



SECURING DATA WITH BLOCK CHAIN AND AI

S Rameshbabu¹, Sainath Suryavanshi², Mohammed Arifuddin Sahil³, Shivarathri Ravi⁴,
Singireddy Sriramreddy⁵

^{2,3,4,5} UG Scholars, Department of CSE, AVN Institute of Engineering and
Technology, Hyderabad, Telangana, India.

¹ Assistant Professor, Department of CSE, AVN Institute of Engineering and Technology,
Hyderabad, Telangana, India.

ABSTRACT

Data is the input for various artificial intelligence (AI) algorithms to mine valuable features, yet data in Internet is scattered everywhere and controlled by different stakeholders who cannot believe in each other, and usage of the data in complex cyberspace is difficult to authorize or to validate. As a result, it is very difficult to enable data sharing in cyberspace for the real big data, as well as a real powerful AI.

In this paper, we propose the *SecNet*, an architecture that can enable secure data storing, computing, and sharing in the large-scale Internet environment, aiming at a more secure cyberspace with real big data and thus enhanced AI with plenty of data source, by integrating three key components:

- 1) Block chain-based data sharing with ownership guarantee, which enables trusted data sharing in the large-scale environment to form real big data;
- 2) AI-based secure computing platform to produce more intelligent security rules, which helps to construct a more trusted cyberspace;
- 3) Trusted value-exchange mechanism for purchasing security service,

providing a way for participants to gain economic rewards when giving out their data or service, which promotes the data sharing and thus achieves better performance of AI.

Moreover, we discuss the typical use scenario of SecNet as well as its potentially alternative way to deploy, as well as analyze its effectiveness from the aspect of network security and economic revenue.

INTRODUCTION

About the Project:-

With the development of information technologies, the trend of integrating cyber, physical and social (CPS) systems to a highly unified information society, rather than just a digital Internet, is becoming increasingly obvious. In such an information society, data is the asset of its owner, and its usage should be under the full control of its owner, although this is not the common case.

Given data is undoubtedly the oil of the information society, almost every big company wants to collect data as much as possible, for their future competitiveness. An increasing amount of personal data, including location information, web-searching behavior, user calls, user preference, is being silently collected by the built-in sensors



inside the products from those big companies, which brings in huge risk on privacy leakage of data owners. Moreover, the usage of those data is out of control of their owners, since currently there is not a reliable way to record how the data is used and by who, and thus has little methods to trace or punish the violators who abuse those data. That is, lack of ability to effectively manage data makes it very difficult for an individual to control the potential risks associated with the collected data. For example, once the data has been collected by a third party (e.g., a big company), the lack of access to this data hinders an individual to understand or manage the risks related to the collected data from him. Meanwhile, the lack of immutable recording for the usage of data increases the risk to abuse them.

If there is an efficient and trusted way to collect and merge the data scattered across the whole CPS to form real big data, the performance of artificial intelligence (AI) will be significantly improved since AI can handle massive amount of data including huge information at the same time, which would bring in great benefits (e.g., achieving enhanced security for data) and even makes AI gaining the ability to exceed human capabilities in more areas. According to the research, if given large amount of data in an order of magnitude more scale, even the simplest AI algorithm currently (e.g., perceptrons from the 1950s) can achieve fanciest performance to beat many state-of-the-art technologies today. The key lies in how to make data sharing trusted and secured. Fortunately, the blockchain technologies may

be the promising way to achieve this goal, via consensus mechanisms throughout the network to guarantee data sharing in a tamper-proof way embedded with economic incentives. Thus, AI can be further empowered by blockchain-protected data sharing. As a result, enhanced AI can provide better performance and security for data.

In this paper, we aim at securing data by combining blockchain and AI together, and design a **Secure Networking** architecture (termed as **SecNet**) to significantly improve the security of data sharing, and then the security of the whole network, even the whole CPS.

In SecNet, to protect data, one of the biggest challenges is where and how to store data, because users have to give their data to service providers if they want to use certain services or applications. This is caused by the inherent coupling of user data and application in current service mechanisms, which significantly hinders the development of data protection and application innovation. Inspired by the concept of Personal Data Store (PDS) from openPDS and the Private Data Center (PDC) from HyperNet, *SecNet* finally inherits and adopts PDC instead of PDS, as PDC is more suitable to deploy and to deal with this problem, since it provides more secure and intelligent data storage system via physical entities instead of software-based algorithms as in openPDS. Each PDC actually serves as a secured as well as centralized physical space for each SecNet user where his/her data lives in. Embedding PDC into SecNet would allow users to monitor and reason about what and why their data is used as well as by who, meaning



the users can truly control every operation on their own data and achieve fine-grained management on access behaviors for data. Actually, besides PDC, other choices can also be applied for the data storing in SecNet according to certain requirements.

