IFIP SEC '17 (Rome, Italy)

Towards Automated Classification of Firmware Images and Identification of Embedded Devices

Andrei Costin (University of Jyvaskyla)
Apostolis Zarras (TUM)
Aurelien Francillon (EURECOM)

Agenda

- **Introduction**
- Contributions
- Firmware Classification
- Device Fingerprinting
- Conclusions and Future Work
- Acknowledgements and Q&A

Introduction

- IoT and embedded devices
  - Increasingly present in any computing environment
  - May be *vulnerable/exploitable*
  - Rely on *network connectivity*
  - Often administered through *web interfaces*
  - Depend on and run *firmware packages*

Introduction

- IoT and embedded firmware packages
  - Software that runs on intended *IoT and embedded devices*
  - Contain many software features and modules
  - May contain *bugs/vulnerabilities*
  - Can yield richer knowledge if analyzed in similar clusters rather than alone
    - E.g., diffing consecutive versions and patches

Introduction

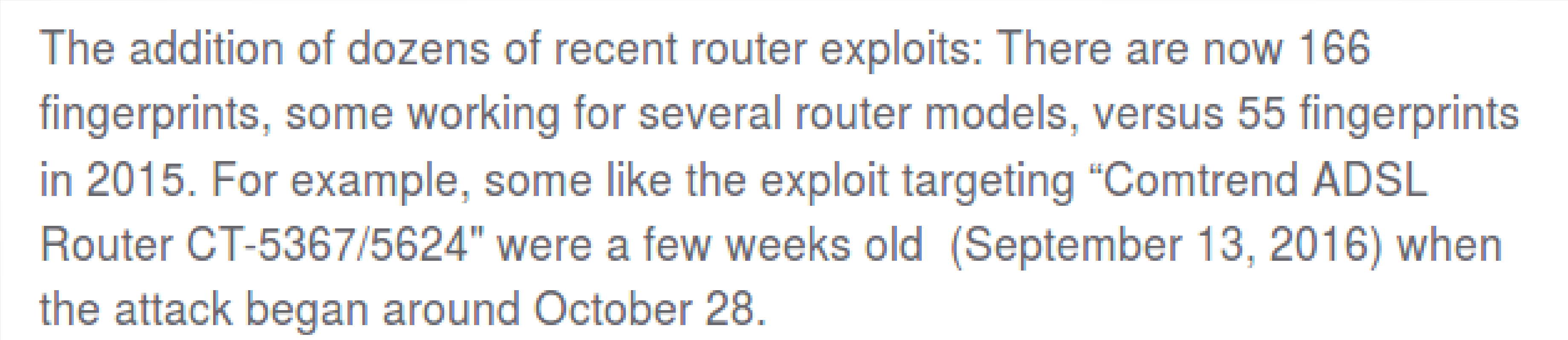
- The number of IoT devices in 2016 was around 6-7 billions [GAR15]
- The number of IoT firmware packages in 2014 was at least in the range of hundreds of thousands [COS14]
- Manual analysis and triage does not scale

Introduction: Research problems

- We formulate the following research problems
  - How to automatically label the brand and the model of the device for which the firmware is intended
  - How to automatically identify the vendor, the model, and the firmware version of an arbitrary web-enabled online device

