A Greek (Privacy) Tragedy: The Introduction of Social Security Numbers in Greece

Eleni Gessiou
Institute of Computer Science
Foundation for Research and
Technology - Hellas (FORTH)
gessiou@ics.forth.gr

Alexandros Labrinidis
Dept. of Computer Science
University of Pittsburgh
labrinid@cs.pitt.edu

Sotiris Ioannidis
Institute of Computer Science
Foundation for Research and
Technology - Hellas (FORTH)
sotiris@ics.forth.gr

ABSTRACT

We highlight the privacy issues that have arisen from the introduction of the Greek Social Security Number (AMKA), in connection with the availability of personally identifiable information on Greek web sites. In particular, we identify privacy problems with the current AMKA setup and present data from a web study we conducted in May 2009, exposing these problems. Given the anticipated ubiquity of AMKA in Greece in the future, along the lines of the Social Security Number in the US.

1. INTRODUCTION

The introduction of the Social Security Number (SSN) in the USA happened in 1936 [6]; its original purpose was to track individuals' accounts within the New Deal Social Security program. The SSN has since come to be used as a unique identifier for individuals within the United States for a variety of purposes, from the Internal Revenue Service (IRS) to student IDs at Universities (up until a decade ago). The SSN is generally required by financial institutions to set up bank accounts, credit cards, and obtain loans, partially because it is assumed that no one except the person to whom it was issued will know it. However, the ubiquity of the SSN and its use as an authentication mechanism for financial transactions have intensified the problem of identity theft [8, 9], prompting legislation against using the SSN as an identifier that is made publicly available (e.g., on the face of student ID cards at Universities [5]).

Greece recently introduced its own Social Security Number, called *AMKA*, that will become mandatory for all transactions relating to employment and insurance starting October 2009 (pushed back from the initial July 2009 deadline). Essentially, this number is used as the first step in the effort to modernize the numerous public pension and insurance plans. Although it is not expected to replace the national tax payer ID, already in place, we strongly believe that this number can take a path similar to that of the SSN in the US and become the ubiquitous and de-facto identifier number for Greek citizens in the near future, further increasing its importance.

Given the importance of AMKA, its current design and implementation are troubling, since they constitute likely failures in terms of privacy and on the potential for identity theft, especially in conjunction with personally identifiable information, that is publicly available on Greek web sites.

Although this paper represents the first scientific study that identifies these problems, while this study was being conducted, a number of Greek blogs have already pointed out the potential privacy issues with AMKA [4, 1]. A few others have also discovered the AMKA numbers (or their non-existence) of political leaders in Greece [3, 2], further illustrating the privacy concerns.

Contributions: This paper makes the following contributions:

- performs a study on the availability of personally identifiable information in Greek web sites.
- highlights the seriousness of the problem with the design and implementation of the Greek SSN (AMKA), given such availability.
- illustrates the potential to obtain someone's AMKA by presenting the results of a small-scale study (using both public figures and private citizens).
- proposes solutions to address the current shortcomings of AMKA.

This paper confirms the lack of awareness on very typical privacy issues with large-scale systems and the lack of guidelines or best practices to that effect, or at the very least, the lack of adoption of privacy standards in the real world.

2. AVAILABILITY OF PERSONALLY IDENTIFIABLE INFORMATION ON GREEK WEB SITES

The first part of our study is to quantify the availability of personally identifiable information in Greek web sites. Specifically: "*First Name*", "*Last Name*", "*Father's First Name*", "*Mother's First Name*", "*DoB*" (Date of Birth), "*ID#*" (National ID number) and "*Tax ID#*" (Taxpayer ID number). To quantify this, we queried Google with all possible permutations of these personal details, limiting our search

WPES'09, November 9, 2009, Chicago, Illinois, USA.
Copyright 2009 ACM 978-1-60558-783-7/09/11 ...$10.00.

