Tufts University - COMP 116

Computer System Security

Assessing Aadhaar

Confidentiality amid the world's largest biometric identification effort

Paul Wullenweber

December 10, 2018

Abstract

India's biometric identity verification system is close to completing its first decade of service. As the largest such effort in the world today, it has seen rapid adoption. While it has undoubtedly had an enormous impact, confidentiality and user data security do not seem to have been completely central to the system's design. This paper will point out vulnerabilities in the enrolment process and discuss the resulting challenges to data integrity as well as examine issues stemming from information disclosure and the public handling of Aadhaar details before scrutinizing the system's security with regard to trusted access to its central database. By analysing both core Aadhaar infrastructure and peripheral factors, multiple potential weaknesses that have been exposed and exploited throughout the recent years are taken into account. This way, the current state of confidentiality and integrity of the Aadhaar initiative is documented.

1 Introduction

In India, the world's largest biometric database is only a few months away from its 10th anniversary. The Unique Identification Authority of India (UIDAI) was set up on January 28, 2009, with the mandate to provide every citizen with a 12-digit unique identification number. This number, the so-called Aadhaar ID, is meant to clearly establish a citizen's identity to both public and private agencies across India [21]. Since its launch, the scheme's user base has grown to cover more than 90% of the country's population, with more than 1.2 billion generated Aadhaar numbers[1] [2]. Aadhaar is now the de-facto standard of identity verification accepted for banking services, mobile communication, education and healthcare [20]. A brief overview of the service's development is displayed in Figure 1.

The primary use case of Aadhaar lies in the distribution of government benefits and handouts. Biometric identification is perceived as a silver bullet when it comes to preventing welfare fraud. Unfortunately, this also means that invasive biometric identification schemes are now mandatory for receiving basic entitlements in India [20].

[1] As of December 2018

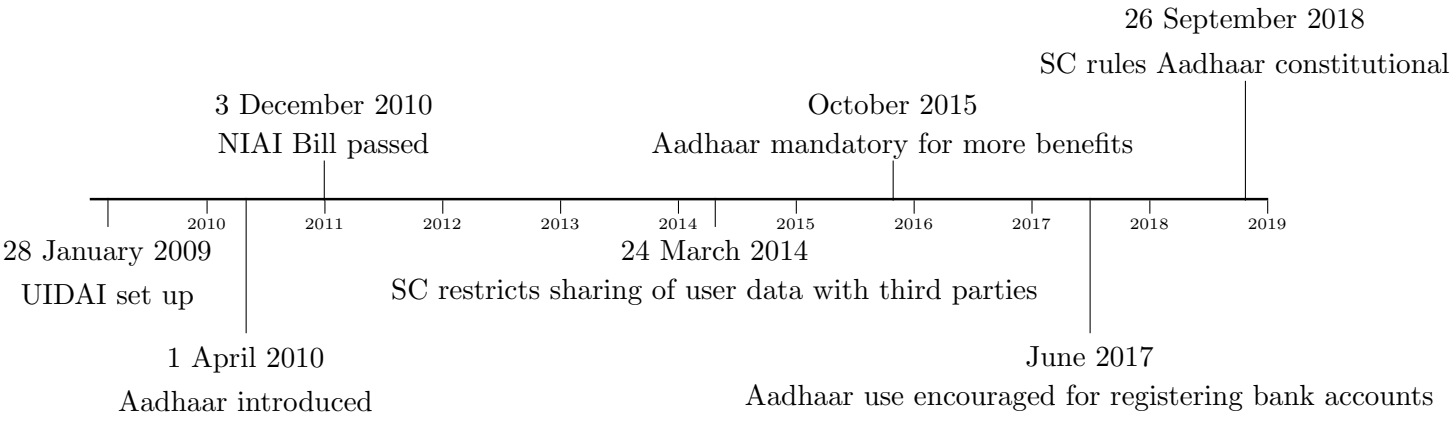

Figure 1: Aadhaar timeline [23]

Even in cases where Aadhaar identification is not yet mandatory, it is often made to look as though it were, for example when activating SIM cards [19, 24]. India's supreme court (SC) has only recently begun to order telecommunication companies to allow customers to register new SIM cards with different kinds of identification instead of their Aadhaar IDs [17, 19].

