Preserving the User's Privacy in Web Search Engines

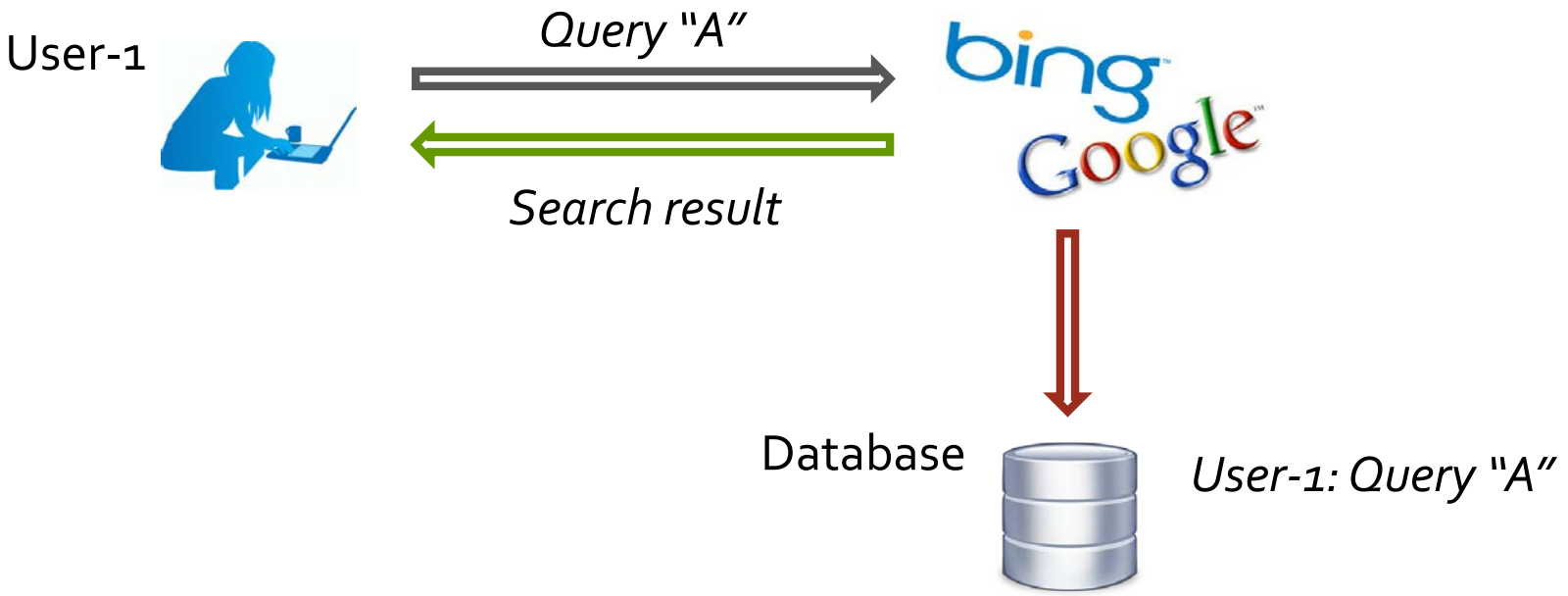
Alexandre Viejo

Universitat Rovira i Virgili

Contents

Introduction

Web search engines (WSEs) answer user queries but they also generate and store query logs.



Introduction

Example: AOL Query Log; 20 million web search queries (500k users) (http://www.cim.mcgill.ca/~dudek/206/Logs/AOL-user-ct-collection/)

FicheroAOL_user-ct-test-collection-01.txt

```
AnonID  Query   QueryTime   ItemRank    ClickURL
142 rentdirect.com  2006-03-01 07:17:12
142 www.prescriptionfortime.com 2006-03-12 12:31:06
142 staple.com  2006-03-17 21:19:29
142 staple.com  2006-03-17 21:19:45
142 www.newyorklawyersite.com   2006-03-18 08:02:58
142 www.newyorklawyersite.com   2006-03-18 08:03:09
142 westchester.gov 2006-03-20 03:55:57 1   http://www.westchestergov.com
142 space.comhttp   2006-03-24 20:51:24
142 dfdf    2006-03-24 22:23:07
142 dfdf    2006-03-24 22:23:14
142 vaniqa.comh 2006-03-25 23:27:12
142 www.collegeucla.edu 2006-04-03 21:12:14
142 www.elaorg  2006-04-03 21:25:20
142 207 ad2d 530    2006-04-08 01:31:04
142 207 ad2d 530    2006-04-08 01:31:14 1   http://www.courts.state.ny.us
142 broadway.vera.org   2006-04-08 08:38:23
142 broadway.vera.org   2006-04-08 08:38:31
142 vera.org    2006-04-08 08:38:42 1   http://www.vera.org
```

Introduction

Example: AOL Query Log; obtaining interests and other data.

```
12482826    wicomico civic center   2006-03-18 11:23:29 1   http://www.wicomicociviccenter.org
12482826    effects of extasy   2006-03-24 12:06:13 1   http://www.chillpharm.com
12482826    get your ex eating out of the palm of your hand let him see what he's missing   2006-03-24 14:07:43
12482826    improve your looks for your ex  2006-03-24 14:10:43 4   http://www.naturalhealthweb.com
12482826    improve your looks for your exboyfriend 2006-03-24 14:19:43
12482826    improve your looks for your exboyfriend 2006-03-24 14:19:45 8   http://www.mynippon.com
12482826    show your exboyfriend what he's missing by looking hotter   2006-03-24 14:22:53
12482826    tips on how to look sexy and get your exboyfriend back  2006-03-24 14:35:30 6   http://www.bosshair.com
12482826    new hott hairdos    2006-03-24 14:45:22
12482826    tips on spiral curling and other cute hairstyles    2006-03-24 14:46:07
12482826    juelz santana   2006-04-07 12:14:51 1   http://www.santanastown.com
12482826    juelz santana clothes   2006-04-07 12:22:43 4   http://shopping.yahoo.com
12482826    crazy lyrics by kc&jojo 2006-05-10 10:54:22 1   http://www.geocities.com
```

13 queries from user **12482826**, what can we get from this stuff?

Introduction

Example: AOL Query Log; obtaining interests and other data.

