Apposcopy: Semantics-Based Detection of Android Malware Through Static Analysis

Yu Feng, Saswat Anand, Isil Dillig, Alex Aiken

University of Texas at Austin    Stanford University
Motivation—Our setting

Why Android?
Motivation-Our setting

Why Android?

Prevalence of Mobile Malware Infections, 2013

2014 Threat Landscape Report, Fortinet
Motivation-Our setting

Why Android?

Prevalence of Mobile Malware Infections, 2013

2013 Android Mobile Malware Samples

2014 Threat Landscape Report, Fortinet
Motivation-damage

50% of Android malware are trying to steal your personal data

http://securelist.com/analysis/quarterly-malware-reports/37163/it-threat-evolution-q2-2013/
Motivation-damage

50% of Android malware are trying to steal your personal data

http://securelist.com/analysis/quarterly-malware-reports/37163/it-threat-evolution-q2-2013/
50% of Android malware are trying to steal your personal data

http://securelist.com/analysis/quarterly-malware-reports/37163/it-threat-evolution-q2-2013/
Motivation-damage

50% of Android malware are trying to steal your personal data

http://securelist.com/analysis/quarterly-malware-reports/37163/it-threat-evolution-q2-2013/
Motivation-Taint analysis

Existing approach 1

Enck et al. 2012
Motivation-Taint analysis

Existing approach 1

Enck et al. 2012
Motivation-Taint analysis

Source
- Call Log
- Contact List
- Credit Card

Propagation

Existing approach 1

Enck et al. 2012
Motivation - Taint analysis

Source
- Call Log
- Contact List
- Credit Card

Propagation

Sink
- Internet
- I/O
- SMS

Existing approach 1

Enck et al. 2012
Motivation-Taint analysis

Source
- Call Log
- Contact List
- Credit Card

Propagation
- ...
- ...
- ...

Sink
- Internet
- I/O
- SMS

Existing approach 1

Enck et al. 2012
Motivation-Taint analysis

Source
- Call Log
- Contact List
- Credit Card

Propagation
- ...
- ...

Sink
- Internet
- I/O
- SMS

Existing approach 1

Enck et al. 2012
Motivation-Taint analysis

Source
- Call Log
- Contact List
- Credit Card

Propagation
- ...

Sink
- Internet
- I/O
- SMS

Existing approach 1

Enck et al. 2012
Motivation-Taint analysis

Source
- Call Log
- Contact List
- Credit Card

Propagation
- ... (interconnected)

Sink
- Internet
- I/O
- SMS

Existing approach 1

Enck et al. 2012
Motivation - Taint analysis

Source

Call Log → Contact List → Credit Card

Propagation

Call Log → Internet

Sink

Internet → I/O → SMS

Existing approach 1

Enck et al. 2012
Motivation-Taint analysis

Source
- Call Log
- Contact List
- Credit Card

Propagation
- ...
- ...
- ...
- ...
- ...
- ...

Sink
- Internet
- I/O
- SMS

Existing approach 1
Enck et al. 2012
Motivation - Taint analysis
Motivation-Taint analysis

Pros: Exposing apps that leak sensitive data in a sound way.
Motivation-Taint analysis

Pros: Exposing apps that leak sensitive data in a sound way.

Cons: Block legitimate apps
Motivation-Signature-based

- Signature-based malware detectors
- Specific sequence of instructions
- Certain string values, e.g., method or variable names (e.g., ‘zjService’)

Existing approach 2

Griffin et al. 2009
Motivation-Signature-based
Motivation-Signature-based

Pros: Represent a corpus of malware through finite signatures
Motivation-Signature-based

Pros: Represent a corpus of malware through finite signatures

Cons: Update signature frequently; Obfuscation by bytecode transformation
Goal-Putting two together?

Taint analysis

Signature-based
Goal-Putting two together?

Taint analysis

Signature-based
Goal-Putting two together?

Taint analysis  Signature-based

Fewer false positives
Goal-Putting two together?

