Enhanced Safe and Secure Distributed Healthcare System Through Efficient Duplication Detection Approach

J. Prassanna¹, Jani L. Anbarasi⁵, R. Prabhakaran¹, Kanchana Devi V³, Rajiv Vincent¹, R. Manikandan², Ambeshwar Kumar²

School of Computer Science and Engineering (SCOPE), VIT Chennai Campus, India; ‘School of Computing, SASTRA Deemed University, Thanjavur, India.

ABSTRACT

Introduction: Currently the Manual Health Care System works on a Referral basis where a Village Patient is referred to District Level Hospital for further treatment and then to State/National Level Hospitals if not cured. The patient needs to carry all the medical prescription records manually for each level and there is no data security. Missing of any data will lead to undergo a tedious process.

Objectives: This study aims to analyse the drawbacks in the existing healthcare system in the rural areas and proposed an effective secured assistive health care system for rural areas.

Methods: Local village structure and the patient information were studied and an application is developed to store the patient's information in a secured way. The secured data is distributed to the state and national level hospitals during treatment. This allows the early diagnosis and treatment of a person in villages.

Results: The system helps to monitor and update the entire patient record securely using encryption. Avoids duplicated file processing and stores the data in Distributed Servers by enabling high maintenance and security thereby maintaining better health care system.

Conclusion: The proposed distributed system will reduce the manual work and help to maintain the patient record in sync at the village-district-state-national level. This distributed system is a procedural healthcare system. It satisfies the patient requirements and it stores the prescription data and information about patients in a distributed system for easy maintenance of data. This system has enabled the security system with auditing files, which are stored in distributed servers.

Key Words: e-Health Care Management, Distributed Storage, Patient Information, Progressive Sorted Neighbourhood Method

INTRODUCTION

Nowadays, when a person is affected by some disease, getting proper treatment and to recover is a tedious process. This process may be easy for people living in metro cities with high medical facilities. But for people living in rural areas with a less medical facility, it is a nightmare for them. In a situation where someone falls sick in a rural area, with all factors kept constant, the village health teams attend to the patient first and if they cannot handle the situation, the person is referred to a health centre ii which is up to one level and up the chain till the national referral hospital in case, the patient is not cured. The manual healthcare facility of the current villages is shown in figure 1.

Figure 1: Manual of Existing Health Care System.
The hierarchy where the rural village people reach out for medical treatments are detailed below:

A. Village Health Teams: The immediate contact for people living in a rural area would be Village Health Team I. Each village is supposed to have this Village Health Team I but in some cases, they do not have some drugs/lab facility for treating common diseases. This is due to limited transportation to the inner areas of the village.

B. Village Health Team II: A health Team II facility, serves little better than Health Team I, should be able to treat common diseases like malaria, typhoid etc. It is to be led by a Nurse, working with a midwife, two nursing assistants and a health assistant by Village Committee.

C. Health Centre III (District Level): These centres are at District Level, led by a senior clinical officer. This will have a general outpatient clinic where patients get treatment for common diseases. This will also have a maternity ward with a Gynaecologist. It should also have a functioning laboratory. This centre will have most of the equipment for treating minor diseases.

D. Health Centre IV (State Level): This level of health facility serves at State Level. A health centre IV is a hospital. This should have the services found at health centre III, but in addition to this, it will have wards for admitting patients. It should have a senior Doctor and an operation theatre for carrying out emergency operations.

E. Health Centre V (National/International Level): At the top of the healthcare, the chain is the national referral hospital. This is the place where the end to end treatment is provided to the Patients by Brainy Medical Experts. Major/Complex diseases which cannot be treated at Village/District/State Levels will be successfully treated.

At each of the levels, personal and health records are obtained from the patients. In a case where a patient is referred to a higher level, the records and history have to be re-obtained. This in turn slows down the whole treatment process. This health care work aims at uploading patients’ records and disease and prescription details. Each patient’s personal and disease details who visit the village health care will be uploaded to the system. The purpose of the project is to completely shut down manual process and automate in the system. The purpose of the project is to develop a health care project with user-friendly that can be used by a common health assistant. The health assistant will be able to enter the patient’s data and disease details and store it in the database. The system will validate to enter all patient fields as those are mandatory.