Mother's FN	DoB	ID#	Tax ID#	Total
yes	yes	yes	no	50
no	yes	yes	no	1,724
yes	yes	no	no	1,983
no	yes	no	no	3,843
yes	no	no	yes	4,244
yes	no	yes	no	4,895
no	no	no	yes	15,806
no	no	yes	no	22,099
yes	no	no	no	63,211

Table 1: Availability of personally identifiable information for regular citizens on Greek web sites. A person's First Name, Last Name, and Father's First Name were present in all searches, so we omitted the columns for those attributes since they would all be *yes*. FN stands for First Name.

to `.gr` web sites. The results were filetype-restricted to `.xls` files, because large-scale publication of personally identifiable information happens primarily through spreadsheet files.

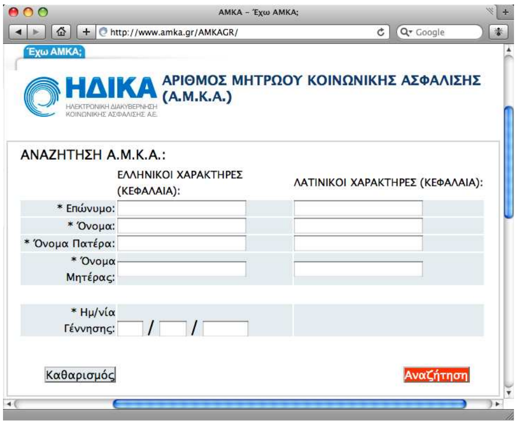
Table 1 shows the amount of personally identifiable information that is available in Greek sites in spreadsheet format[1]. The fields "First Name", "Last Name" and "Father's First Name" are omitted because they are always included in each row. The fields shown in the table were not always requested in the search query; the indication "yes" indicates their presence in the query and the indication "no" their absence. One may notice that half of the possible permutations in Table 1 are missing, this is because they returned no results. Note that the "Total#" should be viewed as the minimum volume of data that can easily be found on the Greek Web.

If we consider that an individual can be identified uniquely by their full name, their father's and mother's first name, then the fact that a malicious party can identify over 60 thousand people, is quite worrying. Moreover, it is surprising that more than 15 thousand taxpayer ID numbers are publicly available by simply querying a search engine.

Although knowing someone's first/last name, their father's and mother's first name, and their date of birth (i.e., for the 1,983 people in the third row of Table 1) might not seem to pose a serious privacy risk, the introduction of AMKA makes it quite dangerous. The fact that there is a way to find out a person's AMKA, as we describe in Section 3, by knowing these five pieces of information about them, triggered our curiosity about AMKA and motivated this paper.

3. ABOUT AMKA

AMKA has the following 11-digit format: `YYMMDDxxxyz`, where the first 6 digits encode the person's date of birth (`YYMMDD`), the following 4 digits are a *sequence number* for people born on that date (`xxxy`) and the last digit is a control digit (`z`). The sex of the person is encoded in the last digit of the sequence number (`y` of `xxxy`): even digits are assigned to women and odd digits are assigned to men[2]. This results in disclosure of both the date of birth and the sex of a person by solely looking at their AMKA.



ΑΡΙΘΜΟΣ ΜΗΤΡΩΟΥ ΚΟΙΝΩΝΙΚΗΣ ΑΣΦΑΛΙΣΗΣ (Α.Μ.Κ.Α.)

ΑΝΑΖΗΤΗΣΗ Α.Μ.Κ.Α.:

	ΕΛΛΗΝΙΚΟΙ ΧΑΡΑΚΤΗΡΕΣ (ΚΕΦΑΛΑΙΑ):	ΛΑΤΙΝΙΚΟΙ ΧΑΡΑΚΤΗΡΕΣ (ΚΕΦΑΛΑΙΑ):
* Επώνυμο:		
* Όνομα:		
* Όνομα Πατέρα:		
* Όνομα Μητέρας:		
* Ημ/νία Γέννησης:	/ /	

Figure 1: Screenshot of the AMKA's Web site form.

