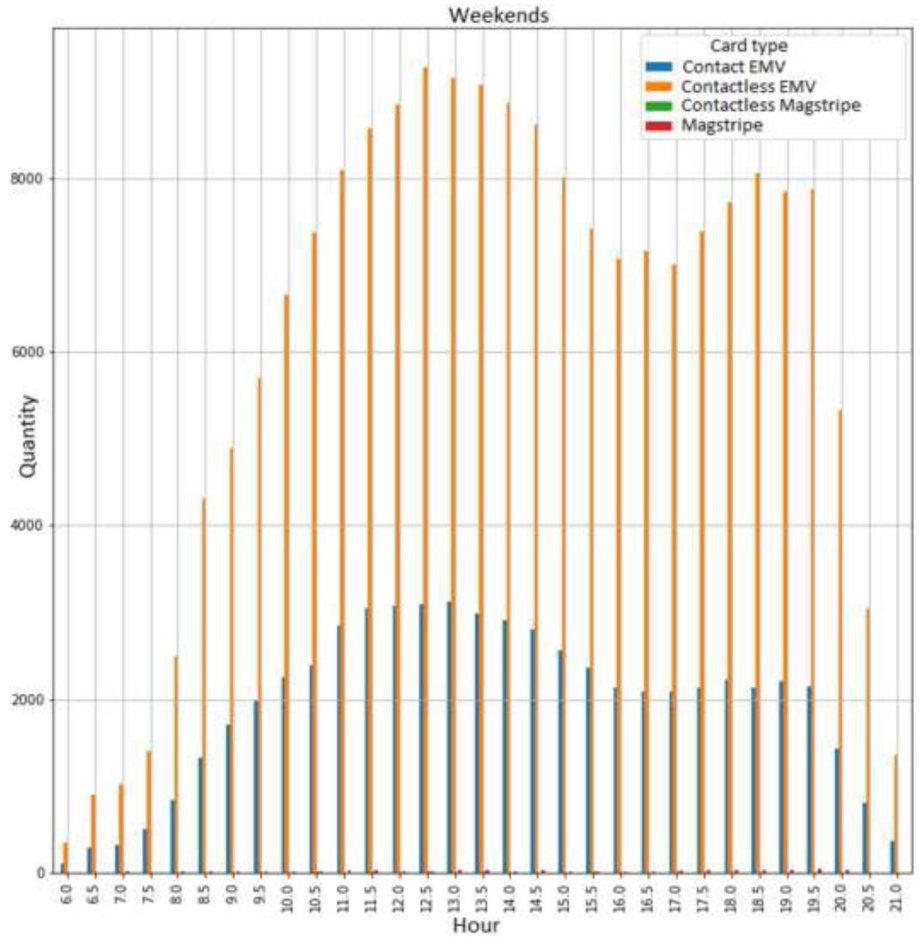

Fig. 8 Quantity of transactions in time with a breakdown per card type, during weekends

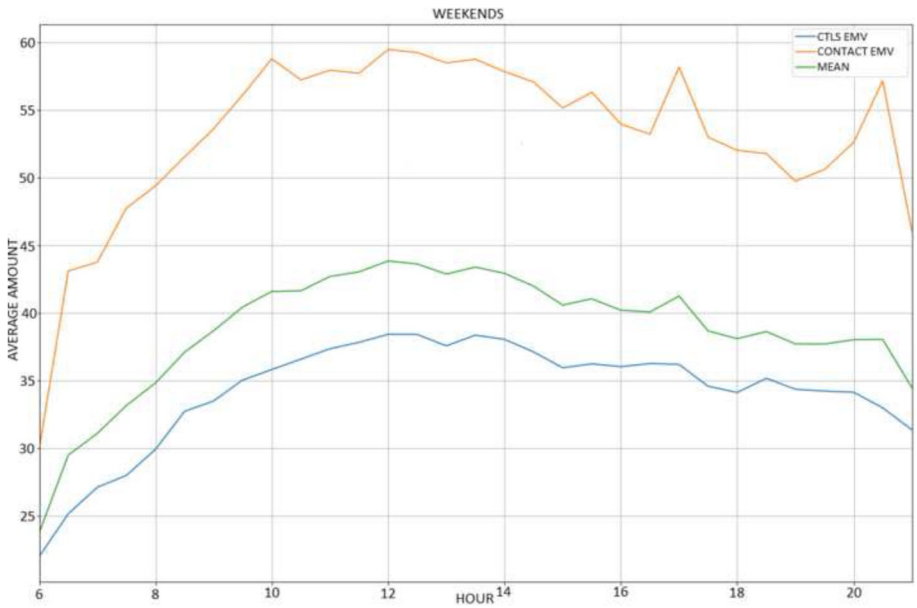

Fig. 9 Average transaction amount in time with a breakdown per card type, during weekends

transactions with PIN verification. In such a situation, it usually takes around 10 s to authorize the transaction.

