

Using Model Checking to Generate Test Cases for Android Applications

Ana Rosario Espada María del Mar Gallardo
Alberto Salmerón Pedro Merino

Dept. Lenguajes y Ciencias de la Computación
E.T.S.I. Informática University of Málaga*

[anarosario,gallardo,salmeron,pedro]@lcc.uma.es

The behavior of mobile devices is highly non deterministic and barely predictable due to the interaction of the user with its applications. In consequence, analyzing the correctness of applications running on a smartphone involves dealing with the complexity of its environment. In this paper, we propose the use of *model-based testing* to describe the potential behaviors of users interacting with mobile applications. These behaviors are modeled by composing specially-designed state machines. These composed state machines can be exhaustively explored using a model checking tool to automatically generate all possible user interactions. Each generated trace model checker can be interpreted as a test case to drive a runtime analysis of actual applications. We have implemented a tool that follows the proposed methodology to analyze ANDROID devices using the model checker SPIN as the exhaustive generator of test cases.

1 Introduction

At present, smartphone technology is ubiquitous and changes constantly. Users use their mobiles not only as phones, but as compact computers, able to concurrently provide services which are rapidly created, updated, renewed and distributed. In this scenario of continuous evolution, different operating systems have been developed such as SYMBIAN, IOS, WINDOWS PHONE and ANDROID, which allow phones to support more and more complex applications. These platforms define new models of execution, quite different from those used by non-mobile devices. For instance, one of the most defining characteristics of these systems is their open and event-driven nature. Mobile devices execute a continuous cycle that consists of first, waiting for the user input and second, producing a response according to that input. In addition, the internal structure of mobile systems is constructed from a complex combination of applications, which enable users to easily navigate through them. Thus, although, at a lower level, the execution of applications on a mobile device involves the concurrent execution of several processes (for instance, in ANDROID, applications are JAVA processes executing on the underlying LINUX operating system), the way these applications interact with each other and with the environment does not correspond with the standard interleaving based concurrency model.

It is clear that the execution of applications on these new operating systems, such as ANDROID [1], may lead to the appearance of undesirable bugs which may cause the phone to malfunction. For example, mobile devices may display the typical errors of concurrent systems such as violations of safety and liveness properties. However, there are other bugs inherent to the particular concurrency model supported by the new platforms which are not directly analyzable using current verification technologies. For

*Work partially supported by grants P11-TIC-07659 (Regional Government of Andalusia), TIN2012-35669 (Spanish Ministry of Economy and Competitiveness), UMA-Agilent Technologies 808/47.3868- 4Green and the AUIP as sponsor of the Scholarship Program Academic Mobility.

example, applications could incorrectly implement the life cycles of their activities or services (in the case of ANDROID), or may misbehave upon the arrival of unexpected external events. In addition, conversion errors, unhandled exceptions, errors of incompatibility API and I/O interaction errors as described in [16] may also appear.

Different techniques for analyzing the execution of mobile platforms have been proposed. Verification approaches such as model checking [9] can be applied to the software for mobile devices [22, 21, 19]. Model checking is based on a exhaustive generation of all the inter leavings for the threads/processes. A major problem to apply this technique to the real code, like mobile applications, is the need to construct a model of the underlying operating system or libraries [10, 7, 11]. The open nature of these platforms, which are continuously interacting with an unspecified environment, makes other analysis techniques such as *testing*, *monitoring*, and *runtime verification* more suitable to check bugs without too much extra effort to model the operating system or the libraries. There have been several recent proposals [12, 24, 18] for testing in this framework with commercial tools [4, 2]. In these approaches, test cases are randomly generated with tools such as MONKEY and MONKEYRUNNER [1].

Testing and runtime verification maybe also combined, as described in [6], to construct verification tools for mobile applications [20, 26]. On the one hand, the careful selection of test cases guides the execution of the device, while, on the other, the runtime verification module implements observers devoted to analyzing the traces produced by the device. The runtime verification module was already addressed by some authors of this paper in [14]. Here we focus on describing how the generation of test cases may be carried out following the *model-based testing* approach [25] supported by model checking tools.

Our proposal is based on the idea that although the interaction between the user and the mobile device is completely undetermined, each application is associated with a set of *intended user behaviors* which define the *common ways* of using the application. For each application, or more precisely, for each application view, we use state machines to construct a *non deterministic model* representing the expected use of the view/application. This state machine is built semiautomatically, with information provided by the expert (the app designer or tester) and by ANDROID supporting tools like UIAUTOMATORVIEWER. Then, all these view models may be conveniently composed to construct a *non deterministic model* of the user interaction with a subset of mobile applications of interest. Due to the way of building the state machines, each execution trace of the *composed state machine* corresponds to a possible realistic use of the mobile. Thus, the generation of test cases is reduced to the generation of all possible behaviors of the composed machine, which may be carried out by a model checking tool. Although the methodology proposed does not depend on the underlying mobile operating system, the tool has been built on the assumption that the operating system is ANDROID.

The paper provides two main contributions. The first one is the formal definition of a special type of state machine that models the expected user interaction with the mobile application. The approach to modeling is completely modular in the sense that adding (or removing) new view state machines to incorporate (eliminate) user behaviors does not affect the rest of state machines that have already been defined. The second one is a method to employ the explicit model checker SPIN [15] that takes the composed state machine as input and produces a significant set of test cases that generate traces for runtime verification tools. We have constructed a tool chain which implements both modeling and test generation phases to show the feasibility of the approach in practice.

The rest of the paper is organized as follows. Section 2 describes our approach to using model checking for test case generation. Section 3 introduces the testing platform that we are developing. Section 4 provides a formal description of the behaviour of composed state machines. Section 5 uses well known ANDROID applications to describe how our approach for test case generation is implemented. Section 6 gives a comparison with related work. Last section summarizes conclusions and future work.

```

1 mtype = { state_init, state_1, state_2 };
2 typedef Device { byte transitions[MAX_TR]; short index; bool finish; }
3 Device devices[DEVICES];
4 mtype state[DEVICES];
5 active proctype traceCloser() provided (devices[DEVA].finish && devices[DEVB].finish) {
6   end_tc: outputTransitions()
7 }
8 active proctype device_A() {
9   state[DEVA] = state_init;
10  do
11    :: state[DEVA] == state_init -> transition(DEVA, BUTTON_1); state[DEVA] = state_1
12    :: state[DEVA] == state_1 -> transition(DEVA, SWIPE); state[DEVA] = state_1
13    :: state[DEVA] == state_1 -> transition(DEVA, BUTTON_2); state[DEVA] = state_2
14    :: state[DEVA] == state_2 -> transition(DEVA, MESSAGE); break
15    :: state[DEVA] == state_2 -> transition(DEVA, BACK); break
16  od;
17  devices[DEVA].finish = true;
18 }
19 active proctype device_B() {
20   state[DEVB] = state_init;
21   ...
22   devices[DEVB].finish = true;
23 }

```

Listing 1: Sample PROMELA specification for test generation

2 Model checking for test case generation

SPIN [15] is a model checker that can be used to verify the correctness of concurrent software systems modeled using the specification language PROMELA. The focus of the tool is on the design and validation of computer protocols, although it has been applied to other areas. SPIN can check the occurrence of a property over all possible executions of a system specification, and provide counterexamples when violations are found.