2 To the Community

Biometric identification is becoming more and more common around the globe. Although the details vary across countries, this usually implies assigning individuals ID numbers, the possession of which the owner usually verifies by having their biometric characteristics scanned. Therefore, almost all such efforts rely on databases containing ID numbers and some representation of the corresponding organic details, usually along with the option of storing much more related, e.g. demographic, data. While biometrics seem to offer an efficient way of identifying individuals, such schemes place vital personal information in the hand of the state, thus potentially facilitating misuse or privacy violations [18]. Even if the government agencies in charge of storing such data are to be trusted, databases holding vast amounts of highly valuable personal information will always be prime targets for individuals with nefarious intents.

The main problem with biometric identification is summarised very well by the Electronic Frontier Foundation (EFF): "If you lose a biometric, you've lost it for life". Unlike losing credit card information, biometric information is unique and cannot be made invalid or be replaced [9]. Systems that aim to employ biometric identification on a national scale thus need to be designed for security from the ground up, as even a single breach of such a system might result in identity fraud on an equally large scale, with no efficient remediation possibilities.

It is likely to be only a matter of time until biometric identification systems will be employed elsewhere. Given the fact that Aadhaar is one of the first systems to roll out on a truly

universal scale, it is vital to assess this attempt and learn from its strengths and weaknesses in order to improve future implementations.

3 Vulnerabilities and their exploitation

Like other biometric identification schemes, Aadhaar can be dissected into several different layers defined by the EFF, as depicted in Table 1. In the data collection stage, a reference sample of the user's fingerprints or iris scans are taken using Aadhaar's Enrollment Client Multi-Platform (ECMP), along with the user's current demographic data. This reference sample, against which future input will be compared, is then transmitted to the UIDAI's Central Identity Repository (CIDR), where all of the collected information is stored. Once a new entry is submitted to the CIDR, a process called (Demographic) De-Duplication is used to eliminate the possibility of an individual appearing multiple times in the database [4]. When verifying an individual's identity, their organic characteristics are scanned and compared against the records present in the CIDR[2]. The identity is successfully verified if a match is found. A scheme called electronic Know-Your-Customer (e-KYC) can be used to retrieve demographic data upon the presentation of an ID or biometric details [7].

Table 1: EFF biometrics characteristics [9], Aadhaar counterparts and attack surfaces

EFF Category	Aadhaar Process	Attack Surface
Data Collection	Enrolment	ECMP, Hardware, Human Operators
Transmission	Transmission to CIDR	ECMP, e-KYC, Network
Signal Processing	Fingerprint / Iris Scanning	ECMP, Hardware, Fake Biometrics
Decision	Matching of scan and CIDR	ECMP, Fake Biometrics
Storage	CIDR	CIDR, Access Control

The traditional structure of Aadhaars identity verification process means that multiple attack vectors related to the different layers have to be considered when gauging the security of the service. Given the fact that Aadhaar IDs are now commonly used to identify individuals in need of government aid, there is a clear incentive for fraudulent registrations in order to qualify for public handouts. Even though Aadhaar was launched with the goal of reducing opportunities of identity theft, the large number of activities that now require Aadhaar identification, such as land transfers, procuring passports, getting loans, casting votes, applying for other IDs, or obtaining food rations, have drastically increased the value of forged registrations [3]. In the following sections, I detail three separate attack surfaces that can impair Aadhaar's confidentiality, availability or integrity and thus reduce its general effectiveness as an identification service when exploited.

3.1 Attacking the ECMP

The Aadhaar enrolment process was designed to quickly grow the user base, without a rigorous focus security or user privacy. This is most apparent in the ECMP software used to enrol

[2]Compare Figure 2 in the appendix

new users. In theory, anyone with a laptop and biometric scanning hardware can set up an enrolling station, as the ECMP client is publicly available for download. To safeguard the quality of the data in the CIDR however, operators need to be registered with the UIDAI, have to enter their login credentials and then verify their identity using biometric details when registering citizens. Since operating enrolling stations is potentially very lucrative, e.g. operators fraudulently registering individuals so that they qualify for handouts in return for a bribe, the enrolment process came under attack very quickly.

