Agenda

● Personal Background
● Mobile Malware
  ○ Geographical Mobile Markets
● Intro to Android and APK package
● MEDS (Malware Evolution Discovery System)
  ○ Creation Phylogenetic (Lineage) Trees
  ○ Predicting Generativeness
● Summary/Future Work
● Antonio Cesar Vargas
  ○ **M.Sc**, John Jay College of Criminal Justice
    ■ Digital Forensics and Cybersecurity
  ○ **B.Sc**, Queens College
    ■ Computer Science
  ○ Interests
    ■ Python
    ■ Malware
    ■ Memory Forensics
2014 10th Birthday Mobile Malware
Mobile Malware

- **Android**
  - Prefered Malware Creators
  - Dominant Mobile Platform

- **Threats**
  - Ransomware, botnet, personal/financial information theft
Tenth Anniversary Mobile Malware

- **2004**
  - Symbian OS
    - Cabir, Trojan.Mos, Skulls
- **2006**
  - Cross Platform Mobile Malware (Symbian and Blackberry)
    - Redbrowser, FlexiSpy
Tenth Anniversary Mobile Malware

• 2010
  ○ Cross Platform Mobile Malware (Symbian and Android)
    ■ ZeusMitmo
• 2011
  ○ Android
    ■ Geinimi, RootCager
Why are mobile devices attractive?

**Surveillance**
- Audio
- Camera
- Call logs
- Location
- SMS messages

**Impersonation**
- SMS redirection
- Sending email messages
- Posting to social media

**Financial**
- Sending premium rate SMS messages
- Stealing transaction authentication numbers (TANs)
- Extortion via ransomware
- Fake antivirus
- Making expensive calls

**Botnet activity**
- Launching DDoS attacks
- Click fraud
- Sending premium rate SMS messages

**Data theft**
- Account details
- Contacts
- Call logs
- Phone number
- Stealing data via app vulnerabilities
- Stealing international mobile equipment identity number (IMEI)

**Android malware instances seen by SophosLabs**

**Smartphones lost every minute in the U.S.**

**Average cost of a U.S. data breach in 2012**

**Price charged by the Android Defender ransomware**

---

1. Source: 2013 Cost of Data Breach Study, Ponemon Institute
Life Companion
Dependence Mobile Devices

- Business Purposes
- Everyday Needs
Something for everyone!!

● Cybercriminals
  ○ Data Theft
  ○ Botnet Activity
  ○ Personal/Financial Information Theft

● Government Entities
  ○ Surveillance
  ○ Tactical Operations
Most Malware is not ‘new’

- Repacks
- Incremental Updates
- Business Model
  - Malware Headquarters
    - Startup Business
  - Governmental Intrusion and Remote Monitoring Solutions
    - Gamma International--FinFisher Suite
Malware HQ (Industrial Business)

- Dragon Lady Investigation
  - Lookout
Dragon Lady Findings

- **Android Malware HQ**
  - Startup
  - Organized
- **Constant Releases** of Malware Families
  - Agile Approach
- **Affiliate Marketers**
  - Distribution
  - Customization
Why is this important?

- **Mobile Malware Visibility**
  - Specific Geographical Regions (China and Russia)
  - Third Party App Stores
New Internet Citizens

- Experience the Internet through a mobile phone
- Third World Countries
MOBILE THREATS
MALWARE, ADWARE, CHARGEWARE
N. American Encounter Rates in 2013.

The United States and Canada have comparable threat encounter rates while mobile users in Mexico have an elevated risk of encountering adware.
Malware Preferred Mobile Platform

- Android: 98.05%
- J2ME: 1.55%
- SymbOS: 0.27%
- Sonstige: 0.13%

Mobile malware distribution by platform
Malware Prefered Mobile Platform

- Android: 98.05%
- J2ME: 1.55%
- SymbOS: 0.27%
- Sonstige: 0.13%

Mobile malware distribution by platform
Android Architecture
APK Package
Android Applications (APK)

- Zip Format Archive
  - AndroidManifest.xml
  - classes.dex
    - Java Code
    - Dalvik VM
  - Meta information
    - SSL Certificate (Self Signed)
  - Resources
## APK Internals