We use the SPIN model checker in our approach for automatically generating test cases from application models in the following way. First, each device will be represented by a single PROMELA process, which contains a state machine that models the applications contained on that device. The state machines themselves are written as loops, where each branch corresponds to a transition triggered by an event. The current state of each state machine (stored as a global PROMELA variable) determines which branches are active and may be taken. The right hand side of each branch records the transition and updates the current state. This PROMELA specification is explored exhaustively by SPIN in order to generate all possible test cases described by the application model, taking all possible alternatives when there is more than one active branch at the same time.

Listing 1 shows an example of a PROMELA specification that follows the approach outlined above. This example contains two devices and with their corresponding state machine (`device_A()` in line 8 and `device_B()` in line 19), with two states plus the initial state. The transition function is used to record the user or system transition associated with each branch. In order to complete a test case, all devices must have finished their respective state machines (lines 17 and 22), usually when the do loop is exited (lines 14 and 15). This enables the `traceCloser` process to be executed due to its schedulability restrictions (line 5), which prints the transitions of the generated test case.

In addition to the current state (line 4), this PROMELA specification also keeps a list of the transitions taken on the test currently being generated (line 2). The purpose of this data structure is twofold. On the one hand, `outputTransitions` will print the trace stored here. On the other hand, the history of the

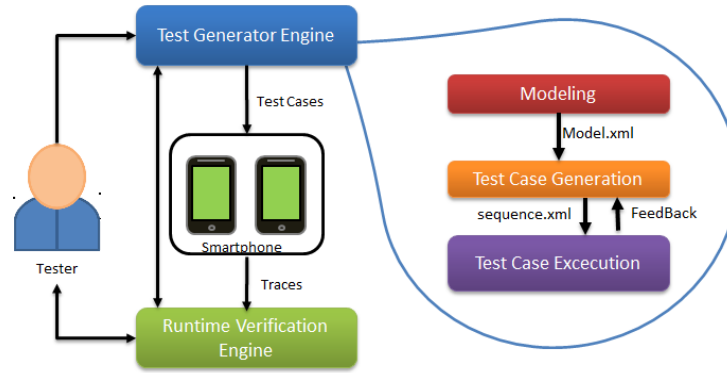


Figure 1: Architecture

current trace is kept inside the SPIN's global state, which is taken into consideration when deciding if a state has already been visited. Thus, the same transition may be taken more than once if possible (e.g. line 12), since the history of the states will be different. However, this requires the maximum depth of exploration to be bounded by the `MAX_TR` constant (line 2).

3 Architecture of the platform

Figure 1 shows the general structure of tools that combine testing and runtime verification techniques to analyze the behavior of applications running on mobile devices. The bottom side uses observers/monitors to analyze the resulting execution traces and verify whether they comply with the expected properties as implemented in the tool DRAGONFLY [14, 13]. The top side of the figure shows the generation of test cases considered in this paper. The Tester is the expert responsible for modeling the behavior of the applications to be analyzed using a state chart diagram. These models may be constructed as part of the design phase of the applications, and are characterized by their compositional nature: functionality can be added to an existing view without essentially altering the existing behavior.

Figure 2 shows the complete process of our actual proposal for test generation and execution, which is divided into three main modules:

- *Modeling.* UIAUTOMATORVIEWER tool from ANDROID tools extracts the controls definition in each view of the ANDROID application under analysis. Then, the controls definition and the state chart diagrams are associated into a `Model.xml` file with a given structure.
- *Test Case Generation.* Creates a test case generator per XML file model into a PROMELA file. The SPIN model checker [15] performs an exhaustive search of all valid paths in the model using the method explained in Section 2, which correspond to test cases, and generates an XML file for each one with the appropriate sequence of user input events.
- *Test Case Execution.* Generates each test class provided using the valid paths described into XML file by the test case generating module. Then, they can be executed by the ANDROID framework, and sends them to the devices to be executed using the UIAUTOMATOR tool which is an extension of JUnit tool using to write user interfaces test cases for ANDROID.

The following sections provide details about the internal behavior of the Modeling and Test Case Generator modules, which are the aim of this work.

