

# Bloom Cookies: Web Search Personalization without User Tracking

Nitesh Mor, Oriana Riva, Suman Nath, John Kubiawicz

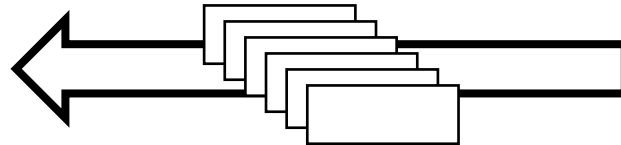
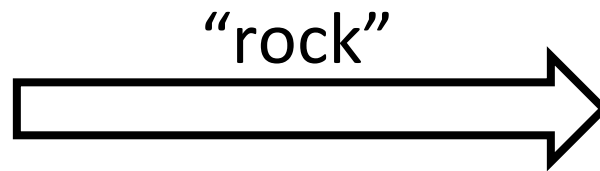


Network and Distributed System Security Symposium, 2015

# Web Search



**Geologist Bob**



search results



**Search Engine**

# Search results

## [Rock music - Wikipedia, the free encyclopedia](#)

[en.wikipedia.org/wiki/Rock\\_music](http://en.wikipedia.org/wiki/Rock_music) ▼

Rock music is a genre of popular music that originated as "rock and roll" in the United States in the 1950s, and developed into a range of different styles ...

[Characteristics](#) · [Origins](#) · [Golden age](#) · [Progression](#) · [Punk era](#) · [Alternative](#)

## [ROCK.COM : MUSIC](#)

[www.rock.com](http://www.rock.com) ▼

rock.com music products ... TOP SELLING PRODUCTS. Pink Floyd Album Covers, Pink Floyd Cards; KISS Logo Embroidered Patch

## [RockAuto Parts Catalog](#)

[www.rockauto.com](http://www.rockauto.com) ▼

RockAuto ships auto parts and body parts to over 3 million manufacturers to customers' doors worldwide, all at warehouse prices. Easy to use parts catalog.

[Parts Catalog](#) · [RockAuto Auto](#) · [Chevrolet](#)

## [Rock | Define Rock at Dictionary.com](#)

[dictionary.reference.com/browse/rock](http://dictionary.reference.com/browse/rock) ▼

noun 1. a large mass of stone forming a hill, cliff, promontory, or the like. 2. Geology. mineral matter of variable composition, consolidated or unconsolidated ...

## [Dwayne Johnson](#)

Actor

Dwayne Douglas Johnson, also known by his ring name The Rock, is an American and Canadian actor, produc...



## [Rock \(geology\) - Wikipedia, the free encyclopedia](#)

[en.wikipedia.org/wiki/Rock\\_\(geology\)](http://en.wikipedia.org/wiki/Rock_(geology)) ▼

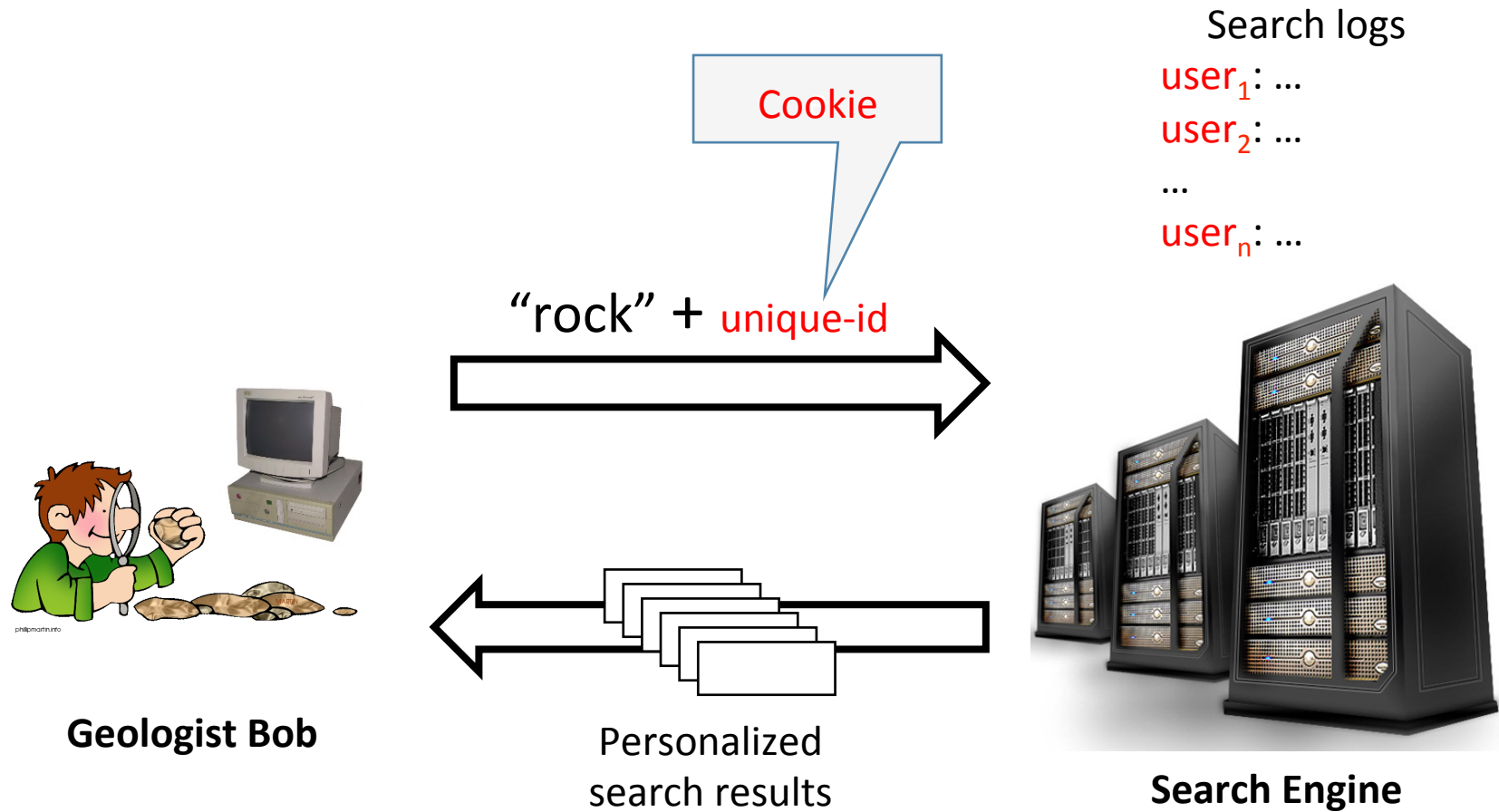
In geology, rock is a naturally occurring solid aggregate of one or more minerals or mineraloids. For example, the common rock granite is a combination of ...

