Web Privacy

CS463/ECE424

University of Illinois

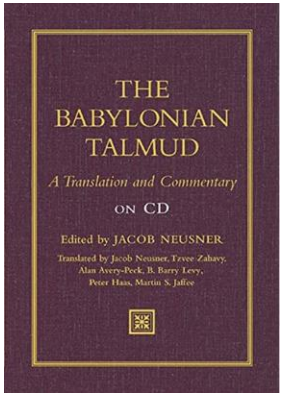
What is “Privacy”?

- 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

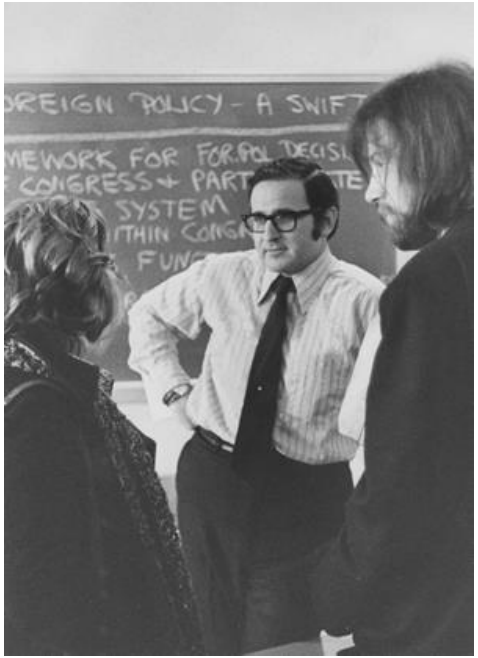
If Alice responds to a phishing email and reveals her personal banking password to a hacker who accesses her bank account, then Alice is a victim of a security breach.

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



WESTIN, A., AND HARRIS LOUIS & ASSOCIATES. Harris-Equifax Consumer Privacy Survey. Tech. rep., 1991. Conducted for Equifax Inc. (1,255 adults of the U.S. public)

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.

Survey Results

	1990	1991
Fundamentalists	46%	41%
Unconcerned	17%	20%
Pragmatists	36%	39%

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

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

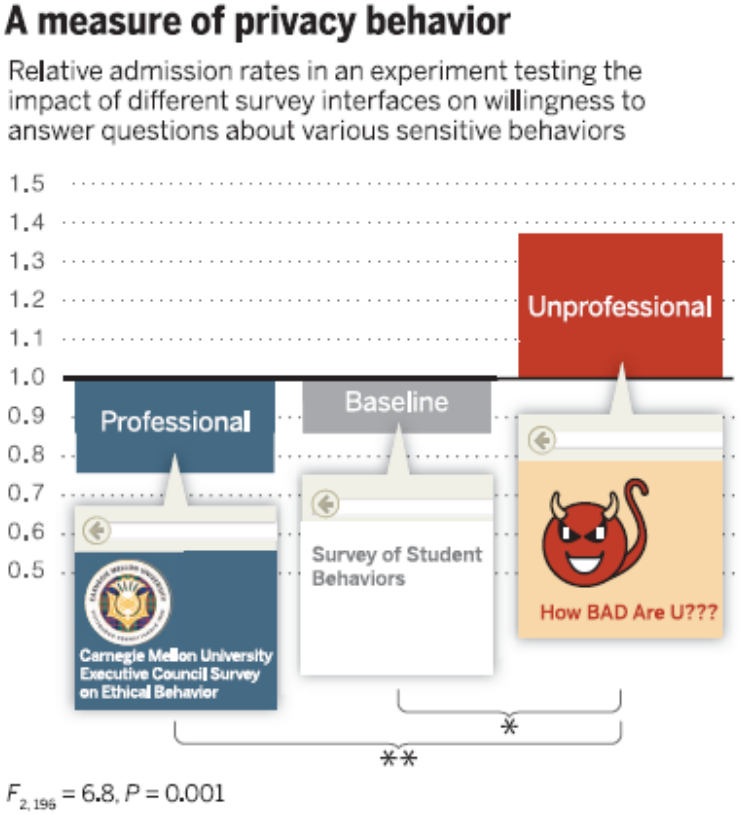
- Not always, consider this endowment effect experiment:
- Anonymous gift cards worth $10 vs. identified cards worth $12.
- Do you want to trade?

- 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

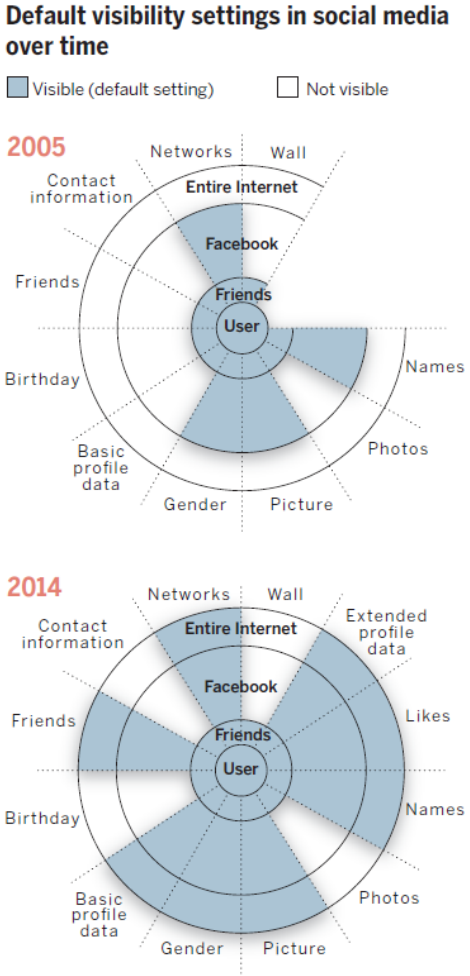
- Although desire for privacy is widespread, individual seek clues in their environment about rules
  - Giving personal info to a hospital vs. to random mobile apps
  - Do what others do

- But clues can lead to worrisome places
  - Revealing private information to a bot



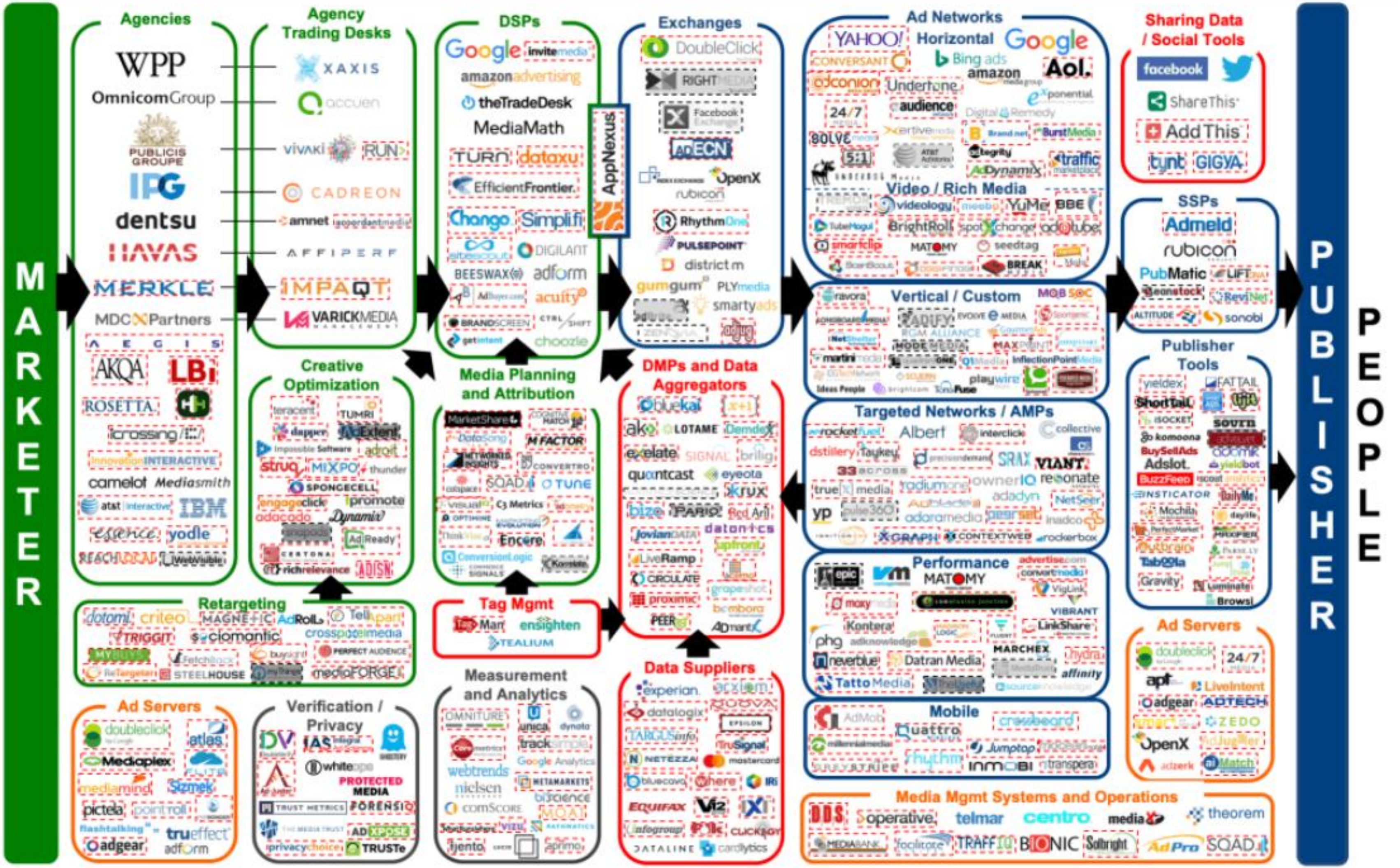
Malleability and Influence

- 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
- **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.

