Privacy Risks from Public Data Sources

Vassilis Prevelakis

Technische Universität Braunschweig

Joint work with

Zacharias Tzermias and Sotiris Ioannidis

ICS/FORTH (Greece)

Background

- Personal & private data are important
  - can be used to identify - track you
  - impersonate you
  - harass you
  - etc.
- Criminals want them all
  - eBay hack: millions of account details stolen
- Not a theoretical threat
  - Mat Honnan had his entire on-line presence destroyed
    - even his cell phone was deactivated
  - This data is used for over-the-phone authentication

Private Data and the State

- The state obsession with data is even more worrying
  - if the state wants something they get it
    - FBI and Lavabit
    - David Miranda and Snowden
    - remember: if the request is illegal, the state can change the law
  - the state is careless with data
    - Jet Blue case in 2004 (passenger profiling data)
    - HM Revenue & Customs lost 25M records (2 CDs sent by post in 2007)
    - Greek Tax Authority fined for leaking tax records (2013).
    - Israel’s voter registration database
  - Drive to link databases to discover offenders
    - water bills to find swimming pools
    - electricity bills to find occupied apartments (but declared empty)
  - Greek government “tax card”

Case Study

- Greek Government Data Sources
  - created and operated entirely by the Greek Government
  - uncoordinated
    - each serves different purpose
    - operated by different department
    - no common authentication framework
  - Major Repositories
    - Tax Registration Number Database
    - Diavgeia (transparency)
    - Voter Registration
    - National Health Service (AMKA)

Tax Registration Number DB

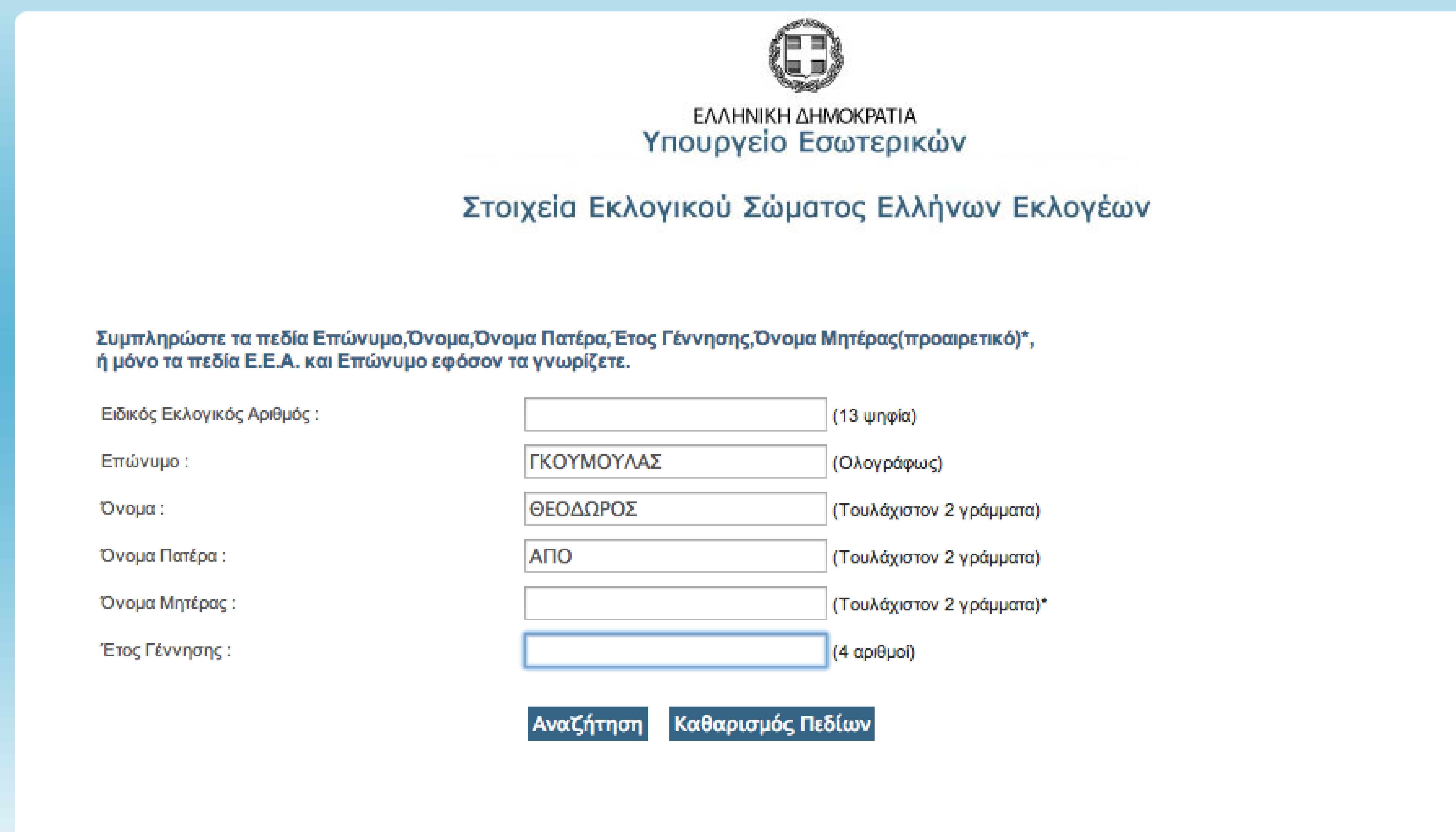
- Created to allow confirmation of Greek TRNs (AFM)
  - everybody in Greece needs a TRN, persons and companies
  - database provided access to records of companies and persons working free lance
    - but once you are in, you stay in
    - some records refer to people who are dead
  - user submits TRN and gets name, address, occupation, etc.
  - no authentication (since changed)
  - But search space is limited
    - TRN comprises 9 digits (9th is guard)
    - Search space is sparse, with clustering
  - Result: database was scraped (multiple times)
    - now its on Google