[Classification](#) · [Human use](#)

Ambiguous Query



# State-of-the-art Personalization



Search logs indexed by unique-id identifying individual users.  
User-profiles created by mining search logs.

# Personalized search results

## [Rock \(geology\) - Wikipedia, the free encyclopedia](#)

[en.wikipedia.org/wiki/Rock\\_\(geology\)](http://en.wikipedia.org/wiki/Rock_(geology)) ▼

In geology, **rock** is a naturally occurring solid aggregate of one or more minerals or mineraloids. For example, the common **rock** granite is a combination of ...

[Classification](#) · [Human use](#)

## [Rock music - Wikipedia, the free encyclopedia](#)

[en.wikipedia.org/wiki/Rock\\_music](http://en.wikipedia.org/wiki/Rock_music) ▼

**Rock music** is a genre of popular music that originated as "rock and roll" in the United States in the 1950s, and developed into a range of different styles ...

[Characteristics](#) · [Origins](#) · [Golden age](#) · [Progression](#) · [Punk era](#) · [Alternative](#)

## [ROCK.COM : MUSIC](#)

[www.rock.com](http://www.rock.com) ▼

rock.com music products ... TOP SELLING PRODUCTS. Pink Floyd Album Cover Playing Cards; KISS Logo Embroidered Patch

## [RockAuto Parts Catalog](#)

[www.rockauto.com](http://www.rockauto.com) ▼

**RockAuto** ships **auto** parts and body parts from over 300 manufacturers to customers' doors worldwide, all at warehouse prices. Easy to use parts catalog.

[Parts Catalog](#) · [RockAuto Auto Parts](#) · [Car List](#) · [Chevrolet](#)

## [Rock | Define Rock at Dictionary.com](#)

[dictionary.reference.com/browse/rock](http://dictionary.reference.com/browse/rock) ▼

noun 1. a large mass of stone forming a hill, cliff, promontory, or the like. 2. Geology. mineral matter of variable composition, consolidated or unconsolidated ...