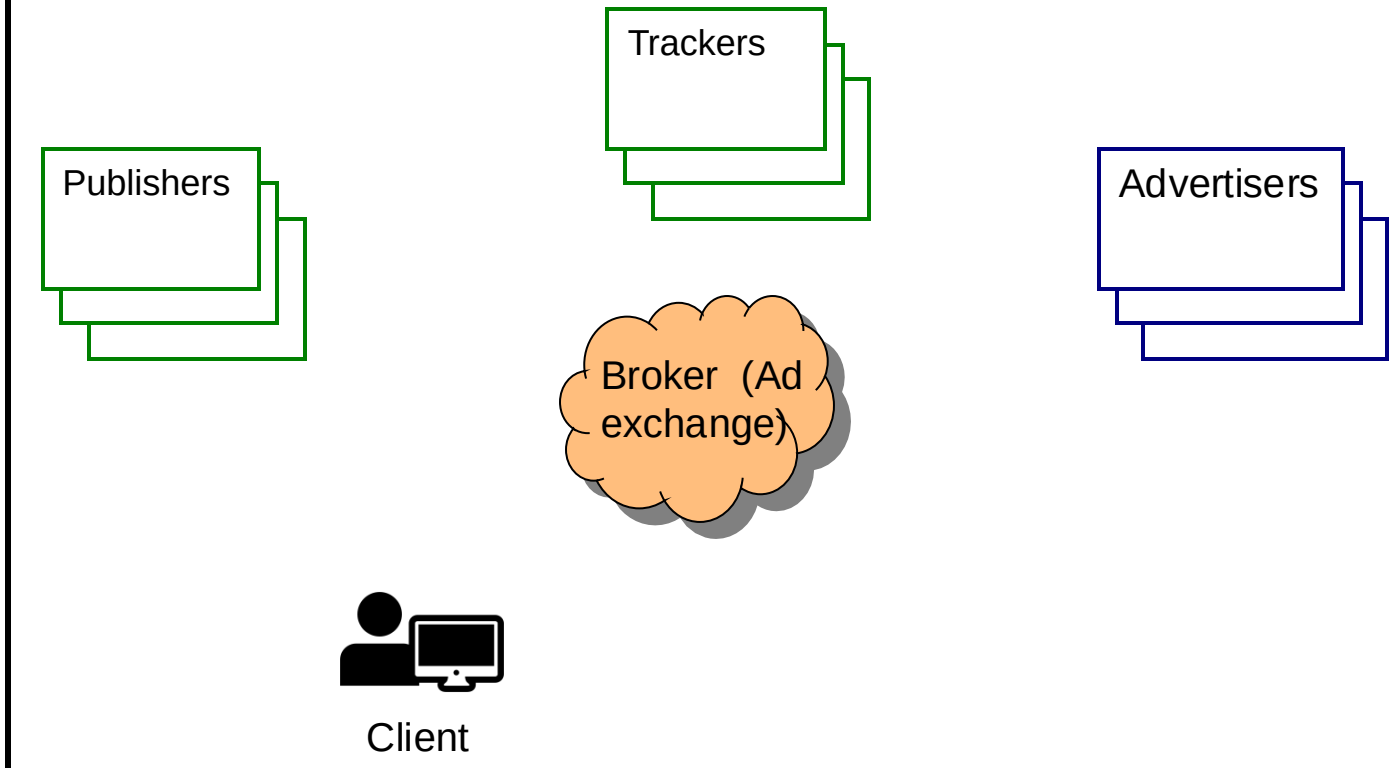


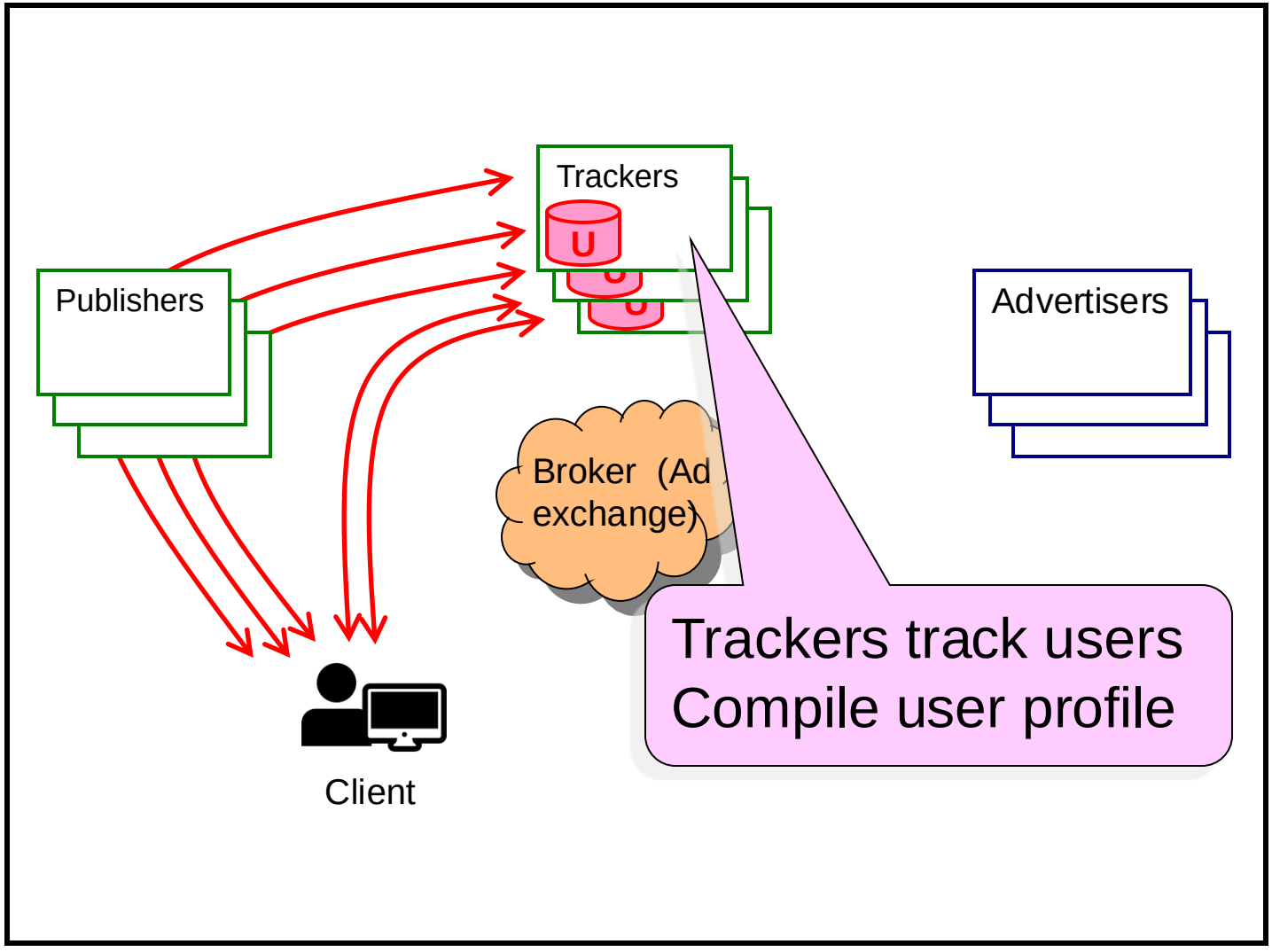
MARKETER
Agencies
Agency Trading Desks
DSPs
AppNexus
Exchanges
Ad Networks
Horizontal
Video / Rich Media
Vertical / Custom
Targeted Networks / AMPs
Performance
Mobile
Sharing Data / Social Tools
SSPs
Publisher Tools
Ad Servers
Creative Optimization
Media Planning and Attribution
DMPs and Data Aggregators
Retargeting
Tag Mgmt
Data Suppliers
Ad Servers
Verification / Privacy
Measurement and Analytics
Media Mgmt Systems and Operations
PUBLISHER
PEOPLE

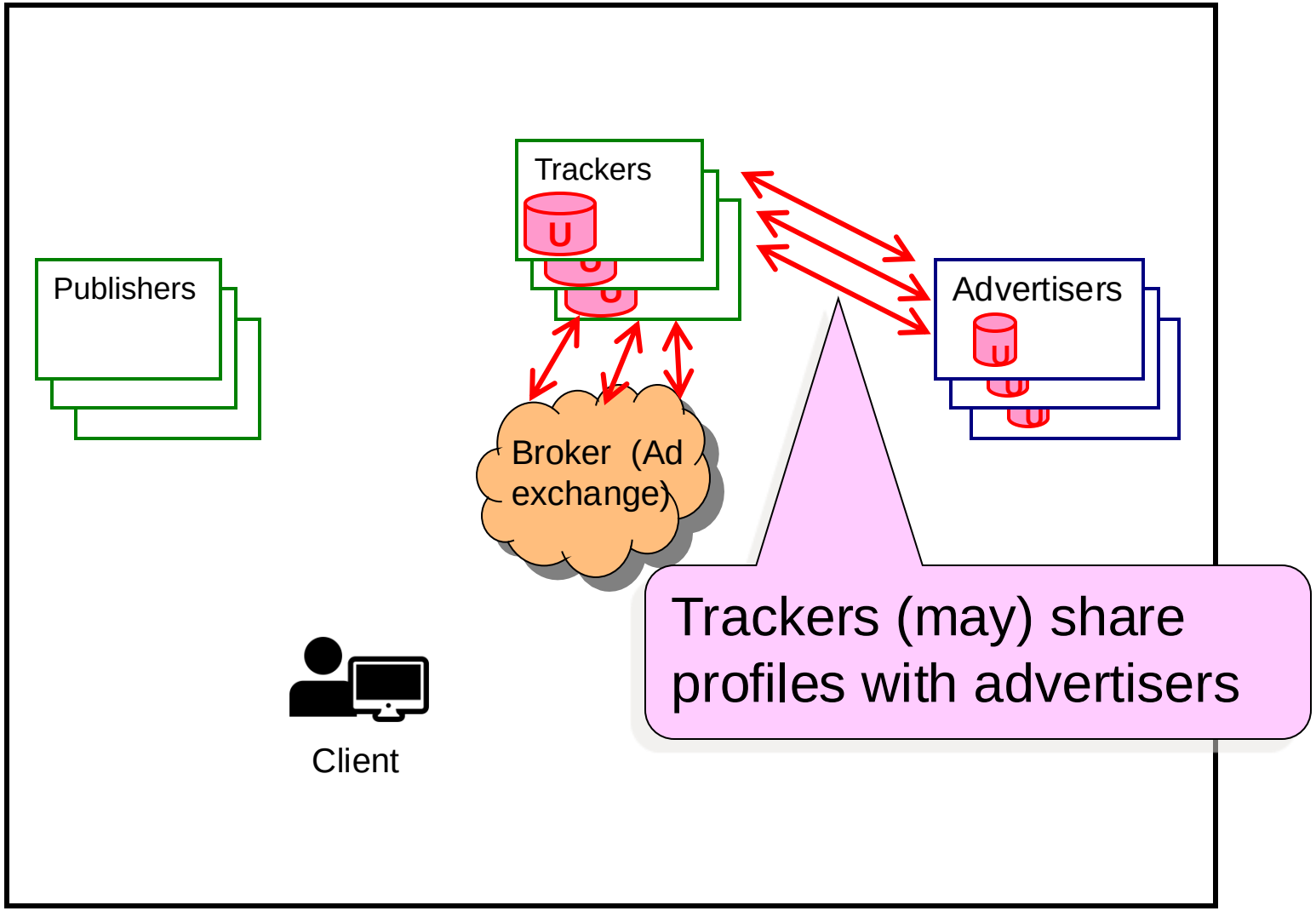
Online Behavioral Advertising (OBA)

