Capturing and Storing Aadhaar Data at Local Bodies Using Distributed Database Systems- An Alternative Approach to Implement Aadhaar In India

Amar Jeet Singh[1], Dharam Pal Singh[2]
[1]Professor, Dept of Computer Science
[2]Dept of Computer Science
[1, 2] HPU, Shimla-171005

Abstract- *Information and Communication Technology (ICT) has transformed almost every aspect of society and one of ICT's potential transformative roles is to change government functions and make governance citizen-centric. The Government of India has envisioned Aadhaar as a biometric solution to provide every resident of India a Unique Identification and to usher India to digitally empowered country. The identification data of the residents is kept in the centralized database in the current Aadhaar architecture. Keeping the data at centralized repository may pose various bottlenecks because inherently, the resident centric data is generated at geographically dispersed places where people live and it is to be used for availing government services at the place of living. This paper proposes an alternative Aadhaar architecture using distributed database approach to the existing centralized one so that resident centric data be captured, stored and used at distributed places at the points of its origin.*

Keywords- Digital India, e-Governance, Central Identity Data Repository (CIDR), Personal Identification Data (PID), Biometrics, Digital Divide, Distributed Database Systems, Locality of Reference.

I. INTRODUCTION

In India, Identification of individuals is a challenging task. In the diverse landscapes of India, tell-tale signs of identity are everywhere like amulets, headgear, signs on the forehead and dresses. These symbols may indicate the belonging to a particular region and identification of an individual in terms of the caste, creed and class. However, this doesn't establish the unique identity of a person [1]. Aadhaar under the Unique Identification Authority of India (UIDAI) is a 'flagship' program in identification process of individuals in order to meet out various obligations of Government of India towards extending welfare services to its citizens. The Supreme Court of India had allowed the governments to use Aadhaar card number for the purposes like welfare schemes and for "non-benefit" purposes like filing of income tax returns and opening of bank accounts [2]. To transform the entire system of availing public services to an electronic mode, the Government of India initiated the very ambitious program **'**Digital India' with the vision to transform India into a digitally empowered society and knowledge economy. It has centered around three key vision areas: Digital Infrastructure as a utility to every resident, government services on demand and digital empowerment of residents [3]. This requires the beneficiaries to be uniquely identified over the network. As a prerequisite for identification every beneficiary should have in place a prior data about him captured and stored and a process for identification. To provide unique identification to the residents of India, the Government of India raised Unique Identification Authority of India (UIDAI) and popularly it is known as Aadhaar. This paper discusses the identification evolution in India in section 2. The research Methodology adopted in the research is described in section 3. The process of UIDAI or Aadhaar is discussed in Section 4 along with various bottlenecks to the existing system. Section 5 elucidates problem statement. Section 6 gives the objectives to be realized. Section 7 is the proposed solution using distribution of Aadhaar data. The section 8 gives findings and concluding remarks.

II. EVOLUTION OF IDENTIFICATION SYSTEMS IN INDIA

To provide unique identification to the residents' in India, the Government of India (GoI) undertook various initiatives. In 1993, the Election Commission of India issued the photo identity cards as an identification document for an individual. Subsequently in 2003, the GoI issued Multipurpose National Identity Cards (MNIC) [4]. These initiatives remained isolated to particular themes and never acted as an integrated identification systems. There are many cases of duplicate and fake identity creations. In 2009, the Government of India launched Aadhaar by establishing UIDAI with the aim to provide every resident of India a biometric based unique identity by issuing a unique identification number (UID) to be known as Aadhaar. UIDAI has the following

components. (i) Central Identity Data Repository (CIDR),the core building block of Aadhaar which holds the personal identification data (both biometric and demographic information) of every resident of India at a single and central place. CIDR is responsible for issuing Unique Identification Numbers (Aadhaar) and to ensure de-duplication of data. (ii) Authentication User Agency (AUA) asan organization or an entity that uses Aadhaar authentication as part of its applications to provide services to residents - it includes Government Departments, banks, and other public or private organizations. (iii) Authentication Service Agency (ASA) asan organization/ entity providing secure leased line connectivity to UIDAI's data centers for transmitting authentication requests from various AUAs [5]. (iv) Enrolling agencies are the entities hired by the State Governments to perform enrolment functions which capture data about individuals near the places of their living; the places people can approach easily i.e. schools, colleges, market places, health centers, panchayat offices etc. [6]. These agencies will directly interact with residents and enroll them into the CIDR. (v) Aadhaar holder is the person who has enrolled himself/ herself already in to the UIDAI and contains the valid Aadhaar Number.