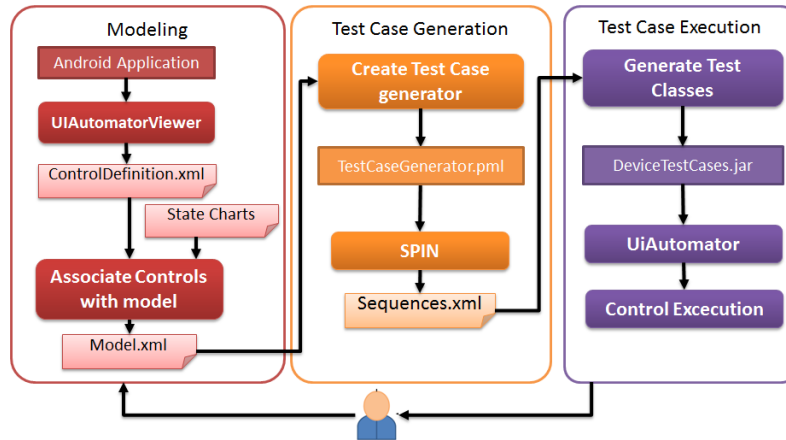


Figure 2: Test generation and execution process

4 Formal Description of models

In the following description, we define the behaviour of mobile applications through the composition of state machines at different abstraction levels. The lowest level is composed of *view state machines*. A view corresponds to a mobile screen, with its buttons, text fields, etc.. through which users may interact with the device. When the view is active, users may fire events through its interface. A *view state machine* models the possible behaviors of the user when he/she is making use of the view. These *behaviors* coincide with the sequence of events fired by the user. Sometimes one of these events makes a different view becomes active. We have modeled this control transfer between views through the *composition relation* of view state machines from which *device state machines* are constructed. Device state machines use the *connection states* to switch from the current active view to a different view. In this formalization, the specific applications to which each view belongs have not been taken into account, that is, we only model the transfer from one view to another, irrespective of whether both views belong to the same application. In the sequel, we use symbols $\vec{\rightarrow} / \vec{\rightarrow}_i$ to denote the transition relation of the view state machines M/M_i . In addition, symbol $\vec{\rightarrow}_c$ defines the binary relation which allows us to connect view state machines. Finally, $\vec{\rightarrow}_d$ represents the transition relation of the device state machine which is constructed from relations $\vec{\rightarrow} / \vec{\rightarrow}_i$ and $\vec{\rightarrow}_c$.

Since ANDROID applications are event driven, we may consider that each *test case* corresponds to the sequence of events fired which drive the mobile behaviour. In the formal description, events are the labels of transitions ($\vec{\rightarrow} / \vec{\rightarrow}_i, \vec{\rightarrow}_c, \vec{\rightarrow}_d$) and have the natural meaning. For instance, $s \xrightarrow{e} s'$ means that event e must be fired to be able to transit from s to s' .

4.1 View state machines

Definition 1 A view state machine is a tuple $M = \langle \Sigma, I, \vec{\rightarrow}, E, C, F \rangle$, where Σ is a finite set of states, $I \subseteq \Sigma$ are the initial states, $C \subseteq \Sigma$ are the so-called connection states, $F \subseteq \Sigma$ is the set of final states, E is the set of user events, and $\vec{\rightarrow} \subseteq \Sigma \times E \times \Sigma$ is the labelled transition relation. Sets I , C and F are mutually disjoint.

Final states are states from which it is not possible to evolve. *Connection states* are states from which it is possible to transit a different state machine. These states are essential to model the switch between

typical views of smart phone devices. Usually, when a new view is called, the execution of the system is supposed to return to the view caller. To take this behavior into account, we assume that each connection state $s \in C$ has a related state $return(s) \in \Sigma$ which represents the state to be returned when the new view invoked from s has finished its execution.

We partition set of events E into two disjunct sets: the set of *user events*, denoted as E^+ , which contains events such as pressing a button, etc., and the set of *system events*, denoted as E^- , which includes, for instance, events corresponding to system responses to user requests. In the following, we use e^+ , e^- to represent user events and system events, and we use e to refer to events which may be of any of both types.

View state machines are deterministic in the sense that if $s \xrightarrow{e} s_1$, and $s \xrightarrow{e'} s_2$ and $e = e'$, then $s_1 = s_2$ that is, the machine defines, at most, a transition for each pair state/input event.

We now define the notion of flow (an execution in a view state machine), and the test cases generated from flows.

Definition 2 Given a view state machine $M = \langle \Sigma, I, \bar{\rightarrow}, E, C, F \rangle$, we define the set $Flow(M) = \{s_0 \xrightarrow{e_1} s_1 \xrightarrow{e_2} \dots \xrightarrow{e_n} s_n | s_0 \in I, s_n \in F \cup C\}$ of all sequences of states, allowed by M , starting at an initial state of M , and ending at a final or connection state of M . The length of a flow is the number of its states. Given a flow of length n , $\phi = s_0 \xrightarrow{e_1} \dots \xrightarrow{e_n} s_n \in Flow(M)$, the sequence of events determined by ϕ (the test case) is $test(\phi) = e_1 \dots e_n$. We define the set of test cases allowed by M as $TC(M) = \{test(\phi) | \phi \in Flow(M)\}$.

According with Definition 2, test cases are finite sequences of user and system events. For instance, sequence $e_1^+ \cdot e_2^+ \cdot e_3^- \cdot e_4^+$ represents a test case where the user first fires events e_1^+ and e_2^+ , then the system fires e_3^- , and finally user fires e_4^+ . Thus, user and system events are similarly dealt with during the generation of test cases. The difference between them is of importance when test cases are transformed into code to be executed on the mobile as described in Section 5. User events will be transformed into non-blocking calls to methods that simulate the real occurrence of the event, while system events will correspond to calls to blocking methods which wait for the arrival of the system event.

4.2 Composition of view state machines

In this section, we describe how view state machines are composed to construct flows that navigate through different views representing realistic ways of using a mobile.

We first define the transition between different view state machines. This transition is realized through the *binary relation* \mathcal{R} defined between the connection and initial states. The idea is as follows. Assume that the flow in execution belongs to a view state machine M_i , and that a connection state cs of M_i has been reached. If relation \mathcal{R} defines a transition from cs to some initial state of other machine M_j , the flow could jump from M_i to M_j , and proceed following the transition relation of M_j . This jump implies the change in the activity visible in the device from M_i to M_j . In the sequel, we call *active* the view state machine which is visible in the device, and *create* to the rest of view state machines which have been created but are not currently visible in the device.

Given a finite family of state machines $M_i = \langle \Sigma_i, I_i, \bar{\rightarrow}_i, E_i, C_i, F_i \rangle$ we define $\Sigma = \cup_{i=1}^n \Sigma_i$, $I = \cup_{i=1}^n I_i$, $E = \cup_{i=1}^n E_i$, $C = \cup_{i=1}^n C_i$, and $F = \cup_{i=1}^n F_i$. In addition, we denote with $\mathcal{E} \subseteq E$ the set of *call events* that provoke the switch between active view state machines.

Definition 3 Let us assume a finite family of state machines, $M_i = \langle \Sigma_i, I_i, \bar{\rightarrow}_i, E_i, C_i, F_i \rangle$. The connection of view state machines M_1, \dots, M_n is given by a binary relation $\mathcal{R}(M_1, \dots, M_n) \subseteq C \times \mathcal{E} \times I$, that connects connection states with initial states. In the following, we denote 3-uples (s_i, e, s_j) of $\mathcal{R}(M_1, \dots, M_n)$ as $s_i \xrightarrow{e}_c s_j$. Observe that source and target machines i and j may coincide.