Next, we will scrutinize possible gain of time that can be accomplished by incorporating contextual risk management system that increases CVL limit dynamically. Figures 12 and 13 present the theoretical maximal gain of time if CVL limit will vary from 0PLN to 200PLN, for all types of transactions and for contactless transactions, respectively. One must note that there is an inflection point at the level of 50PLN, because CVL limit for contactless transaction in Poland is equal to that amount. From those figures, we can see a few interesting regularities, for example:

1. By making CVL limit for contact transactions equal to that for contactless transactions (50PLN), we could gain more than 200 h (around 12,000 min).
2. By increasing CVL limit only for contactless transactions to 70PLN, we could gain more than 117 h (above 7000 min).

As we can see, even a small change in the CVL limit will bring benefits related to the processing time of the transaction.

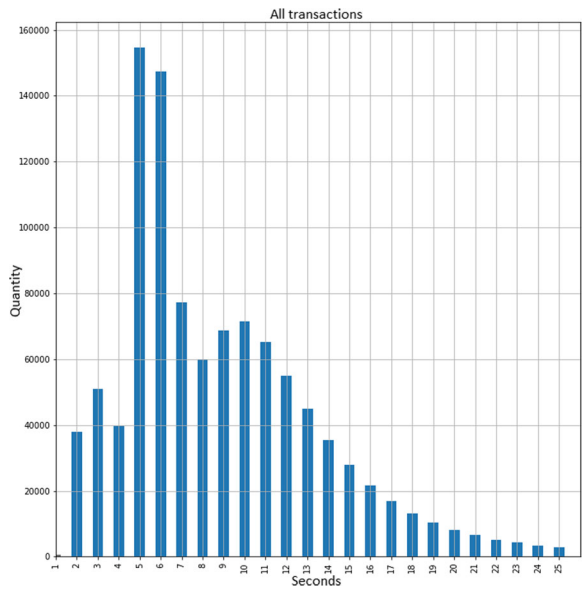

Fig. 10 Processing diagram of all transactions

4.3 Analysis of transactions' traces

Table 3 presents collected transactions' traces together with their quantities. The majority of transactions are processed without any Cardholder Verification and are authorized online. There are also three frequently happening transactions: Online Authorization with Online PIN verification, Offline Authorization without Cardholder Verification, and Online Authorization with Offline PIN verification. The level of other transactions' flows is negligibly low. It is really an important observation from a designer's contextual risk management system point of view. He can decide, for example, to ignore other transactions' flows and focus his algorithms to work only for commonly happening transactions. Such a simplification can have positive impact on performance of whole system and on cost of its implementation.

