16th May, 2017

Clarification on the Information Security Practices of Aadhaar Report

We are releasing an updated version of the report, *Information Security Practices of Aadhaar (or lack thereof): A documentation of public availability of Aadhaar Numbers with sensitive personal financial information*.[1] In our report, we pointed the public disclosures by four government portals of Aadhaar Numbers and other demographic and financial data of beneficiaries of these projects. Our intent was to highlight the extent of public disclosures made by these portals and the ripe opportunity they presented for perpetrating identity as well as financial fraud. We clarify the reasons for changes made to report and also use this post to respond to some of the questions raised by key stakeholders and on social media.

Q1: Does the report claim or suggest that there has been a security breach in UIDAI's CIDR?

Since the publication of the report, we have noticed that media report have misunderstood or misrepresented our findings to suggest that the UIDAI's central repository, or the biometric database has been breached. Nowhere in our report has this been suggested and neither should this be inferred from the data in the report. For instance, the UIDAI CEO, Dr. Ajay Bhushan Pandey[2] and TRAI Chairperson Dr. R.S. Sharma[3] have written or been quoted in articles clarifying that there has been no breach of the CIDR. We completely agree with both Dr. Pandey and Dr. Sharma that the CIDR has not been breached, nor is this suggested anywhere in the report.

Q2: Where the publication of Aadhaar numbers and other data leaks or public disclosure?

The term 'leaks' was originally used 22 times in the report, this has led to reports that security measures on these portals were compromised. However, the intention behind the publication of this sensitive data was greater transparency and no access control measure was in place therefore this is best characterized as an illegal data disclosure or publication and not a breach or a leak. However, we must note here that the dictionary meanings of the word, leak, include 'intentional disclosure of secret information[4].

Further, in the original version our report, we had clearly stated that:

> *It is also important to appreciate that despite the trending twitter hashtag, #AadhaarLeaks, and media reports referring to these instance of Aadhaar data as leaks, these are, in fact, not cases of leaking of confidential data available only for private or controlled access. These are cases where the data in question has not been treated as confidential at all, and the government agencies in question have, in fact, taken pains to publish them. Rather than leaks or security breaches, these are wilful and intentional instances of treating Aadhaar Numbers and other PII as publicly shareable data by the custodians of the data.*

The presence of this paragraph in the report clearly indicates both our intent to classify this as data disclosure as well the intent of demonstrating the harm that such practices can have ramifications on the private information of citizens at large. The new report retains this paragraph, which should have additional context in the form of these clarifications.

Q3. Why are we updating the report?

The one significant change made in the updated version of the report is to change the phrasing of cases of publication of Aadhaar Numbers and other PII from 'leaks' or 'leakages' to 'public disclosure'. As researchers, we are sensitive to feedback we receive on our work and attempt to correct and call out misrepresentations of our work[5], to the extent possible. In this case, we felt that the word 'leaks' while not incorrect, was being misunderstood.

It is important to note that as far as the data subject is concerned, the term leak, breach or disclosure makes no difference as far as potential for identity fraud or financial fraud is concerned, including harm that may occur to the subject due to such disclosure, both prior to the report being published or in the time after the report entered the public domain.

Q4. What steps did we take to ensure that the PII published by the portals in question were taken down? What was the time period available to the concerned departments between CIS notifying them of the publicly available datasets and the report being published? Was this time period sufficient for them to carry out remedial action?

In the time period between our documentation of the portals in question in March-April 2017, and the publication of the report on May 1, 2017, we noticed that some of the portals had already masked the Aadhaar Numbers and other data. For the portals that had not done so, we reported these public disclosures prior to the publication of our report to the concerned government departments. We would like to note these retrospective mitigating steps may have little impact on the damage already caused by these public disclosures, as there was more than sufficient time for anyone looking at these sources of data closely to download detailed XLS files with sensitive PII, and misuse this information for fraud, prior to the report being published. *Further, following the principles of accountability and transparency, we also notified the affected government departments, including the UIDAI, of the public disclosures made by these portals prior to publishing the report on our website.* This ensured they were thus given ample opportunity to take down the sensitive data from the public websites.

Following this intimation, during our pre-publication checks, we discovered all the datasets mentioned in the report has been taken down by the concerned departments, prior to the report being published in the public domain. Considering that the datasets were publicly available not due to software vulnerabilities but due to a proactive choice to publish the data, this notice was sufficient to allow them to remove this information. Only when we were sure that the sensitive information had been taken down by the concerned departments, did we go ahead with the publication of report. At the time of publication, none of the entities had acknowledged our intimation or responded to us in any other form. With no additional harm to the data subjects possible due to the information having already been taken down, we considered the attention that would be garnered from an independently verifiable report (containing the URLs and tools) would aid in more robust security and privacy conscious practices in the future.

Q5. What was CIS's intent behind the act of including the URLs and tools used in the research report?

