An Observe-and-Detect methodology for the security and functional testing of smart card applications
Germain Jolly, Sylvain Vernois, Christophe Rosenberger

To cite this version:
Germain Jolly, Sylvain Vernois, Christophe Rosenberger. An Observe-and-Detect methodology for the security and functional testing of smart card applications. International Conference on Information Systems Security and Privacy (ICISSP), Feb 2016, Rome, Italy. hal-01286831

HAL Id: hal-01286831
https://hal.archives-ouvertes.fr/hal-01286831
Submitted on 20 Mar 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
An Observe-and-Detect methodology
for the security and functional testing of smart card applications

Germain Jolly¹, Sylvain Vernois¹ and Christophe Rosenberger¹

¹Université de Caen Basse Normandie; ENSICAEN; UMR 6072 GREYC, Caen, France
{germain.jolly, sylvain.vernois, christophe.rosenberger}@ensicaen.fr

Keywords: Security, Analysis, Smart Card application, Observation, Detection, Evaluation, WSCT Framework

Abstract: Smart cards are tamper resistant devices but vulnerabilities are sometimes discovered. We address in this paper the security and the functional testing of embedded applications in smart cards. We propose an original methodology for the evaluation of applications and we show its benefit by comparing it to a classical certification process. The proposed method is based on the observation of the APDU (Application Protocol Data unit) communication with the smart card. Some specific properties are verified as a complementary method in the evaluation process and allows the on-the-fly detection of an anomaly and the reasons that triggered this anomaly during the test. Here are presented two uses of this method: a simple use to illustrate the use of properties to verify an implementation of an application and a more complex illustration by applying the fuzzing method to show what we can obtain with the proposed approach, i.e. an analysis of an anomaly.

1 INTRODUCTION

While smart cards have invaded our daily lives, the question of the quality of the card application arises. EMVCo reports that 1.62 billion payment cards and 23.8 million terminals in circulation globally are based on EMV technology (EMVCo, 2013). The development of an application on a smart card is not an easy task (Rankl, 2007). Not clear enough specifications or having a too large number of important variables make the development of smart card implementation difficult. Testing aims at discovering anomalies in actual behavior of a system compared to what is expected. Moreover, it is not possible to test an application in a smart card over the full range of input variables. The full testing would cost too much in time and significant resources.

For functional and security testing, it is necessary to validate each step and each element of a smart card. It is unrealistic to use a secure and certified component with a poorly implemented program. The program must also be validated and certified. The definition of the criteria for evaluating security allowed product certification (Wallace et al., 1996). Recognition of these criteria allows mutual recognition between different structures and different countries. A certified product is tested not only by certification laboratories but also by the manufacturer during its development. The certification allows recognizing the level of security of a product by an external laboratory. EMVCo is providing the EMVCo Security Evaluation Process to its members (Alliance, 2011). It ensures a robust security foundation for smart card and related products. The Card Type Approval process allows testing compliance with the EMV specifications. Another document called Card Type Approval process provides an increased level of confidence. To summarize, EMVCo provides documents to test, evaluate and approve a smart card but also a list of Approved Security Evaluation Laboratories (EMVCo, 2012). For validation purposes, meaningful I/O sequences are generated.

While in the industrial field, evaluation is especially enabled by testing tools, academic research has proposed many different methods. There are several kinds of verification and validation (V&V) methods: control (technical notice, reviews), analysis (mathematical verification) or test (focus on input/output). Most of the academic research works consider white-box systems (known source code and secrets).

- Static analysis is an analysis method ((Distefano and Parkinson J, 2008), (Philippaerts et al., 2014) or (Ahrendt et al., 2005)) allowing to ensure that defined coding practices are being followed.
- The model checking is a method consisting of verifying if a model of an abstraction or a system follows the specifications ((Sabatier and Lartigue,
The model checking is used to check all possible paths of the state machine defining an application on smart cards.

- The test method allows to discover errors at all levels of systems ((Philipps et al., 2003), (van Weelden et al., 2005)). We can also discern the structural test from the functional test. This method is the most used method in the industrial field. The main problem of the test is to build data input that will provide the higher test coverage.