Taint analysis
Signature-based

Fewer false positives
Resist common obfuscation.
Goal

- A high-level signature language for describing semantic characteristics of Android malware families. Such as:
  - Control-flow properties
  - Data-flow properties
Goal

• Powerful static analyses for deciding if a given app matches signature of a malware family.
  • Control-flow properties matching: Inter-Component Call Graph Construction.
  • Data-flow properties matching: Taint analysis
Android Background

Let’s take a detour before we go through the technical details...
Android Background

- Android Components
  - Activity
  - Service
  - Broadcast Receiver
  - Content Provider
Android Background

- Android Components
  - Activity
  - Service
  - Broadcast Receiver
  - Content Provider
Android Background

- Android Components
  - Activity
  - Service
  - Broadcast Receiver
  - Content Provider
Android Background

- Android Components
  - Activity
  - Service
  - Broadcast Receiver
  - Content Provider
Android Background

- Android Components
  - Activity
  - Service
  - Broadcast Receiver
  - Content Provider
Android Background

Inter-Component Communication

Component A

Intent
action
data
category

Intent Filter
action
data
category

Component B
Android Background

An example of Inter-Component Communication

http://www.edureka.co/blog/android-interview-questions-answers-for-beginners/
Key Ideas
Key Ideas

- Control-flow properties
- Data-flow properties
Key Ideas

Can Activity A launch Service B?

- Control-flow properties
- Data-flow properties
Key Ideas

Control-flow properties

Can Activity A launch Service B?

Data-flow properties

Can Receiver C send my credit card number through Internet?
Key Ideas

Can Activity A launch Service B?

Can Receiver C send my credit card number through Internet?

Our signature should reflect the inter-component communication!
System Overview

Apposcopy
System Overview

Apposcopy

Malware Spec
System Overview

Apposcopy

Signature Language

Malware Spec
System Overview

Apposcopy

Signature Language

Malware Spec
Our Approach
Our Approach

An instance of GoldDream malware
SHA256: 3e72cc3c0db3513a29ff53e27726fb9277c7d2f13661cf0dfca8eb34dc690074
Our Approach

An instance of GoldDream malware
SHA256: 3e72cc3c0db3513a29ff53e27726fb9277c7d2f13661cf0dfca8eb34dc690074

GoldDream malware specification:

“It will register a receiver so that it will be notified for certain system events such as when a SMS message is received, or when there is an incoming/outgoing phone call.”

“Upon these events, the malware launches a background service without user's knowledge.”

“GoldDream will collect the IMSI and IMEI of the device.”

“Transport the collected information to a remote server.”

----- GoldDream malware report: http://www.csc.ncsu.edu/faculty/jiang/GoldDream/
Our Approach

GoldDream Signature

1. GDEvent(SMS RECEIVED).
2. GDEvent(NEW OUTGOING CALL).
3. GoldDream :- receiver(r),
   icc(SYSTEM, r, e, _), GDEvent(e),
   service(s), icc*(r, s),
   flow(s, Deviceld, s, Internet),
   flow(s, SubscriberId, s, Internet).
Our Approach

GoldDream Signature

1. GDEvent(SMS_RECEIVED).
2. GDEvent(NEW_OUTGOING).
3. GoldDream :- receiver(r),
4. icc(SYSTEM, r, e, _), GDEvent(e),
5. service(s), icc*(r, s),
6. flow(s, DevicedId, s, Internet),
7. flow(s, SubscriberId, s, Internet).
Our Approach

GoldDream Signature

1. GDEvent(SMS_RECEIEVED).
2. GDEvent(NEW_OUTGOING).
3. GoldDream :- receiver(r),
   4. icc(SYSTEM, r, e, _), GDEvent(e),
   5. service(s), icc*(r, s),
   6. flow(s, Deviceld, s, Internet),
   7. flow(s, Subscriberld, s, Internet).
Our Approach

GoldDream Signature

1. GDEvent(SMS RECEIVED).
2. GDEvent(NEW OUTGOING).
3. GoldDream :- receiver(r),
   4. icc(SYSTEM, r, e, _), GDEvent(e),
   5. service(s), icc*(r, s),
   6. flow(s, Deviceld, s, Internet),
   7. flow(s, SubscriberId, s, Internet).
Our Approach

Signature matching procedure:
Our Approach

Signature matching procedure:
Our Approach

Signature matching procedure:

Malware Signature
Our Approach

Signature matching procedure:

Malware Signature
Our Approach

Signature matching procedure:

Control-flow Properties

Malware Signature
Our Approach

Signature matching procedure:

Control-flow Properties

Data-flow Properties

Malware Signature
Our Approach

Signature matching procedure:
Our Approach

• Data-flow properties matching through Static taint analysis.
  • Credit card number flows to Internet
  • Device Id flows through SMS
  • ...

