id:analytics®

WHITEPAPER

Exploring the Impact of SSN Randomization

March 2014

Table of Contents

Introduction

An individual's Social Security number (SSN) is synonymous with his or her identity to most organizations in the United States. The SSN was originally designed to track individual contributions to the Social Security Program. However, the absence of a better alternative has propagated the nine digit number across private industries and the public sector to become a near universal identifier. In the U.S., SSNs are the nucleus of identity management in both private and public sectors; driving policies for identity management, authentication and compliance activities. Institutions rely on a static definition of SSN legitimacy to manage identity issues. Any change to how the number is issued has the potential to generate a fundamental shift in the collective understanding of U.S. identity and may significantly impact the ability of risk managers to effectively evaluate and mitigate risk.

Beginning in the mid-1990s, the growth of online new-account applications and customer self-service options fueled a dramatic increase in identity-related fraud. While these electronic channels drove an improved customer experience, their inherent anonymity had unintended consequences. Legitimate consumers enjoy the increased automation available via Internet and mobile channels while fraudsters enjoy the ability to more freely misrepresent identity information. As the volume of identity-fraud incidents continued to grow, enterprises and public agencies began to voice concerns that the structured nature of the SSN was facilitating this behavior, inadvertently allowing criminals to use publicly available information to predict an individual's SSN. In response, the Social Security Administration (SSA) instituted a major change to the way it issues new SSNs.

The original system for issuing SSNs relied on chronology and geography to assign significance to the numeration and order of the nine-digit identifiers. On June 25, 2011, the system was abandoned for a new scheme where the digits began to be issued fully randomized. Since that time, SSNs issued to newborns, immigrants, victims of fraud, and victims of domestic violence no longer reflect any of the significance that allowed fraudsters to abuse the number. While this change appears to have succeeded in reducing the opportunity for fraudsters to predict a victim's SSN, the policy change has inadvertently compromised a series of traditional identity-management practices. Most risk managers rely on the previous structure to determine SSN validity, and it appears as though the policy change has created a new set of vulnerabilities for fraudsters to exploit.

Two years of data collected and studied by ID:A Labs, a multidisciplinary group on mathematicians, computer scientists, economists, financial experts, cognitive scientists and advisors from ID Analytics, suggest that the majority of risk managers have chosen to abandon traditional identity-management practices, opting to increase their fraud exposure to avoid inadvertent, disparate treatment of the affected groups. The randomization of SSN issuance appears to be dramatically impacting how firms manage the risk posed by new and existing identities. Risk managers should consider reviewing their current identity-proofing policies and studying the extent to which this policy change is affecting risk performance.

Background

Originally formulated in 1935 to track individual contributions to the Social Security Program, Social Security numbers are issued by the SSA to U.S. citizens, permanent residents and temporary working residents. SSNs never expire, are never reclaimed, seldom reissued, and with a few rare exceptions, are unique and issued to a single individual. While having a SSN is not mandatory, the Internal Revenue Service requires all corporations to collect SSN data from their employees and contractors for tax reporting purposes, and parents are required to submit a SSN for any dependent children they claim. As a result, most natural born citizens receive a SSN in early childhood and immigrants apply for one soon after entering the country. Despite its original intent, the uniqueness and broad coverage of the SSN has propagated it as a near universal identifier across government agencies and private organizations. Because individuals are often asked to submit their SSN when opening a bank account, applying for credit, or receiving medical care, it is nearly impossible to engage in daily life without a SSN.

The SSN is structured in three blocks with the format of "###-##-####". Prior to SSN randomization, the actual numbers and their ordering conveyed significant information. The first three digits, known as the *area number,* corresponded to a particular geographic region of an individual's mailing address. The next two digits, known as the *group number,* were assigned in a nonconsecutive yet predictable order within each distinct area number. The final four digits were determined serially and issued in order of application. The SSA regularly published the highest group number that had been issued for a given area code. This effectively divided the set of total possible numbers into issued and unissued ranges. That is, prior to SSN randomization, firms could use public information to determine whether an asserted SSN had been issued. Assertions of an SSN that fell outside of the issued range appeared highly suspect and were either typos or indicative of misuse.

