SESSION ID: BR-W04

You are what you click: Using Decoys to Identify Mobile Device Attackers

Dr. Salvatore J. Stolfo
Professor
Columbia University
Intrusion Detection Lab
New York, NY

Joel Peterson
Researcher, Columbia University & Software Systems Researcher
Allure Security Technology
New York, NY
The old ways don’t work….

It’s not what you know (twice), it’s what you do….
Age of Collaboration
“Anywhereization”
Rise in BYOD

- 2 billion smartphone users projected for 2015
- 61% of employees use smartphones for work
Enterprise is borderless...and vulnerable

- Devices carry sensitive corporate data:
  - 59% of employees using BYOD haven’t told their employers

- Lost or stolen:
  - 113 smart phones lost every minute
  - 1 laptop is stolen every 53 seconds

- Mobile security fails:
  - 34% of consumers fail to activate security mechanisms on their mobile devices
Security should be designed for the people who use it

- Easy to understand controls
- Transparent to the user
- Seamless and continuous authentication
Are you you?

- With patented machine-learning technology, RUU learns how you use your device and creates a personalized behavioral profile that continuously and seamlessly authenticates.

- If unusual behavior is detected, it’s prompted to ask, “Are you you?”
What are decoys?

- Enticing, believable but bogus data, documents, files, and other types of fake but realistic media
- Touch a decoy and send a beacon alert signal
What’s the difference?
Document is *beaconized*
Touch a decoy file, a data loss alert is emailed.
Enticing decoy files in the cloud, too!

One is real, the others aren’t – Can YOU tell?
The Hypothesis

- We all search uniquely on our own machines....that is a user biometric captured by a behavior model computed by a machine learning algorithm.

- Decoys are a powerful tool to detect intruders who do not know the real content of a target victim’s file system.

The two together detect masqueraders and provide accurate active and continuous authentication.
(Sidebar: Decoys can also be used to detect…)

- Hackers who hijack sessions from other legitimate users
- Embedded APT actors whose malware behaves abnormally
- But, let’s return our attention to Active and Continuous Authentication of users…
Phase 1
DARPA Active Authentication - Desktop
DARPA Phase 1 Goals

Solution: Active Authentication

- An open solution that provides **meaningful** and **continual** authentication to DoD’s computer systems leveraging that which makes up you.

Continuous authentication using:
- Multiple modalities in a rotating fashion
- Multiple authentications initiated each minute
- Open architecture to bring in future modalities

**New Authentication Modalities**

<table>
<thead>
<tr>
<th></th>
<th>Phase 1</th>
<th>Phase 2</th>
<th>Phase 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maximum False Rejections after five (5) scans</td>
<td>1/week</td>
<td>1/month</td>
<td>1/month</td>
</tr>
<tr>
<td>True Positive Rate for each scan</td>
<td>80%</td>
<td>80%</td>
<td>85%</td>
</tr>
<tr>
<td>Usability of modality within the population of DoD personnel</td>
<td>90%</td>
<td>90%</td>
<td>95%</td>
</tr>
</tbody>
</table>
RUU Baseline Architecture

- Alerts/Alarms
- User Challenges
- Continuous Model Testing
- User Behavior Model
- User Touches of Files and Decoy Files
- User Search of File System
- Sensing OS-level Events
- USER COMMANDS AND ACTIVITY

Authenticating User

Learning User Search Behavior

User Actions
Initial baseline scientific user study of accuracy of modeling user behavior

- Model baseline Volunteer Human Subject behavior; detect deviations from normal use. Generative model: inference, prediction, clustering, sampling, etc.

- Behavior biometrics: set of measurements on interactions between the Volunteer Human Subject and the system.

- Biometrics measurements based on OS events caused by Volunteer Human Subject action:
  - Process creation, deletion, manipulation.
  - File creation, deletion, renaming, etc.
  - Process window touches.
  - Registry key creation, manipulation, deletion.

- Four minute sliding window of measurement used.

