ADMAP: Attribution Data Matching Protocol

A Data Clean Room Interoperability Protocol for Attribution Measurement

Version 1.0

Please email support@iabtechlab.com for public comments and questions. This document is available online at *https://iabtechlab.com/admap*

About This Document

This document describes a standard, the *Attribution Data Matching Protocol (ADMAP)* specification for a well-defined use case to support interoperability for Data Clean Room (DCR) Providers and their clients. The well-defined use case is that an advertiser wants to measure and compare the performance of their campaigns across various publishers, ad networks, channels, and platforms. We recommend the "Data Clean Rooms Guidance and Recommended Practices" document as a preread to become familiar with DCRs and their functions and better understand the context of this document.

This document describes the specification for implementing a matching operation between parties and the supporting mechanisms to use the output of the operation to attribute and measure the matched events. The standard will enable Data Clean Room (DCR) Providers to implement well-defined, consistent, and reliable mechanisms to support their customers and enable advertisers and publishers to interoperate with different DCR Providers and business partners.

This document is primarily intended for a technical audience, in particular for engineers and product managers working with first-party data and interested in implementing the mechanisms described herein. The key takeaways for readers are:

- Understand the privacy and security goals in a DCR specific to two-party matching.
- Understand how to support attribution measurement that meets privacy goals through the end use of the outputs.
- How to structure and format the inputs for mapping and attribution operations and how to deploy the outputs.
- Understand potential threat vectors and collusion scenarios that can compromise privacy and security goals.

This document is developed by the IAB Tech Lab Rearc Addressability Working Group. This is the second in a series of DCR interoperability standards. IAB Tech Lab will develop specifications for other well-defined advertising use cases for DCRs in the future.

Note: *The use of words or phrases 'Privacy", "Private", "Security", "Control", "Processing", "Personal Data", "PII" in this document is generic and does not refer to definitions in any specific regulation e.g. GDPR or CCPA.*

License

Significant Contributors

Edik Mitelman, *AppsFlyer;* Tal Inbar, *AppsFlyer*; Andrew Knox, *Decentriq*; Andrei Lapets, *Magnite*, Frederick Jansen, *Magnite*; Brian May, *Dstillery*; Devon DeBlasio, *InfoSum*

IAB Tech Lab Lead

Miguel Morales, Director Addressability & Privacy Enhancing Technologies (PETs)
Shailley Singh, EVP Product & COO, IAB Tech Lab

About IAB Tech Lab

The IAB Technology Laboratory is a nonprofit research and development consortium charged with producing and helping companies implement global industry technical standards and solutions. The goal of the Tech Lab is to reduce friction associated with the digital advertising and marketing supply chain while contributing to the safe growth of an industry.

The IAB Tech Lab spearheads the development of technical standards, creates and maintains a code library to assist in rapid, cost-effective implementation of IAB standards, and establishes a test platform for companies to evaluate the compatibility of their technology solutions with IAB standards, which for 18 years have been the foundation for interoperability and profitable growth in the digital advertising supply chain. Further details about the IAB Technology Lab can be found at https://iabtechlab.com.

Disclaimer

THE STANDARDS, THE SPECIFICATIONS, THE MEASUREMENT GUIDELINES, AND ANY OTHER MATERIALS OR SERVICES PROVIDED TO OR USED BY YOU HEREUNDER (THE "PRODUCTS AND SERVICES") ARE PROVIDED "AS IS" AND "AS AVAILABLE," AND IAB TECHNOLOGY LABORATORY, INC. ("TECH LAB") MAKES NO WARRANTY WITH RESPECT TO THE SAME AND HEREBY DISCLAIMS ANY AND ALL EXPRESS, IMPLIED, OR STATUTORY WARRANTIES, INCLUDING, WITHOUT LIMITATION, ANY WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE, AVAILABILITY, ERROR-FREE OR UNINTERRUPTED OPERATION, AND ANY WARRANTIES ARISING FROM A COURSE OF DEALING, COURSE OF PERFORMANCE, OR USAGE OF TRADE. TO THE EXTENT THAT TECH LAB MAY NOT AS A MATTER OF APPLICABLE LAW DISCLAIM ANY IMPLIED WARRANTY, THE SCOPE AND DURATION OF SUCH WARRANTY WILL BE THE MINIMUM PERMITTED UNDER SUCH LAW. THE PRODUCTS AND SERVICES DO NOT CONSTITUTE BUSINESS OR LEGAL ADVICE. TECH LAB DOES NOT WARRANT THAT THE PRODUCTS AND SERVICES PROVIDED TO OR USED BY YOU HEREUNDER SHALL CAUSE YOU AND/OR YOUR PRODUCTS OR SERVICES TO BE IN COMPLIANCE WITH ANY APPLICABLE LAWS, REGULATIONS, OR SELF-REGULATORY FRAMEWORKS, AND YOU ARE SOLELY RESPONSIBLE FOR COMPLIANCE WITH THE SAME, INCLUDING, BUT NOT LIMITED TO, DATA PROTECTION LAWS, SUCH AS THE PERSONAL INFORMATION PROTECTION AND ELECTRONIC DOCUMENTS ACT (CANADA), THE DATA PROTECTION DIRECTIVE (EU), THE E-PRIVACY DIRECTIVE (EU), THE GENERAL DATA PROTECTION REGULATION (EU), AND THE E-PRIVACY REGULATION (EU) AS AND WHEN THEY BECOME EFFECTIVE.

Glossary

Term	Description
Attribution	The process of identifying and assigning credit to the specific exposure that contributed to a conversion.
Audience	Group of people with a common set of characteristics whom an advertiser wants to show an ad. More specifically this is a list or group of customers or individuals that is most likely to purchase a given product or service from an advertiser.
Blinding	A technique used to enhance privacy and security during the process of encryption or signing by obscuring the data being processed. The primary goal of blinding is to prevent unintended disclosure of sensitive information and to ensure that the party performing the operation cannot see the actual data involved.
Collusion	A scenario in which two or more parties involved in an operation or protocol are sharing information with each other in ways that contravene established standards.
Conversion	When a user completes an action on an advertiser's digital property in response to an ad shown on a publisher's platform.
Data Clean Room (DCR)	A secure, centralized repository where advertisers and publishers can share sensitive data, restricting raw access and offering only limited, aggregated insights into the stored information.
Encoding	The process of transforming data into a specific format to ensure its confidentiality, integrity, and security during storage or transmission
Engagement	Any interaction a user has with an ad beyond just viewing it, indicating that the user has actively responded and shown interest or intent.
Ephemeral email	A temporary email address that is created for a short period of time

