Gabriel H. Nunes

MACQUARIE University
SYDNEY·AUSTRALIA

Dr. Natasha Fernandes
Prof. Annabelle McIver

# Quantitative Information Flow for Privacy Analysis

**Quantitative Information Flow** (QIF) is a decision- and information-theoretic framework based on **Formal Methods** that facilitates the analysis of **complex computational systems**.

The **Topics API** is being proposed by Google as a more privacy-friendly alternative to **third-party cookies** for **Interest-Based Advertisement** (IBA).

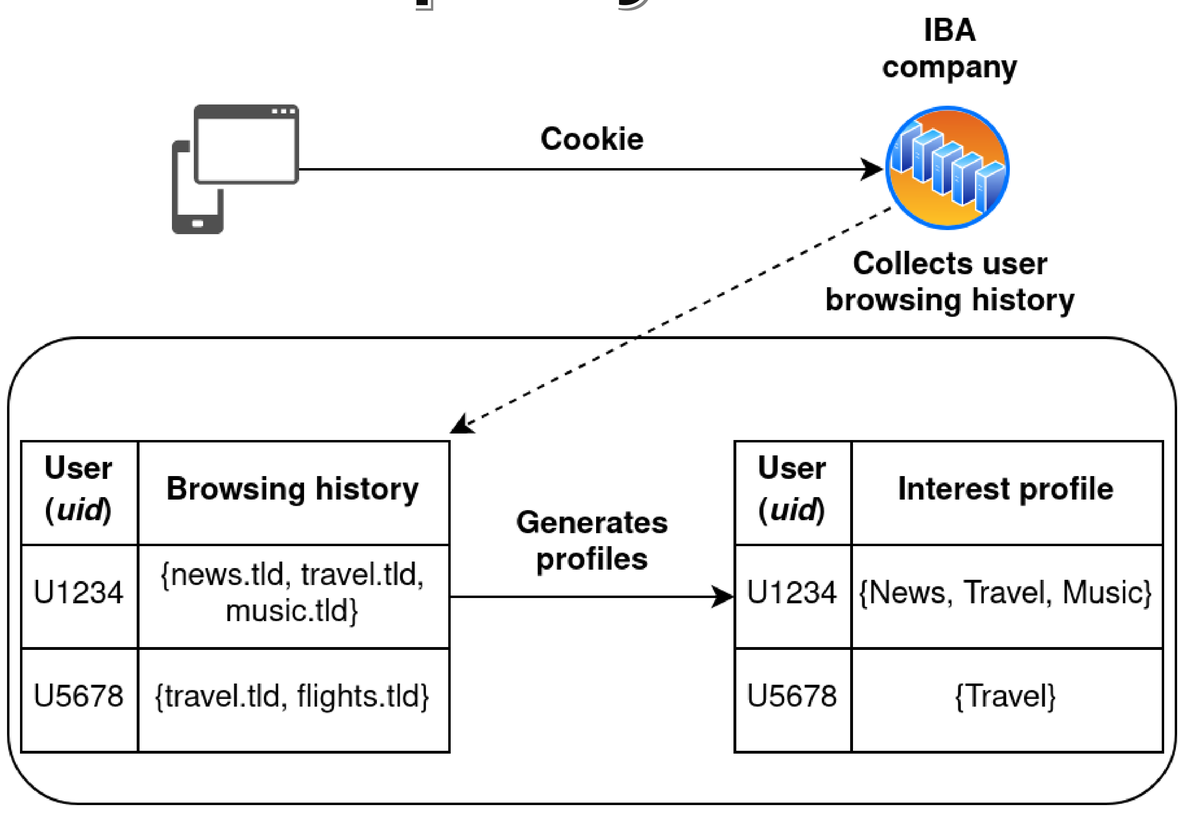
**Third-party cookies** allow the precise **tracking** of individuals' Internet browsing histories.

The **Topics API** represents an individual as a **set of top topics** of interest derived from their browsing history and a topics taxonomy. API callers can learn those topics or a **random topic** from the whole taxonomy with **5% chance**.