The trust-less relationship between different data stakeholders significantly thwarts the data sharing in the whole Internet, thus the data used for AI training or analyzing is limited in amount as well as partial in variety. Fortunately, the rise of Blockchain technologies bring in a hopeful, efficient and effective way to enable trust data sharing in trust-less environment, which can help AI make more accurate decisions due to the real big data collected from more places in the Internet. SecNet leverages the emerging blockchain technologies to prevent the abuse of data, and to enable trusted data sharing in trust-less or even untrusted environment. For instance, it can enable cooperations between different edge computing paradigms to work together to improve the whole system performance of edge networks. The reason why blockchain can enable trusted mechanisms is that it can provide a transparent, tamper-proof metadata infrastructure to seriously recode all the usage of data.

Thus, SecNet introduces blockchain-based data sharing mechanisms with ownership guarantee, where any data ready for sharing should be registered into a blockchain, named Data Recording Blockchain (DRB), to announce its availability for sharing. Each access behavior on data by other parties (not the data owner) should also be validated and

recorded in this chain. In addition, the authenticity and integrity of data can only be validated by DRB as well. Besides, SecNet enables economic incentive between different entities if they share data or exchange security service, by embedding smart contract on data to trigger automatic and tamper-proof value exchange. In this way, SecNet guarantees the data security and encourages data sharing throughout the CPS. Furthermore, data is the fuel of AI, and it can greatly help to improve the performance of AI algorithms if data can be efficiently networked and properly fused. Enabling data sharing across multiple service providers can be a way to maximize the utilization of scattered data in separate entities with potential conflicts of interest, which can enable a more powerful AI. Given enough data and blockchain-based smart contract on secure data sharing, it is not surprised that AI can become one of the most powerful technologies and tools to improve cyber security, since it can check huge amount of data more quickly to save time, and identify and mitigate threats more rapidly, and meanwhile give more accurate prediction and decision support on security rules that a PDC should deploy. Besides, embedded with Machine Learning inside, AI can constantly learn patterns by applying existing data or artificial data generated by GAN to improve its strategies over time, to strengthen its ability on identifying any deviation on data or behaviors on a 24/7/365 basis. SecNet can apply these advanced AI technologies into its Operation Support System (OSS) to adaptively identify more suspicious data-related behaviors, even they are never seen



before. In addition, swarm intelligence can be used in SecNet to further improve the data security, by collecting different security knowledge from huge amount of intelligent agents scattered everywhere in the CPS, with the help of trusted exchange mechanisms for incentive tokens.

LITERATURE REVIEW

LITERATURE SURVEY

1. **Hyper connected network:** A decentralized trusted computing and networking paradigm

Abstract:

With the development of the Internet of Things, a complex CPS system has emerged and is becoming a promising information infrastructure. In the CPS system, the loss of control over user data has become a very serious challenge, making it difficult to protect privacy, boost innovation, and guarantee data sovereignty. In this article, we propose HyperNet, a novel decentralized trusted computing and networking paradigm, to meet the challenge of loss of control over data. HyperNet is composed of the intelligent PDC, which is considered as the digital clone of a human individual; the decentralized trusted connection between any entities based on blockchain as well as smart contract; and the UDI platform, enabling secure digital object management and an identifier-driven routing mechanism. HyperNet has the capability of protecting data sovereignty, and has the potential to transform the current

communication-based information system to the future data-oriented information society.

2. Lightweight RFID protocol for medical privacy protection in IoT

Abstract:

Traditional medical privacy data are at a serious risk of disclosure, and many related cases have occurred over the years. For example, personal medical privacy data can be easily leaked to insurance companies, which not only compromises the privacy of individuals, but also hinders the healthy development of the medical industry. With the continuous improvement of cloud computing and big data technologies, the Internet of Things technology has been rapidly developed. Radio frequency identification (RFID) is one of the core technologies of the Internet of Things. The application of the RFID system to the medical system can effectively solve this problem of medical privacy. RFID tags in the system can collect useful information and conduct data exchange and processing with a back-end server through the reader. The whole process of information interaction is mainly in the form of ciphertext. In the context of the Internet of Things, the paper presents a lightweight RFID medical privacy protection scheme. The scheme ensures security privacy of the collected data via secure authentication. The security analysis and evaluation of the scheme indicate that the protocol can effectively prevent the risk of medical privacy data being easily leaked.