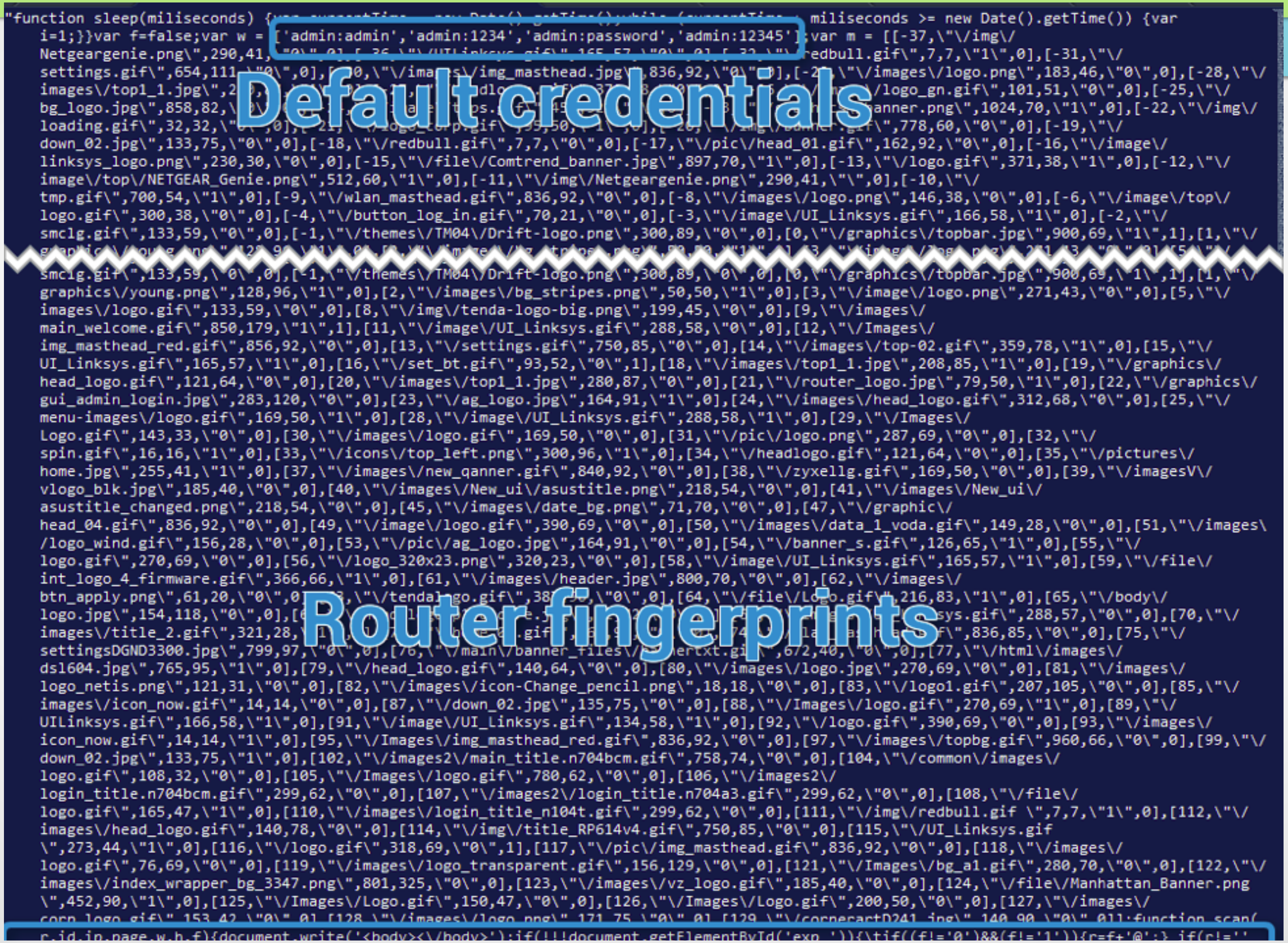
Introduction: Real-world attacks

- "DNSChanger EK" (Dec 2016) [PRO16]
  - Also "CSRF (Cross-Site Request Forgery) SOHO Pharming" (2015)



The addition of dozens of recent router exploits: There are now 166 fingerprints, some working for several router models, versus 55 fingerprints in 2015. For example, some like the exploit targeting “Comtrend ADSL Router CT-5367/5624" were a few weeks old (September 13, 2016) when the attack began around October 28.

