CIS 700/002 : Special Topics : A Large-Scale Analysis of the Security of Embedded Firmwares

Teng Zhang

CIS 700/002: Security of EMBS/CPS/IoT

Department of Computer and Information Science
School of Engineering and Applied Science
University of Pennsylvania

Feb 17 2017

Introduction

- Firmware: software embedded in the hardware device
  - bugs and misconfigurations
  - not updated
  - bad security reputation
- This paper presents a horizontal, large-scale analysis of security of firmware images
  - manual analysis is slow and does not scale well
  - the same vulnerability may be present in different devices or even in different types of devices
    - introduced by the same development or integration vendors
  - global understanding of security of embedded systems

Methodology

- Static analysis
  - scales better
  - no need to require access to the physical devices
- Process
  - collection of firmware images
  - implementation of a distributed architecture to unpack and run static analysis on the collected firmware images
  - implementation of correlation engine to compare and find similarities among objects collected
    - “propagate” vulnerabilities

Challenges

- Building a representative dataset
  - the embedded systems environment is heterogeneous
  - lack of centralized points of collection makes it difficult to gather a large and triaged dataset
- Firmware identification
  - formats are diverse
  - hard to extract meta data
- Unpacking and custom formats
  - locating and extracting important functional blocks from compressed images
  - monolithic images are most challenging
    - the bootloader, the operating, system kernel, the applications, and other resources are combined together in a single memory image

Challenges(cont.)

- Scalability and computational limits
  - correlating information across multiple images
  - one-to-one comparison of each pair of unpacked files
    - use fuzzy hash to compare
    - time-consuming even distributed architectures are used
- Results confirmation
  - possible vulnerabilities are found through static analysis but hard to test
  - need tedious manual work

Architecture

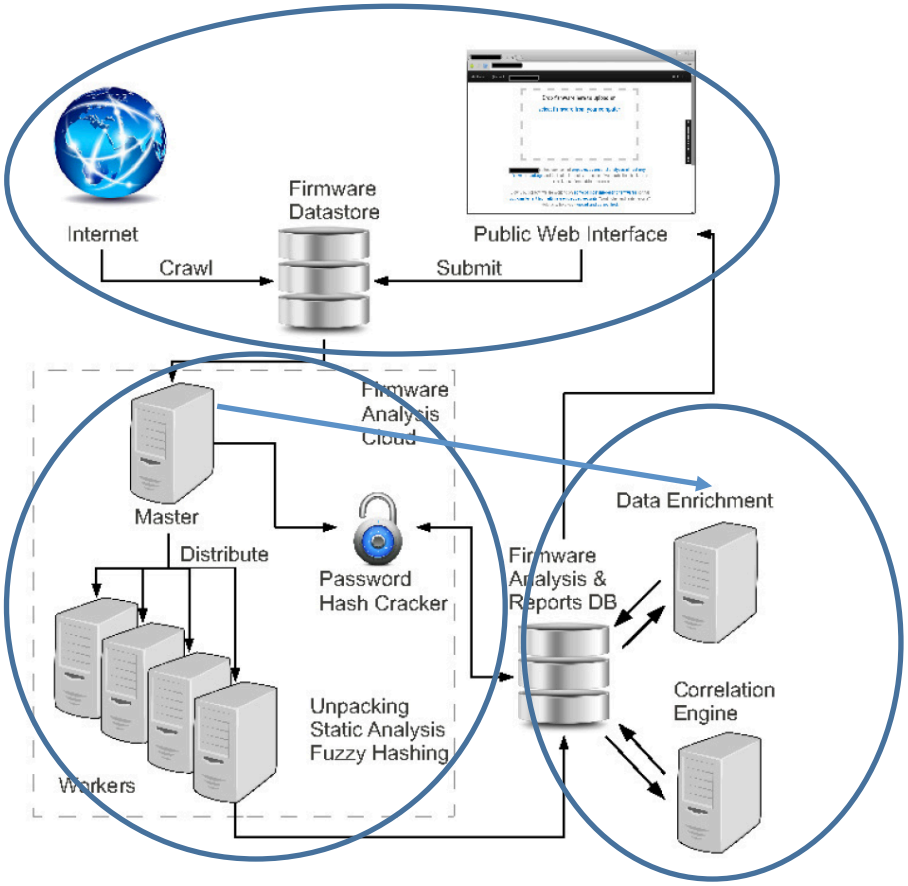


Figure 1: Architecture of the entire system.

