On Unique Personal Identifiers

Rodrigo Borges Nogueira, Ricardo Staciarini Puttini
University of Brasilia, Brazil

Abstract

Countries around the world have established different practices when designing, implementing and maintaining their identification systems to keep a registry of all their population. On analyzing the features of many of these systems, one can clearly recognize the unique personal identifier particularities in each setup. In this paper, we intend to reason about the design decisions related to the personal identifier, considering it as an attribute of an individual and prepared to be used in a wide range of applications, from daily social interactions to e-government and banking.

1. Introduction

Many countries in the world have developed identification systems with very different assumptions and design decisions, mainly based on its own history, political structures, culture and technological advancements. Even in consolidated political-economic unions like the European Union, there are many different systems setups and some projects are trying to establish levels of interoperability and integration, like the STORK Project [1].

There are frameworks and well-documented success cases of policies, methodologies and technologies related to identification registry systems. For example, the Indian Aadhaar project [2] has made publicly available many details about their project. Estonia is also a good example of a well-establish Id Project [3] that fostered other systems such as the Internet voting.

A small, but important, component of an identification registry system is the design of a strategy for the unique personal identifier, which we will refer to as UPId for short. This component seems to be neglected in some setups, given that problems arise from bad choices on how an UPId is constructed. We have identified examples where the identifier design has changed over time in order to address privacy issues or political and social decisions.

It is debatable and controversial how national identification registry projects affects citizen's privacy and how it might be used as a surveillance mechanism, so we consider this matter out of scope of this paper. The problem of privacy related to personal attributes and the use of personal identifiers by private and public sectors has already been consistently addressed in many papers and official reports or documents, as in [4] and [5]. We solely intend to analyze a proper design for the UPId, in such a way that it can be used as an attribute of any person, like the name or birthdate, for instance.

Privacy is a big and complex topic itself. In times of huge amounts of information flowing on the Internet, and suspicions of country level spying, privacy has gaining increasing attention in both political and technical circles. In the context of identification registry systems, many debates also arise, but we will consider a restricted point of view, by looking only at the unique personal identifiers in the context of the principle of data minimization.

The principle of data minimization is present in most personal data protection legislations around the globe. Those legal frameworks are mainly based on the OECD definition of the principle [6]. The later basically states that personal data should be collected only if it is essential for a requested service completion and under user permission.

2. Motivation

Curious and unfortunate facts surround the existence of unique personal identifiers around the world. For example, in South Africa, from the 1950 Population Registration Act No 30 until the end of the Apartheid regime, the person race was an explicit part of the UPId, along with other personal information, namely birthdate, sex and a flag for citizenship [7].

In 2009, researchers showed that the Social Security Number (SSN), commonly used as an official Id document in the United States, could be predicted from public data [8], compromising the sensitiveness of the identifier. In 2011, the Social Security Administration (SSA) has changed the way SSN numbers are generated, adopting a randomization method.

The Estonian Id Project, who claims to be "…by far the most highly-developed national ID card system in the world." [3], like many other countries, established an UPId containing personal information, such as sex, birthdate and place of birth. There is a free tool available on the Internet that uses a citizen personal information to guess the Estonian ID card number and provides a Search button in the webpage to conveniently search for the generated ID on the Internet [9].

Many other countries generate an UPId based on personal information. However, it results on an attribute of an individual that relates to other attributes. Given the potential of usage in a broad variety of applications, the personal identifier might be a source of information leakage, because it carries information that might not be required.

The common issue with the three real identifiers exemplified in this section is the presence of semantic information in the UPId. The ability to autocomplete some fields in a form, given the UPId, cannot justify the availability of some extra personal information, possibly without the knowledge of the user. Additionally, the usage of meaningful data results on low entropy of the identifier, requiring a bigger UPId to represent uniquely an individual of a population. To better illustrate this issue, we take as an example the Mexican CURP, Clave Única de Registro de Población [10]. This unique national identifier is composed by a sequence of eighteen characters, which includes letters of the person name and the place of birth. It also includes numerical data such as the birthdate. Besides the low entropy and possible privacy issues, the CURP design might have usability issues as well, because the simple verbal communication of the CURP is harder than most other numerical identifiers around the world.