When a new view is created, the call event may specify some parameters that determine how it must be started or finished. For instance, if the view has already been created, the caller may choose whether to reuse the previously created view or, to the contrary, create a new one. Additionally, when the new created view has finished the execution, the caller view may automatically become active or not. Boolean functions $reuse, auto_return : \mathcal{E} \rightarrow \{false, true\}$ establish these parameters for the call events. Although there are other parameters that can be defined in the call events, these two are sufficient to describe the mobile behaviour.

We now define the *device state machine* which composes the behavior displayed by the view state machines using the connection relation.

Definition 4 Let us assume a finite family of state machines, $M_i = \langle \Sigma_i, I_i, \bar{\rightarrow}_i, E_i, C_i, F_i \rangle$, and a connection relation of M_1, \dots, M_n , $\mathcal{R}(M_1, \dots, M_n)$, as defined above. The device state machine

$$\mathcal{D} = M_1 ||| \dots ||| M_n ||| \mathcal{R}(M_1, \dots, M_n)$$

is defined as the state machine $\langle \Sigma \times \Sigma^* \times \mathcal{E}^*, I, \bar{\rightarrow}_d, E, F \rangle$ where

1. Σ^* is the set of finite sequences of states of Σ , and \mathcal{E}^* is the set of finite sequences of call events.
2. Transition relation $\bar{\rightarrow}_d$ is defined by the rules below.

We call *configurations* the states of device state machines. A configuration is a 3-uple $\langle s, h, eh \rangle$ where s is the current state of the active view state machine, i.e., the view visible in the mobile. Sequence h is the stack of states $s_1 \cdot s_2 \cdot \dots \cdot s_n$ which constitute the history of the view machines which have been created (and have not been yet destroyed) in the device but which are not currently visible. Each state s_i of $s_1 \cdot s_2 \cdot \dots \cdot s_n$ is a connection state of a view state machine which was active, but a transition from s_i to another view machine took place, and the view state machine of s_i became inactive. Finally $eh = e_1 \cdot \dots \cdot e_n$ is the history of events that have provoked a view switch in the current execution. Thus, $e_i \in \mathcal{E}$ is the event which fired the transition from state s_i to an initial state of another view state machine. In the following, ε represents the *empty* (event) history.

The evolution of configurations is given by the transition relation $\bar{\rightarrow}_d$ defined by the rules in Figure 3. Relation $\bar{\rightarrow}_d$ is constructed from the transition relations of view state machines $\bar{\rightarrow}_i$, and the binary connection relation $\bar{\rightarrow}_c$. In these rules, given a history of states $s_1 \cdot \dots \cdot s_n$ and the index j of a view state machine M_j , function $top : \Sigma^* \times \mathcal{N} \rightarrow \Sigma \cup \{\perp\}$ returns the last state of the view state machine M_j in the sequence $s_1 \cdot \dots \cdot s_n$. That is, $top(s_1 \cdot \dots \cdot s_n, j)$ returns s_k , if $1 \leq k \leq n$ is the biggest index such that $s_k \in \Sigma_j$, or \perp , if such a state does not exist.

Rule **R1** states that a transition inside a view state machine M_i corresponds to a transition in the device state machine. Rules **R2**, **R3** model a transition from machine M_i to machine M_j when both the new state s' and the event e are added to the state and event histories of the current system configuration. Rule **R2** is applied when event e does not involve reusing a previously created view ($reuse(e)$ is false), while **R3** applies when a view of M_j , should have been reused ($reuse(e)$ is true), but the current state history does not contain one ($top(s_1 \cdot \dots \cdot s_n, j) = \perp$). Rule **R4** defines a transition from machine M_i to M_j by reusing a previously created flow of M_j ($reuse(e)$ is true) which is stored in the configuration history ($top(s_1 \cdot \dots \cdot s_n, j) = s_k$). Finally, **R5** defines the case when the flow of the current active view has finished, and the execution must continue with the view stored at the top of the state history. Otherwise, that is, if $auto_return(e)$ returns false, the current configuration $\langle s, h, eh \rangle$ cannot evolve.

$$\begin{array}{ll}
\mathbf{R1.} \frac{s \xrightarrow{e}_i s'}{\langle s, h, eh \rangle \xrightarrow{e}_d \langle s', h, eh \rangle} & \mathbf{R2.} \frac{s \in C_i, s \xrightarrow{e}_c s', \neg reuse(e)}{\langle s, h, eh \rangle \xrightarrow{e}_d \langle s', h \cdot return(s), eh \cdot e \rangle} \\
\mathbf{R3.} \frac{s \in C_i, s' \in I_j, s \xrightarrow{e}_c s', reuse(e), top(s_1 \cdots s_n, j) = \perp}{\langle s, h, eh \rangle \xrightarrow{e}_d \langle s', h \cdot return(s), eh \cdot e \rangle} & \\
\mathbf{R4.} \frac{s \in C_i, s' \in I_j, s \xrightarrow{e}_c s', reuse(e), top(s_1 \cdots s_n, j) = s_k}{\langle s, s_1 \cdots s_n, e_1 \cdots e_n \rangle \xrightarrow{e}_d \langle s_k, s_1 \cdots s_{k-1}, e_1 \cdots e_{k-1} \rangle} & \\
\mathbf{R5.} \frac{s \in F_i, auto_return(e)}{\langle s, h \cdot s', eh \cdot e \rangle \xrightarrow{}_d \langle s', h, eh \rangle} & \\
\mathbf{R6.} \frac{c_0 \xrightarrow{e^+}_d c_1}{\langle c_0, c'_0, dh \rangle \xrightarrow{e^+}_{d||d'} \langle c_1, c'_0, dh + \{e^+\} \rangle} & \mathbf{R7.} \frac{c'_0 \xrightarrow{e^-}_{d'} c'_1, e^+ \in dh}{\langle c_0, c'_0, dh \rangle \xrightarrow{e^-}_{d||d'} \langle c_0, c'_1, dh - \{e^+\} \rangle}
\end{array}$$

Figure 3: Transition relation rules

Definition 5 Given a device state machine

$$\begin{aligned}
\mathcal{D} &= M_1 ||| \cdots ||| M_n ||| \mathcal{R}(M_1, \dots, M_n) \\
&= \langle \Sigma \times \Sigma^* \times \mathcal{E}^*, I, \xrightarrow{}_d, E \cup \mathcal{E}, F \rangle
\end{aligned}$$

1. the trace-based semantics determined by \mathcal{D} ($\mathcal{O}(\mathcal{D})$) is given by the set of finite/infinite sequences of configurations (flows) produced by the transition relation $\xrightarrow{}_d$ from an initial state, that is, $\mathcal{O}(\mathcal{D}) = \{ \langle s_0, \varepsilon, \varepsilon \rangle \xrightarrow{e_0}_d \langle s_1, h_1, eh_1 \rangle \cdots | s_0 \in I \}$.
2. Given a flow $\phi = c_0 \xrightarrow{e_1}_d c_1 \xrightarrow{e_2}_d c_2 \cdots \in \mathcal{O}(\mathcal{D})$, the test case determined by ϕ is the sequence of events $test(\phi) = e_1 \cdot e_2 \cdots$.
3. The set of test cases determined by a set of flows \mathcal{T} is $TC(\mathcal{T}) = \{ test(t) | t \in \mathcal{T} \}$.