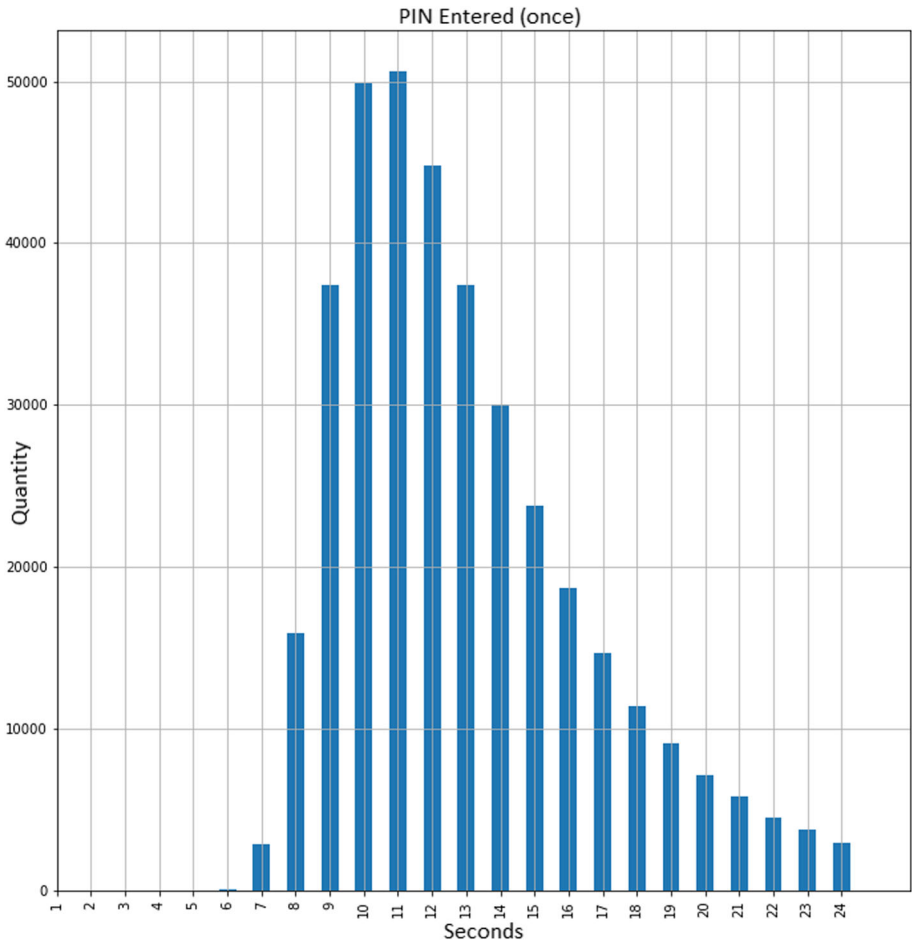

Fig. 11 Processing diagram of transactions with PIN verification

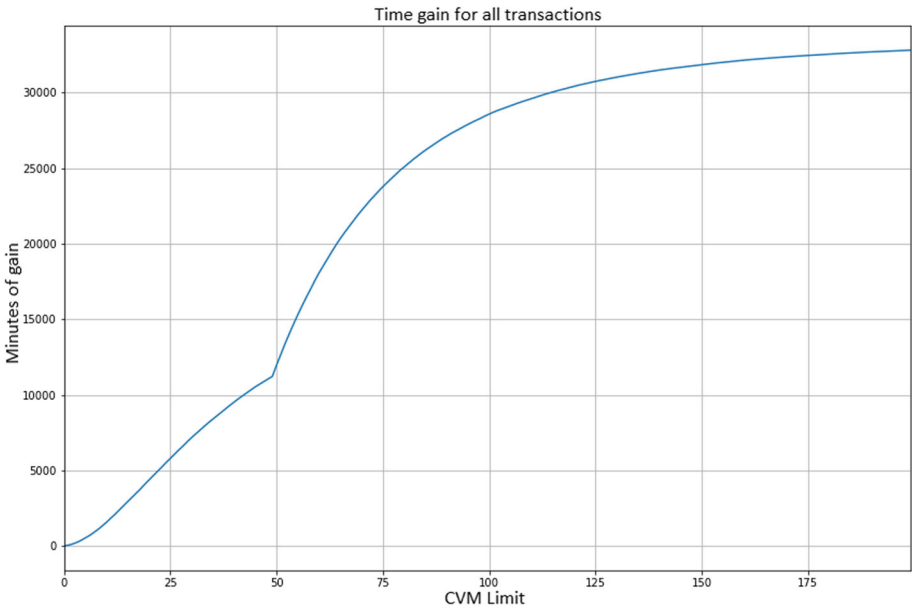

Fig. 12 Theoretical gain of time depending on CVL limit for all types of transactions