3. Personal Identity Number Properties

We propose the analysis of eight essential properties when designing an UPId. Some of them may be applied under social, political or cultural context of a group or society. However, we do not intend to analyze any of these matters, as they can lead to very complex scenarios based sometimes on ideological opinions.

3.1. Codification

In order to propose a convenient codification, we need to identify common use case scenarios of a unique personal identifier. Among them, it is expected that a user presents his/her UPId to a counterparty in the following situations, which we will refer to as *personal transferring* from now on:

- Printed, stamped in a document or even as a handwritten note;
- By communicating it verbally, as in a phone call;
- Through methods like sign language or Braille system;
- Typing it in a keyboard or numpad, as in an ATM or a phone keypad.

Despite the fact that an alphanumeric identifier needs fewer characters to represent uniquely each member of a population, the numerical representation might be the best choice in most cases. Numbers are easier to recognize, especially among illiterate people. Additionally, the sound of some letters can be confusing. This is a well-known issue and addressed by methods like the ICAO Radiotelephony Procedure [11].

3.2. Identifier domain

The unique personal identifier domain should be sufficiently large to last for an arbitrarily long period of time. It is necessary to prospect the population growth in that period in order to determine how big the domain should be. At the same time, the later should be small enough for better usability. Other factors can affect the decision about the size of an UPId domain, such as the possibility of reuse, reserved subdomains or even the visual symmetry of the identifier, because of the representation format.

In the case of a nationwide identification registry system design, it is sometimes appropriate to consider the existence of other entities under the management of the same system, such as businesses, a public service position or sector, an indian tribe and others. These entities, represented by real people or a group of people, may leverage some functionalities, logistics and legal framework established for the identification system.

For any entity other than people, it is helpful to define a different identifier domain to enable a clear visual distinction of the entity type.

3.3. Visual Format

The visual format of an identifier helps the recognition of characters groups, making the memorization and the personal transferring easier and more efficient. The format is basically the combination of the identifier characters in groups and a separator used between those groups.

A group of characters in an identifier should fit comfortably in the short-term memory of an average individual. Fortunately, psychologists have been studying the short-term memory capabilities for many years and, especially from [12], we can establish the number four as an appropriate choice for the size of a group of characters in a personal identifier. People got used on representing telephone numbers in groups of four digits and big numbers in groups of three digits, for example.

The immediate visual recognition of an UPId is a quality to be pursued when designing its format, in such a way that this format works as the visual identity of the UPId. The right separator for the groups of characters depends on the local culture. For example, Americans use commas as thousands separator when representing a number, but most Latin America countries use a dot. The hyphen can be the right choice for many cases.

3.4. Identifier Statuses

The UPId is an *agnostic* identifier, with the potential of use in different contexts. In order to improve interoperability and maintainability of systems and policies, we establish only four different statuses for a personal identifier:

- Valid: UPId is chosen from the proper domain related to the entity, it follows the format (optional) and the checksum (see section **Error! Reference source not found.**) is correct;
- Invalid: If any of the statements above do not stand;
- Active: Identifier linked to an individual and able to engage on transactions in the identification system;
- Inactive: Identifier linked to an individual, but not available to be used, due to a number of situations, such as no use after an arbitrary long time.

Any other status may exist in a separate business context. All active and inactive identifiers are valid.

3.5. Referencing

When referencing an entity system-wide, the corresponding unique identifier should not be used as a key. For this purpose, a more convenient way is to use a surrogate key [13]. When comparing the later to an UPId, we can identify that both concepts have the uniqueness property in common, but other properties may not hold in some scenarios.

Ideally, an UPId is immutable after the assignment to an individual. However, it is a personal attribute with its own semantics and subjected to the rules of a continuously mutable world. Thus, the identifier immutability cannot be always guaranteed. Moreover, an UPId can be manipulated by the system and change its format or domain over time, for example.

