Privacy and Security in Aadhaar

[1]Shaurya Shekhar, [2]Vasantha W B
[1]Student, [2]Professor
[1]VIT University, Vellore, Tamil Nadu, India
[2]SCOPE, VIT University, Vellore, Tamil Nadu, India

Abstract— **The Aadhaar project of the Government of India is the most ambitious program in the world aimed at issuing unique 12 digit numbers to every Indian as well as recording their biometrics for authentication services. A data leak would be potentially disastrous and would constitute a major breach of privacy as well. Such a leak can take place at the application level, network level and the storage level. Data pertaining to the number of data breaches in the past decade have been analyzed to emphasize the importance of a secure ecosystem for such an ambitious project. A new methodology has also been studied which will help in heightening the security of the Aadhaar ecosystem and safeguarding the privacy of the people better.**

Index Terms—**Aadhaar, UIDAI, Security, Privacy, Data**

I. Introduction

The Unique Identification Authority of India (UIDAI), a statutory body established by the Government of India is responsible for the rollout of the Aadhaar project. Aadhaar is a 12 digit unique identity number being issued to every person after his/ her successful enrollment into the Aadhaar database [1]. The highlight of this database is the simultaneous recording of every individuals biometrics (iris scans and fingerprints). A successful enrollment is confirmed only when the quality of the biometrics captured meet certain specifications and when they pass the redundancy check, which checks if the same biometrics exist in the system already [5][2]. This helps in weeding out fake enrollments and also in maintaining the quality of the data being captured by the system. As of November 2017, 1.19 billion Aadhaars have been generated at a total expenditure of around 48 billion Rupees. It is also said that 99% of all adults have been covered by this program. Since, a huge amount of sensitive information will be stored in this repository of data, it is thus, important that the security of this delicate ecosystem be impenetrable.

The objective of this paper is to understand the various aspects of the Aadhaar initiative with respect to security and privacy. Aadhaar being the worlds largest repository of biometrics will be used for identification services such as authentication and authorization by many third party independent services. This paper tries to quantify the huge increase in number of data breaches over the past decade, to be able to emphasize the importance of having a safe and secure ecosystem for Aahdaar. This paper also tries to present an idea which would reduce the chances of private data of individuals being leaked.

II. Cases of Data Breaches In Big Data Systems

To be able to understand the extent to which the number of security breaches in big data systems has increased, we have made use of RStudio and the R language. Data pertaining to security breaches from 2004 to 2017 was collected in the form of a .CSV file [3]. This CSV file was then loaded into RStudio using the read_csv() file which is found in the readr library. This raw data was in the form of:

Raw_Data(Entity, Year, Records, Organization Type, Method)

The various columns of this data frame represent:

- Entity: The organization affected
- Year: Year of security breach
- Records: Total number of records affected (leaked/ modified)
- Organization Type: Industry of the organization affected
- Method: What type of data breach it was (poor security, inside job, hacked, etc.)

Due to the gathered raw data having incomplete data in certain tuples, it is important for us to clean this data before we begin analyzing it. To remove the incomplete tuples from this data frame, we make use of the complete.cases() functionality, which removes the tuples which have garbage values or “NA” in any of their fields.

After having cleaned the data available, we now segregate the data on the basis of the year of data breach, after which we summarize the number of records affected each year and store it in a ‘YearWiseRecord’ data frame. This data frame is then visualized using the barplot() function which is found in the ggplot2 library.

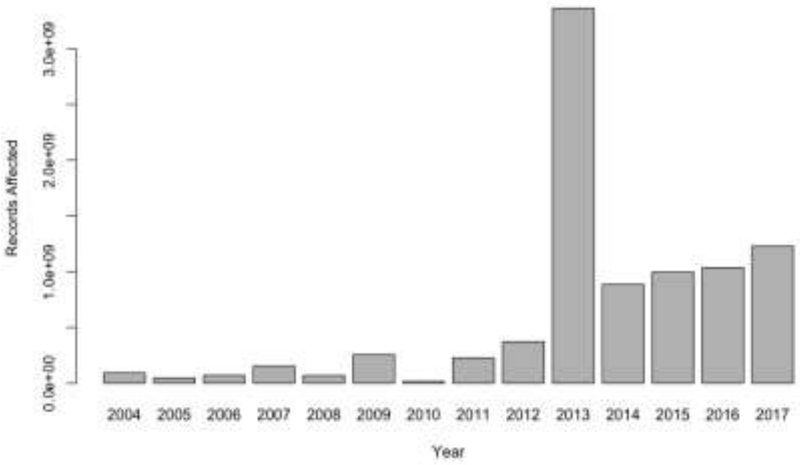

Figure 1: Graph of number of records affected every year from 2004 – 2017

As can be seen in Figure 1, there has been a consistent increase in the number of records being affected due to instances of data breaches since 2011 onwards. The concept of Big Data gained prominence around the 2010s, where-in it was discovered that storage of large amounts of information and subsequent analysis of it could yield large amounts of business intelligence. This is around the time, multinational corporations started investing in huge data warehouses to be able to store terabytes of information, thus creating targets for data breaches.

