463.12 Web Privacy

Computer Security II
CS463/ECE424
University of Illinois
Security News of the Day

- https://www.chromium.org/Home/chromium-privacy/privacy-sandbox
- https://digiday.com/marketing/wtf-googles-privacy-sandbox/

Google is done with cookies, but that doesn’t mean it’s done tracking you

A third-party cookie ban won’t hurt the search giant’s healthy first-party data ad business.

By Sara Morrison  |  Mar 3, 2021, 5:50pm EST

The basics
Google said yesterday that it would phase out the use of third-party cookies in its Chrome browser by 2022.

WTF is Google’s Privacy Sandbox?
That’s where Google’s Federated Learning of Cohorts (FLoC) comes in, which Google says is a “privacy-first” and “interest-based” advertising technology. With FLoC, Chrome will keep track of a user’s browsing habits across the web, and then place the user in various audiences, or “cohorts,” based on those habits. Advertisers will then target their ads to cohorts, rather than an individual user. So if you’re looking for a
What is “Privacy”?

• Difficult-to-define concept.

• Commonly refers to the desires and expectations of individuals about how, when, what, and to whom information about themselves is revealed to others.

• Contrasts with security, which concerns protection against malicious attackers seeking to steal data or disrupt operations.
Privacy Norms are Universal but Variable

Scholars have uncovered evidence of privacy-seeking behaviors across peoples and cultures separated by time and space: from ancient Rome and Greece to pre-industrialized Javanese, Balinese, and Tuareg societies.

- The Quran instructs against spying on one another;
- The Talmud advises home-builders to position windows so that they do not directly face those of one’s neighbors;
- The Bible relates how Adam and Eve discovered their nakedness after eating the fruit of knowledge and covered themselves in shame from the prying eyes of God.
Example

- **Privacy**
  If Alice tells her friend Bob a personal fact about herself in confidence, like, say, that she has been diagnosed with cancer, and then Bob mentions this in a Facebook comment, Alice may well consider this a violation of her privacy.

- **Security**
  If Alice responds to a phishing email and reveals her personal banking password to a hacker who accesses her bank account, then Alice is a victim of a security breach.

  This is probably also a privacy breach since Alice probably would not reveal her bank balance to a stranger.
Alan Westin Surveys

Influential classification and surveys from early 1990s contrast three types of attitudes towards privacy

1. Privacy Fundamentalists
2. Privacy Unconcerned
3. Privacy Pragmatists

Questionaire

Respondents were asked

1. Whether they are very concerned about threats to their personal privacy today,
2. Whether they agree strongly that business organizations seek excessively personal information from consumers,
3. Whether they agree strongly that the Federal government since Watergate is still invading the citizen’s privacy, and
4. Whether they agree that consumers have lost all control over circulation of their information.
Questionaire

Respondents were asked

1. Whether they are very concerned about threats to their personal privacy today,

2. Whether they agree strongly that business organizations seek excessively personal information from consumers,

3. Whether they agree strongly that the Federal government since Watergate is still invading the citizen’s privacy, and

4. Whether they agree that consumers have lost all control over circulation of their information.
Privacy Fundamentalists

• Fundamentalists are
  – Generally distrustful of organizations that ask for their personal information,
  – Worried about the accuracy of computerized information and additional uses made of it, and
  – In favor of new laws and regulatory actions to spell out privacy rights and provide enforceable remedies.

• They generally choose privacy controls over consumer-service benefits when these compete with each other.
Privacy Unconcerned

Privacy unconcerned

• Assert that they do not see what the “privacy fuss” is all about,
• Support the benefits of most organizational programs over warnings about privacy abuse,
• Have little problem with supplying their personal information to government authorities or businesses, and
• See no need for creating another government bureaucracy (a “Federal Big Brother) to protect someone’s privacy.
Privacy Pragmatists

A personal who is guided more by practical considerations than by ideals

- Privacy pragmatists
  - Weigh the value to them and society of business or government programs calling for personal info,
  - Examine the relevance and social propriety of the information sought,
  - Want to know the potential risks to privacy or security of their information,
  - Look to see whether fair information practices are being widely enough observed,

- They use these and similar factors to decide whether they will agree or disagree with specific information activities.
  - Key factor: whether they trust the industry or company involved

- Privacy pragmatists favor voluntary standards and consumer choice over legislation and government enforcement.
  - But they will back legislation when they think not enough is being done (or meaningfully done) by voluntary means.
<table>
<thead>
<tr>
<th></th>
<th>1990</th>
<th>1991</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fundamentalists</td>
<td>46%</td>
<td>41%</td>
</tr>
<tr>
<td>Unconcerned</td>
<td>17%</td>
<td>20%</td>
</tr>
<tr>
<td>Pragmatists</td>
<td>36%</td>
<td>39%</td>
</tr>
</tbody>
</table>
Privacy Paradox

• There arose doubts about the power of attitudinal scales to predict actual privacy behavior.

