DIGITAL ADVERTISING IN A PARADIGM WITHOUT 3RD PARTY COOKIES

IMDA PET SANDBOX – META CASE STUDY

INFOCOMM MEDIA DEVELOPMENT AUTHORITY

Contents

Business Case

1. **Tracking technologies, like 3rd party cookies, are presently a mainstay** in the digital advertising ecosystem. Publishers, advertisers and adtech firms rely on the collection and sharing of user / device identifiers to analyse how consumers can be shown advertisements best aligned to their online interests or activities.

2. However, the ecosystem is preparing for **a paradigm in which the collection of user / device identifiers[1] that are linkable across apps / websites is no longer feasible**, and trust in the ecosystem is low. A prominent avenue where solutions to measure attribution of digital ads without tracking technologies is actively discussed is the World Wide Web Consortium (W3C)'s Private Advertising Technology Community Group, or "PAT-CG".

3. **Meta and Mozilla**, as members of PAT-CG, have proposed a solution – **"Interoperable Private Attribution" or IPA**. It uses a combination of multiparty computing (MPC), aggregation, differential privacy (DP) and write-only identifiers to enable attribution measurement. The solution aims to measure advertising outcomes based on impressions shown on publisher website(s)/app(s) and conversions occurring on an advertiser website/app.

Methodology

4. **'Last Touch Attribution'** was the use case chosen for the proof of concept (POC)**,** which is one of the most basic digital marketing models used by advertisers to try to understand how much value was generated by showing their ads on specific apps or websites.

5. The POC will address the following three objectives:

   i. Find the technical bounds of running queries with IPA on high volume of records

   ii. Assess the usefulness of ad measurements computed through IPA

   iii. Identify the necessary conditions of compliance and governance

[1] Identifiers which can be considered personal data if the organization collecting them is able to identify an individual from the identifiers collected or from those identifiers and other information the organization has or is likely to have access to

6. **A consortium of organizations** participated in the POC to implement the solution architecture, perform tests to measure last touch attribution from synthetically generated data resembling ad campaigns and assess the feasibility of IPA on real world computing and communication infrastructure. The consortium partners and their roles are summarized in the Table 1 below.

Table 1 – POC consortium

Role	Organization
IPA engineering & development	Meta and Mozilla
Attribution Measurement (Adtech)	Kevel
Multiparty Computation Helper Parties	Digital Trust Centre at Nanyang Technological University, Akamai and Cybernetica

7. **The use case flow** for the POC was as follows:

   i. An advertiser runs an advertisement campaign to publish its advertisement on a specific website ("publisher").

   ii. The publisher displays the advertisement to its users.

   iii. Several users view the advertisement, of which some of would go to the advertiser's website and make a purchase (i.e. conversion).

   iv. An Adtech entity, supporting the advertiser, measures the attributed conversion value for the ad campaign (i.e. the total value of purchases made by users within 7 days of seeing an ad on the publisher's website/app).

8. The 3 **Helper Parties servers** were chosen such that they were operated by different organizations, used different cloud services and located in different countries (see Table 2). This was done to measure the performance of a maximally secure setup, where there is no single

point of failure (e.g. personal data[2] remains protected even in a scenario where a single cloud provider is compromised, or a single country raids a data center).

Table 2 – HP Servers

	NTU	Akamai	Cybernetica
Cloud Provider	AWS	Akamai	Azure
Location	Copenhagen (Denmark)	Frankfurt (Germany)	Gavle (Sweden)

Solution Architecture

9. **Seven key steps were defined in the POC solution architecture** for the attribution report to be generated (see Figure 1 below):

   - Step 1 – A unique browser or device key is created for the user upon installation of browser or mobile operating system. This key is cached in the browser and will not be accessible by browser vendor or any third party. Browsers expose an API (namely, *getEncryptedMatchKey*) which will "shred" the cached key into 3 random "secret shares". Pairs of these "secret shares" are then encrypted using Helper Parties' public keys. We call the result of this operation "Encrypted Secret Share Pairs" or ESSP. Each Helper Party can only decrypt its assigned ESSP.

   - Steps 2 & 3 – When an impression occurs at the publisher or a conversion occurs at the advertiser, the publisher or advertisers call *getEncryptedMatchKey* to obtain the ESSPs for that user. The publisher and advertiser also collect the relevant activity data needed for the attribution measurement. The activity data are appended to the ESSPs.