III. RESEARCH METHODOLOGY

The domain of research is the Unique Identification of citizens of India in an efficient and timely manner. Unique Identification of individuals over the network is indispensable in order to provide e- Governance services to the citizens of the country. The data were collected from various secondary sources such as books, journals, annual reports, websites, articles, online news, Government of India websites and Government of Himachal Pradesh website. The current system of identification of individuals i.e. Aadhaar (the centralized system) has been studied and various shortcomings toward the identification of individuals were documented. Keeping in view the aspects such as physiography of India, the digital divide within the country; a new decentralized architecture of Aadhaar has been proposed (Figure 3) and a description of the identification system with new decentralized architecture is laid down in this research paper and for this the Distributed Database Systems have been taken in to account.

IV. AADHAAR AS AN IDENTIFICATION SYSTEM WITH BIOMETRICS

In contrast to the photo identity cards and MNIC, Aadhaar has evolved with the applications of biometrics as a solution to the identification process. Biometric recognition refers to the automatic recognition of individuals based on their physiological and/or behavioral characteristics. By using biometrics, it is possible to confirm/ establish an individual's identity based on "who he/she is", rather than by "what he/she possesses" (e.g., an ID card) or "what he/she remembers" (e.g., a password) [7]. The biometric systems are widely employed for the authentication of individual's identity using their unique physiological/ biological characteristics. A biometric is a trait that is part of us, rather than something we know. Some examples include our height, weight, the shape of our hand, the pattern of our voice, veins, retina or iris, our face and the fingerprints [8]. With the Government of India initiatives towards Digital India, the biometrics is massively used for authentication rather than identification. These two terms, authentication and identification seems to be the same; however, both serve the different purposes. In order to provide the welfare services of the Government or to give access to hardware/ software remotely in an electronic mode, everything requires first, the answer to the question 'Am I who I say I am?' in 'yes or no' with high degree of accuracy. The process to answer this type of question is called 'Authentication'. It works by comparing an actual physical biometric data as read by a scanner at any terminal on the network with a previously stored piece of biometric data in the main data repository. Biometric authentication performs the same function as a password or PIN number. However, it is deemed to be more secure and accurate as it involves measurements performed on a physical characteristic, which is hard to fake. Authentication is therefore 'one to one' comparison between scanned and stored data. When a new person is to be enrolled, the answer to the question 'who am I?' is sought with high degree of accuracy to ensure de-duplication of information in the main data repository. The process to answer this type of question is called 'Identification'. The biometric data of the new enrollee captured at any terminal is compared with every single biometric data already stored in the main centralized data repository. Identification therefore, involves a long series of authentications. To enroll a new person in to the identification system requires the answer to the question 'Who am I?' to be 'You are not known'. Once such an answer has been given, a person's biometric data can be added to the library of identities in the main data repository. Identification is therefore 'one to many' comparisons between scanned and stored data [9].

4.1 ENROLMENT PROCESS IN UIDAI

Enrolling agencies enroll the residents in to UIDAI for providing them Aadhaar, the unique identity number (Figure 1). The Personal Identification Data of the residents is captured and is sent to the centralized database CIDR using ASA and the enrollee is assigned an enrollment number at the time of enrollment. The Personal Identification Data is processed for authentication at CIDR and the Aadhaar is delivered after a couple of days or months sometimes; via

Department of Posts or alternatively through e-Aadhaar from the resident portal of UIDAI.