3. Amber: Decoupling user data from Web applications

Abstract:

User-generated content is becoming increasingly common on the Web, but current web applications isolate their users' data, enabling only restricted sharing and cross-service integration. We believe users should be able to share their data seamlessly between their applications and with other users. To that end, we propose Amber, an architecture that decouples users' data from applications, while providing applications with powerful global queries to find user data. We demonstrate how multi-user applications, such as e-mail, can use these global queries to efficiently collect and monitor relevant data created by other users. Amber puts users in control of which applications they use with their data and with whom it is shared, and enables a new class of applications by removing the artificial partitioning of users' data by application.

4. Enhancing selectivity in big data

Abstract:

Today's companies collect immense amounts of personal data and enable wide access to it within the company. This exposes the data to external hackers and privacy-transgressing employees. This study shows that, for a wide and important class of workloads, only a fraction of the data is needed to approach state-of-the-art accuracy. We propose selective data systems that are designed to pinpoint the data that is valuable for a company's current and evolving workloads. These systems limit data exposure by setting aside the data that is not truly valuable.

SYSTEM ANALYSIS

EXISTING SYSTEM:-

An increasing amount of personal data, including location information, web-searching behavior, user calls, user preference, is being silently collected by the built-in sensors inside the products from those big companies, which brings in huge risk on privacy leakage of data owners. Moreover, the usage of those data is out of control of their owners, since currently there is not a reliable way to record how the data is used and by who, and thus has little methods to trace or punish the violators who abuse those data. That is, lack of ability to effectively manage data makes it very difficult for an individual to control the potential risks associated with the collected data.

In cyber world everything is dependent on data and all Artificial Intelligence algorithms discover knowledge from past data only, for example in online shopping application users review data is very important for new comers to take decision on which product to purchase or not to purchase, we can take many examples like health care to know good hospitals or education institutions etc. Not all cyber data can be made publicly available such as Patient Health Data which contains patient disease details and contact information and if such data available publicly then there is no security for that patient data.

Now a days all service providers such as online social networks or cloud storage will store some type of users data and they can sale that data to other organization for their



own benefits and user has no control on his data as that data is saved on third party servers.

PROPOSED SYSTEM:-

To overcome from above issue author has describe concept called Private Data Centres (PDC) with Blockchain and AI technique to provide security to user's data. In this technique 3 functions will work which describe below

- 1) **Block chain:** Block chain-based data sharing with ownership guarantee, which enables trusted data sharing in the large-scale environment to form real big data. In this technique users can define access control which means which user has permission to access data and which user cannot access data and Blockchain object will be generate on that access data and allow only those users to access data which has permissions. In Blockchain object user will add/subscribe share data and give permission.
- 2) **Artificial Intelligence:** AI-based secure computing platform to produce more intelligent security rules, which helps to construct a more trusted cyberspace. AI work similar to human brain and responsible to execute logic to check whether requesting user has permission to access shared data. If access is available then AI allow Block chain to display share data otherwise ignore request.
- 3) **Rewards:** In this technique all users who is sharing the data will earn rewards point upon any user access his data. Trusted value-exchange mechanism for purchasing security service, providing a

way for participants to gain economic rewards when giving out their data or service, which promotes the data sharing and thus achieves better performance of AI.

IMPLEMENTATION

MODULES:

This project consists of two modules

- 1) **Patients:** Patients first create his profile with all disease details and then select desired hospital with whom he wishes to share/subscribe data. While creating profile application will create Blockchain object with allowable permission and it will allow only those hospitals to access data.

Patient Login: Patient can login to application with his profile id and check total rewards he earned from sharing data.

- 2) **Hospital:** Hospital1 and Hospital2 are using in this application as two organizations with whom patient can share data. At a time any hospital can login to application and then enter search string as disease name.

AI algorithm will take input disease string and then perform search operation on all patients to get similar disease patients and then check whether this hospital has permission to access that patient data or not, if hospital has access permission then it will display those patients records to that hospital.