Diavgeia

- Web site to promote transparency in Government procurement
  - All government organisations must post on Diavgeia their purchasing decisions
- Good idea in principle
  - name and shame for big spenders of state funds
- But too detailed
  - treasure trove of personal information in hiring decisions
- No authentication
  - everybody has full access
  - allowed the creation of “value-added” front ends

Voter Registration DB

- Online site to answer “where do I vote?” question
- Weak authentication
  - asks for a few personal details (name, year of birth etc)
- Available all the time
  - even when no elections have been scheduled
- Provides full record of voter

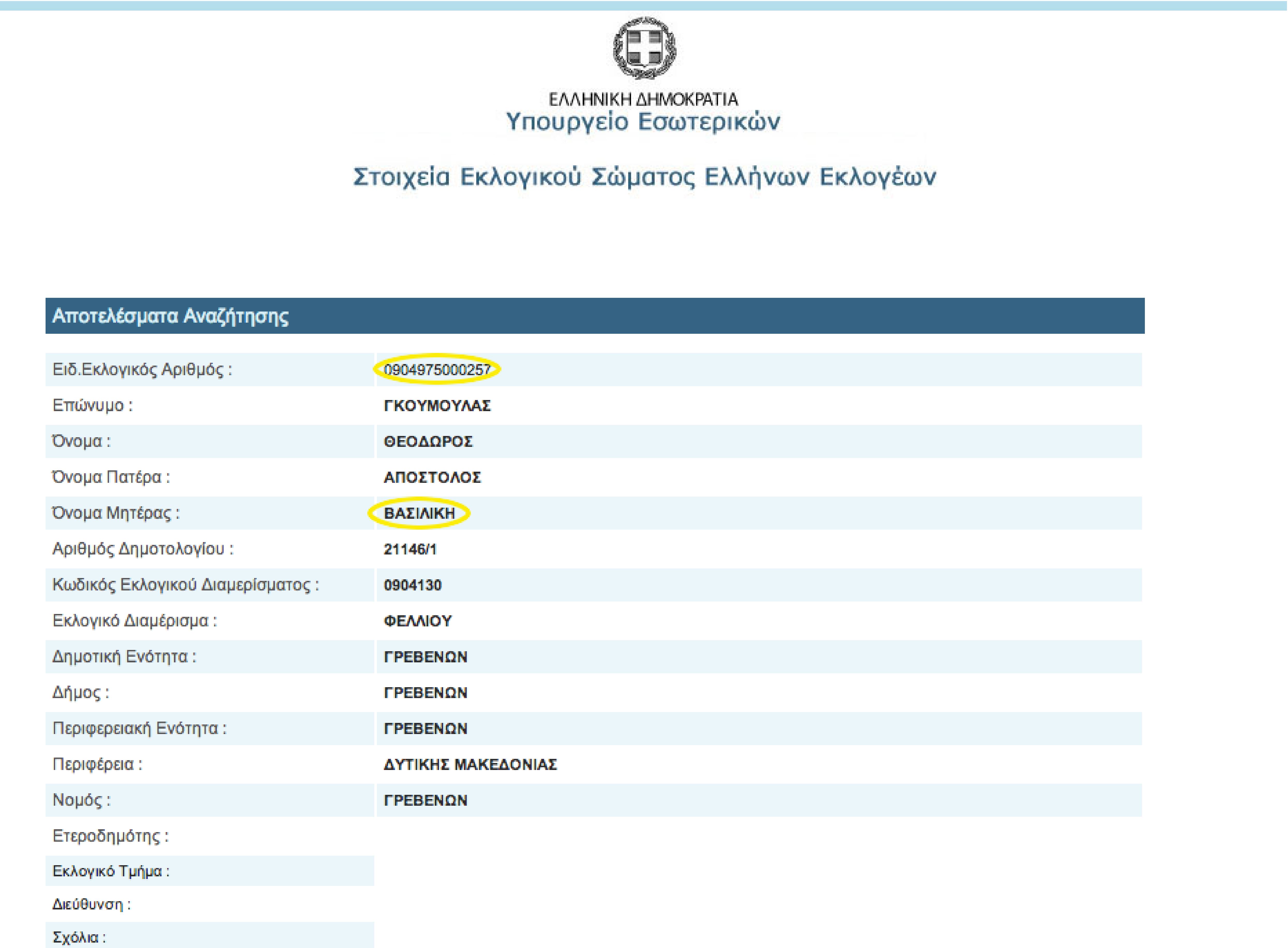


ΕΛΛΗΝΙΚΗ ΔΗΜΟΚΡΑΤΙΑ
Υπουργείο Εσωτερικών

Στοιχεία Εκλογικού Σώματος Ελλήνων Εκλογέων

**Συμπληρώστε τα πεδία Επώνυμο,Όνομα,Όνομα Πατέρα,Έτος Γέννησης,Όνομα Μητέρας(προαιρετικό)*,
ή μόνο τα πεδία Ε.Ε.Α. και Επώνυμο εφόσον τα γνωρίζετε.**

Ειδικός Εκλογικός Αριθμός :		(13 ψηφία)
Επώνυμο :	ΓΚΟΥΜΟΥΛΑΣ	(Ολογράφως)
Όνομα :	ΘΕΟΔΩΡΟΣ	(Τουλάχιστον 2 γράμματα)
Όνομα Πατέρα :	ΑΠΟ	(Τουλάχιστον 2 γράμματα)
Όνομα Μητέρας :		(Τουλάχιστον 2 γράμματα)*
Έτος Γέννησης :		(4 αριθμοί)



ΕΛΛΗΝΙΚΗ ΔΗΜΟΚΡΑΤΙΑ

Υπουργείο Εσωτερικών

Στοιχεία Εκλογικού Σώματος Ελλήνων Εκλογέων

Αποτελέσματα Αναζήτησης

Ειδ.Εκλογικός Αριθμός :	0904975000257
Επώνυμο :	ΓΚΟΥΜΟΥΛΑΣ
Όνομα :	ΘΕΟΔΩΡΟΣ
Όνομα Πατέρα :	ΑΠΟΣΤΟΛΟΣ
Όνομα Μητέρας :	ΒΑΣΙΛΙΚΗ
Αριθμός Δημοτολογίου :	21146/1
Κωδικός Εκλογικού Διαμερίσματος :	0904130
Εκλογικό Διαμέρισμα :	ΦΕΛΛΙΟΥ
Δημοτική Ενότητα :	ΓΡΕΒΕΝΩΝ
Δήμος :	ΓΡΕΒΕΝΩΝ
Περιφερειακή Ενότητα :	ΓΡΕΒΕΝΩΝ
Περιφέρεια :	ΔΥΤΙΚΗΣ ΜΑΚΕΔΟΝΙΑΣ
Νομός :	ΓΡΕΒΕΝΩΝ
Ετεροδημότης :	
Εκλογικό Τμήμα :	
Διεύθυνση :	
Σχόλια :	

NHS Number DB

- One of three distinct id numbers for Greek residents
  - national id card, AMKA, TRN
  - best practices advise to use separate numbers
    - only works if we don’t always store all three together
    - but, almost always at least 2 out of 3 are kept