The second, and even more troubling failure in the implementation of AMKA, is in the way a person can find what their AMKA is. Currently, there is a heavily advertised web form in place (`http://www.amka.gr/AMKAGR/`), which asks for the following information:

- First Name (῾Ονομα)
- Last Name (Επὶθετο)
- Father's First Name (῾Ονομα πατὲρα)
- Mother's First Name (῾Ονομα μητὲρας)
- Date of birth (Ημερομηνὶα γὲννησης)

to provide a person's AMKA (Figure 1).

Although this can be seen as a big convenience for Greek citizens, to establish whether they have an AMKA (and what that number is) or not (and need to apply for one), it is also a problem in terms of potentially exposing many individuals to identity theft. A malicious third party can misuse such a system in three ways:

1. Find the AMKA of public figures, by obtaining the required data already available on the Web (e.g., in wikipedia or their Facebook profiles).

2. Find the AMKA of citizens, by gathering personally identifiable information that has already been (improperly) published about them on the Web, as we showed in the previous Section.

3. Find the AMKA of citizens for whom not all of the five above fields are known, simply by brute-force guessing.

Before testing these ideas, we wanted to be certain that such actions would not violate the web site's acceptable use policy, despite our investigation being done purely for academic purposes. To our surprise, we found that the link to the acceptable use policy was broken. In fact, it seems that it has not yet been supplied (as of June 19, 2009), since the link currently points to `http://www.amka.gr/#`, i.e., a placeholder.

4. METHODOLOGY AND RESULTS

In this section we present the methodology we used to conduct our study. Our goal was to discover whether it is

[1] Personally identifiable information was also found for the mother of one of the authors, while conducting this study!

[2] This information is included in the welcome letter sent out to some AMKA recipients and was mentioned in [1, 4].

possible to find the AMKA of a person and how much effort is needed to do so.

The data sets we used for our study include personal details of both public figures and regular citizens, found on the Web. As mentioned earlier we need the five attributes - “First Name”, “Last Name”, “Father’s First Name”, “Mother’s First Name”, “DoB” - in order to query the system for an AMKA. Finding all five of them for each individual would be ideal but not always the case, as Table 1 shows. That led us to divide our data-set in two categories: (i) *Exact Data*, for cases when we have all five fields for a person, and (ii) *Non-exact Data*, for when one or more fields are missing.

4.1 Exact Data

The *Exact Data* set consists of the personal details of individuals (public figures and private citizens) for whom we were able to collect all five of the required attributes. The strategy we followed was straightforward; we simply queried the system for each individual. To accomplish this with the minimum amount of effort we built a script, written in python, that automated the entire procedure. This script issued one HTTP POST request per individual with the appropriate fields filled-in and parsed the AMKA when available in the results web page.

	Total	with AMKA	%
Public	18	10	55.5
Private	1,983	1,490	75.1

Table 2: Results of the *Exact Data* set.

The results of the *Exact Data* set are shown in Table 2. The first row contains the results after querying for 18 selected Greek public figures. We collected information about them from public web sites (e.g., wikipedia, their personal pages, fan sites, etc.), all discoverable through Google. As far as the private citizens were concerned, we managed to collect `.xls` documents, mainly from two different sites, which contained all the personal information required for searching, that is “DoB”, “Mother’s First Name” etc. Most of these lists were found on a hospital’s website and their entries referred to nurses and midwives. As we can see, 3 out of 4 already had an AMKA which we collected successfully.

We should note that failure to find a person’s AMKA when having all of the five required fields can be attributed to one of three reasons:

1. The person’s data available on the Web is incorrect.
2. The person’s data entered in the AMKA database is incorrect.
3. The person does not have an AMKA number yet. This seems the most possible reason for the majority of our cases and, of course, will change when AMKA becomes mandatory (in October 2009).

