INFORMATION TECHNOLOGY

UDC 004.056.52/004.432(045)

Vasyl Melnyk¹
Katerina Melnyk²
Oksana Zhyharevych³

STATIC ANALYSIS OF SOURCE CODE MODELED FOR JAVA-PROGRAMS CONTAINING APPLICATIONS WITH ANDROID SECURITY

Lutsk National Technical University
75, Lvivska st., Lutsk, 43018, Ukraine
E-mails:¹melnyk_v_m@yahoo.com; ²ekaterinamelnik@gmail.com; ³oz_lutsk@mail.ru

Abstract. A static analysis techniques were combined with model-based deductive verification using solvers of the static model theory (SMT) to create a framework that, given an aspect of analysis of the source code, automatically generated with an analyzer outputting a conclusion information about this aspect. The analyzer is generated by translating of a program collecting semantic to outlined formula in first order over a few multiple submitted theories. The underscore can be looked as some set of holes or contexts corresponding to the uninterpreted APIs invoked in the program. As the program makes an import of the packages and uses classes’ methods of these packages, it is importing the semantics of API invocations in first order assertion. The analyzer is using these assertions as models and their first logic order formula incorporates the specification behavior (its negation) of the described programs. A solver of SMT-LIB formula is treated as the combined formula for “constrain” and “solve” it. The “solved” form can be used for logic errors (security) identification Android-based Java-programs. The properties of Android security are represented as constraint and analysis aims to show the respecting for these constraints.

Keywords: Android security; Java-program; static analysis; static model theory; program code.

1. Introduction

Software systems stenically manage mission-critical activities in organizations which rely on dependable, situation award and delivered in time classified or sensitive information. Information streams in such organizations are usually machined by common-built, with open source code, and traditionally cooperated security software. Such programs may be unlicensed with malicious code or vulnerability containing, which can be used for insider or outsider to seize a confidential data, reclassify documents, make some destroyed actions, and modify valuable information. So, the reliability and correct work of described above software to drive these systems have become important issues for today.

Program-independent errors in software systems like overflows of buffer or null dereferences can be used by malicious applications to make unsecured holes, through which unsecured confidential data can be captured. Most of these bugs are detected too late, when destructive effects are already appeared [1] complicating the task of runtime preceding mechanisms of fault handling for ensuring data recovery. A more important is the lack of peculiar tools for logical dependencies of program-dependent errors detecting in applications. A testing of a well-known incident list resulting from software glitches opens that logical application-depended errors were the causes in [2,3] works. A lot of these logical faults were hard and difficult to reveal (opposite to simple errors) with using observed testing methods [4] alone.

In this paper we combine the static analysis techniques with model-based deductive verification using solvers of the static model theory (SMT) to create a framework that, given an aspect of analysis of the source code, automatically generated with an analyzer outputting the conclusion information about this aspect. The analyzer is generated by program translation for collecting semantic to outlined formula in first logic order over a few multiple submitted theories. The underscore can be looked as some set of holes or contexts corresponding to the uninterpreted APIs invoked in the program. As the program makes an import of the packages and uses classes’ methods of those packages, it is importing the semantics of API invocations in first order assertion. The analyzer is using these assertions as models and their first logic order formula incorporates the specification behavior (its negation)
of the described programs. A solver of SMT-LIB formula is treated as the combined formula for “constrain” and “solve” it. The “solved” form can be used for logic errors (security) identification of Android-based Java-programs. The properties of Android security are represented as constraint and analysis aims to show the respecting for these constraints.

2. Related works

Software verification and validation techniques for software can be divided in three categories. The first one includes informal methods such as testing and monitoring. These techniques are scaling well by the most used technique in practice to validate software systems. The testing accounts for forty to sixty percent for efforts of developments [4, 5]. The traditional methods for software program testing [6], however, do not allow for formal specification and verification of high logical properties that need to satisfy the system. In the area of security critical software where exponential goes up in the number of possible situations to be dealt is inevitable, traditional testing methods can be with difficulty used to provide some confidence level. The second category of methods for software verification and validation includes traditional formal methods such as verification model and providing of theorems are sometimes too heavy and rarely can be used in practice without considerable manual work.

