Privacy-Preserving Multi-Touch Attribution at TikTok

Abhishek Tiwari

Citation: *A. Tiwari*, ”Privacy-Preserving Multi-Touch Attribution at TikTok”, Abhishek Tiwari, 2024. doi:10.59350/rhrwf-tbs74

Published on: October 22, 2024

Multi-touch attribution is considered as holy grail in advertising industry. As advertisers are targeting users with multiple advertisements across different platforms and publishers, understanding how each of these touch points contributes to conversion is crucial—but this understanding has traditionally come at the cost of user privacy. In a previous blog post(see [1]) we briefly covered Mozilla's Private Attribution API(see [2]) that aims to provide advertisers with conversion data without compromising user privacy. In recent PEPR '24 Conference, TikTok team talked about how they are integrating Differential Privacy (DP) into their ad measurement systems (see [3]). This blog post is notes from the talk.

The Challenge of Multi-Touch Attribution

Modern advertising campaigns span across multiple platforms and publishers, creating a complex web of user interactions. Consider this common scenario: a user sees a search ad, later encounters a social media advertisement, and finally converts after clicking a display ad on a news website. Traditionally, tracking these interactions requires cross-site user tracking through technologies like pixels and cookies, raising significant privacy concerns.

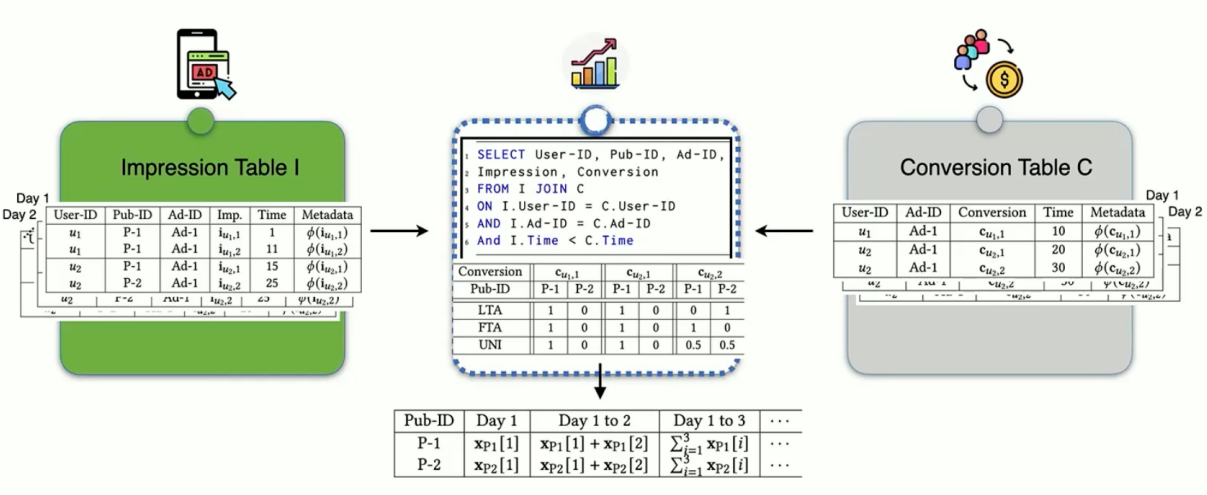

Figure 1: An example of ad measurement tables

The complexity of this tracking becomes apparent when considering the vast amount of data being collected and correlated across different platforms. Each user interaction generates multiple data points, from initial ad impressions to final conversion events, all of which need to be linked together to create a coherent understanding of the user journey.

Current Attribution Landscape and Privacy Concerns

Traditional multi-touch attribution systems rely on combining two critical data sets. The first is impression data, which records when users encounter ads across different platforms. A typical impression record might show that a user viewed a Nike advertisement at 1 PM on May 3rd. The second dataset is conversion data, tracking user actions on advertiser websites. This might include records showing that the same user purchased Nike sneakers for $60 on May 6th.

The correlation of these datasets enables advertisers to calculate return on investment (ROI) and distribute credit across different advertising touch points. This information is crucial for optimising campaign spending across platforms and understanding the effectiveness of different advertising channels.

However, this approach faces mounting challenges in today's privacy-conscious environment. Regulatory pressure, exemplified by GDPR and similar regulations, has placed strict limitations on cross-site tracking. Platform-level changes, such as Apple's App Tracking Transparency initiative, have further restricted data collection capabilities. Perhaps most significantly, there's growing user awareness and resistance to personal data collection practices.