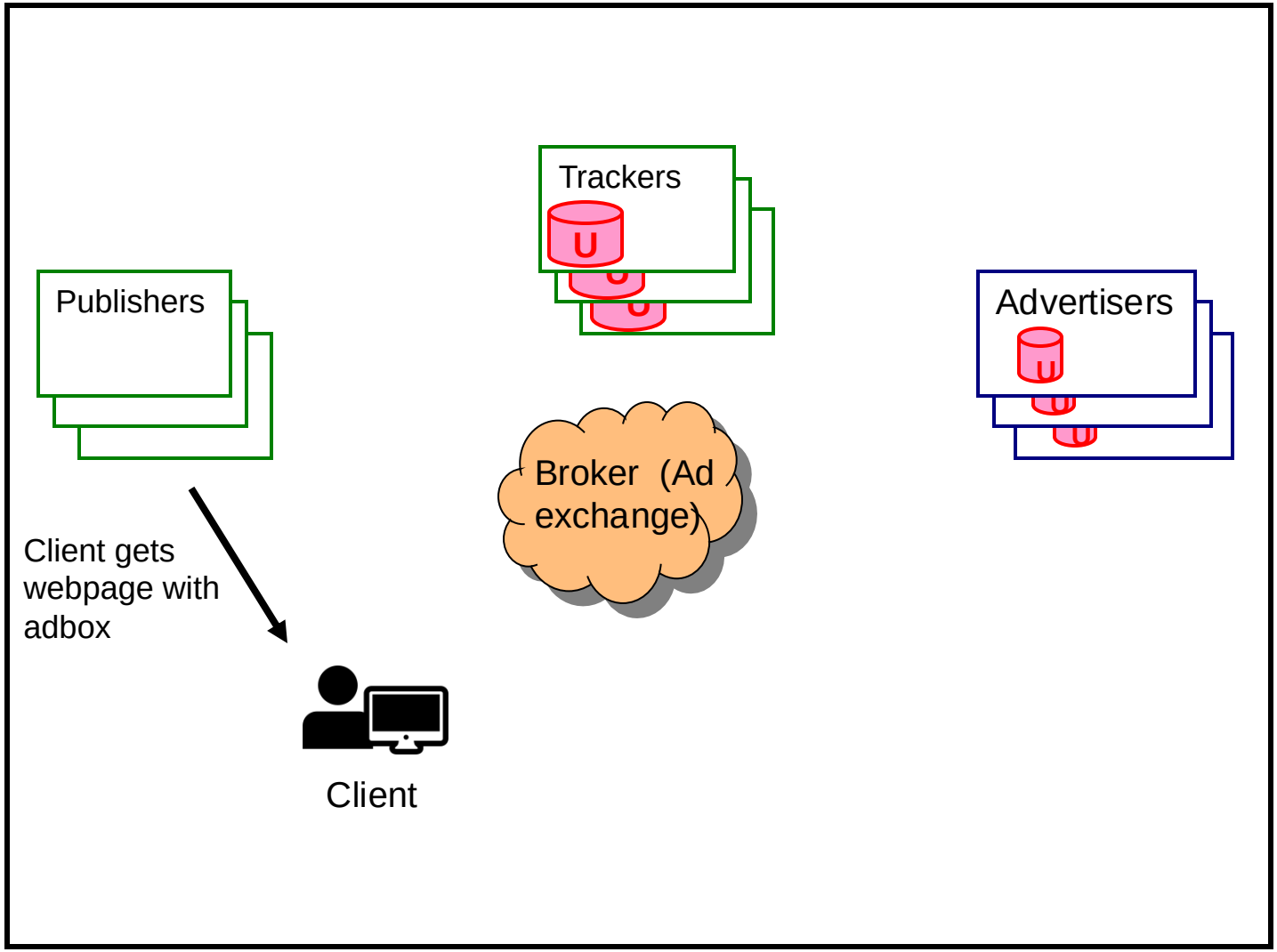
- Publishers embed a link for the broker on their webpages.
  - The sole purpose of these broker links is user tracking.
- When a user views the webpage, the user's browser contacts the broker's servers.
- This enables the broker to track the user across all partnering publishers.
- The broker then runs 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.