Thus, a flow $\phi \in \mathcal{O}(\mathcal{D})$ consists of a sequence of view state machine flows (Definition 2) connected through *connection states*. Flow ϕ may finish at a final state of some view state machine, or may be infinite. The *length* $|\phi|$ of a flow ϕ is the number of its states (configurations), if it is finite, or ∞ , otherwise. Given a flow $\phi = c_0 \xrightarrow{e_1}_d c_1 \xrightarrow{e_2}_d c_2 \cdots \in \mathcal{O}(\mathcal{D})$, we define the truncated flow of n , ϕ^n , as ϕ iff $|\phi| \leq n$ or $\phi^n = c_0 \xrightarrow{e_1}_d c_1 \xrightarrow{e_2}_d c_2 \cdots \xrightarrow{e_{n-1}}_d c_{n-1}$, otherwise. Considering this, we define the set of traces $\mathcal{O}^n(\mathcal{D})$ as the set all traces of $\mathcal{O}(\mathcal{D})$ truncated up to length n , that is, $\mathcal{O}^n(\mathcal{D}) = \{ \phi^n | \phi \in \mathcal{O}(\mathcal{D}) \}$.

Observe that the state space of device state machines is not finite because configurations include the state and event histories which may have arbitrary lengths. In addition, the state space generated when an explicit model checker is constructing all the flows allowed by a device state machine is non-finite not only due to the state and event histories, but also because the matching algorithm, carried out during the state space search, must take into account both the current state of the flow and the history of the previous states of the flow. This allows that, for instance, flows $\phi_1 = \langle s_0, \varepsilon, \varepsilon \rangle \xrightarrow{e_1^+}_d \langle s_1, \varepsilon, \varepsilon \rangle \xrightarrow{e_2^+}_d \langle s_2, \varepsilon, \varepsilon \rangle \xrightarrow{e_3^+}_d \langle s_3, \varepsilon, \varepsilon \rangle$ and $\phi_2 = \langle s_0, \varepsilon, \varepsilon \rangle \xrightarrow{e_4^+}_d \langle s_4, \varepsilon, \varepsilon \rangle \xrightarrow{e_1^+}_d \langle s_1, \varepsilon, \varepsilon \rangle \xrightarrow{e_2^+}_d \langle s_2, \varepsilon, \varepsilon \rangle \xrightarrow{e_3^+}_d \langle s_3, \varepsilon, \varepsilon \rangle$ can be both generated by the model checker although when constructing ϕ_2 state s_1 has been already visited as explained in Section 2.

In consequence, the models of device state machines are not, in general, state finite which means that, the model checking process does not, in general, terminate. In the current implementation, we have solved this problem by bounding the depth of the execution flows analyzed generating $\mathcal{O}^n(\mathcal{D})$ for some fixed n .

4.2.1 Composing several devices

The extension of the state machine model to several devices is carried out by composing the device state machines by interleaving. Thus, if $c_0 \xrightarrow{e_1}_d c_1$ and $c'_0 \xrightarrow{e'_1}_{d'} c'_1$ are transitions in devices \mathcal{D} and \mathcal{D}' , respectively, then they allow the two following transitions, $\langle c_0, c'_0 \rangle \xrightarrow{e_1}_{d||d'} \langle c_1, c'_0 \rangle$ and $\langle c_0, c'_0 \rangle \xrightarrow{e'_1}_{d||d'} \langle c_0, c'_1 \rangle$ in the interleaved composition of \mathcal{D} and \mathcal{D}' . The communication between both devices is modeled by a user event in the sender device (the device that starts the communication), and a system event in the receiver device (the device that expects the message).

This is described in the last two rules of Figure 3. Rule **R6** handles the transition from the sender, and **R7** handles the transition in the receiver. Note that dh denotes the set of system events produced but not yet consumed. Thus, for instance, using the previous example, if $e_1 = e_1^+$ is an event that implies a communication from \mathcal{D} to \mathcal{D}' , and $e'_1 = e_1^-$ is the corresponding event to be read by \mathcal{D}' from \mathcal{D} , we would generate the test cases $e_1^+ \cdot e_1^-$ and $e_1^- \cdot e_1^+$. Note that in the second test case, the method that implements the transition for the receiver event will suspend the execution of \mathcal{D}' until event e_1^+ is fired by \mathcal{D} .

In addition, when dealing with more than one device, we make use of model checking optimization techniques such as partial order reduction [15] to avoid the generation of different test cases that correspond to a single feasible interaction between the devices.

5 Case Study

In this section, we describe how the behavior of mobile applications is modeled and how tests cases are automatically generated from these models.

5.1 Modeling

We first need to construct an abstract model of the system to be analyzed, using statecharts and following the notions of view and device state machines given in Section 4. This model should include the relevant user interactions for the tests we want to perform. For instance, a test case which is affected by whether the GPS is on or off may include user interactions to change its status, while other tests may not need those interactions. A test case generator will be created from this model using automatic transformations. This modeling step can be performed separately from the design of the application, or in combination as is custom in other model-driven tools such as IBM Rational Rhapsody [3]. In addition, the controls of each screen have to be extracted and modeled, so that transitions on the state machines can be tied to actions performed on these controls.