Input vs. Output Privacy

The current privacy challenges cannot be addressed through simple anonymization or data collection minimisation. Modern advertising measurement requires systems that provide formal privacy guarantees while maintaining utility. This complex requirement demands a sophisticated approach to both input and output privacy protection.

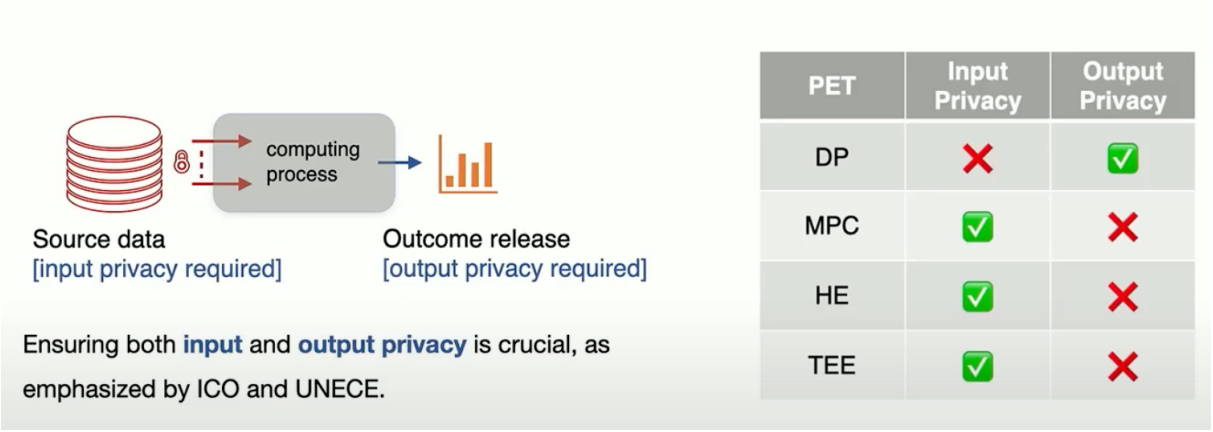

Figure 2: Input Privacy vs. Output Privacy. Credits TikTok team.

Input privacy challenges encompass the protection of raw user data during collection and processing,

the secure matching of cross-platform data, and the prevention of unauthorised access to individual user journeys. These challenges are fundamental to building trust in the measurement system.

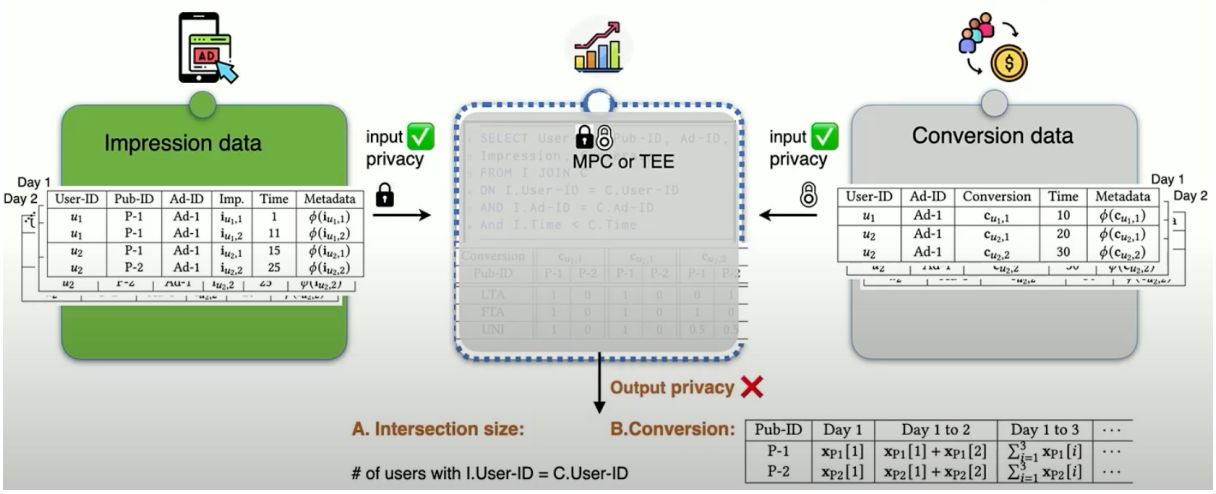

Figure 3: Input privacy guarantee but no output privacy guarantee

Output privacy challenges are equally important but often overlooked. They include ensuring that aggregate reports don't leak individual user information, protecting intermediate computation results, and maintaining user-level privacy guarantees across multiple reports. Without proper output privacy protection, even systems with strong input privacy can leak sensitive information through their results.