Term	Description
	and is designed for single-use or limited-use scenarios. These email addresses allow users to receive messages without revealing their permanent email addresses, enhancing privacy and security.
Exposure	An instance when an ad is displayed on a user's screen, regardless of whether the user interacts with it.
First-party Data Sets	Data acquired by an organization as a result of an individual's interaction with the organization either online on their website or mobile app or connected device or offline in their physical locations or by mail or phone.
ID Resolution	A service that matches multiple identifiers across various systems and devices to provide a unified view of an entity using sophisticated algorithms and diverse data sources.
Identity Partner or Provider	An organization that maintains an individual, household or device level unique identification that can be used to perform a match between two or more organizations' data sets.
Impression	A single instance of an ad creative being displayed on a user's screen.
Noise	Random data that is added to the output of a query or computation to obscure the influence of individual data entries.
Normalization	The process of standardizing and organizing data from different origins into a consistent format or structure. This is crucial when integrating data from various systems, databases, or applications to ensure that the data can be effectively analyzed and utilized.
Personally Identifying Information (PII)	Any data that can, independently or in combination, be used to identify a person, either directly or indirectly.
PETs	Privacy enhancing technologies (PETs) are technology solutions that use one or more of the privacy technologies (differential privacy, secure multi party compute and on device learning) to accomplish

Term	Description
	complex data processing functions in digital advertising without revealing the individual, household or device level personal information to parties that do not already have them.
Private Set Intersection (PSI)	A secure, multi-party computation, cryptographic technique that allows two parties holding sets of data to compare encrypted versions of these data sets in order to determine the intersection, while not revealing what data does not overlap.
Rounding	The process of adjusting a numerical value to a nearby, often simpler, or more convenient value, usually by increasing or decreasing it to the nearest specified unit. Rounding helps make query results more interpretable while maintaining privacy by obscuring exact values.
SHA256 hash	A 256 bit hash value generated from a given input value. The same input will always result in the same hash and the resulting hash cannot be used to directly recover the original value. There is also a high probability of the hash being unique for a given set of inputs.
Third party	A party to an interaction that has no direct relationship with the individual involved.
Thresholding	A technique used to ensure that the output of a query on a dataset does not reveal too much information about any individual entry. It aims to provide strong privacy guarantees when releasing aggregated data or statistics, ensuring that the inclusion or exclusion of a single individual's data does not significantly affect the outcome of the query.
Trusted Execution Environment (TEE)	A secure compute environment that provides guarantees about the consistency of operations and data security and is trusted not to allow for information leakage.

Table of Contents

Overview

Attribution Data Matching Protocol (ADMAP) is a privacy-centric Data Clean Room (DCR) protocol that enables advertisers and publishers to collaborate and generate attribution reporting by safely sharing user conversions and engagements respectively, without disclosing their user's personal information or data.

Document Organization

The remainder of this document is organized in four parts:

1. The **Technical Requirements** section describes the privacy and security goals of ADMAP. It provides a blueprint architecture describing the participants, as well as the input and output requirements of the different components of the protocols.

2. The **Mapping Protocol** section presents the high-level requirements that a mapping component implementation must satisfy. The purpose of the mapping component is to identify and define common identifiers between advertiser and publisher data to ensure that both can deterministically map their respective first-party data to common match keys.

3. The **Attribution Protocol** sections present the data formats, architecture, and workflows that constitute an attribution protocol, focusing on overall design, security features, input and output formats, and intended usage. The attribution component is responsible for computing the attribution based on the input events and mapping component identifiers. It also includes aggregation and reporting components for generating the privacy safe output from the computed attribution measurements.

4. The **Reference Implementations** section presents example implementations of protocols and protocol components. This includes two reference implementations: one that leverages private set intersection (PSI) and one that leverages a trusted execution environment (TEE).

5. The **Collusions and Threats** section provides threat vectors that must be considered by any component designs adhering to this specification.

Technical Requirements

ADMAP is a DCR protocol for computing the intersection of user records within datasets provided by advertisers and publishers for the purpose of attributing user engagements with advertisements on a publisher's digital media properties (e.g. websites, mobile app or CTV app) to user conversions on advertiser's digital media properties (e.g. website or an app). The two parties are typically an advertiser and a publisher/ad network or their delegated vendors and the protocol is implemented by their designated DCRs.

Privacy and Security Design Goals

In this section, we describe the design goals related to the transfer of information between the participants (advertisers and publishers) involved in the protocol. Solutions based on the protocol must document how they achieve these privacy and security design goals.

Given a list of users with PII known to an advertiser, and a separate list of users with PII known to a publisher, a protocol solution maps the two lists to a common identifier space. This common identifier space will then allow impression event data to be joined with associated conversion event data to ultimately generate attribution reports.

Design Goal 1 - Security of PII
The solution protects the end user's PII data throughout the operation using encryption. This means that participants with whom the *end-user* **has not shared their PII directly** should not be able to learn any end user's PII.

Design Goal 2 - Privacy of User Identity
The solution prevents each participant from learning the identity of end users that are **not** part of their own contributed input data set.

Design Goal 3 - Privacy of Group Membership
The solution prevents each participant from learning which end users they contributed are members in the computed overlap.

Protocol Architecture and Participants

The participants, their roles and interactions, and the overall attribution methodology must be agreed upon by both the advertiser and publishers. While we specifically name "publishers" and "advertisers", in practice it is common for designated vendors to

participate on behalf of one or both of the principal participants. For instance, an advertiser may delegate the responsibility to a measurement provider, media agency, demand side platform, customer data platform, DCR provider or various combinations of the above providers. While this may have operational implications, it does not affect the protocol meaningfully: vendors supporting the advertiser are considered part of the advertiser from the protocol's point of view, and similarly, vendors supporting publishers are considered part of the respective publishers.