User 12482826 is interested/related to (*knowledge base: the Web*):

Wicomico Youth and Civic Center

Arena

Directions

The Wicomico Youth and Civic Center is a multipurpose arena located in Salisbury, Maryland, USA. It contains 28,000 square feet of space and can seat 2,500 for banquets, 1,600 for theater concerts and ... Wikipedia

Address: 500 Glen Ave, Salisbury, MD 21804, United States

Capacity: 5,130

Phone: +1 410-548-4900

K-Ci & JoJo

Musical Group

K-Ci & JoJo are an American R&B duo, consisting of brothers Cedric "K-Ci" Hailey and Joel "JoJo" Hailey, Natives of Monroe, North Carolina, they are also members of the chart-topping R&B group Jodeci with the DeGrate brothers—Donald and Dalvin. Wikipedia

Juelz Santana

Rapper

LaRon Louis James, better known by his stage name Juelz Santana, is an American rapper and actor. He hails from Harlem, New York City and is member of East Coast hip hop group The Diplomats. Wikipedia

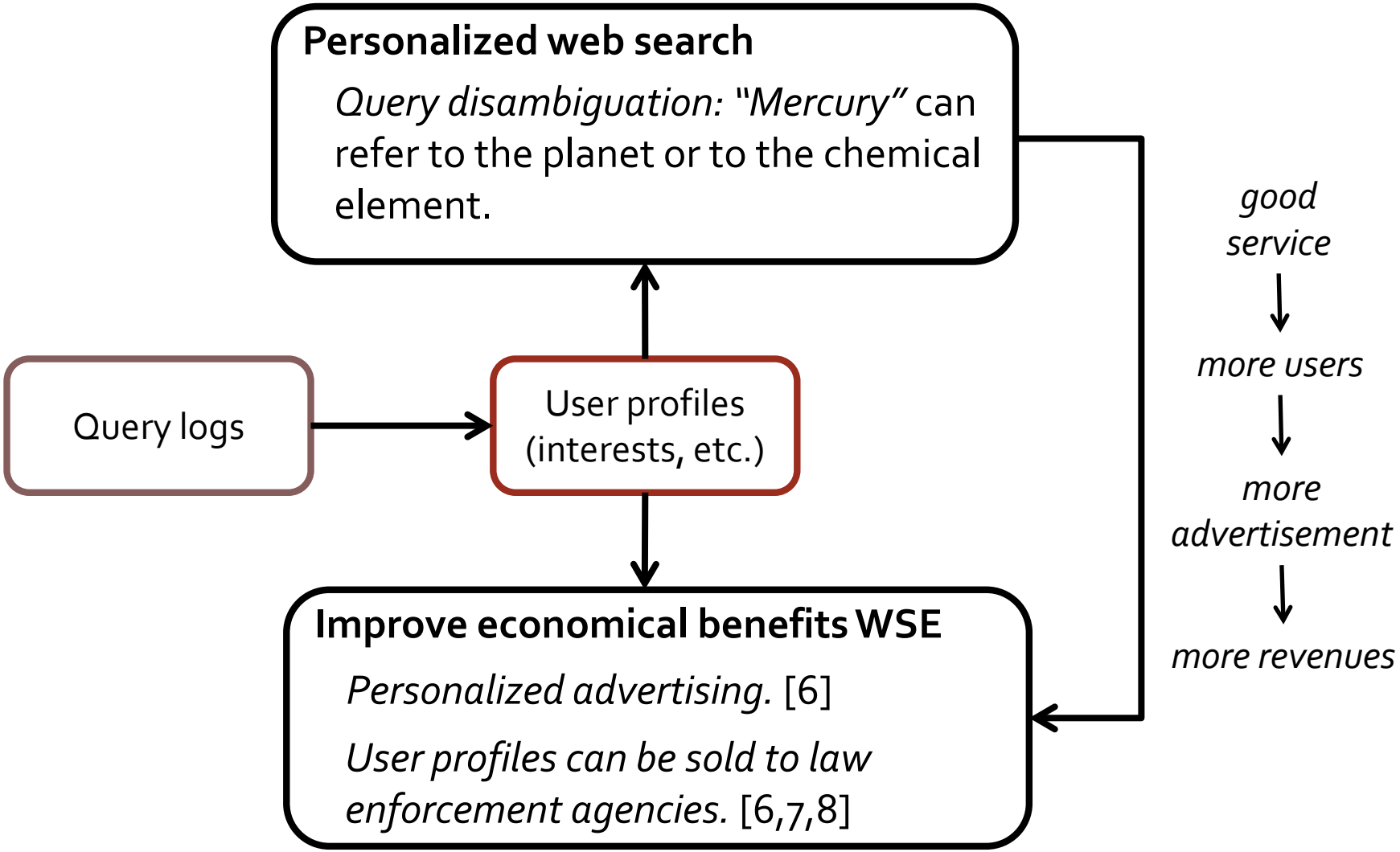
MDMA

Drug

MDMA is an empathogenic drug of the phenethylamine and amphetamine classes of drugs. MDMA has become widely known as "ecstasy", usually referring to its street form, although this term may also include the presence of possible adulterants. Wikipedia

also: “cute hairstyles” and “looking hotter to get her boyfriend back”.

Introduction



Introduction

Building user profiles requires identifying the users. How does it happen?

Gathering *pseudo-identifiers*: [11]

- IP addresses.
- Browser cookies.
- Browser search bars (e.g. Google Toolbar).
- Browser version and configuration data (device fingerprinting):
  P. Eckersley, "How unique is your web browser?" [10]

Introduction

Next problem: Can pseudo-identifiers reveal the real identity of users?

An Internet Service Provider (ISP) can connect the IP address linked to a bunch of queries with the complete name of the user who submitted them.

A user who logs in to an account associated to a WSE and submits queries, enables the WSE to link these queries with that account.

A user can submit a query about personal information which identifies her uniquely: her name, national ID, etc. (vanity search)

A single query might not reveal the real identity of a certain user but the aggregation of several queries might do it → Thelma Arnold case [14].

Introduction

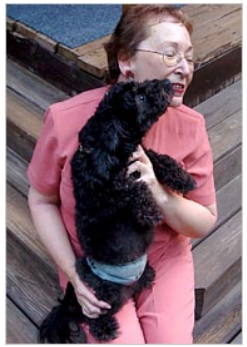
Thelma Arnold case [14]:

Among a list of 20 million Web search queries collected by AOL and **released on the Internet** is user **No. 4417749** (number assigned by AOL to protect the user's anonymity).

4417749 conducted hundreds of searches over a three-month period on topics ranging from:

"60 single men"
"dog that urinates on everything"
"landscapers in Lilburn, Georgia"
several people with the last name "Arnold"

The data trail pointed to **Thelma Arnold**, a 62-year-old widow who lives in Lilburn, Ga.



Introduction

WSEs know our identity, our interests, etc. What is the big deal?

User profiles may contain sensitive information like diseases, sexual tendencies, economical status, etc –> **privacy threat**.

This information is stored in a database far from our control. Is it safe enough?

In the AOL scandal [14], 20 million queries made by thousands of users were **publicly disclosed for research purposes**.

Users profiles can also be stolen by hackers.

Users profiles can also be disclosed by error.

Or they can be directly sold by the WSEs.

Conclusion: there is room for privacy-preserving schemes that enable the privacy-aware users to work with WSEs.

Current proposals: two approaches

1) Conceal the real identity of the user in front of the WSE.

Using a dynamic IP and a plain web browser without cookies is a simple example of this.