Introduction: Real-world attacks



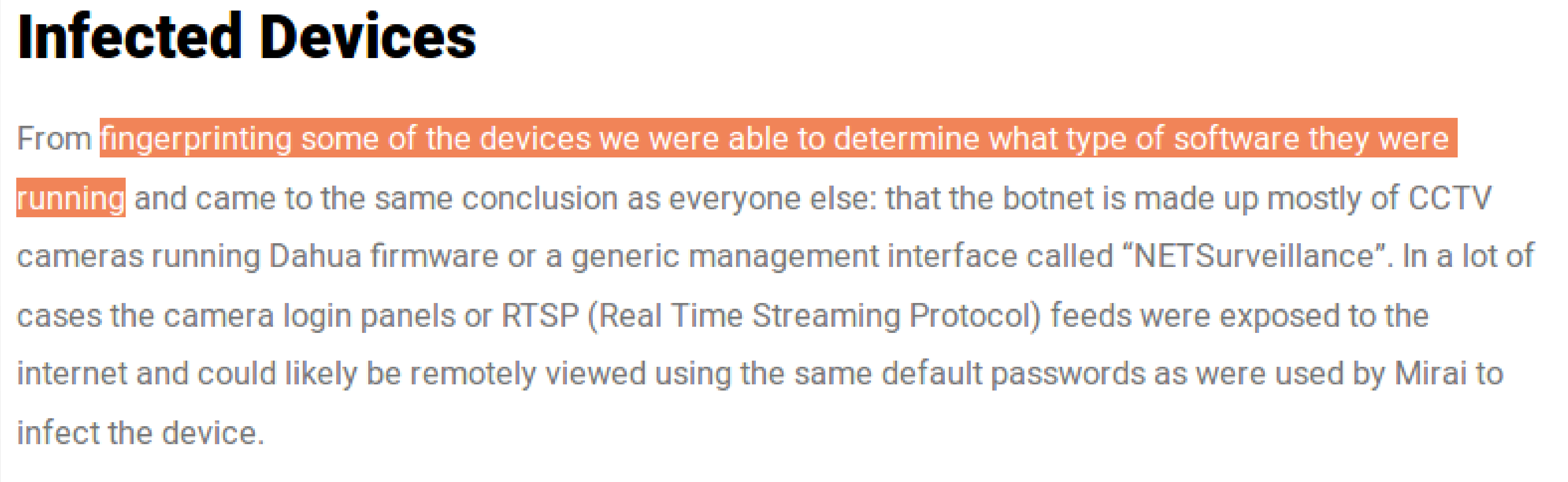
Default credentials

Router fingerprints

```
"function sleep(miliseconds) {var currentTime = new Date().getTime();while (currentTime + miliseconds >= new Date().getTime()) {var
    i=1;}}var f=false;var w = ['admin:admin','admin:1234','admin:password','admin:12345'];var m = [[-37,\"\/img\/
    Netgeargenie.png\",290,41,\"0\",0],[-36,\"\/UILinksys.gif\",165,57,\"0\",0],[-32,\"\/redbull.gif\",7,7,\"1\",0],[-31,\"\/
    settings.gif\",654,111,\"0\",0],[-30,\"\/images\/img_masthead.jpg\",836,92,\"0\",0],[-29,\"\/images\/logo.png\",183,46,\"0\",0],[-28,\"\/
    images\/top1_1.jpg\",2[illegible]dlo[illegible]\/logo_gn.gif\",101,51,\"0\",0],[-25,\"\/
    bg_logo.jpg\",858,82,\"0\[illegible]anner.png\",1024,70,\"1\",0],[-22,\"\/img\/
    loading.gif\",32,32,\"0\",0],[-21,\"\/logo_corp.gif\",[illegible],50,\"1\",0],[-20,\"\/img\/banner.gif\",778,60,\"0\",0],[-19,\"\/
    down_02.jpg\",133,75,\"0\",0],[-18,\"\/redbull.gif\",7,7,\"0\",0],[-17,\"\/pic\/head_01.gif\",162,92,\"0\",0],[-16,\"\/image\/
    linksys_logo.png\",230,30,\"0\",0],[-15,\"\/file\/Comtrend_banner.jpg\",897,70,\"1\",0],[-13,\"\/logo.gif\",371,38,\"1\",0],[-12,\"\/
    image\/top\/NETGEAR_Genie.png\",512,60,\"1\",0],[-11,\"\/img\/Netgeargenie.png\",290,41,\"\",0],[-10,\"\/
    tmp.gif\",700,54,\"1\",0],[-9,\"\/wlan_masthead.gif\",836,92,\"0\",0],[-8,\"\/images\/logo.png\",146,38,\"0\",0],[-6,\"\/image\/top\/
    logo.gif\",300,38,\"0\",0],[-4,\"\/button_log_in.gif\",70,21,\"0\",0],[-3,\"\/image\/UI_Linksys.gif\",166,58,\"1\",0],[-2,\"\/
    smclg.gif\",133,59,\"0\",0],[-1,\"\/themes\/TM04\/Drift-logo.png\",300,89,\"0\",0],[0,\"\/graphics\/topbar.jpg\",900,69,\"1\",1],[1,\"\/
    graphics\/young.png\",128,96,\"1\",0],[2,\"\/images\/bg_stripes.png\",50,50,\"1\",0],[3,\"\/image\/logo.png\",271,43,\"0\",0],[5,\"\/
    smclg.gif\",133,59,\"0\",0],[-1,\"\/themes\/TM04\/Drift-logo.png\",300,89,\"0\",0],[0,\"\/graphics\/topbar.jpg\",900,69,\"1\",1],[1,\"\/
    graphics\/young.png\",128,96,\"1\",0],[2,\"\/images\/bg_stripes.png\",50,50,\"1\",0],[3,\"\/image\/logo.png\",271,43,\"0\",0],[5,\"\/
    images\/logo.gif\",133,59,\"0\",0],[8,\"\/img\/tenda-logo-big.png\",199,45,\"0\",0],[9,\"\/images\/