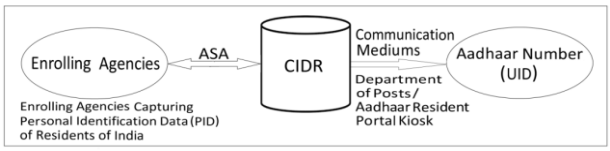

Figure 1.Enrolment Process in UIDAI [4].

4.2 AUTHENTICATION PROCESS

Aadhaar involves the authentication of individual's identity with the applications of biometrics. In authentication process, the Aadhaar Number, along with other attributes (demographic /biometrics/ OTP) is submitted to CIDR for ensuring the authenticity of the credentials of an individual (Figure 2). The CIDR verifies whether the data submitted matches the data available in CIDR and responds with an answer in "Yes or No". No personal identity information is returned as a part of the response [5]. In order to give financial benefits to the residents under the government welfare services, the process of authentication is carried out by various Authentication User Agencies all over the country and these agencies sought for the quick response in 'Yes or No' from the CIDR to accomplish the job. Also, to enroll residents in to the identification system, a series of the authentication processes are required to be accomplished.

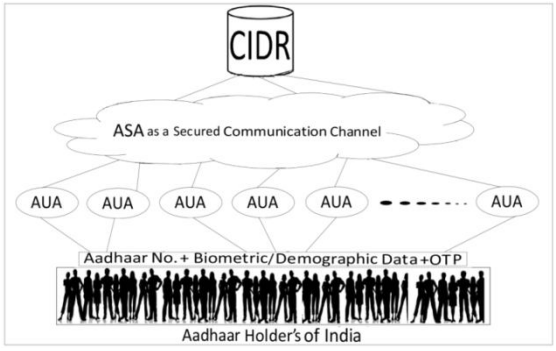

Figure 2.Authentication Process in UIDAI [5].

4.3 BOTTLENECKS

In the process of enrolment and authentication, in practice there may arise a large number of bottlenecks.

4.3.1 POOR NETWORK CONNECTIVITY

For Aadhaar to make it completely online system; the internet connectivity has to be 24x7. However, the internet connectivity in many parts of rural India is absent and the cities of India are also suffering with the poor internet connectivity [10]. The studies have revealed that only 20% of the population of India is connected to the internet [11]. The National Sample Survey Organization (NSSO) in its report on expenditure in 2009-10, said that in India only about 0.4 percent of rural households had access to internet at homes as compared to about six percent of urban households, showing the great digital divide in India [12].

4.3.2 RIOTS, MILITANT ATTACKS

These are the manmade situations and such situation damage the state of Internet in the affected areas and causes the internet shutdown. The law enforcement agencies many a times, suspend the services of Internet in order to control the situations and to maintain law and order [13].

4.3.3 NATURAL DISASTERS

Natural disasters result in more downtime than the hackers. A report published by the European Union Agency for Network and Information Security (ENISA) revealed that outages due to storm and heavy snowfall lasted the longest, around 36 hours on an average [14]. Even the Google has not been spared by these calamities; the power loss due to lightning strikes at a power grid caused the data loss from disks of one of the data centers of Google at Belgium and was the cause of four-day cloud outage [15]. Floods, tsunamis and earth quakes hit India several times causing huge destruction in terms of lives and infrastructure. This leaves the affected regions deprived from the internetconnectivity with the remaining world.

4.3.4 HARDWARE FAILURE

Another cause of the service disruption is the hardware failure. In 2011, Amazon's Cloud crash disaster held down the sites of dozens of high profile companies for days and permanently destroyed many customers' data. However, the size of the data destroyed was apparently small relative to the data stored by Amazon, but a small loss on a percentage basis for Amazon, obviously, could be ruinous for numerous companies. Moreover, this may be the end of some of the companies [16].

4.3.5 POWER SUPPLY FAILURE

In April 2015, up to 1.5 million customers were without digital services following an electrical failure at a data center in Ireland [17]. ENISA in its study has reported that on an average 2.8 million user connections were affected due to power supply failure in the year 2012 [14].

4.3.6 VIRUS ATTACKS

Sensitive data is always in the risk zone of virus attacks. Computer systems at energy firm RasGas in Qatar have been taken offline by a computer virus. The attack forced the RasGas firm to shut down its website and e-mail systems. A similar attack also hit oil giant Aramco causing huge destruction in terms of money, time and production [18].