An UPId is subjected to the issues mentioned above and those issues are not expected to happen with a surrogate key, by definition.

3.6. Privacy

Having or not some personal data embedded on an UPId is yet another expression of the tradeoff usability × security: one can automatically fill up some information other than the pure identity attribute, just by typing the identifier characters in a proper field of a web form, for example. On the other hand, it is possible that some not requested information leak when an UPId is used.

A personal identifier carrying any personal information may not comply the principle of data minimization, as some unsolicited personal data can be acquired by a counterparty without user consent. Frequently, most users do not even recall which piece of information are present in his/her personal identifier.

Many national UPId composed by personal data can be deduced with a relatively high probability from information such as birthday, place of birth and gender. Besides, roughly speaking, their uniqueness is constructed from non-unique data attached to some random or sequential characters to treat matching cases.

Like any other attribute of an individual, the UPId has its own semantics and it is used in specific contexts. Then, it is appropriate to make it independent from other attributes to avoid changes in its structure after a nationwide deployment, for instance. The South African Id and American SSN are two examples of this situation, as presented in the Motivation section. In the case of the Mexican CURP, a simple change in the name, due to marriage, would result in an inconsistent personal identifier, despite of whether or not it would imply in practical issues. As a more general example, the changing in an attribute domain, such as gender, would require changes in many existing UPId.

Considering the issues exposed above, it is easy to see that generating personal identifiers at random is a sound design decision. The sequential generation is not recommended, because it could leak information about someone's age and it would be easy to guess an active UPId.

The usage of a (pseudo)random generator for the UPId has another advantage: it complicates the guessing of an active UPId. It is possible to keep an arbitrary probability of finding by chance this UPId as a function of the identifier domain size and the number of active identifiers. On increasing the domain size, we can obtain a lower probability of UPId guessing, although the hardness of a corresponding UPId trial and error process does not offer a cryptographic security level, as it would result in an impractical identifier system.

The concept of pseudonyms might be handy for preserving the privacy of a real personal identifier in some contexts. As a volatile and revocable identifier, the pseudonyms domain is bigger and separated from the UPId domain. In addition, it is usually generated at random. Many works have addressed the use of pseudonyms, as in [14].

Finally, we note that the discussion about (pseudo)random generators and algorithmic randomness is a broad area of research, as in [15].

3.7. Checksum

In personal identifiers, the checksum is a redundant information commonly presented as the last character in a sequence of characters. It only

checks if there are errors in the sequence, but do not correct those errors, as it would require more information attached to the identifier. There are a number of algorithms used to calculate a checksum, some of basic weighted sum and others more elaborated, as in [16].

The existence of a checksum is especially relevant to avoid that a user inputs an UPId that matches another valid or active one. In addition, it detects whether a user has typed the identifier correctly before data submission in the frontend validator of an online system, for example, in order to save some network and server processing resources in case of a mistype. The checksum is also helpful in offline systems, where the UPId active state validation and other business specific validation are not available. Many other typical use case scenarios of an UPId may not validate the checksum most times, as in a personal transferring, for example, even when a simple checksum algorithm is used.

Nowadays, most national identification systems functionalities are online with low response times. The generated network traffic and processing limitations may not justify the existence of an additional character in an UPId, as in some decades ago. Recent technologies can speed up the remote validation or grant some level of correctness to input data, such as AJAX or web sockets capabilities, auto completion mechanisms available in web browsers or in soft keyboards of many mobile devices and server-side in-memory databases. Then, the online validation of a personal identifier on the server side is cheap and fast.

Therefore, an UPId is usually submitted along with some data for authentication purposes, such as a password or biometrics. In this scenario, one can question whether a checksum is necessary or not, despite of the tradition of its existence in official documents in many countries.