## [Dwayne Johnson](#)

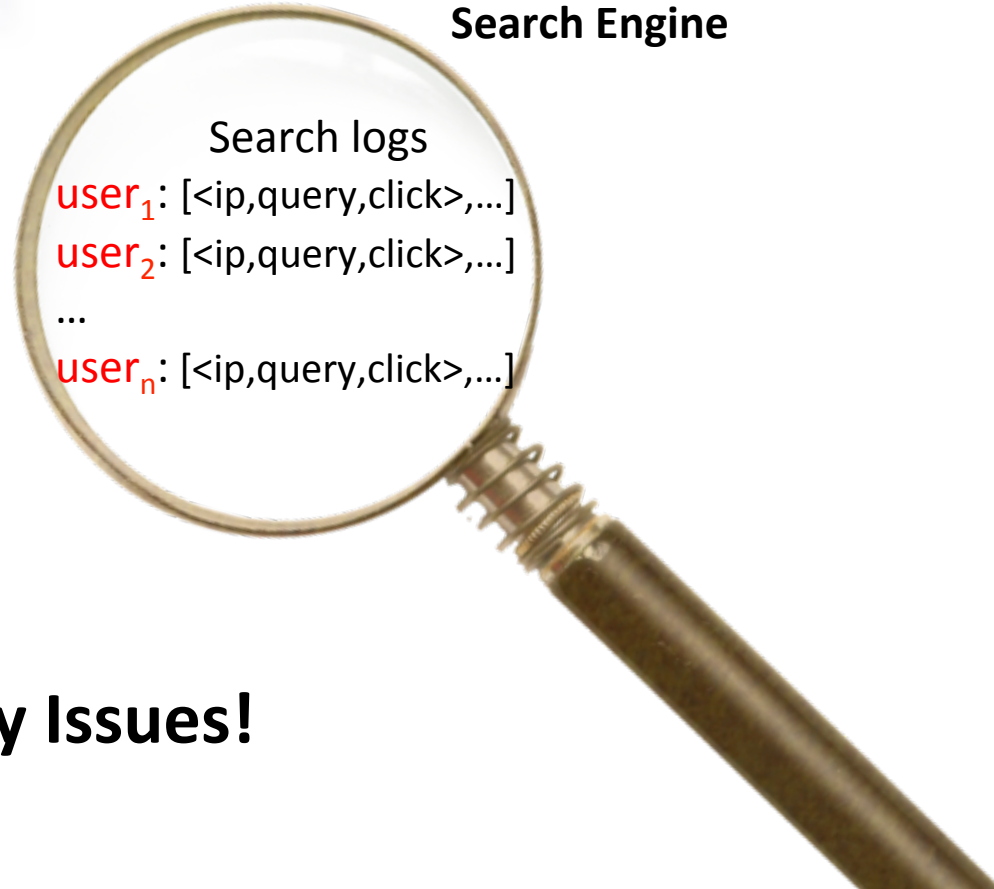
Actor

Dwayne Douglas Johnson, also known by his ring name The Rock, is an American and Canadian actor, produc...





Search Engine



Search logs

**user<sub>1</sub>**: [<ip,query,click>,...]

**user<sub>2</sub>**: [<ip,query,click>,...]

...

**user<sub>n</sub>**: [<ip,query,click>,...]

**Endless Privacy Issues!**

# What could go wrong?

## A Face Is Exposed for AOL Searcher No. 4417749

By MICHAEL BARBARO and TOM ZELLER Jr.  
Published: August 9, 2006

Buried in a list of 20 million Web search queries collected by AOL and recently released on the Internet is user No. 4417749. The number was assigned by the company to protect the searcher's anonymity, but it was not much of a shield.

✉ SIGN IN TO E-MAIL THIS

🖨 PRINT

📄 SINGLE PAGE

📄 REPRINTS

Linking  
queries  
together



Erik S. Lesser for The New York Times

Thelma Arnold's identity was betrayed by AOL records of her Web searches, like ones for her dog, Dudley, who clearly has a problem.

No. 4417749 conducted hundreds of searches over a three-month period on topics ranging from "numb fingers" to "60 single men" to "dog that urinates on everything."

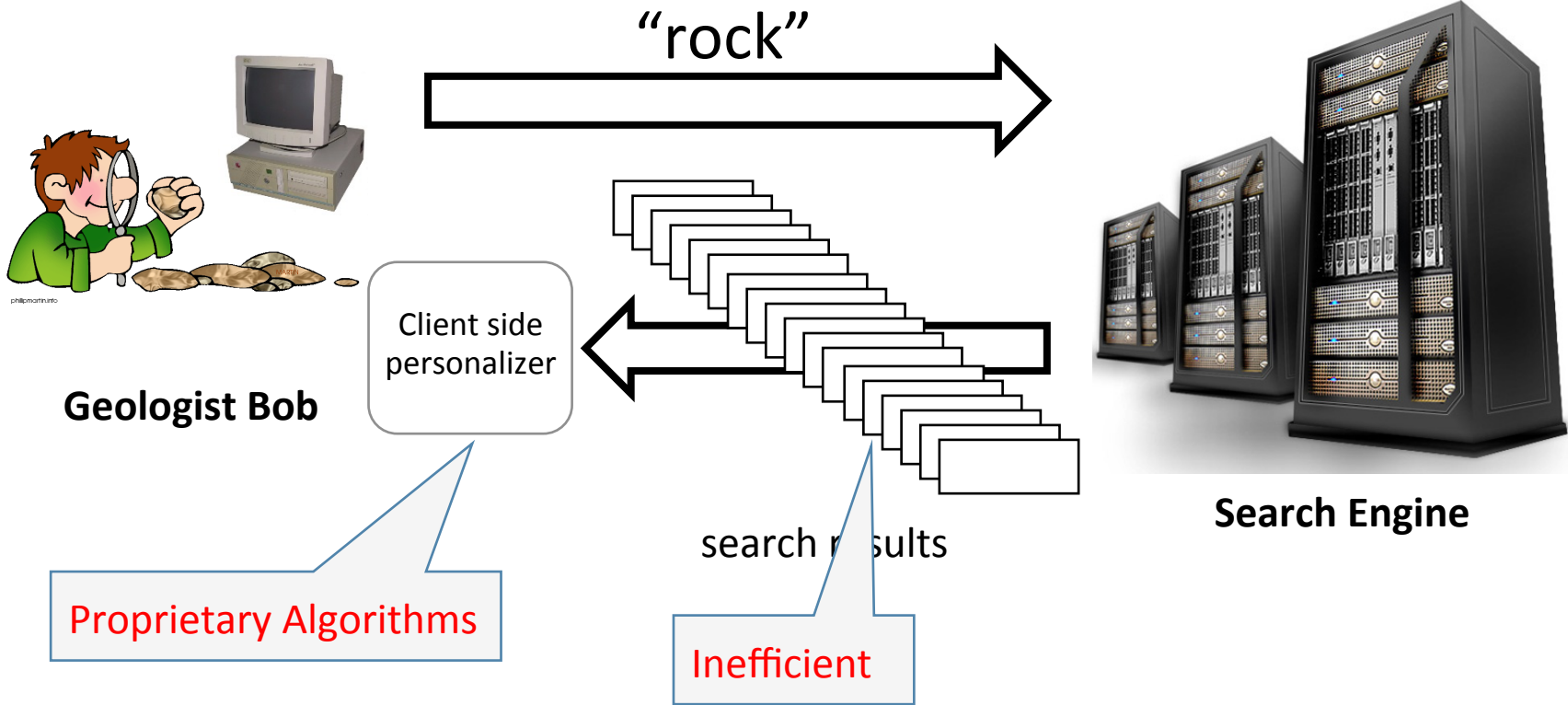
And search by search, click by click, the identity of AOL user No. 4417749 became easier to discern. There are queries for "landscapers in Lilburn, Ga," several people with the last name Arnold and "homes sold in shadow lake subdivision gwinnett county georgia."