Other methods include the use of **anonymizing proxies** (e.g. Tor [15]) **+ HTML header filters** (e.g. Privoxy).

This approach pursuits **total user anonymity** –> queries cannot be linked to the users and the WSEs cannot build profiles –> good for the **privacy**, bad for the **usefulness** (WSEs cannot provide personalized web search).

2) Distort the user profile by submitting fake queries to the WSE.

This is based on submitting fake queries to the WSE together with legitimate ones –> the user profile will contain a **mix** of real and fake interests –> real sensitive data cannot be unequivocally identified.

This approach enables us to build profiles with a **trade-off** between privacy and usefulness.

Distorting user profiles

There are two main categories: *multi-party* and *single-party*

1) Multi-party protocols (or p2p protocols).

Require **external entities** (e.g., human users, central servers, etc.).Users submit queries generated by other users.

Good: Fake queries are real queries generated by real users –> difficult to detect.

Bad: **Slow response time** and **availability problems**. It is difficult to control the contents of the fake queries.

2) Single-party protocols (or stand-alone protocols).

Work directly in the computer of the user. Fake queries are synthetic.

Good: Full control over the contents of the fake queries. No response time or availability issues.

Bad: Fake queries can be detected as "computer-generated".

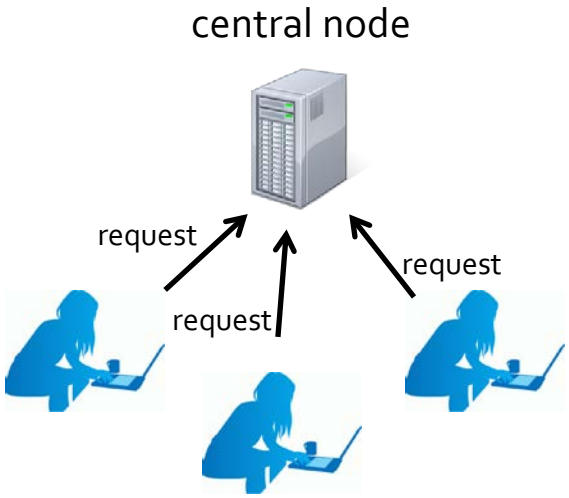
Multi-party protocols

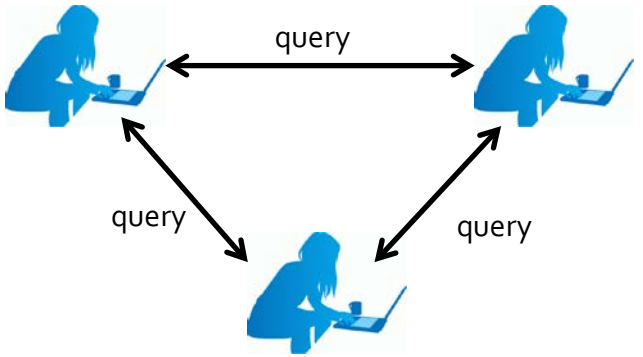
1) Jordi Castellà-Roca, Alexandre Viejo, Jordi Herrera-Joancomartí, "Preserving User's Privacy in Web Search Engines", *Computer Communications*, vol. 32, no. 13-14, pp. 1541-1551, 2009.

Protocol Overview:

Step-1: Users willing to submit a search query get in touch via a central server.

Step-2: Users form a p2p group of "n" users. All queries are shuffled and distributed (we use: ElGamal encryption, ElGamal re-masking, permutation).



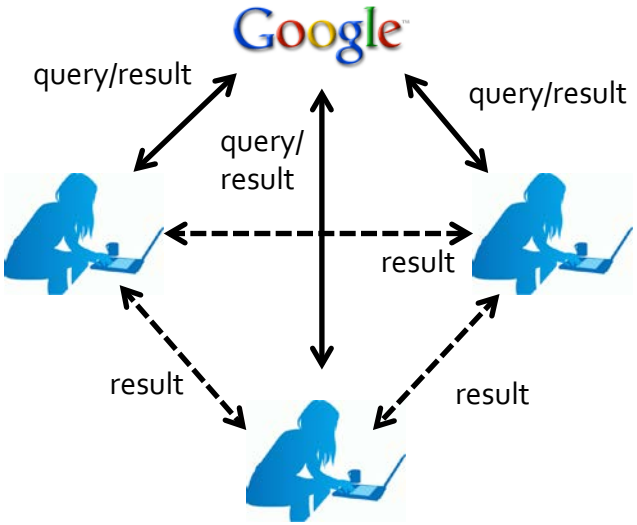


Multi-party protocols

1) Jordi Castellà-Roca, Alexandre Viejo, Jordi Herrera-Joancomartí, "Preserving User's Privacy in Web Search Engines", *Computer Communications*, vol. 32, no. 13-14, pp. 1541-1551, 2009.

Protocol Overview:

Step-3: Users submit their assigned query, get the answer and distribute it to the whole group. Each user gets the answer to her specific query (and n-1 additional answers which are discarded).



Multi-party protocols

1) Jordi Castellà-Roca, Alexandre Viejo, Jordi Herrera-Joancomartí, "Preserving User's Privacy in Web Search Engines", *Computer Communications*, vol. 32, no. 13-14, pp. 1541-1551, 2009.

Discussion:

The central node is a clear bottle-neck and building a group for each query is time consuming (but **dynamic groups** are good for the privacy!!).

Real tests with groups of 3 users give a response time of **5,2 sec**. **3,2 sec.** in:

C. Romero-Tris, A. Viejo, J. Castellà-Roca, "Improving query delay in private web search", *Int. Workshop on Securing Information in Distributed Environments and Ubiquitous Systems (SIDEUS'11), 2011*

Only considers **semi-honest** adversaries –> everyone follows the specified protocol. The following two papers address this point:

Y, Lindell, E. Waisbard, "Private web search with malicious adversaries", Proceedings of the 10th international conference on Privacy enhancing technologies (PETS'10), pp. 220–235, 2010.