- Fuzzing is an automatic generation of a large number of test cases ((Lancia, 2011), (Bekrar et al., 2012) or (Alimi et al., 2014). The test is used to verify one specific path in the state machine of the application. This approach allows an automatic generation of test cases. The problem is that this method costs smart cards (as they are not scheduled to accept so many transactions). The main drawback of this approach is also the difficulty to fully understand the scenario that permits to generate an attack.

The presented method here allows both verification of the complete application on a smart card, or at least provides an accurate documentation of a detected anomaly. The method detects an anomaly relative to the expected behavior and provides accurate documentation of the reasons that triggered this error. The proposed method detects vulnerabilities through the observation of the APDU communications. It improves the documentation comparing to the results of known evaluation methods as the test method used to verify, validate and evaluate a smart card. We would like to mix the efficiency of the whitebox methods with the simplicity of the blackbox ones.

In the second section, we provide some necessary background on smart cards and their certifications. The methodology is defined in the third section. Section 4 deals with the results obtained by creating a proof of concept using the WSCT framework for the payment application and the second use of the method mixed with a fuzzing method.

2 BACKGROUND

We focus on the smart card application. For the payment, EMVCo provides EMV (Europay MasterCard Visa) specifications and the CPA (Common Payment Application) specification (Watanabe et al., 2006). Moreover, proprietary specification is added to develop a smart card application.

In this work, the MasterCard M/CHIP specification, an EMV-compliant specification for Mastercard smart cards, is studied. The application contained on the chip can take several states (selected performed, GPO performed, ...). The evolution of the state of the application is allowed by sending commands APDU (select, get data, GPO, ...) and receiving responses APDU (9000, 6283, ...). A machine state is given by Mastercard to illustrate the M/CHIP application in its proprietary specification.

The payment application can pass from a state to another only by receiving and responding to a series of command/response pairs, this is the concept of application state. We cannot only consider the acceptance of a command and its response without considering its current state. A terminal communicates with the smart card by sending commands and receiving responses. In the figure 1, we can see the structure of APDU commands and APDU responses according to ISO/IEC 7816. CLA indicates the type of a command, INS a specific command, P1 and P2 are parameters for a command, LC indicates the length of the UDC which is optional data and LE indicates the length of the expected data. Expected data, only contained in the response if LE is in the command, are the UDR. Finally, SW1 and SW2 are mandatory in the response and are the command processing status.

During the life of a product, we have to ask some questions (Radatz et al., 1990):

- **Verification**: Are we building the product right?
- **Validation**: Are we building the right product?
- **Evaluation**: How can we interpret the verification and validation?
- **Certification**: Are the evaluation results approved by a specific firm?

To evaluate and validate the product, several tests are made during the different steps of the life of the smart card (Rankl and Effing, 2010). All aspects of the card (application, chip and personalization data) must be validated to ensure a secure product. Once designed and validated, the card is certified. The certification allows recognizing the level of security of a product by an external laboratory.