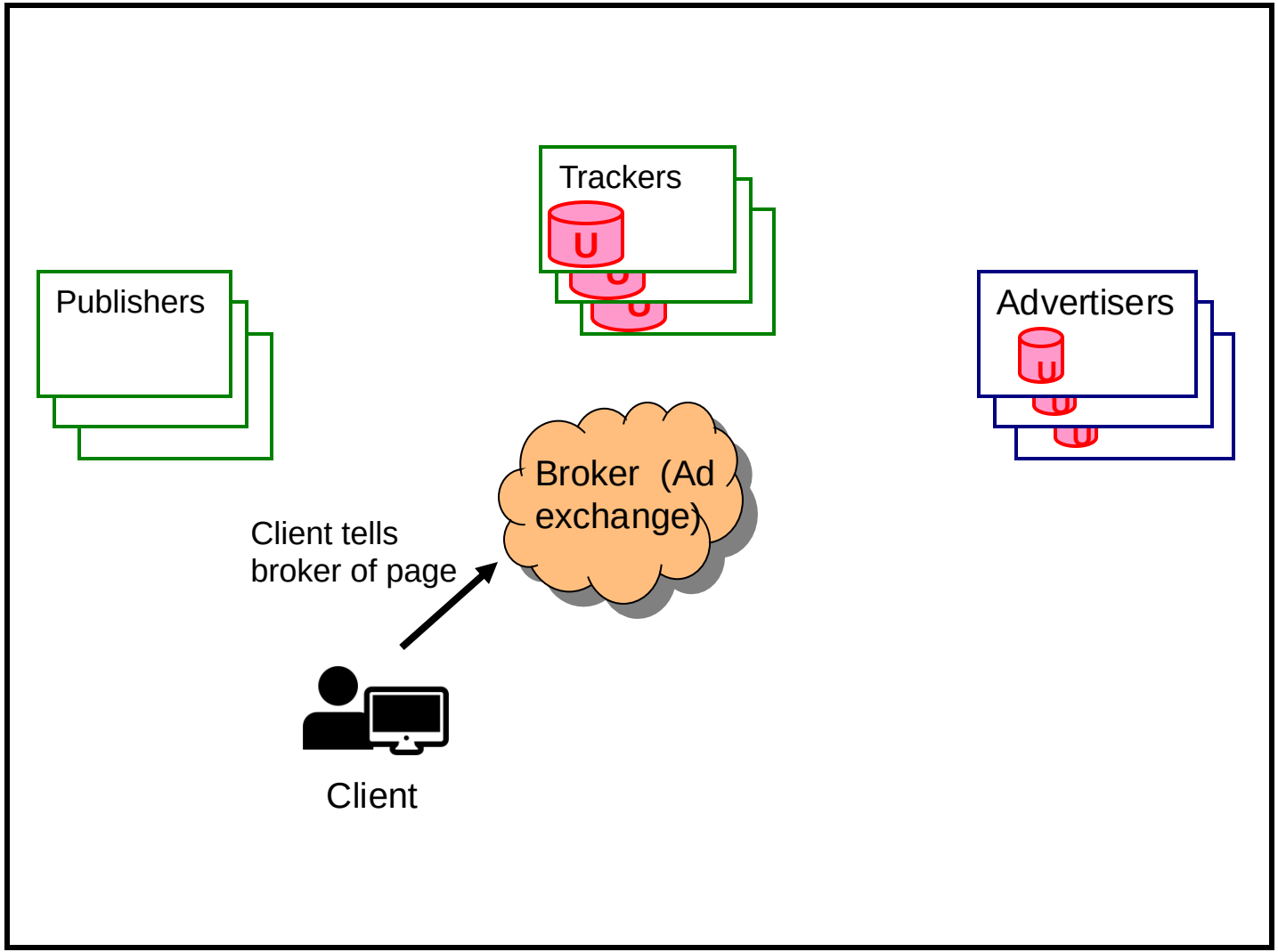
OBA Auction Scenario

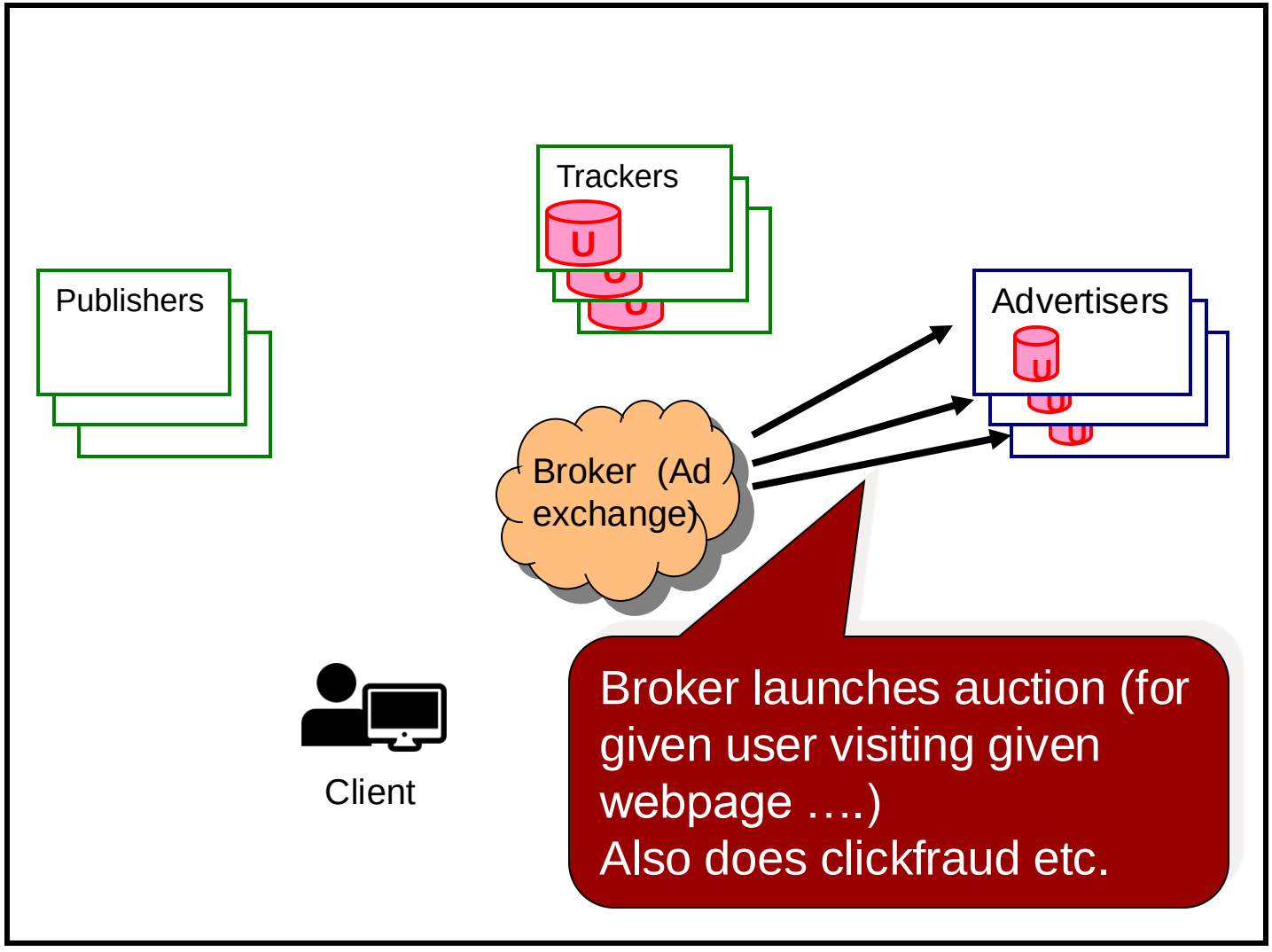


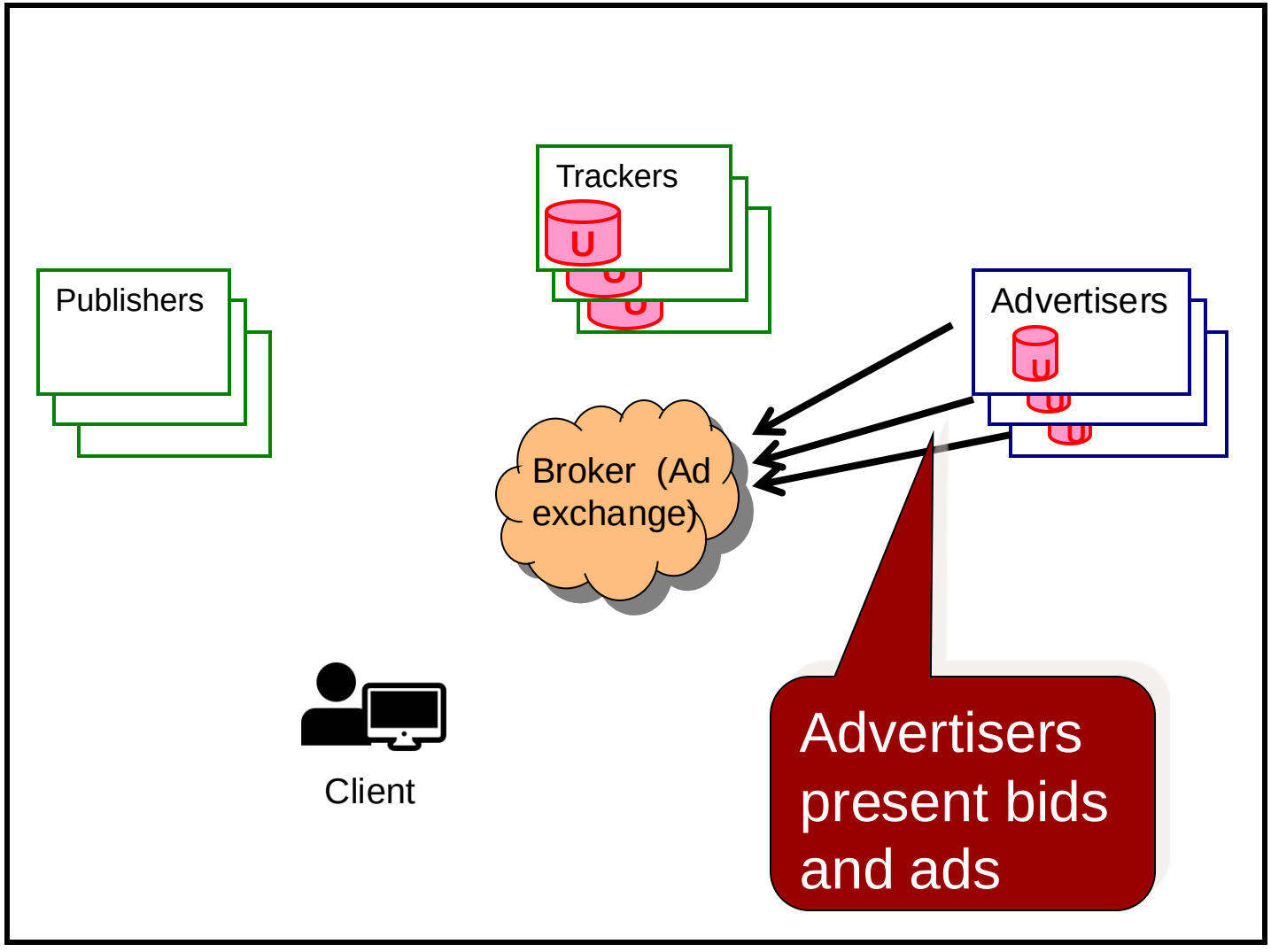

Trackers
U
U
U
Publishers
Advertisers
Broker (Ad exchange)
Trackers track users
Compile user profile
Client

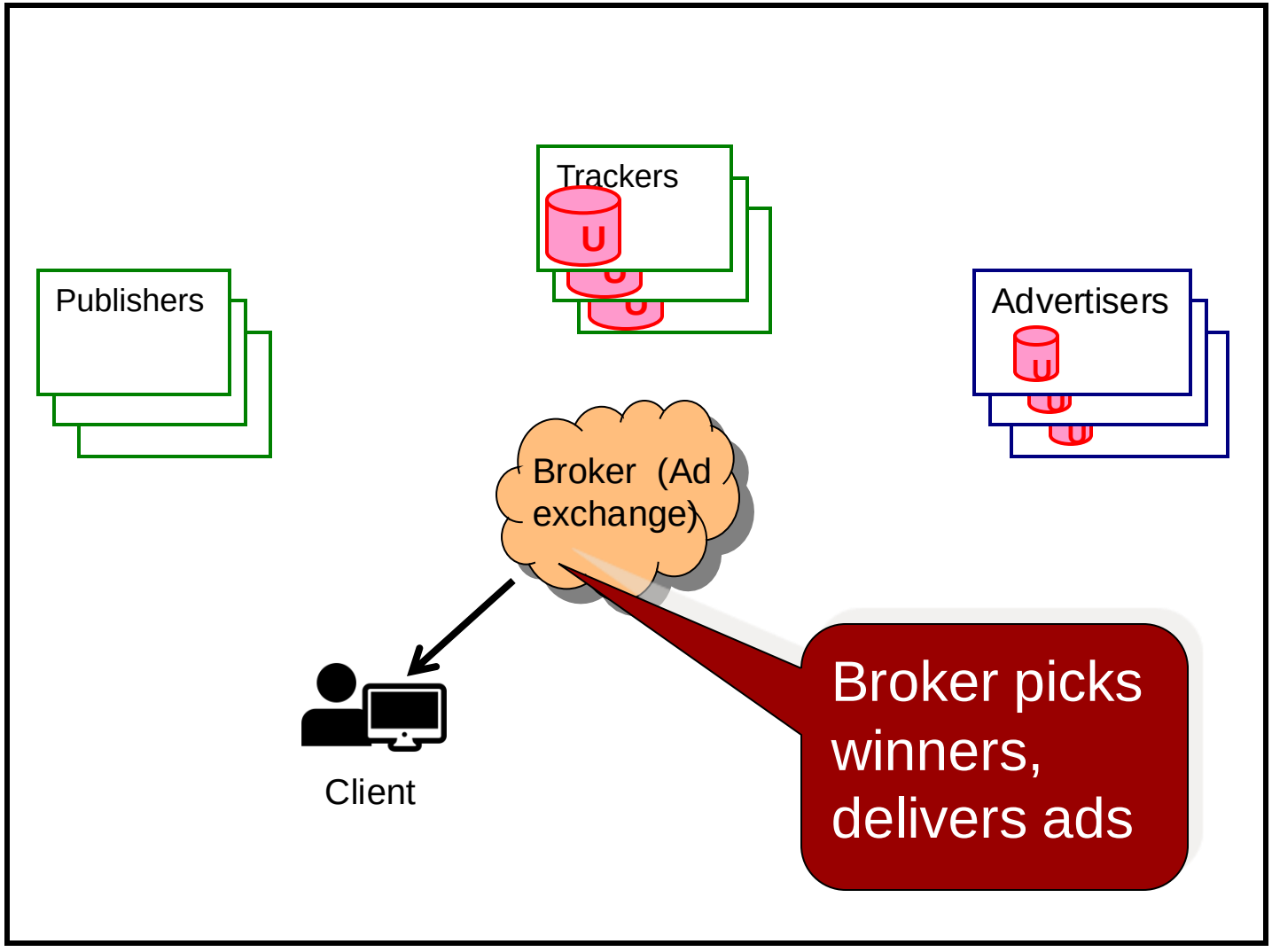

Trackers
U
Publishers
Advertisers
U
Broker (Ad exchange)
Trackers (may) share profiles with advertisers
Client