- RUU1 dataset: 18 Volunteer Human Subjects at Columbia University, measured over the course of five weeks. Captured in 2011. RUU2 and RUU200 datasets delivered Sept 2013.
## Fisher Linear Discriminatant Analysis

<table>
<thead>
<tr>
<th>Feature</th>
<th>FLD Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of unique processes</td>
<td>0.0359</td>
</tr>
<tr>
<td>Number of delete key actions</td>
<td>0.0018</td>
</tr>
<tr>
<td>Number of processes created</td>
<td>0.0015</td>
</tr>
<tr>
<td>Number of files touched</td>
<td>0.0013</td>
</tr>
<tr>
<td>Number of registry flush key actions</td>
<td>0.0012</td>
</tr>
<tr>
<td>Number of user touches</td>
<td>0.0011</td>
</tr>
<tr>
<td>Number of registry key queries</td>
<td>0.0011</td>
</tr>
<tr>
<td>Number of registry value queries</td>
<td>0.0010</td>
</tr>
<tr>
<td>Number of processes destroyed</td>
<td>0.0010</td>
</tr>
<tr>
<td>Number of open key actions</td>
<td>0.0010</td>
</tr>
<tr>
<td>Number of manual search actions</td>
<td>0.0009</td>
</tr>
</tbody>
</table>
Accuracy Improvements ... choose wisely

Accuracy over the initial RUU dataset. GMM model with Fisher features, improved accuracy and faster.
True Positive Rate increases with training

Accuracy improves over time. As more data is observed, the accuracy of the user’s model improves. And…
False Positive Rate Decays, too…

Maintaining and improving model performance over time is an important goal. Continuous learning methods work well.
Phase 2 – Desktop

Sensor Improvements

Automatic Decoy Placement

Larger formal user study to
detect masqueraders
RUU Host Sensors

Phase 1

- Volunteer Human Subject data acquisition uploaded to server for analytics and performance bundled with Decoy Document Distribution
  - Identify most discriminating features
  - Measure decoy touch behavior

Phase 2

- Volunteer Human Subject data acquisition on local host for automatic analysis and active authentication with mitigation strategy, also bundled with Decoy Document Distribution
  - Continuous learning
  - Automate Decoy Placement
  - Self-measurement of performance
  - Re-authentication strategies
RUU Sensor Identity Engine – 10 Dimensions works well
Learns User Search Behavior and OS-level Behavior Modeling

Gaussian Mixture Model. Trained automatically

Ten dimensions in real model

Multidimensional behavior measurements
RUU Decoy Distribution

How to deploy decoys in scale throughout an organization?

Manual placement
- Tedious
- Requires survey of Volunteer Human Subject habits

Alternative approach
- Distribute via an automated application
- Decoy Document Distributor (DDT)
Decoy Document Distributor (DDT)

- Fetches decoys from server
DDT Analyzes User’s file system

- Automatic deployment of decoys to strategic file locations
RUU Average Decoy Touch Rate of real user

- Most decoy touches are caused by initial deployment.
- Curiosity decays rapidly!
Masquerader Detection Accuracy with user models and decoys: Average ROC

RUU models vs. masquerader data. Influential factors: masqueraders used “smash and grab.” (They didn’t play games.)
Accuracy of detecting masqueraders over time is consistently high

Human subject activities are scaled as a percentage of capture progress (0%-100%). Average performance across all users.
Accuracy translated to detection latency – users emit observables at different rates

- Evaluation interval: 3 minutes
- Active authentication corresponds to Bernoulli trial: Probability that masquerader evades detection in 5 consecutive evaluations is less than 5%.
- Detection within 15 minutes with 95% confidence

### Experiment

- Overall Average Attacker Detection Across All Users
- 160 Users
- 1 week average capture period

### Experiment Results

- 95% detection accuracy at 1% false positive rate
- Constraint: 1 FP per 40 hour work week
  - Fifteen minutes until detection with 95% confidence