The SSA does not retain any biometric information when it issues a SSN, making it next to impossible to know if the individual who asserts a SSN is the legitimate holder. The legitimacy of SSN ownership was exacerbated by the structured nature of the SSN, as it allowed fraudsters some ability to predict an individual's SSN given knowledge of his or her date and location of birth[1], allowing fraudsters to represent the number as their own. Fraudsters have long targeted SSNs as the only unique consumer identifier that public and private institutions trust for identity proofing, despite its vulnerabilities.

The SSA's decision to randomize the issuance of new numbers addressed the predictability of an individual's SSN, but had the unintended consequence of reducing the information available to risk managers. It is now nearly impossible for risk managers to distinguish between legitimately issued numbers and those that are being illegitimately asserted. Prior to randomization, a SSN in the unissued range sent a strong and explicit signal to risk managers. Whether the individual asserted the invalid credential with malicious intent, as a benign attempt to escape a marred past, or as typographical error, it was clear that something was not right with the stated information. Organizations realized that SSNs in the unissued range were often associated with high risk, and required remediation. The tools and policies they developed to prevent identity fraud reflected these learnings and often applied more stringent, yet warranted, scrutiny toward individuals asserting SSNs from the unissued range.

Unfortunately, post-randomization of these traditional policies can inappropriately flag legitimate SSNs issued after July 2011. To avoid applying disparate treatment to legitimate consumers, risk managers have been forced to become more agnostic towards SSNs coming from the previously unissued range, forfeiting an effective method for combatting fraud and constraining their ability to effectively manage risk. Organizations must now consider which techniques will help account for this change, particularly as the risk for exposure to attacks is expected to grow as more randomized SSNs are issued.

[1]Acquisti & Gross, 2009

Tracking SSNs in the ID Network

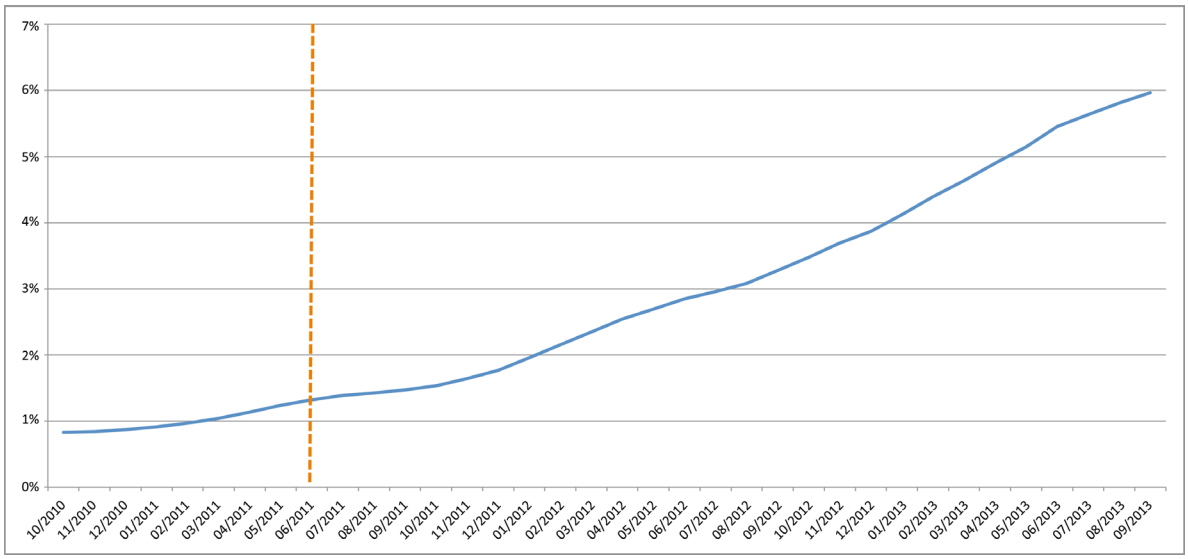
ID Analytics' ID Network® is a consortium of consumer behavioral data built through the contributions of more than 250 enterprise clients. As of Q1 2014, six of the top ten U.S. financial service institutions, three of the top four U.S. wireless carriers, and seven of the top ten U.S. credit card issuers contribute identity and outcome data into the ID Network, often in real-time. To date, the ID Network has amassed over one trillion aggregated data elements, more than 3.3 million client-confirmed frauds, and is growing at a rate of 54 million new identity elements every day.