Our Approach

Output of Taint Analysis:

com.sjgo.client.zjService:
- $SimSerialNumber -> !INTERNET
- $DeviceId -> !INTERNET
- $SubscriberId -> !INTERNET
- $DeviceId -> !sendTextMessage
- $SubscriberId -> !sendTextMessage

cxboy.android.game.fivelink.FiveLink:
- $ID -> !INTERNET
- $MODEL -> !INTERNET

net.youmi.android.AdActivity:
- $DeviceId -> !WebView
- $ExternalStorage -> !WebView
Our Approach

Output of Taint Analysis:

com.sjgo.client.zjService:
- $SimSerialNumber -> !INTERNET
- $DeviceId -> !INTERNET
- $SubscriberId -> !INTERNET
- $DeviceId -> !sendTextMessage
- $SubscriberId -> !sendTextMessage

cxboy.android.game.fiveInk.FiveLink:
- $ID -> !INTERNET
- $MODEL -> !INTERNET

net.youmi.android.AdActivity:
- $DeviceId -> !WebView
- $ExternalStorage -> !WebView
Our Approach

Output of Taint Analysis:

com.sjgo.client.zjService:
$SimSerialNumber -> !INTERNET
$DeviceId -> !INTERNET
$SubscriberId -> !INTERNET
$DeviceId -> !sendTextMessage
$SubscriberId -> !sendTextMessage

cxboy.android.game.fiveInk.FiveLink:
$ID -> !INTERNET
$MODEL -> !INTERNET

net.youmi.android.AdActivity:
$DeviceId -> !WebView
$ExternalStorage -> !WebView
Our Approach

Signature matching procedure:

- Control-flow Properties
- Data-flow Properties

= Malware Signature
Our Approach

Control-flow properties matching through ICCG Construction.
ICCG: Inter-Component Call Graph, an high-level abstraction for Android application

Intent analysis: Resolve the target components
Our Approach

Partial ICCG for current example
Our Approach

Partial ICCG for current example
Our Approach

Partial ICCG for current example

1. GDEvent(SMS RECEIVED).
2. GDEvent(NEW OUTGOING CALL).
3. GoldDream :- receiver(r),
4. icc(SYSTEM, r, e, _), GDEvent(e),
5. service(s), icc*(r, s),
6. flow(s, Deviceld, s, Internet),
7. flow(s, SubscriberId, s, Internet).
Our Approach

Signature matching procedure:

Control-flow Properties

Data-flow Properties

Malware Signature
Implementation

Apposcopy
Implementation

Apposcopy

- Pointer Analysis
- Call Graph
- Taint Analysis
- Intent Analysis
- ICCG Construction
- Build-in Predicates
Implementation

Apposcopy

Signature Language

Malware Spec
Implementation

Apposcopy

Signature Language

Malware Spec
Implementation

- Pointer Analysis
- Taint Analysis
- ICCG Construction
- Build-in Predicates
- Call Graph
- Intent Analysis
- Malware Spec
- Signature Language
- Apposcopy

Match or not
Experiments

• Our experiments are trying to answer three questions:

  • **RQ1**: Can Apposcopy pinpoint malware?

  • Malware from Android Malware Genome Project.

http://www.malgenomeproject.org/
## Experiments

**Malware in Android Genome project**

<table>
<thead>
<tr>
<th>Malware Family</th>
<th>#Samples</th>
<th>FN</th>
<th>FP</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>DroidKungFu</td>
<td>444</td>
<td>15</td>
<td>0</td>
<td>96.6%</td>
</tr>
<tr>
<td>AnserverBot</td>
<td>184</td>
<td>2</td>
<td>0</td>
<td>98.9%</td>
</tr>
<tr>
<td>BaseBridge</td>
<td>121</td>
<td>75</td>
<td>0</td>
<td>38%</td>
</tr>
<tr>
<td>Geinimi</td>
<td>68</td>
<td>2</td>
<td>2</td>
<td>97.1%</td>
</tr>
<tr>
<td>DroidDreamLight</td>
<td>46</td>
<td>0</td>
<td>0</td>
<td>100%</td>
</tr>
<tr>
<td>GoldDream</td>
<td>46</td>
<td>1</td>
<td>0</td>
<td>97.8%</td>
</tr>
<tr>
<td>Pjapps</td>
<td>43</td>
<td>7</td>
<td>0</td>
<td>83.7%</td>
</tr>
<tr>
<td>ADRD</td>
<td>22</td>
<td>0</td>
<td>0</td>
<td>100%</td>
</tr>
<tr>
<td>jSMSHider</td>
<td>16</td>
<td>0</td>
<td>0</td>
<td>100%</td>
</tr>
<tr>
<td>DroidDream</td>
<td>14</td>
<td>1</td>
<td>0</td>
<td>92.9%</td>
</tr>
<tr>
<td>Bgserv</td>
<td>9</td>
<td>0</td>
<td>0</td>
<td>100%</td>
</tr>
<tr>
<td>BeanBot</td>
<td>8</td>
<td>0</td>
<td>0</td>
<td>100%</td>
</tr>
<tr>
<td>GingerMaster</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>100%</td>
</tr>
<tr>
<td>CoinPirate</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>100%</td>
</tr>
<tr>
<td>DroidCoupon</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>100%</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>1027</strong></td>
<td><strong>103</strong></td>
<td><strong>2</strong></td>
<td><strong>90%</strong></td>
</tr>
</tbody>
</table>
## Experiments