To compare the public availability of SSNs in the US versus AMKA numbers in Greece, we used Google to assist us in collecting as many SSNs as possible. Our search resulted in the collection of only 270 SSNs. If we take into account the population of USA, which is in excess of 300 million and the population of Greece which is approximately 11 million, the percentage difference between the number of AMKAs and the number of SSNs that are publicly available is huge. Additionally we were able to find 27 SSNs of people that were deceased. This is an alarming fact because a malicious party can use one of these SSNs for illegal purposes, presenting to be the legal owner of it with less chances of getting detected.

4.2 Non-exact Data

Despite the fact that in most of the cases it was possible to find all the required personal details needed to query for an individual’s AMKA, there were cases that only some of them could be found. During our study we faced two distinct categories of missing attributes, either missing “Mother’s First Name” or “DoB” (day and month – the year was known). The way we dealt with both cases was brute-forcing the missing attribute. Doing so in the second case was trivial. We simply tried all possible combinations of day and month. In the former case we created a list of the most common Greek female names, 287 in number and tried them (see Table 3).

	Total	with AMKA	%
Public-NM	7	3	42.8
Public-ND	5	2	40.0
Private-NM	149	33	22.1

Table 3: Results of the *Non-exact Data* set. The -NM suffix stands for “No Mother’s First Name” and the -ND stands for “No Date of birth”.

The probability of guessing the missing date of birth is one, so in our case we are certain that only two public figures out of five have AMKA. Unfortunately, in the missing mother’s name case, we are not able to measure the success of our brute-forcing technique because we do not know the exact fraction of individuals that does have AMKA. We believe however, that finding the AMKA of almost one out of two public figures is a very high success rate, especially if we consider that it is almost the same as in the case where all personal details were available. Finally, we should note here that the low rate in the last row of Table 3 may be due to another reason. The data-set we used mainly consists of driving license candidates. This indicates that the majority of them are young people, probably 18 year olds, that may have not applied for an AMKA yet.

Using a larger subset of Greek female names we could have seen an increase in the success rate, but this was not the main goal of this study. Our goal was a proof-of-concept discovery of AMKAs and not an exhaustive search style database extraction.

Table 4 shows a breakdown by profession for the public figures for whom we found an AMKA. The majority of them are politicians, but we also found the AMKA of a celebrity. We also found the AMKA of three journalists from major TV stations and two AMKAs of famous Greek athletes.

Politicians	9
Journalists	3
Athletes	2
Celebrities	1

Table 4: Break down of public figures, for whom we discovered their AMKA, by profession.

4.3 Observations and Limitations

During the course of our study, we noticed some interesting characteristics and updates in the behavior of the AMKA web form (Figure 1) that we find worth mentioning.

First of all, we observed that if the day and the month of a date of birth are left blank, the system responds with a new form asking for the individual's Taxpayer or National ID number. An additional observation is that if someone fills the year of the field "DoB" incorrectly then the system does not returnl an AMKA. We infer from these two observations that the company which is responsible for setting up the system is making some effort in securing the citizens' AMKA. Also, the system recognizes synonyms of many common Greek names and thus assists users of the web form.

At some point however, while conducting our experiments we noticed the following odd behavior. We would fill all the fields normally except the "DoB". There we would only fill in the year of birth and the website would still return the AMKA. Fortunately, in the interest of privacy of Greek citizens, this only lasted for a few days in May 2009 and then was reverted back to responding with a new form asking for the individual's Taxpayer or National ID number, when the date of birth field was incorrect or partial. Another error we noticed was in the case that only the month was missing from the "DoB" field. In that case we get a pop-up with the following message "ORA-01858 COMMUNICATE WITH THE ADMINISTRATOR". This is clearly a database error [7], and we identified it to be: "a non-numeric character found where a digit was expected".