The first attacks were carried out by individuals using stolen stolen credentials and forged fingerprint moulds to gain operator privileges, or authorised operators sharing their credentials with rogue individuals. In one incident, perpetrators thus managed to fraudulently register more than 870 individuals in the 'physically disabled category', qualifying them for handouts [6].

Given the client model approach, the next set of attacks targeted the ECMP software itself. It seems to have been rushed into service: in its early versions, it did not even encrypt the data packets containing the collected information queued for uploading, potentially allowing anyone with access to the machine to steal both demographic and biometric data[3]. By reverse-engineering the ECMP[4], attackers were able to circumvent or disable operator biometrics checks and could spoof the GPS location of the station, thus enabling anyone with the required knowledge and access to an operator's login credentials to contribute to the CIDR [26]. In an early version, a bug allowed anyone to circumvent the fingerprint check, as it accepted any fingerprint after initially rejecting it twice [6]. The first unauthorized enrollment centers emerged in 2016 [13]. In order to combat rogue operators, subsequent updates introduced additional security measures such as iris authentication and code tampering detection. However, as of April 2018, GPS and fingerprint checks remained bypassed [11], with the required login credentials of legitimate operators available for purchase on the black market for as little as $35 [16].

All of these attacks have the objective of enabling unauthorized individuals to add new data to the CIDR or modify existing records, thus impairing Aadhaar's integrity. Once a fraudulent entry has been successfully submitted to the CIDR, it can be used indefinitely by the person responsible for it. Furthermore, given that stolen or shared credentials along with possibly spoofed GPS data were used in many cases, it is very difficult to distinguish between legitimate and illegitimate entries, especially given the massive size of the Aadhaar database. Such activities can therefore prove detrimental to a system that aims to provide accurate and trustworthy identity verification.

3.2 Trolling the former UIDAI chief

Aadhaar's quick rise in importance and its rapid adoption have led to serious security issues in the enrolment process. This is only one issue that the system is grappling with, however. Unfortunately, there still seem to be frequent misconceptions about Aadhaar's purpose when actually deployed in the field. The idea is simple: having a reliable database populated with demographic and biometric data, against which users can easily verify their identity, allows for quick identification. When deployed and used in public life, however, misconceptions seem

[3] Due to unreliable internet access in some regions, the collected information is stored on the machine used for enrolment until a network connection is established

[4] It is written in Java and does not seem to use obfuscation, thus facilitating decompilation [26]

to be rife. One recent example is the decision of R.S. Sharma, former chief of the UIDAI, to publicise his Aadhaar ID on Twitter, along with the challenge of showing him what harm could be done with this information. Although apparently convinced of the security of the system, this publicity stunt backfired spectacularly: within days, his mobile number linked with Aadhaar, his secondary phone number, postal address, email address, date of birth, voter details including his father's name and his Permanent Account Number (PAN)[5] were leaked and shared online [25].

This episode summarises the misconceptions concerning Aadhaar quite well. In order to identify yourself using the service, you need to know your Aadhaar ID and present biometric characteristics, e.g. fingerprints or iris scans, that match the ones on record. In a sense, this resembles the two-factor-authentication of the web: to identify yourself, you have to have both *something you know* and *something you have*. In Aadhaar's case, the 12-digit random ID represents the former, while biometrics represent the latter. In theory, Mr. Sharma could rest easy, knowing that without his biometric details, no one would be able to use his ID for fraudulent activities. In his case, the Aadhaar system itself or the CIDR were not breached. His information was leaked due to information disclosure by third party databases, which allowed looking up information by Aadhaar ID. This is why it is important not to unnecessarily expose this ID, as there does not seem to be a consensus about what to use the ID for or when to require additional verification. In the past, there have been numerous incidents where government agencies have leaked Aadhaar related information through public websites, including demographic details and IDs. In 2017 alone, a high number of such incidents has been recorded [1].

