htw Hochschule für Technik und Wirtschaft Berlin

University of Applied Sciences

HOCHSCHULE FÜR TECHNIK UND WIRTSCHAFT BERLIN

Master Thesis

Community friendly tracking of Hardware Target Distributions and Software releases

Stefan Venz

Fachbereich I – Ingenieurwissenschaften - Energie und Information
der Hochschule für Technik und Wirtschaft Berlin

Submitted To:
Prof. Dr.-Ing. Thomas Scheffler, HTW Berlin
Prof. Dr.-Ing. Thomas Hühn, Hochschule Nordhausen

Berlin, March 2021

Abstract

In community-based software development projects for multi-platform systems, such as OpenWrt, Tasmota, etc., developers currently have very limited information about the platforms on which and on how many platforms their software is used. Such data could improve the software development process significantly. Therefore this thesis will discuss ...

- Research suitable transmission technologies

- system anonymization

- System identification

- data collection

- data obfuscation

- data transmission

- frequency of transmission

- elimination of duplication

- reference implementation

- data collection disclaimer

- user acceptance

- data evaluation

- first results

Zusammenfassung

Ich versichere an Eides statt, dass ich diese Masterarbeit selbständig verfasst und nur die angegebenen Quellen und Hilfsmittel verwendet habe. Alle wörtlich oder inhaltlich übernommenen Stellen habe ich als solche gekennzeichnet. Die vorliegende Arbeit wurde weder in der vorliegenden noch einer modifizierten Fassung einer dritten, in- oder ausländischen Fakultät als Prüfungsleistung, oder zum Erlangen eines akademischen Grades vorgelegt.

Datum	Stefan Venz

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Developing a multi-platform application is a challenging task. Especially if there is no, or only sparse telemetry and deployment data.

This chapter will discuss the problems developer are currently facing in 1.1 and how we intend to contribute with this thesis to solve this problem. Furthermore an overview of the thesis structure is provided in 1.3

1.1 Problem Statement

In community-based software development projects for multi-platform systems, such as OpenWrt, Tasmota and similar, the access to information on which and on how many platforms the software is used is very limited to developers. Such statistics on the temporal spread of the hardware platforms used the end user (device life-cycle), can improve the software development process. Likewise the process of troubleshooting issues, development of patches and features can be prioritized based on:

- Which platforms are currently most used and in need of additional testing

- Detect end-of-life platforms to remove from development cycle

- Identify systems with software faults

- Distribution of software updates

As the testing process is intense in software engineering due to compile, flash and boot, as well as functionality testing targets with higher distributions could be prioritized. This would allow to roll out e.g. security updates to as many users as possible in short time.

Additionally developer of open source projects, like Internet of Things (IoT) applications, could receive useful information on the update behavior of their user base. Based on these data, warnings to update the firmware of the connected device, could be issued on the control application of the device.
This could lead to an overall increase in network security and hinder the creation of new botnets.

Collecting hard- and software data on user devices is a delicate action. The collected data needs to be anonymized, so no conclusion about the user can be drawn from the data set. Therefore the collected information should exclude sensitive data, or should be anonymized. In addition, the user location should not be derivable from the transnmitted data or the transmission itself. At the same time the data should be obscured from any eavesdropper during the client server communication.
Furthermore, a valid database of software and hardware distribution for developers should emerge. In addition, each device should be recorded and included in the statistics only once. Thus a manipulation of the statistics is made more difficult.

1.2   Contributions

In this thesis we discuss known Systems for data collection and transmission anonymous. Based on the discussion we provide a blueprint for the given problem. In addition an open source implementation will be provided as reference. This will include data collection, transmission on the client site, as well as receiving data and evaluation on the server side

1.3   Thesis outline

The rest of this thesis is organized as follows. In Chapter 2 we discuss related work to our research. Chapter 3 discusses the decision of the software design process and how we are going to solve the stated problem. In chapter The 4. chapter discusses our measurement methodology and in chapter 5 we present the results of our measurement and software design. In chapter 6 we conclude this thesis and identify possibilities for future research and improvements.

Background and Related Work

2.1 Related Software

2.2 Data Anonymization

2.3 Data Transmission

2.4 Summary

In this chapter we summarised previous work on related topics. This allows us to narrow our research. Therefor we concentrate our research on ¡technology¿ as we want to achieve.

<div align="right">

Chapter 3

</div>

<div align="center">

Software Design

</div>

3.1 Parameter

3.2 Data Collection

3.3 Data Transmission

3.4   Reference Implementation

3.5   Summary

In this chapter, we gave an introduction to the data collection application (3.2 and described the concept of data transmission in (section 3.3). We also introduced the reference implementation for the proposed solution in section 3.4.

<div align="right">

Chapter 4

</div>

<div align="right">

Measurements

</div>

To evaluate and debug the conTest framework, we conducted multiple measurements with the Minstrel algorithm. In this chapter we list the parameter we monitored during our measurements and show which limitations our test environment setup faces. In the last part of this chapter we describe the scenarios we would like to use for our measurements.

4.1   Measurement Limitations

4.2   Evaluation Setup

4.3   Summary

In this the section 3.1 we introduced the parameter for our measurements. While we are able to ... we can only measure ... . Section 4.1 provides an overview of the limitations our data set faces. Finally we provide three channel condition scenarios. For this thesis we are going to concentrate on a scenario with improving and decreasing channel conditions.

In the next section we present our evaluation of the collected data from the setup introduced in 4.2.

<div align="right">

Chapter 5

Results

</div>

In this chapter we provide information of our evaluation steps and discuss the results of our measurement setup described in **??**. The small data set allows us to verify our testbed but not the usefulness of Minstrel for medical applications. Therefore further measurements would be needed.

5.1 Anonymization Results

5.2 Telemetry statistics

5.3 Summary

As expected, the user systems are fully anonymized ... and several different devices reported to our servers

Chapter 6

Conclusion and Outlook

6.1   Summary

6.2   Future Directions

While we have shown significant gains, ...