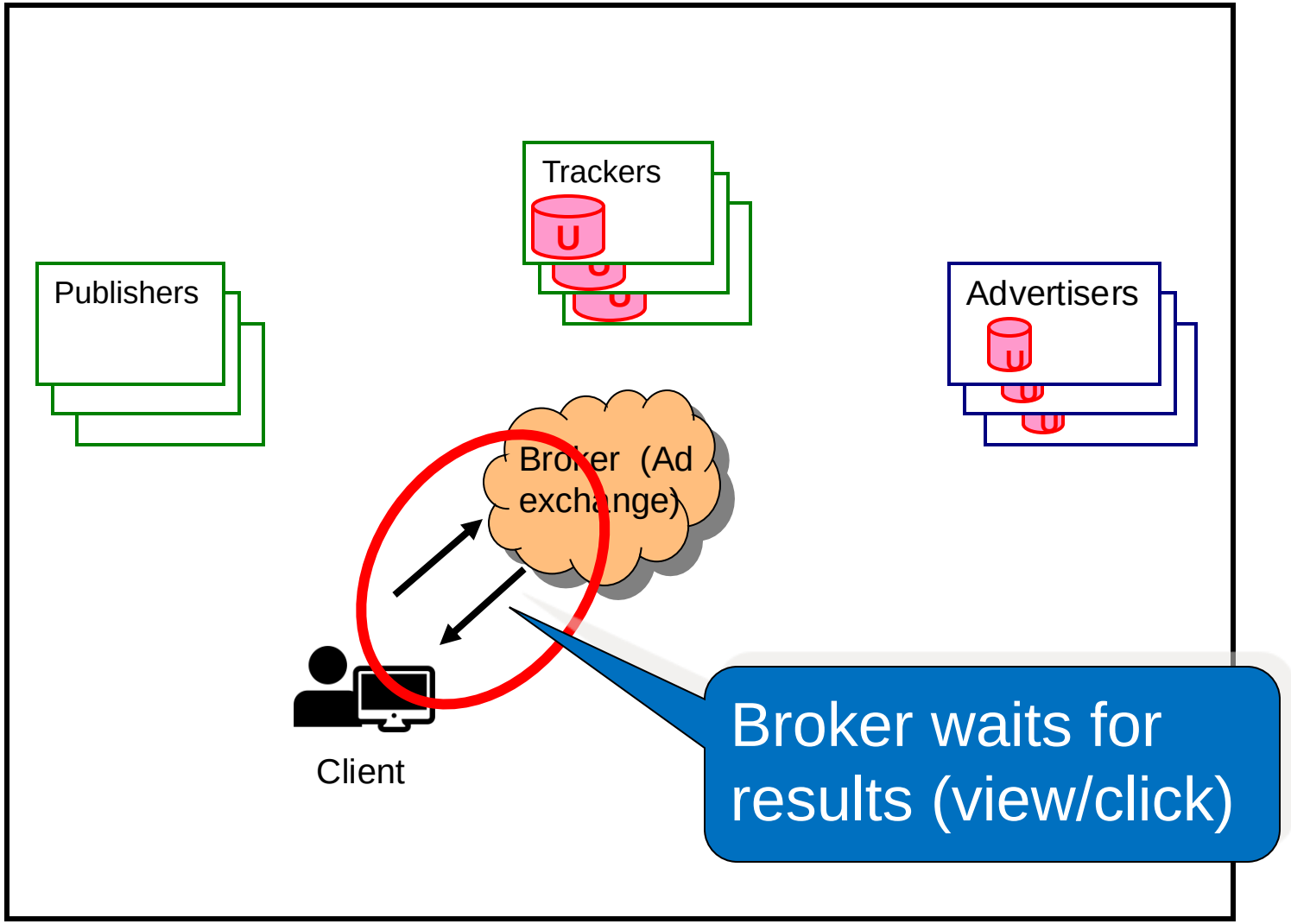

Trackers
Publishers
Advertisers
Broker (Ad exchange)
Client gets webpage with adbox
Client

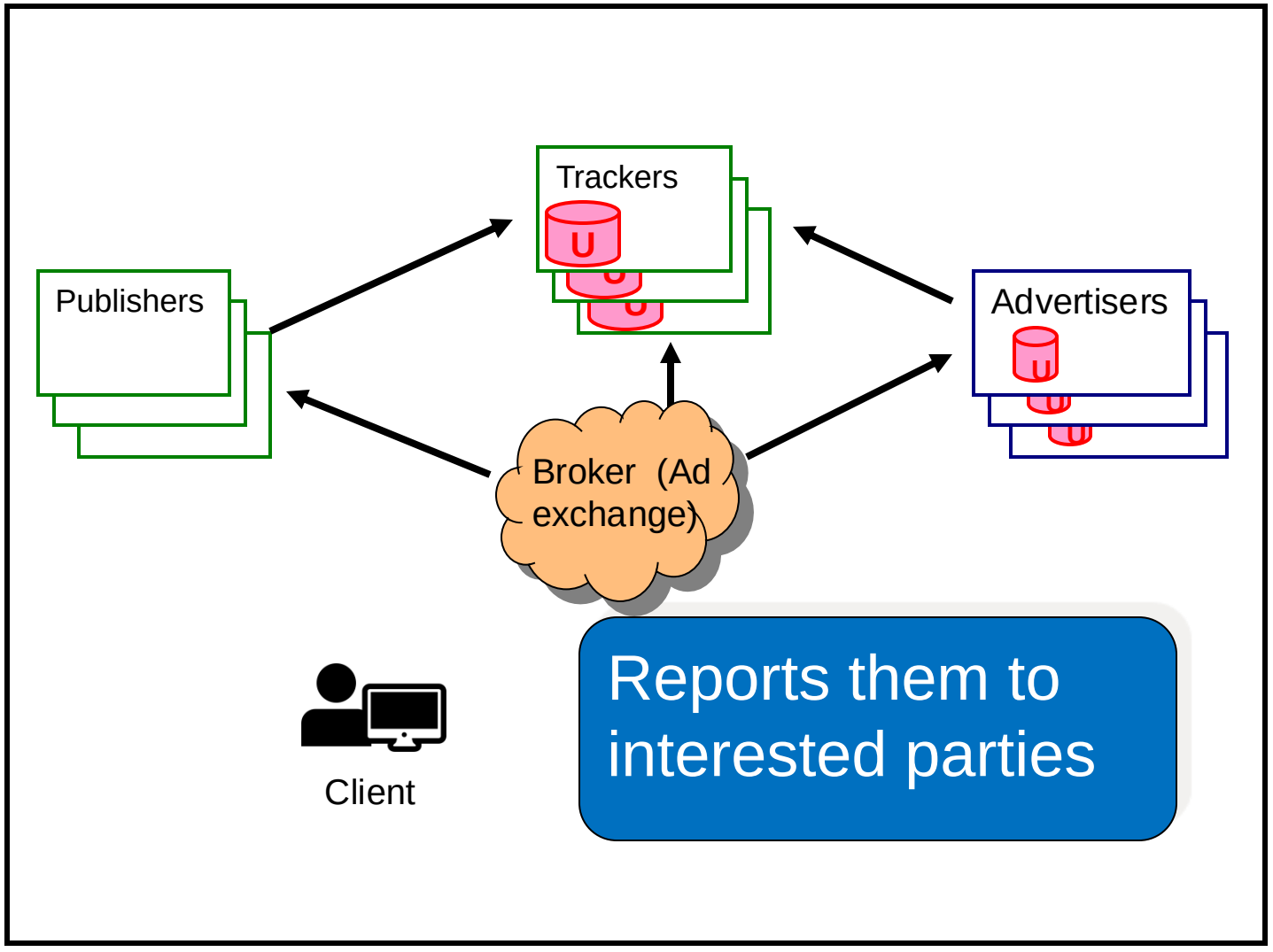

Trackers
Publishers
Advertisers
Broker (Ad exchange)
Client tells broker of page
Client


Trackers
Publishers
Advertisers
Broker (Ad exchange)
Client
Broker launches auction (for given user visiting given webpage ….)
Also does clickfraud etc.


Trackers
Publishers
Advertisers
Broker (Ad exchange)
Client
Advertisers present bids and ads


Trackers
Publishers
Advertisers
Broker (Ad exchange)
Client
Broker picks winners, delivers ads


Trackers
Publishers
Advertisers
Broker (Ad exchange)
Client
Broker waits for results (view/click)


Trackers
Publishers
Advertisers
Broker (Ad exchange)
Client
Reports them to interested parties

“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 ($(G_n \times U_n) / (G_c \times U_c)$)
  - B_c , G_c , U_c are for the current (winning/highest) bid
  - B_n , G_n , U_n are for the next highest bid
  - CPC must <= B_c

Quality Scores

- Historical click performance of the ad
- Landing page quality
- …….

Known at broker (call it G)

- Relevance to the user
- User click through rates
- …….