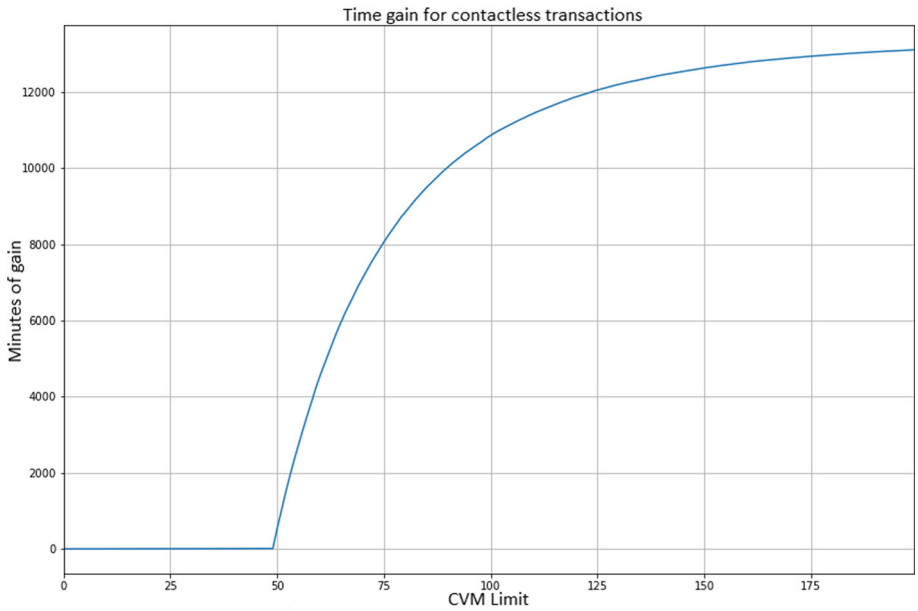

Fig. 13 Theoretical gain of time depending on CVL limit for contactless transactions

4.4 Analysis of clients' loyalty

During our experiment, we collected transaction traces for 189,898 unique card tokens. Figure 14 presents how many card tokens has been used to perform certain number of transactions. As we can see, almost 40% (73,229 pcs.) of unique tokens performed only one transaction. Two transactions have been performed by more than 16% (30,577 pcs.) tokens and so on. In our opinion, such a situation could happen because:

- The experiment has been conducted in Poland, near the Polish-German border, where there are a lot of tourists visiting this area and buying things occasionally.
- Nowadays, majority of Cardholders are using more than one payment card. From the designer's contextual risk management system perspective, it is worth to foresee the development of some web service where customers can log-in and link several payment cards to one account (for loyalty purposes) and design his system to operate on the level of a client rather than on pure token.

Despite the previous results, we observed quite a big set of loyal clients. We denoted loyal client a person, who performed at least 10 payment transactions using his card during our experiment. It turned out that there were 21,125 tokens with such many performed transac-

Table 3 Transactions' traces collected during experiment

Trace	Meaning	Quantity
CRS_CR_ONR	Onl. Auth. without CVM	500,925
CRS_CR_PONS_PONE_ONR	Onl. Auth. with Onl. PIN	262,762
CRS_CR_OFA	Offl. Auth. without CVM	152,103
CRS_CR_POFS_POFV_ONR	Onl. Auth. with Offl. PIN	126,478
CRS_CR_PONS_PONC	Onl. PIN canceled	1554
CRS_CR_CP_ONR	Online Auth. with CDCVM	1041
CRS_CR_POFS_POFC_OFD	Offl. PIN canceled	704
CRS_CR_ONR_SS_SV	Onl. Auth. with Signature	691
CRS_CR_POFS_POFV_POFF_POFS_POFV_ONR	Onl. Auth. with Offl. PIN verified on 2nd try	595
CRS_CR_POFS_POFV_OFD	Declined Offl. by the card, after Offl. PIN verification	258
Other (25 traces)	Other transactions' traces	1271