• This discrepancy between attitudes and behaviors has become known as the “privacy paradox.”

• At least three factors seem relevant
  – Uncertainty
  – Context Dependence
  – Malleability and Influence

[Acquisti15]
Uncertainty

- People are often uncertain about which privacy choices to make.
- Exacerbating factors of technology include incomplete and asymmetric information.
- Experiments often show disregard for privacy versus other factors.

Not always, consider this endowment effect experiment:
- Anonymous gift cards worth $10 vs. identified cards worth $12.
- Do you want to trade?

- Of the subjects who originally held the less valuable but anonymous card, five times as many (52.1%) chose it and kept it over the other card than did those who originally held the more valuable card (9.7%).
- This suggests that people value privacy more when they have it than when they do not.
Context Dependence

• Although desire for privacy is widespread, individual seek clues in their environment about rules
  – Giving personal info to a hospital vs. to random mobile apps
  – Do what others do

• But clues can lead to worrisome places
  – Revealing private information to a bot
Malleability and Influence

• Because they are not sure what they want and are often led by subtle clues in their privacy decisions, individuals are subject to manipulation.
  – Posting a privacy policy that users do not read may lead to comfort
  – Offering many privacy choices can lead to less privacy
  – Default configuration
Display Advertising

- **Advertisers** want the ads for their products to reach plausible clients.
  - e.g., ford, coca cola, mcdonalds
- **Publishers** are willing to place ads on their webpages, but they expect to be paid
  - eg., the New York Times website
- **Brokers** mediate between advertisers, publishers, and clients: the advertisers send their ads and bids to the broker, who then distributes them to the publishers’ webpages, which are then viewed by users.
  - eg. Google and Yahoo
- **Users**: may click on an ad if they find it relevant to their needs or interests.
- For every ad viewed in the **pay-per-view (PPV)** advertising model or clicked in the **pay-per-click (PPC)** advertising model, the advertiser pays the broker who in turn pays the publisher.
Online Behavioral Advertising (OBA)

- Publishers embed a link for the broker on their webpages.
  - The sole purpose of these broker links is user tracking.
- When a user views the webpage, the user’s browser contacts the broker’s servers.
- This enables the broker to track the user across all partnering publishers.
- The broker then runs its own algorithm over the tracked data to decide which ad to present on the publisher’s page.
- This tracking practice poses a threat to the privacy of users:
  - Research has shown that it is often possible to link tracked information to personally identifying information of the user.
OBA Auction Scenario

- Publishers
- Trackers (Broker (Ad exchange))
- Advertisers
- Client
Trackers track users
Compile user profile
Trackers (may) share profiles with advertisers
Client gets webpage with adbox
Client tells broker of page

Broker (Ad exchange)

Client

Publishers

Trackers

Advertisers
Broker launches auction (for given user visiting given webpage ....)
Also does clickfraud etc.
Advertisements present bids and ads through brokers (Ad exchange). Advertisers and trackers interact with brokers.
Broker (Ad exchange) picks winners, delivers ads.
Broker waits for results (view/click)
Publishers → Trackers → Broker (Ad exchange) → Advertisers

Reports them to interested parties

Client
“Second Price Auctions” for Advertising

How do we select which ads to show to which users and how do advertisers pay Cost Per Click (CPC) on the ads they provide?

– Bid $B$
– Global quality score (at broker) $G$
– User quality score (at user) $U$
– Rank = $B \times G \times U$
– CPC = $B_n \left( \frac{(G_n \times U_n)}{(G_c \times U_c)} \right)$
  
  Where $B_n$, $G_n$, $U_n$ are for the next lowest bid and $B_c$, $G_c$, $U_c$ are for the current (winning) bid
  
  CPC must $\leq B_c$
Quality Scores

- Historical click performance of the ad
- Landing page quality
- ........
- Relevance to the user
- User click through rates
- ........

$Known at broker$ (call it $G$)

$Known at user$ (call it $U$)
Is OBA a Violation of Privacy?