C. Romero-Tris, J.Castellà-Roca, A. Viejo, "Multi-party private web search with untrusted partners", *7th Int. Conference on Security and Privacy in Communication Networks (SecureComm'11), 2011.*

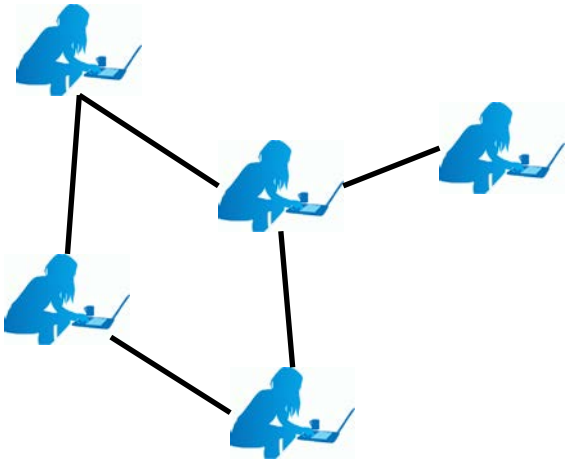
This proposal generates a distorted profile that will contain **random fake interests** (only 1/n will correspond to legitimate interests). The **resulting profile will be very useless.**

Multi-party protocols

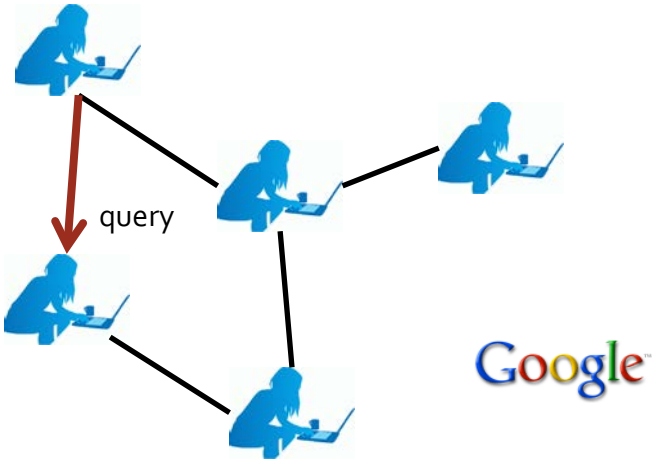
2) Alexandre Viejo, Jordi Castellà-Roca, "Using Social Networks to Distort Users' Profiles Generated by Web Search Engines", *Computer Networks, vol. 54, no. 9, pp. 1343-1357, 2010.*

Overview:

Step-0: Users are *logically* and **permanently** connected using an already deployed **social network**.



Step-1: A user willing to submit a query can submit it directly to google or forward it to one of her friends in the social network (a heuristic is defined for this purpose).

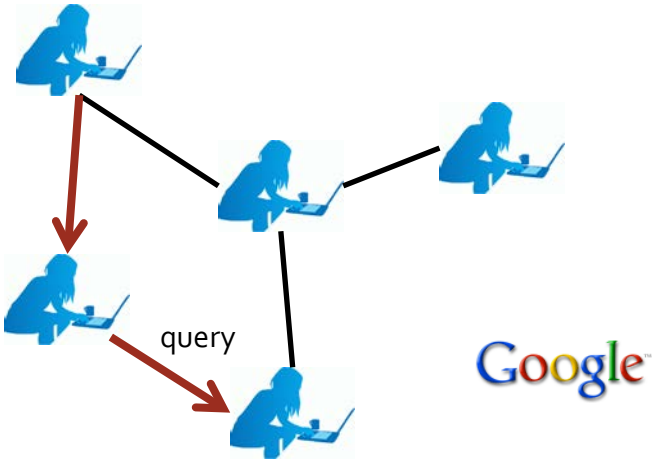


Multi-party protocols

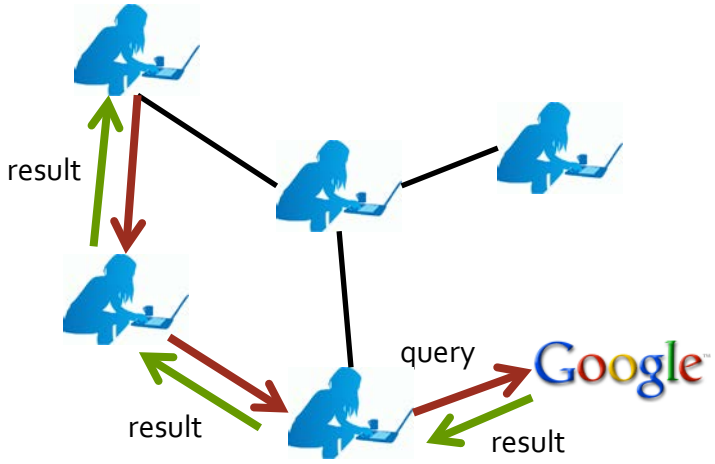
2) Alexandre Viejo, Jordi Castellà-Roca, "Using Social Networks to Distort Users' Profiles Generated by Web Search Engines", *Computer Networks, vol. 54, no. 9, pp. 1343-1357, 2010.*

Overview:

Step-2: the friend in turn can submit the query to Google or forward it again to another friend (a heuristic decides that).



Step-3: there is a user that submits the query to the WSE. The answer is forwarded following the reverse path.



Multi-party protocols

2) Alexandre Viejo, Jordi Castellà-Roca, "Using Social Networks to Distort Users' Profiles Generated by Web Search Engines", *Computer Networks, vol. 54, no. 9, pp. 1343-1357, 2010.*

Discussion:

It is a variation of "Crowds" .

Reiter, M., Rubin, A., 1998. Crowds: anonymity for web transactions. ACM Transactions on Information and System Security 1 (1), 66–92.

Benefits: i) use of already deployed social networks; ii) considers that to conceal a user her queries have to be uniformly distributed among the rest of the users [28].

The source of the query is not known, each user only knows the predecessor and the successor in the path → Privacy.

Weak against the **predecessor attack** [29](like Crowds).

The use of already established groups **reduces the response time**.

Use of **groups of friends → better profile usefulness** because friends **are expected to share similar interests** (proof is not provided).

Moving to single-party protocols

General problems of multi-party protocols:

Response time in seconds (**3,2 seconds** in the best situation).

Direct query to Google: **300 ms**.

We depend on others, will they be available? will they collaborate?

We submit queries of other users to the WSE. Are we **comfortable** with that? are the contents of these queries **useful** to generate our desired distorted profile?

Single-party schemes seem more promising:

They are suited to provide fast response times.

They do not suffer availability problems.

They have full control over the contents of the fake queries → they can control the level of detail of the profile build by the WSE –> we can improve profile usefulness.

Problem: synthetic queries are "detectable" –> **we have to work on that**.

Single-party protocols

1) Howe, D., & Nissenbaum, H. (2009). Trackmenot: Resisting surveillance in web search. *Lessons from the Identity Trail: Anonymity, Privacy, and Identity in a Networked Society, 23 , 417-436.*

Overview and Discussion:

TrackMeNot is a Firefox plugin that **periodically issues randomized search-queries** to the WSE. It hides real user queries in a **bunch of real/fake queries**.

Fake queries are **selected in a random way from blog entries or news headlines.**

Example of fake queries (gathered from the log file):

```
[QUERY] engine=google | query='inside microsoft autopilot nadella'
[QUERY] engine=google | query= 'meet genius sochi opening'
[QUERY] engine=google | query='thai election rejected'"
[QUERY] engine=google | query='stripped business'
[QUERY] engine=google | query='snipers'
[QUERY] engine=google | query='sports live updates from winter'
```

This scheme generates a user profile containing **a mix of legitimate and fully random interests** –> random profile –> **useless profile**.