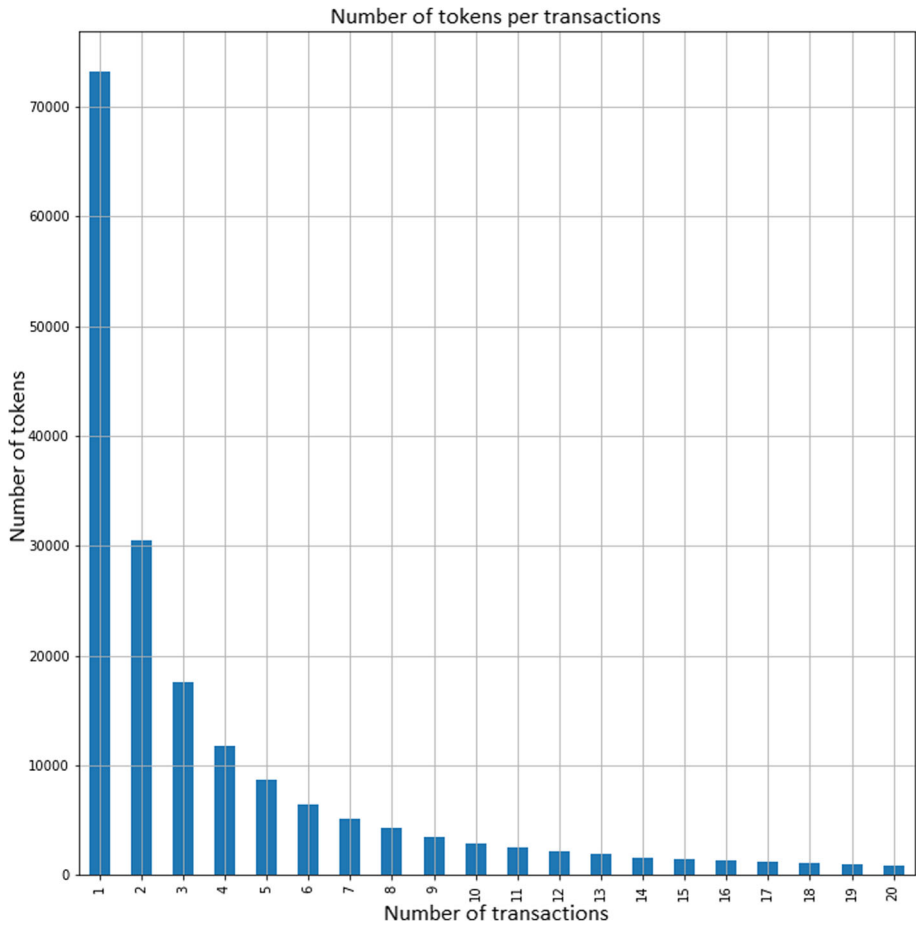

Fig. 14 Number of clients that performed certain number of transactions

tions, what is more than 10% of all cards. After that, we checked the proportions of loyal clients in certain stores. Figure 15 shows the number of loyal clients per store. As we can see, the level of loyalty across the stores varies. It is a good source of information that can help to prepare effective marketing strategy. The sum of all presented tokens is 19,042; it means that generally, loyal customers of one retail chain are usually making their transaction in one store.

5 Conclusion and future work

Transaction traces collected in various points of payment ecosystem are a valuable source of information for diverse analyses. In this paper, we introduced our approach to collect transaction traces at the level of payment terminal, which can inspire a new field of research. We also described our proposition how to evaluate those data with currently available tools and techniques. After that, we presented our experiment which involved the following: performance of changes in productive payment application, data collection, and its analysis. Based on the results of the experiment, we formulated important recommendations, especially for designers of contextual risk management systems for payment transactions, such as

(1) Design your system only for contact EMV and contactless EMV cards. The level of rest of the cards is negligibly low (see, e.g., Fig. 6).
(2) Design your system to be able to operate differently, depending on current time and day of the week.
(3) It is worth to design such a system even for small increase of CVL Limit (see Section 4.2).
(4) Design your reputation system to handle only most common transaction flows (see the level of other transaction flows in Table 3).
(5) Design your system to operate on a client level, not on a card token level (see Section 4.4).
(6) Perform analogous trial before configuring and launching such a system in new location.

It is worth to mention that gathered productive data can act as reference dataset for design and simulation of similar systems and solutions.