Keeping the data at a single place in a centralized database, as is the case with CIDR of Aadhaar, it is always vulnerable to the disruption of services. If any of these problems or combination occurs; there will be disruption in the information flow from CIDR to the remote place where the individual's data is being accessed. Thus, the AUAs, enrolment agencies as well as the residents of that particular region will have to wait for the system to recover. This may result in no authentications; no creation of new UIDs and eventually the e-Governance services will not be available.

V. PROBLEM STATEMENT

The CIDR is accessed simultaneously through a large number of AUA's across the country to authenticate individuals in order to grant them access to the physical resources or to benefit them under the welfare services of the government, such as Public Distribution System (PDS), Direct Benefit Transfer (DBT), Rashtriya Swasthya Bima Yojna(RSBY) etc. All this require the instant response in 'yes or no' from the CIDR to the request made by an AUA. For example, if a person at Lahul- Spiti (remote snow-covered populated terrain) of Himachal Pradesh is to be benefitted with DBTL, he has to be authenticated by the CIDR over the internet via ASA; as CIDR is not located at Lahul- Spiti. Most of the time, Lahul- Spiti remains disconnected with the remaining world both physically and electronically [19]. This will directly affect the authentications and will result in the poor e- Governance. More than thousands of the electronic transactions take place across the country 24x7 hours a day and this requires as many quick responses from the CIDR. Along with authentications, CIDR is quite busy in assigning new UIDs to the residents. If sometimes, centralized CIDR is unable to give just-in-time response to the various AUAs, the transactions will be stopped. The following problems may arise due to the bottlenecks discussed in section 4.3: (i) there is a probability that CIDR is unable to respond instantly at the request of an AUA for authentication. (ii) It may affect badly the process of identification i.e. the process to provide Aadhaar to a new enrollee. To provide Aadhaar to a new enrollee, the system would effectively have to compare the personal identification data of new enrollee against the personal identification data of the people already enrolled in to the CIDR. This means that to uniquely enroll an individual into a population of x million people, x million individual authentications would effectively to be performed, to give an accurate and unambiguous answer to the question 'Who am I?'.

Therefore, keeping data at centralized CIDR may result in disruption of services.

VI. OBJECTIVE

a) To speed up the process of authentication and identification.
b) To propose an alternative architecture of Aadhaar database (CIDR) based on DDBMS; that performs well under the practical scenario as discussed in 4.3.

VII. PROPOSED SOLUTION WITH DISTRIBUTION OF AADHAAR DATA AT LOCAL BODIES

In a distributed database, overall performance is greatly influenced by the amount of communication required to execute an operation. Faster response time can be achieved if all the data accessed by a transaction is available at the site where the transaction is initiated [20]. The citizens avail most of the welfare services at the place of their living. Also, the personal identification data about citizens is used frequently again and again in order to provide them the Government services at their doorsteps.

The citizen centric data is generated at local bodies such as Panchayats, Notified Area Councils (NACs) and Municipal Corporations. As a usual practice, only the demographic data of an individual is kept at these places. In order to provide him/ her unique ID, his/ her biometric data along with this demographic data is to be sent to CIDR. The proposed solution suggests that the personal identification data should be kept at the place of its origin i.e. at distributed places, where it is more likely to be used in future for all kind of electronic transactions. The Locality of reference can be adopted as a model for the Distributed database management System of Aadhaar because most of the time the residents' data is used at the places they reside [21]. This is also applicable particularly to avail government services by the residents as these services are rendered where they live. In contrast to existing Aadhaar approach of keeping data at centralized CIDR, a distributed approach can be adopted using concepts of Distributed Database Systems (Figure 3). A distributed database is a collection of multiple logically related databases dispersed at geographically distant places and connected to each other through computer networks. Distributed systems assure the availability of information and easy access to the data that is stored at geographically distant places. It helps to get rid of the drawbacks of centralized scheme where any kind of system failure may lead to huge loss to the organizations [22]. In the proposed solution, the

complete data about the residents may be stored at local level and only the data required for Aadhaar purpose may be shared. This will provide autonomy to the local databases. If in some applications, the resident's data is required at different places then it can be replicated or fragmented for replication at different databases. The existing setup of UIDAI has been divided into a collection of distributed databases which are connected with each other via communication channels. The separate databases have been developed at every level of abstraction ranging from the State level to District level and eventually to the local level i.e. Gram Panchayats, Notified Area Councils (NACs) and Municipal Corporations (MCs). The issues like manageability, accessibility, availability and security of the residents' personal identification data would be addressed efficiently.