Even when typed without the corresponding authentication data, it is possible to generate personal identifiers without a checksum by using (pseudo)random generators, so that, in a given instant in time, all active UPId are uniformly distributed in their domain. In this case, one can produce $\mathrm{UPId}_1, \mathrm{UPId}_2 \in F_{10}^{(n)}$, where $F_{10}^{(n)}$ is the base-10 domain of all UPId of length n, such that the Hamming distance [17] is, for example, $d(\mathrm{UPId}_1, \mathrm{UPId}_2) = 2$. This value for d allows the detection of all single-digit errors, which is the most common mistake. To catch more errors, we just consider an arbitrary greater value for the Hamming distance among all active UPId.

Recall that we have discussed in the previous section about the suitability of bigger domains to reduce the probability of guessing an active UPId. When deciding on do not implement a checksum, one can leverage this growth of the domain to properly generate uniformly distributed identifiers. However, on guessing any active $\mathrm{UPId_n}$, given an active $\mathrm{UPId_a}$, an attacker can ignore all neighbors $\mathrm{UPId_x}$ in the designed Hamming distance D, i.e., where $d(\mathrm{UPId}_x, \mathrm{UPId}_a) < D$. Then, it might be necessary to increase even more the UPId domain size to reach a minimum accepted level of guessing probability.

3.8. Exclusion filters

An exclusion filter drops any personal identifier that matches the rules implemented in the filter. Many reasons can justify the creation of such filters; some are even unexpected, like those based on culture or religious beliefs, for instance. As an illustration, we can cite the American SSN: "No SSNs with an area number of 666 have been or will be assigned." [18]. The area number is represented in the first set of three digits in the SSN.

A rule of this type is most likely to be applied to avoid identifiers with too many repetitions of a single digit or identifiers that might produce a sense of privilege in some individuals, like those with an UPId 0011221100, for example. In this context, the exclusion rule definitions are arbitrary and they intend to provide *virtually random* numbers.

The application of wide exclusion filters will dramatically reduce the available identifiers in a domain, demanding bigger numbers to represent an individual.

4. Proposal for a unique personal identifier

Considering all characteristics we have discussed about personal identifiers, we can propose a worldwide unique personal identifier. It would be used globally in a federated environment where each country manages its own identification system. The principles of identifiers generation are also appropriate in other scenarios, such as a federated environment of universities sharing services for their students.

We can propose different modes of UPId generation and distribution. It will be described three strategies.

In a global identifier, no personal attribute is expected as a component of the UPId. However, we intend to consider one exception, which is the nationality or the political-economic union embracing that nationality. In general terms, this attribute will represent a state or entity recognized outside its borders, sovereign and sufficiently empowered to establish policies accepted by a population whose members are supposed to be identified by UPIds managed in the context of the state/entity.

In order to abstract the concepts of nation, country, state, etc., we will simply use the expression *identity issuer*.

4.1. Explicit location in the identifier

For the first strategy, on submitting an UPId in a foreign service, a user could simply use, for example, an ISO 3166-1 (or other unique naming standard) code of his/her country, concatenated to the personal identifier. Then, for example, an identifier 123456 of a person from Brazil could be represented as br123456.

The advantage of this strategy is the UPIds decentralized generation. It is not necessary to create a global set of rules to enforce the generation of those identifiers in order to guarantee their uniqueness globally. On the other side, the presence of an explicit country identifier might not be intended.

4.2. Centralized identifier issuer

In another strategy, a central UPIds generator creates groups of unique (pseudo)random numbers to each identity issuer. Each of those group of numbers can be delivered in encrypted and digitally signed packages. In this scenario, a central and trusted generator authority is necessary. Additionally, it is possible to generate identifiers not directly linked to a specific identity issuer, i.e., without an embedded attribute. We cannot establish whether this feature is an advantage or not, because it will depend on the decisions of each identity issuer as a sovereign entity. Then, it might be necessary to embed a piece of information in the identifier to represent that identity issuer.

4.3. Decentralized strategy

In an UPId, a personal attribute is *visually distinguishable* when an average person is able to tell apart, with minimal effort, the value of such attribute. As an example, one can easily checks the value of the nationality attribute if it is a fixed code embedded in a specific position of an UPId, or if it is within a range of predetermined values.