According to ISO 8402 (for Standardization, 1994), quality is defined as the totality of features and characteristics of a product or service that bear on
its ability to satisfy the stated or implied needs. The two most important attributes of quality are robustness and freedom from defects and errors. Concerning this field, the test is mostly used for the evaluation of a smart card and the evaluation of a software is called software evaluation. The purpose of testing is always to find defects and errors, and the entire testing process must be oriented towards this purpose. Tests can never be complete, since it is never possible to fully work through all possible permutations. We obtain a desired level of quality. As tests can never fully cover everything, it is difficult to define a completion criterion for test development. The rule of thumb is that the same amount of effort should be spent on test development as on product development (i.e. try to achieve full requirements coverage with the tests). We obtain a desired level of quality. As tests can never fully cover everything, it is difficult to define a completion criterion for test development. The rule of thumb is that the same amount of effort should be spent on test development as on product development (i.e. try to achieve full requirements coverage with the tests). We obtain a desired level of quality. As tests can never fully cover everything, it is difficult to define a completion criterion for test development. The rule of thumb is that the same amount of effort should be spent on test development as on product development (i.e. try to achieve full requirements coverage with the tests). We obtain a desired level of quality. As tests can never fully cover everything, it is difficult to define a completion criterion for test development. The rule of thumb is that the same amount of effort should be spent on test development as on product development (i.e. try to achieve full requirements coverage with the tests). We obtain a desired level of quality. As tests can never fully cover everything, it is difficult to define a completion criterion for test development. The rule of thumb is that the same amount of effort should be spent on test development as on product development (i.e. try to achieve full requirements coverage with the tests). We obtain a desired level of quality. As tests can never fully cover everything, it is difficult to define a completion criterion for test development. The rule of thumb is that the same amount of effort should be spent on test development as on product development (i.e. try to achieve full requirements coverage with the tests). We obtain a desired level of quality. As tests can never fully cover everything, it is difficult to define a completion criterion for test development. The rule of thumb is that the same amount of effort should be spent on test development as on product development (i.e. try to achieve full requirements coverage with the tests). We obtain a desired level of quality. As tests can never fully cover everything, it is difficult to define a completion criterion for test development. The rule of thumb is that the same amount of effort should be spent on test development as on product development (i.e. try to achieve full requirements coverage with the tests). We obtain a desired level of quality. As tests can never fully cover everything, it is difficult to define a completion criterion for test development. The rule of thumb is that the same amount of effort should be spent on test development as on product development (i.e. try to achieve full requirements coverage with the tests). We obtain a desired level of quality. As tests can never fully cover everything, it is difficult to define a completion criterion for test development. The rule of thumb is that the same amount of effort should be spent on test development as on product development (i.e. try to achieve full requirements coverage with the tests). We obtain a desired level of quality. As tests can never fully cover everything, it is difficult to define a completion criterion for test development. The rule of thumb is that the same amount of effort should be spent on test development as on product development (i.e. try to achieve full requirements coverage with the tests). We obtain a desired level of quality. As tests can never fully cover everything, it is difficult to define a completion criterion for test development. The rule of thumb is that the same amount of effort should be spent on test development as on product development (i.e. try to achieve full requirements coverage with the tests). We obtain a desired level of quality. As tests can never fully cover everything, it is difficult to define a completion criterion for test development. The rule of thumb is that the same amount of effort should be spent on test development as on product development (i.e. try to achieve full requirements coverage with the tests). We obtain a desired level of quality. As tests can never fully cover everything, it is difficult to define a completion criterion for test development. The rule of thumb is that the same amount of effort should be spent on test development as on product development (i.e. try to achieve full requirements coverage with the tests). We obtain a desired level of quality. As tests can never fully cover everything, it is difficult to define a completion criterion for test development. The rule of thumb is that the same amount of effort should be spent on test development as on product development (i.e. try to achieve full requirements coverage with the tests). We obtain a desired level of quality. As tests can never fully cover everything, it is difficult to define a completion criterion for test development. The rule of thumb is that the same amount of effort should be spent on test development as on product development (i.e. try to achieve full requirements coverage with the tests). We obtain a desired level of quality. As tests can never fully cover everything, it is difficult to define a completion criterion for test development. The rule of thumb is that the same amount of effort should be spent on test development as on product development (i.e. try to achieve full requirements coverage with the tests). We obtain a desired level of quality. As tests can never fully cover everything, it is difficult to define a completion criterion for test development. The rule of thumb is that the same amount of effort should be spent on test development as on product development (i.e. try to achieve full requirements coverage with the tests). We obtain a desired level of quality. As tests can never fully cover everything, it is difficult to define a completion criterion for test development. The rule of thumb is that the same amount of effort should be spent on test development as on product development (i.e. try to achieve full requirements coverage with the tests). We obtain a desired level of quality. As tests can never fully cover everything, it is difficult to define a completion criterion for test development. The rule of thumb is that the same amount of effort should be spent on test development as on product development (i.e. try to achieve full requirements coverage with the tests). We obtain a desired level of quality. As tests can never fully cover everything, it is difficult to define a completion criterion for test development. The rule of thumb is that the same amount of effort should be spent on test development as on product development (i.e. try to achieve full requirements coverage with the tests). We obtain a desired level of quality. As tests can never fully cover everything, it is difficult to define a completion criterion for test development. The rule of thumb is that the same amount of effort should be spent on test development as on product development (i.e. try to achieve full requirements coverage with the tests). We obtain a desired level of quality. As tests can never fully cover everything, it is difficult to define a completion criterion for test development. The rule of thumb is that the same amount of effort should be spent on test development as on product development (i.e. try to achieve full requirements coverage with the tests). We obtain a desired level of quality. As tests can never fully cover everything, it is difficult to define a completion criterion for test development. The rule of thumb is that the same amount of effort should be spent on test development as on product development (i.e. try to achieve full requirements coverage with the tests). We obtain a desired level of quality. As tests can never fully cover everything, it is difficult to define a completion criterion for test development. The rule of thumb is that the same amount of effort should be spent on test development as on product development (i.e. try to achieve full requirements coverage with the tests). We obtain a desired level of quality. As tests can never fully cover everything, it is difficult to define a completion criterion for test development. The rule of thumb is that the same amount of effort should be spent on test development as on product development (i.e. try to achieve full requirements coverage with the tests). We obtain a desired level of quality. As tests can never fully cover everything, it is difficult to define a completion criterion for test development. The rule of thumb is that the same amount of effort should be spent on test development as on product development (i.e. try to achieve full requirements coverage with the tests).