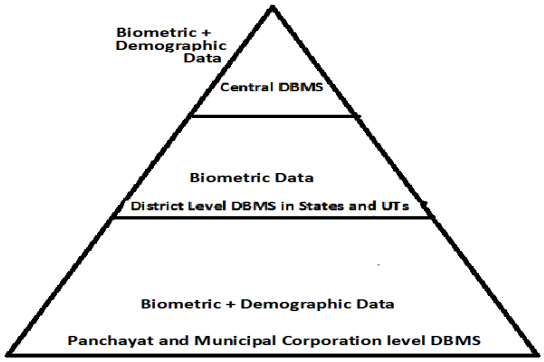

Figure3.Proposed Decentralized Architecture for Aadhaar with faster response time and fault Tolerance.

7.1 Authentication & Identification of Personal Identification Data (PID) with the Proposed System

The personal Identification Data (PID) can now be authenticated at the place of the living of residents. The authentications will become relatively easy and fast.Mostly, all the electronic transactions are initiated by the local AUAs. This means that the credentials of an individual are mostly required to be authenticated by the local AUAs. The AUAs are likely to find the related information more easily and relatively fast at local databases rather than at the central database. The answer to the question 'Am I who I say I am?' will be given instantly, as the response is to be generated by a local DBMS. For the authentication of an individual's identity, his credentials willbe searched at the local database first. If the search remains unsuccessful then it will be searched for, in the database at higher level.

Identification, in the decentralized approach will be done by sending the scanned biometrics of the new enrollee to the databases distributed over the country. Scanned biometrics will now be compared in parallel for authentication against every single biometrics stored in multiple small databases rather than with a big central database, the CIDR. This will lead to relatively fast responses than CIDR. In the decentralized approach, the system will be able to provide the Aadhaar immediately at the time of enrollment. Whenever a new person is to be enrolled into the UIDAI, his personal identification data can directly be added to any local database, not necessarily to CIDR. His/ her information will automatically be synchronized back and forth. This will definitely help to achieve the promises of the Digital India.

VIII. FINDINGS AND CONCLUSIONS

The existing Aadhaar approach is to put resident's data in centralized database CIDR whereas in this paper a decentralized architecture for Aadhaar has been proposed using distributed systems approach. This has been done by keeping in view the bottlenecks that may arise in authenticating the credentials of an individual. Also, because mostly residents will use government services at the places of their living and due to locality of reference it is advisable to keep data at nearby decentralized places for improving response time. The distributed database approach can be used to design databases at Local Level, State Level and National Level by replicating data if there is need to do so. Further study is needed to implement the idea given in this paper.

IX. ACKNOWLEDGEMENT

This paper is an abridged version of the paper presented in the two days National Seminar 'Innovations in Governance: Multidisciplinary Perspectives' Titled "A Distributed Database Approach: A Novel Way to Realize Digital India Through Aadhaar" held at University School of Open Learning, Panjab University, Chandigarh on dated March 30-31, 2016. The efforts have been made to incorporate the feedback given by participants attending the presentation. The authors are thankful to the conference organizers as well as the participants.

REFERENCES

[1] ShankkarAiyar (2017), Aadhaar: A Biometric History of India's 12 Digit Revolution, Westland Publications Ltd.
[2] http://www.financialexpress.com/india-news/aadhaar-card-deadline-extended-to- december-31-after-right-to-privacy-order-issued-by-supreme-court/833891/ accessed on 22/12/2017.
[3] About TheProgramme Digital India, Retrievedfrom http://www.digitalindia .gov.in/content/about-programme.
[4] UIDAI Strategy Overview, Creating a Unique Identity Number for every resident in India. (2010, April).