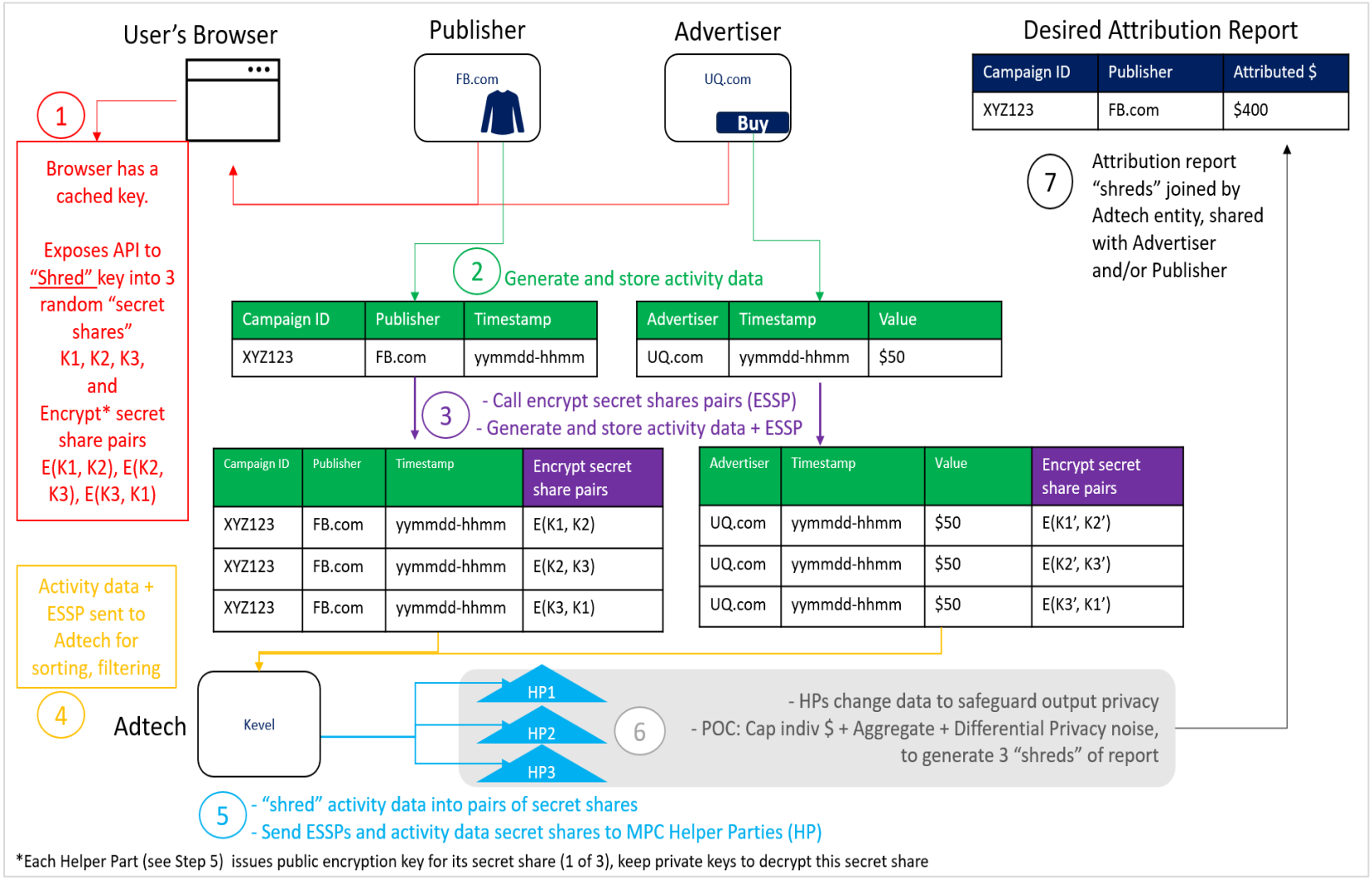
- Steps 4 & 5 – The dataset of activity data with ESSPs are sent to an Adtech entity for filtering and sorting as desired by the type of attribution report. The activity data is further "shredded" into pairs of secret shares and, along with ESSPs, sent to three Helper Parties. This marks the start of the MPC process.