The Distributed Health Care will have Patient’s data distributed across other Level Hospitals. This makes the process easy as the patient’s data can be fetched anywhere and anytime without any manual process. Since the Patient’s data is distributed across all systems it is important to maintain security and patient’s medical records should be kept confidential. In the proposed Distributed Health Care scheme, Security is maintained using Encryption and Decryption techniques. It allows only authorized persons to update any patient’s record and the duplicate Records are discarded when the user tries to modify the Prescription details. If the patient’s records are not duplicated then the data is stored in the database. By doing this deduplication is avoided and data redundancy is maintained. The purpose of the project is to overcome the below limitations of Existing Healthcare Process.

- In VHC, Patient Details are maintained in Files/Papers which leads to loss of data
- Information Sharing Difficulty. Village – District – State – National Level Hospitals cannot share the patient’s data.
- District/State/National Level hospitals maintain a System to upload patient records but the data are not distributed across levels.
- There is no Security for Patient Data. Anyone can access the Health Care Application and updates the patient’s data which leads to Security Violation.
- Patient’s Prescription data entered can be duplicated by anyone which causes serious Illegal problems

The centralized systems will be handled by an admin at each hospital levels (village/district/state/national). This health care system is cost-effective, user friendly, simple and speedy. It deals with the collection of a patient’s personal information and disease details. Traditionally, all these work are done manually which slows down the entire process. The main function of the system is to connect to village-district-state-national level hospitals using distributed servers. Each level hospital is maintained using authorized login and view/ update patient details and retrieves these details as and when required. This health care system input contains patient personal details, disease details; while system output is to get these details on key request. The project detects duplicate data entries and discards the data without saving into the system. This makes the system more secure thereby maintaining the history of the patient/disease details in a secured manner without allowing illegal activities.

**RELATED WORK**

Hospitals can also be seen as high-tech and information-intensive organizations. The use of a distributed network for the solution of computer problem is often referred to in distributed computing. In distributed computing, a problem is divided into several tasks, where each are solved by one or more computers that transmit messages to each other.
The method of recognizing multiple versions of equivalent real-world entities is duplication detection. Today, duplicate detection methods need to process larger and larger datasets in an increasingly short period. Progressive duplicate detection algorithms that significantly help us to efficiently identify the duplicates. This algorithm works perfectly even when the execution time is less. By reporting most results well in advance of traditional approaches, they maximize the overall process gain.

PSNM includes input features like a data reference D, Maggie sorting algorithm, minimum and maximum window size W and Iteration I. This helps to identify the close promising neighbours. The quality of data is therefore often affected by many factors. Two new progressive duplicate algorithms are processed in the proposed work which works best in small and almost clean datasets. The PSNM sorts the records using a sorting key. The key is calculated through the extraction of appropriate attributes or sections of the values of the attributes. A proper discriminatory authority should be given for identifying records which are considered to be duplicates.

To manage large data sets, a fixed size W window is used to process over the dataset in a sequence. Compared with the previous w−1 records, any new record enters the window to find matching records. Usage of a multi-pass system to sort data sets with different sort key for each one-pass not once, but multiple times.

Levenshtein is the best known character-based similarity metric defined as the minimum number of inserts, removals, or replacements needed for translating one string to the next. Although character-based measurements work well to the approximate distance between strings that vary due to typographical mistakes or abbreviations, they are less expensive and accurate in the case of larger strings. If several words are added, deleted or transposed, the differences between equivalent strings are expressed, character-based metrics can assign high costs to non-corresponding string segments producing low similarity values for strings with a couple of words.

**PROPOSED SYSTEM**

The proposed system is a distributed system where patients can easily follow up and the patient’s records that remain distributed across different hospital levels (village/district/state/national). These systems share the resources locally as well as globally. In the health care distributed systems the clients are nothing but the different local hospitals which in turn are distributed across all global level hospitals. Since the medical records are distributed across all hospitals, there will be openness in the data resource sharing which helps to reduce the manual tedious process. The proposed scheme, patient’s records are kept in sync if the patients are referred to the next level hospital which is easy to follow up. Their history is easily obtained and health personnel can easily make decisions based on the accessed record. This also saves time because the records do not have to be obtained from the patient afresh. On another note, if the patient has to store their records personally, there is no guarantee that they will keep them safe so the patient’s records are always available at hand when required.