<table>
<thead>
<tr>
<th>Frequency</th>
<th>Total Samples</th>
<th>FP Req.</th>
<th>Acc.</th>
<th>Evals</th>
<th>TTD</th>
</tr>
</thead>
<tbody>
<tr>
<td>1m</td>
<td>2400</td>
<td>0.042%</td>
<td>49.55%</td>
<td>5</td>
<td>5m</td>
</tr>
<tr>
<td>2m</td>
<td>1200</td>
<td>0.083%</td>
<td>50.29%</td>
<td>5</td>
<td>10m</td>
</tr>
<tr>
<td>3m</td>
<td>800</td>
<td>0.125%</td>
<td>51.46%</td>
<td>5</td>
<td>15m</td>
</tr>
<tr>
<td>4m</td>
<td>600</td>
<td>0.167%</td>
<td>53.11%</td>
<td>4</td>
<td>16m</td>
</tr>
<tr>
<td>5m</td>
<td>480</td>
<td>0.208%</td>
<td>54.00%</td>
<td>4</td>
<td>20m</td>
</tr>
</tbody>
</table>

Time until detection (TTD) given evaluation frequency for a 40-hour work week.
Discussion – user model alone works, too

Masquerader ID and number of decoy touches by masquerader

- Masqueraders had higher than normal volumes of activity; exhibited “smash and grab“ behavior
- 10 decoys were distributed randomly on the test environment
- Nearly every masquerader touched several decoys, didn’t matter where they were placed
- Some touched no decoys, but were still detected

Masquerader detection even without decoy touches!
What about mitigation?

- De-authenticate and challenge the user to re-authenticate
  - This also provides an opportunity to update and improve the user model, ground truth is revealed
- Several possible re-authentication strategies, here’s one…

What do you do when you detect a masquerader?
RUU Secondary Authentication: When Desktop Locks

Secondary Authentication

- Time-based One-time Password Algorithm for secondary authentication (RFC 6238)

When installing RUU the user is prompted to enable secondary authentication.

Google Authentication running on an iPhone as the authentication agent.

When RUU locks and the user re-authenticates the secondary authentication is requested.
Monitoring and displaying RUU Sensor Performance: System Monitoring in scale for BYOB Management

Using internal pipeline, we can create monitors for different components and visualize them instantly.
mRUU – Mobile Phones
Decoy Apps
Decoy Clouds
mRUU Study

IRB-Approved User Studies

- January 2014: Pilot study
  - Preliminary Activity Collector
  - Users gathered from Accenture and Columbia University
  - Used to inform modeling approach

- July-August 2014: Full scale user study with 53 Accenture users
  - Fully developed activity collector
    - More efficient
    - Collection of auxiliary activity data
  - Used for final Identity Engine design and accuracy analysis
mRUU Study Results

Participant Upload Distributions
mRUU Update

- Implemented Identity Engine using adapted modeling technique which incorporates:
  - Activity hotspots
  - Temporal information
  - Location information
mRUU Location Based Modeling of User App Behavior

Location Based Sub-Modeling

GPS Hotspot Detection

Columbia University in the City of New York

RSA Conference 2015
Accurate Modeling of user app behavior

Modeling where you use Apps is very accurate
mRUU Study Results - classification accuracy with no FP

- Behavior eval every 2 min
- 4 hours total = 120 Evals/day
- Goal: 1 FP/day = 0.00833

<table>
<thead>
<tr>
<th>False-Positives per day</th>
<th>Percent of Foreign behavior identified</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>62%</td>
</tr>
<tr>
<td>2</td>
<td>70%</td>
</tr>
<tr>
<td>3</td>
<td>78%</td>
</tr>
<tr>
<td>4</td>
<td>80%</td>
</tr>
</tbody>
</table>
mRUU Study Results