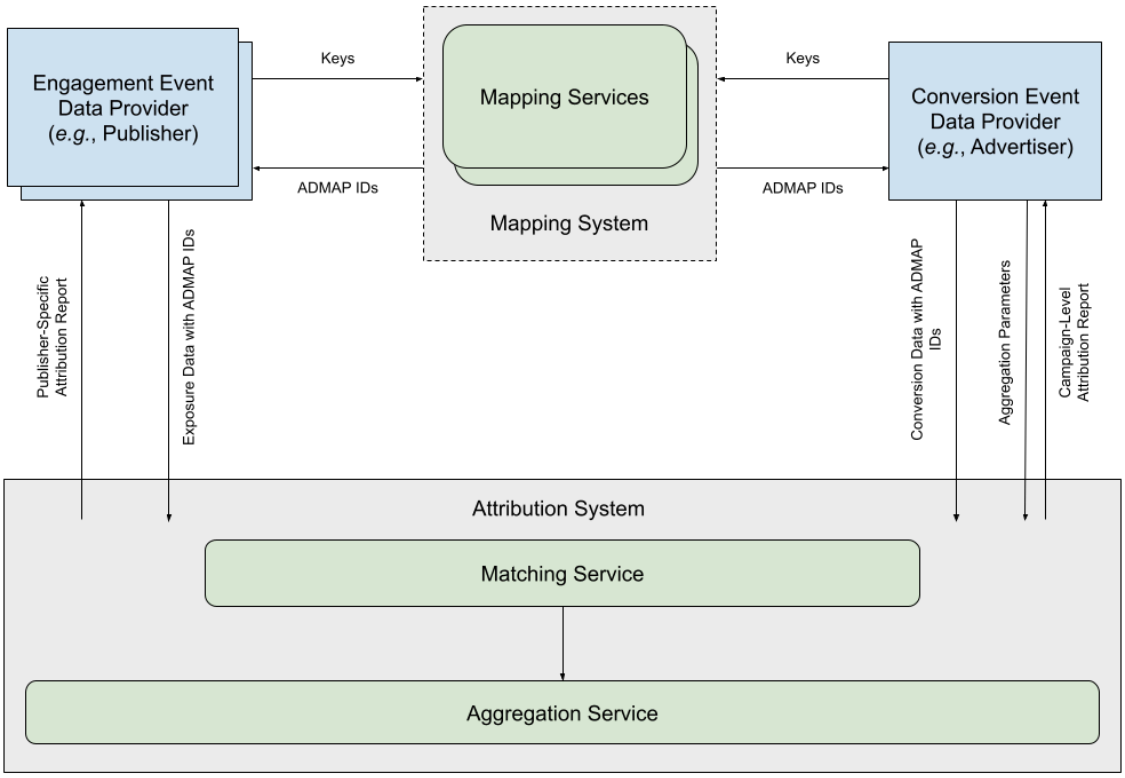


Figure 1: Blueprint of the architecture, depicting principal data flows.

In order to describe the participants in ADMaP, we refer to the blueprint of the architecture in Figure 1.

Advertiser

The advertiser is the entity that wants to attribute their conversions to a specific network identified by common identifiers (e.g. identity providers, universal ids) or PII records (e.g. IP address, email addresses, phone numbers). The list of users may be, for example, the advertiser's new or existing customers, and the identifiers PII records may have been obtained through either online or offline means.

The advertiser may be the advertiser organization itself, or a delegated organization acting on behalf of the advertiser, such as a technology vendor. Possible types of

vendors here may include Data Collaboration Platforms (DCP), Attribution Measurement Platforms, Mobile Measurement Platforms (MMP), Data Management Platforms (DMPs), Customer Data Platforms (CDPs), etc. For the purposes of this proposal, we shall not distinguish between various types of delegated vendors, since they are trusted by and are at the discretion of the advertiser.

In this protocol, we consider the specific scenario where the advertiser wants to attribute and measure conversions of users to interactions with digital media properties controlled by a publisher.

Publisher

The publisher is the entity that has an user audience, some percentage of which may be linked to persistent user identifiers such as email addresses or phone numbers. The publisher controls digital media properties (e.g. websites, mobile and CTV applications) which support digital advertisements.

The publisher wants to enable an advertiser to attribute the success of their advertising campaigns and measure the performance on the publisher's digital media properties for users overlapping with the advertiser's list.

The publisher supports, in various ways, advertisers' efforts to gather data for, and measure the performance of, campaigns that include the publisher's digital media properties; this support is generally focused on members of the publisher's audience who are also potential customers of the advertiser.

The publisher may be the digital media property owner itself, or a technology vendor acting on behalf of the digital media property owner. The list of vendors media owners engage with is similar to those that an advertiser may use. We assume that the advertiser and publisher, if they are delegating aspects of the process described here, would be working with different vendors and that no single participant would have direct access to PII records from both the advertiser and publisher involved in the protocol. In this protocol we do not distinguish between various types of delegated vendors, since they are trusted by and are at the discretion of the publisher.

Mapping System Operator

It is possible for advertisers and engagement data providers to have a pre-existing agreement for mapping data, such as using an email address. If not, a DCR or identity solution provider may operate a mapping system (readers may wish to refer to the IAB Tech Lab's Identity Solutions Guidance). The mapping system operator may enforce

contractual obligations to only decrypt a specific subset of the data relevant to the measurement objectives.

Attribution System Operator(s)

Some architectures enabling the described process, such as the one depicted in Figure 1, could require or benefit from the help of a third-party system for matching two or more data sets that have been mapped in a compatible way (whether via a mapping system operator or some other means). Where a third-party attribution-stage matching system is involved, we must consider the third-party entity operating that system and its relationship with the other participants involved in enabling ADMAP. Solution designers must also consider the privacy and security design goals as they relate to such a third-party operator.

End User

While not pictured in the blueprint architecture shown in Figure 1, the end user is the consumer or user of the publisher's digital property, a person that owns the PII record (e.g. email address) that it has voluntarily and separately shared directly, with both the advertiser and the publisher, and that accesses the publisher's controlled media properties where advertisements are displayed.

Mapping Protocol

To perform a match between two data sets, it is necessary to have a common identity or match key. Mapping protocol is a process that creates a common identity space between the advertiser and the publisher data sets. The mapping system can also integrate with ID resolution vendors to help create a common ID link between advertiser and publisher.

A mapping workflow may not be required between an advertiser and publisher in some cases, such as when they

1) already share a common ID space for their respective data sets via an existing ID solution,
2) have already determined a mapping within a common DCR, or
3) have already determined a mapping for their respective data sets in the clear by sharing data directly.[1]

In any case, we describe a mapping protocol to help advertiser and publisher arrive at a common shared id space and also protect the personal information being directly used in matching and attribution processes there by protecting the user's PII from being directly exposed to different systems.

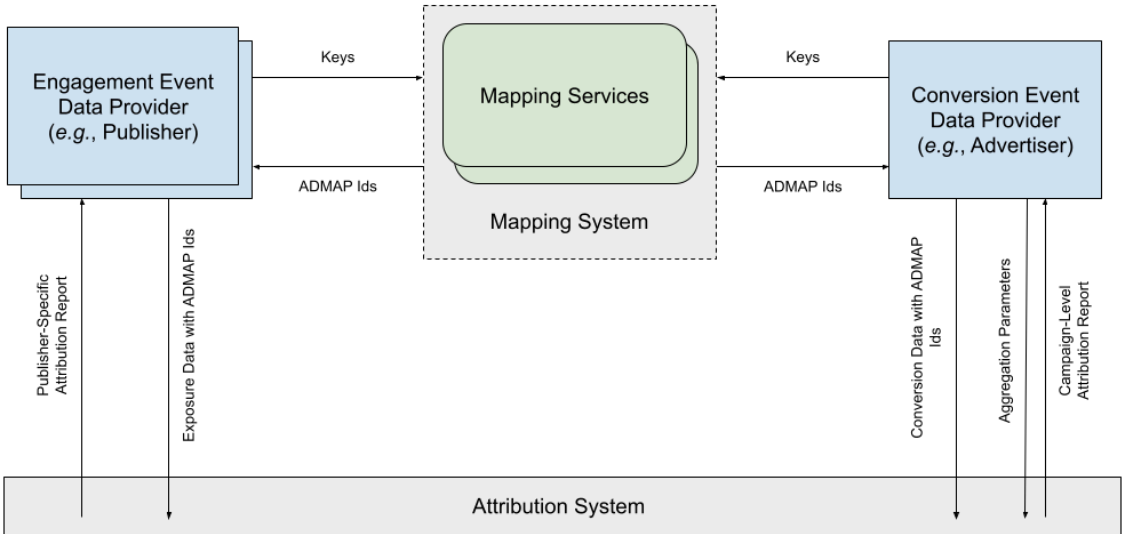


Figure 2: Blueprint of Mapping System and associated data flows, depicting principal data flows.

