COLLABORATION ON FIRST PARTY DATA TO ENABLE CUSTOMER ACTIVATION

IMDA PET SANDBOX – SPH MEDIA CASE STUDY

INFOCOMM MEDIA DEVELOPMENT AUTHORITY

Contents

Use Case Background

1. To ensure that consumers are served relevant advertisements, data on customer behaviour and activity is essential. 3rd party cookies have traditionally performed this role, enabling **advertisers to track and link the behaviour of users across websites** so that they understand the profile of customers they serve advertisements to. However, growing global privacy concerns have led to the need to find alternatives to 3rd party cookies for audience analytics and advertising.

2. As an alternative way to understand consumer activity and behaviour without using 3rd party cookies, publishers and advertisers need a mechanism or tool that enables them to learn more about their consumers while preserving the privacy of the consumer, i.e. not being able to re-identify or discern who the individual corresponding to a user may be.

3. In this context, SPH Media is exploring the use of Privacy Enhancing Technologies (PETs), particularly the use of Trusted Execution Environments (TEEs). TEEs would allow advertising partners to serve relevant and targeted advertisements to potential customers, without collecting or learning any personal identifiable information about them.

4. The effectiveness of the TEE in balancing privacy with utility could create new business value for both advertising partners as well as platform owners like SPH Media.

Use Case Details

5. <u>The POC will address the following objectives:</u>
    a. Evaluate the effectiveness of TEE to securely and privately match common customers and generate a lookalike list (LAL)
    b. Understand relevant data protection safeguards put in place to enable customer activation

6. <u>POC stakeholders</u>
    a. **SPH Media (Publisher) –** Owns platform where advertisements would be shown
    b. **Global Wealth Manager ("GWM") (Advertiser) –** Intends to increase advertising-effectiveness by tapping on SPH Media's customer database to find new customers.
    c. **Data Management Platform ("DMP") Partner –** Partners SPH Media to ensure that the relevant advertisement is shown to the right audience within the LAL

d. **Decentriq (TEE Solution Provider) –** Provider of PETs solution built on TEE technology

7. There are two key phases to the POC: **1) Identifying common customers between SPH Media and GWM to create Lookalike Audiences and 2) Customer Activation.**

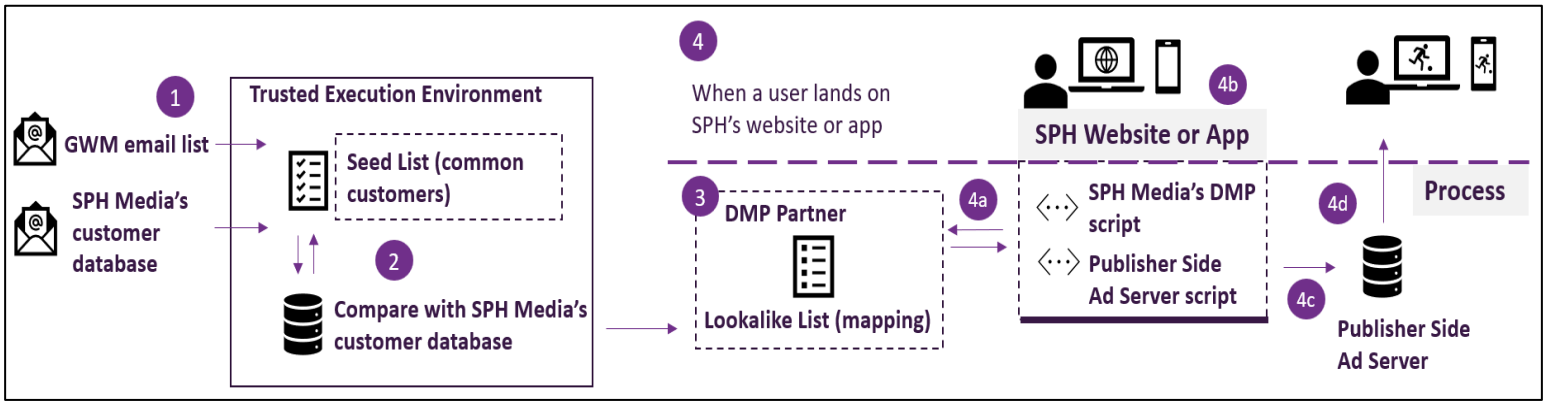
    a. In phase 1, customer identifiers i.e. email addresses from SPH Media and the GWM are hashed by both Advertiser and Publisher using the same algorithm to enable the identification of common customers without revealing or exchanging personal information of such customers. The matched seed list is used to generate a lookalike list (LAL) of the common customers. This matching and model generation is done within the TEE solution provided by Decentriq.
    b. In phase 2, when a user from the LAL lands on SPH Media's website or app, they would be shown an advertisement.

