PRACTICAL GUIDANCE TO TIKTOK

Background

1. TikTok Pte Ltd ("**TikTok**") is a social media company that offers advertisement publishing services for businesses. As part of a proof-of-concept ("**POC**"), TikTok has developed an open-source solution "PrivacyGo" that uses various privacy enhancing technologies ("**PETs**"), to measure the success of targeted advertising while ensuring the privacy of individual users.

2. Key stakeholders involved in this POC are as follows:

   a. TikTok (**"Publisher"**) – holds identifiers (e.g., mobile phone numbers, email addresses, ID for advertisers ("**IDFA**")) corresponding to its users who have viewed the advertising campaign.

   b. A mobility service provider (**"Advertiser"**) – holds identifiers (e.g., mobile phone numbers, email addresses, IDFA) and their corresponding spending amounts on the advertiser's website/application when a purchase has been made.

3. A brief description of the key steps involved in the POC is as follows:

   a. **Data preparation.** Publisher and Advertiser will each identify relevant datasets to be used for the analysis. Both parties will each add randomly generated "dummy records" to their datasets and shuffle the rows of the data records to garble the original orientation.

   b. **Encryption of datasets.**

      i. Identifiers within both Publisher and Advertiser datasets are double encrypted using elliptic curve cryptography ("**ECC**")[1] (i.e. customer IDs are encrypted with both parties' private keys).

      ii. Accompanying attributes of each data record (e.g. impression timestamp, conversion timestamp and conversion value) are also

[1] ECC is an asymmetric cryptographic technique based on the arithmetic operations of elliptic curves over finite fields. It allows for faster processing and lower key bit count vs. other forms of asymmetric cryptography such as RSA encryption that relies on factoring of large numbers

homomorphically encrypted (“**HE**”) using the Pallier encryption algorithm[2].

c. **Finding common customers.** Private Set Intersection (“**PSI**”) is then conducted to find common customers. At the end of this step, the Publisher and Advertiser would each hold a dataset with only matched customers and their accompanying attributes in the encrypted form.

d. **Generating secret shares for computation.** The encrypted accompanying attributes in the datasets will be “shredded” into two secret shares (i.e. encrypted value of the attribute “impression timestamp” would be “shredded” into secret share 1 and 2, with secret share 1 held by Publisher and secret share 2 by the Advertiser). Publisher and Advertiser would each hold an encrypted secret share of all 3 accompanying attributes.

e. **Compute total conversion value.** Based on the agreed criteria between parties in determining what would constitute a conversion (e.g. parties agree that difference between impression timestamp and conversion timestamp must be < 10 seconds to constitute a conversion), relevant records meeting these criteria will be used for the computation. The computation then takes place to determine the total conversion value for the advertising campaign based on the conversion values of the relevant records. The computation hides input from both parties through Secure Multi-Party Computing (“**SMPC**”)’s additive secret sharing technique which is implemented using the ABY framework[3]. During the computation of the aggregated conversion value through SMPC, Differential Privacy (“**DP**”), i.e. noise, is also added. This is accomplished through an Oblivious Transfer (“**OT**”) protocol[9] to mathematically distribute noise between the parties for addition to their respective secret share values.

4. TikTok sought Practical Guidance from the Personal Data Protection Commission (“**PDPC**”) on whether the data shared between the Advertiser and Publisher in the POC constitutes personal data such that the data protection obligations under the Personal Data Protection Act 2012 (“**PDPA**”) would apply.

PDPC’s assessment

Whether there is sharing of personal data between Advertiser and Publisher in the POC

[2] Pallier encryption is an asymmetric homomorphic cryptographic technique defined in the ISO/IEC 18033-6 standard

[3] ABY Framework allows mixed-protocol SMPC schemes based on Arithmetic sharing, Boolean sharing, and Yao’s garbled circuits by making available best practice solutions in a secure two-party computation setup - http://dx.doi.org/10.14722/ndss.2015.23113

5. Personal data is defined in Section 2 of the PDPA to refer to data, whether true or not, about an individual who can be identified (a) from that data; or (b) from that data and other information to which the organization has or is likely to have access.

6. PDPC is of the view that in this POC, there is no serious possibility that each party will be able to identify individuals from the data that it receives from the other party, or that a third party who obtains the shared data will be able to identify individuals from it. Hence, the data sharing between the parties does not constitute a disclosure or collection of personal data under the PDPA for the disclosing party and the receiving party respectively. In coming to this view, the PDPC considered the following:

a. The adding of "noise" by adding dummy records and shuffling records;

b. The use of effective encryption techniques to prevent recipients from gaining access to information or insights relating to the common customers;

c. That each party keeps its encryption keys and secret share "shred" confidential from the other party and any other persons; and

d. That the final output is an aggregated conversion value that is obtained through Secure Multi-Party Computing (SMPC) where neither party would be able to attribute or link to any identifiable individual.

7. TikTok and the Advertiser should ensure that the measures and security arrangements implemented to prevent the risk of re-identification and protect against data protection threats remain effective and up-to-date. This includes keeping the PETs used in the POC updated with prevailing industry-recognised processes and standards, and ensuring "cryptographic agility" by replacing cryptography algorithms that are found to be vulnerable.

END OF DOCUMENT