1 Using statistical and machine learning to help institutions detect suspicious access to electronic health records

The objective of this article was to determine where machine learning and statistical methods could help identify illegal or suspicious access to electronical health records. Many Health Care Organizations (HCO) prioritize their investigations by monitoring patient records deemed to be very important persons (VIPs) 26 features were collected what could be useful to determine suspicious access. 1291 labeled events (which is very small for a training set comparining the the amounts of data generated in a single day) were used to train and validated on 58 truly inappropriate events. As a conclution it showed that this method can be helpful for security and compliance specialists but there are limitations with LR and SVM models since organizations, who are generating documents, are very different in their organizational structure, workflows, programs and other interinstitutional differences. Usually there are only few information security officers or appropriate compliance specialists in the organization who have to analyze the access and data logs. Due to the large amount of logs and data, it is nearly impossible to audit all the information and assess if any document access was unneeded or data been misused. [1]

2 Detecting Inappropriate Access to Electronic Health Records Using Collaborative Filtering

Two hospital data and Amazon Access data taken into Machine learning algorithms. One dataset is collected for a prior study - Raw data was 34.1 million accesses from two hospitals databases over a 6 months period in 2009. Different types of data is prepared - one portion is labeled and other is non-labeled but with different types of features. Second dataset is Amazon Access data, which is publicy available but not directly related to electronic health records but it has enough similarities to justify using the same model for identifing illegal access. Data is labeled. Liner and logistic regression are strong baselines and perform good on all performance metrics. Collaborative

filtering is also used. As a result ideal end-system would be fully automated. Previous works and models and also this work is a direction to the ideal solution but still far from it because of the complexity and a lot of input is needed from human side to predict correctly. [2]

3 The Personal Health Working Group: final report

Group of private and governmental health information experts gathered and their main assignment was to examine potential benefits of personal health records. Also concerns and issues with these, that needs attention. Shortly people are concerned about their EHR security and privacy. Report makes a summary of the findings and brings out requirements for making a national system supporting personal health records. [3]

4 A Research Agenda for Personal Health Records (PHRs)

Estimated 70 million people in US have access to some form of personal health record (PHR). One survey says 75 percent of those people would use communicate with their physicians if they have the possibility. 60 percent of the people would look up their health records and test results electronically when possible. 91 percent of the people are concerned about their health records privacy and security. This article gives a short overview of the PHR use in the US - function, security, privacy, architecture, adaption. [4]

5 Reviewing the benefits and costs of electronic health records and associated patient safety technologies.

Electronic health record (EHR) systems allow to improve the quality of patient care at the hospital, while reducing costs and enabling research for doctors and other involved individuals. [5]

6 Flagging and Ranking Suspicious Accesses in Electronic Health Record Systems

Increased accessibility to the protected health information in EHR systems can be of a threat for misuse and abuse by authorized users. [6] In this work Supervised machine learning is tried to use. Data is used from a EHR system with 8 million accesses audit log which was collected with a week. Data short overview: 7.5 million total accesses, 6.9 million are repeated access, 21 thousand are self-access, 710 thousand non-self access, 13 thousand unique employees, 152 thousand unique patients, 2.1 thousand unique departments. Hypothesized that a high-risk audit rule holds merit when the observed frequency at which it fires is higher than what would occur due to routine daily behavior.

7 Authorisation and access control for electronic health record systems

An EHR is fundamentally a collaborative information system, which, traditionally, is protected through proactive strategies, such as fine-grained access control technologies. [7] Article focuses more on the security and access part of the EHR, for how someone can get access, roles and rights to access and view EHR systems and how the rights structure should look like. This is not related so much to finding anomalies or suspicious action in the system. Gives overview how to build authorization and access control.

8 A Study of Access Control Requirements for Healthcare Systems Based on Audit Trails from Access Logs

Focuses on 8 hospitals access logs in Norway to uncover if audit trails have information about real user needs that would be beneficial for designing a better system and access controll mechanisms and also find information on any misuse of the data. 8 hospitals data for a period of a month (March 2006). 2 logs - access logs (every time a document is opened with information about user, patient and document) and Actualization and emergency

logs (when electronic personal record is opened through actualization or document opened using emergency access). The dynamics of patient care, in combination with the difficulty in predicting who needs access to a patient's medical record when, make it challenging to deploy such fine-grained control schema without triggering a substantial quantity of false alerts and slowing care workflows when the information is needed the most. [8] This study make recommendations to the EHR access control requirements through the analysis of access logs and their use through explainations and analysis. Machine learning is not used in this but gives some thoughts about the features what to look in access logs and if a machine learning algorithm can use this information.

9 Access Control: how can it improve patients' healthcare?

Despite acknowledging the potential for insider threats, HCOs typically do not instantiate fine-grained controls. [9] Access control models are reviewed and how much of them have beem implied in the hospitals EHR systems.

10 Using statistical and machine learning to help institutions detect suspicious access to electronic health records

Large amount of investigations are started from a patient complaint. [10] This study is more database style auditing using sql queries and bringing together patient and doctor visits or information and analyzing for any misuse. Doesnt use machine learning.

11 Detecting Anomalous User Behaviors in Workflow-Driven Web Applications

Data-centric workflow driven web application is vurnerable to two types of threats - request integrity attacks (vulnerabilities in the implementation of business logic) and guideline violations (privilege misuse scenarios where

business logic and policies are too complex to be accurately defined and enforced) [11]

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.

[2] A. K. Menon, X. Jiang, J. Kim, J. Vaidya, and L. Ohno-Machado, "Detecting inappropriate access to electronic health records using collaborative filtering," *Machine Learning*, vol. 95, pp. 87–101, Apr 2014.

[3] C. for Health Personal Health Working Group *et al.*, "The personal health working group: final report," *Markle Foundation*, 2003.

[4] D. C. Kaelber, A. K. Jha, D. Johnston, B. Middleton, and D. W. Bates, "A Research Agenda for Personal Health Records (PHRs)," *Journal of the American Medical Informatics Association*, vol. 15, pp. 729–736, 11 2008.

[5] N. Menachemi and R. G. Brooks, "Reviewing the benefits and costs of electronic health records and associated patient safety technologies," *Journal of Medical Systems*, vol. 30, pp. 159–168, Jun 2006.

[6] M. S. Hedda, *Flagging and Ranking Suspicious Accesses in Electronic Health Record Systems*. PhD thesis, Vanderbilt University, 2018.

[7] B. Blobel, "Authorisation and access control for electronic health record systems," *International Journal of Medical Informatics*, vol. 73, no. 3, pp. 251 – 257, 2004. Realizing Security into the Electronic Health Record.

[8] L. Rostad and O. Edsberg, "A study of access control requirements for healthcare systems based on audit trails from access logs," in *2006 22nd Annual Computer Security Applications Conference (ACSAC'06)*, pp. 175–186, Dec 2006.

[9] A. FERREIRAabd, C.-C. Ricardo, L. Antunes, and D. Chadwick, "Access control: how can it improve patients' healthcare?," *Medical and care compunetics*, vol. 4, no. 4, p. 65, 2007.

[10] D. Fabbri and K. LeFevre, "Explanation-based auditing," *Proc. VLDB Endow.*, vol. 5, pp. 1–12, Sept. 2011.

[11] X. Li, Y. Xue, and B. Malin, "Detecting anomalous user behaviors in workflow-driven web applications," in *2012 IEEE 31st Symposium on Reliable Distributed Systems*, pp. 1–10, Oct 2012.