3 PROPOSED EVALUATION METHODOLOGY

3.1 Objectives

From a certified card, we get a collection of properties, local and required behaviors (first, this collection is created manually but we are working on this topic to automatize the generation). These properties are used by the Observer to determine if a card application is correct. In order to reach this goal, a modified smart card, containing a modified smart card application of the certified one, is used with a terminal. In our case, it is a terminal doing a nominal EMV transaction. The Observer therefore takes as input the communication between the tested card and terminal and the collection of properties to check. Note that the verification of the state machine is possible. During the communication, the Observer recovers transmitted data and detects an anomaly on the fly. Finally, a report is created.

3.2 Property Language

To detect abnormal behaviors and give some documentation about the reasons that triggered those behaviors, we define the properties that correspond to required and local behaviors of an application. The section defines the property language.

3.2.1 Basis

Unlike formal methods such as model checking or proof of theorem approaches, we are checking properties while observing the transmitted communication during the transaction. In fact, in studies like (Posegga and Vogt, 1998) (Sabatier and Lartigue, 1999) (Lanet and Requet, 2000) (Jacobs et al., 2004) or (Haneberg et al., 2007), the verification needs a formal model or the access to the source code. With our language, we can define the required behavior (local and global) using only the transmitted data.

3.2.2 Definitions

The language allows us to define both a state machine, allowing to follow the actual behavior of the application and also to define the properties as defined in (Jolly et al., 2014). We can follow the global behavior of the smart card but also detect local changes in the application behavior.

To define this behavior, we have two types of elements. First, we have the element representing the APDU communication. The APDU data is defined by:

- **Instruction**: represents the CLA and INS bytes of a command APDU
- **Parameters**: represents the P1 and P2 bytes of a command APDU
- **Status**: represents the SW1 and SW2 bytes of a response APDU

The second type is the logical relation between two or more elements. The logical relation between APDU are defined by:

- **And**: elements are all true
- **Or**: one element is true
- **Nor**: elements are all false

In addition to the previous elements, defining a machine state, i.e. a global behavior of a smart card application, requires these tags:

- **cardstate**: the name of the state of the application used to follow the actual behavior of the application.
  - name: name of the cardstate
  - default: the initial state of the machine state
- **transition**: allows modifying the actual state of the smart card application
  - name: name of the transition
  - from: previous cardstate
  - to: next cardstate
• **from** / **to**: to construct the transitions between cardstates
  – **nom**: name of the beginning cardstate

For the definition of property, we need the following elements to define properties:

• **property**: required and local behavior of the smart card
  – **name**: name of the property
  – **explanation**: description of the property

• **step**: indicates the order of elements

For each of these elements, we can work on transmitted data (i.e. UDC or UDR data). We added tags to recover the plain transmitted data. If we recover TLV data, we should be able to work with only one or several specific values.