We use a scenario with two applications, *Facebook* and *YouTube*. This scenario is composed of three views (*HomeView*, *CommentView* and *MovieView*), which describe the behavior of a user placing a link to a YouTube video in a Facebook comment, and watching this or other videos in YouTube. The state machines can be modeled using UML as shown in Figure 4. These state machines include additional information required to correlate them with the applications and their views. An XML definition of the model can then be automatically generated from these state machines. Listing 2 shows part of this XML

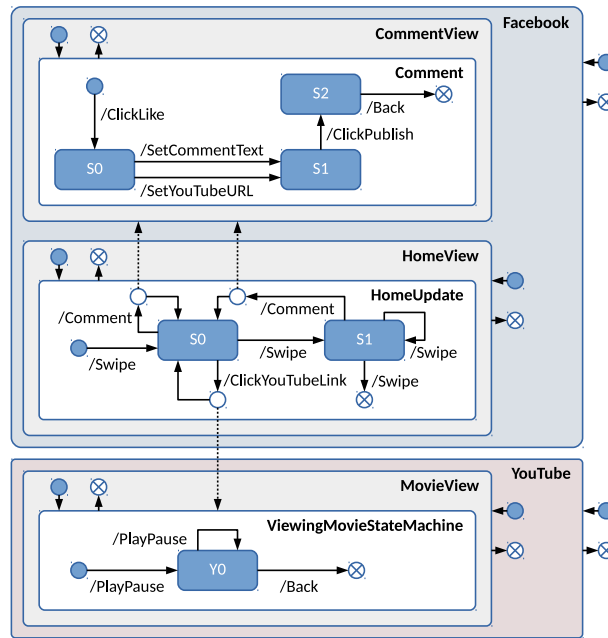


Figure 4: Facebook and YouTube model



Figure 5: Identifying control groups

definition¹. In particular, it contains the state machine associated with the *Home* view of the Facebook app. Each state machine may define several states and transitions. In addition to simple transitions between states of the same state machine, it is also possible to define transitions that call another state machine and, upon its termination, continue in the caller machine. The *type* and *through* attributes identify the type of transition and the state machine to call (if any). Listing 2 provides examples of both simple (line 10) and complex (lines 9 and 11) transitions. Each transition has a unique ID within its view that is used to identify transitions, and also declares the user or system event that triggers it.

The events that fire the transitions in Figure 4 are the user actions performed on controls placed in visible views. We organize controls into *group of controls* according to the actions associated with each. Figure 5 shows some of the control groups that have been identified in the *CommentView* View. For instance, the *Comment* group could represent any of the text fields to write a comment, and the *clickYouTubeLink* identifies links to YouTube videos.

These control groups are declared in a *control definition* file with the help of the UIAUTOMA-

¹More complete versions of this and other parts of the model are available online at <http://morse.uma.es>.

```

1 <Application name="Facebook" package="com.facebook.android">
2   <Views>
3     <View name="HomeView" controlsFile="Home.xml" >
4       <StateMachines>
5         <StateMachine name="HomeUpdate">
6           <States><State name="S0"/><State name="S1"/></States>
7           <Transitions>
8             <Transition ID="1" event="Swipe" prev="" next="S0" type="Simple"/>
9             <Transition ID="2" event="Comment" prev="S0" next="S0"
10              through="CommentView" type="View"/>
11             <Transition ID="3" event="Swipe" prev="S0" next="S1" type="Simple"/>
12             <Transition ID="4" event="ClickYouTubeLink" prev="S0" next="S0"
13              through="ViewingMovieStateMachine" type="StateMachine"/>
14             <Transition ID="5" event="Swipe" prev="S1" next="S1" type="Simple"/>
15             <Transition ID="6" event="Comment" prev="S1" next="S0" through="CommentView"
              type="View"/>
              <Transition ID="7" event="Swipe" prev="S1" next="" type="Simple"/>
              ...

```

Listing 2: State machine configuration

```

1 <node index="0" text="" testGroup="" ...
2   <node index="0" ...
3     <node testGroup="clicLike" IsFixedValue="" PatternOrValue="" index="0" text="Like"
4       resource-id="id/feed_feedback_like_container" clickable="true"
5       long-clickable="false" password="false" ... />

```

Listing 3: Control group definition

TORVIEWER tool [1]. This tool analyzes each view without requiring its source code, and generates an UIX (XML) file containing the hierarchy of controls in the view. Listing 3 shows part of the generated file for the *Facebook* application. The attributes associated with each control in the UIX file include the kind of actions that the control supports, such as `clickable` or `scrollable`. The UIX file is then customized to bring together the controls which belong to the same group by setting the `controlGroup` attribute. Some controls accept parameters which may also be included as attributes in this file. For instance, the values introduced in text fields may be fixed or generated automatically according to some pattern.

5.2 Test case generation

We are now ready to generate the corresponding test cases in an exhaustive manner. The XML file is automatically transformed into a PROMELA specification that follows the same principles described in Section 2, but with a few additions to accommodate the structure of ANDROID applications. Each device is still represented by a single process, but their state machines are defined in separate *inlines*, one per each app, view and state machine, which can then be composed. In addition, there may be device-specific app and view inlines, since views and state machines can be assigned to a particular device. In the simplest form of composition, device processes call their app inlines, app inlines call their view inlines, and view inlines call their state machine inlines. On the other hand, a state machine may call another view or state machine. When this happens, the state of the previous state machine should be stored such that when the new one is finished, the state is correctly restored. To support this we introduce a *backstack* data structure, where the state of the current state machine is always at the top of the stack.

Listing 4 shows a simplified extract of the PROMELA specification generated for the Facebook and

```

1 typedef Backstack { mtype states[MAX_BK]; short index; }
2 Backstack backstacks[DEVICES];
3 #define currentBackstack    devices[device].backstack
4 #define currentState       currentBackstack.states[currentBackstack.index]
5 active proctype device_219dcac41() {
6     if
7         true -> app_219dcac41_Facebook(D_219dcac41);
8         true -> app_219dcac41_YouTube(D_219dcac41);
9     fi;
10    devices[D_219dcac41].finished = true
11 }
12 inline statemachine_Facebook_HomeView_HomeUpdate(device) {
13     currentState = State_Facebook_HomeView_HomeUpdate_init;
14     pushToBackstack(device, State_Facebook_HomeView_HomeUpdate_init);
15     do
16         :: currentState == State_Facebook_HomeView_HomeUpdate_S0 ->
17             transition(device, VIEW_HomeView, 2);
18             view_Facebook_CommentView(device);
19             currentState = State_Facebook_HomeView_HomeUpdate_S0
20         :: currentState == State_Facebook_HomeView_HomeUpdate_S0 ->
21             transition(device, VIEW_HomeView, 4);
22             statemachine_YouTube_MovieView_ViewingMovieStateMachine(device);
23             currentState = State_Facebook_HomeView_HomeUpdate_S0
24     ...
25     od;
26     popFromBackstack(device)
27 }

```

Listing 4: PROMELA specification for Facebook and YouTube

YouTube example. The backstack data structure is shown on line 2. A new element is pushed to or popped from the backstack at the beginning or end of a state machine, respectively (lines 14 and 26), while `currentState` always point to the top of this stack for each device.