According to [30], an aware WSE can detect fake queries analyzing the grammatical construction and the semantics.

Single-party protocols

2) J. Domingo-Ferrer, A. Solanas, J. Castellà-Roca, "*h(k)-Private Information* Retrieval from Privacy-Uncooperative Queryable Databases", *Journal of Online Information Review, vol. 33, no. 4, pp. 1468-4527,* 2009.

Overview and Discussion:

GooPIR only works with **one-term queries**. A **thesaurus** is used to obtain the fake terms according to their **frequency of appearance.**

This system **submits a unique** query to the WSE that contains fake terms together with the *legitimate ones.* **All terms are permuted**.

Example: legitimate query –> "university" (frequency: 0.0064)

Final query submitted: **number** OR **university** OR **better** OR **man**

Freq.: 0.0062 Freq.: 0.0069 Freq.: 0.0064

Resulting profile contains **a mix of legitimate and fully random interests** –> random profile –> **useless profile**.

An aware WSE can detect fake queries using semantics [30].

Single-party protocols

3) Y. Xu, B. Zhang, Z. Chen, K. Wang, "Privacy-enhancing personalized web search", *Proc. of 16th international conference on World Wide Web*, pp. 591-600, 2007.

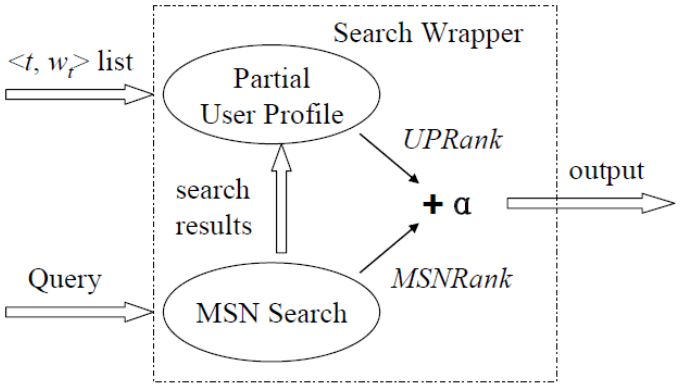
Overview and Discussion:

Scheme that requires **deep changes at the WSE side** but focuses on the trade-off between privacy and profile usefulness.

Users can **choose the content and degree of detail** of the profile information which is exposed to the WSE.

User submits **her query and a partial user profile managed by herself,** the WSE personalizes the results using this information.

Module **Search Wrapper** at the WSE side:

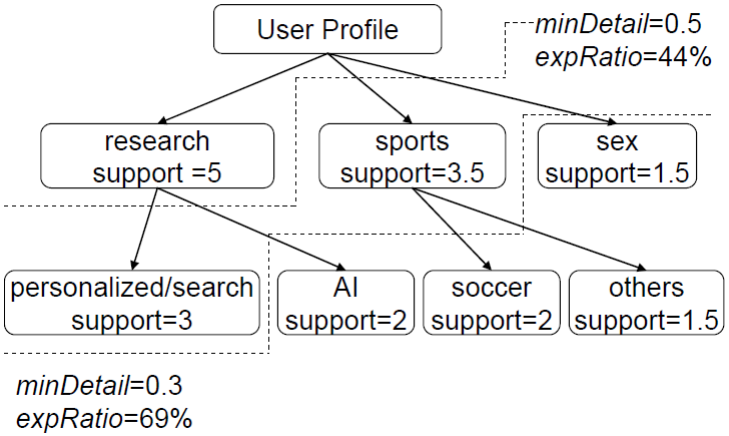


Single-party protocols

3) Y. Xu, B. Zhang, Z. Chen, K. Wang, "Privacy-enhancing personalized web search", *Proc. of 16th international conference on World Wide Web*, pp. 591-600, 2007.

Overview and Discussion:

The partial user profile submitted by the user is organized as a tree. The user selects the level of detail of the information shown to the WSE:



Authors assume that the WSE does not store any additional information from the user –> **semi-honest** WSE–> it follows the specified protocol.

Deep changes at the WSE + Semi-honest WSE –> quite unrealistic.

Single-party protocols

4) A. Viejo, J. Castellà-Roca, O. Bernadó and J. M. Mateo-Sanz (2012). "Single-Party Private Web Search", In *Proc. of the 10th annual conference on privacy, security and trust (PST'12).*

Overview and Discussion:

Scheme that enables the users to decide the **semantic distance** between legitimate and fake interests –> **trade-off between privacy and useful profile**.

Does not require changes at the server side.

The system generates *m* fake queries which are submitted at the same time or with a certain delay together with the authentic one.

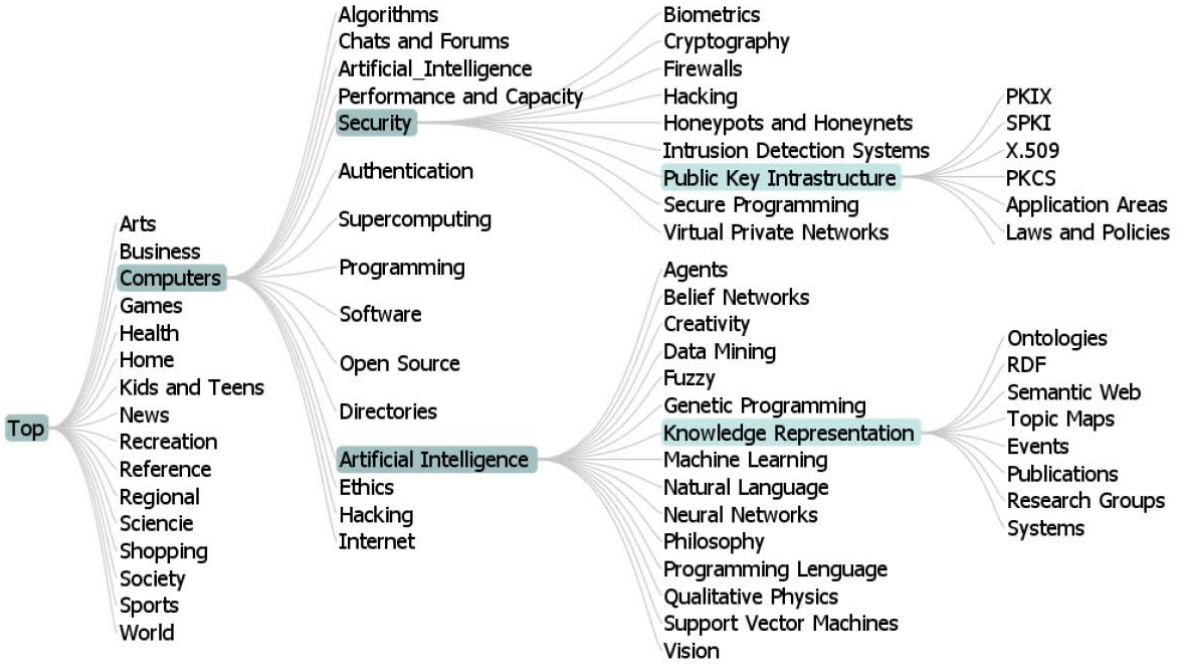
Fake queries are generated using **ODP** (Open Directory Project) [26]. This is a **knowledge base** that allows us to **semantically interpret** the original interests and control the **distance between fake interests and authentic ones**.