Firmware acquisition and storage

- Two ways of collecting images
  - web crawler
  - website for user submission
- web crawler
  - initialize crawler with well-known manufacturers
  - use public FTP indexing engines to search for files with keywords related to firmware images
  - use Google Custom Search Engines (GCSE) to create customized search engines

Unpacking and analysis

- Unpacking Frameworks
  - Binary Analysis Toolkit (BAT) and its extension is used

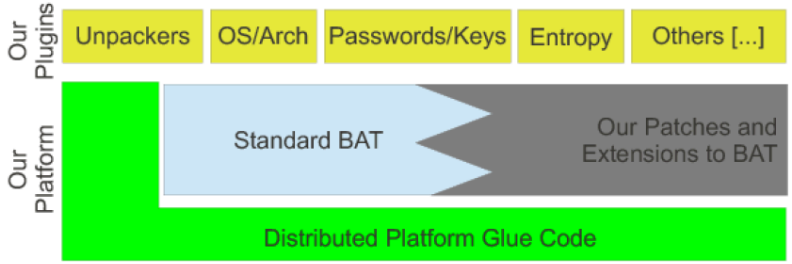


Figure 2: Architecture of a single worker node.

- Password Hash Cracking
  - John The Ripper is used to crack password
  - Run on GPU machines
- Parallelizing the Unpacking and Analysis

Correlation Engine

- Find similarities between different firmware images
- Comparison is made along four different dimensions
  - shared credential
  - shared self-signed certificates
  - common keywords
  - fuzzy hashes

Correlation Engine(cont.)

- Hard coded passwords and self-signed certificates exist in the firmware
  - hint for the strong connection between firmware images
    - CCTV systems from two different vendors have the same default non-trivial password
- Keywords are specific strings extracted by static analysis
  - common backdoor functionalities
  - common compilation and SDK paths
    - cluster images of different devices

Correlation Engine(cont.)

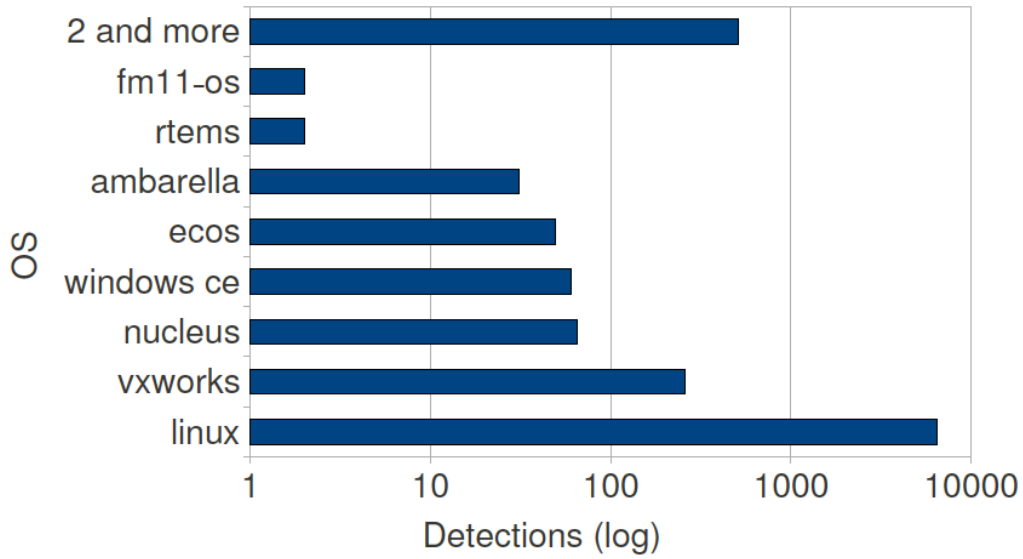
- Fuzzy hashes
  - ability to compare two distinctly different items and determine a fundamental level of similarity between the two
- Engine computes both *ssdeep* and *sdhash* to tell the similarities between files

Data Enrichment

- Extend the knowledge base about firmware images
- Extract info from firmware
  - <title> tag of web pages
  - authentication realms of web servers
- Correlate SSL certificates
  - find SSL certificates in the firmware
  - compare with certificates collected by ZMap
    - ZMap is a powerful tool for internet-scale scan

Dataset and Result

- Overall statistics
  - 32,356 images
  - 63% ARM devices, 7% MIPs devices
  - 86% Linux OS



Dataset and Result(cont.)

- Password Hashes
  - /etc/passwd and /etc/shadow are targets of attackers
  - 100 distinct password hashes covers 681 distinct firmware images belonging to 27 venders
  - 58 of them are recovered belonging to 538 firmware images

Dataset and Result(cont.)

- Certificates and Private RSA Keys Statistics
  - many devices contains self-signed certificates and private keys
  - 41 self-signed certificates with RSA keys were obtained
    - 35,000 devices use these certificates
    - both certificates and keys should be regenerated or the https would be broken

Dataset and Result(cont.)