3.3 The Hyderabad SIM Card scheme

There is another incident that demonstrates the ways in which initiatives like Aadhaar can facilitate identity fraud. Throughout the first half of 2018, a SIM card distributor in Hyderabad used publicly available information to activate more than 6,000 cards by impersonating unsuspecting individuals. In order to activate the cards, the distributor had to have two things: an individual's Aadhaar number and their matching biometrics. In order for the system to work as intended, these two components should never be stored side by side, except for in the CIDR. Unfortunately, more and more public outlets require Aadhaar information on official documents, storing them themselves. In the case of the Andra Pradesh subregistrar's office, such records were even made available online: against a small fee, property registration documents containing the IDs and fingerprints of the buyer, the seller and those of two required witnesses could be obtained by any interested party. Given the relatively high market price of pre-activated SIM cards, the distributor was able to turn a profit by paying for the details, ordering moulds made from the fingerprints in the documents[6] and using a legitimate Aadhaar verification device to register the SIM cards for the given identities [14]. He was only caught because he used the same verification terminal for most of his activations. Instead of activating SIM cards, he could have also used the stolen identities to open bank accounts or apply for PAN numbers.

The UIDAI responded by allowing users to link a mobile phone number to their Aadhaar ID, to which one-time-passcodes (OTP) are sent upon any verification requests. In essence,

[5] 10-character alpha-numeric identifier, used for tax purposes and identification

[6] A process developed by the CCC to circumvent Apple's TouchID was used [10]

this change just added another layer to the Aadhaar security scheme: aside from the ID and the biometric details, users are enabled to provide *something that they temporarily have*, making a full identity takeover much more difficult to achieve[7]. Although this measure has the potential to drive down identity theft, it will only benefit those that own a phone and know how to activate the option.

This incident shows again that vulnerabilities do not necessarily need to arise from flaws in the implementation of Aadhaar. The problem here is that government agencies do not treat Aadhaar IDs or even biometrics with the care and scrutiny that should be applied, especially given the verification architecture behind Aadhaar. By neglecting the dangers of publishing such sensitive information, these agencies facilitated identity theft. Without a proper understanding of how the different factors of Aadhaar authentication are combined to provide security, incidents such as this one are likely to occur again.

3.4 Unauthorized API access

The UIDAI insists that the CIDR has never been directly compromised, remains secure and "can't be breached" [27, 22]. Unfortunately, a recent incident shows that breaching Aadhaar is not a prerequisite for unauthorised access to the database. In the spring of 2017, a software engineer managed to gain access to an API used by Aadhaar's e-hospital service to query user demographics when presented with an Aadhaar ID [15]. Two vulnerabilities facilitated fraudulent access: first, a hardcoded default token in the implementation of the API and second, the use of HTTP for queries, which allowed for the password to be intercepted using a proxy. The same default token was also hardcoded into the e-hospital Android app[8], allowing it to be found by static analysis [12]. Having gained access to the API, the engineer then published an app of his own, allowing anyone to enter an Aadhaar ID and query the CIDR for the demographic details associated with it by routing the request through the e-hospital API. Given that the requests to the CIDR seemed to originate from the e-hospital system, it took more than four months for the hack to be discovered, during which he was able to earn more than 40,000 Rupees through advertisements placed in the app [15].

This incident shows how difficult it is to fully secure all attack surfaces of a system on the scale of Aadhaar. Even though critical infrastructure such as the CIDR has not yet been breached directly, attacks such as this have shown multiple times that serious data theft can occur when trusted parties[9] insufficiently secure their systems. The confidentiality of Aadhaar information does not only rest solely with the UIDAI, but with all those partners that are trusted with access to the stored data.

