Obfuscation-Based Private Web Search

Claudia Diaz
KU Leuven ESAT/COSIC

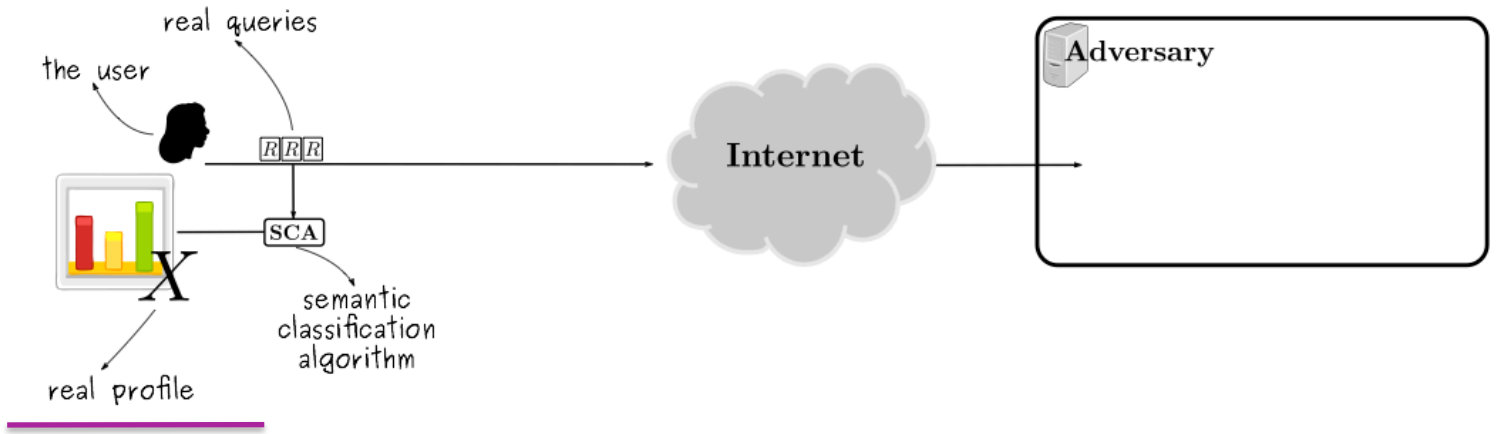
Obfuscation Symposium, 15 February 2014

Balsa, Troncoso, Diaz, (2012): “Obfuscation-Based Private Web Search” IEEE S&P 2012

