TALLINN UNIVERSITY OF TECHNOLOGY
Faculty of Information Technology
Department of Cyber Security

IVCM09/14
Urmo Lihten 143912IVCM

DETECTION OF PROCESS ABUSE AND DATA REQUEST MISUSE ON ELECTRONIC HEALTH RECORD SYSTEM BASED ON REQUEST LOGS

Master Thesis

Supervisor: Firstname Lastname PhD
Co-Supervisor: Firstname Lastname MSc

Tallinn 2020

TALLINNA TEHNIKAÜLIKOOL
Infotehnoloogia teaduskond
Arvutitehnika instituut

IAY70LT
Firstname Lastname 123456 ABCD

PROTSESSI KÕRVALEKALLETE JA ANDMEPÄRINGUTE VÄÄRKASUTUSE AVASTAMINE TERVISE INFOSÜSTEEMIS PÄRINGU LOGIDE PÕHJAL

Magistritöö

Juhendaja: Firstname Lastname PhD
Kaasjuhendaja: Firstname Lastname MSc

Tallinn <year>

Author's declaration of originality

I hereby certify that I am the sole author of this thesis and that no part of this thesis has been published or submitted for publication. All works and major viewpoints of the other authors, data from other sources of literature and elsewhere used for writing this paper have been referenced.

Author: Urmo Lihten

February 8, 2020

Abstract

Here goes your abstract...

The thesis is in English and contains 14 pages of text, 5 chapters, 23 figures, 8 tables.

Annotatsioon

Annotatsioon on lõputöö kohustuslik osa, mis annab lugejale ülevaate töö eesmärkidest, olulisematest käsitletud probleemidest ning tähtsamatest tulemustest ja järeldustest. Annotatsioon on töö lühitutvustus, mis ei selgita ega põhjenda midagi, küll aga kajastab piisavalt töö sisu. Inglisekeelset annotatsiooni nimetatakse Abstract, venekeelset aga

Sõltuvalt töö põhikeelest, esitatakse töös järgmised annotatsioonid:

- kui töö põhikeel on eesti keel, siis esitatakse annotatsioon eesti keeles mahuga $\frac{1}{2}$ A4 lehekülge ja annotatsioon *Abstract* inglise keeles mahuga vähemalt 1 A4 lehekülg;
- kui töö põhikeel on inglise keel, siis esitatakse annotatsioon (Abstract) inglise keeles mahuga $\frac{1}{2}$ A4 lehekülge ja annotatsioon eesti keeles mahuga vähemalt 1 A4 lehekülg;

Annotatsiooni viimane lõik on kohustuslik ja omab järgmist sõnastust:

Lõputöö on kirjutatud inglise keeles ning sisaldab teksti 14 leheküljel, 5 peatükki, 23 joonist, 8 tabelit.

Glossary of Terms and Abbreviations

ATI	TTÜ Arvutitehnika instituut
DPI	*Dots per inch*, punkti tolli kohta

Contents

List of Figures

List of Tables

1. Introduction

Estonian Health Information System gives doctors the ability to send patient data to centralized information system. From there other medical workers and also the patient can view entered documents and information. This also gives doctors and nurses access to private data, when they are doing examinations and other procedures to their patients.

1.1. Problem statement

Since medical staff can view peoples medical historyin Health Information System, this poses security threat of misusing the queried data. All that is needed, to see the information, is the persons identification code.

It is hard to determine, if patient has really turned to them for medical help or not. When in an emergency and patient is un-cooperative or in such state unable to communicate, patients identity has to be confirmed without an persons consent. Permission or rights to view patients data is usully given, when the person turns to the doctor with medical issue. Meaning, the permission is not spefically given in the information system and thus allowing view data knowing just the persons personal This allows medical staff to open any persons medical history and view it at any given time whether the person has any medical relationship to that medical staff or does not.

When a persons private data such as medical information is viewed and used, then there has to be a reason. Even if it wrongly done but is still explainable (wrong identification code submission into the system by accident due to similarities, typing wrongly by mistake or third person has given wrong patient identification code by accident).

To solve this problem is to detect health records data misusage and errors in the process as early as possible by analyzing Health Information System logs what include data requests and documents sent. Learning about the different processes in which different queries has to be made within patient treatment and data forwarded. This gives the possibility to detect processes and its anomalies - queries and documents sent when not needed or out of the ordinary. Upon problem detection healthcare service provider can be contacted and be questioned, if action was intententional or not. Also to find out the reason. If the misuse is very serious, proper action has to be taken by proper authorities.

2. Approach overview

Estonian Health Information System logs every data query and document sent to it as requests. Every response is either data from queried documents and/or from subqueries to other data providers or approvement, that document or data is saved. Before the request reaches to the database, there are multiple layers of services that receive the request and examine it, if the organization is allowed to send it, if its properly constructed to its standard, if the syntax of query is valid and if subqueries to other information systems is needed. When some part of the checks and validations fail to accept the query, proper error message is sent as an response. If data query is too large or query requests data for large period then information system might cancel the query if it takes a long time to respond or its unable to respond. Requests and responses are sent as XML SOAP messages. These contain different object identification codes to classify each document and query.

Usually every request is made following a certain process. This is agreed upon on an organizational and national level. In information system the process model might be different and needs to be found out. For this process mining tools and machine learning techniques could help to create process models and check conformance. Also detect anomalies is data usage.

Every request has to be parsed for certain data fields, which give input for the process mining tools to form a process model. After that, a conformance check can be made to find anomalies and machine learning helps to find out data misusages.

2.1. Data cleansing

2.2. Data representation

2.3. ...

3. Discovering process models

3.1. Clustering request types

4. Conclusion

References

[1] A. A. Boxwala, J. Kim, J. M. Grillo, and L. Ohno-Machado, "Using statistical and machine learning to help institutions detect suspicious access to electronic health records," *Journal of the American Medical Informatics Association*, vol. 18, pp. 498–505, 07 2011.