It did not take much investigating to follow that data trail to Thelma Arnold, a 62-year-old widow who lives in Lilburn, Ga., frequently researches her friends' medical ailments and loves her three dogs. "Those are my searches," she said, after a reporter read part of the list to her.

<http://www.nytimes.com/2006/08/09/technology/09aol.html>

# Personalization on client side?

Don't send any unique-id to server, maintain user history on client-side.



Search engines don't want proprietary algorithms on client side.  
More search results need to be sent to client.  
Impractical!

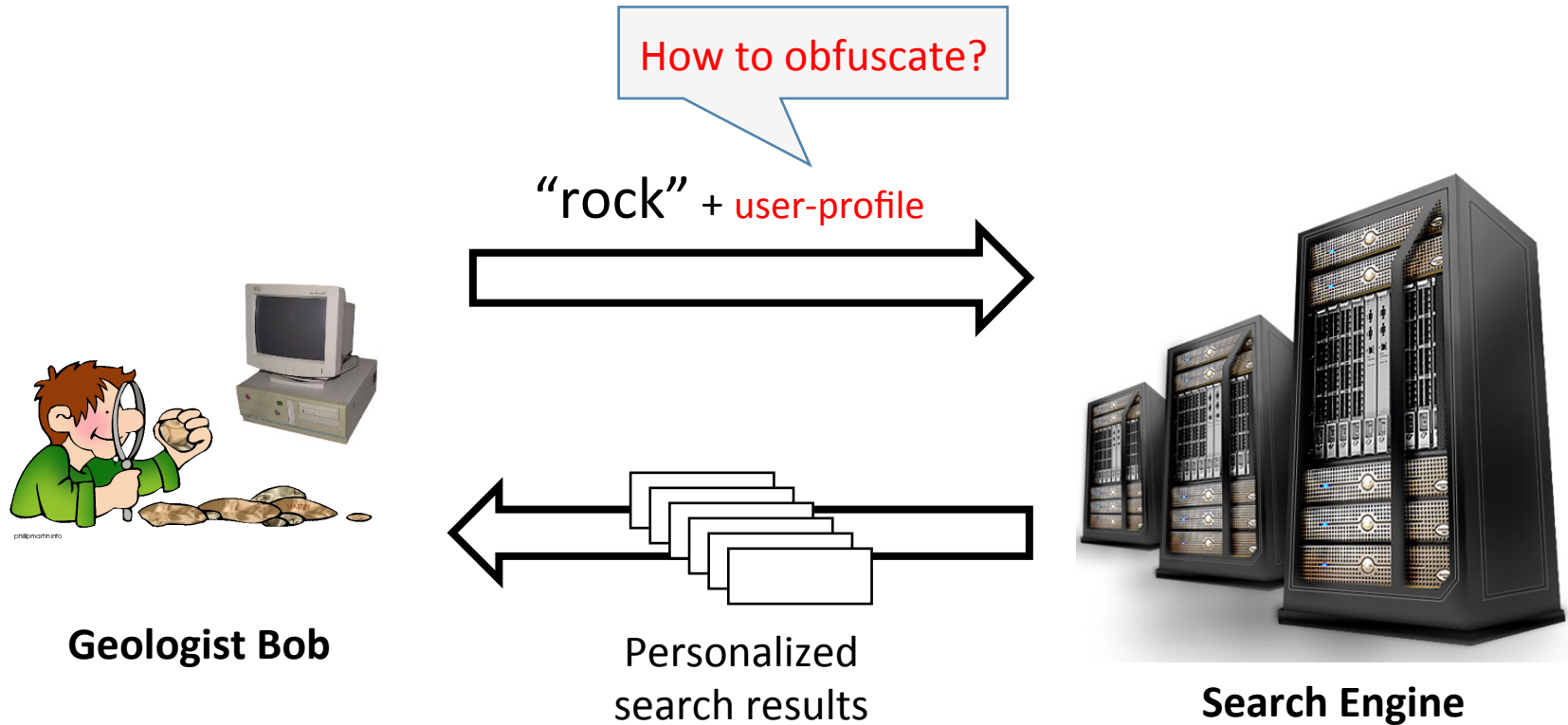


# Send only user profile?

Create *user profile* on client side, instead of server.

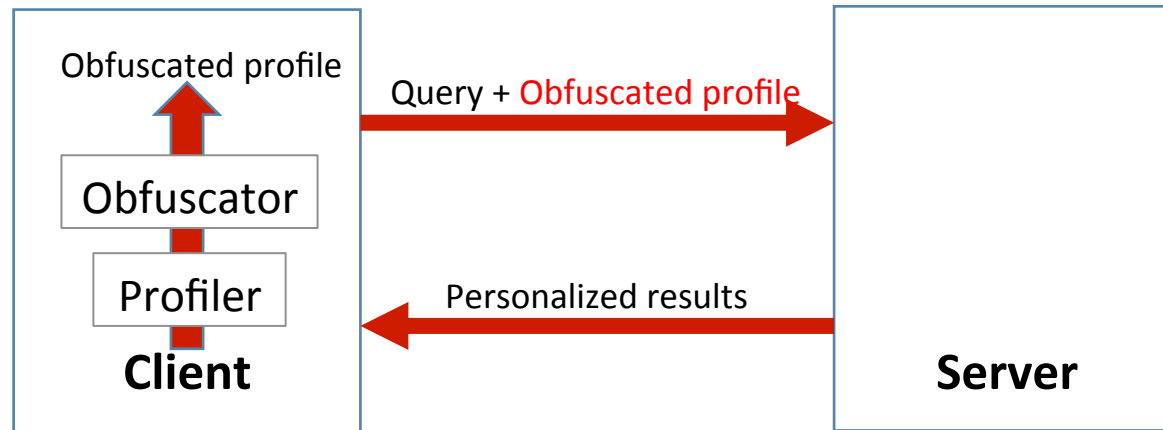