ID Analytics leveraged the cross-industry data in the ID Network to determine how SSN randomization is impacting lenders and service providers. The analysis primarily focused on SSNs from the unissued range; evaluating the frequency of assertion before and after randomization; the rate at which members of the ID Network are booking (approving) applications with these SSNs; and the rate at which these new accounts result in fraudulent losses.

As of January 2014, the ID Network had visibility into over 3.6 million SSNs classified as coming from the unissued range and were asserted during consumer-initiated interactions (applications for loans, account management activities, etc.). This population includes illegitimate, unissued SSNs asserted before randomization and a mix of illegitimate and legitimate, randomized SSNs asserted since July 2011. As would be expected, ID Analytics has observed a rapid increase in the proportion of newly asserted SSNs (SSNs not previously seen in the ID Network) coming from the unissued range since randomization took effect (See Figure 1).

Figure 1.

Percent of Newly Asserted SSNs Coming from the Unissued Range



Approximately one percent of newly asserted SSNs came from the unissued range prior to randomization. These observations represented both benign misuse from unintended errors (e.g., transposition of numbers, transcription mistakes), as well as malicious behavior resulting from intentional obfuscation. After randomization, ID Analytics observes a rapid acceleration in the assertion of SSNs from the unissued range. Nearly six percent of all newly asserted SSNs came from this range by the end of 2013, comprised of a growing population of individuals asserting legitimate credentials in addition to cases of mistakes and misuse.

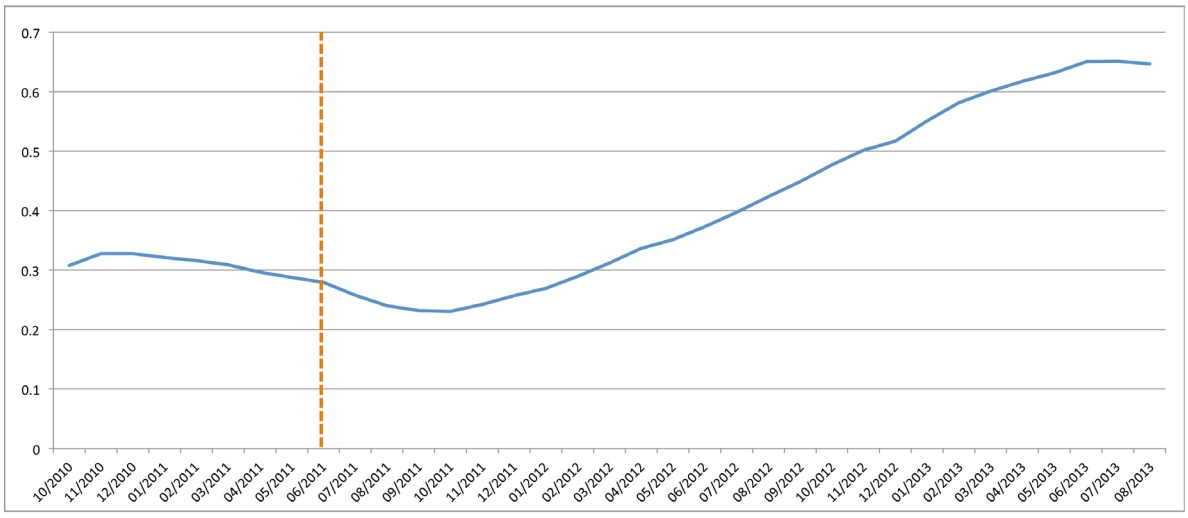
While the majority of randomized SSNs issued to date are to newborns, these individuals are not yet engaged in financial activity and are not observed in the ID Network. Therefore, the majority of individuals asserting legitimate SSNs from the unissued range are likely immigrants applying for loans and services. This cohort of legitimate consumers will continue to be overrepresented by new immigrant populations until newborns, with randomized SSNs, gradually become credit active.