Even when the nationality, for example, is required as a component in a personal identifier, we do not want a fixed value for it, because this value can harm the entropy, integrity or domain's distribution of the identifiers in case the attribute is no longer required or convenient in the future.

4.3.1. Requirements for a decentralized strategy. Let us propose some requirements for UPIds generation in our federated scenario, as follows:

- Each identity issuer has to be able to generate identifiers for its population;
- A recognized and sovereign entity code is embedded in the identifier;
- The identification numbers do not have any visually distinguishable information;

To accomplish these requirements, a setup phase is necessary in order to have every identity issuer agreeing on some public information, which are used to establish the rules for the UPIds generation.

4.3.2. Identifiers domain properties. Let D be the global domain of all possible identifiers UPId_{c_i} generated by the n identity issuers $c_i \in C$, where C is the set of all issuers with cardinality $|C| = n$ and $1 \leq i \leq n$. Also, let f_i be an arbitrary constant used to define the size of a subdomain $D_i \subset D$. The set D_i contains all identifiers for the population in c_i; and the estimated c_i population at an arbitrary point in the future is α_i. The following relations should hold in our scenario:

$$\bigcup_i D_i \subseteq \mathrm{D} \quad (1)$$

$$|D_i| \geq f_i . \sigma_i \quad (2)$$

$$\mathrm{UPId}_{c_i} \in D_i \quad (3)$$

The arbitrary value of the constant f_i is such that the cardinality of D_i makes the guessing of an active UPId_{c_i} more difficult, as we described in section 3.6. It might be appropriate to establish a small range of possible values for f_i, or even a unique constant F, as disparate values have a global impact on the size of D, resulting on bigger identifiers.

Note we are more interested in establishing the size of the subdomain D_i, than in defining the actual values for D_i. This is due to the way identifiers are generated. Each issuer will be able to generate identifiers without any interaction with a central authority or with any other issuer. An identifier is guaranteed to be unique among all c_i in our decentralized generation strategy.

4.3.3. The setup phase. In the setup phase, a central authority distributes a set P_{c_i} of prime numbers to each identity issuer c_i, where $P_{c_i} \neq \emptyset$ and $P_{c_i} \subset P_{c_j}$ for all $i < j$: $1 \leq i, j \leq n$, i.e., a c_i owns a nonempty set P_{c_i} which is a subset of P_{c_j}, the prime numbers set owned by c_j. The strict set order $P_{c_1} < \cdots < P_{c_i} < P_{c_j} < \cdots < P_{c_n}$ is a result of the estimated population number in each identity issuer, such that $\alpha_1 > \cdots > \alpha_i > \alpha_j > \cdots > \alpha_n$. That means the identity issuer c_1 with the bigger population α_1 owns the set P_{c_1}. The assignment of elements to any set $P_{c_i} = \left\{p_{c_i,1}, \ldots, p_{c_i,|P_{c_i}|}\right\}$, will follow the consecutive prime number sequence, starting with the prime number $p_{c_i,1} = 2$ up to a prime $p_{c_i,|P_{c_i}|}$. Before showing how construct P_{c_i}, let's understand the basic idea of an UPId_{c_i} generation.

4.3.4. Identifier generation. Any $\text{UPId}_{c_i} \in D_i$ is such that:

$$\text{UPId}_{c_i} = p_{c_i,1}^{\gamma_1}.p_{c_i,2}^{\gamma_2}\dots p_{c_i,k}^{\gamma_k} = \prod_{j=1}^{k} p_{c_i,j}^{\gamma_j} \qquad (4)$$

In (4), $p_{c_i,j} \in P_{c_i}$, $\gamma_j \in \mathbb{N}^*$ and $k \leq |P_{c_i}|$.