Known at user (call it U)

Is OBA a Violation of Privacy?

- Linking behavioral data with unique identifiers
- Detail and scope of data collection
- Sensitivity of data
- Impact on operability
- Difficult to opt 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

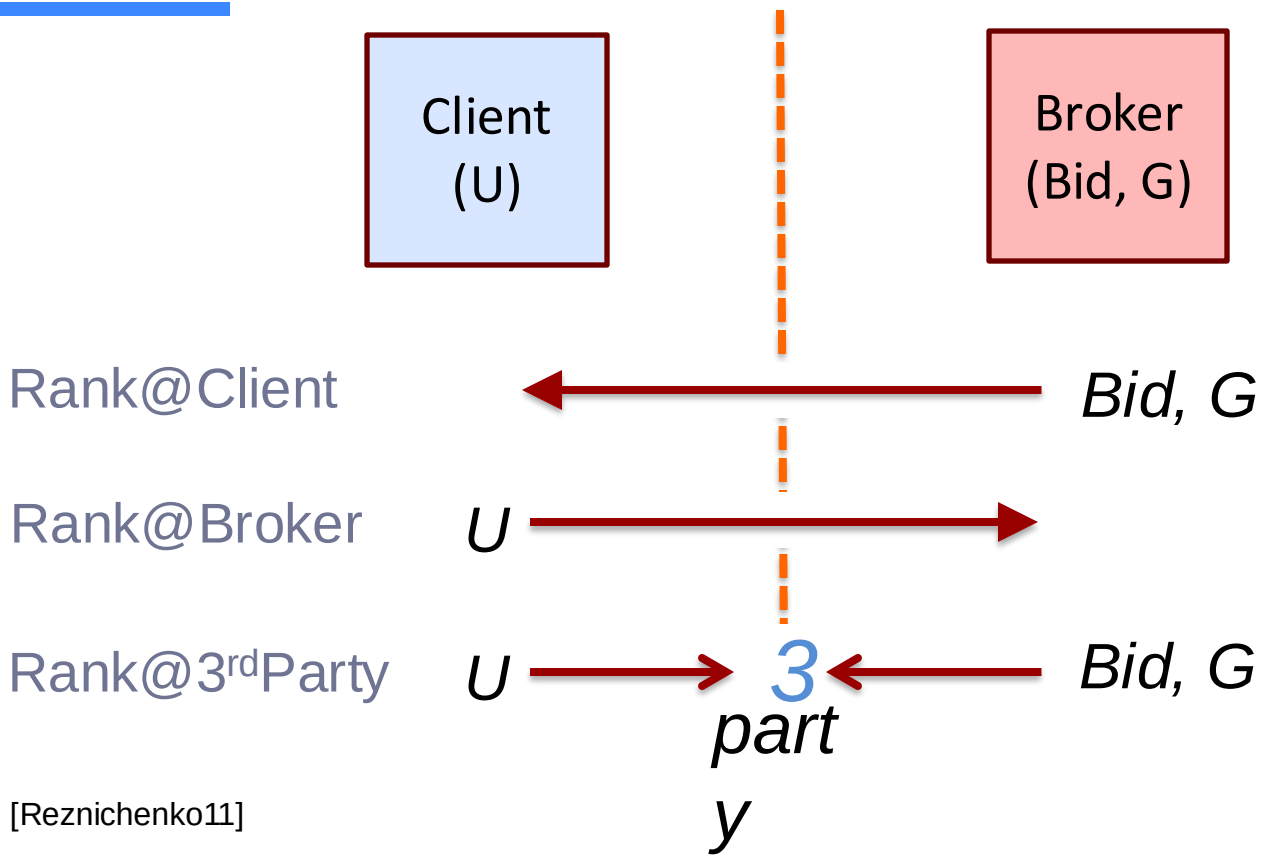
Advertisers, publishers, and brokers have their own concerns.

- 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

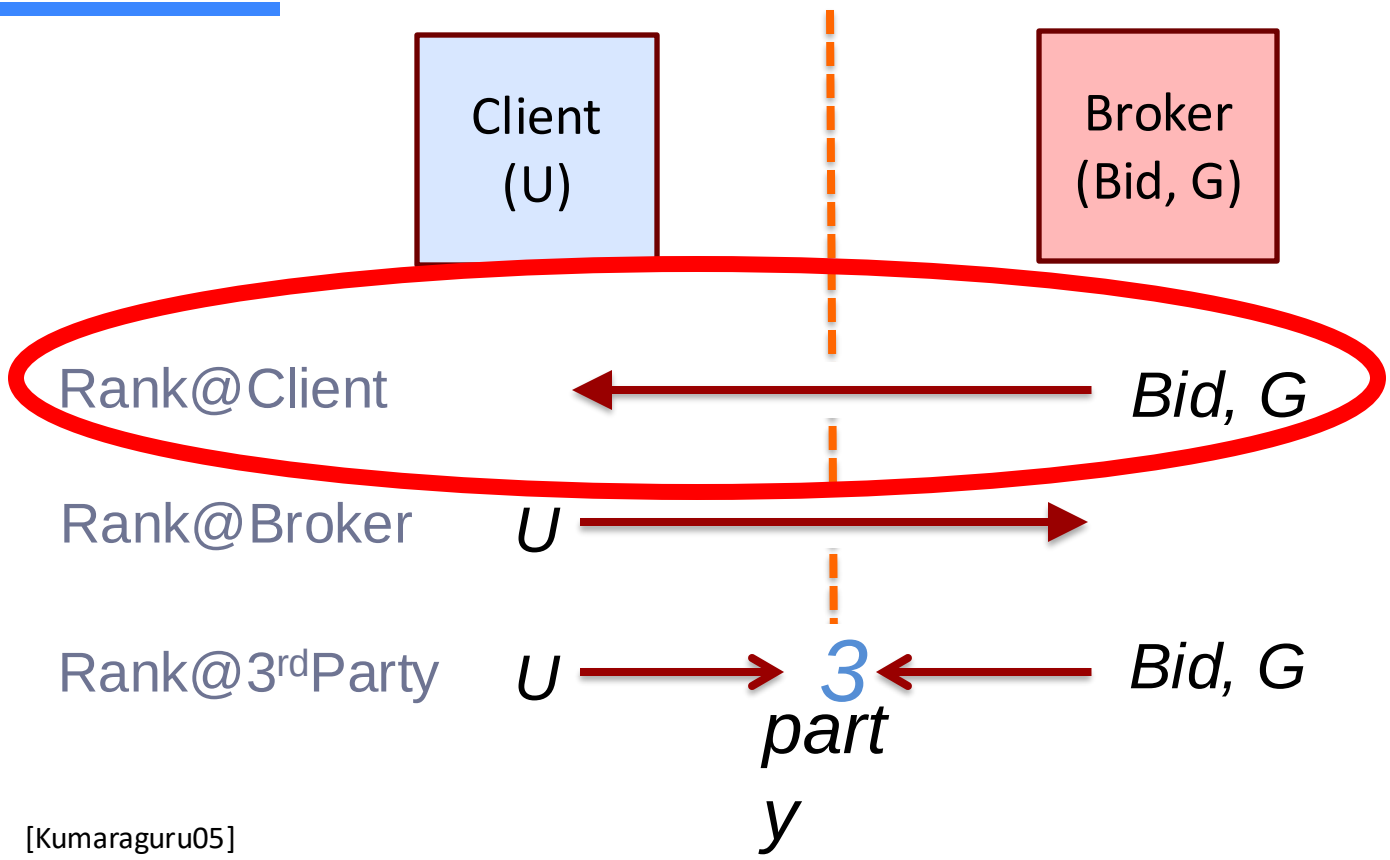
- Encourage trust in the data server (“don’t be evil”)
  - Current situation, but there are concerns …
- Make decisions on the client device so private information is never shared with the data server
  - Don’t send U to the server
- Transform data from the client so that privacy risks of the transformed 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

