New Approaches on Malware-Detection on Mobile Devices

Michael Gröning INET RG - HAW Hamburg

November, 30th 2010



Outline

- Malware Threats on Mobile Phones
- Approaches on Malware Detection
- Our Approach on Anomaly-Detection



- People use mobile phones differently than 3 years ago
 - Facebook, Twitter, Banking, E-Mail....
- Way more resources in a mobile handset
- Security on desktopsystems has improved
 - \rightarrow less interesting?



 $((\mathbf{o}))$

A Smartphone Platform in 2010:



© Pierre Alan Lepetit, CC-BY, Wikimedia Commons



- 1 GHz Cortex A8
- 512 MB Ram
- 802.11bgn-Wireless
- 7.2 MBps HSDPA
- 32 GB Flash-Memory
- 132g
- Desktop-Related OS: (Linux, OS X/iOS, Palm OS)

- Valuable data on the handset
- Targeted attacks on specific persons
- Tracking of users via GPS/Location Based Services
- Tracking of communication behavior of user



Open Source Components

Software vendors don't cover all attack vectors

OEMs hinder deployment of patches



Attack-Vectors on Smartphone Platforms:

- Malicious Apps (dialer, spyware...)
- Cellular Baseband (SMS,MMS, rogue basestations...)
- WiFi Baseband/Services (Bluetooth,WLAN)
- OS / 3rd Party Libraries (Linux, OS X, PDF, SQL, Drivers...)
- Browser (Webkit is standard on most systems)
- Network-Attacks over IP-Layer (e.g. XMPP, Bonjour)
- Chained Exploits (e.g. first use malicious *.pdf then start local root-Exploit)



On the Plus-Side:

- Modern Mobile Platforms were developed with security in mind, not as an afterthought.
- Tighter Control of Software-Platforms adds barriers for malware (App-Store, Reviewing)
- Carrier-Networks can (in theory) add to additional security



Outline:

- Possible solutions to the problem
 - Signature based detection
 - Behavior based detection
 - Cooperative approach
- Survey of real malware detection software



Signature based Malware Detection

- Scanning of Data against signatures of known malware
 - \rightarrow unknown malware is not detected
 - \rightarrow regular updates of Sig-DB are necessary
- No protection against behavioral attacks.
- Unreliable for hidden malware



Behavior based Approach:

- Scanning for behavior of application:
 - \rightarrow Scanning of Data on Phone
 - \rightarrow Suspicious network traffic/SMS
- Scanning for behavior of handset:
 - \rightarrow Handset is active while in standby mode
 - \rightarrow Devices in Action whithout associated Application (Bluetooth, GPS)



Server/Cloud based Approaches:

- Putting the Workload away from the phone:
 - \rightarrow improved battery life
 - \rightarrow no updates on phone necessary
- Cooperative Approach:
 - \rightarrow Other nodes profit from scan-results
 - \rightarrow Node can be warned before attack happens



How does real malware detection Software work?

Two leading anti-malware apps in Android-Market have been analysed:

- No impact on battery runtime and small size
 - \rightarrow No big DB of signatures
 - \rightarrow No scan of running software/processes
- Both rely heavily on cloud-services

 \rightarrow Cooperative/centralized approach?

 Focussed on rogue applications, no scan of data or network-traffic!



Summary:

- Traffic and Data is ignored by most approaches.
 - \rightarrow more difficult than scanning of Apps
- Software seems to be blind to Attacks over unsolicited network traffic



Combining different approaches:

- Preliminary scanning on Handset
- Suspicious data is forwarded to cloud service

 \rightarrow frees local resources for more $% \left({\left| {{{\rm{s}}} \right|_{\rm{s}}} \right)$ intense scanning of traffic or data



Requirements for a 1st Malware detection stage:

- Focused on data & traffic instead of apps
- Lightweight & Simple
- False positives are possible
 - \rightarrow may be discovered in 2nd stage
- Should be able to find different types of attacks in different environments.
- Should be platform agnostic



Our Proposal:

- **Entropy-Fingerprinting!**
 - Fast and lightweight
 - Can be implemented on all platforms
 - Allows to detect anomalies in data-streams
 - Does not need to understand semantics of processed data



Entropy-Fingerprinting of Data-Stream:

- Different types of data have different entropysignature
- Local differences in entropy can point to suspicious Data or hidden Shellcode.
- Shellcode has special characteristics (NOP-Sleds, Landing Zones, lots of system-calls...)



Characteristics of Data:

- Most Data transferred over Network is either compressed or text-based:
 - → Compressed Data has high entropy-values
 - \rightarrow Entropy of Text is significantly lower

How do entropy-values of different types of data compare?



Entropy-Fingerprinting of Data-Stream:

- Different types of data have different entropysignature
- Local differences in entropy can point to suspicious Data or hidden Shellcode.
- Shellcode has special characteristics (NOP-Sleds, Landing Zones, lots of system-calls...)



Entropy Distribution in Data-Types



Entropy Distribution in Data-Types



Entropy Distribution in Data-Types





- Entropy-fingerprinting seems promising for malware-detection in compressed Datatypes.
- Results for non-compressed Datatypes is inconclusive

 \rightarrow More Work needed



Questions?



References

- Oberheide et al.: CloudAV: N-Version Antivirus in the Network Cloud, Proc. of the 17th USENIX Security Symposium . San Jose, CA, July 2008
- Liang Xie et al.: pBMDS: a behavior-based malware detection system for cellphone devices. Proc. of the 3rd ACM conference on Wireless network security. (WiSec '10), pp. 37-48, March 2010.
- Conti et al.: Automated mapping of large binary objects using primitive fragment type classification. Proc. of the Tenth Annual DFRWS Conference, pp S3-S12, August 2010