>> ["apple.com", "radioshack.com", "frys.com", "bestbuy.com", "bmw.com", ...]

Send only curated information.



Profile obfuscation required. Otherwise profile acts as a unique-id.

# System Architecture



**Threat Model:** Adversarial server trying to link user profiles.

- A well configured web-browser:
  - No web-cookies
  - No client-side scripts
  - No search toolbars
  - ...
- No attacks based on search keywords.
- Only additional information available to server: obfuscated profile

# Evaluation Metrics

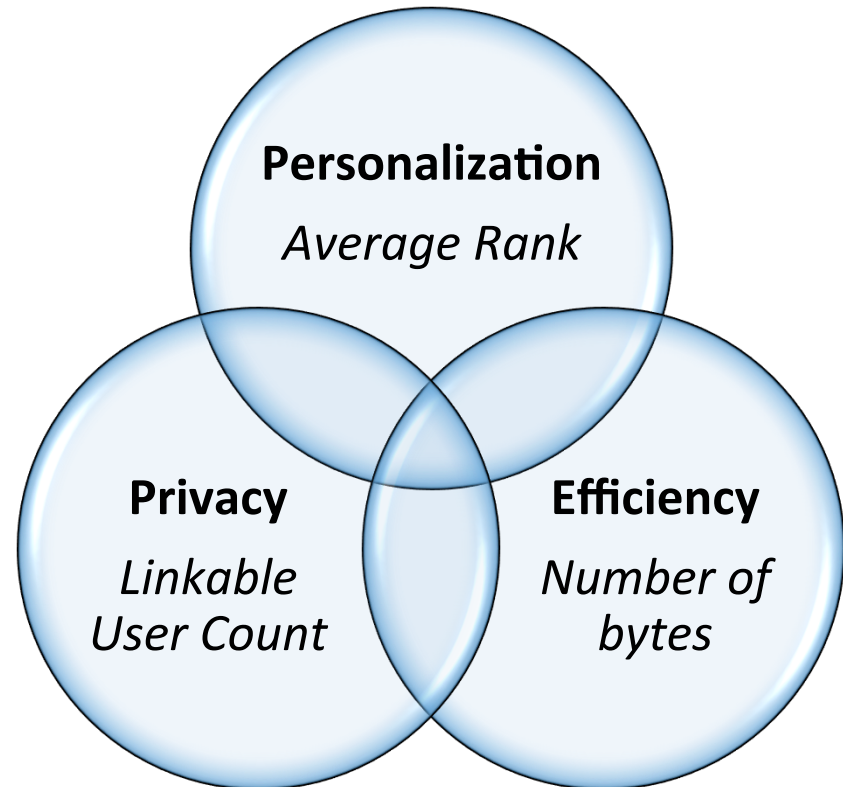
## Personalization:

**Average Rank:** Average position of the URL clicked by a user in the displayed search results.

## Privacy:

**Average User Unlinkability:** Average measure of how non-linkable a user's profile from  $T1$  is to a set of user profiles from  $T2$ .