Checking of the model includes an automatic approach for verification more successful while dealing with finite state systems. It suffers not only from infamous state explosions problem but it also need the model performance for program software.

Third method category for program software verification and validation is settled on static analysis techniques and abstract interpretation [7,8]. The static analysis refers to automatic displaying behavior technique for programs during the time of compiling. While static analysis tools have touched with a great practical success, and have been integrated with state of the newer compilers, such tools can reveal only small and simple errors due to the lack of their deductive strength. So, traditional tools of static analysis cannot reveal a deadlock presence or the violation of common exclusion in concurred programs. The abstract interpretation is a technique for the program semantic collection, comparison and combination. It successfully has been used for inferring of program runtime properties that may be used for program optimization. On the next the most successful approaches for program analysis are reviewed. In the last years a lot of work has been made in the area of software static analysis. Some tools of static analysis, such as described in articles [9-14] are making lightweight analysis of data flow. In the [15] conference article the data flow analysis is provided for verification, described in "metalanguage", designed for checking "automata" encoding. Astree is a static program analyzer directed to confirm the runtime error absence in the installed programs, and can handle only "safe" C-subset rather than all C-language. It concerns only for separate runtime errors but general program properties. In [16] the linear relation analysis is used to reveal invariant linear inequalities numerical program variables. Their methods have been used for verify (analyze delays) in synchronous programs written in "Lustre-language". A few technique approaches have been considered to provide an approximate answer to the validation problem as enlargement, convex approximations and factoring of Cartesian [17]. These approximations are realized with using of part of polyhedral library. In the article [18] the predicate abstraction is used to analyze hybrid systems. In this method the finite abstraction of hybrid automated system is creating with priory using initial predicates taken of the user.

3. Android platform

Android is the software to develop mobile devices. It includes an operating system, key applications and middle-connection software. The main feature of the Android is that an application can use elements of other programs. To rich this, the system has to start the process when any of the other programs part is need. Android has no any single point of entrance like function main(). They have essential components, which the system can instantiate and start if necessary. There are four main component types: activity, services, audio receivers and connect providers.

Android has an aim to active the first three components. For activities and services, the intent is the pair: <action_name, data>, to display the previous action, which receiver has to except and data to process. In the framework of the represented program analysis this aimed object is keyword used to make data flow analysis and call of analysis function.

Android-architecture provides created applications with phone functions and protects users to minimize mistake consequences and harmful software. As it was said, in Android an application may exchange its data
and functionality with other software, and these accesses have to be controlled carefully to provide the security. Android permissions are rights given to programs to perform such functions like photocopy, GPS using and phone call making. When applications are installing, they accept the unique identifier UID, and every application always runs under its name with the particle appointed device. Application UID is using for protection its data exchange with other applications.

4. Model architecture

Fig. 1 shows architecture of static analysis proposed on the model basis. The abstract collecting semantic of Java-program are represented as "marked" coercions. "Marked" can be considered as a set of holes and contexts corresponding to no interpreted APIs, i.e. API from library with unknown semantics. As program imports the packages and uses classes' methods in imported packages, the semantics of API invocations are importing here as first order assertions (constraints). These assertions are models used to "unmark" constraints of the abstract collecting semantics i.e. for "fill in" uninterpreted APIs "holes". Aspects of analysis are specifying as constraints.

![Model architecture diagram](image)

A. Verification of the permission

Android permissions can be separated into different security levels [20, 21], which are defined in the Manifest.xml file. The framework examines this file and reveals permission information. The necessary list of permissions should be formed from analyzing of APIs invoked in the program and can be compared with permission information P taken from file Mainifest.xml. The android API calls can be mapped with necessary permission list used in [22]. If the every permission is provided we conclude that application has no permission violations. Otherwise, API calls which have no provided appropriate permissions will be considered as with return of -1 in the following analysis.