- Step 6 – The Helper Parties process the activity data i.e. cap each attribution value contribution, aggregate the data and add DP noise (for output privacy[3]), and each send their shreds of the processed data to the Adtech entity.

- Step 7 – Adtech entity then merges the shreds of the reports to generate the Attribution Report. The Attribution Report is shared with the Advertisers and Publisher.

[3] Output privacy reduces the risk of an individual user's contribution to be learnt precisely from the Attribution Report.

Figure 1 – Solution Architecture



*Each Helper Part (see Step 5) issues public encryption key for its secret share (1 of 3), keep private keys to decrypt this secret share

10. **A combination of 3 safeguards** was tested in the POC which enable attribution measurement without disclosure of identifiers *and* protect reidentification of individual or small group of individuals through the attribution reports.

    i. **Match key provided as ESSPs**: The key cached on the user's browser or mobile device is not disclosed to any other entity in its original form. Instead, the key is "shredded" into 3 ESSPs on the user's browser or mobile device and only these ESSPs are provided to the publisher and advertiser.

    ii. **Helper Parties behaving as 'Honest Majority'**: For the POC, it was assumed that the helper parties are not colluding to reconstruct the "shreds" of activity data or ESSPs, and that a suitable governance structure will be in place to ensure non-collusion between any two of the three Helper Parties. Based on this 'honest majority' structure, no single Helper Party learns useful information about the match key or activity data from its own set of "shreds'' received.

    iii. **Output privacy in Attribution Report**: The POC considered scenarios when an attribution report could potentially be used to re-identify an individual. For example:

    - An individual has such an unusually large purchase value, that a total alone would reveal them as being one of the users who contributed to the total.
    - A query is constructed to sieve out an individual with non-zero contribution from among multiple users.
    - Multiple queries are constructed, which differ by only one user's contributions, such that subtracting one total from the other would single out the contribution of that one user.

    As such, individual attribution value contribution of each user was capped, the capped contributions were aggregated and then noise was added to the data for differential privacy guarantee.

Regulatory Considerations

11. **Clarifications were sought from Singapore Personal Data Protection Commission** (PDPC), based on the scope of this POC and the design of the IPA solution architecture, on whether the platform (browser or mobile device vendor) and the participating website/app (e.g. advertiser, publisher, adtech entity) were considered to have collected, used or disclosed personal data for the generation of the attribution report, and if any additional safeguards are required to ensure compliance to the PDPA.

12. PDPC provided guidance based on the assumption that the activity data fields shared by the Publisher and Advertiser with the Adtech Entity for attribution measurement involved personal data.

    The key PETs involved in the IPA architecture include:

    - Anonymisation of the generated browser/device key by "shredding" and encrypting the browser/device key into ESSPs accessible only by respective Helper Parties.
    - Anonymisation of activity data by shredding of activity data into pairs of secret shares accessible only by respective Helper Parties based on the IPA's protocol.

    PDPC used a risk-based approach, in that data that has sufficiently low risk of re-identifying any individual will be considered anonymised data under the PDPA[4]. Based on the design of the IPA POC, no Helper Party will be able to know the browser/device key or activity data in its entirety, unless any Helper Party is able to access and combine any two of the three secret share pairs. The guidance below provides PDPC's recommendations on some of the technical, governance and process safeguards that can be put in place by each stakeholder to lower the risk of re-identification.

13. Guidance for Publishers and Advertisers

    i. **Publisher and Advertisers are Data Controllers:** Generation of the attribution report is for both the Publisher's and Advertiser's purposes. The Publisher and Advertiser are data controllers (DCs) of activity data which they each collect from individuals, a subset of

[4] Refer to Section on Anonymisation in Advisory Guidelines on the Personal Data Protection Act for Selected Topics and PDPC's Guide to Basic Anonymisation.

which they would share with the Adtech Entity. As DCs, they should assess and minimise any downstream risks of re-identifying any individual by other stakeholders or unauthorised parties from the activity data shared and the generated attribution report.

ii. **Data Minimisation:** In determining the activity data fields to be used in generating the attribution report, both the Publisher and Advertiser should apply the principle of data minimisation. Based on the intended structure of the attribution reports, they are to select only data fields that are relevant to the reports. Where possible, they should remove any direct or common identifiers that are tagged to any individual (e.g., campaign IDs specific to individuals or customer IDs) and consider using activity data fields that are less likely to identify any individual.