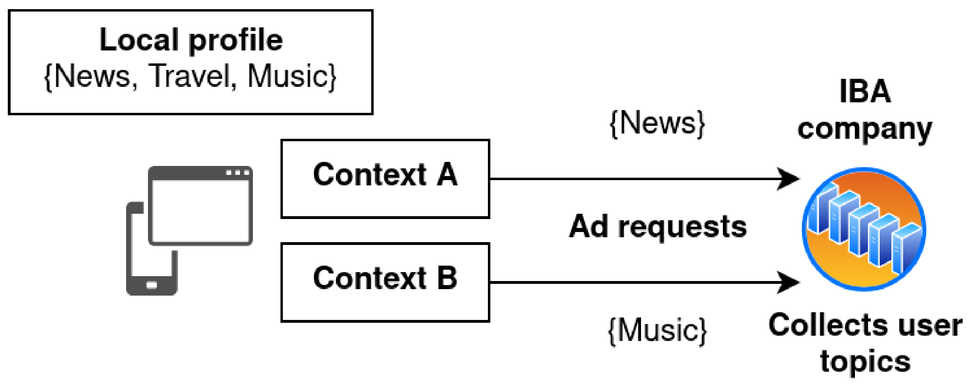
We use **QIF** to precisely **measure** the **privacy and utility** effects of each aspect of the **Topics API**, and to **verify** Google's privacy-related **claims**.

"Users should be able to understand the API, recognize what is being communicated about them, and have clear controls."

## Third-party cookies




## Topics API




## Assumption

"If the user is known to be the same across **colluding sites** (e.g. because they're logged into each with a **persistent identifier**), then it is possible for those sites **to join their topics for the user** together."

*BH*: Browsing History.

| $C_{BH}$ | $\left\{\begin{matrix}\text{news.tld}\\\text{travel.tld}\\\text{music.tld}\end{matrix}\right\}$ | $\left\{\begin{matrix}\text{travel.tld}\\\text{flights.tld}\\\text{music.tld}\end{matrix}\right\}$ |
|---|---|---|
| Alice | 1 | 0 |
| Bob | 0 | 1 |

*G*: Generalization.

| $C_{G}$ | $\left\{\begin{matrix}\text{News}\\\text{Travel}\\\text{Music}\end{matrix}\right\}$ | $\left\{\begin{matrix}\text{Travel}\\\text{Music}\end{matrix}\right\}$ |
|---|---|---|
| $\left\{\begin{matrix}\text{news.tld}\\\text{travel.tld}\\\text{music.tld}\end{matrix}\right\}$ | 1 | 0 |
| $\left\{\begin{matrix}\text{travel.tld}\\\text{flights.tld}\\\text{music.tld}\end{matrix}\right\}$ | 0 | 1 |

Channel composition: internal choice.

| $C_{BN\oplus_{0.95}DP}$ | Music | News | Travel |
|---|---|---|---|
| $\left\{\begin{matrix}\text{News}\\\text{Travel}\\\text{Music}\end{matrix}\right\}$ | $\frac{2.95}{6}$ | $\frac{2.95}{6}$ | $\frac{0.05}{3}$ |
| $\left\{\begin{matrix}\text{Travel}\\\text{Music}\end{matrix}\right\}$ | $\frac{2.95}{6}$ | $\frac{0.05}{3}$ | $\frac{2.95}{6}$ |

95% $\oplus_{0.95}$ 5%

*BN*: Bounded Noise.

| $C_{BN}$ | Music | News | Travel |
|---|---|---|---|
| $\left\{\begin{matrix}\text{News}\\\text{Travel}\\\text{Music}\end{matrix}\right\}$ | $\frac{1}{2}$ | $\frac{1}{2}$ | 0 |
| $\left\{\begin{matrix}\text{Travel}\\\text{Music}\end{matrix}\right\}$ | $\frac{1}{2}$ | 0 | $\frac{1}{2}$ |

*DP*: Differential Privacy.

| $C_{DP}$ | Music | News | Travel |
|---|---|---|---|
| $\left\{\begin{matrix}\text{News}\\\text{Travel}\\\text{Music}\end{matrix}\right\}$ | $\frac{1}{3}$ | $\frac{1}{3}$ | $\frac{1}{3}$ |
| $\left\{\begin{matrix}\text{Travel}\\\text{Music}\end{matrix}\right\}$ | $\frac{1}{3}$ | $\frac{1}{3}$ | $\frac{1}{3}$ |

## Privacy Methods

- **Generalization:**
  Many-to-one mapping.
  May provide *k*-anonymity.
- **Bounded Noise:**
  One-to-many mapping.
  May provide *k*-anonymity.
- **Differential Privacy:**
  State-of-the-art noise-addition.
  Provides plausible deniability.

"It must be difficult to re-identify [a] significant numbers of users across sites using just the API."

"The 5% noise is introduced [...] to provide some amount of plausible deniability."

## Channel Capacity (Bayes Leakage)

$$r+\frac{m(1-r)}{k}$$

Secret reconstruction

## Differential Privacy (ε)

$$\ln\left(1+\frac{m(1-r)}{kr}\right)$$

Plausible deniability

- *m*: total number of topics in the taxonomy.
- *k*: total number of topics available to be reported by each user.
- *r*: probability of reporting a random topic from the whole taxonomy.