Below is the code example to create
Block chain object with patient data
blockchain = Blockchain() //creating
block chain object

```
x = '{"Patient_id":"' + str(count) + "',  
"patient_name":"' + name + "',  
"age":"' + age + "',  
"problem_desc":"' + problem + "',  
"profile_date":"' + str(current_time) + "',  
"access_data":"' + str(access) + "', "gender":  
"'+ gender + '"}' //creating access with  
input data
```

```
blockchain.add_new_transaction(json.  
loads(x)) //adding transaction to  
blockchain
```

```
hash = blockchain.mine()//mining  
transaction to generate hash value
```

In above code see comment to
understand

SYSTEM DESIGN

System Architecture:-

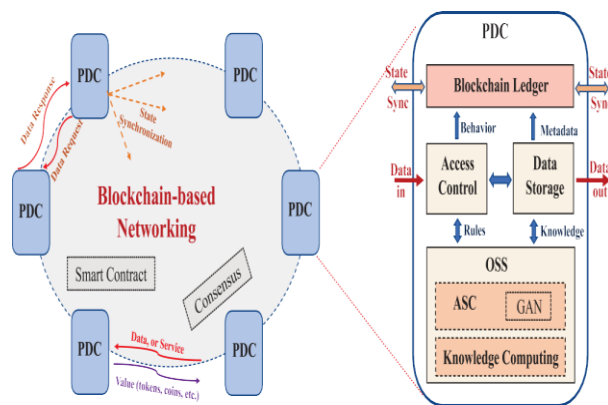


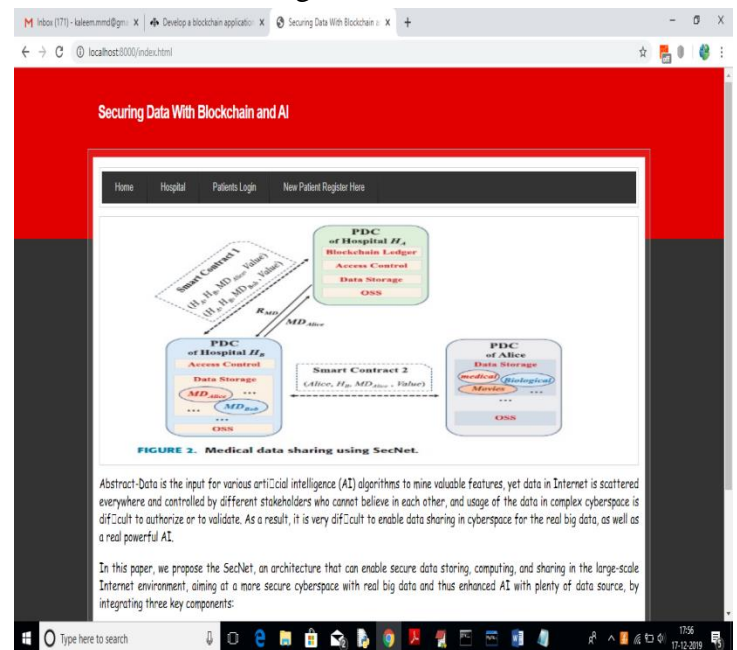
Fig. System Architecture

SCREENSHOTS

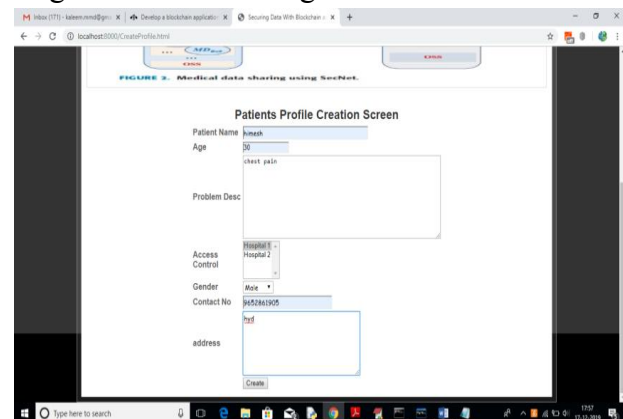
First create database in MYSQL by copying
content from 'DB.txt' file and paste in
MYSQL.