Single-party protocols

4) A. Viejo, J. Castellà-Roca, O. Bernadó and J. M. Mateo-Sanz (2012). "Single-Party Private Web Search", In *Proc. of the 10th annual conference on privacy, security and trust (PST'12).*

Overview and Discussion:

ODP is a knowledge-base constructed and maintained by volunteer editors. Its purpose is to list and categorize web sites. Manually **created categories are classified following a tree structure** and associated with related web resources.

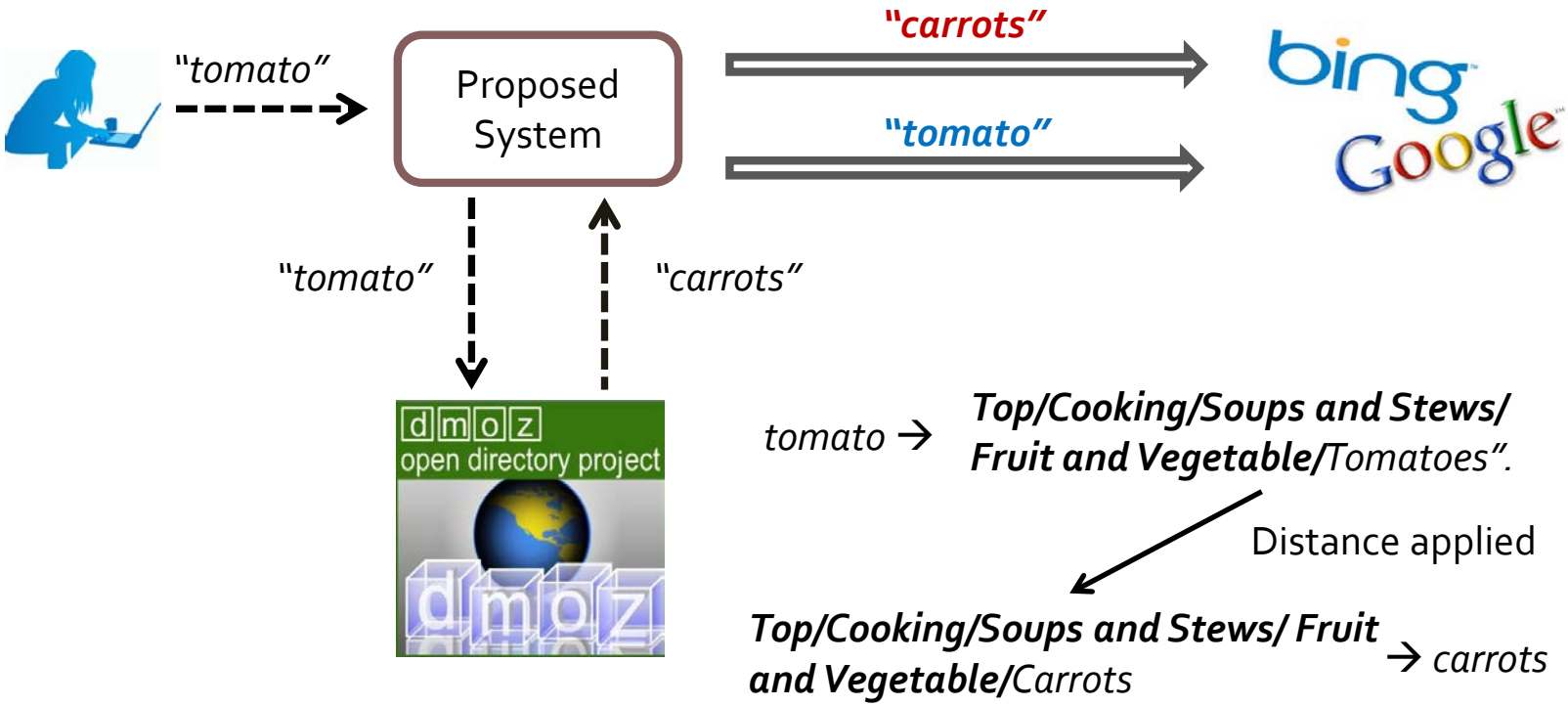


Single-party protocols

4) A. Viejo, J. Castellà-Roca, O. Bernadó and J. M. Mateo-Sanz (2012). "Single-Party Private Web Search", In *Proc. of the 10th annual conference on privacy, security and trust (PST'12).*

Overview and Discussion:

Workflow example assuming m=1 (one fake query for each legitimate one):

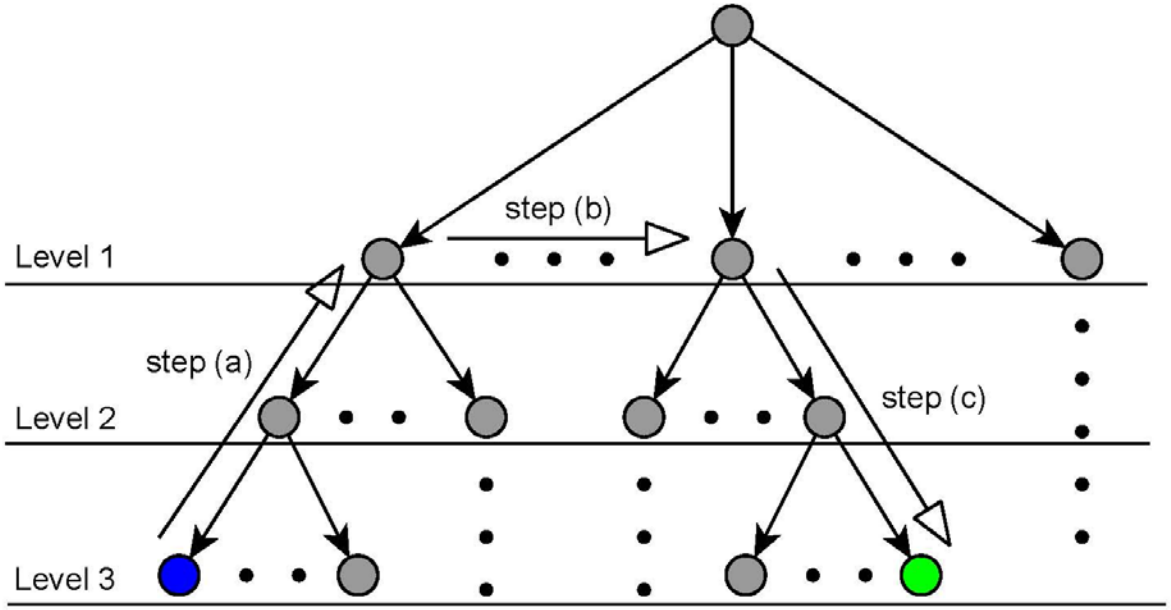


Single-party protocols

4) A. Viejo, J. Castellà-Roca, O. Bernadó and J. M. Mateo-Sanz (2012). "Single-Party Private Web Search", In *Proc. of the 10th annual conference on privacy, security and trust (PST'12).*