iii. **Structure and Quantity of the Attribution Reports:** In determining the structure and quantity of the attribution reports to be generated by the Adtech Entity, the Publisher and Advertiser should consider whether the generated reports (individually or in aggregation) may result in disclosure of any of their customers' personal data and take reasonable measures to reduce the risk of such individual linkage (e.g. transaction value made at Advertiser's website/app by a specific individual which is tagged with a specific campaign ID is disclosed to Publisher via the generated report(s)). In cases where the generated attribution report(s) can be used to reveal personal data about an individual to the other stakeholder (i.e. Advertiser or Publisher), it will be considered disclosure of personal data for which consent is required, unless any of the exceptions provided in the PDPA apply.

14. Guidance for Adtech Entity

i. **Data Intermediary role.** PDPC viewed the Adtech Entity as a data intermediary (DI) that processes personal data (i.e., sorting, filtering and "shredding" the activity data) on behalf of and for the purposes of both the Publisher and Advertiser. PDPC has given guidance that express consent is not necessary for an organisation to share personal data with its DI to process personal data on its behalf, provided that the personal data is not used by the DI for other purposes without the consent of the individual[5]. As such, consent is not

[5] See PDPC's Guide to Data Sharing, paragraph 1.8.

required for the Adtech Entity to collect (from the Publisher and Advertiser), sort, filter and "shred" the activity data for the purposes of generating the attribution report for the Publisher and Advertiser.

ii. **Obligations of Data Intermediary.** Nevertheless, as a DI, the Adtech Entity will be subject to the Protection, Retention Limitation and Data Breach Notification Obligations under the PDPA[6]. For avoidance of doubt, where the Adtech Entity uses the activity data beyond what is required and agreed with the Publisher and Advertiser, the Adtech Entity will be considered a DC in relation to the activity data, and all PDPA Obligations will apply (including the need to obtain consent from the individual to collect the activity data from the Publisher and Advertiser).

15. Guidance for Platform Provider (Browser or Mobile Device Vendor)

i. **Browser/Device Key**. In the IPA implementation, the browser/device key is intended to be kept hidden from any parties and will not be combined with any data that the Platform Provider or any other third parties may have, to identify the user. As such, the Platform Provider's generation of the browser/device key will not constitute collection of personal data. The "shredding" and encryption of the unique browser/device key will also not constitute use of personal data, and the Data Protection Provisions under the PDPA will not apply.

16. Guidance for Helper Parties

i. **Output considered anonymised data.** PDPC considers the output from PET implementation to be anonymised data, so long as the risk of re-constructing the browser/device key and activity data from the data remains reasonably low. This risk should be assessed in conjunction with any technical, governance and contractual safeguards implemented system-wide in the IPA implementation.

[6] For instance, the DI will need to ensure that the personal data it collects and anonymises on behalf of the Publisher and Advertiser is adequately protected, and not retain the personal data for periods longer than necessary. The DI is also required to notify DCs without undue delay from the time it has credible grounds to believe that the data breach has occurred.

ii. **Processing anonymised data.** Based on the activities the Helper Parties undertook in the POC (i.e., collecting activity data secret shares and corresponding ESSP from the Adtech Entity, and processing the data as described in the POC), the Helper Parties will be considered to be processing anonymised data and thus not be subject to the PDPA.

17. Guidance on Additional Safeguards

i. **Lower the risk of identification from persistent key.** Given that the browser/device key is designed to be permanent and unique, it has the characteristics of an identifier which could be used by various websites/apps to combine other information about the user. This increases the likelihood of the browser/device key being personal data. Additional safeguards that lower the risk of identification may include, for instance:
   - Ensuring that the browser/device key is generated and used only for the attribution report. It may also be worthwhile considering whether a temporary browser/device key can be deployed instead (e.g., imposing a validity period and re-generation cycle for each browser/device key).
   - Ensuring that the techniques used in the "shredding" of browser/device key and activity data are sufficiently robust to prevent the same key "shreds" from being generated at both Publisher and Advertiser's end, as well as threat actors from being able to execute an attack (e.g., rainbow table attack) to precompute possible key "shreds" and combinations. Where possible, these techniques (including encryption) should be aligned with industry standards (e.g., using encryption protocols widely accepted by industry to be secure).