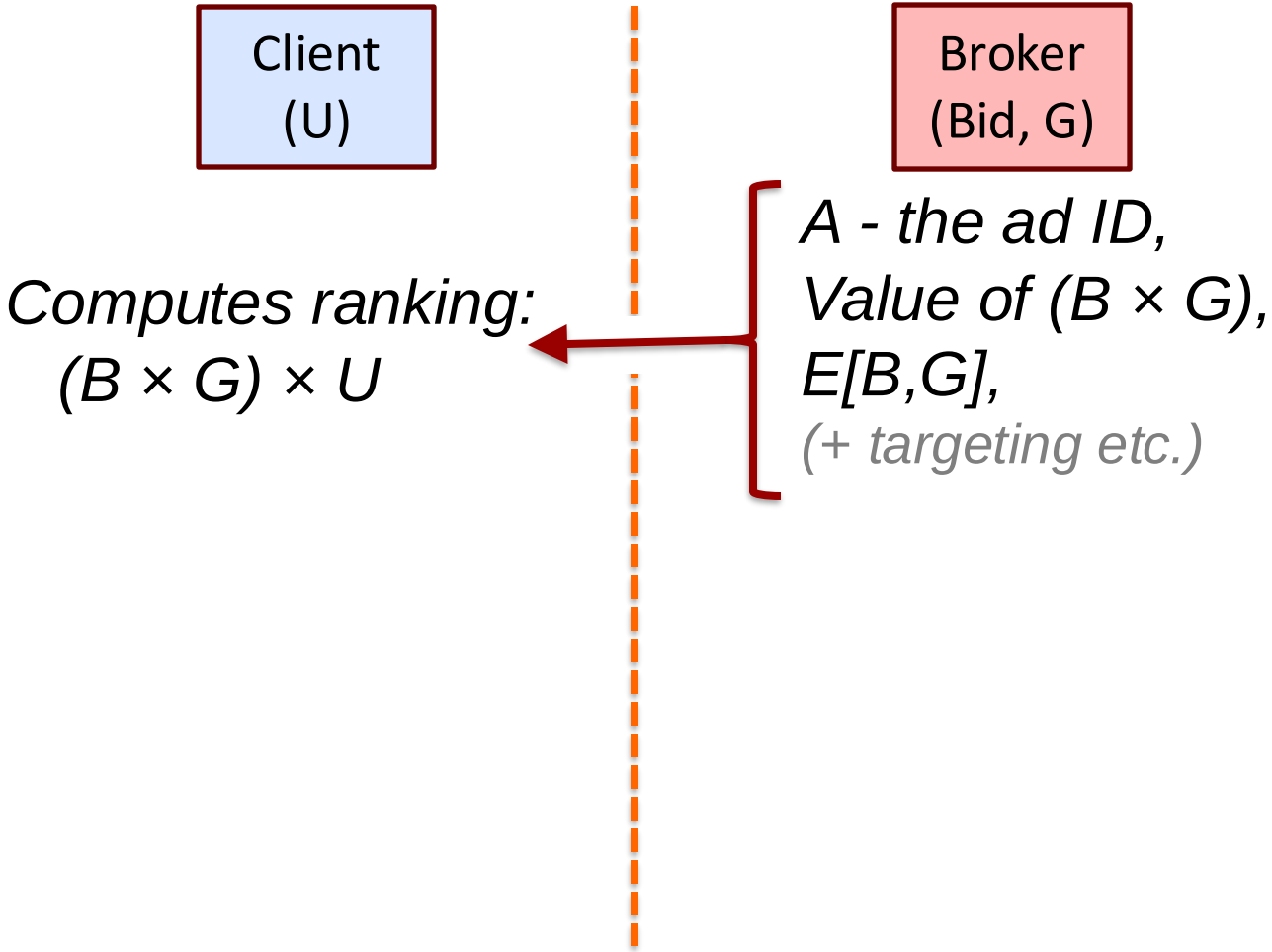


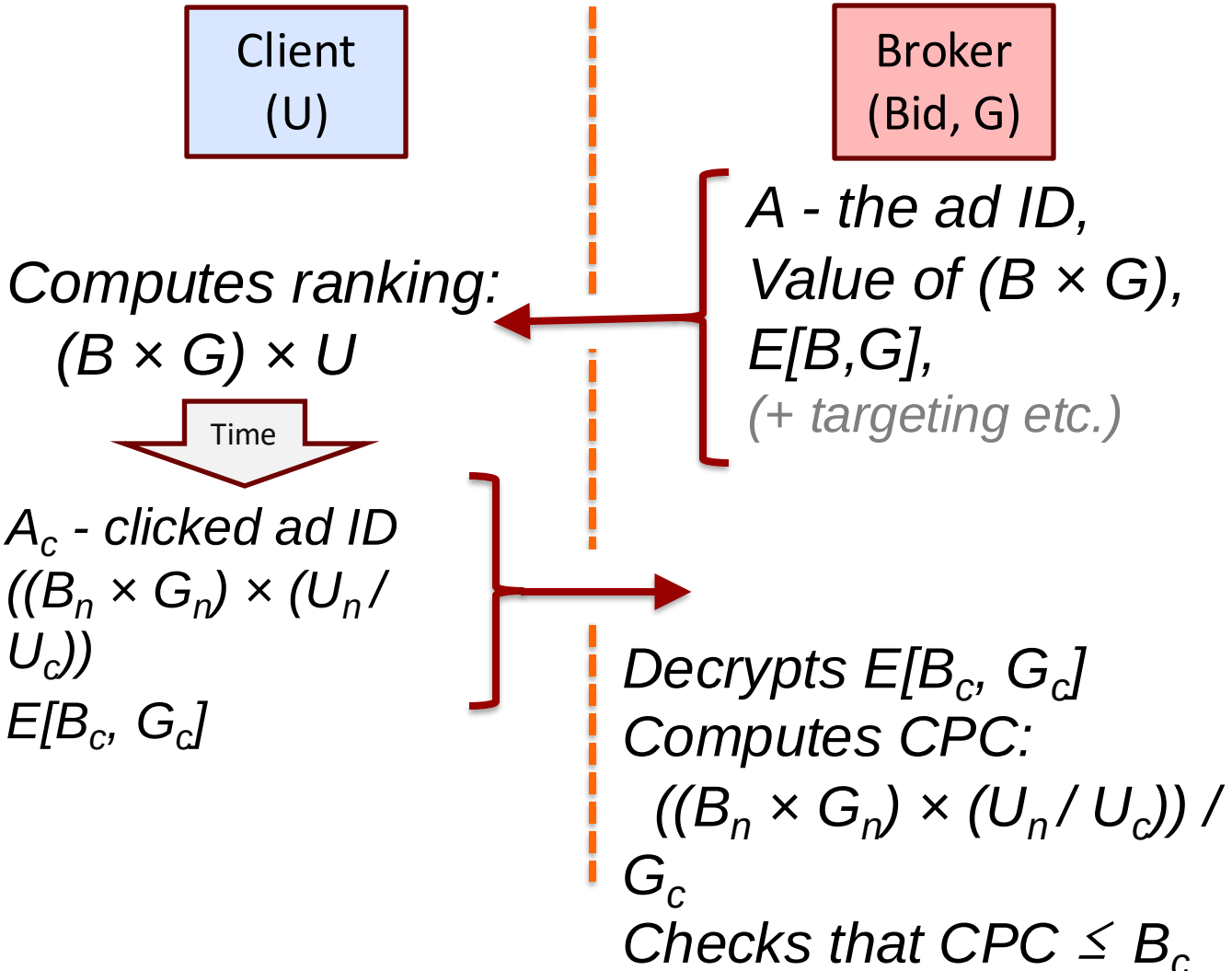
[Reznichenko11]

Three flavors of Non-Tracking auctions



[Kumaraguru05]


Client
(U)
Broker
(Bid, G)
A - the ad ID,
Value of (B × G),
E[B,G],
(+ targeting etc.)
Computes ranking:
(B × G) × U

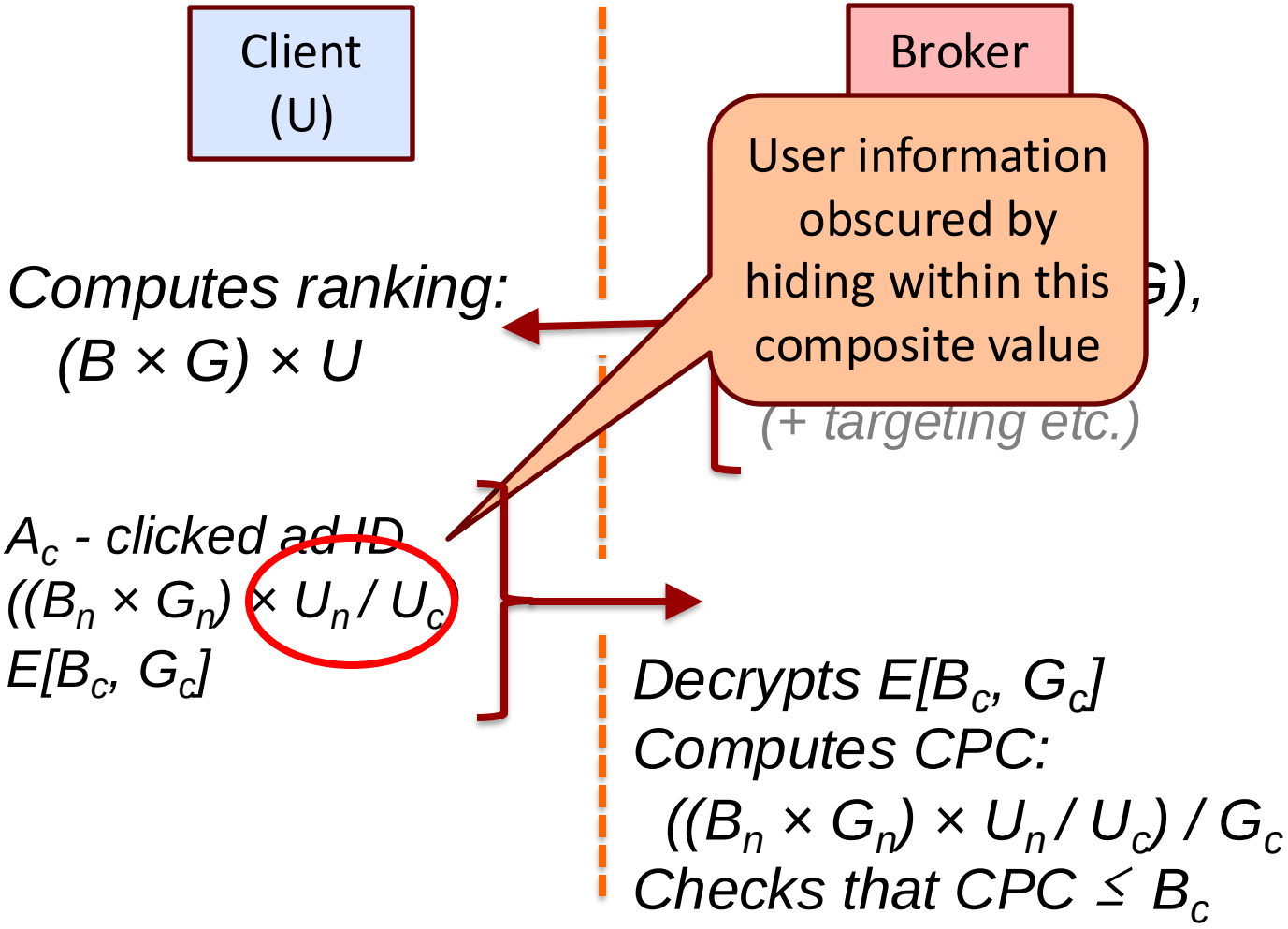

Client
(U)
Broker
(Bid, G)
A - the ad ID,
Value of (B × G),
E[B,G],
(+ targeting etc.)
Computes ranking:
(B × G) × U
Time
A_c - clicked ad ID
((B_n × G_n) × (U_n / U_c))
E[B_c, G_c]
Decrypts E[B_c, G_c]
Computes CPC:
((B_n × G_n) × (U_n / U_c)) / G_c
Checks that CPC ≤ B_c