The TikTok team demonstrated that effective privacy-preserving attribution requires a comprehensive approach combining multiple privacy-enhancing technologies. At its core, their attribution framework implements dual privacy protection through both input and output safeguards.

When an ad publisher and advertiser need to match their user data for attribution purposes, they face what the TikTok team calls a "dual curiosity" problem. The ad publisher wants to know if their users converted on the advertiser's platform, as this information could help improve their advertising models. Similarly, advertisers are interested in understanding user behavior on the publisher's platform. This mutual interest in each other's user data creates a complex privacy challenge that goes beyond simple input privacy protection.

Input privacy is achieved through the implementation of secure multi-party computation (MPC) and trusted execution environments (TEE). These technologies ensure that raw data remains protected throughout the processing pipeline. Output privacy is guaranteed through the careful application of differential privacy, with optimised noise addition techniques that maintain utility while providing strong privacy guarantees.

Real-time reporting optimization represents another crucial component of the framework. The chal-

lenge lies in balancing the need for frequent reporting with privacy requirements. Through sophisticated composition techniques, the TikTok team has developed methods to minimize the impact of noise addition while maintaining strong privacy guarantees. This enables advertisers to make timely decisions based on accurate data without compromising user privacy.

The Matching Size Vulnerability

TikTok team's also shared insight about a subtle but significant privacy leak through matching size information. Consider a scenario where an ad publisher has impression data showing a user clicked on a Nike shoes advertisement at 1 PM on May 3rd. The publisher might be curious about whether this user completed a purchase on the advertiser's website. While input privacy protections like TEE or MPC might seem sufficient, the research revealed a key vulnerability.

Intersection Cardinality Protection

Traditional Private Set Intersection (PSI) protocols based on elliptic curve cryptography typically reveal this intersection size as part of their operation. Recent studies in [4] and [5] revealed these PSI protocols disclosing intersection size might unintentionally leak membership information about the parties' sets. The TikTok team's research confirms that even when using secure computation methods like PSI, the mere revelation of intersection cardinality—the number of users that match between platforms—is typically revealed to both parties (publishers and advertisers) can lead to significant privacy leaks.