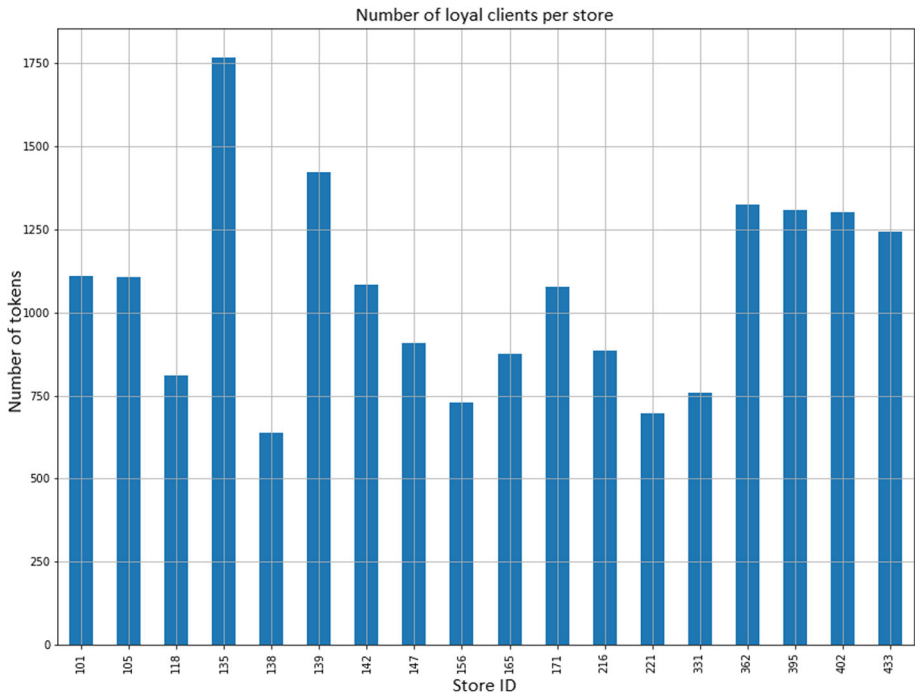

Fig. 15 Number of loyal clients (that made more than 10 transactions), per store

In future work, we would like to create an engine to simulate decision making algorithms from the contextual risk management systems, especially to validate and customize an algorithm proposed in [42]. Moreover, it would be valuable to perform an analogous experiment in a different region of the country, which is not impacted by many occasional consumers and tourists.

Abbreviations

CDCVM: Consumer device cardholder verification method; CUP: China UnionPay; CVL: Cardholder verification limit; CVM: Cardholder verification method; EFT: Electronic fund transfer; EMV: Europay mastercard visa; EU: European union; GDPR: General data protection regulation; HCE: Host card emulation; IDE: Integrated development environment; JCB: Japan credit bureau; NFC: Near field communication; PAN: Primary account number; PCI: Payment card industry; PCI DSS: Payment card industry data security standard; POS: Point of sale; TMS: Terminal management system; SSS: Strong secret salt

Acknowledgements

We are indebted to our commercial collaborator, who prefers to remain anonymous, for allowing us to perform our experiment.

Authors' contributions

Both authors contributed to the design, implementation, and the writing of the article. Albert Sitek ran the experiment. Both authors read and approved the final manuscript.

Competing interests

The authors declare that they have no competing interests.

Publisher's Note

Received: 21 November 2017 Accepted: 13 April 2018
Published online: 27 April 2018

References

1. Maciá-Fernández, G, Camacho, J, Magán-Carrión, R, García-Teodoro, P, Therón, R (2018). Ugr'16: a new dataset for the evaluation of cyclostationarity-based network IDSs. *Computers & Security*, *73*, 411–424.
2. Beyond point of sale data. Looking forward, not backwards for demand forecasting. http://www.gxs.fr/wp-content/uploads/wp_beyond_point_of_sale_data.pdf. Accessed 20 Feb 2018.
3. Williams, BD, & Waller, MA (2011). Top-down versus bottom-up demand forecasts: the value of shared point-of-sale data in the retail supply chain. *Journal of Business Logistics*, *32*(1), 17–26. https://doi.org/10.1111/j.2158-1592.2011.01002.x.
4. Nachtmann, H, Waller, MA, Rieske, DW (2010). The impact of point-of-sale data inaccuracy and inventory record data errors. *Journal of Business Logistics*, *31*(1), 149–158. https://doi.org/10.1002/j.2158-1592.2010.tb00132.x.
5. Bekarian, N, Payne-Sturges, D, Edmondson, S, Chism, B, Woodruff, TJ (2006). Use of point-of-sale data to track usage patterns of residential pesticides: methodology development. *Environmental Health*, *5*(1), 15. https://doi.org/10.1186/1476-069X-5-15.
6. World Payment Report (2017). https://www.capgemini.com/fr-fr/wp-content/uploads/sites/2/2017/10/world-payments-report-2017_year-end_final_web-002.pdf.
7. Information about payment cards 2nd quarter 2017 (in Polish): National Bank of Poland. http://www.nbp.pl/systemplatniczy/karty/q_02_2017.pdf.
8. EMV card-present transaction percentage. https://www.emvco.com/about/deployment-statistics/. Accessed 20 Feb 2018.
9. EMVCo: EMV specifications. http://www.emvco.com/specifications.aspx. Accessed 20 Feb 2018.
10. Press information about HCE development on the market. http://www.bankier.pl/wiadomosc/Eksperci-Platnosci-HCE-to-rynkowy-przelom-3323308.html. Accessed 20 Feb 2018.
11. Android pay homepage. https://www.android.com/intl/pl_pl/pay/. Accessed 20 Feb 2018.
12. Samsung pay homepage. http://www.samsung.com/us/samsung-pay/. Accessed 20 Feb 2018.
13. Host card emulation. https://en.wikipedia.org/wiki/Host_card_emulation. Accessed 20 Feb 2018.
14. Near field communication. http://nfc-forum.org/what-is-nfc/. Accessed 20 Feb 2018.