<table>
<thead>
<tr>
<th>Length</th>
<th>Date</th>
<th>Time</th>
<th>Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>1651</td>
<td>2014-02-05</td>
<td>16:37</td>
<td>res/drawable/ic_launcher.png</td>
</tr>
<tr>
<td>174398</td>
<td>2014-02-05</td>
<td>16:37</td>
<td>res/raw/debiancacerts.bks</td>
</tr>
<tr>
<td>966562</td>
<td>2012-10-26</td>
<td>08:10</td>
<td>res/raw/geoip.mp3</td>
</tr>
<tr>
<td>198652</td>
<td>2014-02-05</td>
<td>16:37</td>
<td>res/raw/iptables</td>
</tr>
<tr>
<td>130204</td>
<td>2014-02-05</td>
<td>16:37</td>
<td>res/raw/iptables_g1</td>
</tr>
<tr>
<td>134308</td>
<td>2014-02-05</td>
<td>16:37</td>
<td>res/raw/iptables_nl</td>
</tr>
<tr>
<td>963309</td>
<td>2014-02-05</td>
<td>16:37</td>
<td>res/raw/obfsproxy</td>
</tr>
<tr>
<td>346549</td>
<td>2014-02-05</td>
<td>16:37</td>
<td>res/raw/privoxy</td>
</tr>
<tr>
<td>883</td>
<td>2014-02-05</td>
<td>16:37</td>
<td>res/raw/privoxy_config</td>
</tr>
<tr>
<td>2074073</td>
<td>2012-10-26</td>
<td>08:10</td>
<td>res/raw/tor.mp3</td>
</tr>
<tr>
<td>324</td>
<td>2014-02-05</td>
<td>16:37</td>
<td>res/raw/torrc</td>
</tr>
<tr>
<td>291</td>
<td>2014-02-05</td>
<td>16:37</td>
<td>res/raw/torrcollector</td>
</tr>
<tr>
<td>480</td>
<td>2014-02-05</td>
<td>16:37</td>
<td>res/xml/policies.xml</td>
</tr>
<tr>
<td>6104</td>
<td>2014-02-05</td>
<td>16:37</td>
<td>AndroidManifest.xml</td>
</tr>
<tr>
<td>5648</td>
<td>2014-02-05</td>
<td>16:37</td>
<td>resources.arsc</td>
</tr>
<tr>
<td>1651</td>
<td>2014-02-05</td>
<td>16:37</td>
<td>res/drawable-hdpi/ic_launcher.png</td>
</tr>
<tr>
<td>887</td>
<td>2014-02-05</td>
<td>16:37</td>
<td>res/drawable-ldpi/ic_launcher.png</td>
</tr>
<tr>
<td>1133</td>
<td>2014-02-05</td>
<td>16:37</td>
<td>res/drawable-mdpi/ic_launcher.png</td>
</tr>
<tr>
<td>2597</td>
<td>2014-02-05</td>
<td>16:37</td>
<td>res/drawable-xdpi/ic_launcher.png</td>
</tr>
<tr>
<td>2917032</td>
<td>2014-02-05</td>
<td>16:37</td>
<td>classes.dex</td>
</tr>
<tr>
<td>35890</td>
<td>2014-02-05</td>
<td>16:37</td>
<td>info/guardianproject/onionkit/trust/StrongTrustManager.java.underreview.txt</td>
</tr>
<tr>
<td>4224</td>
<td>2014-02-05</td>
<td>16:37</td>
<td>ch/boyer/httpclientandroidlib/impl/conn/tscmc/doc-files/tscmc-structure.png</td>
</tr>
<tr>
<td>1795</td>
<td>2014-02-05</td>
<td>16:37</td>
<td>META-INF/MANIFEST.MF</td>
</tr>
<tr>
<td>1848</td>
<td>2014-02-05</td>
<td>16:37</td>
<td>META-INF/CERT.SF</td>
</tr>
<tr>
<td>1203</td>
<td>2014-02-05</td>
<td>16:37</td>
<td>META-INF/CERT.RSA</td>
</tr>
</tbody>
</table>

**Configuration Setup**

- **Executable Code**

- **25 files**
AndroidManifest.xml

- Content Provider (Provides a seamless interface to store and retrieve data in the android app)
- Broadcast Receiver (Used for listening to messages/events in the android system)
- Services (Long running operations in background without UI)
- Activity (A single screen where user can interact)
- Intent (An abstraction/message describing the operation to be performed)
- Package name, min-sdk, permissions etc

AndroidManifest.xml (The Glue)
MEDS (Malware Evolution)

- Malware Evolution
  - Android Malware
  - Similarity Percentage
    - Approximate Matching
      - Creation Approximation
      - Phylogenetetic/Lineage Tree
Similarity of object?