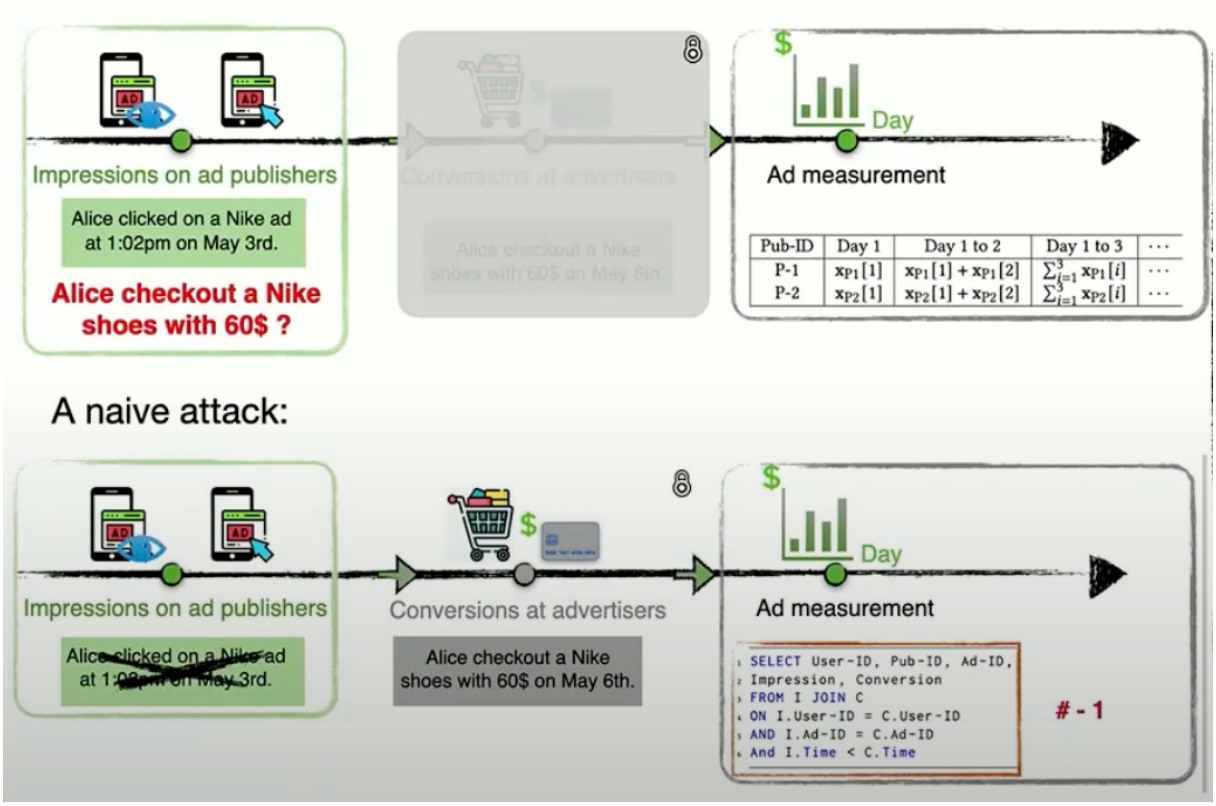

Figure 4: Differential attacks

The team demonstrated that this seemingly innocuous information could be exploited through what they call “differential attacks.” The attack works as follows: An ad publisher can systematically remove specific user impression data from their dataset and observe changes in the intersection size during subsequent matching operations. If removing a particular user’s impression data causes the intersection size to decrease by one, this reveals that the user had a matching conversion event on the advertiser’s side. The TikTok team’s tests using real-world data demonstrated membership leakage rates of up to 0.4%, which was further increased to approximately 2.5% with optimised attack strategies.

Two-Party Privacy Protection System

A key take away from TikTok’s privacy-preserving framework is its approach to two-party dual-sided differential privacy in the context of advertising measurement. This system addresses a critical challenge in cross-platform advertising: both publishers and advertisers have legitimate interests in protecting their respective user data while still enabling effective measurement.

MPC-DualDP

To address this vulnerability, the team used an approach combining distributed differential privacy with secure multi-party computation. The solution introduces carefully calibrated dummy data before the matching process begins. Both parties share a common dummy data size using a synchronised random seed, generating a set of dummy records that get shuffled with the real data. By appending these dummy records to the ID columns before matching, the system can effectively mask the true intersection sizes while maintaining the accuracy of the final measurement results.