B. Constraints, solvers of SMT-LIB formula, satisfiability

Constraints are special formulas in first logic order [23], and a constraint system itself specifies the syntax and constraints semantics. The solver of the constraint implements a checking algorithm to verify the satisfiability/consistency of constraint sets using constraint theory, i.e., revealing if there exists a variable assignment that satisfies the constraints. A solver uses constraint theory axioms together with simplification rules as rewritten rules for the constraints transforming to a normal form which is called the "solved" form. The last constraint which results from the computation will be named the answer.

C. SMT-LIB formulas and Yices

Satisfiability Modulo Theories (SMT) libraries (SMT-LIB) [24] provide a framework for the first order formulas satisfiability checking in agreement with some background logical theories. SMT provides background theories standard description used in SMT systems. It gives a common input and output languages for SMT formula solvers. "Yices" described in [19] is effective SMT-LIB formula solver, that decides the arbitrary formulas satisfiability containing uninterpreted function symbols with equality, linear real and integer
arithmetic, scalar types, recursive data types, tipples, records, extensional arrays, fixed bit-vectors, lambda expressions, and quantifiers.

5. Java programs inferring collecting semantics

We have to perform intraprocedural and interprocedural analysis to analyze deep logical properties of Java-based programs. Using data flow analysis of the source code in intraprocedural analysis, in a series of steps is built a constraint system for capturing its collecting semantics. In the way of the interprocedural analysis the call graph has to be built and some external rules have to be defined for relating the different API invocations and detecting if the analyzed code breaks these rules.

In the analysis framework the following sequence of steps (SSA) has to be followed to check if the program satisfies an analysis aspect defined by the user. 1) The dataflow analysis of the Java source code has to be performed and its collecting semantic has to be generated, 2) The static single assignment graph of the program, based on the dataflow analysis results, has to be generated, 3) The SSA graph has to be convert to the SMT-LIB formulas (see below), 4) On the last step the models of uninterpreted API invocations have to be imported as first order assertions.

The above steps can be illustrated in the following code:

```java
1 class udhpced
2 {
3    int getSocket (int listen_mode)
4    {
5        int fd = 0;
6        if(listen_mode == 2) {
7            fd = listen_socket();
8        }
9    }
10   else {
11      fd = raw_socket();
12    }
13 }
14 }
15 }
```

In the program shown above listen_mode is the input of the user, listen_socket() is a return method with positive integer, raw_socket() is a method, providing of operated system, due to return specific positive integer. Provided data flow analysis of the source code is shown in fig 2. The integer number in the graph of the data flow indicates the code line number of each statement.

After line 12 the variable value fd can be as return value of listen_socket() or raw_socket(), because the above program has two branches. During the compilation time we can’t determine, what way the control will pass. So, we consider that fd value is `{ listen_mode = 2 ∧ fd = listen_socket(); listen_mode ≠ 2 ∧ fd = raw_socket() }, where the semicolon present disjunction. To build SMT-LIB logic formulas for capturing the collection program semantics, we have to convert the program to static single assignment graph as shown in fig. 3.

![Fig. 2. Flow of Data](image)

![Fig. 3. Static Single Assignment for the program](image)
semantics are incorporated and if API has no permission provided we confirm that API returns -1. Else an assertion, characterizing API from site of the model library is imported. For example, for listen_socket and raw_socket, the assertion (assert (and (> listen_socket 1) (= raw_socket 1))) is imported.

The main program condition considered above is $fd > 1$. Verifying, if this condition holds is considered an aspect of the analysis for this program, which is included SMT-LIB formula, to characterize the collecting semantics of the program by adding the conjunct ($< fd4 1$). The combined SMT-LIB formula was unsatisfyable by the solver in [19], which indicates that the program satisfies the specification.