[1] In the case of (3), an advertiser and publisher may be comfortable sharing first-party data but not transaction data, in which case the attribution workflow is still required.

Mapping Protocol Inputs

We specify the types of inputs that must be provided by both the advertiser and publisher to the mapping workflow.

- **Key Type:** The type of key value being provided such as email address or phone number. Participants can agree on additional key types, and we may standardize additional key types in a future revision of this document.
- **Key Value:** The advertiser and publisher must each prepare a list of keys. The key lists could be ordered, as required by the mapping implementation.

  The type of each key may consist of personally identifiable information (PII), such as for example an email address, and the encoding of such keys – which is not explicitly specified by this protocol – must be agreed to by both parties ahead of time. We specify two types of standard PII keys and their expected normalization and encoding in Table 1 below.

Key Type	Normalization & Encoding	Example
Email address	(i) Leading and trailing spaces trimmed (ii) ASCII characters converted to lowercase (iii) SHA256 hashed (iv) No hashing salt	b4c9a289323b21a01c3e940f150eb9b8c542587f1abfd8f0e1cc1ffc5e475514
Phone number	(i) E.164 normalized (maximum of 15 digits) (ii) No spaces, hyphens, parentheses, or other special characters (iii) SHA256 hashed (iv) No hashing salt	c1d3756a586b6f0d419b3e3d1b328674fbc6c4b842367ee7ded780390fc548ae

Table 1: Key Types and Encodings

It is important to note that the mapping workflow must specify additional security measures on the input data, such as salting, encryption, *etc.* The normalization and encoding specified in Table 1 are therefore not designed to provide security, but rather to enable a consistent means for mapping keys belonging to different parties.

We recognize that an increasing number of users are moving to ephemeral email addresses for specific purposes, and that technology platforms are adding features to make this easier. In Table 1 we propose a default normalization and encoding scheme for the email address key type, but participants are welcome to agree on additional normalization rules, as long as they are the same across all participants for mapping purposes.

Mapping Protocol Outputs

The mapping solution must generate the following outputs **for both the advertiser and the publisher**:

- **Space ID:** a randomly assigned value which is used to create a common compartmentalized grouping space between the advertiser and the publisher.
    - The returned Space ID is used to verify that the key values being used in the downstream Attribution Protocol are matched against other key values generated from the same grouping space.
    - It should be used by the advertiser and publisher to reference the grouping space when configuring an ad campaign in ad attribution systems (such as when working with DSPs or SSPs). Space ID is only generated once per mapping job and can be the same across multiple mapping jobs.
    - Space ID may optionally be manually generated by the advertiser in coordination with the publisher.
    - Note that for mapping systems that allow updates/deltas, it is generally acceptable for the mapping to change over time (especially additions to the mapping as deltas). Currently, the protocol does not include “version” of a mapping, since both sides will generally want to just always want to use the latest version available to them at the moment. If they agree that they do want to be precise about a specific version/cut of a mapping, they can handle this by assigning different Space IDs to different versions of the same mapping, even though those different versions exist in the same mapping space.

- **Key Value:** the original encoded user key value from the participant's input data set.
- **ADMAP ID:** the user id generated by the mapping system which links the advertiser's user to the publisher's user. It is a randomly generated unique number for each row of advertiser and publisher dataset. When there is a match for provided PII, the ADMAP ID is the same for both publisher and advertiser matching rows.
    - There should always be a 1:1 relationship between each key and an ADMAP ID (i.e. each value appears exactly once in the entire file). Each participant receives back one row for each row submitted. For each participant, their original key value is listed in the first column, the ADMAP ID is listed in the second column.
    - In the case that no link between advertiser and publisher was found for a row an ADMAP ID should be randomly generated so that audience membership information is not leaked.

A table like the one in the examples below will be generated for each publisher providing input records to the advertiser and, though advertiser key values must only be included once in any given table, the same advertiser key value may be included in multiple tables.

Examples Input and Output

Simple Scenario

Example advertiser input:

Key Type	Key Value
email	hash(jane@example.com)

Example publisher input:

Key Type	Key Value
email	hash(jane@example.com)

Example advertiser output:

Space ID: 123

Key Value	ADMAP ID
hash(jane@example.com)	043e3ef5e76e423c9d3d3a5605bcb8904496c65

	4fcd16180386298150d8e1b8f

Example publisher output:

Space ID: 123

Key Value	ADMAP ID
hash(jane@example.com)	043e3ef5e76e423c9d3d3a5605bcb8904496c65 4fcd16180386298150d8e1b8f

Complex Scenario - Using ID Resolution Provider

Example advertiser input:

Key Type	Key Value
email	hash(jane@example.com)

Example publisher input:

Key Type	Key Value
email	hash(jane@email.com)

Example advertiser output:

Space ID: 123

Key Value	ADMAP ID
hash(jane@example.com)	043e3ef5e76e423c9d3d3a5605bcb8904496c65 4fcd16180386298150d8e1b8f

Example publisher output:

Space ID: 123

Key Value	ADMAP ID
hash(jane@email.com)	043e3ef5e76e423c9d3d3a5605bcb8904496c65 4fcd16180386298150d8e1b8f

Attribution Protocol Inputs and Outputs

The attribution protocol is the process of ingesting exposure events from publishers along with conversion events from advertisers and computing attributions.

Attribution Protocol Inputs

The attribution data input is compounded from two types of data sets: advertiser conversions and publisher exposures.

The attribution protocol only requires a core set of fields which are:

- **Space ID:** The grouping space between the advertiser and publisher.
    - If the mapping system is not used, Space ID can be manually generated by the advertiser in coordination with the publisher.
    - Space ID must be unique per advertiser and publisher relationship.
    - Space ID is included in every row and its value may be different within the same input table.
- **Key Type:** The type of key value being provided such as email address or phone number.
    - Multiple key types and ADMAP IDs may appear in the same row.
- **ADMAP ID:** ADMAP ID is the key which is used to join the events between advertiser and publisher.
    - ADMAP ID is generated programmatically by the mapping system as described in the “Mapping Protocol” section above. In the case that a mapping system was used, the advertiser/publisher must then merge the ADMAP ID into their input events.
    - If the mapping system is not used, ADMAP ID is the original **key value**, which is the user’s personal information which can be optionally encoded as described in the “Mapping Inputs” section above.
    - Multiple key types and ADMAP IDs may appear for the same row.

Events must also include additional information necessary to generate reports about when and where the attribution occurred. These additional fields are not explicitly defined by this protocol but example schemas are given below.

Advertiser Conversions

Example advertiser conversion event schema:

Field	Type	Obfuscated	Example

Space ID	string	No	123
Key Type	ENUM	No	Possible values: ‘email’ / ’phone’ / ’ip’ / ’advertising_id’
ADMAP ID	string	Yes	043e3ef5e76e423c9d3d3a5605bcb8904496c654fcd16180386298150d8e1b8f
Timestamp	Integer	No	EPOCH timestamp. 1714933914
Event Metadata	List of fields	No	

Example advertiser conversion:

Field	Value
Space ID	123
Key Type	‘email’
ADMAP ID	043e3ef5e76e423c9d3d3a5605bcb8904496c654fcd16180386298150d8e1b8f
Timestamp	EPOCH timestamp. 1714933914
Event Name	‘Purchase’
Event Revenue	50.0
Event Currency	‘USD’

Note: In this example, there is a single ADMAP ID but there is an option to have a list of pairs of mapping key types and ADMAP values.