• **cdata**: allows to recover the plain udc data

• **rdata**: allows to recover the plain udr data

• **Tag**: allows to verify the value of a particular TLV data, contained in a response APDU
  – **name**: name of the tag (example: 0x9F27)
  – **value**: value of the TLV data (example: 0x40)

• **mask**: allows to apply a mask on the values

• **call**: allows to call a function to manage the data for a particular property (example: checking of the validity of a specific data). This last element permits to improve the verification by adding some functionality to the properties and particularly on the studied data that can be encrypted and can be added to every step of a property.

### 3.3 Development tools

Many tools exist to help verifying and validating smart cards. Here are several independent or academic tools able to work with smart cards:

- **SmartCard Framework**, an independent tool to easily develop a smart card application (Rouit, 2011)
- **OpenSCDP**, a collection of tools for the development, test and deployment of smart card (CardContact, 2012)
- **Javaemvreader**, a tool to communicate with, and read data from an EMV smart card (Sasc, 2014)
- **PCSC_sharp**, an independent tool in order to work with smart cards (Mueller, 2012)
- **CardPeek**, an independent tool to read the contents of smart cards (Pannetrat, 2010)

- **OPAL**, a library for developing tools for Smart Card research and education (Bkakria et al., 2011)
- **WSCT**, a framework able to easily explore and work with smart cards (Vernois and Alimi, 2010)


<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>SmartCard Framework</td>
<td>*</td>
<td>**</td>
<td>*</td>
</tr>
<tr>
<td>OpenSCDP</td>
<td>**</td>
<td>***</td>
<td>**</td>
</tr>
<tr>
<td>Javaemvreader</td>
<td>**</td>
<td>**</td>
<td>*</td>
</tr>
<tr>
<td>PCSC_Sharp</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>OPAL</td>
<td>**</td>
<td>**</td>
<td>***</td>
</tr>
<tr>
<td>CardPeek</td>
<td>*</td>
<td>**</td>
<td>***</td>
</tr>
<tr>
<td>WSCT</td>
<td>***</td>
<td>**</td>
<td>***</td>
</tr>
</tbody>
</table>

* Functionalities contain observation, extension and replay mechanisms. The more there are stars like *, the more the tool is interesting for this feature for the Functionnalities, Documentation and Maintenance criteria.

### 3.4 Discussion

**WSCT** (Vernois and Alimi, 2010) is a framework developed by Sylvain Vernois et al. at Greyc for several years. It is written in C# and allows to work with smart cards, e.g. for exploration and finding fault on smart cards. The two main purposes of this tool are to provide:

- an **object-oriented API** to access a smart card reader.
- an **evolutive GUI** (with creation of plugins) to manipulate different types of smart cards.

Using this framework, we have easily created two plugins. Indeed, WSCT framework has an easy extension mechanism. We choose to create separate plugins to emphasize that the Observer platform is not linked to the terminal. The Observer can be used on every kind of terminal (nominal or test transaction).

We used this tool to develop the proposed property language in order to observe communication, thanks to the possibilities of the framework, and detect anomalies, thank our language.

### 4 ILLUSTRATION

Here, we are showing two studies done with this observation tool. The first deals with EMV transaction, i.e. we are verifying an existing implementation of the EMV transaction. The second is about the use of the properties to analyze Fuzzing results on a PME
implementation, an e-wallet application. We are focusing on the smart card application using a correct and non-malicious smart card reader. For each study, the reference to see the divergence with the correct behavior is the collection of properties (a set of assertions based on the transmitted data between the certified smart card and the terminal) that can be partial or complete. In our study, we use specific, local and required behaviors in order to check a specific part of the application.

4. Observation of the EMV transaction

4.1 Observation of the EMV transaction

4.1.1 View of the platform

Two plugins are needed for our method: one to communicate with the smart card and one to observe the communication and do the properties verification. We add an automat to follow the behavior of this smart card. We decided to add the machine state of the EMV application based on (Aarts et al., 2013). The inputs of the Observer are communication APDU and the selected properties to verify.