Contact List Accuracy

When contacts list accessed
Ineffective – too few samples

<table>
<thead>
<tr>
<th>User</th>
<th>Days</th>
<th>Hours</th>
<th>Apps</th>
<th>Contact List</th>
<th>GPS</th>
<th>Phone</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>6</td>
<td>11:20</td>
<td>2,718,911</td>
<td>3,728</td>
<td>2,051</td>
<td>12</td>
</tr>
<tr>
<td>2</td>
<td>22</td>
<td>20:12</td>
<td>3,490,217</td>
<td>143</td>
<td>7,940</td>
<td>107</td>
</tr>
<tr>
<td>3</td>
<td>24</td>
<td>0:41</td>
<td>15,425,996</td>
<td>2,861</td>
<td>13,938</td>
<td>151</td>
</tr>
<tr>
<td>4</td>
<td>3</td>
<td>3:19</td>
<td>3,767,563</td>
<td>83</td>
<td>1,870</td>
<td>10</td>
</tr>
<tr>
<td>5</td>
<td>27</td>
<td>17:44</td>
<td>9,415,586</td>
<td>412</td>
<td>16,235</td>
<td>249</td>
</tr>
<tr>
<td>6</td>
<td>20</td>
<td>5:49</td>
<td>3,142,314</td>
<td>10</td>
<td>3,506</td>
<td>0</td>
</tr>
<tr>
<td>7</td>
<td>27</td>
<td>12:00</td>
<td>5,628,139</td>
<td>2,353</td>
<td>7,082</td>
<td>368</td>
</tr>
<tr>
<td>8</td>
<td>10</td>
<td>11:06</td>
<td>9,311,562</td>
<td>139</td>
<td>6,255</td>
<td>53</td>
</tr>
<tr>
<td>9</td>
<td>90</td>
<td>0:07</td>
<td>9,793,582</td>
<td>16,261</td>
<td>8,840</td>
<td>197</td>
</tr>
<tr>
<td>10</td>
<td>46</td>
<td>21:27</td>
<td>14,548,709</td>
<td>717</td>
<td>8,965</td>
<td>176</td>
</tr>
<tr>
<td>11</td>
<td>14</td>
<td>21:13</td>
<td>4,659,372</td>
<td>72</td>
<td>2,871</td>
<td>7</td>
</tr>
<tr>
<td>12</td>
<td>27</td>
<td>13:28</td>
<td>35,406,045</td>
<td>131</td>
<td>16,170</td>
<td>27</td>
</tr>
<tr>
<td>13</td>
<td>26</td>
<td>22:09</td>
<td>27,127,850</td>
<td>1,081</td>
<td>15,252</td>
<td>360</td>
</tr>
<tr>
<td>14</td>
<td>16</td>
<td>21:31</td>
<td>7,335,354</td>
<td>863</td>
<td>7,829</td>
<td>109</td>
</tr>
<tr>
<td>15</td>
<td>24</td>
<td>19:32</td>
<td>5,216,493</td>
<td>77</td>
<td>12,157</td>
<td>0</td>
</tr>
<tr>
<td>16</td>
<td>24</td>
<td>1:54</td>
<td>17,703,599</td>
<td>4,189</td>
<td>13,290</td>
<td>265</td>
</tr>
<tr>
<td>17</td>
<td>3</td>
<td>10:38</td>
<td>2,739,994</td>
<td>103</td>
<td>2,029</td>
<td>26</td>
</tr>
<tr>
<td>18</td>
<td>19</td>
<td>21:54</td>
<td>12,941,415</td>
<td>318</td>
<td>8,095</td>
<td>34</td>
</tr>
<tr>
<td>19</td>
<td>8</td>
<td>21:22</td>
<td>8,623,558</td>
<td>131</td>
<td>5,239</td>
<td>49</td>
</tr>
<tr>
<td>20</td>
<td>24</td>
<td>6:08</td>
<td>9,884,105</td>
<td>518</td>
<td>14,189</td>
<td>376</td>
</tr>
<tr>
<td>21</td>
<td>27</td>
<td>13:41</td>
<td>14,385,298</td>
<td>99</td>
<td>15,899</td>
<td>7</td>
</tr>
<tr>
<td>22</td>
<td>26</td>
<td>6:33</td>
<td>23,098,123</td>
<td>835</td>
<td>14,694</td>
<td>480</td>
</tr>
</tbody>
</table>

Contacts List Model Average ROC

Columbia University
IN THE CITY OF NEW YORK
mRUU User Study

Conclusions

- Users’ mobile application usage habits can successfully be used to derive behavioral biometric identifiers.
- The discriminative power of application usage patterns can be augmented using temporal and geographic information.
- Additional usage data (e.g., contacts, etc.) provides poor discriminative measurements.
Introducing Decoy Apps and Decoy Clouds

- The mobile RUU app automatically creates decoy apps from unused apps or downloads strategic decoy apps.