Solution Architecture

The Decentriq solution

8. Decentriq provides the technology where private data can be used for collaboration without being shared, for a specific limited purpose. The solution leverages the use of TEE, which takes advantage of hardware security features and cryptographic protocols to ensure:
    a. **Confidentiality and Integrity** - Data is stored and processed within the hardware chip in a secure manner, such that the data can neither be exfiltrated nor modified
    b. **Authentication and Access Control** – Strict controls and rules in place to ensure only specific input data and/or persons are allowed view results
    c. **Verifiable Logic** - Data is only decrypted for pre-approved software and functions in advance
    d. **Remote attestation** – A mechanism to verify the integrity of the processing environment within the TEE, to ensure it has not been tampered with.

POC Overview



Simplified POC flow

Phase 1: Selection of Lookalike Audiences

9. **Step 1** – GWM uploads a list of hashed email addresses to the TEE, to identify common customers with SPH Media. SPH Media uploads a list of its customers' hashed email addresses, corresponding generated **SPH User** ID (**SPH ID**) and corresponding segment IDs which represent the interest areas of each customer (e.g., segment ID 001 to represent soccer, ID 002 for fashion etc) As emails are hashed similarly, identification of common customers is possible without revealing or exchanging personal information of such customers between SPH Media and GWM.

10. **Step 2** - The list of matched common customers between GWM and SPH Media and its corresponding customer segment IDs form the "Seed List". This list contains the base profile of customers, which will be used by SPH Media to further select other customers within its database of customers, who will then be tagged as lookalike audiences. Matching of common customers, creation of the seed list and selection of lookalike audiences happen with the TEE, and at no point is any of this information visible or accessible to any party.

How TEE works:

- Two key computations occur within the TEE: (1) matching of customer using hashed emails from both parties and (2) training and inference of a machine learning Lookalike model in order to find lookalike audiences based on the matched common customer attributes. TEE's secure environment prevents both parties from accessing any interim results during the computation, thus protecting the identity of matched customers.

- The TEE serves as a secure enclave to allow data to be computed in an encrypted state. TEE prevents tampering of the data by creating a secure execution environment segregated from the rest of the operating environment through hardware encryption technology. Only authorised codes can manipulate data within the TEE, and this is verifiable by the remote attestation within the TEE. Throughout the process, the data remains encrypted and inaccessible to anyone including the GWM, SPH Media, Decentriq (the provider of the cleanroom solution) and Microsoft (the cloud provider).
- The TEE leverages Intel SGX (Software Guard Extensions) and AMD SEV-SNP (Secure Encrypted Virtualization Secure Nested Paging) technologies for hardware-based isolation[1] and attestation[2] to ensure that data and code are not accessible by attackers even if the main operating system is compromised. TEEs protect code and data from malicious software, privilege escalation and rogue admins.
- The solution also adheres to confidential computing standards established by the Confidential Computing Consortium (CCC) to ensure robust data isolation and verifiable processing through remote attestation.
- Furthermore, the maintenance of SOC2 Type 2 compliance by Decentriq provides additional assurance of security controls.

11. **Step 3** – The output from the TEE is a list of Lookalike Audiences, known as the Lookalike List (LAL). The LAL consists of will comprise a list of the customers' **SPH User ID** (**SPH ID**) and the corresponding segment ID. The Advertiser does not have access to the seed list nor the "lookalike customers" list. The list is shared with and stored by SPH Media's DMP partner.

Phase 2: Serving the advertisement:

12. **Step 4** – SPH Media's website and app are enabled with two scripts: 1) SPH Media's DMP Script and 2) SPH Media's Publisher-side Ad Server (PAS) Script that would programmatically ensure individuals – if they are from the LAL - are shown the right advertisement.

13. **Step 4a – b** – Whenever a user lands on SPH Media's website or app, the SPH Media DMP Script confirms that the user is within the LAL and if so ensures that the user's corresponding segment information, i.e. segment ID is picked up.

[1] SGX uses a hardware-enforced approach to create enclaves, which are isolated memory regions within the CPU's memory. These enclaves are protected by memory encryption and access controls, ensuring that data and code within them are not accessible by other parts of the system.
[2] The attestation allows a remote party to verify the integrity and authenticity of the code running within an enclave. This process involves generating cryptographic reports about the enclave signed using a hard-ware-protected key, which can be used to verify its configuration and integrity.

14. **Step 4c - d** – The PAS script then picks up the segment ID and checks against active advertisement campaigns within its servers.[3] If there are any active ad campaigns targeting the specific segment ID, the corresponding advertisement from the campaign is then shown to the user.

Regulatory Learnings

15. SPH Media sought Practical Guidance (Guidance) from the Personal Data Protection Commission (PDPC) on the following:

    a. Whether the Publisher and Advertiser would be considered to have disclosed personal data to each other by uploading of their respective list of customer information for processing within the TEE; and
    b. Whether the SPH IDs and customer segment IDs that the Publisher shares with the DMP partner constitute personal data under the PDPA.