When a village health care assistant adds patient record it is added to the distributed servers using encryption key. When the high-level health care system need to search and view the patient personal/medical record, it can be accessed only after requesting a key from the admin user. The admin user has to activate the key and the decryption key will be sent to the authorized persons. Using this decryption key the data can be viewed/prescription can be updated. If anyone tries to duplicate/modify the patient’s prescription data then that can be detected and the data modification can be discarded. This is done using progressive duplicate detection algorithms namely progressive sorted neighbourhood method (PSNM) which performs best on small and almost clean datasets.

If the district/state/national level user tries to upload/duplicate a patient medical record, the system validates and avoids data insertion/modification. The admin monitors the entire file uploading and downloading behaviour. If it finds any duplicated file processing, it disagrees with the files to store. This ensures that the patient’s secured records are not modified by unauthorized professionals. Only authorized admin user can modify the patient’s record. The patient records are stored in distributed systems with multiple level storage based on the hospital level to maintain the integrity of data. Since the patient’s records are stored in distributed server this benefits in sparing efforts on heavy data maintenance and management. This proposed system maintains the patient information with data integrity and also enables high maintenance and better health care system which is designed especially for serving rural areas in any country. With this distributed health care system in place, the aged patients will be benefited as they don’t need to carry every medical record manually. Since this is a secured system there are no chances of patient’s record leakage. The Figure 2 shows the use of Encryption and Decryption Keys in the Distributed Health Care System.

In this Distributed Health Care System, the data will be stored in Distributed Servers. All the uploaded patient’s records are uploaded to Distributed Servers in an Encrypted Format. We have Distributed Servers at each Level. When a patient is referred to the next level hospital, the patient’s record can be fetched from the referred Level.
• Village/District Level - Distributed Server 1
• State Level - Distributed Server 2
• National Level - Distributed Server

2. When a key request is given by the national user, the admin will activate key using server 3.

RESULTS AND DISCUSSION

This distributed health care project allows patients to follow up medical records easily and to maintain records. With this system, patients' records are kept in sync if the patients are referred to the next level hospital which is easy to follow up. Their history is easily obtained and health personnel can easily make decisions based on the accessed record. This also saves time because the records do not have to be obtained from the patient afresh. On another note, if the patient has to store their records personally, there is no guarantee that they will keep them safe so the patient’s records are always available at hand when required. In this distributed health care system, the data will be stored in distributed servers. All the uploaded patient’s records are uploaded to distributed servers in an encrypted format. We have distributed servers at each level. When a patient is referred to the next level hospital, the patient’s record can be fetched from the referred level.

When a doctor has prescribed wrong medicines to patients which leads to serious side effects, the doctor can misuse the authorized logins and update the patient data in the system. To avoid these kinds of illegal activities, this system is implemented with high security by using the key for every file processing. When a patient record is duplicated by any level user, the duplicated data will be discarded. The details will be updated only when there is no duplication/modification. This proposed system maintains the patient information with data security and also enables high maintenance and a better health care system. Since this is a highly secured system there are no chances of patient’s record leakage. This system is designed especially for serving rural areas in any country. With this distributed health care system in place, many rural areas will be benefited as the patients don’t need to carry every medical record manually.

CONCLUSION AND FUTURE WORK

This distributed health care system acts as data storage using distributed servers. Patient’s records are uploaded to distributed servers in an encrypted format. We have distributed servers at each level. When a patient is referred to the next level hospital, the patient’s record can be fetched from the referred level. This distributed system has secured login for all health care centre admin user and the data can be protected using encryption and decryption keys. When a village health care assistant adds patient record it is added to the distributed servers using the encryption key. When the high-level health care system need to view the patient personal/medical

AUTHENTICATED PROCESS FLOW OF THE PROPOSED SYSTEM

The Patient name, Mobile-no, Date of birth, Blood group, basic information’s like Sex, Height, Weight, location and Symptoms are updated and stored in the Distributed Servers. The uploaded file is encrypted using the BEM (Bit Exchanging Method). For downloading each file the request has to be sent which is received by the Auditor and verified and the Decryption Key is sent to the Authorized User. District level hospital has the function of monitoring details of incoming patients and also it can update the prescription and status of the patient. But it needs to get Decryption Key from Admin to get patient’s details. It has following fields: Patient name, Email-Id, Mobile_no, Dateofbirth, Age, Blood group, BP, Weight, Sex, Symptoms, Patient location, Health specialist, Hospital code, Hospital Name, prescription and status. If the patient is not cured at this level, it will forward it to the next State Level hospital.

