A Privacy framework for DNS big data

3 October 2014
Cristian Hesselman, Jelte Jansen, Maarten Wullink,
Karin Vink en Maarten Simon
SIDN

- "nl" (Registry voor Nederland)
- 5.5M domain names, >1.600 registrars
- > 1.000.000.000 DNS queries per day
- Private foundation with public task
SIDN Labs

• R&D team SIDN

• Improve services of SIDN

• Center of expertise

• Improve security of Internet in the Netherlands

• Facilitates external research
Privacy Framework

• What?

• Why?

• How?
Privacy Framework: Why?

• Public service that is vital to Dutch society and economy

• Keep trust and confidence in SIDN as the operator for .nl

• Responsibility to be proactive in the field

• SIDN wants to act transparently
Privacy Framework: Innovations

• Introduces Privacy management to the use of DNS data

• Integrates legal, technical and organisational aspects of privacy management
ENTRADA: DNS Big Data Platform

• ENhanced Top-level domain Resilience through Advanced Data Analysis

• Goal: Develop and evaluate big data applications
  • To Safeguard stability of '.nl'
  • To increase the safety of the (Dutch) Internet
  • To Detect botnets and abuse
  • Non-goal: commercial use

• What about privacy?
ENTRADA: DNS Big Data Platform

User Interfaces

Services en applicaties

DNS auto-config service

statistics dashboard

ENTRADA platform

Algoritmes

Basisfaciliteiten

Storage

DNS

Bronnen (intern en extern)

HTTP

EPP

ATLAS

...
DNS

1. User
2. ISP network
3. Resolvers
4. TLD Operator (e.g., .nl)
5. Child operator (example.nl)
6. www.example.nl
7. HTTP
8. ENTRADA Platform
9. Root operator
(Potential) Personal Data in DNS Queries

• IP Address

• Queried name

• 'other'
  • Timestamps
  • Protocol flags
  • Etc.
• Dutch Data Protection Act (Wet Bescherming Persoonsgegevens, WBP)

• Personal Data:
  • 'any piece of information regarding an identified or identifiable natural person'

• Processing:
  • ‘any action or sequence of actions involving personal data, including but not restricted to the collection, recording, sorting, [...] deletion or destruction of such data’
Requirements for Processing

• Public Function
• Contractual obligation
• Legitimate Basis
• Explicit consent
• Purpose Limitation
  • Personal data may only be used for the purpose for which it was collected
• Special Personal Data explicitly forbidden
  • Religion
  • Political views
  • Etc.
WBP and ENTRADA

• We are not using 'Public Function': too weak
  • besides, we are not government

• Explicit consent not possible
  • So we need to be completely transparent

• Legitimate basis + Purpose Limitation
  • The goal is for the benefit of the users themselves
WBP and ENTRADA: IP Addresses

• Can't simply anonymize them

• Most addresses are from resolvers (shared by users)

• BUT: Resolvers may be 'home' systems
Number of Queries per Resolver per Day
Number of Queries per Resolver per Day

![Bar chart showing the number of queries per resolver per day. The x-axis represents the average number of queries per day, ranging from $10^0$ to $10^8$, and the y-axis represents the number of resolvers. The chart shows a decline in the number of resolvers as the average number of queries increases.]
WBP and ENTRADA: IP Addresses

• Most individual resolvers are 'home' resolvers
  • Few users, so addresses likely to be personal data

• 'Big' resolvers either ISP or domainers
  • In the second case, still personal data

• Better metrics are future work
  • Problem: to decide whether the address is personal data, you need to process it
WBP and ENTRADA: Queried names

• Single query does not say much
  • 'www.universityoftilburg.nl'
  • Not even associated with Tilburg University ;)

• Combined data can be considered personal
  • Query patterns, pre-fetching

• Query names may include other personal data
  • Personal names (firstname.lastname.mycloud.nl)
  • IP addresses (192.0.2.1.customer-adsl.example.nl)

• Can also be combined with IP address of resolver (previous slides)
WBP and ENTRADA: IP Addresses in Queried names
Privacy Framework: Requirements

- Purpose limitation
  - Per type of use of the data (i.e. per application)

- Verifiable
  - Transparency

- Simple

- Extensible
Privacy Framework: Overview

ENTRADA privacy framework

- PEP-U: Security and stability services and dashboards
- PEP-A: Data analysis algorithms
- PEP-S: Storage
- PEP-C: Collection

Legal and organisational
ENTRADA data platform (technical)

R&D licence
Template
Author (Application Developer)
Draft Policy
Privacy Board
Policy

Adjustments

Template

nl name servers

DNS packets (PCAP)

Database queries

DNS queries and responses

Resolvers

ENTRADA data platform (technical)

Legal and organisational

ENTRADA privacy framework

Privacy Board

Author (Application Developer)

Draft Policy

Adjustments

Policy

Template

R&D licence

nl name servers

DNS packets (PCAP)

Database queries

DNS queries and responses

Resolvers
Privacy Framework: Policies

• One policy per application

• Policy describes:
  • Purpose
  • Data that is used
  • Filters on the data
  • Access to the data
  • Type of application (Research vs. Production)
  • Other security measures
Privacy Framework: Data Filters at PEPs

- Anonimization

- Pseudonimization

- Aggregation

- Etc.
Privacy Framework: Privacy Board

• Reviews and approves policies

• Members:
  • Legal
  • Technical
  • Organisational

• Publishes approved policies
Privacy Framework: Position Paper

• Currently at https://sidnlabs.nl
Privacy Framework: Conclusions

• DNS Data can be personal data

• DNS Data processing needs privacy-protecting measures

• Not 'just' technical
Privacy Framework: Future work

• Solicit feedback and discussion (hi!)

• Better metrics for 'public' vs 'private' resolvers

• How to incorporate the policy system when sharing data

• Keep eye on new laws (EU Data protection regulation, for one)

• Apply the framework to other types of data
Got questions?

Jelte Jansen
jelte.jansen@sidn.nl

sidn.nl | sidnlabs.nl