Publisher Exposures

Example publisher engagement event schema:

Field	Type	Obfuscated	Example
Space ID	string	No	123
Key Type	ENUM	No	Possible values: ‘email’ / ’phone’ / ’ip’ / ’advertising_id’
ADMAP ID	string	Yes	043e3ef5e76e423c9d3d3a5605bcb8904496c654fcd16180386298150d8e1b8f

Timestamp	Integer	No	EPOCH timestamp. 1714933800
Type	ENUM	No	Possible values: ‘click’/ ‘view’ / ‘engaged_view’
Campaign Metadata	List of fields	No	campaign name, ad set name, ad, etc

Example publisher exposure:

Field	Value
Space ID	123
Key Type	‘email’
ADMAP ID	043e3ef5e76e423c9d3d3a5605bcb8904496c654fcd16180386298150d8e1b8f
Timestamp	1714933800
Type	‘click’
Campaign Source	‘Ad system’
Campaign Name	‘Red shoes’

Note: In this example, there is a single match key but it is possible to have a list of pairs of match key types and values.

Attribution Protocol Outputs

The attribution system's output is an aggregated attribution result. The data includes advertiser conversion data and publisher/ad network campaign data. The mapping key is not included in the attribution results.

This protocol does not specify a strict output schema. Below an example schema is provided.

Attribution Data - The attribution system generates attribution data and provides it to either the publisher or the advertiser to enable campaign measurement and optimization of the mapped users.

Example:

Field	Type	Obfuscated	Example

Date	Integer	No	EPOCH timestamp in days. 171493000
Campaign Name	string	No	‘Red shoes’
Campaign ID	Integer	No	abc123
Number of Conversions	integer	No	5
Total Revenue	float	No	2500
Currency	ENUM	No	‘USD’

Attribution Protocol Methodology and Architecture

Overview

Below we describe a way in which an attribution system may be implemented.

The attribution system architecture has two main components:

1. Matching service - Responsible for finding the intersection between the network engagements and the advertiser app conversion (taking into consideration the attribution parameters such as the lookback window).
2. Aggregation service - Responsible for aggregating the matching results for a specific day. The advertiser and the network can set different dimensions of the aggregated data.

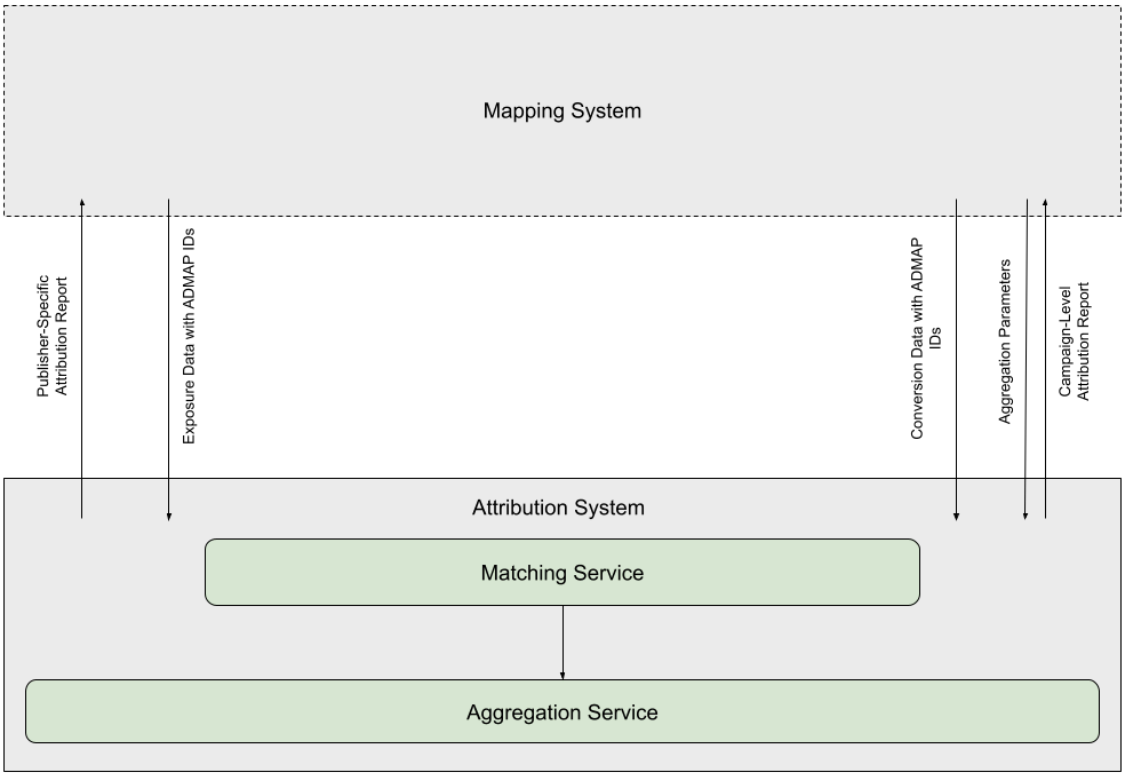


Figure 3: Attribution system architecture blueprint.

Attribution Matching Algorithm

The attribution matching algorithm is not based only on key matching. It takes into consideration two main factors:

- The priority of an engagement in the attribution waterfall.
- The time between the interaction with the ad and the conversion itself

Engagement Event Types

There are many types of engagements and each one of them describes different interactions with the advertisement. Each engagement shows a different level of intent of the user which should impact the priority that this engagement gets in the attribution waterfall. Clicks are active engagements where users take a specific action by clicking on the ad, showing intentional interest, whereas Impressions are passive views when the ad is displayed to users, potentially influencing their decision-making without direct action.

Engagement Type	Description	Priority
click_to_open	An ad click that redirects the user to open the advertiser app/website.	1
click_to_interact	A click that keeps the user in the same context, engaging with the ad, and does NOT direct that user to the advertiser app/website. Example: Like/share in a social media	2
engagement	An engaged view occurs when a meaningful view takes place within skippable videos. For example, when a user watches a skippable video ad for at least 5 seconds after the skip option appears.	2
view	An ad view is when an ad is rendered, and a viewable (as per viewability definition) impression takes place.	3
listen	Listening is when a user listens to an audio ad, such as a podcast ad.	3

Note: This is only an example list and priorities. Each advertiser should determine the engagement types they want to track and the priorities. The lowest priority number is the highest priority. Depending on the platform (Web / Mobile / CTV), there could be more engagement types.

The attribution algorithm uses a waterfall approach based on the priority of each engagement type.