According to the *Fundamental Theorem of Arithmetic*, we know that the factorization of any number, in this case an UPId_{c_i}, is a unique product of prime numbers, also known as factors. Then, a unique personal identifier can be generated by combining the prime numbers in P_{c_i}. A number of different methods apply. In [19], it is presented an efficient algorithm to generate a random factored number. Additionally, considering that an UPId_{c_i} has a relatively small size, it is viable to use less efficient methods, such as brute force or factorization algorithms to find an UPId_{c_i} close to a randomly chosen $r \in_R D$. It is not known whether a polynomial time integer factorization algorithm exists, but a practical size of an UPId_{c_i} can make the process viable with standard computer power.

4.3.5. Identifiers domain construction. In order to choose the prime numbers of P_{c_i}, we consider a counting problem that we solve using a computational brute force approach. It can be optimized in a number of ways to save computational resources.

We state the problem as: *Construct the sets P_{c_i} calculating incrementally the value of $|D_i|$ by computing how many $UPId_{c_i} \in D$ can be generated by the product of prime numbers in P_{c_i}.*

Firstly, construct P_{c_1} adding the prime 2 to this set. We count how many numbers satisfy $2^{\gamma} \in D$, $\gamma \in \mathbb{N}^*$, which is $\log|D|$ in this case. Then, do the same for the next consecutive prime 3 and for the products $2^{\gamma_1}.3^{\gamma_2} \in D$, $\gamma_1, \gamma_2 \in \mathbb{N}^*$. In general, count all possible products $\prod_{j=1}^{k} p_{c_1,j}^{\gamma_j}$, $p_{c_1,j} \in P_{c_1}$, $\gamma_j \in \mathbb{N}^*$ and $k \leq |P_{c_1}|$. At some point, we will find a $p_{c_1,j} \in P_{c_1}$ such that the cumulative counting of all generated composites is at least $|D_1|$.

Now, for P_{c_2}, start by adding to this set all elements in P_{c_1} and the consecutive prime number greater than the greatest prime number $p_{c_1,j} \in P_{c_1}$, $j \leq |P_{c_1}|$ and start the counting process to construct P_{c_2}. However, the counting with this set will consider only the prime numbers products not generated by the previous counting process with the primes of the set P_{c_1}. This can be accomplished just by considering products with at least one factor $p \in P_{c_2}$ such that $p \notin P_{c_1}$.

Then, we construct all P_{c_i}, $i > 1$ by counting prime products $\prod_{j=1}^{k} p_{c_i,j}^{\gamma_j} . \prod_{j'=1}^{k'} p_{c_{i-1},j'}^{\gamma'_{j'}}$, where $p_{c_i,j} \in P_{c_i}$, $p_{c_{i-1},j'} \in P_{c_{i-1}}$, $\gamma_j, \gamma'_{j'} \in \mathbb{N}^*$, $k \leq |P_{c_i}|$, $k' \leq |P_{c_{i-1}}|$ and $P_{c_{i-1}} \subset P_{c_i}$, i.e., count prime products that use at least one prime factor $p_{c_i,j} \in P_{c_i}$ that were not used by composites generated with the elements of prime sets previously constructed. For any P_{c_i}, any generated $\text{UPId}_{c_i} \in D_i$ has at least one factor $p \in P_{c_i}$ such that $p \notin P_{c_j}$, for all $j < i: 1 \leq i, j \leq n$. This is the guarantee of uniqueness of any $\text{UPId}_{c_i} \in D_i$ in the context of the entire identifiers generation system.

Considering the practical size of an UPId, one can factorize the identifier number in order to identify its identity issuer c_i. It is just necessary to check the greatest factor p of the identifier and search the corresponding P_{c_i} such that $p \in P_{c_i}$ and $p \notin P_{c_j}$, for all $j < i: 1 \leq i, j \leq n$. The greatest prime factor of an identifier works as a signature of the corresponding identity issuer.

We remark that the method described favors the maintainability of the identifier generation, because an eventual expansion of the UPId domain is easily achieved by adding new non-consecutive prime numbers to a set P_{c_i}. Besides, the generated numbers are not visually distinguishable, as they may look like a uniform distribution for users. However, it is easy to identify an identity issuer computationally.