    main_welcome.gif\",850,179,\"1\",1],[11,\"\/image\/UI_Linksys.gif\",288,58,\"0\",0],[12,\"\/Images\/
    img_masthead_red.gif\",856,92,\"0\",0],[13,\"\/settings.gif\",750,85,\"0\",0],[14,\"\/images\/top-02.gif\",359,78,\"1\",0],[15,\"\/
    UI_Linksys.gif\",165,57,\"1\",0],[16,\"\/set_bt.gif\",93,52,\"0\",1],[18,\"\/images\/top1_1.jpg\",208,85,\"1\",0],[19,\"\/graphics\/
    head_logo.gif\",121,64,\"0\",0],[20,\"\/images\/top1_1.jpg\",280,87,\"0\",0],[21,\"\/router_logo.jpg\",79,50,\"1\",0],[22,\"\/graphics\/
    gui_admin_login.jpg\",283,120,\"0\",0],[23,\"\/ag_logo.jpg\",164,91,\"1\",0],[24,\"\/images\/head_logo.gif\",312,68,\"0\",0],[25,\"\/
    menu-images\/logo.gif\",169,50,\"1\",0],[28,\"\/image\/UI_Linksys.gif\",288,58,\"1\",0],[29,\"\/Images\/
    Logo.gif\",143,33,\"0\",0],[30,\"\/images\/logo.gif\",169,50,\"0\",0],[31,\"\/pic\/logo.png\",287,69,\"0\",0],[32,\"\/
    spin.gif\",16,16,\"1\",0],[33,\"\/icons\/top_left.png\",300,96,\"1\",0],[34,\"\/headlogo.gif\",121,64,\"0\",0],[35,\"\/pictures\/
    home.jpg\",255,41,\"1\",0],[37,\"\/images\/new_qanner.gif\",840,92,\"0\",0],[38,\"\/zyxellg.gif\",169,50,\"0\",0],[39,\"\/imagesV\/
    vlogo_blk.jpg\",185,40,\"0\",0],[40,\"\/images\/New_ui\/asustitle.png\",218,54,\"0\",0],[41,\"\/images\/New_ui\/
    asustitle_changed.png\",218,54,\"0\",0],[45,\"\/images\/date_bg.png\",71,70,\"0\",0],[47,\"\/graphic\/
    head_04.gif\",836,92,\"0\",0],[49,\"\/image\/logo.gif\",390,69,\"0\",0],[50,\"\/images\/data_1_voda.gif\",149,28,\"0\",0],[51,\"\/images\
    /logo_wind.gif\",156,28,\"0\",0],[53,\"\/pic\/ag_logo.jpg\",164,91,\"0\",0],[54,\"\/banner_s.gif\",126,65,\"1\",0],[55,\"\/
    logo.gif\",270,69,\"0\",0],[56,\"\/logo_320x23.png\",320,23,\"0\",0],[58,\"\/image\/UI_Linksys.gif\",165,57,\"1\",0],[59,\"\/file\/
    int_logo_4_firmware.gif\",366,66,\"1\",0],[61,\"\/images\/header.jpg\",800,70,\"0\",0],[62,\"\/images\/
    btn_apply.png\",61,20,\"0\",0],[63,\"\/tendalogo.gif\",38[illegible],\"0\",0],[64,\"\/file\/Logo.gif\",216,83,\"1\",0],[65,\"\/body\/
    logo.jpg\",154,118,\"0\",0],[6[illegible]sys.gif\",288,57,\"0\",0],[70,\"\/
    images\/title_2.gif\",321,28,\"1\",0],[[illegible].gif\",[illegible]\",836,85,\"0\",0],[75,\"\/