Lookback Window

The lookback window is the maximum period of time after an ad engagement occurs within which a conversion can be attributed to the ad. Conversions that take place after the lookback window are considered organic. The attribution system has separate lookback window configurations per engagement type that should be controlled by the advertiser. The attribution algorithm takes into consideration the time between the interaction with the ad and the conversion itself.

Attribution Method

Advertisers and publishers need to agree on a methodology

- Rules-based examples of different attribution methods
    - Last Touch
    - First and Last (U-shaped)
    - Even Credit
    - Custom Heuristic (to cover anything else)
    - Time Decay
- Modeled
    - Custom Model (i.e. we have no opinion on models, that's up to the attribution system provider)

All methodologies also need to specify different parameters that are essential to the method, such as the lookback window, engagement type, priority, or any other relevant parameter.

See specific examples in the Reference Implementation section.

Matching Service and Aggregation Service Interface

The data that flows between the matching service and the aggregation service should include all the attributed and non-attributed conversions. This document doesn't define the structure or the way this data will be stored or transferred to the aggregation service.

Reference Implementations

In order to explain how the protocol and different components will work together, we outline reference designs of two systems, each which would work interoperably with each other.

Both proposed systems are designed to be operated by a third-party operator, the DCR. Both aim to satisfy the protocol's privacy and security design goals and are designed according to the input and output requirements described above.

We describe reference designs for the following types of systems and assisted by a third-party DCR operator:

- Mapping using Private Set Intersection (PSI)
- Attribution using Trusted Execution Environments (TEE)

We do not claim that the two proposed designs are the only possible system designs that can satisfy the protocol's security requirements. To that end, it is our intention to explore, present, and evaluate additional open designs in future versions of this document.

Mapping Using Private Set Intersection

We present a mapping system using a Diffie-Hellman private set intersection protocol based on elliptic-curve cryptography and leveraging an untrusted helper server. We refer to the untrusted helper server as the *facilitator*. The facilitator performs a join operation on encrypted match key records and also generates the mapping system outputs. The facilitator could be implemented by a Data Clean Room.

Figure 4 depicts the overall mapping workflow. It should be noted that in order to achieve correctness in outputs, participants are assumed to be *honest-but-curious*.

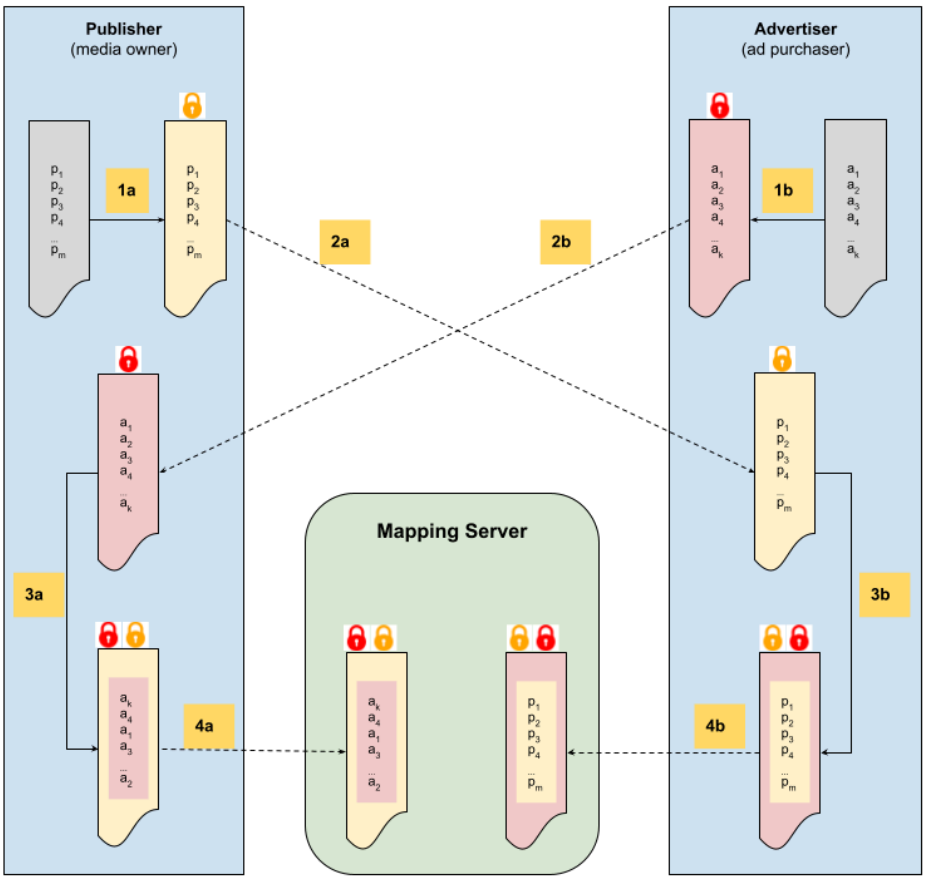


Figure 4: Mapping workflow using EC-DH-PSI and a facilitator.

The steps required to execute the mapping workflow are annotated in Figure 4 and described below.

1. Both the publisher (Figure 4, step 1a) and the advertiser (Figure 4, step 1b) each separately *blind* their input records using their own private keys.

2. The parties exchange their blinded data sets (Figure 4, steps 2a and 2b).

3. The publisher, upon receiving a once-blinded dataset from the advertiser, proceeds to blind it a second time with its own private key and shuffles the twice-blinded records (Figure 4, step 3a). The advertiser, upon receiving a once-blinded dataset from the publisher, proceeds to blind it a second time with its own private key (Figure 4, step 3b), but does not shuffle the records.

4. The parties each upload their twice-blinded datasets to the facilitator (Figure 4, steps 4a and 4b).

The facilitator then proceeds to perform the join on the twice-blinded datasets and computes the mapping outputs.

Note that the advertiser's records are shuffled by the publisher prior to step 4, whereas the publisher's records are **not** shuffled at any step, and their order is maintained throughout. The preservation of the order of the publisher's match key records enables the facilitator to generate the mapping output in the same order, as required by the attribution protocol.

Blinding
The blinding steps are performed using the ristretto255 group, implemented with the elliptic curve Curve25519. The blinding operations are commutative, such that two records twice blinded in opposing order can be compared by the PSI server.

Every workflow requires that each of the parties (advertiser and publisher) generate a new secret key, a scalar *k*, embed each input match key x_i into a ristretto point X_i, and perform the blinding function by computing points kX_i on the elliptic curve Curve25519.

For clarity, if *A* is the ordered set of input records from the advertiser, and *P* is the ordered set of input records from the publisher, then note that each of the parties (advertiser and publisher) perform the blinding with their own keys k_a and k_p which remain secret to them. If the advertiser's secret key is k_a and the publisher's secret key is k_p, then:

- The twice-blinded dataset in step 4a consists of:
  all points $k_p k_a A_i$ on Curve25519, where A_i is the ristretto point embedding the match key $a_i \in A$

- The twice-blinded dataset in step 4b consists of:
  all points $k_a k_p P_i$ on Curve25519, where P_i is the ristretto point embedding the match key $p_i \in P$

Attribution Using TEE Server

We present an attribution system involving a matching component that leverages a Trusted Execution Environment (TEE) to restrict access to the advertiser's and publisher's provided key records.