There is no definitive way to differentiate between the intentional abuse of unissued SSNs and the assertion of legitimately issued SSNs from the unissued range in the absence of a structured SSN. The implication of these observations is significant. Prior to randomization, the use of a SSN from the unissued range was one of the most effective fraud-detection signals. That signal is now primarily reflective of behavior from one of the most coveted and protected consumer segments in the U.S.: immigrants. Risk managers must now understand how the SSA's policy change has confounded their ability to accurately separate the legitimate assertion of new SSNs, primarily by immigrants, from illegitimate assertions from identity fraudsters.

Prior to randomization, ID Analytics observed a relatively lower booking rate for applications asserting a SSN in the unissued range. This was explained by the higher risk these applications represented and the more stringent level of remediation that they required. Post-randomization, it is expected that the difference in booking rates would narrow as the risk profiles between the issued and unissued range converge. Any persistence in the disparity may highlight legacy procedures that have yet to adjust to the new SSA issuance policies. Examining the ratio of booking rate in the unissued range to that of the overall population (Figure 2) is one way to investigate this concern. Ratio values closer to zero reflect a large spread in booking rate, while values closer to one suggest a minor difference in booking rate between the two groups.

Figure 2.

Ratio of Booking Rate in Overall Population to Booking Rate of Individuals w/ Previously Unissued SSNs

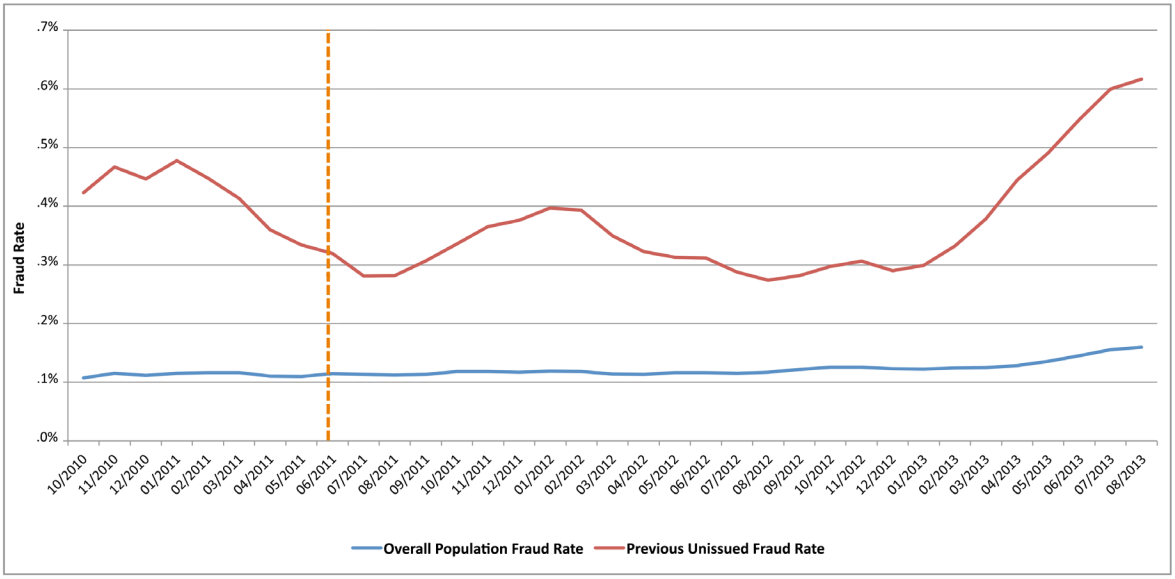


Beginning in late 2010, ID Analytics observes an initial decrease in the booking rate, which extends through September 2011. This corresponds to an increasing spread in the booking rate between the unissued range and the overall population, and reflects policies that may assign more stringent treatment to individuals in the unissued range. Despite a brief lag period after randomization went into effect, relative booking rates in the unissued range begin to increase in late 2011 through 2013. This trend likely corresponds to risk managers recognizing that legitimate, randomized SSNs do not represent the same risk previously conveyed by the unissued range, and updating their procedures to reflect these learnings. While this reflects a positive development for consumers – individuals with legitimate, randomized SSNs are not being unduly subjected to increased scrutiny – it also implies that risk managers have abandoned a powerful fraud mitigation technique.

