A TWO PARTY PRIVACY PRESERVING SET INTERSECTION PROTOCOL AGAINST MALICIOUS USERS

S.Kanahatharani, M.Kanimozhi and R.Harish, UG Scholar
Dr.P.Suresh, Assistant Professor
Kongu Engineering College, Erode, India
tharanisenthil30@gmail.com, psuresh@kongu.ac.in

Abstract

An Electronic Medical Record(EMR) is a computerized medical record created in an organization that delivers care, such as hospital or lab. Electronic medical records tend to be a part of a local stand-alone health information system that allows storage, retrieval and modification of records. Using an EMR to read and write a patient's record is not only possible through a workstation but depending on the type of system and health care settings may also be possible through mobile devices that are handwriting capable. Electronic medical record (EMR) systems have enabled healthcare providers to collect detailed patient information from the primary care domain. At the same time, longitudinal data from EMRs are increasingly combined with bio-repositories to generate personalized clinical decision support protocols. A Bio-repository is a biological materials repository that collects, processes, stores, and distributes bio-specimens to support future scientific investigation. Bio-repositories can contain or manage specimens from animals, including humans, and many other living organisms. Vertebrates, invertebrates, arthropods, and other life-forms are just a few of the many classes of living organisms which can be studied by preserving and storing samples taken. Emerging policies encourage investigators to disseminate such data in a de-identified form for reuse and collaboration, but organizations are hesitant to do so because they fear such actions will jeopardize patient privacy.

1. INTRODUCTION

Recently, EMRs have been combined with bio-repositories to enable is designated as disease positive or negative . As these studies mature, they will support personalized clinical decision support tools and will require longitudinal data to understand how treatment influences a phenotype. Meanwhile, there are challenges to conducting GWAS on a scale necessary to institute changes in healthcare. First, to generate appropriate statistical power, scientists may require access to populations larger than those available in local EMR systems . Second, the cost of a GWAS - incurred in the setup and application of software to process medical records as well as in genome sequencing - is non-trivial Thus, it can be difficult for scientists to generate novel, or validate published, associations. To mitigate this problem, the U.S. National Institutes of Health (NIH) encourages investigators to share data from NIH-supported GWAS into the Database of Genotypes and Phenotypes (dbGaP) . This, however, may lead to privacy breaches if genome-wide association studies (GWAS) with clinical phenomena in the hopes of tailoring healthcare to genetic variants. EMR-based GWAS have focused on static phenotypes; i.e., where a patient patients' clinical or genomic information is associated with their identities

EXISTING SYSTEM

To anonymize patient records in transactional form, a privacy principle is introduced to ensure sets of potentially identifying diagnosis codes are protected from re-identification, while remaining useful for GWAS validations. To enforce this principle, they proposed an algorithm That employs generalization and suppression to group semantically close diagnosis codes together in a way that enhances data utility.

Drawbacks Of Existing System

It is note that the re-identification problem is not addressed by access control and encryption-based methods because data need to be shared beyond a small number of authorized recipients.

PROPOSED SYSTEM

Unlike existing system, the proposed approach differs from the aforementioned research along two principal dimensions. First, it addressed re-identification in longitudinal data publishing. Second, it groups diagnosis codes together, the framework is based on grouping of records, which has been shown to be highly effective in retaining data utility due to the direct identification of records being anonymize.

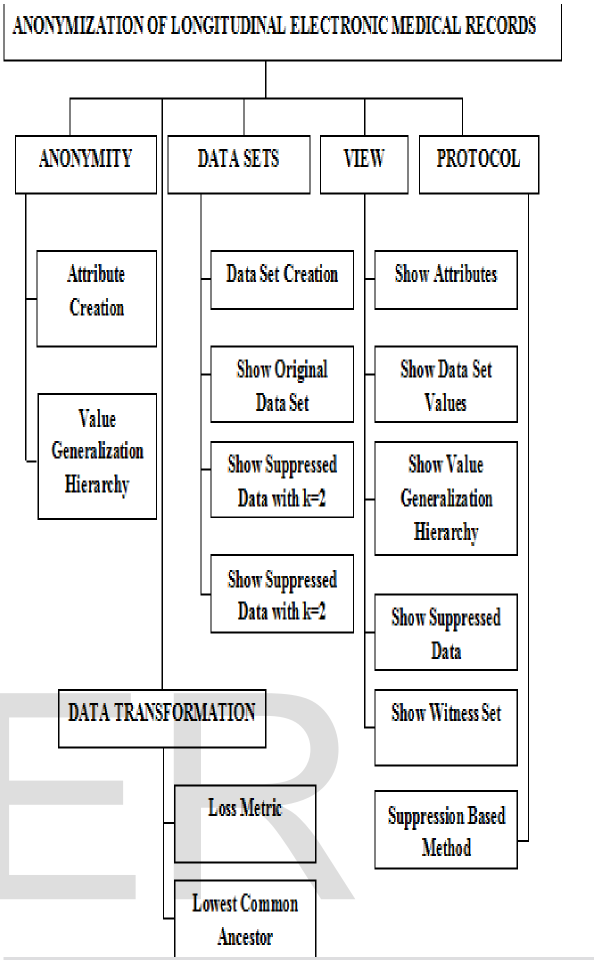
Advantages Of Proposed System

Sharing patient data while providing computational privacy guarantees. The approach uses sequence alignment and clustering-based heuristics to anonymize longitudinal patient records.

Loss Metric

In this form, Information Loss occurred by replacing a node with its ancestor is found out. The No. of leaf nodes in the sub tree rooted by Ai, No. of leaf nodes in the sub tree rooted by Aj and Domain Size of Attribute A are details given as input in this form. The loss metric value is the output obtained from this form.

SYSTEM FLOW DIAGRAM



CONCLUSION

This work was motivated by the growing need to disseminate patient- specific longitudinal data in a privacy- preserving manner. It introduces the first approach to sharing such data while providing computational privacy guarantees.

REFERENCES

[1] L. Sweeney, “k-anonymity: A model for protecting privacy,” International Journal of Uncertainty, Fuzziness and Knowledge- Based Systems, vol. 10, no.

5, pp. 557–570, 2002.

[2] B. C. M. Fung, K. Wang, R. Chen, and P. S. Yu, “Privacy-preserving data publishing: A survey of recent developments,” ACM Computing Surveys, vol. 42, no. 4, pp. 1–53, 2010.

[3] B.-C. Chen, D. Kifer, K. LeFevre, and A. Machanavajjhala, “Privacy-preserving data publishing,” Foundations and Trends in Databases, vol. 2, no. 1-2, pp. 1–167, 2009. R. S. Sandhu, E. J. Coyne, H. L. Feinstein, and C. E. Youman, “Role-based access control models,” IEEE Computer, vol. 29, no. 2, pp. 38–47, 1996.

[4] B.Pinkas, “Cryptographic techniques for privacy-preserving data mining,” ACM Special Interest Group on Knowledge Discovery and Data Mining Explorations, vol. 4, no. 2, pp. 12–19, 2002