In settings file change port no from 3308 to
3306 and in 'views.py' file also change port
no to 3306

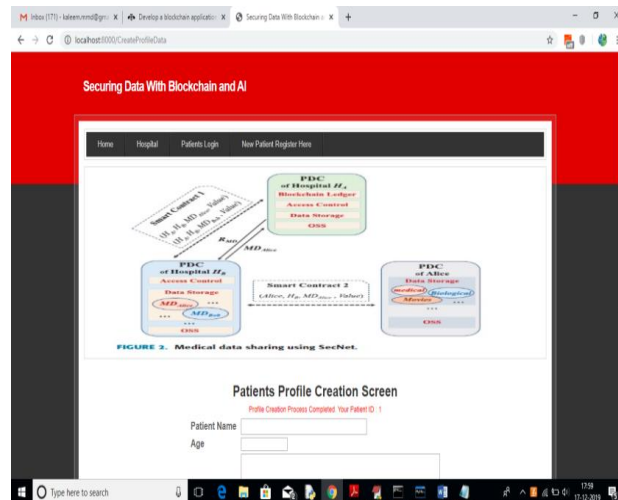
Deploy code on DJANGO and start server
and run in browser to get below screen



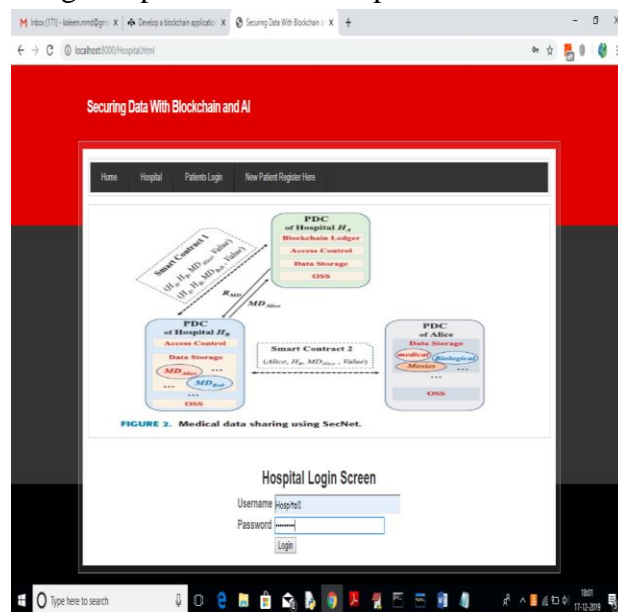
In above screen click on 'New Patient
Register Here' link to get below screen



In above screen I am adding patient disease details and selecting 'Hospital1' to share my data and if you want to share with two hospitals then hold 'CTRL' key and select both hospitals to give permission. Now press 'Create' button to create profile

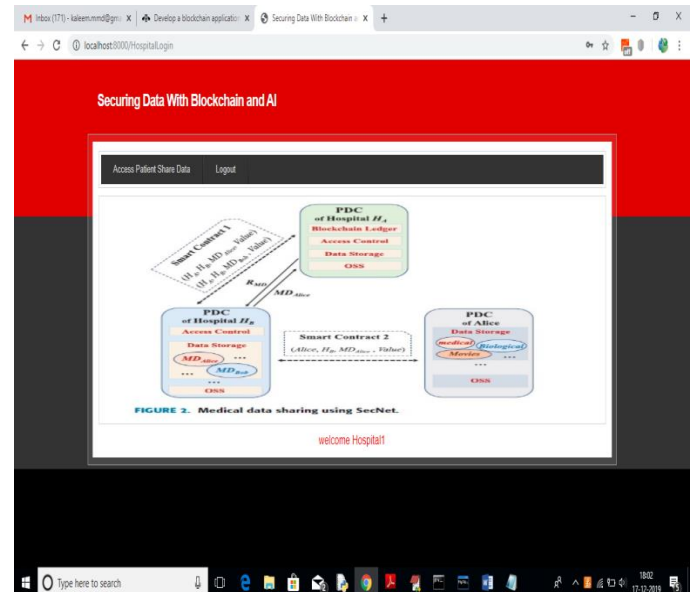


In above screen one patient is created with patient ID 1 and now Hospital 1 can login and search and access this patient data as patient has given permission to Hospital1