We also noticed that there is a limitation and a variation in the number of characters in the fields "First Name", "Last Name", "Father's First Name" and "Mother's First Name". The first two fields can hold up to 20 characters, the "Father's First Name" field can hold up to 15 characters and the "Mother's First Name" field can only hold up to 10 characters. The number of characters limitation may create problems in case of multiple first or last names, which is sometimes the case in Greece [1].

Another weakness of AMKA is the fact that only two digits of the 11-digit format of AMKA encode the person's year of birth. This may cause problems since it is impossible to tell if someone was born in 1908 or in 2008, solely by looking at their AMKA. As we already mentioned, the sex of a person is encoded in the next digit - odd digits are assigned to men and even digits are assigned to women. We found however, two cases that this was not the case. Both cases were for women, the digit was odd instead of even, and we believe to be in error.

4.4 Scenarios

Armed with the knowledge of a citizen's AMKA, an adversary has the opportunity to conduct several malicious and dangerous actions.

Private Data Confirmation A curious, or worse, malicious party, can use the online interface to confirm a citizens birth date, parents' names, etc. They can then use this information for other activities, from "harmless" gossip about someone's age, to impersonation.

Identity Confirmation An adversary may use the website to confirm guesses. Essentially one can plug in random values and get back confirmed identities, using the site as an oracle. If guessing is easy (hence cheap) you can actually use the interface to mine out the entire database. One can then use these identities, as before, for impersonation.

False Medical Payments Depending on how AMKA is used it may be possible to permit the generation of false medical payment records (medical exams, prescriptions, hospital visits, etc.) using harvested AMKAs. If a malicious party acquires a large number of them, they may be able to charge minute amounts in medical claims that will not trigger any alarm, but still generate substantial income.[3]

Identity Spoofing If AMKA starts being used for identity purposes, in the same way the National ID card, the passport number or the Taxpayer ID are currently being used, it can open a whole new way to commit identity spoofing with all the related problems this can cause.

Future Uses Finally, it is unclear at this point what other uses AMKA may have in the future. Therefore it is hard to estimate the full privacy impact it will have on the Greek citizens.

5. CONCLUSIONS

In this paper we presented the privacy issues that have arisen with the introduction of the Greek Social Security Number (AMKA). The privacy concerns stem from the fact that possibly malicious parties can extract Greek citizen AMKAs from the online portal. We also conducted a more general study on the availability of personally identifiable information on Greek sites in contrast to US sites and, through a small-scale study, found that there is a lot of it publicly available. To address these issues there are a number of steps one can take:

- Modify the AMKA web form as to always also require a person's Taxpayer ID or National ID for authentication purposes, not only in case of errors as is now. This simple solution should eliminate 100% of the AMKAs found in our study, as fas as public figures concerned.
- Provide a way for citizens to be taken off this online lookup service, with proper safe-guards to avoid DOS-style attack.
- Educate people to not post sensitive information about themselves or others on the web.

This work was supported in part by the Marie Curie Actions - Reintegration Grants project PASS.

6. REFERENCES

[1] http://blog.postmaster.gr/2009/05/20/on-amka/.

[2] http://dcorfu.blogspot.com/2009/05/o.html.

[3] http://lesvos.wordpress.com/2009/05/21/amka/.

[4] http://malvumaldit.wordpress.com/2009/05/06/e-government-center-for-social-security-societe-anonyme/.

[5] http://www.insidehighered.com/news/2007/07/24/idnumbers.

[6] http://www.ssa.gov/history/1930.html.

[7] http://www.techonthenet.com/oracle/errors/ora01858.php.

[8] H. Berghel. Identity theft, social security numbers, and the web. *Commun. ACM*, 43(2):17–21, 2000.

[9] S. L. Garfinkel. Risks of social security numbers. *Commun. ACM*, 38(10):146, 1995.

[3] http://www.nytimes.com/2009/06/13/health/13patient.html?hpw