    settingsDGND3300.jpg\",799,97,\"0\",0],[76,\"\/main\/banner_files\/[illegible]ertxt.gif\",672,40,\"0\",0],[77,\"\/html\/images\/
    dsl604.jpg\",765,95,\"1\",0],[79,\"\/head_logo.gif\",140,64,\"0\",0],[80,\"\/images\/logo.jpg\",270,69,\"0\",0],[81,\"\/images\/
    logo_netis.png\",121,31,\"0\",0],[82,\"\/images\/icon-Change_pencil.png\",18,18,\"0\",0],[83,\"\/logo1.gif\",207,105,\"0\",0],[85,\"\/
    images\/icon_now.gif\",14,14,\"0\",0],[87,\"\/down_02.jpg\",135,75,\"0\",0],[88,\"\/Images\/logo.gif\",270,69,\"1\",0],[89,\"\/
    UILinksys.gif\",166,58,\"1\",0],[91,\"\/image\/UI_Linksys.gif\",134,58,\"1\",0],[92,\"\/logo.gif\",390,69,\"0\",0],[93,\"\/images\/
    icon_now.gif\",14,14,\"1\",0],[95,\"\/Images\/img_masthead_red.gif\",836,92,\"0\",0],[97,\"\/images\/topbg.gif\",960,66,\"0\",0],[99,\"\/
    down_02.jpg\",133,75,\"1\",0],[102,\"\/images2\/main_title.n704bcm.gif\",758,74,\"0\",0],[104,\"\/common\/images\/
    logo.gif\",108,32,\"0\",0],[105,\"\/Images\/logo.gif\",780,62,\"0\",0],[106,\"\/images2\/
    login_title.n704bcm.gif\",299,62,\"0\",0],[107,\"\/images2\/login_title.n704a3.gif\",299,62,\"0\",0],[108,\"\/file\/
    logo.gif\",165,47,\"1\",0],[110,\"\/images\/login_title_n104t.gif\",299,62,\"0\",0],[111,\"\/img\/redbull.gif \",7,7,\"1\",0],[112,\"\/
    images\/head_logo.gif\",140,78,\"0\",0],[114,\"\/img\/title_RP614v4.gif\",750,85,\"0\",0],[115,\"\/UI_Linksys.gif
    \",273,44,\"1\",0],[116,\"\/logo.gif\",318,69,\"0\",1],[117,\"\/pic\/img_masthead.gif\",836,92,\"0\",0],[118,\"\/images\/
    logo.gif\",76,69,\"0\",0],[119,\"\/images\/logo_transparent.gif\",156,129,\"0\",0],[121,\"\/Images\/bg_a1.gif\",280,70,\"0\",0],[122,\"\/
    images\/index_wrapper_bg_3347.png\",801,325,\"0\",0],[123,\"\/images\/vz_logo.gif\",185,40,\"0\",0],[124,\"\/file\/Manhattan_Banner.png
    \",452,90,\"1\",0],[125,\"\/Images\/Logo.gif\",150,47,\"0\",0],[126,\"\/Images\/Logo.gif\",200,50,\"0\",0],[127,\"\/images\/
    corp_logo.gif\",153,42,\"0\",0],[128,\"\/images\/logo.png\",171,75,\"0\",0],[129,\"\/cornerartD241.jpg\",140,90,\"0\",0]];function scan(
    c,id,ip,page,w,h,f){document.write('<body><\/body>');if(!!!document.getElementById('exp_')){\tif((f!='0')&&(f!='1')){c=f+'@';} if(c!=''
```