15. EMV transaction steps. https://www.level2kernel.com/flow-chart.html. Accessed 20 Feb 2018.
16. What electronic payments reveal about you to lenders. https://www.creditcards.com/credit-card-news/credit-card-purchase-privacy-1282.php. Accessed 20 Feb 2018.
17. Aleskerov, E, Freisleben, B, Rao, B (1997). CARDWATCH: A neural network based database mining system for credit card fraud detection, In *IEEE/IAFE Conference on Computational Intelligence for Financial Engineering, Proceedings (CIFEr)* (pp. 220–226): IEEE.
18. Patidar, R, & Sharma, L (2011). Credit card fraud detection using neural network. *International Journal of Soft Computing and Engineering (IJSCE)*, *1*, 32–38.
19. Bolton, RJ, & Hand, D (2001). Unsupervised profiling methods for fraud detection, In *Conference on Credit Scoring and Credit Control*, *7*. Edinburgh.
20. Tasoulis, DK, Weston, DJ, Adams, NM, Hand, DJ (2008). Mining information from plastic card transaction streams. *Advances in Data Analysis and Classification*, *2*(1), 45-62.
21. Seeja, KR, & Zareapoor, M (2014). Fraudminer: a novel credit card fraud detection model based on frequent itemset mining. *The Scientific World Journal*. https://doi.org/10.1155/2014/252797.
22. Gadi, MFA, Wang, X, do Lago, AP (2008). Credit card fraud detection with artificial immune system. In PJ Bentley, D Lee, S Jung (Eds.), *Artificial Immune Systems: 7th International Conference, ICARIS 2008, Phuket, Thailand, August 10-13, 2008. Proceedings*. https://doi.org/10.1007/978-3-540-85072-4_11 (pp. 119–131). Berlin, Heidelberg: Springer.
23. Ogwueleka, F (2011). Data mining application in credit card fraud detection system. *Journal of Engineering Science and Technology*, *6*, 311–322.
24. Zaslavsky, V, & Strizhak, A (2006). Credit card fraud detection using self-organizing maps. *Information & Security: An International Journal*, *18*, 48–63.
25. Quah, JTS, & Sriganesh, M (2007). Real time credit card fraud detection using computational intelligence, In *2007 International Joint Conference on Neural Networks*. https://doi.org/10.1109/IJCNN.2007.4371071 (pp. 863–868): IEEE.
26. Iyer, D, Mohanpurkar, A, Janardhan, S, Rathod, D, Sardeshmukh, A (2011). Credit card fraud detection using hidden Markov model, In *2011 World Congress on Information and Communication Technologies*. https://doi.org/10.1109/WICT.2011.6141395 (pp. 1062–1066): IEEE.
27. Panigrahi, S, Kundu, A, Sural, S, Majumdar, A (2009). Credit card fraud detection: a fusion approach using dempster–shafer theory and bayesian learning. *Information Fusion*, *10*, 354–363.
28. Sánchez, D, Vila, M, Cerda, L, Serrano, J (2009). Association rules applied to credit card fraud detection. *Expert Systems with Applications*, *36*, 3630–3640.
29. Duman, E, & Ozcelik, MH (2011). Detecting credit card fraud by genetic algorithm and scatter search. *Expert Systems with Applications*, *38*(10), 13057–13063. https://doi.org/10.1016/j.eswa.2011.04.110.
30. Weston, DJ, Hand, D, Adams, N, Whitrow, C, Juszczak, P (2008). Plastic card fraud detection using peer group analysis. *Adv. Data Analysis and Classification*, *2*, 45–62.
31. Juszczak, P, Adams, N, Hand, D, Whitrow, C, Weston, DJ (2008). Off-the-peg and bespoke classifiers for fraud detection. *Computational Statistics & Data Analysis*, *52*, 4521–4532.
32. English Dictionary, definition of context. http://www.dictionary.com/browse/context. Accessed 20 Feb 2018.
33. Gwizdka, J (2000). What's in the context?, In *An extended position paper for CHI 2000 Workshop 11. The what, who, where, when, why and how of context-awareness*. https://doi.org/10.1.1.306.8167. Accessed 20 Feb 2018.
34. Baldauf, M, Dustdar, S, Rosenberg, F (2007). A survey on context-aware systems. *International Journal of Ad Hoc and Ubiquitous Computing*, *2*, 263–277.