General idea of OB-PWS

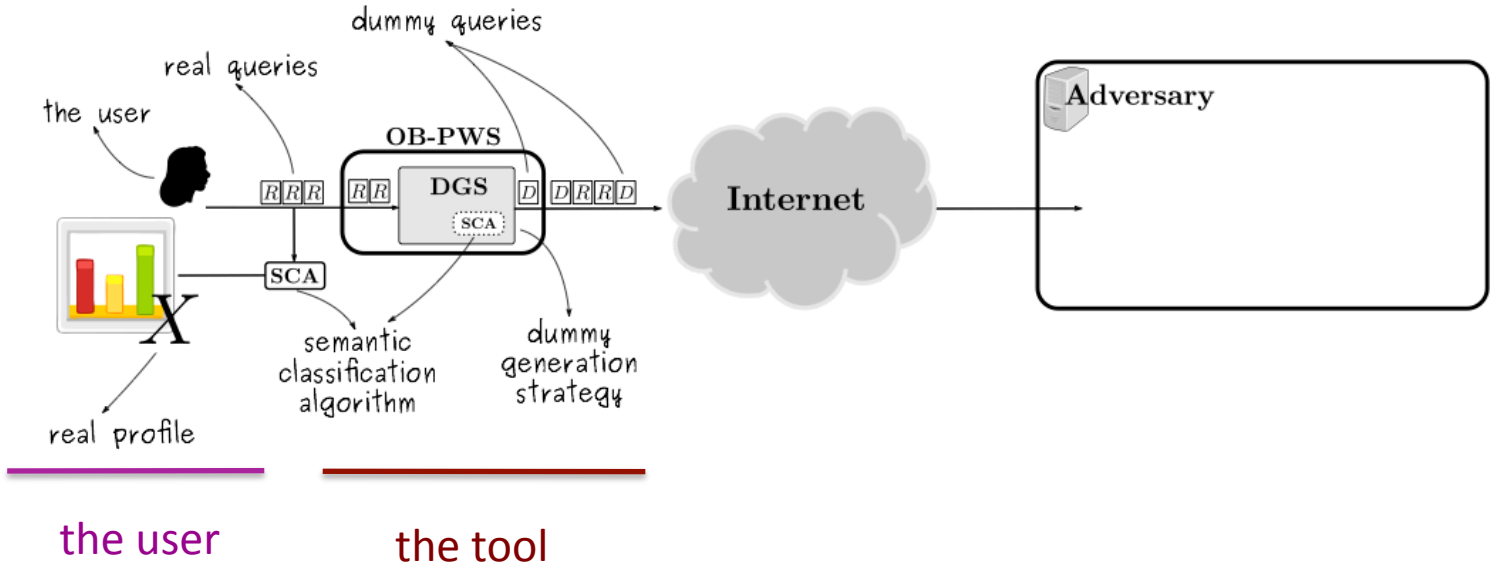
- The problem:
  - Search queries might be very revealing and potentially privacy sensitive
- The adversary:
  - The search engine itself, or 3rd parties with access to search logs
- The obfuscation strategy:
  - Automatically generate “fake” or “dummy” queries to introduce noise in the search logs
- Our contribution:
  - Analyze six proposed tools / obfuscation strategies from a computer security perspective, ie, taking into account a *strategic adversary*