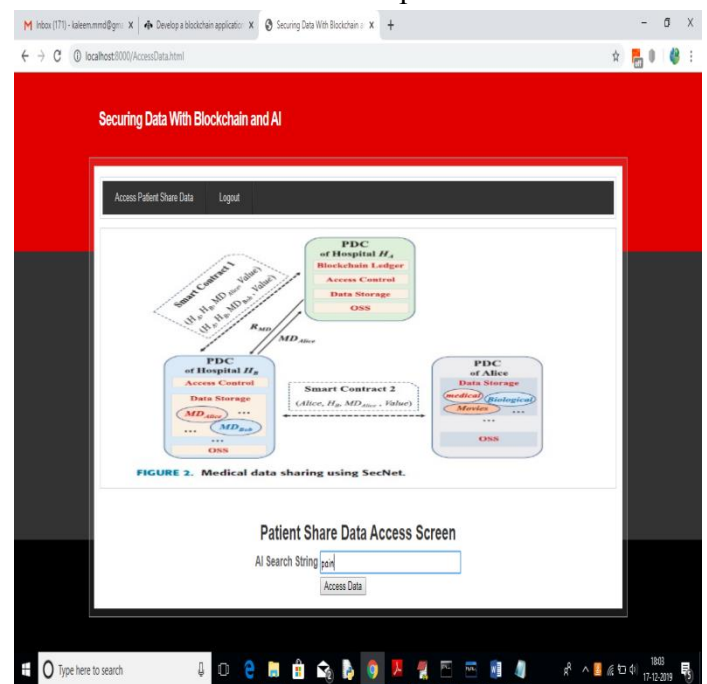


In above screen to login as Hospital1 click on 'Hospital' link to get above screen. Use

'Hospital1' as username and 'Hospital1' as password to login as Hospital1 and use Hospital2 to login as Hospital2. After login will get below screen

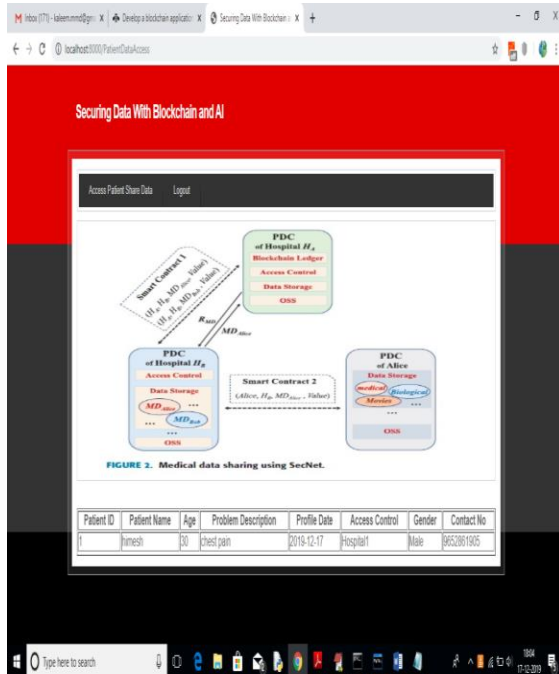


In above screen click on 'Access Patient Share Data' link to search for patient details



In above screen I want to search for all patients who are suffering from 'pain' and

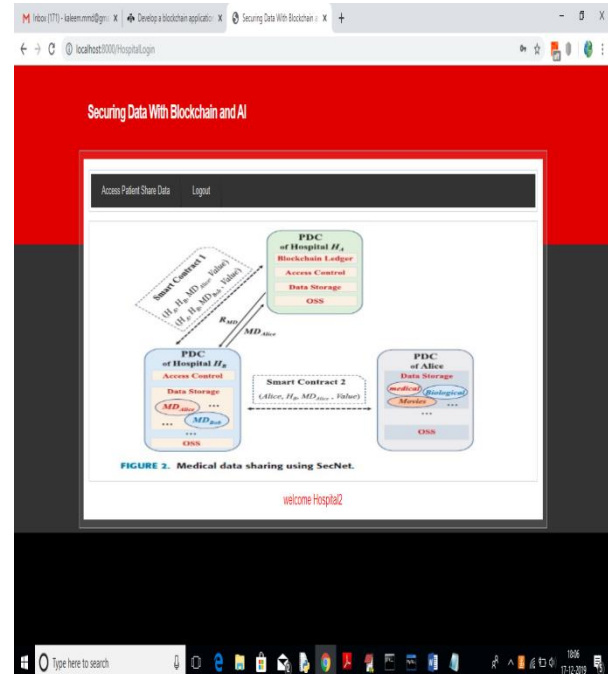
then click on 'Access data' button to get below screen