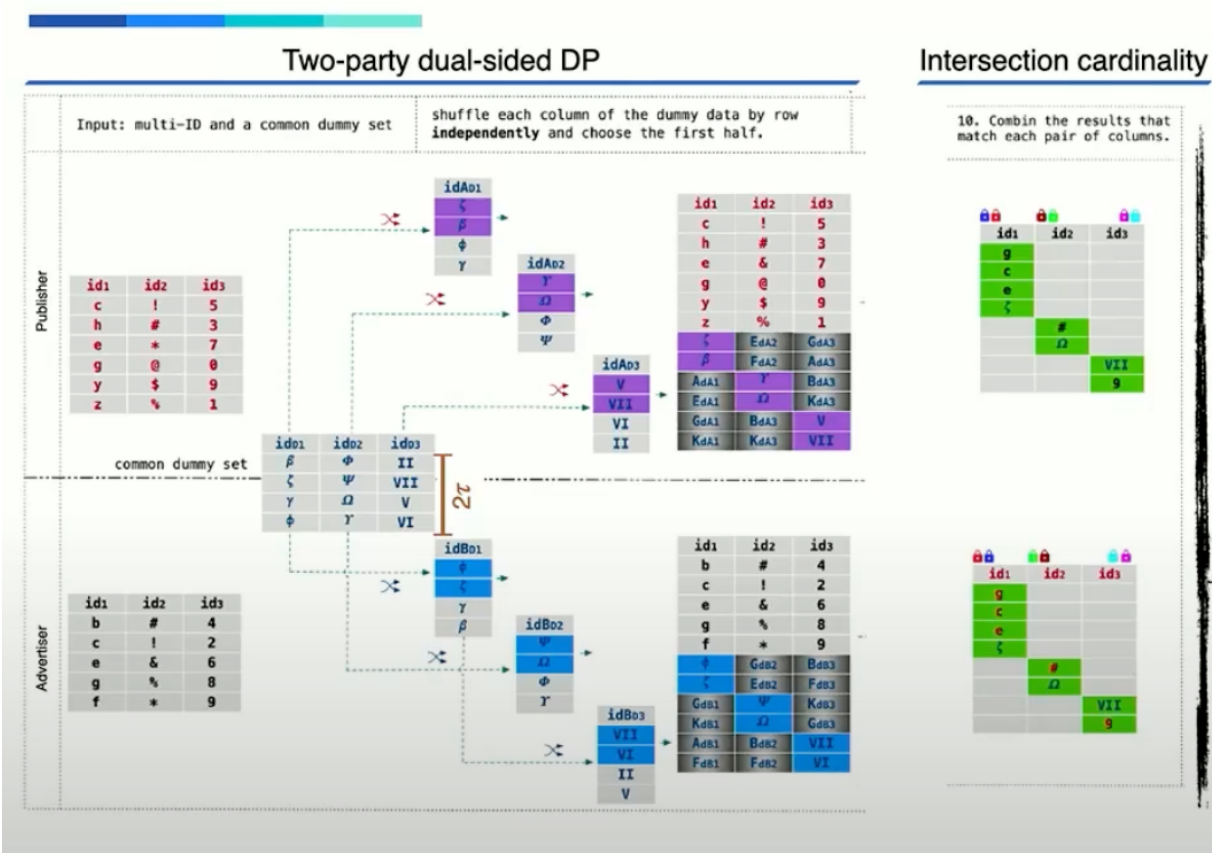

Figure 5: Distributed DP for MPC intermediate results

This approach provides several advantages. It prevents membership inference through intersection size analysis (as long as one of the parties are honest). More importantly, it maintains measurement accuracy by properly handling dummy data.

However, the solution also introduces new challenges, particularly in terms of communication and computation overhead. The addition of dummy data increases both the data transfer requirements and the computational complexity of secure multi-party computations.

Optimising Performance Through Tight Composition

To address the overhead challenges, the TikTok team developed novel approaches to tight differential privacy composition. Through careful analysis and optimisation, they identified tighter composition bounds that significantly reduced the required dummy data size while maintaining the same privacy guarantees. These bound minimise the total noise added over the entire ad campaign, improving accuracy compared to adding independent noise. Experimental results showed that this optimisation could substantially reduce both communication and computation overhead, making the system more practical for real-world deployment.

The team's empirical evaluation demonstrated that their tight composition analysis could achieve the same privacy guarantees with significantly less dummy data compared to standard advanced composition techniques. This improvement is particularly important for streaming or real-time advertising measurement scenarios where computational efficiency is crucial.

PrivacyGo

The TikTok team has made their privacy-preserving technologies available through PrivacyGo, an open-source project that implements these innovative approaches to privacy-enhancing technologies (PETs). The project includes several key components that directly address the challenges discussed above. The DPCA-PSI protocol specifically tackles the intersection size leakage problem in ECDH-style PSI protocols, particularly crucial for multi-ID matching scenarios. The Privacy-Preserving Ads Measurement (PPAM) component leverages DPCA-PSI to enable private ad measurement with encrypted match keys and differential privacy guaranteed matched group sizes. Additionally, the MPC-DualDP component provides a distributed protocol for generating shared differential privacy noise in a two-server setting, addressing the need for collaborative noise generation in secure computations.

{{< youtube OX4X78JRuf4 >}}

Conclusion

TikTok's framework demonstrates that it's possible to achieve both effective attribution and strong user privacy protection through the careful application of privacy-enhancing technologies and optimized implementation strategies.

References

[1] A. Tiwari, “Privacy Preserving Measurement,” 2024, *Abhishek Tiwari*. doi: 10.59350/fnpfz-3v466.

[2] “Outline of a Private Attribution Measurement API [Public],” Mozilla. [Online]. Available: https://docs.google.com/document/d/1QMHkAQ4JiuJkNcyGjAkOikPKNXAzNbQKILqgvSNIAKw

[3] J. Du and S. Zhang, “Designing for User Privacy: Integrating Differential Privacy into Ad Measurement Systems in Practice,” presented at the PEPR24, 2024. Available: https://www.usenix.org/conference/pepr24/presentation/du

[4] X. Guo, Y. Han, Z. Liu, D. Wang, Y. Jia, and J. Li, “Birds of a Feather Flock Together: How Set Bias Helps to Deanonymize You via Revealed Intersection Sizes,” in *USENIX Security Symposium*, 2022. Available: https://www.usenix.org/conference/usenixsecurity22/presentation/guo

[5] J. Powar and A. R. Beresford, “SoK: Managing Risks of Linkage Attacks on Data Privacy,” *PoPETs*, 2023, doi: 10.56553/popets-2023-0043.