The Attribution System can be logically broken down into two subcomponents: The Matching Service and the Aggregation Service. These do not need to be logically separated, but we present them separately here for clarity. The Matching Service joins

on key records and performs attribution (e.g. comparing timestamps and applying attribution methodology) for each conversion. The Aggregation Service is responsible for aggregating the conversion-level results, applying any minimum thresholds or privacy noise, protecting against differential privacy attacks, and otherwise collecting information necessary to create attribution reports for the protocol outputs.

Matching Service

Figure 5 depicts the TEE-based matching data flows.

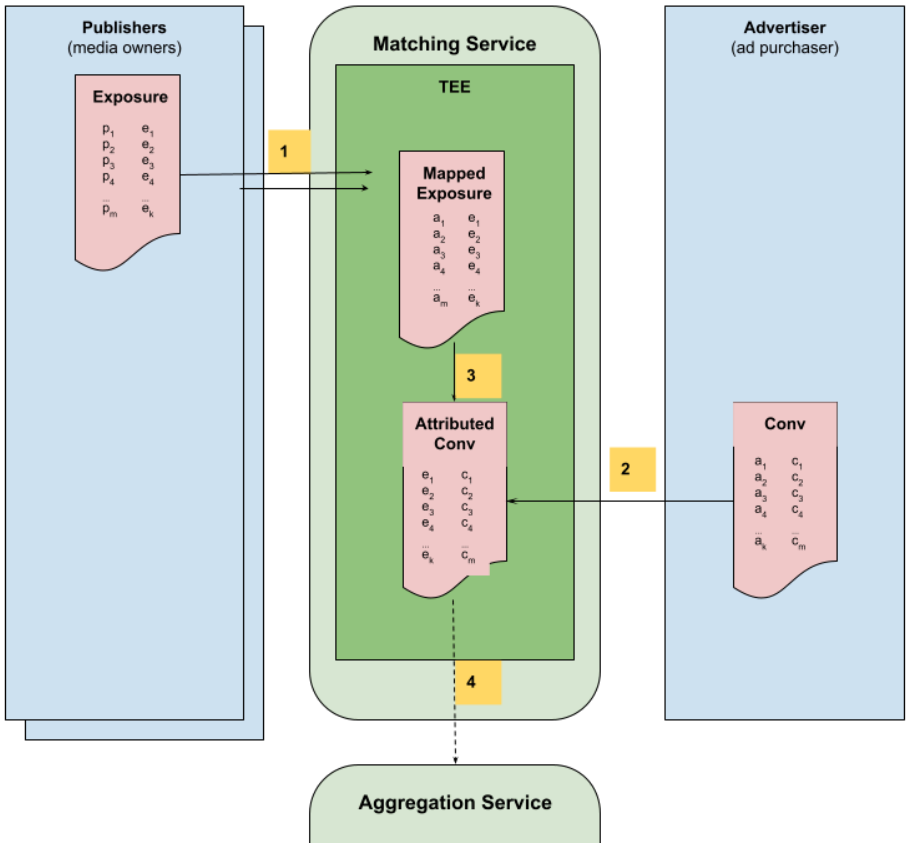


Figure 5: Matching using TEE matching server

The steps required to execute the matching are annotated in Figure 5 and described below.

Before interacting, the publisher and the advertiser establish trust of the TEE by performing a *remote attestation*. This “step 0” is a precursor to any interaction with a TEE running in a secure mode. All participants must have enough ancillary information (e.g. source code that can be used to reproduce the program, or a signature over the program by an authority they trust) to verify that the TEE is credibly remotely attesting to only processing data in the prescribed manner. In this example implementation, data is streamed directly from the participants to the TEE – the mechanisms for verifying the

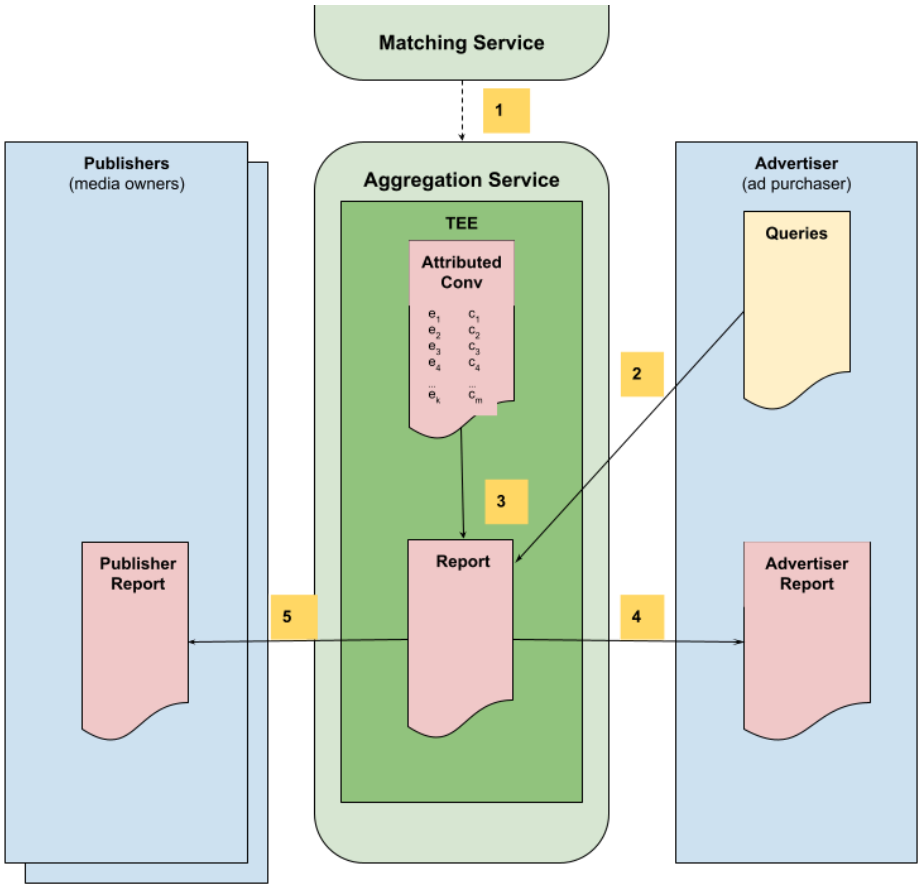
TEE, authenticating the participants, and establishing end-to-end-encrypted communication to the TEE are not described in this reference implementation to keep it focused on the steps that are specifically relevant to computing attribution. However, for the security properties of the system to hold, all of these steps must take place every time data is communicated in or out of the TEE.