**Linkable User Count:** Fraction of users that are linked correctly after a simple linking attack by an adversarial server.



## Efficiency:

**Size:** Size of additional data sent to server

Bing search logs: ~2 months, 1300 users, 264615 queries

# Profile Obfuscation

*State-of-the-art techniques:*

1. Profile Generalization
2. Noise Addition

## Profile Obfuscation – Method 1: Generalization

Generalize profile items to coarser granularity, say high-level interests. Send only these high-level user interests to server.

Example:

["apple.com", "radioshack.com", "frys.com", "bestbuy.com", "bmw.com", ...]

becomes

["computers", "electronics", "cars", ...]

	Personalization loss	Linkable users
Generalization	24%	44.1%

Can we do better?

## Profile Obfuscation – Method 2: Noise Addition


Add randomly chosen fake items picked from a *dictionary* to the profile.

Example:

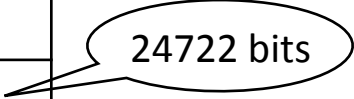
["apple.com", "radioshack.com", "frys.com", "bestbuy.com", "bmw.com", ...]

becomes

["apple.com", "orange.com", "banana.com", "radioshack.com", "ebay.com", "craigslist.com", "frys.com", "fries.com", "bestbuy.com", "goodbuy.com", "wikipedia.org", "facebook.com", "bmw.com", "audi.com", ...]



	Personalization loss	Linkable users
Generalization	24%	44.1%
Noise Addition	1.1%	20.0%



Works okay, but...

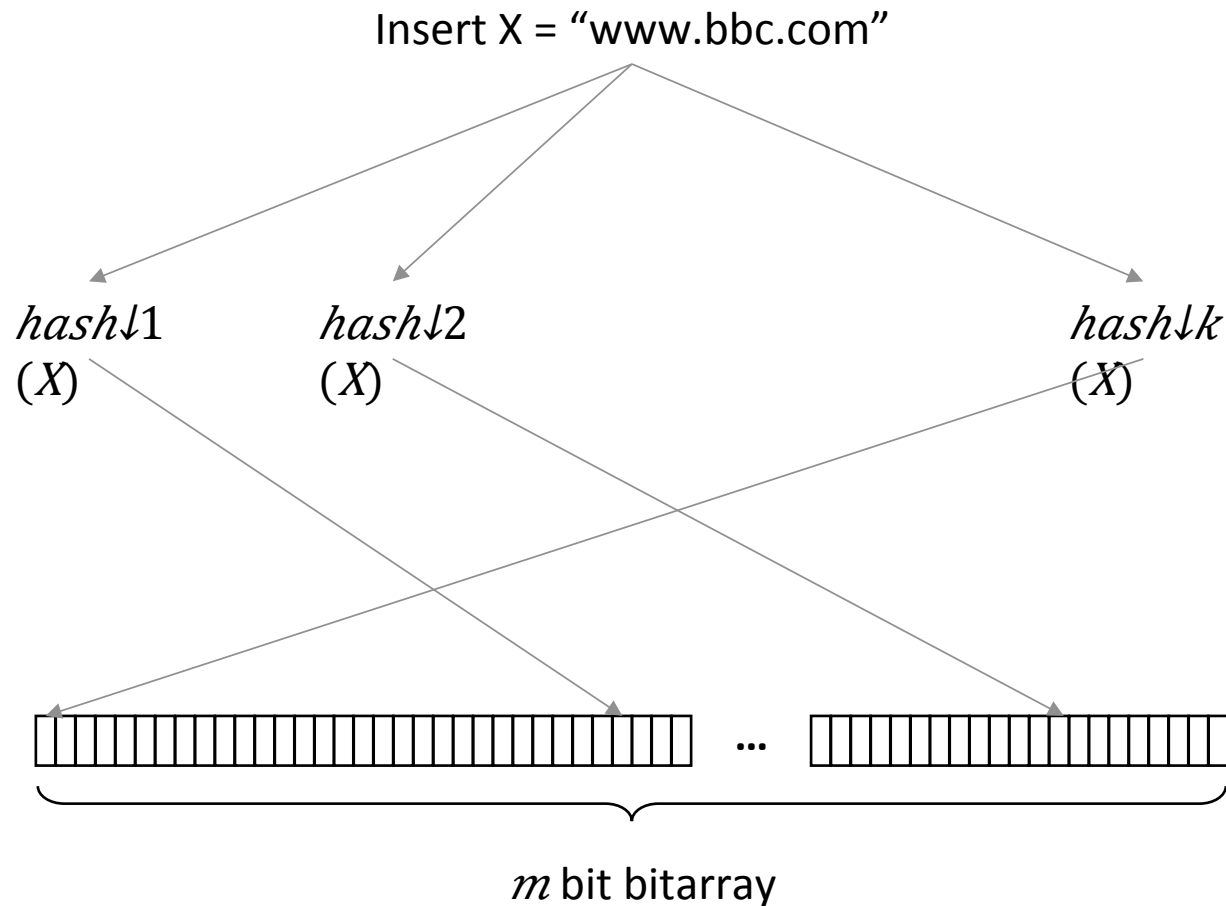
1) Profile size explodes, and 2) needs unbiased dictionary