Introduction: Real-world digital investigations

- „Mapping Mirai: A Botnet Case Study“ (Oct 2016) [MAL16]
  - Mirai – perhaps the most disruptive and well-known DDoS botnet



Infected Devices

From fingerprinting some of the devices we were able to determine what type of software they were running and came to the same conclusion as everyone else: that the botnet is made up mostly of CCTV cameras running Dahua firmware or a generic management interface called “NETSurveillance”. In a lot of cases the camera login panels or RTSP (Real Time Streaming Protocol) feeds were exposed to the internet and could likely be remotely viewed using the same default passwords as were used by Mirai to infect the device.

Introduction: Real-world CVE management

- CVE-2013-5637, CVE-2013-5638 – Consecutive/similar firmware clustering allows proper identification of impacted components [COS14]

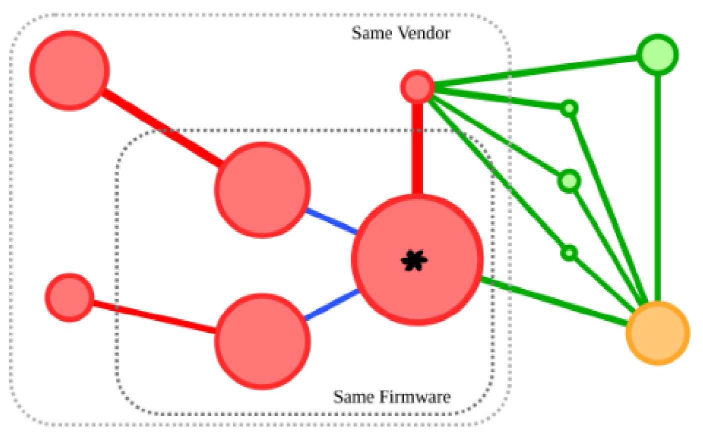


Figure 5: Fuzzy hash clustering and vulnerability propagation. A vulnerability was propagated from a *seed file (*)* to other two files from the same firmware and three files from the same vendor (in red) as well as one file from another vendor (in orange). Also four non-vulnerable files (in green) have a strong correlation with vulnerable files. Edge thickness displays the strength of correlation between files.

Agenda

- Introduction
- **Contributions**
- Firmware Classification
- Device Fingerprinting
- Conclusions and Future Work
- Acknowledgements and Q&A

Contributions

- We propose and study the firmware features and the ML algorithms in the context of firmware classification
- We research the fingerprinting and identification of web-enabled embedded devices and their firmware version
- We present and discuss direct practical applications for both techniques

Agenda

- Introduction
- Contributions
- **Firmware Classification**
- Device Fingerprinting
- Conclusions and Future Work
- Acknowledgements and Q&A

Firmware Classification: Related Work

- Clemens [CLE15]
- Context focused on
  - Explosion of different types of devices and myriad of executable code (firmware, mobile apps, etc.)
  - Automating digital forensic for forensic analysis, reverse engineering, or malware detection
- Their dataset over 16000 code samples from 20 (embedded) architectures
- Their classifiers achieve very high accuracy with relatively small sample sizes

Firmware Classification: Dataset

- Total Firmware Vendors: 13
- Total Firmware Files: 215
- Firmwares Per Vendor: 5(min)/54(max)/16(avg)
- Dataset: www.firmware.re/ml/

Firmware Classification: Features

- Firmware File Size
- Firmware File Content Properties (output of „*ent*“, except bytes frequency)
- Firmware File Strings (class strings, class unique strings)
- Fuzzy Hash Similarity (threshold-based binary value feature)

Firmware Classification: Evaluation

- ML: Decission Tree (DT) and Random Forests (RF) from *sklearn*
- Training/Evaluation points
  - Training sets size 10% and 90% of each firmware class
  - Training sets increment 10% at each evaluation point
- At each training/evaluation point
  - Runs 100 times with new random choice of training set data
  - Runs both DT and RF
  - Runs four different sets of features