35. Munoz, MA, Rodriguez, M, Favela, J, Martínez, A, Gonzalez, V (2003). Context-aware mobile communication in hospitals. *Computer*, *36*, 38–46.
36. Wrona, K, & Gomez, L (2005). Context-aware security and secure context-awareness in ubiquitous computing environments, In *XXI Autumn Meeting of Polish Information Processing Society* (pp. 255–265).
37. Li, W, Joshi, A, Finin, T (2013). Cast: context-aware security and trust framework for mobile ad-hoc networks using policies. *Distributed and Parallel Databases*, *31*(2), 353–376. https://doi.org/10.1007/s10619-012-7113-3.
38. Hulsebosch, RJ, Salden, A, Bargh, M, Ebben, PWG, Reitsma, J (2005). Context sensitive access control, In *Proceedings of ACM Symposium on Access Control Models and Technologies, SACMAT* (pp. 111–119). Stockholm.
39. Covington, MJ, Long, W, Srinivasan, S, Dey, A, Ahamad, M, Abowd, G (2001). Securing context-aware applications using environment roles, In *Proceedings of Sixth ACM Symposium on Access Control Models and Technologies (SACMAT 2001)*. New York: ACM.
40. Zhang, G, & Parashar, M (2004). Context-aware dynamic access control for pervasive computing, In *Proceedings of the Communication Networks and Distributed Systems Modeling and Simulation Conference (CNDS '04)*. San Diego, Calif.
41. Kotulski, Z, Sepczuk, M, Sitek, A, Tunia, MA (2014). Adaptable context management framework for secure network services. *Annales UMCS Informatica*, *14*, 7–30.
42. Sitek, A, & Kotulski, Z (2017). Computer Network Security: 7th International Conference on Mathematical Methods, Models, and Architectures for Computer Network Security, MMM-ACNS 2017, Warsaw, Poland, August 28-30, 2017, Proceedings. In J Rak, J Bay, I Kotenko, L Popyack, V Skormin, K Szczypiorski (Eds.) (pp. 158–170): Springer International Publishing.
43. Fourth report on card fraud. European Central Bank. https://www.ecb.europa.eu/pub/pdf/other/4th_card_fraud_report.en.pdf.
44. Sitek, A (2014). One-time code cardholder verification method in electronic funds transfer transactions, In *Annales UMCS Ser. Informatica vol. 14,2* (pp. 46–59). Lublin: Universitatis Mariae Curie-Sklodowska.
45. Sitek, A, & Kotulski, Z (2015). Contextual management of off-line authorisation in contact EMV transactions. *Telecommun. Rev. Telecommun. News*, *88*((84) 8-9), 953–959.
46. Kotulski, Z, & Szczepinski, W (2010). *Error analysis with application in engineering*. Dordrecht: Springer.
47. Information Supplement: PCI DSS Tokenization Guidelines. Scoping SIG, TokenizationTaskforce PCI Security Standards Council. https://www.pcisecuritystandards.org/documents/Tokenization_Guidelines_Info_Supplement.pdf.
48. Regulation (EU) 2016/679 of the European parliament and of the council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, andrepealingDirective95/46/EC(GeneralDataProtectionRegulation). http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=uriserv:OJ.L_.2016.119.01.0001.01.ENG.
49. Top 10 operational impacts of the GDPR: part 8—pseudonymization. https://iapp.org/news/a/top-10-operational-impacts-of-the-gdpr-part-8-pseudonymization/. Accessed 20 Feb 2018.
50. Louridas, P, & Ebert, C (2013). Embedded analytics and statistics for big data. *IEEE Software*, *30*(6), 33–39. https://doi.org/10.1109/MS.2013.125.
51. Pandas homepage. http://pandas.pydata.org/. Accessed 20 Feb 2018.
52. Numpy homepage. http://www.numpy.org/. Accessed 20 Feb 2018.
53. Matplotlib homepage. https://matplotlib.org/. Accessed 20 Feb 2018.
54. Pérez, F, & Granger, BE (2007). IPython: a system for interactive scientific computing. *Computing in Science and Engineering*, *9*(3), 21–29. https://doi.org/10.1109/MCSE.2007.53.
55. Jupyter IDE homepage. http://jupyter.org/. Accessed 20 Feb 2018.