Lastly, as the numbers of a D_i are distributed over a bigger domain D, the use of a checksum might not be necessary in the context of a c_i.

5. Conclusions

We have analyzed issues of existing personal identifiers of different countries in the world. On highlighting those issues and addressing common problems, we came up with a set of features for a more consistent design of an UPId.

Additionally, we have proposed mechanisms for unique personal identifiers generation in different settings.

6. References

[1] STORK 2.0, Secure Identity Across Borders Linked 2.0. [Online]. Available at: https://www.eid-stork2.eu/ [Accessed 01 Sept 2015].

[2] Government of India, Unique Identification Authority of India. [Online]. Available at: https://uidai.gov.in/ [Accessed 01 Sept 2015].

[3] e-estonia.com The Digital Society. Electronic ID Card.[Online].Available at: https://e-estonia.com/ component/electronic-id-card/ [Accessed 01 Sept 2015]

[4] H. Chives, "Personal Attributes and Privacy. How to ensure that private attribute management is not subverted by datamining", Conference on Communications and Multimedia Security, Sept. 15-18, 2004, Windermere, UK.

[5] WP13, "D13.3: Study on ID number policies," 14 Sep 2007. [Online]. Available at: https://lirias.Kuleuven.be/bitstream/123456789/205522/1/fidis-wp13-del13_3_number_policies_final.pdf. [Accessed 01 Sept 2015].

[6] OECD, "OECD Guidelines Governing the Protection of Privacy and Transborder Flows of Personal Data" 2013.

[7] African History Apartheid Era Laws: Population Registration Act No 30 of 1950 [Online] Available at: http://africanhistory.about.com [Accessed 01 Sept 2015]

[8] Acquisti, Alessandro, and Ralph Gross. "Predicting Social Security numbers from public data." Proceedings of the National academy of sciences 106.27 (2009): 10975-10980.

[9] M. Sidorin, "Estonian ID card number generator," 2015. [Online]. Available at: http://www.uzzersite.eu/labs/estonian-id-card-generator/index.html. [Accessed 01 Sept 2015]

[10] RENAPO (Registro Nacional de Población e Identificación Personal), "Composição identificador CURP," [Online]. Available: http://sistemas.uaeh.edu.mx/dce/admisiones/docs/guia_CURP.pdf [Accessed 01 Sept 2015].

[11] International Civil Aviation Organization, "Alphabet - Radiotelephony," 2005 [Online] Available at: http://www.icao.int/Pages/AlphabetRadiotelephony.aspx [Accessed 01 Sept 2015].

[12] Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. Behavioral and Brain Sciences, 24, 87–114.

[13] R. S. Wazlawick, Object-Oriented Analysis and Design for Information Systems: Modeling with UML, OCL, and IFML, Elsevier Inc., 2013.

[14] A. Pfitzmann and M. Hansen, "Anonymity, Unlinkability, Undetectability, Unobservability, Pseudonymity, and Identity Management – A Consolidated Proposal for Terminology". 15 Feb 2008.

[15] R. G. Downey e D. R. Hirschfeldt, Algorithmic Randomness and Complexity, Springer Science & Business Media, 2010.

[16] Verhoeff, J. "Error Detecting Decimal Codes", Mathematical Centre Tract 29, The Mathematical Centre, Amsterdam, 1969.

[17] R.W. Hamming, "Error Detecting and Correcting Codes", Bell Sys. Tech. Journ., Vol 29, pp 147 – 160, 1950.

[18] Carolyn Puckett, "The Story of the Social Security Number", Social Security Bulletin, Vol. 69 No. 2, 2009. [Online]. Available: http://www.ssa.gov/policy/docs/ssb/v69n2/v69n2p55.html [Accessed 01 Sept 2015].

[19] V. Shoup, A Computational Introduction to Number Theory and Algebra, Cambridge University Press; 2 edition, 2009.