4.1.2 The used protocol

The protocol we used is based on three main points:

- the smart card application and its documentation. Using the documentation, we manually model the local behaviour. Also, it should be mentioned sooner that the properties are encoded with XML. We used our own EMV smart card application certified by a laboratory.

- the application to do the transaction because this method uses APDU communication between a terminal application and the smart card application. We used the WSCT framework to create an EMV explorer able to do the EMV transaction with the smart card application. (It can be simplified by the use of an extern terminal observed by the WSCT functionalities).

- the Observer, an application able to recover APDU communication and detect a bad behavior of the smart card application. It simply compares the transmitted APDU communication with the theoretical behaviors defined in the xml file.

4.1.3 Observation of smart card application

To define the machine state, we need two elements: "cardstate" and "transition". The listing 1 shows a partial machine state of the payment application. We call "selected" the state when the application is select (and so select performed) and "initiated" when the GPO command (initiation part of the transaction) is performed.

Listing 1: Partial machine state of the payment application

```
<machine>
  <cardstate name="Selected"/>
  <cardstate name="Initiated"/>
  <transition text="GPO">
    <from name="selected" dir="Down"/>
    <to name="initiated" dir="Up"/>
    <and>
      <instruction="0x80A8"/>
      <status="0x9000"/>
    </and>
    <transition text="else">
      <from name="initiated" dir="Right"/>
      <to name="selected" dir="Right"/>
    </transition>
  </transition>
</machine>
```

Figure 2 presents an illustration of the machine state observation of the payment application, which is implemented in the Observer plugin.

4.1.4 Example of property detection

To illustrate our study, we take the example from (Jolly et al., 2014). The property P3 represents the behavior between the "selected" and "initiated" states using only one command. The property P3 is defined on three clock cycles, i.e. three pairs APDU command/response. If one wants to check the value of a TLV data, one could have added a TAG verification as previously defined. Here, we checked only the success or the failure of the GPO (Get Processing Options) command between two particular states of
the application. The xml file in the listing 2 is the P3 property usable by the Observer.

Good behavior with three GPO commands:

\[ P_3 : (SW1(1) = 90) \text{ and } (SW2(1) = 00) \text{ and } (SW1(3) \neq 90) \text{ and } (SW2(3) \neq 00) \text{ and } (SW1(5) = 90) \text{ and } (SW2(5) = 00) \text{ or } (INS(0) = A8) \text{ and } (INS(2) = A8) \text{ and } (INS(4) = A8) \]

Listing 2: P3 property

```xml
<property name="Third GPO">
  <step>
    <instruction>0x80A8</instruction>
    <status>0x9000</status>
  </step>
  <step>
    <instruction>0x80A8</instruction>
    <nor>
      <status>9000</status>
    </nor>
  </step>
  <step>
    <instruction>0x80A8</instruction>
  </step>
  <step>
    <instruction>0x80A8</instruction>
    <status>0x9000</status>
  </step>
</property>
```

To understand the scope of the detection properties method, here is a comparison of three possible behaviors of the application:

- **a)** The first case is the result of a successful and ordered test from the selection of the application to the use of Generate AC command. The tester knows the expected behavior of the application comparing to his test. The application responds appropriately to the known configuration of the test. The property is not violated.

- **b)** The second case presents an unusual refusal Generate AC command but also detection of the violation of the property P3, which explains the reason for the anomaly detected by the test. The property violation detection gives an improved documentation about the result of the test.

- **c)** The third case is correct behavior relative to the test result. But the property has been violated. In fact, the application contains an error but it’s not critical comparing to the behavior of the application with the computed test.

### 4.2 Fuzzing analysis of a PME implementation

The same observation method can be used during a fuzzing session to complete the analysis of an implementation. The PME application is an electronic purse. It’s a much more simple application than the EMV one.