• Linking behavioral data with unique identifiers
• Detail and scope of data collection
• OBA based on “negative” assumptions
• Sensitivity of data
• Impact on operability
• Ease of opting out
• Lack of notice
Do-Not-Track Requirements

- **Anonymity**
  
  The broker cannot associate any unit of learned information with Personally Identifiable Information (PII) of any user (including network address). Requires an anonymizing proxy.

  Related to PII (real-world identity)

- **Unlinkability**
  
  The broker cannot associate separate units of learned information with a single (anonymous) client. This prevents a broker from building up a user profile, and then associating it with a known user using externally gathered knowledge.

  Related to linking behaviors across sites
Market Concerns

Advertisers, publishers, and brokers have their own concerns.

- Targeting ads makes ad dollars more effective. Avoid rules that constrain it excessively.
- Privacy rules must not significantly impact performance. Customers are sensitive to delays.
- Privacy rules must respect the competitive concerns of participants. E.g. advertisers compete and do not want their bids disclosed.
- Do not interfere with preventing fraud (such as click fraud).
Strategies for Privacy Protection

• Encourage trust in the data server (“don’t be evil”)
  – Current situation, but there are concerns ...
• Make decisions on the client device so private information is never shared with the data server
  – Don’t send U to the server
• Transform data from the client so that privacy risks of the transformed data on the server are limited
  – Send something derived from U but not U itself
• Use a trusted third party
Three flavors of Non-Tracking auctions

[Reznichenko11]
Three flavors of Non-Tracking auctions

Rank@Client

Broker (Bid, G)

Rank@Broker

Bid, G

Rank@3^{rd} Party

3 party

[Kumaraguru05]
Computes ranking: 

$$ (B \times G) \times U $$

A - the ad ID,
Value of \((B \times G)\),
\(E[B,G]\),
(+ targeting etc.)
A - the ad ID, Value of ($B \times G$), $E[B,G]$, (+ targeting etc.)

Computes ranking: $(B \times G) \times U$

$A_c$ - clicked ad ID

$((B_n \times G_n) \times (U_n / U_c))$

$E[B_c, G_c]$

Decrypts $E[B_c, G_c]$

Computes CPC:

$((B_n \times G_n) \times (U_n / U_c)) / G_c$

Checks that CPC $\leq B_c$

Client (U)

Broker (Bid, G)
Decrypts $E[B_c, G_c]$

Computes CPC:

$$CPC = B_n \left( \frac{G_n \times U_n}{G_c \times U_c} \right)$$

Checks that $CPC \leq B_c$
A clicked ad ID

((B \times G) \times \frac{U_n}{U_c})

Computes ranking:

(B \times G) \times U

User information obscured by hiding within this composite value (+ targeting etc.)

Decrypts E[B_c, G_c]

Computes CPC:

\frac{((B_n \times G_n) \times U_n)}{U_c} / G_c

Checks that CPC \leq B_c
A - the ad ID, Value of $(B \times G)$, $E[B,G]$, (+ targeting etc.)

Computes ranking: $(B \times G) \times U$

Client (U)

Broker (Bid, G)

$A_c$ - clicked ad ID
$(B_n \times G_n) \times U_n / U_c$

Decrypts $E[B_c, G_c]$

Computes CPC:
$((B_n \times G_n) \times U_n / U_c) / G_c$

Checks that CPC $\leq B_c$
A - the ad ID,
Value of \((B \times G)\),
\(E[B,G]\),
\((+\text{targeting etc.)})

Nuance: \(B_c\) and \(G_c\) may have changed between ranking and CPC calculation

Deciphers \(E[B_c, G_c]\)

Computes CPC:
\[
\frac{((B_n \times G_n) \times U_n / U_c)}{G_c}
\]

Checks that \(\text{CPC} \leq B_c\)
Privad: a privacy preserving ad system

- ~4800 active users per day
- Only in October
  - ad requests 1.1M
  - ads 9.5M
  - ad views 790K
  - ad clicks 417
  - product purchases 4
Privad: a privacy preserving ad system

• CTR comparable to Google “display text ads”

• “Search ads” achieve better CTR than “display ads”

• Majority of users who opted-in are tech-savvy with many of them having ad blocking software, browse in private mode and rarely click on ads
Reading

Discussion Questions

• What kind of privacy person are you? Why? Can you think of a case where you (or others) were influenced by factors like context?

• Analyze the likely views of users from various privacy profiles and factors with respect to OBA.
  – Example: what might be the role of manipulation and influence?