We can also see that the year 2013 stands out in particular, as it has the largest number of records being affected amongst all the years. This is due to the Yahoo database hack due to which approximately 3000000000 records were affected. It remains one of the most massive data breaches till date, and thus, contributes to the year 2013 being an outlier.
We have now established that there has been a steady spurt in the number of records being affected by data breaches over the past decade, we can also safely assume that being the worlds largest repository of demographics and biometrics, the Aadhaar database would also be an ideal target for miscreants and cyber terrorists.

III. Aadhaar Usage For Authentication And Identity Verification And The Risk Involved

One of the principle uses of the Aadhaar database has been to be able to provide identification services to independent third-party agencies such as banks, State Governments, etc. To be able to clearly understand this, we need to differentiate between identity and authentication.

Table 1: Difference Between Identification & Authentication

Identification	Authentication
1. Who are you? 2. Sometimes public information 3. Active user interaction might not necessarily be involved Eg: During opening of bank account	1. Proof of claim of identity 2. Must always be private 3. Active user interaction must always be involved Eg: During financial transaction

Authentication services are provided by UIDAI by means of Yes/ No responses from the central repository (CIDR). Authentication User Agencies (AUAs) [1] represent the third party independent agencies which make use of these services, and have to make use of Authentication Service Agencies (ASAs) [1] which have direct secure line connectivity to the UIDAI database. Multiple AUAs can make use of a single ASA. The information packet which needs to be verified contains various demographic details and in some cases, biometric samples, which is transmitted to the CIDR. After matching, a simple message of "Yes or No" is returned signaling successful authentication or unsuccessful authentication respectively.

The UIDAI provides 5 different types of authentication services. They are:

1. Type 1 Authentication: Matching of Aadhaar number with demographic details.
2. Type 2 Authentication: Makes use of OTP sent to mobile number/ email ID.
3. Type 3 Authentication: Makes use of one of the biometric (iris/ fingerprint).
4. Type 4 Authentication: Makes use of OTP and one of the biometrics.
5. Type 5 Authentication: Makes use of OTP and both the biometrics.

Since, all the authentication types involve use of the Aadhaar number, it may be considered a global identifier cutting across multiple domains. This leads us to the main concern of correlation and connection of identities over multiple by means of the global identifier. Such correlation would contribute to authentication without consent, which is a very dangerous privacy breach. This could lead to illegal profiling, tracking and surveillance of individuals without any legal sanction.

IV. A Temporary Identifier – Virtual ID

Since it has been pointed out, that a massive problem arises from the fact that every individuals Aadhaar number is a global identifier, it is proposed that a method to keep this global identifier private at all times be developed. Keeping this in mind, it is proposed that the concept of slave identifiers be introduced. A number of these slave identifiers can be be linked to a single master identifier (the Aadhaar number) and each of them can be used individually for every authentication purpose. Such connections shall be unidirectional in nature only and it will be impossible to decode the global identifier from any of its local identifiers.

The domain of such local identifiers will be restricted to one particular domain itself and thus, linking of these identifiers across multiple domains would not be possible. To enhance the security of this local identifier, they will be valid only for a particular time period for the time of generation and also only for single time usage. These two features will also undermine the temporary aspect of these local identifiers.

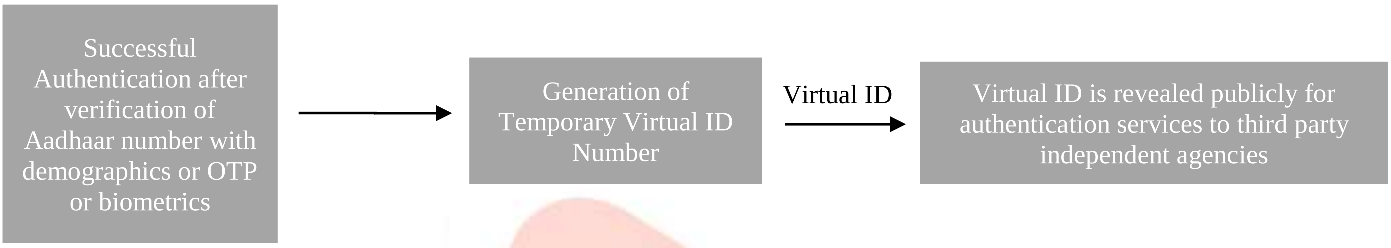
Successful Authentication after verification of Aadhaar number with demographics or OTP or biometrics
Generation of Temporary Virtual ID Number
Virtual ID
Virtual ID is revealed publicly for authentication services to third party independent agencies

Figure 2: Architecture of Proposed Virtual ID

The generation of such local identifiers can be done online by using an application or through a website. After successful self-authentication, a virtual ID is generated, which can then be presented to an AUA for authentication services. This further restricts the amount of your personal details in the public domain.