1. Each publisher sends an exposure table to the TEE. This must be keyed by a publisher id (*p*), and contain the other ad exposure information described in the inputs section (*e*). Different publishers may use different publisher id spaces.
2. The advertiser sends the conversion table to the TEE. This must be keyed by the advertiser id (*a*), and contain the other information described in the inputs section (*c*). Unlike the exposure data added in step 2, all advertiser data must use the same advertiser id space.
3. The TEE groups conversion/exposure events by Space ID and joins the exposure table to the conversion table, on the advertiser id (*a*). It then performs attribution based on previously agreed methodology, attributing each conversion to zero or more exposures. Note that information after this process is still individual level information about specific ad exposures and conversions, but no longer contains explicit identifiers (they are discarded after being used to join). The TEE also calculates summary statistics for unattributed conversions and ad exposures (not pictured).
4. The TEE sends the attributed conversion table and ancillary summary tables to a TEE being used by the Aggregation Service.

Note that it is important that the Matching Service only sends outputs to the Aggregation Service. As noted in step 5, information at this stage is still individual level information about specific ad viewing events and individual conversions by individual consumers. While no longer explicitly keyed by an identifier, revealing this information to some of the participants at this stage would allow them to easily reverse engineer which individual person each row is about. For these reasons, it must be sent only to the Aggregation Service, which will provide the necessary aggregation and other techniques to summarize the data into anonymous attribution reports.

Aggregation Service

Figure 6 depicts the TEE-based aggregation data flows.



Step-by-step

1. The Matching Service sends the attributed conversion table to the Aggregation Service TEE, as well as any summary information about unattributed conversions and exposures (not pictured).
2. The advertiser submits queries that will be used to construct attribution reports. This may include details such as: which time period to be considered, specific attribution methodology to apply (e.g. last click vs. even credit), which campaigns to be considered, which breakdown reports to be included, or any other configurable options in the attribution report.
3. The TEE prepares an attribution report based on the query, as well as any other rules or limits established by the system. A typical report may be to aggregate all conversions for the same campaign, over the past month, with breakdowns per day and per market. This is also the step where minimum thresholds, rounding, noise, or any other aggregate privacy measures are applied.
4. The TEE sends the report to the advertiser.
5. The TEE sends publisher-specific reports to each participating publisher. These have the same information as the advertiser report, but filtered down to only include the information about that specific publisher.

Calculating Match Rates

In order to mitigate against Information Leakage via Match Rates, the Aggregation Service shall apply thresholding, rounding, and introduce noise to calculated match rates prior to providing them in output to the advertiser and publisher.

Collusions and Threats

This section provides threat vectors that must be considered by any component designs adhering to this specification. The proposed attribution protocol and matching system designs are analyzed in regards to various collusion scenarios and threats.

Collusion Scenarios

We use the term *collusion* to mean a scenario where two or more protocol participants share information. This can be due to malicious intent, or because they happen to be commonly owned and operated. For example, a publisher may also own and operate an SSP platform. In some cases, a media company may own and operate both an SSP and a DSP and at the same time assume the role of publisher in this protocol's operation. In the latter case, the media company may not be malicious, but we must consider the implications of information sharing among a subset of participants insofar as protocol privacy and security design goals are concerned. We therefore propose that:

- The mapping protocol designs consider the following *collusion* scenarios, when a matching system operator is required by the proposed matching system:
    - Publisher and matching system operator are sharing information
    - Advertiser and matching system operator are sharing information

- The attribution protocol component designs consider the following *collusion* scenarios:
    - Publisher and attribution system are sharing information
    - Advertiser and attribution system are sharing information

- Additionally, we consider the matching system operator (when a matching system operator is required) and attribution system sharing information.

Mapping System Collusion Scenarios

In general, the risks associated with a particular mapping approach would need to be analyzed by considering the implementation details of that approach. In this document, we present such an analysis for the reference implementation for mapping that leverages PSI.

Table 2 shows the implications of the advertiser/publisher colluding with the mapping system operator on the proposed PSI server mapping system design, when used in conjunction with the outlined attribution protocol.

Collusion Scenario	Impact on Design Goal 1: Security of PII	Impact on Design Goal 2: Privacy of User Identity	Impact on Design Goal 3: Privacy of Audience Membership	Notes
Publisher and mapping system operator are sharing information	Unaffected The publisher and the mapping system operator cannot learn the PII of *any* end user of the advertiser.	Unaffected The publisher cannot learn the PII of *any* end user of the advertiser that are not in the overlap.	**Affected** The mapping system can share the unencrypted labels with the publisher, which can infer the keys associated with the members of the shared audience.	The publisher could cheat by forcing the mapping system to generate incorrect labels and/or return incorrect overlap rates. This could make the advertiser bid on incorrect ad requests.
Advertiser and mapping system operator are sharing information	Unaffected The advertiser and the mapping system operator cannot learn the PII of *any* end user of the publisher.	Unaffected The advertiser cannot learn the PII of *any* end user of the publisher that are not in the overlap.	Unaffected The advertiser's double blinded records are shuffled by the publisher. Therefore, the advertiser cannot learn which of its end users are in the overlap.	The advertiser could cheat by forcing the mapping system to generate incorrect labels and/or return incorrect overlap rates. Our expectation is that this is low risk since it does not provide any advantage to the advertiser.

Table 2: PSI Server Mapping: Impact of Collusion Scenarios

Attribution System Collusion Scenarios

Table 4 shows the implications of the advertiser/publisher colluding with the attribution system on the outlined attribution protocol.

Collusion Scenario	Impact on Design Goal 1: Security of PII	Impact on Design Goal 2: Privacy of User Identity	Impact on Design Goal 3: Privacy of Audience Membership	Notes
Publisher and attribution system are sharing information	Unaffected The publisher and the ad attribution	Unaffected The publisher cannot learn the PII of *any*	**Affected** The ad attribution system can	

	system operator cannot learn the PII of *any* end user of the advertiser.	end user of the advertiser that are not in the overlap.	share decrypted labels with the publisher, which can infer the match keys associated with the members of the matched audience.	
Advertiser and ad attribution system are sharing information	Unaffected The advertiser and the ad attribution system operator cannot learn the PII of *any* end user of the publisher.	Unaffected The advertiser cannot learn the PII of *any* end user of the publisher that are not in the overlap.	Unaffected The ad attribution system can share decrypted labels with the advertiser. However, the advertiser cannot infer the match keys associated with ad requests and hence cannot infer the members of the matched audience.	The advertiser can learn which ad requests are positive (including for mappings which it is not a participant of).

Table 4: attribution Protocol: Impact of Collusion Scenarios

Threats

In the context of this document, a threat is an activity that can be performed by one or more protocol participants in order to exploit the proposed mechanisms such that our privacy and security design goals are violated. We document and comment on potential threats, attacks, and their possible mitigations below.

Information Leakage via Match Rates

The overlap rates computed by the mapping system and shared as outputs with both the advertiser and the publisher parties could enable one or both of the parties to test for the presence of individuals within the list of matched users.

For example, an advertiser may perform multiple successive matches with a publisher using ADMAP, taking special care to insert and remove an individual PII match key record from its inputs, and observe the outputted match rate to determine whether the

added or removed record is present in the publisher's inputted records. This would violate **Design Goal 3 - Privacy of Audience Membership**.

Matching system designers could introduce noise, rounding, and/or minimum thresholds to the match rate results, thereby mitigating the effectiveness of this attack in practice.