In above screen Hospital1 is getting details of patient and Hospital2 not having permission so it will not get details. To see this logout and login as 'Hospital2'



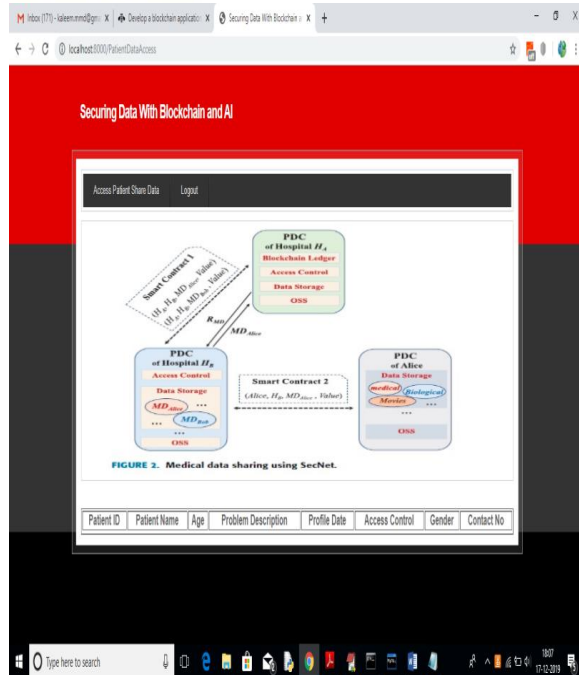
In above screen 'Hospital2' is login, after login will get below screen



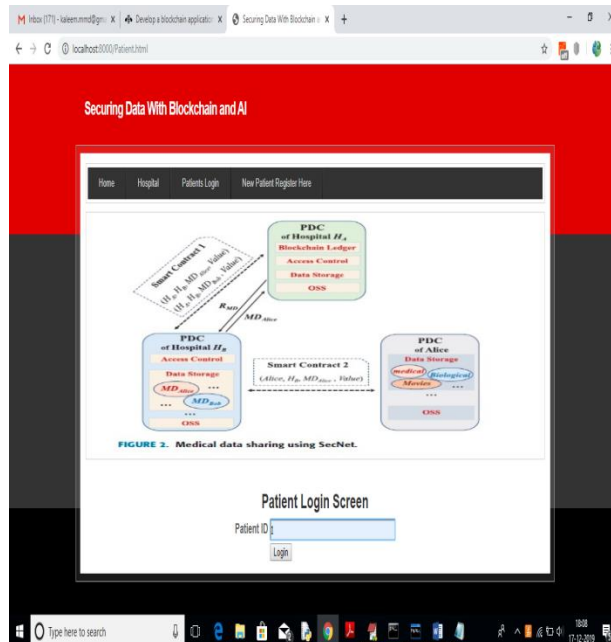
Now click on 'Access Patient Share Data' link and search for same pain disease



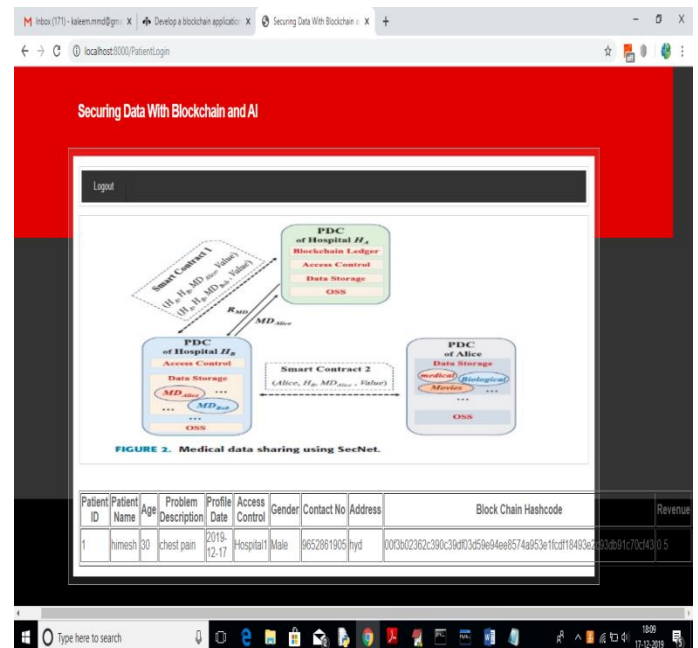
For above query will get below result



In above screen no patient details are showing as Hospital2 not having permission. So block chain allow only those users to access data who has permission. Now logout and login as patient by entering patient id in below screen



After login will get below details for patient 1



In above screen we can see patient all details and hash code generated by block chain and in last column we can see patient reward revenue as 0.5 and it will get update upon every access from hospital user.