Overview and Discussion:

How the distance is applied (parameters fixed by the user):



Single-party protocols

4) A. Viejo, J. Castellà-Roca, O. Bernadó and J. M. Mateo-Sanz (2012). "Single-Party Private Web Search", In *Proc. of the 10th annual conference on privacy, security and trust (PST'12).*

Overview and Discussion:

This scheme **does not deal properly with queries containing multiple terms** (we name these ones "complex queries").

We evaluated the AOL query log and we found that around **60% of the legitimate queries** submitted by real users **contained more than one word** –> clear problem that requires attention.

We provide **support for complex queries** in:

Sánchez, D., Castellà-Roca, J., Viejo, A. (2013). Knowledge-based scheme to create privacy-preserving but semantically-related queries for web search engines. *Information Sciences, 218 , 17-30.*

Use of Natural Language Processing (NLP) tools to **detect the Noun-Phrases (NPs)** (assumed to be the units of meaning) of the legitimate queries.

Single-party protocols

4) A. Viejo, J. Castellà-Roca, O. Bernadó and J. M. Mateo-Sanz (2012). "Single-Party Private Web Search", In *Proc. of the 10th annual conference on privacy, security and trust (PST'12).*

Overview and Discussion:

We provide **support for complex queries** in:

Sánchez, D., Castellà-Roca, J., Viejo, A. (2013). Knowledge-based scheme to create privacy-preserving but semantically-related queries for web search engines. *Information Sciences, 218 , 17-30.*

The notion of **Information Content (IC)** is used **select the NP** (from all the extracted NPs) **that provides the highest quantity of information.**

Note that, the IC of a term *t* is computed as the inverse of the probability of finding *t* in a certain knowledge base (WordNet [31] and ODP in our case).

A **generic** term like "Europe" has a **low IC**.
A **specific** term like "Tom Cruise" has a **high IC**.

The main interest of the whole complex query is assumed to be the interest of the most informative NP –> the rest of the protocol works in a similar way to (4).

Single-party protocols

4) A. Viejo, J. Castellà-Roca, O. Bernadó and J. M. Mateo-Sanz (2012). "Single-Party Private Web Search", In *Proc. of the 10th annual conference on privacy, security and trust (PST'12).*

Overview and Discussion:

(4) And other related works assume that **past queries effectively reflect the real interests of the user** and can be used to **create utility-preserving fake queries**.

This assumption does not always hold:

1. Circumstantially, users may submit several queries related to a certain topic which is quite far from their real interests.

2. Users can submit quite inaccurate queries to the WSE expecting better suggestions –> inaccuracy adds bias.

3. If 2 users share the same web search system the past queries will reflect the aggregated interests of both of them.

4. Queries have no fixed structure –> it is difficult to extract accurate interests.

Other sources of data to obtain the interests of the user should be investigated.

Single-party protocols

5) A. Viejo, D. Sánchez, (2013). Providing useful and private web search by means of social network profiling. In *Proc. of the 11th annual conference on privacy, security and trust (PST'13).*

Overview and Discussion:

Legitimate interests are extracted from their **social network account** (e.g., Twitter) instead of past searches.

Assumption: Micro-interests (detailed interests) are very useful to provide personalized web search but they are dangerous from the privacy point of view. **Macro-interests** (general interests) are useful to provide personalized web search and they do not disclose enough sensitive information.

Fake queries: 1) Hide real **micro-interests** among fake but realistic interests (privacy). 2) Force the WSE to build a user profile with the real **macro-interests** of the user (usefulness).

A profile is a set of categories (science, sports, etc) with a relative weight.

We consider a **local profile** (extracted from Twitter) and a **public profile** (local representation of the user profile that, **we assume**, is being built by the WSE) .

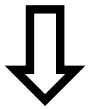
Single-party protocols

5) A. Viejo, D. Sánchez, (2013). Providing useful and private web search by means of social network profiling. In *Proc. of the 11th annual conference on privacy, security and trust (PST'13).*

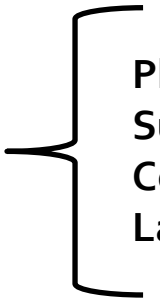
Overview and Discussion:

Step-1: The system builds the local profile the user's social network account.

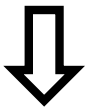
"Planet found orbiting sun-like star"
"Communications satellite falls after launch"



Extract and classify noun-phrases (NLP + ODP)



Planet → Science
Sun-like star → Science
Comm. satellite → Business
Launch → Science