4 Conclusion & Outlook

In the implementation of complex systems such as Aadhaar, where so much is at stake because of the use of irreplaceable biometric data, security should be factored in at every level. Instead of rushing a half-baked product into public use, state funded entities such as the UIDAI should question their every decision with regard to protecting the privacy of the many individuals

[7] As of 2018, only 10% of respondents in a UIDAI poll knew about the OTP option [5]

[8] Compare Figure 3 in the appendix

[9] As of August 2017, there were 27 entities with 'leased line' (API) access to the CIDR, known as Authentication Service Agencies (ASAs) [12]

that are more and more often forced or nudged into registering for their services. Some of the mentioned incidents could have been prevented by security audits. The development of a threat model that takes into account the incentives for attacking different layers of the system could also strengthen its resilience. The mentioned misconceptions about the different authentication factors are likely to fade as politics, society and bureaucracy adjust to Aadhaar' approach to identification.

Unfortunately, the UIDAI's user privacy track record does not bode well for the immediate future. While every new leak is likely to provoke additional security measures and a public outcry, officials are always quick to insist on the security of the system as a whole. Again and again, it has become evident that this is not enough, however. Nevertheless, India is unlikely to do away with Aadhaar anytime soon, especially now that more than 90% of its population has been enrolled. Time will tell whether this is is the right direction to head into. In the meantime, one can only hope that privacy and security will gain in importance and be iteratively strengthened along the way.

References

[1] *#AadhaarLeaks: A List of Aadhaar Data Leaks*. Apr. 24, 2017. URL: https://www.medianama.com/2017/04/223-aadhaar-leaks-database/ (visited on 12/04/2018).

[2] *Aadhaar Dashboard*. URL: https://uidai.gov.in/aadhaar_dashboard/index.php (visited on 12/04/2018).

[3] *Aadhaar Fraud Is Not Only Real, But Is Worth More Closely Examining*. URL: https://thewire.in/economy/aadhaar-fraud-uidai (visited on 12/05/2018).

[4] *Aadhaar Generation - Unique Identification Authority of India — Government of India*. URL: https://www.uidai.gov.in/enrolment-update/aadhaar-enrolment/aadhaar-generation.html (visited on 12/05/2018).

[5] *Aadhaar Platform — State of Aadhaar*. URL: https://stateofaadhaar.in/focus_area/overview-architecture/ (visited on 12/06/2018).

[6] *Aadhaar Scam Did Not Stop with Kingpin - Times of India*. URL: https://timesofindia.indiatimes.com/city/hyderabad/Aadhaar-scam-did-not-stop-with-kingpin/articleshow/12916396.cms (visited on 11/08/2018).

[7] Shweta Agrawal, Subhashis Banerjee, and Subodh Sharma. "Privacy and Security of Aadhaar: A Computer Science Perspective". In: *Economic & Political Weekly* 52.37 (2017), pp. 93–102.

[8] *Architecting the World's Largest Biometric Identity System: The Aadhaar Experience — MapR*. URL: https://mapr.com/blog/architecting-worlds-largest-biometric-identity-system-aadhaar-experience/ (visited on 11/03/2018).

[9] *Biometrics: Who's Watching You?* Sept. 14, 2003. URL: https://www.eff.org/wp/biometrics-whos-watching-you (visited on 12/04/2018).

[10] *CCC — Chaos Computer Club Breaks Apple TouchID*. URL: https://www.ccc.de/en/updates/2013/ccc-breaks-apple-touchid (visited on 12/06/2018).

[11] Saikat Datta. *India's Ambitious Digital ID Project Faces New Security Nightmare*. URL: http://www.atimes.com/article/indias-ambitious-digital-id-project-faces-new-security-nightmare/ (visited on 11/08/2018).

[12] Alt News Desk. *Is GOI's National Informatics Centre Also Culpable for Abhinav Srivastav's Aadhaar Data Hack Incident?* Aug. 11, 2017. URL: https://www.altnews.in/gois-national-informatics-centre-also-culpable-abhinav-srivastavs-aadhaar-data-hack-incident/ (visited on 12/07/2018).

[13] Kamalapathi Rao H. *Illegal Aadhaar Centres Spring up All over Telangana*. Feb. 8, 2016. URL: https://www.deccanchronicle.com/nation/current-affairs/080216/illegal-aadhaar-centres-spring-up-all-over-telangana.html (visited on 12/05/2018).