Each sequence of transitions generated by SPIN is translated into a `UiAutomatorTestCase` subclass, where each transition is implemented by a method. This class simulates the actions performed by the user, such as pressing buttons or swiping. The code shown in Listing 5 shows part of a test case obtained from the model of Figure 4, in particular a user adding a link to a YouTube video in a Facebook comment, and later watching that video on the YouTube application. The file is compiled into a `.dex` (ANDROID application binary) file, and then deployed into a ANDROID device and executed using the `adb` tool.

Table 1 provides some quantitative results of the number of test cases generated and the computational effort required, for several scenarios, averaged over three runs. Device 219dcac4 was assigned only the Facebook application, while device 219dcac41 was assigned both, although in both cases other modeled applications may be reached from the assigned ones. The fourth column declares the maximum depth allowed for the test case transitions generated for a device. The fifth column represents the total time spent to generate the test cases from the XML models. The last three columns are stats taken from the SPIN execution, namely the number of SPIN states generated, the size of each state, and the total memory spent, respectively. These results show how adding the YouTube application, which is fairly isolated, has little impact in the results (rows 3 and 4 of data).

6 Comparison with related work

There are other proposals to apply model-based testing to ANDROID applications. Some of them consider that the testing process starts without a precise model of the expected behavior of the applications and

```

1 public class TestDevice1 extends UiAutomatorTestCase {
2     // Transition 2 previous S0 next S0 on view HomeView
3     public void TestFacebookComment2() throws UiObjectNotFoundException {
4         UiObject control = new UiObject(new UiSelector().
5             className("android.widget.TextView").index(1).textContains("Comment"));
6         control.click();
7     }
8     // Transition 4: previous S0 next S0 on view HomeView
9     public void TestFacebookclicYouTubeLink27() throws UiObjectNotFoundException {
10        UiObject control = new UiObject(new
11            UiSelector().className("android.view.View").index(3));
12        control.click();
13    }
14    // Transition 1: previous next Y0 on view MovieView
15    public void TestYouTubeplaypause28() throws UiObjectNotFoundException {
16        UiObject control = new UiObject(new
17            UiSelector().className("android.view.View").index(4));
18        control.click();
19    }
20 }

```

Listing 5: Generated UiAutomatorTestCase

Devices		Configuration		Results				
219dcac4	219dcac41	Backstack	Transitions	# Test Cases	Time (s)	# States	State Size (B)	Memory (MB)
✓		4	20	5641	1.0	307234	84	156.8
✓		4	26	111317	9.0	6063398	92	728.6
	✓	4	20	5660	1.0	307493	84	156.8
	✓	4	26	111342	9.0	6063735	92	728.6
✓	✓	4	10	1872	7.0	4039337	100	560.3
✓	✓	4	12	12180	52.3	28972472	108	3445.2

Table 1: Test case generation results

they focus on techniques to obtain such model. *MobiGUITAR* framework [5] automatically construct a state machine of one application by executing events in the running application and recording a tree with fireable events for each new state. The authors use a "breadth-first" traversal of the apps GUI for open source applications. As far as they are not considering any knowledge on the way of using the application but they are making an exhaustive execution, they need some criteria to assume whether some states are equivalent to prevent state explosion. The *Swift-Hand* technique proposed in [8] employs machine learning to construct an approximated model of the application during the testing process. Their aim is to cover as much behavior as possible, making the execution to enter in unexplored parts of the state space. In our method, we separate test generation from testing and the states in our high-level state machines are limited and differentiated by design. So our models are more compact, and for instance, compared with *MobiGUITAR* we do not need extra work to remove unrealistic test cases. In addition, our approach allows to generate test cases for several applications that interact using ANDROID intents, while the complexity of the runtime based modeling process for *MobiGUITAR* and *Swift-Hand* makes them more suitable for single applications.

Like in our proposal, other works also consider the existence of a formal specification of the applications to start the test generation. In [17] the authors describe how to follow a property-driven method build the models in Alloy, a formal language based on high order logic. In their proposal the role of the model checker in our approach is done they the Alloy analyzer, that generates positive (expected) and negative (undesired) test cases. Like in our approach, they use XML based transformations to translate

the test cases to some executable form in order to activate the applications under test. Apart from the inside technologies (model checking vs constraint solver), the main difference in both proposals is the way to obtain the refined executable model. The Alloy specification in [17] is constructed manually, while the PROMELA specification in our work is done automatically from the high level design of the view state machines. We still need to work in the same case study to get a quantitative comparison on the human and computational effort required in both approaches.

There are other model-based testing tools for Android which are not focused on models that consider the user inputs. For instance, the tool APSET [23] considers manually constructed formal models of vulnerability patterns to generate test cases for ANDROID applications. Test generation is implemented with an ad-hoc algorithm that also considers the compiled code of the application and the configuration files in the ANDROID system.

7 Conclusions and Future Work

ANDROID systems have a complex architecture designed to support the concurrent execution of applications on devices with limited resources. Here we have presented a model-based testing approach for generating test cases for ANDROID applications, which takes into account the way in which these applications interact with the user and with each other. We model the expected user behavior by composing state machines, and then explore this model exhaustively with SPIN to obtain all possible user behaviors, which correspond to test cases. These test cases are then executed in the device simulating the user inputs. In contrast with other approaches that generate random input events, our approach produces realistic user behaviors. Although our tool is currently geared towards ANDROID, the same principles can be applied to analyze applications in other mobile platforms, such as IOS and WINDOWS MOBILE.

The next step of our work will be to connect the generated test cases with a runtime verification monitor DRAGONFLY [14, 13]. In addition, we are working in adding more runtime information, like energy consumption, to perform richer analysis.