Whether the Publisher and Advertiser have disclosed personal data for processing within the TEE

16. In this POC, PDPC notes that the solution is designed such that the Publisher and Advertiser will not be able to access each other's data input to the TEE. Given so, PDPC is of the view that there is likely to be no disclosure of personal data between the Publisher and Advertiser. In particular, we note that the following safeguards have been implemented in the POC to prevent access to the data uploaded to the TEE:

    a. Hashing. Email addresses uploaded into the TEE for matching will be hashed with SHA-256. SHA-256 is a specified secure hashing algorithm under NIST FIPS 180-4[4] which adds an additional layer of protection from unauthorised disclosure/access by third parties

    b. Implementation of TEE solution. The use of TEE solution ensures that data is executed in a secure computing environment and inaccessible to all parties (i.e., Advertiser, Publisher, TEE solution provider), and that only authorised codes/algorithms can be run (and subsequently verified) within the TEE. It

[3] In parallel, SPH Media would have worked with the PAS to ensure that right segment IDs are tagged to any advertisement campaigns

[4] NIST FIPS 180-4 Secure Hash Standard specifies hash algorithms that can be used to generate digests of messages. The digests are used to detect whether messages have been changed since the digests were generated.

leverages technology for hardware-based isolation and attestation to ensure data and code integrity even if the main operating system is compromised.

17. Nevertheless, the Publisher and Advertiser would be considered to have used personal data by uploading their hashed email lists to the TEE and generating both the seed list and the lookalike customer list, and the Data Protection Provisions under the PDPA will apply (e.g., Consent Obligation, Protection Obligation). As part of this POC, both the Advertiser and Publisher may consider relying on PDPA's Business Improvement Exception (BIE) to use its customers' personal data without consent given that the intent is (i) to learn and understand its customers preferences and (ii) to personalise relevant goods and services for the users in the "lookalike" list.

Whether the SPH ID and customer segment IDs that the Publisher shares with the DMP constitute personal data under the PDPA.

18. PDPC is of the view that the SPH IDs and customer segment IDs shared by the Publisher with the DMP will not likely constitute personal data under the PDPA due to the very low risk of re-identification of individuals. While unique to individuals, both the SPH ID and Segment ID are indirect identifiers created by the Publisher which will be meaningful only to the Publisher. The DMP is unlikely to have the ability to link these identifiers to its own customer records without additional information about this individual. In addition, PDPC notes that SPH IDs are temporary IDs related to individuals, which will expire after 30 days from its creation.

19. To reduce the likelihood of re-identification, the DMP partner has implemented an additional transformation step to convert the Publisher-segment ID into DMP-segment ID before sharing with the PAS. This prevents the exposure of the Publisher-segment ID to additional parties, which in turn reduces the likelihood of the Publisher-segment ID from being linked or associated with specific individuals where more information may be gleaned from such categorisation of profiles/segments.

Additional safeguards that can be implemented

20. PDPC is of the view that there are additional safeguards that Parties may wish to consider implementing as part of this POC:

a. Prior to data processing within the TEE, Parties may consider using a shared common salt in the hashing process to prevent rainbow table attacks (pre-computed hash tables). Parties should also ensure they are using current implementations of SHA-256 from well-maintained cryptographic libraries and regularly update these libraries to incorporate security patches and implementation improvements.

b. While TEE solutions offer secure infrastructure, organisations that engage third party TEE solution providers remain responsible under the PDPA for the data processing within the TEE. Both the Publisher and Advertiser may wish to conduct due diligence to assess and validate if the data security afforded by TEE solution, and the data governance policies and practices of the solution provider are sufficient and appropriate for their business needs.

Results and Next Steps

21. A key metric that the GWM wanted to measure to determine the success of the advertisement campaign was click-through and continued navigation through the website (i.e. click through onto website to find out more about certain products). Continued navigation following click through would ensure that there was genuine interest instead of clicking through and immediately navigating away.

22. Two parameters that could impact success criteria are

a. **Availability of "Fresh" SPH audience** (Depending on how often SPH Media refreshed the collected data, e.g. 30, 60, 90 days)
b. **Intended size of LAL** (Size of list could be chosen as a % of SPH's customer database)

23. As there was no benchmark or objective guide to how much both parameters should be adjusted, AB testing was conducted to find the right configuration optimal for the POC:

a. **Scenario 1:** SPH Media's audience availability - refreshed every 60 days; Intended lookalike audience size – 10% of SPH customers. Both impressions and click through rate were found to be low if longer days of data were used, refresh rate was low and intended size was large
b. **Scenario 2**: SPH Media's audience availability – refresh every 30 days; Lookalike size – 3% of SPH customers. This configuration was found to be optimal, where there was

90% “lift” given increased freshness. A lower intended audience size also meant that there was a higher level of accuracy of selected audience with the profile of the initial seed list.

24. In summary, the higher the refresh rate and smaller the size of the intended lookalike audience produced individuals that were more likely to click through and continue to navigate the websites. Whilst this specific configuration worked for the GWM, different advertisers may have different thresholds of what may be deemed acceptable; therefore, tweaking would be necessary for different advertisers.

25. The use of a TEE has shown potential to both the GWM and SPH Media, as a viable means to facilitate new data collaborations in the advertisement tech space. As a technology solution the TEE is user-friendly and can be set up quickly, setting a precedence for potential collaborations with other partners in the future.