Retrieved from https://uidai.gov.in/ UID_PDF/Front_Page_Articles/Documents/Strategy_ Overveiw-001.pdf

[5] Aadhaar Authentication Specification - version 1.6.pdf, UIDAI. (2014, January), Retrieved from https:// auth portal.uidai.gov.in/static/aadhaar_authentication_api_1_6. pdf

[6] UIDAI's MOU with Government of H.P. (2010, May 21). Retrieved from https://uidai.gov.in/images/mou/MOU-HP.pdf.

[7] Anil K. Jain, Arun Ross and SalilPrabhakar. (2004, January).An Introduction to Biometric Recognition.IEEE Transactions on Circuits and Systems for Video Technology, Volume 14, Issue-1, Retrieved from http://www.cse.msu.edu/~rossarun/pubs/RossBioIntro_C SVT2004.pdf

[8] Banking and Biometrics. White Paper, Biometrics Research Group, INC. (2014, November), Retrieved from www.biometricupdate.com

[9] RICHARD HOPKINS. (1999). An Introduction to Biometrics and Large Scale Civilian Identification. International Review Of Law Computers & Technology, Volume 13, NO. 3, PAGES, 337–363, Retrieved from http://newton.ee.auth.gr/bio metrics/images/docs/hopkins.pdf

[10] Rahul Jacob: The Aadhaar tragedy. (2014, June 25). Retrieved from http://www.business-standard. Com / article/opinion /rahul-jacob-the-aadhaar -tragedy-114062501238_1.html

[11] Internet Seen as Positive Influence on Education but Negative on Morality in Emerging and Developing Nations. (2015, March 19). Retrieved from http://www.pewglobal.org/2015/03/19/internet-seen-as-positive-influence-on-education-but-negative-influence-on-morality-in-emerging-and-developing-nations

[12] Internet revolution bypasses rural India: Survey. (2012, May 6). http:// www. thehindu.com/sci-tech/ technology /int ernet/internet-revolution-bypasses-rural -india-survey/article3390353.ece

[13] India: 20 Internet Shutdowns in 2017; June 15, 2017. Retrieved from https: // www.hrw.org /news/2017/06/15/india-20-internet-shutdowns-2017

[14] Dr. Marnix Dekker, ChristofferKarsberg, MatinaLakka. (2013, August). Annual Incident Reports 2012 Analysis of Article 13a annual incident reports. Retrieved from http://www.teleoff. Gov .sk/data /files/35451.pdf

[15] Google loses data as lightning strikes. BBC New. (2015, August 19). Retrieved from http://www.bbc.com / news/technology-33989384

[16] HenryBlodget. (2011, April 28). Amazon's Cloud Crash Disaster Destroyed Many Customers' Data permanently, Retrieved from http://www.businessinsider.com/amazon-lost-data-2011-4?IR=T

[17] Tom Jowitt. (2015, August 27). C4L Resolves Service Outage, Blames Software Glitch. Retrieved fromhttp://www .techweek europe.co.uk /cloud/ datacenter /c4l-resolves-service-outage-175587.

[18] Computer virus hits second energy firm. BBC News. (2012, August 31). Retrieved from http://www.bbc.com /news/technology- 19434920

[19] Kuldeep Chauhan. (2014, April 1). Sissu, Koksar phone exchanges out of order for a month.The Tribune Online edition, Chandigarh, India, Retrieved fromhttp://www.tribuneindia.com/2014/20140401/himachal.htm#15

[20] G. Alonso, A. EI Abbadi, "Partitioned Data Objects in Distributed Databases", in Distributed and Parallel Databases, Vol. 3, Issue-1, pp-5-35, January 1995.

[21] M. Tamer Ozsu, Patrick Valduriez.(2011). Principles of distributed database systems, Third Edition.

[22] Andrew S. Tanenbaum, Maarten Van Steen. (2005). Distributed systems, principles and paradigms.