References

- [1] *Android developers*. [Http://developer.android.com/](http://developer.android.com/).
- [2] *DroidPilot*. [Http://droidpilot.wordpress.com/](http://droidpilot.wordpress.com/).
- [3] *IBM - Software - Rational Rhapsody family*. [Http://www-01.ibm.com/software/awdtools/rhapsody/](http://www-01.ibm.com/software/awdtools/rhapsody/).
- [4] *Robotium*. [Https://code.google.com/p/robotium/](https://code.google.com/p/robotium/).
- [5] Domenico Amalfitano, Anna Rita Fasolino, Porfirio Tramontana, Bryan Ta & Atif Memon (2014): *Mobi-GUITAR – A Tool for Automated Model-Based Testing of Mobile Apps*. *IEEE Software* 99(PrePrints), p. 1, doi:10.1109/MS.2014.55.
- [6] Cyrille Artho, Howard Barringer, Allen Goldberg, Klaus Havelund, Sarfraz Khurshid, Mike Lowry, Corina Pasareanu, Grigore Rosu, Koushik Sen, Willem Visser & Rich Washington (2005): *Combining Test Case Generation and Runtime Verification*. *Theor. Comput. Sci.* 336(2-3), pp. 209–234, doi:10.1016/j.tcs.2004.11.007.
- [7] Pedro de la Cámara, J. Raúl Castro, María del Mar Gallardo & Pedro Merino (2010): *Verification support for ARINC-653-based avionics software*. *Software Testing Verification & Reliability* 21(4), pp. 267–298, doi:10.1002/stvr.422.
- [8] Wontae Choi, George Necula & Koushik Sen (2013): *Guided GUI Testing of Android Apps with Minimal Restart and Approximate Learning*. *SIGPLAN Not.* 48(10), pp. 623–640, doi:10.1145/2544173.2509552.

- [9] Edmund M. Clarke, Jr., Orna Grumberg & Doron A. Peled (1999): *Model Checking*. MIT Press, Cambridge, USA.
- [10] Pedro de la Cámara, María del Mar Gallardo, Pedro Merino & David Sanán (2009): *Checking the reliability of socket based communication software*. *Intl. Journal on Software Tools for Technology Transfer* 11(5), pp. 359–374, doi:10.1007/s10009-009-0112-7.
- [11] M.B. Dwyer, Robby, O. Tkachuk & W. Visser (2004): *Analyzing interaction orderings with model checking*. In: *Automated Software Engineering, 2004. Proceedings. 19th Intl. Conference on*, pp. 154–163, doi:10.1109/ASE.2004.1342733.
- [12] William Enck, Peter Gilbert, Byung-Gon Chun, Landon P. Cox, Jaeyeon Jung, Patrick McDaniel & Anmol N. Sheth (2010): *TaintDroid: An Information-flow Tracking System for Realtime Privacy Monitoring on Smartphones*. In: *Proceedings of the 9th USENIX OSDI, OSDI' 10, USENIX Association, Berkeley, CA, USA*, pp. 1–6, doi:10.1145/2494522. Available at <http://dl.acm.org/citation.cfm?id=1924943.1924971>.
- [13] Ana Rosario Espada, María-del-Mar Gallardo & Damián Adalid (2013): *DRAGONFLY: Encapsulating Android for Instrumentation*. In: *Proceedings of the XIII PROLE13*.
- [14] Ana Rosario Espada, María-del-Mar Gallardo & Damián Adalid (2013): *A Runtime Verification Framework for android Applications*. In: *Proceedings of XXI JCSD*.
- [15] Gerard J. Holzmann (2003): *The SPIN Model Checker: Primer and Reference Manual*. Addison-Wesley Professional.
- [16] Cuixiong Hu & Iulian Neamtiu (2011): *Automating GUI Testing for Android Applications*. In: *Proceedings of the 6th International Workshop on AST, AST '11, ACM, New York, NY, USA*, pp. 77–83, doi:10.1145/1982595.1982612.
- [17] Yiming Jing, Gail-Joon Ahn & Hongxin Hu (2012): *Model-Based Conformance Testing for Android*. In: *Advances in Information and Computer Security - 7th International Workshop on Security, IWSEC 2012, Fukuoka, Japan, November 7-9, 2012. Proceedings*, pp. 1–18, doi:10.1007/978-3-642-34117-5.1.
- [18] William Klieber, Lori Flynn, Amar Bhosale, Limin Jia & Lujo Bauer (2014): *Android Taint Flow Analysis for App Sets*. In: *Proceedings of the 3rd ACM SIGPLAN International Workshop, SOAP '14, ACM, New York, NY, USA*, pp. 1–6, doi:10.1145/2614628.2614633.
- [19] Yepang Liu & Chang Xu (2013): *VeriDroid: Automating Android application verification*. In: *Proceedings Middleware 2013 Doctoral Symposium, ACM*, doi:10.1145/2541534.2541594.
- [20] Aravind Machiry, Rohan Tahiliani & Mayur Naik (2013): *Dynodroid: An Input Generation System for Android Apps*. In: *Proceedings of the 2013 ESEC/FSE, ESEC/FSE 2013, ACM, New York, NY, USA*, pp. 224–234, doi:10.1145/2491411.2491450.
- [21] Peter Mehrlitz, Oksana Tkachuk & Mateusz Ujma (2011): *JPF-AWT: Model checking GUI applications*. *2011 26th IEEE/ACM International Conference ASE 2011* 0, pp. 584–587, doi:10.1109/ASE.2011.6100131.
- [22] Heila van der Merwe, Brink van der Merwe & Willem Visser (2012): *Verifying Android Applications Using Java PathFinder*. *SIGSOFT Softw. Eng. Notes* 37(6), pp. 1–5, doi:10.1145/2382756.2382797.
- [23] Sébastien Salva & StassiaR. Zafimiharisoa (2014): *APSET, an Android aPplication SEcurity Testing tool for detecting intent-based vulnerabilities*. *International Journal on Software Tools for Technology Transfer*, pp. 1–21, doi:10.1007/s10009-014-0303-8.
- [24] Tommi Takala, Mika Katara & Julian Harty (2011): *Experiences of System-Level Model-Based GUI Testing of an Android Application*. In: *Proceedings of the 2011 Fourth IEEE ICST, ICST '11, IEEE Computer Society, Washington, DC, USA*, pp. 377–386, doi:10.1109/ICST.2011.11.
- [25] Mark Utting & Bruno Legeard (2007): *Practical Model-Based Testing: A Tools Approach*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA.
- [26] Hui Ye, Shaoyin Cheng, Lanbo Zhang & Fan Jiang (2013): *DroidFuzzer: Fuzzing the Android Apps with Intent-Filter Tag*. In: *Proceedings 11th International Conference on Advances in MoMM2013, ACM*, doi:10.1145/2536853.2536881.