# **Bloom Cookies**

A compact and privacy-preserving way to encode user-profiles for personalization purposes.

# Bloom filters 101

Space efficient data-structure for set-membership queries

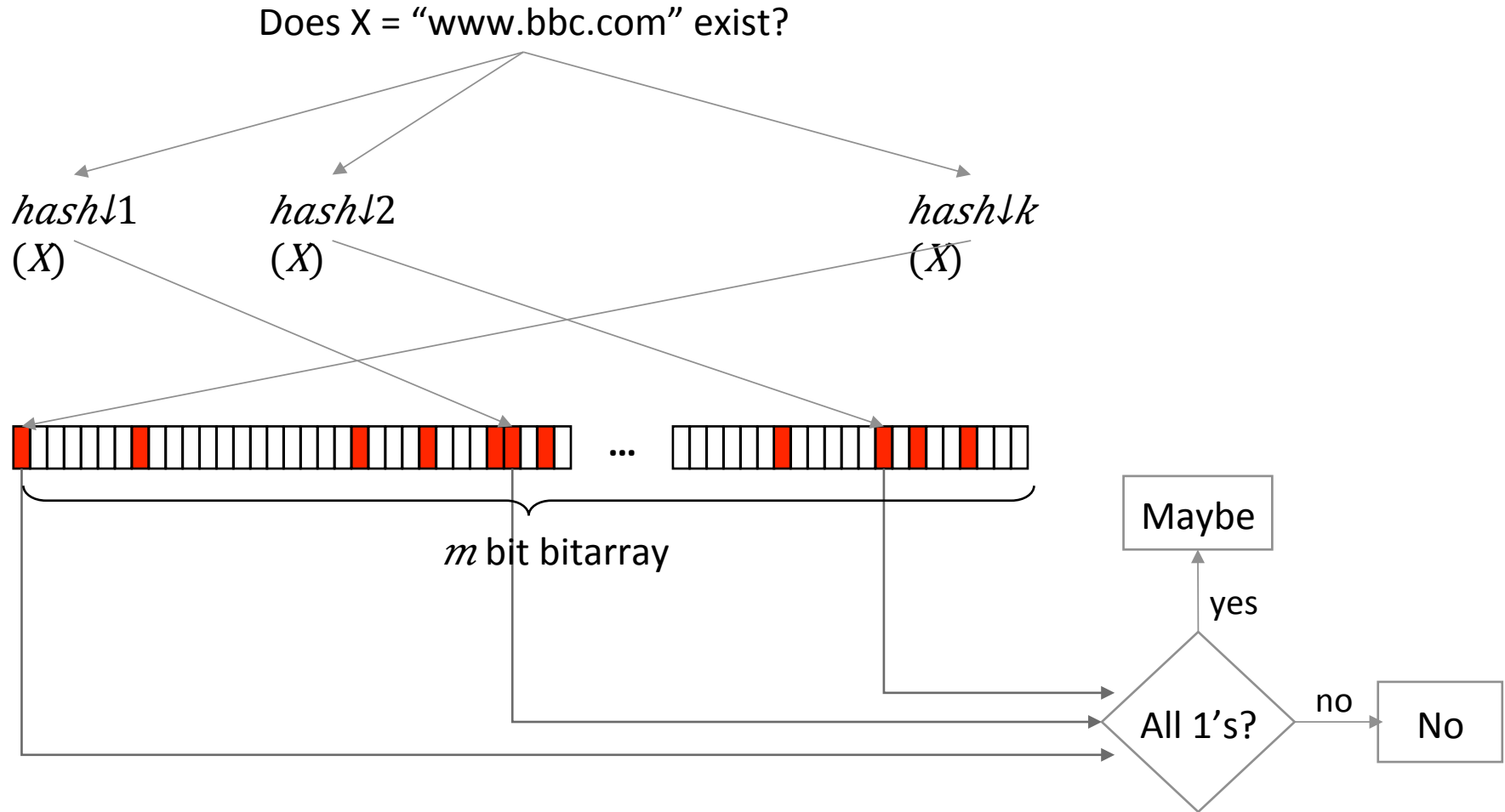


Space efficient.