- Very good at **equality**
  - Hashing (Fingerprint)
- Similarity of two objects?
Approximate Matching

- NIST Special Publication **800-168**
- **Approximate Matching**
  - **Bytewise** (Sequence of bytes)
    - SDHASH
  - **Syntactic** (Internal Structures)
    - AndroidManifest.xml
  - **Semantic** (Contextual Attributes)
Phylogenetic (Lineage) Tree

- **Metadata**
  - Creation Date

- **Approximate Matching Value**
  - AndroidManifest.xml (Syntactic)
  - Dex Files (Bytewise)
MEDS (Discovery System)

● Regression Analysis
  ○ Feature Extraction
    ■ Number of Dangerous Permissions
  ○ Linear and Logistic Regression
    ■ Generativeness
      ● Statistics about what malware will influenced the creation of future malware
Python Implementation

● **Graphs**
  ○ **Pygraphviz**
    ■ Modeling
    ■ Dot Files for visualization

● **SDHash (Bytewise)**
  ○ Python SWIG Binding to C++ library

● **Edit Distance (Syntactic)**
  ○ AndroidManifest.xml
Phylogenetic (Lineage) Tree

● Not A New Idea
  ○ DARPA *(43 Million)* Cyber Genome Project, 2010
    ■ Lockheed Martin
    ■ Invicea Labs *(Cynomix.org)*
    ■ BAE Systems
    ■ Raytheon BBN Technologies
Invincea Labs (Cynomix)
Generative Malware

- Generativeness
  - Predict Future Malware Trends
  - Active Malware
    - Features
    - Vulnerabilities
    - Baseline
Present Samples Scenario

Figure 1.1: 5 Malware Samples May 14-August 2, 2010
Malware Evolution

Figure 1.2: 10 Malware Samples Before and After August 2, 2010
Regression Analysis

- **Malware Features**
  - Number of Dangerous Permissions
  - Number of Receivers

- **Phylogenetic Tree Features**
  - Approximate Matching Value (to parent)
  - Age in second from parent
  - Age in second of the latest child
Dangerous Permissions

Figure 4.2: Number of dangerous permissions and actual generativeness values.
Number of Children

Figure 4.5: Number of Children actual generativeness

$r = 0.074$
$m = 3e-04$
$b = 0.0037$
$n = 250$
Python Implementation

- Regression Analysis Algorithms
  - Octave
    - Oct2py
  - Visualization
    - rpy2
Summary

● MEDS
  ○ **Phylogenetic** Tree Malware
    ■ Evolution Malware
      ● Rapid Development of *detection* and *eradication*
  ○ **Generative** Malware
    ■ Detect *Promiscuous* Malware
    ■ Pro-Active Malware Outbreaks
    ■ Data Science Problem
Summary

- Python very flexible
  - Phylogenetic Malware Tree
  - Machine Learning Algorithms Integration
- **Generativeness (Data Science Problem)**
  - Bias
    - Further research
    - Choose Different Features
Future Work

- **Regression Analysis** improvements
  - scikit-learn
  - Bias Problem
    - Choose Different Features
    - Different Malware Sets
- Better **Visualization**
- More/Different **Malware Samples**
Acknowledgements

● **virusshare.com**
  ○ Malware Samples

● **Candice Quates**
  ○ SDHash Core Developer

● **CUNY**
  ○ Prof. Bilal Khan
    ■ http://www.systemic-inquiry.com
  ○ Jeremy D. Seideman
More Acknowledgments

- **AndroGuard**
  - Anthony Desnos
- **Silvio Cesare**
  - Software Similarity
Thank You !!!

- Questions
- Comments
- Clarifications
@vargasces