[14] *How a SIM Card Operator in Hyderabad Apparently Created His Own Aadhaar Database*. URL: https://thewire.in/tech/aadhaar-database-breach (visited on 11/09/2018).

[15] *IIT Kharagpur Graduate Hacked Aadhaar Data through Digital India App: Police*. Aug. 4, 2017. URL: http://indianexpress.com/article/india/iit-grad-hacked-aadhaar-data-through-digital-india-app-cops-4781447/ (visited on 11/08/2018).

[16] *Illegal Patch Allows Easier Access to India's Aadhaar Biometric Database*. URL: https://www.bleepingcomputer.com/news/security/illegal-patch-allows-easier-access-to-indias-aadhaar-biometric-database/ (visited on 12/05/2018).

[17] Lucas Laursen. *Aadhaar, India's Biometric ID System, Gets Its Day in Court*. Feb. 21, 2018. URL: https://spectrum.ieee.org/computing/software/aadhaar-indias-biometric-id-system-gets-its-day-in-court (visited on 11/03/2018).

[18] *Mandatory National IDs and Biometric Databases*. URL: https://www.eff.org/de/issues/national-ids (visited on 12/04/2018).

[19] *Month after SC Verdict, Telcos Still Using Aadhaar for Issuing Mobile SIM Cards - Times of India*. URL: https://timesofindia.indiatimes.com/india/month-after-sc-verdict-telcos-still-using-aadhaar-for-issuing-mobile-sim-cards/articleshow/66370680.cms (visited on 12/05/2018).

[20] Jyoti Panday. *Aadhaar: Ushering in a Commercialized Era of Surveillance in India*. June 1, 2017. URL: https://www.eff.org/deeplinks/2017/05/aadhaar-ushering-commercialized-era-surveillance-india (visited on 12/04/2018).

[21] Vikas Sharma. "AADHAAR - A Unique Identification Number: Opportunities and Challenges Ahead". In: *Research Cell: An International Journal of Engineering Sciences* 4 (September 2011), p. 10.

[22] Amber Sinha and Srinivas Kodali. *Information Security Practices of Aadhaar (or Lack Thereof): A Documentation of Public Availability of Aadhaar Numbers with Sensitive Personal Financial Information*.

[23] *Supreme Court Extends Aadhaar Linking Deadline: Here's a Five-Year Timeline*. URL: https://www.businesstoday.in/current/economy-politics/aadhaar-linking-deadline-supreme-court-timeline-pan-card-bank-account/story/266066.html (visited on 12/05/2018).

[24] *Telecom Companies Ignore SC, Pester Users on Aadhaar Link - Times of India*. URL: https://timesofindia.indiatimes.com/business/india-business/telecom-companies-ignore-sc-pester-users-on-aadhaar-link/articleshow/63932602.cms (visited on 12/05/2018).

[25] *TRAI Chief Throws Aadhaar Dare, Shares Details Online - Times of India*. URL: https://timesofindia.indiatimes.com/india/trai-chief-throws-aadhaar-dare-shares-details-online/articleshow/65181805.cms (visited on 12/06/2018).

[26] Anand Venkatanarayanan. *Aadhaar — A Self Certified ID*. May 2, 2018. URL: https://medium.com/karana/aadhaar-a-self-certified-id-a63e299b36f5 (visited on 11/08/2018).

[27] J. Venkatesan. *UIDAI CEO Ajay Bhushan Pandey Says Aadhaar Security Can't Be Breached*. Mar. 23, 2018. URL: https://www.deccanchronicle.com/nation/current-affairs/230318/uidai-ceo-ajay-bhushan-pandey-says-aadhaar-security-cant-be-breache.html (visited on 12/06/2018).