Update the Local Profile

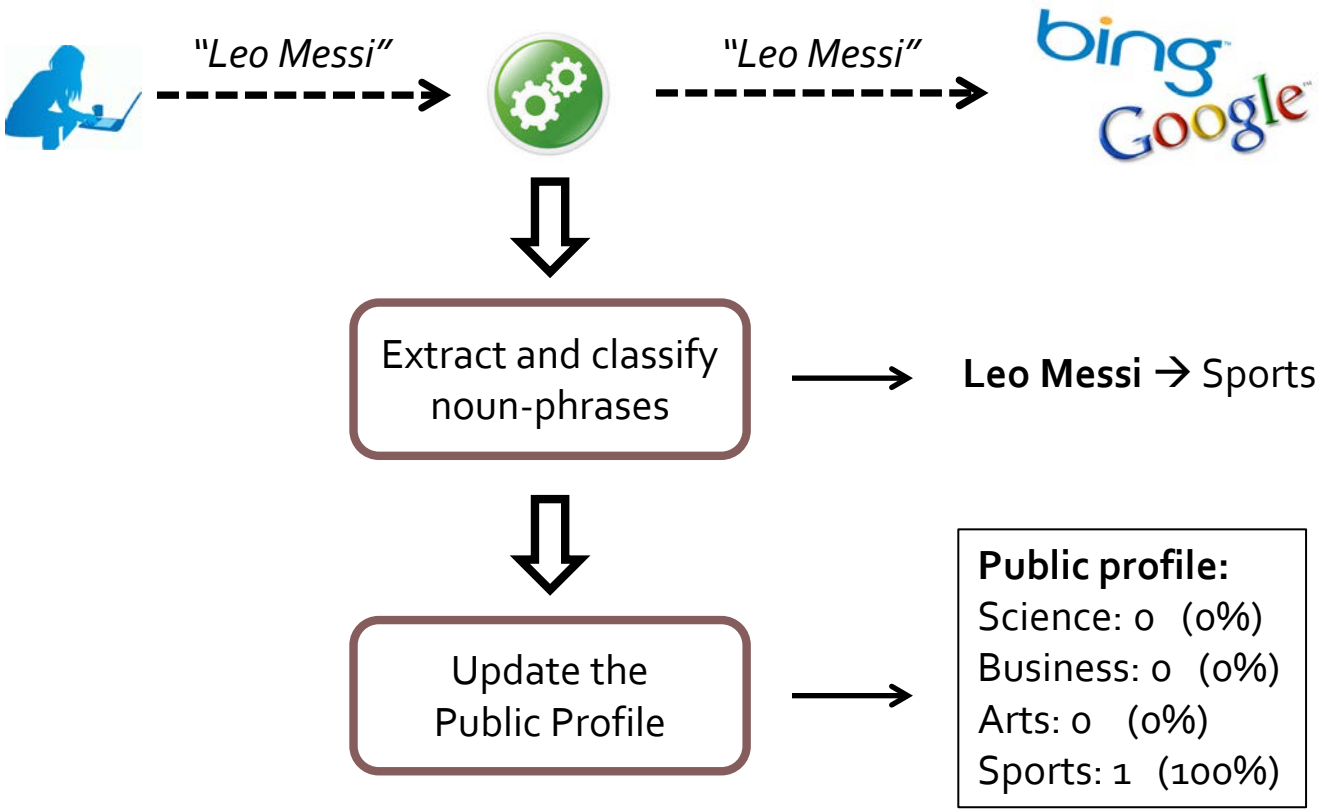
Local profile:
Science: 3 (75%)
Business: 1 (25%)
Arts: 0 (0%)
Sports: 0 (0%)

Single-party protocols

5) A. Viejo, D. Sánchez, (2013). Providing useful and private web search by means of social network profiling. In *Proc. of the 11th annual conference on privacy, security and trust (PST'13).*

Overview and Discussion:

Step-2: User submits a legitimate query.



Single-party protocols

Overview and Discussion:

Step-3: Fake query is generated and submitted.

Select the category of the new fake query →

Local profile:
Science : 75%
Business: 25%
Arts: 0%
Sports: 0%

Public profile:
Science: 0%
Business: 0%
Arts: 0%
Sports: 100%

Difference:
Science: 75%
Business: 25%
Arts: 0%
Sports: -100%

⇩

Obtain the query and submit it →

Science → Climate change

"Climate change" ⇢

bing Google

⇩

Update the Public Profile →

Public profile:
Science: 1 **(50%)**
Business: 0 (0%)
Arts: 0 (0%)
Sports: 1 **(50%)**

Open problems

1) Synthetic fake queries are "detectable" using semantics, grammatical construction, etc –> Find a way to evaluate their "detectability" (i.e, evaluate their quality) and how to improve it.

2) Fake queries are assumed to be submitted in certain periods of time but the specific sending procedure has received little attention. Note that a predictable procedure can be learnt by the WSEs and allow them to detect the fake queries. (can we imitate a human behavior?)

3) It is assumed that the WSE profiles the users following certain categories and a system based on weights. This concept is extracted from the literature on profilers (of social networks or other web 2.0 applications) but there is no proof that a certain WSE profiles its users in that way.

4) A study about the willingness of real users to use all these provided privacy-preserving schemes is strongly required. This would help us to design arguably usable methods.

References

[2] A. Cooper, "A survey of query log privacy-enhancing techniques from a policy perspective", *ACM Transactions on the Web, vol. 2, no. 4. pp.* 1–27, 2008.

[6] S. Hansell, "Increasingly, Internet's Data Trail Leads to Court", *The New York Times, February, 2006.*

[7] N. Summers, "Walking the Cyberbeat", *Newsweek, May, 2009.*

[8] K. Zetter, "Yahoo Issues Takedown Notice for Spying Price List", *Wired,* December, 2009.

[10] P. Eckersley, "How unique is your web browser?", *Proc. of the 10th International Conference on Privacy Enhancing Technologies – PETS'10,* 2010.

[11] S. Ye, F. Wu, R. Pandey, H. Chen, "Noise Injection for Search Privacy Protection", *Technical Report CSE-2008-10 - University of California,* 2008.

[14] M. Barbaro and T. Zeller, "A Face Is Exposed for AOL Searcher No. 4417749", *The New York Times, 2006.*

[15] The Tor Project, 2012. http://www.torproject.org

References

[16] J. Castellà-Roca, A. Viejo, J. Herrera-Joancomartí, "Preserving user's privacy in web search engines", *Computer Communications, vol. 32*, no. 13–14, pp. 1541–1551, 2009.

[18] A. Viejo, J. Castellà-Roca, "Using Social Networks to Distort Users' Profiles Generated by Web Search Engines", *Computer Networks, vol.* 54, no. 9, pp. 1343–1357, 2010.

[20] F. Saint-Jean, A. Johnson, D. Boneh, J. Feigenbaum, "Private Web Search", *Proc. of the ACM workshop on Privacy in electronic society –WPES'07, pp. 84–90, 2007.*

[22] J. Domingo-Ferrer, A. Solanas, J. Castellà-Roca, "*h(k)-Private Information* Retrieval from Privacy-Uncooperative Queryable Databases", *Journal of Online Information Review, vol. 33, no. 4, pp. 1468-4527,* 2009.

[23] TrackMeNot, 2010. http://mrl.nyu.edu/*~dhowe/trackmenot*

[24] Y. Xu, B. Zhang, Z. Chen, K. Wang, "Privacy-enhancing personalized web search", *Proc. of 16th international conference on World Wide Web,* pp. 591-600, 2007.

[26] Open Directory Project, 2012. http://www.dmoz.org

References

[27] Reiter, M., Rubin, A., 1998. Crowds: anonymity for web transactions. ACM Transactions on Information and System Security 1 (1), 66–92.

[28] Sweeney, L., 2002. k-Anonymity: a model for protecting privacy. Int. Journal of Uncertainty Fuzziness Knowledge-based Systems 10 (5), 557–570.

[29] Matthew K. Wright, Micah Adler, And Brian Neil Levine, 2002. The Predecessor Attack: An Analysis of a Threat to Anonymous Communications Systems. *ACM Transactions on Information Systems Security* (ACM Press) **7** (4): 489–522

[30] Balsa, E., Troncoso, C., & Diaz, C. (2012). Ob-pws: Obfuscation-based private web search. In *Proc. of the ieee symposium on security and privacy (sp'12),* 491-505.

[31] Fellbaum, C. (1998). *Wordnet: An electronic lexical database. MIT Press*

Preserving the User's Privacy in Web Search Engines

Alexandre Viejo

Universitat Rovira i Virgili