OB-PWS model

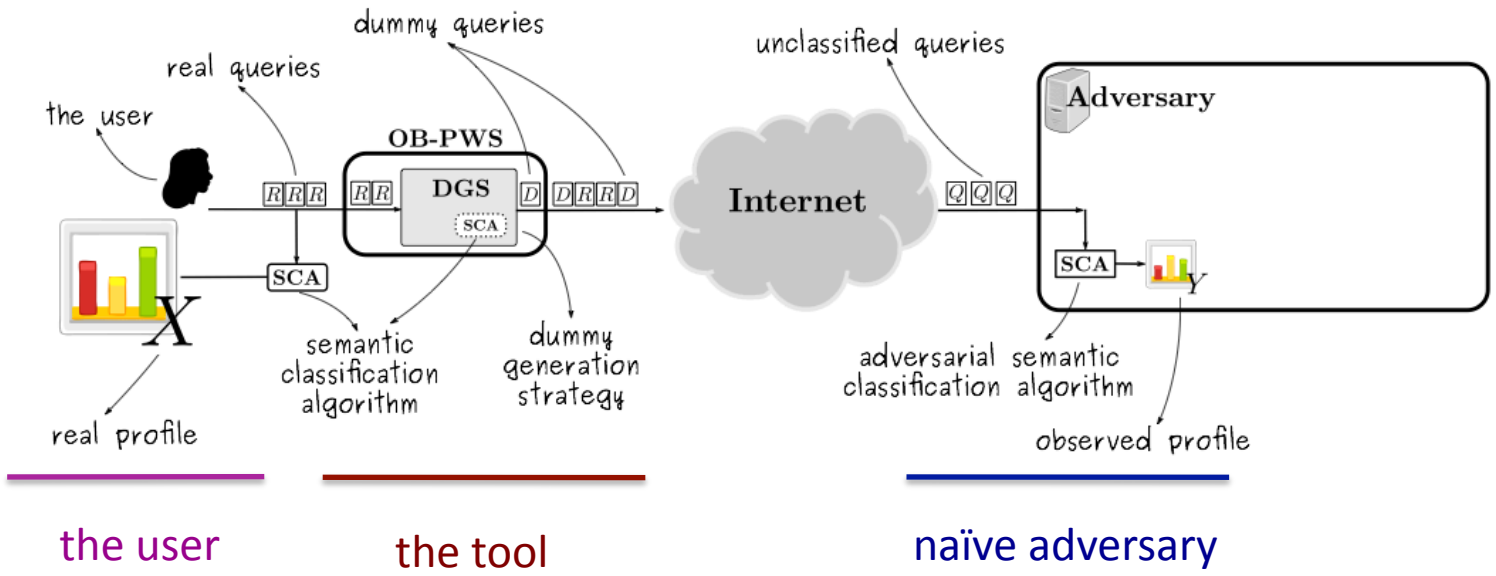


the user

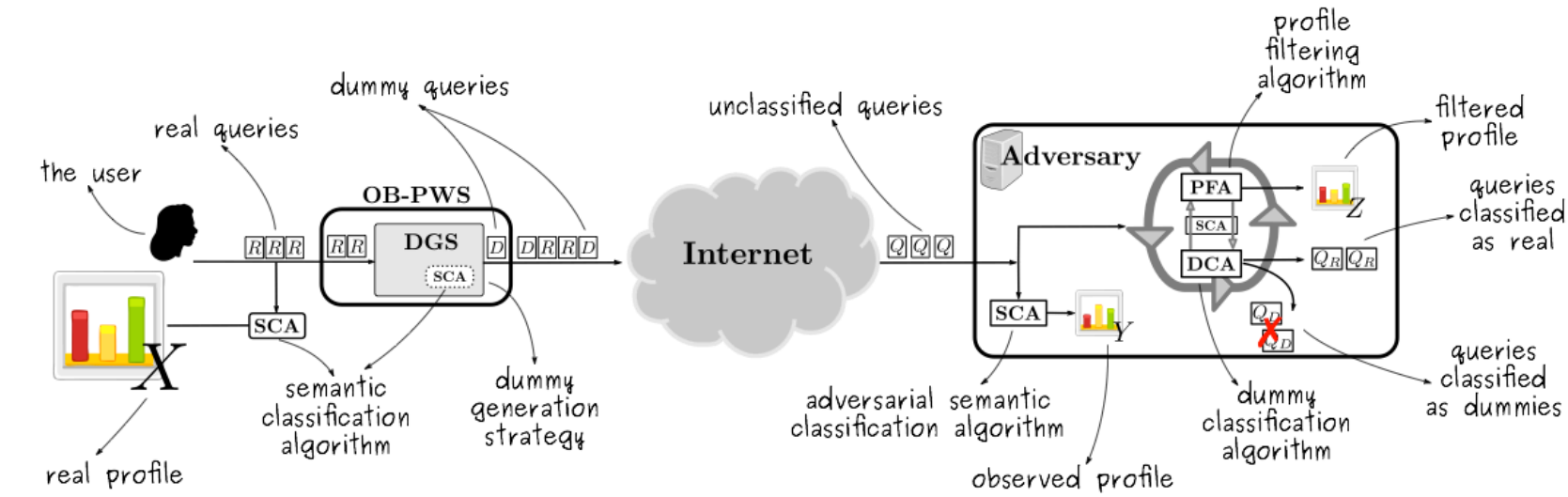
OB-PWS model


dummy queries
real queries
the user
OB-PWS
DGS
SCA
Internet
Adversary
SCA
X
semantic classification algorithm
dummy generation strategy
real profile
the user
the tool

OB-PWS model


dummy queries
real queries
the user
OB-PWS
DGS
SCA
unclassified queries
Adversary
Internet
SCA
X
Y
semantic classification algorithm
dummy generation strategy
adversarial semantic classification algorithm
real profile
observed profile
the user
the tool
naïve adversary