Now, the algorithm for SSA graph program converting to SMT-LIB formulas that captures it’s collecting semantics, will be described down. Let $G = (N, e)$ be the SSA program graph. In this graph, every node represents a statement in the program. The if and loop conditions can be represented here as an edge graph labels, and SMT-LIB formulas can be generated to capture the program collecting semantics using algorithm 1 that formalizes the described above intuition.

Algorithm 1 to convert SSA algorithm into SMT algorithm

for $n \in N$ do
  if $n$ is a simple assignment statement $VAR = EXP$ then
    Create an assertion (assert (= VAR EXP)) in SMT;
  end if
  if $n$ is a assignment statement with API call $VAR = API$ then
    Create an assertion (assert (= VAR API));
  end if
  if $n$ is a $\phi$ function statement then
    Let $v$ be the variable in this statement and the set $W$ be the values of this $\phi$ function;
    Create a disjunction $v = wi$, where $wi \in W$;
  end if
  if $n$ is a function call statement $FUN()$ then
    Create an assertion (assert (= FUN SUMMARY));
  end if
end for

6. Experiments and limitations

The source code from Android Bluetooth ChatServices application was analyzed above. This program builds a Bluetooth network platform to allow for device to exchange data with other devices of Bluetooth. It has three main functionalities: 1) searching of Bluetooth devices, 2) pair and connect the devices 3) transfer data between these devices. The application uses such API calls as: 
BluetoothAdapter.getDefaultAdapter(), BluetoothAdapter.getRemoteDevice(address), BluetoothSocket and BluetoothChatService.

These API models are provided here which are based on the Android Development Documentation. For this application we simply consider, that application can set up the Bluetooth service and can connect to any discovered devices. So, we set up BluetoothChatService to return non-null object, and in the SMT-specification we model the value of API-function as not returned -1. The source code analyzed satisfied specification. For this, a few Android-free source codes of the applications were been downloaded and run the static analysis tool to the source code described above. Some possible program vulnerabilities were been detected from the analysis, are under-mentioned.

Android SMSPopup: 1) in class SmsReceiverService.java there is a false null checker for statement. The branch if(message.isSms() && message.getMessageClass() == MessageClass.CLASS 0) will be never reached,
2) in class SmsPopupUtils.java the method getUnreadSmsCount(Context context) will be never called, 3) there is a system command call Runtime.getRuntime().exec(commandLine.toArray(new String [0])).getInputStream(), this statement may provide a command injection error.

**openGPStracker:** 1) in class Constants.java on line 98, there is a hardcoded password, 2) The function serializeWaypoints() in GpxCreator.java fails to perform a null checker for variable mediaUri (line 440), 3) The expression if(startImmidiatly && mLoggingState == Constants.STOPPED) (line 563) in GPSLoggerService.java is evaluated as true every time, the branch else will be never reached.

**OpenSudoku:** 1) method update() in IMNumpad.java (line 208 and 229) fails to perform a null checker for statement, 2) the method saveToFile() in FileExportTask. Returns of Java in a catch block (line 156) may lead to a lost error of the return value.

In the application Android SMSPopup the command injection error has been detected. The command statement is an array that comes from other function which possibly providing a wrong statement. In openGPStracker it is revealed more permissions than necessary; that may lead to the problem of overprivileged permission. There are also detected that the application openGPStracker is with hardcoded password.

### A. Limitations

In the intraprocedural analysis, the constraint system includes all the possible variable values and this may result in false positives. There are necessary more sharp abstract interpretation methods to provide some more precise analysis. In the interprocedural analysis the summary based functions are modeled and this abstraction loses precision and gives out false negatives. Another problem of described above analysis tool is that it needs developers for external XML files creation to specify the right program properties; these files are difficult to build.