Without a mechanism to differentiate between legitimate, randomized SSNs and cases of misuse, risk managers appear to be erring towards increased bookings to avoid disparate treatment of specific demographics, and accepting the increased risk exposure. If organizations are effectively managing this new paradigm, and are prepared for this new risk, then they should expect a relative decline in fraud rates within the unissued range as randomized SSNs become more prevalent. However, a relative increase in the fraud rate may indicate that these updated risk policies are ineffective and are allowing malicious applications to slip-through.

Figure 3.

Fraud Rate Comparison:
Overall Population Vs. Previously Unissued SSNs



ID Network members report a cyclic decline in the fraud rate within the unissued range beginning in late 2010 and extending through early 2013 (Figure 3). This trend reverses in February 2013 and gives rise to a dramatic increase in fraud rates throughout the remainder of 2013. This coincides with a similar, though far less dramatic, trend in the overall fraud rate. Though the long-term persistence of this trend is unclear, it has several plausible explanations. The initial increase in the unissued range fraud rate may represent less sophisticated attacks that risk managers were able to respond to quickly and effectively. However, the rapid increase beginning in early 2013 suggests that those initial adjustments may not be robust to recent, more sophisticated attacks. Nevertheless, ID Analytics observes evidence that perpetrators of fraud are successfully exploiting the ambiguity of the post-randomization unissued range. The sudden and rapid growth of fraud in this population is of concern and merits ongoing observations to determine the long-term implications of the trend.

Taken as a whole, these analyses highlight three critical and inter-related trends for risk managers:

- The number of previously unissued SSNs being asserted to enterprises is growing
- Enterprises appear to be responding by approving a far higher percentage of previously unissued SSNs
- By loosening restrictions on previously unissued SSNs, enterprises appear to have exposed themselves to an increased level of risk, and have yet to find adequate remediation techniques

Conclusions

SSN randomization presents substantial challenges for any organization that relies on SSN to evaluate identity risk. The proper response will require a concerted, cross-organizational investigation. Risk managers should evaluate the severity of the problem in their own environments, identify areas of strength and vulnerability, and respond with an updated approach.

The first step for risk managers should be to consider how SSNs are used across their organization, and the degree to which current policies and processes are impacted by randomization. They should undertake an initiative to evaluate the tools and processes currently in operation. Moreover, these analyses should revisit any policy updates made in response to SSN randomization, as observations within the ID Network suggest that they may require tuning. Consider the following questions during the review process:

- How does the current identity-proofing process depend on SSN? Has the process been updated to accommodate the SSA's policy change?
- How has randomization affected remediation procedures as part of the new-account onboarding process?
- Who developed the identity verification tools and policies? Were they created internally or via an external vendor? Who is responsible for maintaining these solutions?
- On what data does the process rely? Who provides this data and where does it come from?
- What have vendors done to respond to SSN randomization?
- What solutions are there to distinguish between new, legitimately issued SSNs, benign errors and malicious assertions of SSNs that have never been assigned by the SSA?

The challenges posed by SSN randomization are beyond a simple fix and require significant resources and expertise. In an environment where risk managers are consistently asked to do more with less, determining next steps can be both difficult and confusing. Many organizations may simply lack the ability to respond to SSN randomization internally and in a timely manner.

ID Analytics is here to help. Backed by over a decade of experience in the field of identity risk management, ID Analytics is a thought leader in the evolution of identity and an authority of its use in the information economy. The topic of SSN randomization captured the company's attention from its initial announcement in 2011. ID Analytics has dedicated scientists, analysts, and other experts to research the topic and to regularly monitor the impact of randomization across public and private sectors. One result of these efforts has been a pioneering technology capable of differentiating legitimate SSNs issued to immigrants, typographical errors and cases of deliberate misuse. Under guidance from cross-functional leadership, this advanced technology has been incorporated into new solutions to directly address the challenges of SSN randomization. Armed with technology, research and experience, ID Analytics is ready to collaborate with organizations in confronting the growing challenge of SSN randomization, and regaining the advantage over the perpetrators of identity fraud.

id:analytics®

www.idanalytics.com