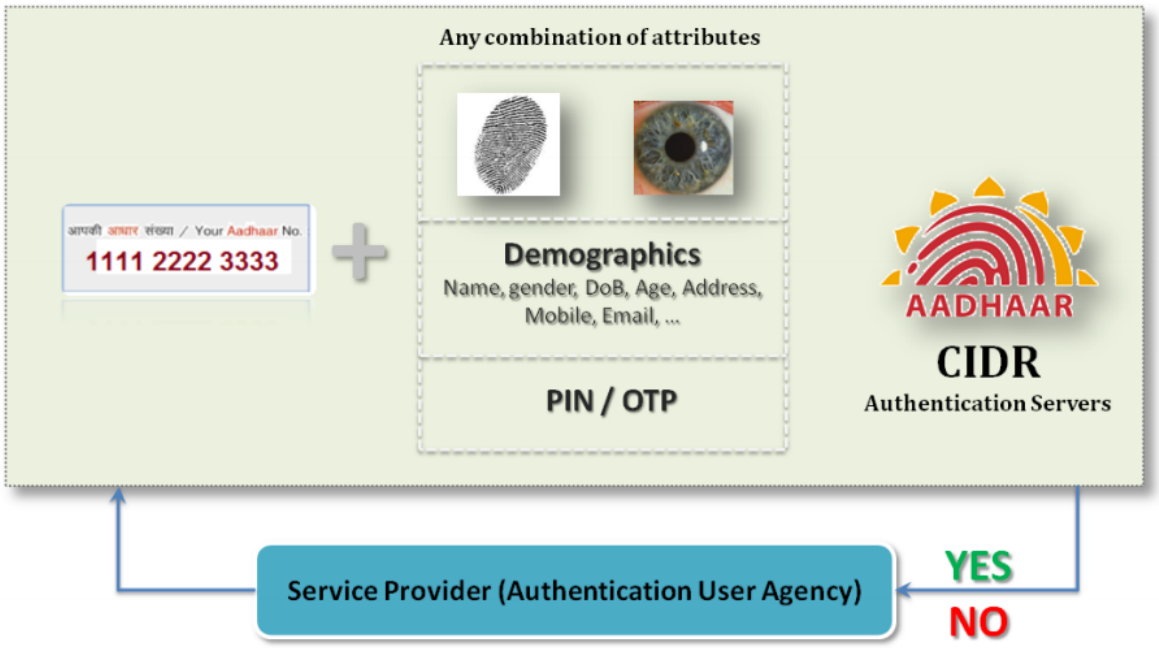

Figure 2: Aadhaar authentication blueprint [8]

```
publishProgress(new String[] { "Loading contents..." });
final Object localObject = new SoapObject("http://orsws/", AuthenticateNonAadhaar.this.METHOD);
((SoapObject)localObject).addProperty(AuthenticateNonAadhaar.this.addProperties("mobileno", AuthenticateNonAadhaar.this.number.getText().toString(), St
((SoapObject)localObject).addProperty(AuthenticateNonAadhaar.this.addProperties("userid", "mobileappors", String.class));
((SoapObject)localObject).addProperty(AuthenticateNonAadhaar.this.addProperties("intoken", "dG9rZW5Ad2ViQGFwcG9pbnQjbmlj", String.class));
try
{
  SoapSerializationEnvelope localSoapSerializationEnvelope = new SoapSerializationEnvelope(110);
  localSoapSerializationEnvelope.setOutputSoapObject(localObject);
  new HttpTransportSE("http://ors.gov.in/ORSServicecontainer/services?wsdl", 60000).call("http://orsws/", localSoapSerializationEnvelope);
  localObject = (SoapPrimitive)localSoapSerializationEnvelope.getResponse();
  AuthenticateNonAadhaar.this.finalresp = ((SoapPrimitive)localObject).toString();
  if (AuthenticateNonAadhaar.this.METHOD.equals("getRawOTP"))
  {
    localObject = new JSONObject(AuthenticateNonAadhaar.this.finalresp);
    if (!((JSONObject)localObject).getString("status").equalsIgnoreCase("Y")) {
      break label212;
    }
```

Figure 3: Hardcoded API token in the Aadhaar e-hospital app, decompiled by Anand Venkatanarayan and Anivar Aravind [12]