#### 4.2.1 Basis of the study

A reference implementation, an own implementation and a set of scripts (one script is a possible transaction) are inputs of the evaluation system. (Items such an XML file representing the application and an XML file containing properties are generated from the reference implementation or associated documentation). The system sends the scripts to both applications and detects a difference between the implementation, i.e. an anomaly is detected. To analyze the reasons for this error, the system uses the violated properties observed by the tool.

The terminal sends the N transaction scripts to the two smart cards. if it detects a difference between our implementation and the reference one, an anomaly is detected. The Analysis tool is composed of three main modules:

- **A)** The Observation Module recovers APDU communication (we used a representation of the evolution of the application state as in figure 3 but it’s informative information).

- **B)** The Detection Module detects a difference between two implementations of an application (the reference implementation and the tested one).

- **C)** The Analysis Module can give more information about a detected anomaly.

#### 4.2.2 Experimental results

To study the addition of the method on an evaluation method such as fuzzing, we have used electronic wallet application called PME application. In order to compute our tests, we used:

- 4 implementations: one is the reference implementation and the others contains errors (defective verification of balance during a debit operation, defective check of pin code and unmanaged blocking of the card when using false pin code)

- 2 scripts of transaction: these scripts allow to compute several different transactions in order to detect the presence of an anomaly (by checking the difference between the responses of the reference implementation and the responses of the tested implementation)

- 5 properties: we use a limited number of properties because we are showing the feasibility of the method on our own smart cards. Each property is a local and required behavior of the smart card application (given by the reference implementation which must be validated).
We may also use the documentation of an application to create these elements. Having a reference application is simpler for our study.

4.3 Explanation of an example

This example explains an anomaly detection: the defective blocking of the card if the user is testing too many pin code. Without blocking the application, the so-called brute force attack would be possible. The script therefore sends over n times a wrong pin code (n is the maximum number of false pin code defined by the application documentation). When the pin code is accepted by the tested smart card but not by the reference smart card, an anomaly is detected. We just need to look the observer to see the violated property to understand the reason of the anomaly. We have, in this example, three properties in the listing 3.

- The first checks that the blocking displayed by the error code is effective, i.e. it is not just a superficial blockage.
- The second ensures that the application sends an error message, equal to 9170, once the counter is zero, i.e. the smart card has processed the information correctly.
- The latest involves a control method. We observe two failed uses of the verify command. In this case, the difference between the two error codes is 1, i.e. the counter is decremented correctly.

Listing 3: Example of explanation by properties

```
<properties>
  <property name="blockageOK">
    <step>
      <instruction instruction="0x9020" />
      <status status="0x9170" />
    </step>
  </property>
  <step>
    <instruction instruction="0x9020" />
    <status status="0x9170" />
  </step>
  <require>
    <call method="control" />
  </require>
</properties>
```

4.4 Advantages and Improvements

The use of properties allows verifying an implementation as an efficient and stand alone evaluation method. The efficiency of this method is due to the efficiency of the description of the application, i.e. a full collection of properties must be defined. We need also have a full documentation of the application or a certified implementation of the application. The method can be used with another evaluation method as fuzzing in order to give more documentation about an error detector by the fuzzing report.

5 CONCLUSIONS

This paper has defined a language property related to the verification approach exposed in this paper. The two main principles of this method are: the communication observation and the property violation detection. Using this language, we have defined the machine state corresponding to our studied smart card application to observe the global behavior. And more interesting, we have defined properties to detect anomalies and the reasons that triggered those anomalies. We have created the observe-to-detect tool using the WSCT framework. Finally, the violation detection allows us to give more documentations about a detected error or an undetected anomaly by a
test plan. We have been able to validate our method with the creation of our own smart card application and our own intern terminal. Moreover, the use of this method mixed with a known evaluation method as the fuzzing shows complementary information about an anomaly. It is easier to explain fuzzing results, i.e. find the reasons of a detected error on a smart card application as the PME application.

The perspectives of this work is to automatize the creation of properties for a specific application as an EMV payment application. Indeed, creating properties manually allows us to validate our method with suitable properties. With an automatic generation, we will be able to have a complete and user configurable collection.

REFERENCES