### 7. Conclusions

All Android programs can communicate to each other through system provided mechanisms such as files, activities, services, broadcast receivers and providers for communications. If developers use one of these techniques they have to be sure, that they communicate with the right entity, because permissions can be easily violate here inadvertently. The analysis program framework described in this paper can help developers to reveal programming errors and the statistic of the permission violations. Developers have to understand what permissions have to be set up correctly and what basic knowledge has to be need in the logic and limited solving. The task to help users to set up easily constraints and logical errors in programs can be accepted for the future work.

### References


Для Java-програм, які містять додатки з безпекою Android

Здійснено поєднання методів статичного аналізу з моделлю дедуктивної перевірки й використанням рішень теорії статичної моделі (ТСМ) для створення основи, яка, враховуючи аспект аналізу вихідного коду, автоматично створюється за допомогою аналізатора, котрий виводить кінцеву інформацію про цей аспект. Аналізатор генерується шляхом перекладу програми для збору семантики з метою викладення формул в першому наближенні на основі кількох представлених теорій. Оскільки програма здійснює імпорт пакетів і використовує класові методи цих пакетів, вона імпортує семантику викликів API в наближені першого порядку. Аналізатор, використовуючи ці наближення як моделі та їх формули першого порядку, залучає поведінку специфікації (його негативність) описаної програми. Рештюння SMT-LIB формули розглядається як комбінована формула для того, щоб її «обмежувати» та «розв’язувати». Форма «розв’язку» може використовуватись для ідентифікації логічних помилок (безпеки) Java-програм на базі Android. Властивості безпеки Android представлено як обмежувальні аналітичні цілі, щоб показати важливість цих обмежень.

Ключові слова: Android-безпека; Java-програма; програмний код; статичний аналіз; статична теорія моделювання

Received 25 November 2014.
В. М. Мельник1, К. В. Мельник2, О. К. Жигаревич3. Статистический анализ исходного кода смоделированный для java-программ содержащих приложения с безопасностью android

Проведено сопоставление методов статического анализа с моделью дедуктивной проверки и использования решений теории статической модели (ТСМ) для создания основания, которая, учитывая аспект анализа исходного кода, автоматически создается с помощью анализатора, выводящего конечную информацию об этом аспекте. Анализатор генерируется путем перевода программы для сбора семантики с целью изложения формул в первом приближении на основании нескольких представленных теорий. Так как программа делает импорт пакетов и использует классовые методы этих пакетов, она импортирует семантику вызовов API в приближении первого порядка. Анализатор, используя эти приближения как модели та их формулы первого порядка, включает поведение спецификации (его отрицательность) описанной программы. Решения SMT-LIB формул рассматриваются как скомбинирована формула для того, чтобы их «ограничивать» и «решать». Форма «решения» может использоваться для идентификации логических ошибок (безопасности) Java-программ на базе Android. Свойства безопасности Android представлены как ограничивающие аналитические цели, чтобы показать важность этих ограничений.

Ключевые слова: Android-безопасность; Java-программа; программный код; статический анализ; статическая теория моделирования

Computer Engineering Department, Lutsk National Technical University, Lutsk, Ukraine.
Research area: computing, programming and sockets.
Publications: 34.
E-mail: melnyk_v_m@yahoo.com

Melnyk Kateryna. PhD in technique. Assistant professor,
Computer Engineering Department, Lutsk National Technical University, Lutsk, Ukraine.
Research area: computational intelligence systems.
Publications: 29.
E-mail: ekaterinamelnik@gmail.com

Zhyharevych Oksana. Assistant professor.
Computer Engineering Department, Lutsk National Technical University, Lutsk, Ukraine.
Education: Lutsk Biotechnical Institute of International Science and Technology, Lutsk, Ukraine.
Research area: computer programming, simulation-based semantics.
Publications: 20.
E-mail: oz_lutsk@mail.ru