OB-PWS model

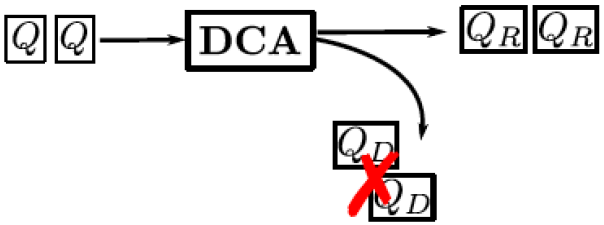


the user | the tool | naïve adversary | strategic adversary

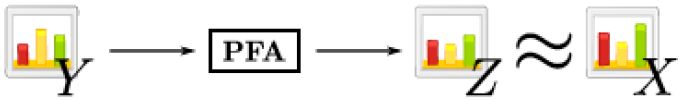
required from a computer security perspective !!

Query-based vs Profile-based

- Property: query indistinguishability, plausible deniability
  - Not possible to tell if a query is real (generated by the user) or dummy (generated automatically)
- Attack strategy
  - Exploit vulnerabilities in the DGS to distinguish real from dummy queries
  - Issue: query topic persistence (in time)



- Property: profile obfuscation, protection from profiling
  - The observed profile (real + dummy queries) should not leak information about the real profile (real queries only)
- Attack strategy
  - Exploit vulnerabilities in the DGS to filter observed profile *Y* and recover the real profile *X*
  - Specifically, exploit any predictability in the modification of profiles
  - Bad metric: real/observed profile (dis)similarity



Interactions between the two levels: concrete instance vs long-term patterns??
- Query distinguishability hurts profile obfuscation
- Ineffective profile obfuscation leads to query distinguishability

Challenges (1)

- How to take into account prior background information available to the adversary?
- How to take into account related visible actions, such as links that have been clicked after the search results have been returned to the user?
- Include “sensitive” queries, yes or no?
  - If no: any query that is “sensitive” is known to be real: usefulness of the tool? self-censorship?
  - If yes: what if they are considered as real by the search engine?

Challenges (2)

- Tool (un)observability
  - if observable: search engine may deny service to users who implement the tool
  - if unobservable: attacks that detect the tool?

- Can we "predict" the strategy of the adversary?
  - Example: inference algorithms (SCA) to create profiles may be secret!
  - Consequences if our assumptions are wrong?
    - Adversary uses a different inference (profiling) algorithm than the one assumed by the tool
      - obfuscation might not achieve its goal – what does it do then?
    - Adversary behaves as naïve instead of strategic
      - takes the fake information as true (!)
      - there is a cost in being strategic

What can we achieve??

(Assuming the technology would work perfectly)

- **Query deniability**, when confronted with a query, eg, by law enforcement: "it wasn't me, it was the tool"

- If **many** people use the tool: overall degradation of the effectiveness of profiling (civil disobedience?)

- If **not many** people use the tool:
  - Sense of satisfaction derived from the search engine having a wrong profile
  - Prevent "manipulations" derived from intimate knowledge of what "makes you tick"
  - **BUT** decisions may still be made on the basis of the noisy profile *as if it was real*
    - does not prevent, eg, discrimination based on profiling, and might even make it worse, if your new profile category provides more disadvantages than your "real" category
    - unknown inference rules

- Something else to keep in mind:
  - even if dummy queries are generated, your real queries are also revealed: at best you can decrease the "weight" of some categories of interest in your profile
    - eg, would not prevent re-identification in the AOL case

Possible ways forward

- Let users indicate the type of profile they would like to present to the search engine and generate the dummy queries accordingly
  - Still requires some knowledge of the profiling algorithms
- Tool for solidarity: generate queries on certain sensitive topics, so that people who *actually* search for those topics get cover from the other tool users
  - requires a critical mass (but possibly smaller than for overall degradation of profile quality)
  - normalize the use of certain “taboo” queries
- Combination of obfuscation and anonymity?