One of the other proposed methods is to embed local-domain identifiers in the master identifier by means of sufficiently strong cryptographic keys. This would allow for bi-directional linkages between the global and local identifiers. Further this could also allow for multi-domain identifiers for special use cases. However, there is a certain degree of security risk involved as even the strongest of cryptographic keys can be broken and decoded eventually.

V. Other Security Risks And Future Works

After close analysis of the Aadhaar ecosystem, it has been found that security loopholes can be found at the following three levels [7][9]:

1. Storage Level [4]: At this level, we refer to the need to safeguard the repositories of data from outside attacks by making use of real-time intrusion avoidance and detection systems [6]. Access control [8]should also be implemented to limit the number of people who have access to sensitive data, thus helping in reducing the number of potential leakage points. It should also be ensured that no sensitive data is ever stored at an unsecure data storage facility.

2. Network Level [4][3]: Since the Aadhaar ecosystem is online throughout, it also very important that the networks being used for transmission of enrollment and authentication data packets be secure and reliable. To ensure, this we suggest the usage of Virtual Private Networks (VPNs) as well as strong cryptography keys [4]. This can also be referred to as Communication Security.

3. Application Level [4]: At the application level, we refer to the enrollment devices as well as the authentication devices. Such authentication servers and application devices must routinely authenticate each other so as to ensure the validity of such connections. Since, it is impossible for UIDAI to check the security at all ‘point-of-service’ devices as well, it is thus, assumed that all such connections are trustworthy and comply with predefined standards.

Third-party security audits should also be conducted regularly to ensure that all security guidelines are being strictly adhered not only in the Aadhaar database but also by client applications. Such comprehensive audits would also help in reducing the risk of insider attacks.

VI. Conclusion

The Aadhaar initiative is the biggest biometric based identity project in the world and if implemented correctly can revolutionize governance in India [10]. It would also help in saving crores of rupees by weeding out fake beneficiaries and thus ensuring that Government mandated funds and benefits reach the sections of society that they are actually meant for. While the benefits of such a project are huge, so are the risks. As has been demonstrated, there has been a steady increase in data breaches and a breach in the Aadhaar database would mean a privacy risk to a billion people.

It is thus, important to ensure that the security infrastructure of the Aadhaar ecosystem is regularly updated and is the best in its class. In this paper, we have also suggested the use of domain specific – temporary - local identifiers to further limit the amount of sensitive personal information in the public domain. Some of the main security leakage points have been identified and measures to safeguard them have also been suggested.

References

[1] Agrawal, S., Banerjee, S., & Sharma, S. (2017). Privacy and security of Aadhaar: a computer science perspective. *Economic & Political Weekly*, *52*(37), 93-102.

[2] Chander, S., & Kush, A. (2010). Unique Identification Number and E-Governance Security. *International Journal of Computing and Business Research*, *1*(1).

[3] Constantine, C. (2014). Big data: an information security context. *Network Security*, *2014*(1), 18-19.

[4] Dayal, M., & Singh, N. (2016). An Anatomization of Aadhaar Card data set–A big data challenge. *Procedia Computer Science*, *85*, 733-739.

[5] Faundez-Zanuy, M. (2006). Biometric security technology. *IEEE Aerospace and Electronic Systems Magazine*, *21*(6), 15-26.

[6] Lafuente, G. (2015). The big data security challenge. *Network security*, *2015*(1), 12-14.

[7] Liu, S., & Silverman, M. (2001). A practical guide to biometric security technology. *IT Professional*, *3*(1), 27-32.

[8] Mahajan, P., Gaba, G., & Chauhan, N. S. (2016). Big Data Security. *IITM Journal of Management and IT*, *7*(1), 89-94.

[9] Moreno, J., Serrano, M. A., & Fernández-Medina, E. (2016). Main issues in big data security. *Future Internet*, *8*(3), 44.

[10] Sharma, V. (2011). Aadhaar-a unique identification number: Opportunities and challenges ahead. *Research Cell: An International Journal of Engineering Science*, *4*(2), 169-176.