State-level hospital has the function of monitoring details of incoming patients and also it can update the prescription and status of the patient. But it needs to get decryption key from admin to get patient’s details. If the patient is not cured at this level, it will forward it to the next national level hospital. The national level hospital has the function of monitoring details of incoming patients and also it can update the prescription and status of the patient. The patient will be cured and the case can be closed. Admin has to log in with their username and password. The key activation process is done by this admin user. When a key request is given by district user, the admin will activate key using server 1. When a key request is given by state user, the admin will activate key using server

Figure 2: Proposed Distributed System using Encryption/Decryption Keys.
record, it can be accessed only after requesting a key from
the admin user and then using decryption key which is sent
to the authorized person. If a person tries to duplicate/modify
the patient’s prescription data then that can be detected and
the data modification can be discarded. Future work for this
project can be to implement more fields like many search
filters, patient’s lab report, pharmacy bills, administrative
details, reports finding the number of people affected with
a disease in a particular location or year etc. This will be
helpful to maintain the complete medical record. The ma-
jor innovation of this project in future is the storage of this
distributed health care system can be moved to the cloud.
Since this current distributed framework is security-based,
duplicate files can be identified by an efficient load-balanced
cloud environment for data integrity maintenance.

ACKNOWLEDGEMENT

Authors acknowledge the immense help received from the
scholars whose articles are cited and included in references
to this manuscript. The authors are also grateful to authors/
editors / publishers of all those articles, journals and books
from where the literature for this article has been reviewed
and discussed.

Conflict of Interest

The Author(s) declare(s) that there is no conflict of interest.

REFERENCES

1. http://shodhganga.inflibnet.ac.in/bitstream/10603/9915/6/06_
chapter%201.pdf

2. Coulouris GF, Dollimore J, Kindberg T. Distributed systems:
concepts and design. Pearson Education; 2005.
3. Yan S, Lee D, Kan MY, Giles LC. Adaptive sorted neighbour-
hood methods for efficient record linkage. InProceedings of the
7th ACM/IEEE-CS joint conference on Digital libraries 2007
Jun 18 (pp. 185-194).
4. Ong S, Pei A. A Comparative Study of Record Matching Algo-
rithms. Pearson Edu. 1998; 8(6): 361-7.
5. Papenbrock T, Heise A, Naumann F. Progressive duplicate
detection. Transac Knowl Data Engi 2014;27(5):1316-29.
6. Levenshtein VI. Binary codes capable of correcting dele-
tions, insertions, and reversals. InSoviet Physics Doklady
1966;10(8):707-710.
7. Dhanya D, Vidhya R, Narendra M, Davamani A, Anbarasi LJ.
Consistency and Reliability of Services in Cloud Environment.
In2017 Int Conf Current Trends in Computer, Electrical, Elec-
tronics and Communication (CTCEEC) 2017; 8:747-751.
8. Raj B, Anbarasi LJ, Narendra M, Subashini VJ. A New Trans-
formation of 3D Models Using Chaotic Encryption Based on
Arnold Cat Map. In Int. Conference on Emer. Internet Working,
Data & Web Technologies 2019; 322-332.
9. Prassanna J, Venkataraman N. Threshold based multi-objective
memonic optimized round Robin scheduling for resource efficient
load balancing in cloud. Mobile Net Appl 2019;24(4):1214-25.
10. Prassanna J, Venkataraman N. Adaptive regressive holt–winters
workload prediction and firefly optimized lottery scheduling for
load balancing in cloud. Wireless Networks. 2019:1-9.
11. Mohanrao M, Karthik S. Intelligent data mining principles with
privacy preserving procedures. In2017 IEEE International Con-
ference on Power, Control, Signals and Instrum. Engg. (ICPCSI)
2017:2513-2516.
12. Kannan RR, Abarna DR, Aswini G, Hemavathy P. Effective pro-
gressive algorithm for duplicate detection on large dataset. Int J
Sci Res 2016;2(2):105-10.