### Malware in Android Genome project

<table>
<thead>
<tr>
<th>Malware Family</th>
<th>#Samples</th>
<th>FN</th>
<th>FP</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>DroidKungFu</td>
<td>444</td>
<td>15</td>
<td>0</td>
<td>96.6%</td>
</tr>
<tr>
<td>AnserverBot</td>
<td>184</td>
<td>2</td>
<td>0</td>
<td>98.9%</td>
</tr>
<tr>
<td>BaseBridge</td>
<td>121</td>
<td>75</td>
<td>0</td>
<td>38%</td>
</tr>
<tr>
<td>Geinimi</td>
<td>68</td>
<td>2</td>
<td>2</td>
<td>97.1%</td>
</tr>
<tr>
<td>DroidDreamLight</td>
<td>46</td>
<td>0</td>
<td>0</td>
<td>100%</td>
</tr>
<tr>
<td>GoldDream</td>
<td>46</td>
<td>1</td>
<td>0</td>
<td>97.8%</td>
</tr>
<tr>
<td>Pjapps</td>
<td>43</td>
<td>7</td>
<td>0</td>
<td>83.7%</td>
</tr>
<tr>
<td>ADRD</td>
<td>22</td>
<td>0</td>
<td>0</td>
<td>100%</td>
</tr>
<tr>
<td>jSMSHider</td>
<td>16</td>
<td>0</td>
<td>0</td>
<td>100%</td>
</tr>
<tr>
<td>DroidDream</td>
<td>14</td>
<td>1</td>
<td>0</td>
<td>92.9%</td>
</tr>
<tr>
<td>Bgserv</td>
<td>9</td>
<td>0</td>
<td>0</td>
<td>100%</td>
</tr>
<tr>
<td>BeanBot</td>
<td>8</td>
<td>0</td>
<td>0</td>
<td>100%</td>
</tr>
<tr>
<td>GingerMaster</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>100%</td>
</tr>
<tr>
<td>CoinPirate</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>100%</td>
</tr>
<tr>
<td>DroidCoupon</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>100%</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>1027</strong></td>
<td><strong>103</strong></td>
<td><strong>2</strong></td>
<td><strong>90%</strong></td>
</tr>
</tbody>
</table>
Experiments

• Our experiments are trying to answer three questions:

  • RQ2: Does Apposcopy report a lot of false positives?

  • Benign apps from Google play.
11,215 'benign' apps from Google Play

11,199 Benign
16 Malicious
Experiments

• Our experiments are trying to answer three questions:

• RQ3: Is Apposcopy resistant to common obfuscations?

• Obfuscated malware
Experiments

Obfuscate existing malware using the ProGuard tool.

Comparison with other tools on obfuscated malware:

- AVG
- Symantec
- ESET
- Dr. Web
- Kaspersky
- Trend Micro
- McAfee
- Apposcopy
Experiments

Obfuscate existing malware using the ProGuard tool.

Comparison with other tools on obfuscated malware:

- AVG
- Symantec
- ESET
- Dr. Web
- Kaspersky
- Trend Micro
- McAfee
- Apposcopy
Summary

• Apposcopy: a new static analysis approach for detecting Android malware
Summary

- Apposcopy: a new static analysis approach for detecting Android malware
- Perform deep static analysis and use a high-level representation (ICCG) to extract both data-flow and control-flow properties.
Related work