Firmware Classification: Evaluation

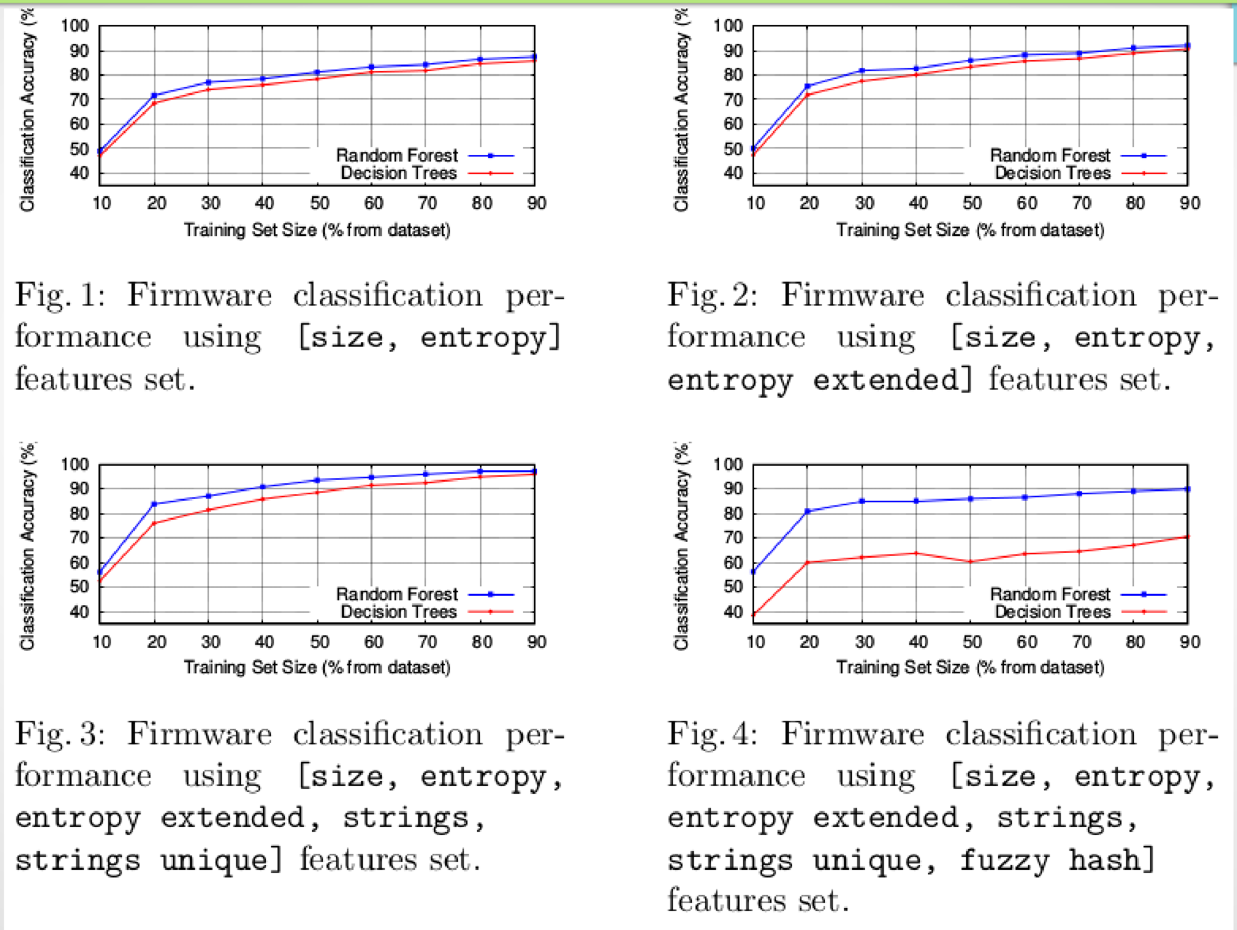


Fig. 1: Firmware classification performance using [size, entropy] features set.

Fig. 2: Firmware classification performance using [size, entropy, entropy extended] features set.

Fig. 3: Firmware classification performance using [size, entropy, entropy extended, strings, strings unique] features set.

Fig. 4: Firmware classification performance using [size, entropy, entropy extended, strings, strings unique, fuzzy hash] features set.