CONCLUSION

In order to leverage AI and blockchain to fit the problem of abusing data, as well as empower AI with the help of blockchain for trusted data management in trust-less environment, we propose the SecNet, which is a new network- ing paradigm focusing on secure data storing, sharing and computing instead of communicating. SecNet provides data ownership guaranteeing with the help of blockchain tech- nologies, and AI-based secure computing platform as well as blockchain-based incentive mechanism, offering paradigm and incentives for data merging and more powerful AI to finally achieve better network security. Moreover, we discuss the typical use scenario of



SecNet in medical care system, and gives alternative ways for employing the storage function of SecNet. Furthermore, we evaluate its improvement on network vulnerability when countering DDoS attacks, and analyze the inventive aspect on encouraging users to share security rules for a more secure network.

In future work, we will explore how to leverage blockchain for the access authorization on data requests, and design secure and detailed smart contracts for data sharing and AI-based computing service in SecNet. In addition, we will model SecNet and analyze its performance through extensive experiments based on advanced platforms (e.g., integrating IPFS and Ethereum to form a SecNet-like architecture).

FUTURE ENHANCEMENT

The proposed framework could serve as a first step to understand new business models and unlock the enormous potential of the future EC ecosystem. There are several future research directions. For example, we will investigate the ME concept in the case when several edge networks cooperate with each other to form an edge/fog federation. Investigating the impacts of the strategic behavior on the efficiency of the ME is another interesting topic. Note that N. Chen et. al. have shown that the gains of buyers for strategic behavior in Fisher markets are small. Additionally, in this work, we implicitly assume the demand of every service is unlimited. It can be verified that we can add the maximum number of requests constraints to the EG program to capture the limited

demand case, and the solution of this modified problem is indeed an ME. However, although the optimal utilities of the services in this case are unique, there can have infinite number of equilibrium prices. We are investigating this problem in our ongoing work. Also, integrating the operation cost of ENs into the proposed ME framework is a subject of our future work. Finally, how to compute market equilibria with more complex utility functions that capture practical aspects such as task moving expenses among ENs and data privacy is an interesting future research direction. It is also interesting to test the performance of the proposed approach on real datasets of an EC system when EC is widely deployed.

REFERENCES

- [1] H. Yin, D. Guo, K. Wang, Z. Jiang, Y. Lyu, and J. Xing, "Hyperconnected network: A decentralized trusted computing and networking paradigm," *IEEE Netw.*, vol. 32, no. 1, pp. 112–117, Jan./Feb. 2018.
- [2] K. Fan, W. Jiang, H. Li, and Y. Yang, "Lightweight RFID protocol for medical privacy protection in IoT," *IEEE Trans Ind. Informat.*, vol. 14, no. 4, pp. 1656–1665, Apr. 2018.
- [3] T. Chajed, J. Gjengset, J. Van Den Hooff, M. F. Kaashoek, J. Mickens, R. Morris, and N. Zeldovich, "Amber: Decoupling user data from Web applications," in *Proc. 15th Workshop Hot Topics Oper. Syst. (HotOS XV)*, Warth-Weiningen, Switzerland, 2015,



IJARST

International Journal For Advanced Research In Science & Technology

A peer reviewed international journal

www.ijarst.in

ISSN: 2457-0362

pp. 1–6.

- [4] M. Lecuyer, R. Spahn, R. Geambasu, T.-K. Huang, and S. Sen, “Enhancing selectivity in big data,” *IEEE Security Privacy*, vol. 16, no. 1, pp. 34–42, Jan./Feb. 2018.
- [5] Y.-A. de Montjoye, E. Shmueli, S. S. Wang, and A. S. Pentland, “openPDS: Protecting the privacy of metadata through SafeAnswers,” *PLoS ONE*, vol. 9, no. 7, 2014, Art. no. e98790.