CIS, as a research organisation, believes in the full and open disclosure of the methodology and also the data where appropriate so that accuracy of our research can be confirmed

through independent replication. Due to CIS's intimation, all the Aadhaar numbers and sensitive personal financial information was either masked, or protected, or taken down by the concerned departments, prior to the publication of the report. All such information remains inaccessible as of this statement. This makes it clear that no additional harm to the data subjects occurred due to the publication of the report and any statement to the contrary is ill-considered.

Q6. Did CIS violate any legal or regulatory provisions in course of its research?

The relevant sections of the Information Technology Act, 2000 that some have argued is applicable in this case are Section 43 - "Penalty and compensation for damage to computer, computer system, etc.", Section 65 - "Tampering with computer source documents", and Section 66 - "Computer related offences", often known as the hacking provisions.

For CIS researchers to have violated either of these three provisions, it would need to be shown that they either could "accesses or secures access to such computer, computer system or computer network or computer resource" and did so "without permission of the owner or any other person who is in charge of a computer, computer system or computer network" [Section 43(a), and Section 66], or "tampered with computer source documents" [Section 65].

Neither of these provisions are applicable in this case, including in the case of changing the public URL string "login" to "nologin" while accessing the NSAP website, as can be seen in page 6 of the report. The reasons for this inapplicability are as follows:

1. **Prior and Proactive Publication** - All the databases mentioned in the report had been proactively published by the concerned government departments and had been available in the public domain for a significant period of time prior to the intimation provided to the government authorities and CIS publishing the report. These datasets were not protected or secured in a manner that would prevent an ordinary member of the public from accessing them.
2. **Public Availability via Search Engines** - In the absence of a robots.txt exemption or alternative means to restrict search engines access, the significant portion of the data was widely crawled and indexed by search engines and available in the to anyone tying in the right keywords on the search engine. This public availability via widely used search engines not only made it far easier for this information to be accessed but also made it possible for it to be discovered inadvertently, significantly compromising the privacy of the affected individuals.
3. **Lack of Access Controls for Sensitive Data** - The datasets used in the study were not guarded by any form of access control, including usernames, passwords or any other unique identifier that controlled access to them in any form. The lack of protection for such sensitive personal and financial information allowed for the data to be accessed without unique knowledge or significant effort, a fact that we have highlighted in the report.

Given that there was no access control placed on the data, that it was publicly indexed by search engines and all of this was enabled via the proactive publication of such data by government departments, it cannot remotely be claimed that access to the data was procured "without permission of the owner or any other person who is in charge of a computer, computer system or computer network". If what CIS researchers did violate the law, then every single person visiting a government website without taking prior approval

from the site's owner would be violating the law as well. Clearly that is not what the law is meant to do and as not been done in this case. Keeping these facts and the law in mind, there is no violation of the Information Technology Act, 2000 due to the research method adopted for this report, the intimation to government authorities prior to publication and the subsequent publication of the report.

Q7: Was there a violation of the law due to the actions of other parties?

While none of the research dealt with the CIDR at the UIDAI or any other information stored with the UIDAI, the proactive publication of such documents (by unrelated government departments) without access controls and allowing it to be indexed by search engines by concerned departments does have ramifications with regard to the Aadhaar (Targeted Delivery of Financial and other Subsidies, benefits and services) Act, 2016 and its Aadhaar (Sharing of Information) Regulations, 2016 (No. 5 of 2016). These liabilities exist between the UIDAI and the concerned departments which published this information, and do not involve CIS in any manner.

Further, the publication of Aadhaar numbers by the government portals was in violation of Section 29(4) of the Aadhaar Act, 2016, under Rule 6 of the Aadhaar (Sharing of Information) Regulations. They were also not mandated by the Right to Information (RTI) Act, under the section on proactive disclosure (Section 4). Indeed, Section 8(1)(j) of the RTI Act specifically states there is no obligation to release personal information which is not related to public activity or for the larger public interest.[6] CIS is not responsible for any of these practices nor for reporting the same via a documented, methodological and open access report.

ENDNOTES

[1] http://cis-india.org/internet-governance/information-security-practices-of-aadhaar-or-lack-thereof-a-documentation-of-public-availability-of-aadhaar-numbers-with-sensitive-personal-financial-information-1

[2] https://www.thequint.com/news-videos/2017/05/13/exclusive-or-data-leaks-from-aadhaar-not-possible-says-uidai-ceo-ajay-bhushan-pandey-pan

[3] http://blogs.economictimes.indiatimes.com/et-commentary/there-has-been-no-aadhaar-data-leak/

[4] https://en.oxforddictionaries.com/definition/leak

[5] https://twitter.com/pranesh/status/863132344890392578

[6] Section 8(1)(j) reads: "Notwithstanding anything contained in this Act, there shall be no obligation to give any citizen information which relates to personal information the disclosure of which has no relationship to any public activity or interest, or which would cause unwarranted invasion of the privacy of the individual unless the Central Public Information Officer or the State Public Information Officer or the appellate authority, as the case may be, is satisfied that the larger public interest justifies the disclosure of such information".