- Packaging Outdated and Vulnerable Software
  - firmware may rely on third-party software which may be outdated
- Building images as root
  - host info may be leaked
- Web Servers Configuration
  - 81 % web servers were configured to run as a privileged user

Case Studies

- Backdoors
  - some backdoors can be detected using simple keyword match
  - one backdoor was first discovered by string search
  - correlation engine is then used to find the same backdoor for two different types of devices
  - These devices rely on the SoC from the same vendor

Case Studies(cont.)

- Private SSL Keys

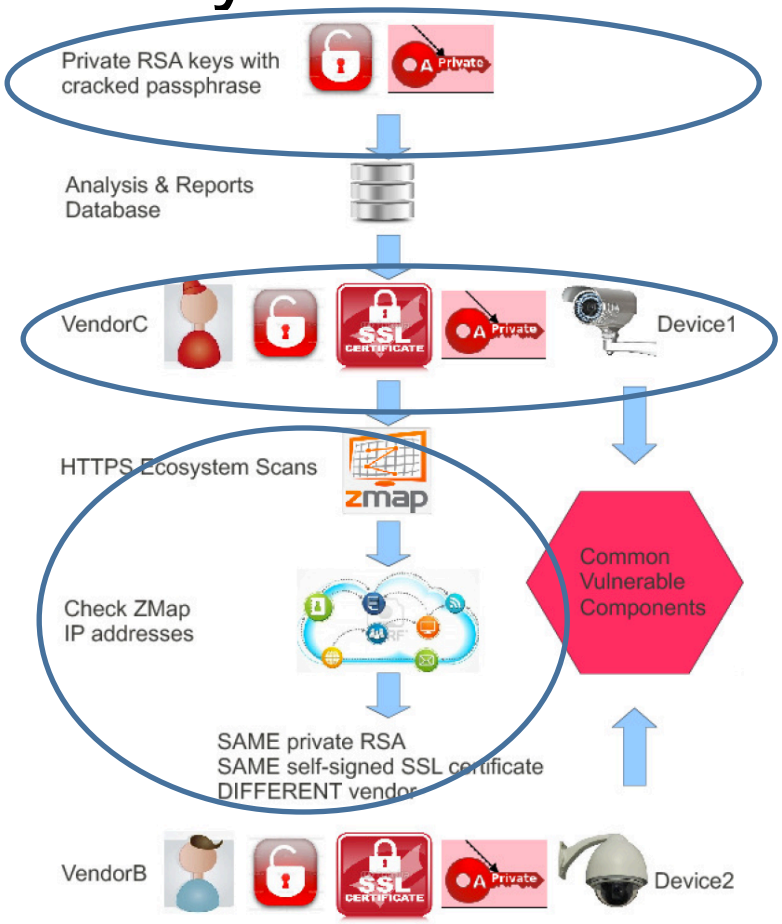
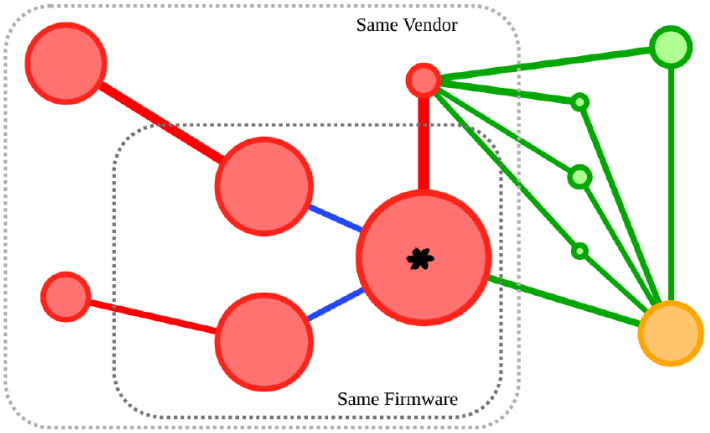


Figure 4: Correlation engine and shared self-signed certificates clustering.

Case Studies(cont.)

- XSS in WiFi Enabled SD Cards
  - Some SD cards have wifi interface with web server
  - A Cross Site Scripting (XSS) vulnerability was found in one of these web interfaces
  - Correlation engine was used to find the same vulnerability on other files or images



Conclusion

- Large-scale analysis of firmware images is useful
  - find the same vulnerabilities from different images or even different vendors
- Firmware images are far from secure
  - bad practices repeatedly appear
  - both manufactures and users should pay much more attention

Related Papers

- Costin, Andrei, Apostolis Zarras, and Aurélien Francillon. "Automated dynamic firmware analysis at scale: a case study on embedded web interfaces." *Proceedings of the 11th ACM on Asia Conference on Computer and Communications Security*. ACM, 2016.
- Roussev, Vassil. "An evaluation of forensic similarity hashes." *digital investigation* 8 (2011): S34-S41.
- Durumeric, Zakir, Eric Wustrow, and J. Alex Halderman. "ZMap: Fast Internet-wide Scanning and Its Security Applications." *Usenix Security*. Vol. 2013. 2013.