- Masqueraders are herded to pre-positioned decoy file system and decoy cloud services when they fail to re-authenticate.
Under the Hood

- Bad Behavior or Touching of decoy apps
  - de-authenticates the user
- Locks the device
- Captures a picture of the current user and records background ambient sound
- Sends an alert out of band to the user
- Re-authenticates by a second factor
  - Failure: Load Decoy Clouds and Decoy file system
  - Capture data on attacker
Decoy Apps are intuitive

- Authentic looking apps that hold fake but enticing information to the adversary
- An attacker does not know what is a Decoy App and what is a Real App
- They are simple to use
- They are simple to understand
- They do not increase resource use
Bloatware is turned into a Security Feature

Numerous Unused Apps available as decoys

![Bar chart showing the number of apps on smartphones in different countries, with labels for Australia, Canada, France, Germany, India, Italy, Japan, Korea, UK, and USA. The chart indicates average number of apps and includes data for the last 30 days and paid apps.]
Onboard unused apps become decoys or strategic decoy apps are installed
Who do you really bank with?

One is real, the others aren’t – Can YOU tell?
Which is your real email client?

One is real, the others aren’t – Can YOU tell?
Who is Your Cloud Provider?

One is real, the others aren’t – Can YOU tell?
Which is your real corporate VPN?

One is real, the others aren’t – Can YOU tell?
Which is your real Facebook?

One is real, the others aren’t – Can YOU tell?

Note: 2-D Passcode!
Recall, enticing decoy files in the cloud, too!

One is real, the others aren’t – Can YOU tell?
How do we do it?

Create Decoys
Any app is easily converted into a decoy

1. Unpack and disassemble .apk file.
2. Copy decoy functionality classes into assembly code folder.
3. Insert code into original classes to run decoy functionality.
4. Insert permissions for decoy functionality into manifest file.
5. Reassemble and repack .apk file.
Touch a decoy app, the phone locks and alerts.
...includes location, picture & recording
Sample Decoy App email alert

From: rapid-cr@email.com
Subject: Beacon Activated
Date: June 18, 2014 at 11:27:37 AM EDT
To: sasl@active-security.com

 Somebody at 0.12.10.0.253 has accessed your beaconized application. 
Open attachments for more details.
Have we offered sufficient gifts to the demo gods?

Demonstration
Alternative Unlock strategy, challenge the user, the phone knows your most recent behavior

With whom did you last chat?

- John Public
- Jane Doe
- Bill Jones
- None of the Above
What city did you last visit?

- Philadelphia
- New York
- Menlo Park
- None of the Above
Answer wrong again...

Configurable mitigation strategies

- Brick the phone
- Upload tracking data
- Unlock and Load a decoy file system
- Alert Security Personnel
DARPA Sponsorship

- DARPA ADAMS – Anomaly Detection at Multiple Scales
  - Insider threat

- DARPA Active Authentication
  - Masquerader/Impersonator threat

- $10 Million of research support, transitioned from Columbia University IDS Lab to Allure Security Technology
The Research Team

Sal Stolfo
Malek Ben Salem
Jon Voris
Yingbo Song
Joel Peterson
Shlomo Hershkop
Apply What You Have Learned Today

◆ Next week you should:
  ◆ Review corporate security policy for BYOD
  ◆ Identify the number of employee phones stolen or compromised
  ◆ Measure how many employees have no security controls on their devices

◆ In the first three months following this presentation you should:
  ◆ Measure employee mobile access to critical corporate infrastructure
  ◆ Evaluate corporate access and authentication controls
  ◆ Explore a deployment strategy for advanced mobile authentication

◆ Within six months you should:
  ◆ Identify and deploy solutions to protect employee mobile devices
Thank you…

◆ **Resources and contact**
  ◆ [www.cs.columbia.edu/ids](http://www.cs.columbia.edu/ids)
  ◆ [www.alluresecurity.com](http://www.alluresecurity.com)

