Differentially Private Ad Conversion Measurement

John Delaney
Google
MA, USA
johnidel@google.com

Badih Ghazi
Google
CA, USA
badihghazi@gmail.com

Charlie Harrison
Google
TX, USA
csharrison@google.com

Christina Ilvento
Google
CA, USA
cilvento@google.com

Ravi Kumar
Google
CA, USA
ravi.k53@gmail.com

Pasin Manurangsi
Google
Thailand
pasin@google.com

Martin Pál
Google
Switzerland
mpal@google.com

Karthik Prabhakar
Google
CA, USA
kprabhakar@google.com

Mariana Raykova
Google
NY, USA
marianar@google.com

ABSTRACT

In this work, we study *ad conversion measurement*, a central functionality in digital advertising, where an advertiser seeks to estimate advertiser website (or mobile app) conversions attributed to ad impressions that users have interacted with on various publisher websites (or mobile apps). Using differential privacy (DP), a notion that has gained in popularity due to its strong mathematical guarantees, we develop a formal framework for private ad conversion measurement. In particular, we define the notion of an operationally valid configuration of the attribution rule, DP adjacency relation, contribution bounding scope and enforcement point. We then provide, for the set of configurations that most commonly arises in practice, a complete characterization, which uncovers a delicate interplay between attribution and privacy.

KEYWORDS

differential privacy, online advertisement, measurement

1 INTRODUCTION

Over the last two decades, numerous attacks have illustrated the privacy risks associated with the release of (aggregated, de-identified) information, across various domains [24, 30, 36]. This has led to the introduction of differential privacy (DP) [17, 18], a rigorous mathematical notion that quantifies the privacy loss of users against arbitrary adversaries observing the algorithm's output (and regardless of the data contributed by other users). While DP has been deployed in several fields including census data release [3], learning frequently typed words [8], and collection of telemetry data [16], there has not been any formal study of it for ad conversion measurement—a core functionality in the digital advertising space—that we undertake in this work.

We now provide a quick overview of the problem space. Let the term *conversion* refer to a valuable action (e.g., a purchase, add-to-cart, newsletter sign up, etc.) on the advertiser website[1] and the term *impression* refer to an ad engagement (e.g., a click on an ad or a view of an ad) by the user on the publisher website (and not merely an ad fetch). In *(ad) conversion measurement*, an advertiser running campaigns on multiple publishers aims to measure the performance of various campaign slices.[2] This includes estimating conversion counts and values (the latter can be numerical, e.g., in US dollars) for different settings of impression and conversion features (e.g., Example 1 below). This information can then be used by the advertiser to estimate the conversion rate and the return on ad spend, to optimize the purchase of ad impressions on guaranteed selling channels, and/or to power real-time bidding systems [12]. The ease of measuring conversions on advertiser websites (or apps) and attributing them to impressions that the same user had interacted with on publisher websites (or apps) is an important reason why online advertising is significantly more efficient than more traditional forms of advertising (e.g., print media, radio and television).

Privacy is a crucial consideration in conversion measurement. Ad measurement is indeed fundamentally a cross-site functionality, involving multiple publishers and advertisers. For the last two decades, non-private approaches to the problem have allowed third parties to track users across websites; this has been enabled on the Web by third-party cookie technology. However, in recent years, a consensus has emerged that these approaches are too invasive for users, and that new privacy-preserving methods for supporting various ad use cases are critically needed. This has led to the decision to deprecate third-party cookies by several browsers including Apple Safari [40], Mozilla Firefox [43], and Google Chrome [34]. Consequently, multiple efforts are currently underway by many browsers, platforms and industry groups to design privacy-preserving APIs

Proceedings on Privacy Enhancing Technologies 2024(2), 124–140
© 2024 Copyright held by the owner/author(s).
https://doi.org/10.56553/popets-2024-0044

[1] Although in this paper we mostly discuss advertiser and publisher *websites*, these could alternatively be *mobile apps*, and the same treatment holds.

[2] A campaign consists of a subset of impressions, associated with the same campaign ID. Eeach impression is also associated with attributes over which an analyst can slice.

that aim to support ad conversion measurement functionalities. [3] These APIs and proposals include the Interoperable Private Attribution (IPA) proposed by Mozilla and Meta, [37], Masked LARk from Microsoft [32], the Privacy Sandbox Attribution Reporting API (ARA) on Chrome [29] and Android [7], Private Click Measurement (PCM) on Safari [41], SKAdNetwork on iOS [1] as well as the recent proposal [42] from Apple. While most of these efforts seek to guarantee DP in order to ensure that sensitive user information cannot be recovered from the output of the API, they still lack an end-to-end formal DP framework. The goal of our work is to develop such a framework, so that proposals in this space are built on a solid mathematical foundation.