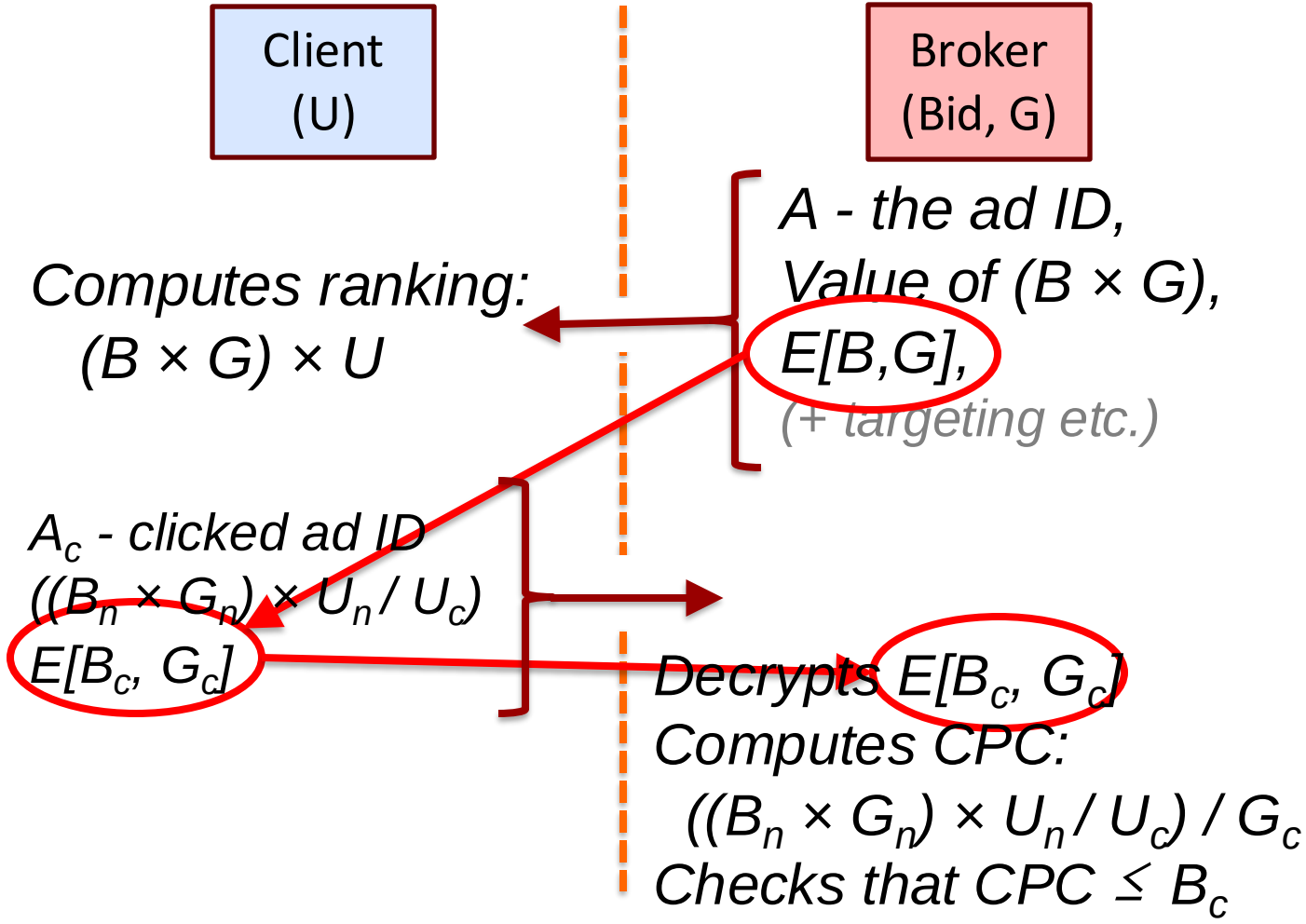
Client
(U)

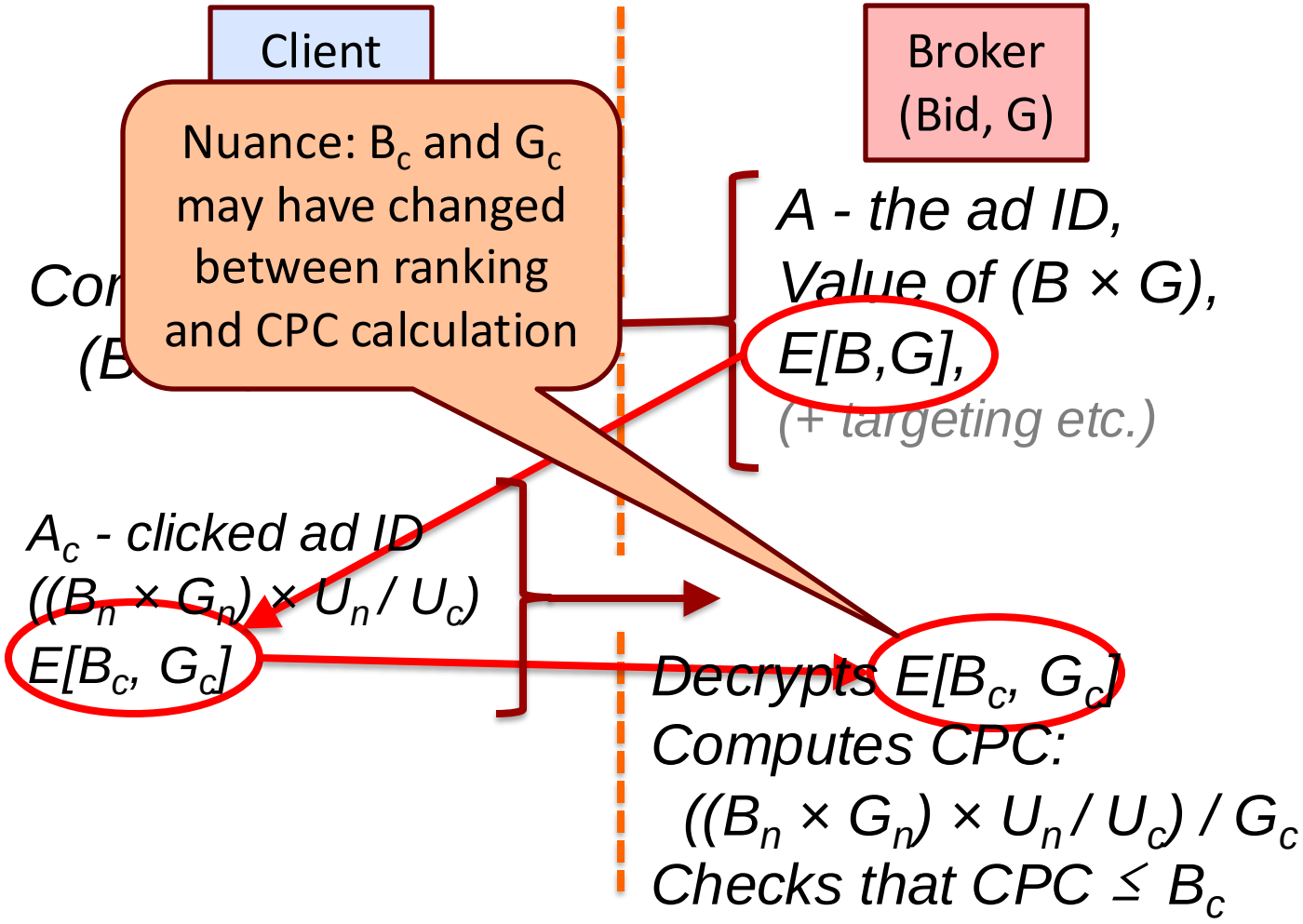
Broker
(Bid, G)

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

Decrypts $E[B_c, G_c]$
Computes CPC:
$((B_n \times G_n) \times U_n / U_c) / G_c$
Checks that CPC ≤ B_c


Client
(U)
Broker
User information obscured by hiding within this composite value
Computes ranking:
(B × G) × U
(+ targeting etc.)
A_c - clicked ad ID
((B_n × G_n) × U_n / U_c)
E[B_c, G_c]
Decrypts E[B_c, G_c]
Computes CPC:
((B_n × G_n) × U_n / U_c) / G_c
Checks that CPC ≤ B_c


Client
(U)
Broker
(Bid, G)
A - the ad ID,
Value of (B × G),
E[B,G],
(+ targeting etc.)
Computes ranking:
(B × G) × U
A_c - clicked ad ID
((B_n × G_n) × U_n / U_c)
E[B_c, G_c]
Decrypts E[B_c, G_c]
Computes CPC:
((B_n × G_n) × U_n / U_c) / G_c
Checks that CPC ≤ B_c


Client
Broker
(Bid, G)
Nuance: B_c and G_c may have changed between ranking and CPC calculation
A - the ad ID,
Value of (B × G),
E[B,G],
(+ targeting etc.)
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 ≤ B_c

Privad: a privacy preserving ad system

- 13K opted-in users in Sep. and Oct. 2013
- ~4800 active users per day
- Only in October
  - ad requests 1.1M
  - ads 9.5M
  - ad views 790K
  - ad clicks 417
  - product purchases 4

Private-by-Design Advertising Meets the Real World, Alexey Reznichenko, Paul Francis, CCS, 2014

Privad: a privacy preserving ad system

- CTR (click-through-rate) 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

- [Kumaraguru05] Privacy Indexes: A Survey of Westin's Studies, Ponnurangam Kumaraguru and Lorrie Faith Cranor. Technical Report CMU-ISRI-5-138, 2005.
- [Acquisti15] Privacy and human behavior in the age of information. Acquisti, Alessandro, Laura Brandimarte, and George Loewenstein. *Science* 347.6221 (2015): 509-514.
- [Reznichenko11] Auctions in Do-Not-Track Compliant Internet Advertising, Alexey Reznichenko, Saikat Guha, and Paul Francis. CCS 2011.

Discussion Questions

- What kind of privacy person are you? Why? Can you think of a case were 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?