# Bloom filters 101

Space efficient data-structure for set-membership queries



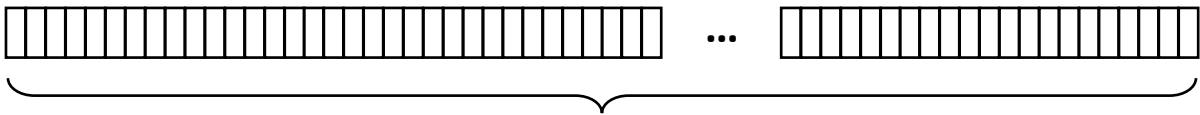
False positives can occur, false negatives can't.

# Bloom Cookies

$$0.0 < L < 1.0$$

```
profile = [ "www.tbs.com", "apple.com", "orange.com", "google.com", "en.wikipedia.org", ... ]  
set rand(bit) in bloom-filter  
for X in profile:  
    insert X in bitarray
```

bloom-filter



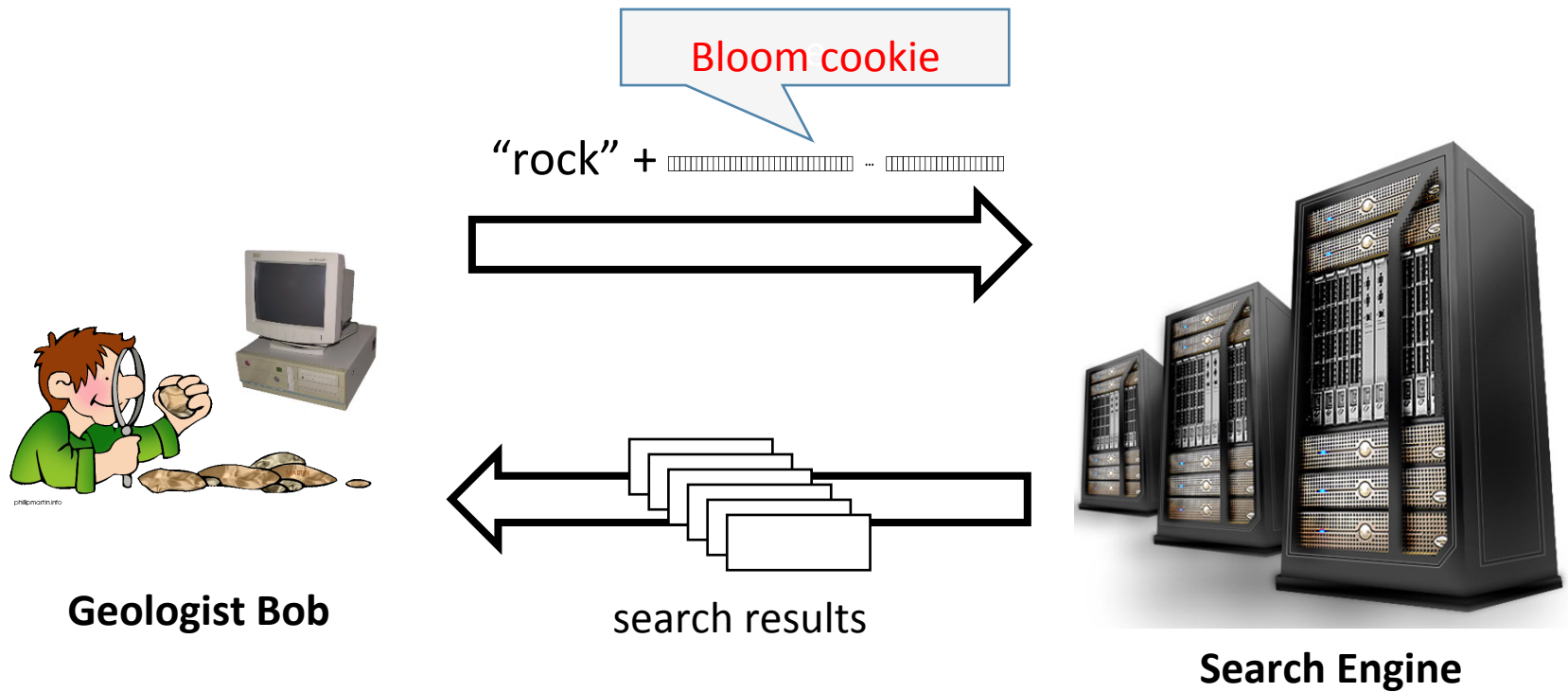
*m* bit bitarray

*Example: L=0.7 => Make sure 70% bits are set.*

- Space efficient.
- Non-deterministic noise by design.
- No dictionary required.

# Bloom Cookies

Contain enough profile information for personalization.



	Personalization loss	Linkable users
Generalization	24%	44.1%
Noise Addition	1.1%	20.0%
Bloom Cookies	3.3%	2.3%

24722 bits

2000 bits



## Summary of our work

- Systematic evaluation of existing profile obfuscation techniques: *Generalization, noise addition*.
- **Bloom cookies:** excellent personalization and privacy tradeoff compared to other methods.
- A method for end-users to configure bloom cookie parameters.

# Questions?

Nitesh Mor

[mor@berkeley.edu](mailto:mor@berkeley.edu)