- Weak authentication
  - asks for names and birthdate
  - a bit more strict than the voter registration
  - Sometimes asks for TRN as well
    - implies that they keep the TRN as well

Putting it all together

- start with a TRN
  - ➢ from that we get first name and last name
  - ➢ maybe father’s name as well
- go to the voter registration database
  - ➢ guess year of birth (no big deal, tiny search space)
  - ➢ may need to guess father’s name (only two letters)
  - ➢ now we know names of father and mother, plus year of birth.
- and then over to the AMKA site
  - ➢ need to guess the date of birth
    - □ but we already know the year
- all your data are belong to us :-)

Tools we used

- Mainly integrated off-the-shelf tools
  - parsing web pages
  - text searches
  - python scripts
  - etc.
- We also looked at other databases
  - e.g. military enlistment database
  - found similar weaknesses
- In some cases we failed
  - address space was too big
  - car registration number database

Analysis

- Databases used different authentication data
  - we used one against the other
- Easy questions
  - year of birth could be guessed
    - and the system confirmed the guess!
- Unlimited number of guesses
  - compare with ATMs
- TRN database wide open
  - no captchas
  - no rate limiting
  - no traffic analysis
  - now they ask for login (but the horse is gone)

Mitigation

- Data sanitization
  - Adopt a “need to know” rule, delete what you don’t need
    - eBay was keeping way too much information about customers
  - Keep information rather than raw data
  - e.g. criminal record vs a binary flag
- Rate limiting - traffic analysis
  - you need to know if someone is scraping your database
- Stop the bots
  - check to see you are talking to a human
- Coordination
  - maybe use a single login mechanism
- Accountability

Privacy Risks from Public Data Sources

QUESTIONS?

Vassilis Prevelakis

Technische Universität Braunschweig

Zacharias Tzermias and Sotiris Ioannidis

ICS/FORTH (Greece)