Firmware Classification: Results

- In summary:
  - RF with „best“ features-set and 50% training reaches 93.5% accuracy
  - „Best“ features-set was *[size, entropy, entropy extended, category strings, category unique strings]*
  - Using only basic features *[size, entropy]* do not even reach 90% accuracy (either RF or DT)
  - As expected
    - Increased training set results in increased accuracy
    - RF more accurate than DT

Agenda

- Introduction
- Contributions
- Firmware Classification
- **Device Fingerprinting**
- Conclusions and Future Work
- Acknowledgements and Q&A

Device Fingerprinting: Related Work

- Samarasinghe and Mannan [SAM16]
- Context focused on the study of weak SSL/TLS in IoT/embedded devices
- Performed IoT/embedded device fingerprinting
- Used HTTPS web-interface and certificates of IoT/embedded devices

Device Fingerprinting: Dataset

- Total Devices: 31
  - Emulated Devices: 27
    - Vendors: 3
    - Functional categories: 7
  - Physical Devices: 4
    - Vendors: 2
    - Functional categories: 4

Device Fingerprinting: Features

- Total Features: 6
- HTTP Web Sitemap
- HTTP Finite-State Machine (FSM)
  - Model able to learn the headers’ order of an HTTP response
  - Use this order to classify an unknown HTTP conversation
- Cryptographic Hashing and Fuzzy Hashing for each sitemap entry
  - HTML Content
  - HTTP Headers

Device Fingerprinting: Evaluation

- Feature ranking/scoring
  - „Majority voting“
  - „Uniform weights“
  - „Non-uniform weights“ (empirical weights)
  - „Score fusion“
  - Future work: use (un)supervised ML

Device Fingerprinting: Results

- In summary:
  - On average 89.4% identification accuracy
  - Cryptographic hash of HTML content most „stable“ feature
  - Fuzzy hash of HTTP headers least „stable“ feature
  - „Majority voting“ yielded most accurate matching

Agenda

- Introduction
- Contributions
- Firmware Classification
- Device Fingerprinting
- **Conclusions and Future Work**
- Acknowledgements and Q&A

Conclusions

- We presented two complementary techniques for IoT firmware/devices
  - Embedded firmware supervised learning and classification
  - Embedded web interface fingerprinted identification
- We achieved average accuracies of 93.5% and 89.4% respectively
- We presented practical use-cases for our techniques
- Our scripts and datasets will be updated at: www.firmware.re/ml/

Future Work

- Larger and more varied datasets for both techniques
- Unsupervised automated firmware emulation and vulnerability discovery [COS16]
- Unsupervised and more scalable ML for both techniques
- Evaluation of more ML algorithms with more parameters and features

Agenda

- Introduction
- Contributions
- Firmware Classification
- Device Fingerprinting
- Conclusions and Future Work
- **Acknowledgements and Q&A**

Acknowledgements

- IFIP SEC '17 organizers, and reviewers for valuable comments
- Prof. Pietro Michiardi for insightful discussions and feedback
- Ala Raddaoui for his early contributions to this study

Q&A

- Questions, suggestions, ideas?

www.firmware.re/ml/

ancostin@jyu.fi

andrei@firmware.re

Twitter: @costinandrei

References

- [COS14] A. Costin, J. Zaddach, A. Francillon, D. Balzarotti, „A Large Scale Analysis of the Security of Embedded Firmwares“, USENIX Security (2014)
- [COS16] Costin, A., Zarras, A., & Francillon, A., "Automated dynamic firmware analysis at scale: a case study on embedded web interfaces", In Proceedings of the 11th ACM on Asia Conference on Computer and Communications Security (2016)
- [CLE15] Clemens, J., "Automatic classification of object code using machine learning", Digital Investigation 14 (2015)
- [SAM16] Samarasinghe, N., Mannan, M., „Short Paper: TLS Ecosystems in Networked Devices vs. Web Servers.“, Financial Crypto (2016)

References

- [ESC16] Eschweiler, S., Yakdan, K., & Gerhards-Padilla, E., „DiscovRE: Efficient cross-architecture identification of bugs in binary code“, 23th Symposium on Network and Distributed System Security (NDSS) (2016)
- [PRO16] „Home Routers Under Attack via Malvertising on Windows, Android Devices“ https://www.proofpoint.com/us/threat-insight/post/home-routers-under-attack-malvertising-windows-android-devices
- [MAL16] „Mapping Mirai: A Botnet Case Study“ https://www.malwaretech.com/2016/10/mapping-mirai-a-botnet-case-study.html
- [GAR15] http://www.gartner.com/newsroom/id/3165317

IFIP SEC '17 (Rome, Italy)

Towards Automated Classification of Firmware Images and Identification of Embedded Devices

Andrei Costin
ancostin@jyu.fi
University of Jyvaskyla, Finland