ii. **Lower the risk of reidentification at Helper Parties.** There are also risks of re-identification of individuals due to the critical role the Helper Parties play in the IPA solution architecture. Additional safeguards that may be put in place to lower the risk of re-identification may include, for instance:
   - Ensuring that Helper Parties do not attempt to collude or re-identify any individual from the anonymised data through contractual means and other governance obligations (e.g. audits). Technical safeguards (e.g., programmatic guardrails) can also be explored to prevent or red-flag possible collusion between Helper Parties.
   - Ensuring that Helper Parties put in place baseline governance and technical implementation measures to protect and secure their secret keys from unauthorised access/compromise (e.g., industry-recognised processes and standards such as ISO and NIST).

Feasibility Assessment

18. **IPA-based Attribution Report retained usefulness.** The ability of IPA to deliver attribution reports useful to adtech or advertiser firms was assessed by how much such reports deviated from a perfectly accurate answer, assuming entities were able to collect accurate device-level linkable data using existing techniques (like 3rd party cookies or mobile advertising IDs). The assessment was based on the relative error in IPA-based attribution values, reflected in the Table 3 below. As expected, the relative error was zero up to the point of "capping" and adding DP "noise", after which the relative error ranged from 0.2% to 3% for a DP epsilon value of 1.

19. **MPC communication latency was acceptable**. High Latency, an issue inherent to MPC based computations distributed across nodes (i.e. Helper Parties) instead of central server, was an issue the project anticipated. Table 3 below describes the number of records queried and the time taken to generate the measurement report. Kevel, the participating adtech organization in this POC, opined that the time to generate such reports was not disruptively high. The consortium identified some ideas to reduce latency, e.g.

    - Locating the Helper Party servers such that they are geographically as close as possible while avoiding vulnerability to easy attacks.
    - Accommodating parallelisation in the IPA code (currently using a single core).
    - Leveraging research work dedicated to reducing network communication in MPC.

Table 3 – Evaluation of latency and usefulness

Query Size	Time to generate report	Deviation before Output Privacy measures	Deviation after Output Privacy measures
100,000	35 mins	0%	-2.0% to 3.0%
500,000	2.6 hours	0%	-0.2 to 0.8%
1,000,000	6.5 hours	0%	0.4% to 0.7%

20. **Management of encryption keys could be more secure.** Post POC, security of keys used to encrypt the match key secret share pairs was discussed by the consortium. As a part of the IPA protocol, the Helper Parties issue the public keys for the encryption of the ESSPs and retain the private keys. These are then used by Helper Parties to decrypt the ESSPs before generating the attribution reports. This could be made more secure by strategies like:

    - In the broadcast of encryption keys, Helper Parties should use methods which protect them against man-in-the-middle attacks and vouch for their provenance. For example, the "Key Consistency and Discovery" protocol described in this IETF memo[7].

    - In the POC, the public keys were part of the docker image that was distributed to the Helper Parties. However, the key management system should be refactored out and separately configured so that the code doesn't have access to the keys.

21. **Cost of upkeep of Helper Parties could be reduced.** While the cost of running the computations was insignificant, this was not the case for the upkeep of idle Virtual Machines and container restarts. In a scaled deployment, the Virtual Machines are expected to be occupied with running many ad tech companies' queries. Further, container restarts could be reduced by providing API to reset Virtual Machines to default configuration.

Conclusions and Next Steps

22. The POC revealed that while PET-based solutions like IPA can be used to reliably measure advertising outcomes in an ecosystem without tracking mechanisms like 3rd party cookies, there are governance processes and technical improvements required to use it at production-level. These could range from encoding best practices for participating entities to safeguard against risks of re-identification, to reducing the time and cost of upkeep of the system.

23. Findings of this POC were presented at PAT-CG (technical consortium of tech companies like Apple, Google and Mozilla) in Seville, Spain on 11 September 2023.

[7] https://datatracker.ietf.org/doc/html/draft-ietf-privacypass-key-consistency

24. Taking the learnings from the POC as well as feedback from PAT-CG, the IPA team will further engage with other regulators to assess their viewpoints on the regulatory considerations discussed above and work to further improve IPA's performance and ease of deployment.