2 MOTIVATION, SETUP & CONTRIBUTIONS

2.1 Ad Conversion Measurement System

We next define the main components of an ad conversion measurement system, and discuss the central concept of *attribution*.

2.1.1 Basic Definitions.

An *impression* event consists of (i) a timestamp, (ii) a publisher id, (iii) an advertiser id, (iv) a user id, and (v) metadata associated with the impression (e.g., type indicating if it is a click or view, format indicating size of ad, etc.).

A *conversion* event consists of (i) a timestamp, (ii) an advertiser id, (iii) a user id, and (iv) metadata associated with the conversion (e.g., a type indicating if it is a purchase or a sign up, a value if it is a purchase etc).

The entities that are involved in ad conversion measurement are:

- *Publisher*: a website on which ad impressions take place.
- *Advertiser*: a website where conversion events take place.
- *Ad tech:* entity (often a third party) used by an advertiser to buy, manage, and measure their digital advertising.

The most common type of functionality in ad conversion measurement can be thought of as a two-step process: first, conversions are assigned to one or more impressions using a fixed *attribution rule*, and then queries are executed on the resulting *attributed dataset.* On a high level, the attribution rule determines how the credit from a conversion is to be divided over the different ad impressions corresponding to the same advertiser and the same user as the conversion. The attributed dataset consists of (impression, conversion) pairs, each corresponding to an attribution; optionally the pair is also associated with a (fractional) credit. (In the cases where all credits are equal, we omit them from the attributed dataset notation for simplicity.) The queries used in the second step include counting the number of conversions attributed to a subset of impressions (and the corresponding conversion rate), as well as the total return on ad spend for that subset.

Example 1. *An example of the set of impression features can include the publisher website, the advertiser website, the ad engagement type (click or view), as well as its time and geographical location. An example set of conversion features can include the advertiser website, the conversion time, and the conversion type (e.g., add-to-cart, purchase, email sign-up). Thus, examples of ad conversion measurement queries include asking for:*

- *The number of conversions on advertiser1.com attributed to ad impressions on publisher1.com which occurred in the UK.*
- *The total value (in US dollars) of purchases on advertiser1.com taking place on August 31st and attributed to ad impressions on publisher1.com.*

,

We next discuss the attribution step in more details.

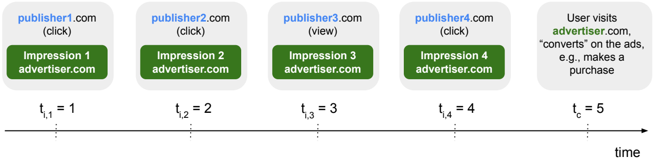

Figure 1: Example Attribution Path. In this case, a user interacts with four ad impressions from the same advertiser, but on four different publishers. The third of these interactions is a view, whereas the others are clicks. For simplicity, we assume that the four impressions and the subsequent conversion occur at equally spaced times.

2.1.2 Attribution Rule.

A key component of any conversion measurement system is the attribution rule. Specifically, consider the setting, illustrated in Figure 1, where a user is exposed to several impressions on multiple publishers before converting on the advertiser website. Which of these impressions should get the credit for driving the conversion? This attribution question has been the subject of numerous studies (see, e.g., [25] and the references therein). In practice, the most popular attribution rules include last-touch attribution (LTA), first-touch attribution (FTA), uniform attribution (UNI), and exponential time decayed attribution (EXP). As the names suggest, first- and last-touch attribution assign all the credit to the first and last impressions on the attribution path, respectively, whereas uniform attribution divides the credit evenly among all the impressions, and exponential time decayed attribution assigns to each impression a credit that decays exponentially as a function of the time gap to the conversion. For an illustration of these different attribution rules for the example in Figure 1, we refer the reader to Table 1.

Attribution Rule	Credit to Click on publisher1	Credit to Click on publisher2	Credit to View on publisher3	Credit to Click on publisher4
LTA	0	0	0	1
FTA	1	0	0	0
UNI	0.25	0.25	0.25	0.25
EXP	0.0667	0.1333	0.2667	0.5333

Table 1: Example Credit Attribution.

We next explain how the values in Table 1 were computed. For LTA, all the credit goes to the “last touch”, which is the user click

[3] We note that, on a high-level, the goals of ad conversion measurement are not at odds with privacy; in fact, advertisers typically want to change their ad spend allocation based on patterns that are significant enough to generalize to large groups of users (as opposed to overfitting to individual user actions).

that took place on publisher4.com. For FTA, all the credit goes to the "first touch", which is the user click on publisher1.com. For UNI, the credit is split equally among the 4 ad impressions (3 clicks and 1 view). For the EXP attribution rule, we assume that the half-life parameter[4] is 1, and that the times of the successive impressions and conversion are all separated by 1 time unit, as shown in Figure 1; thus, the credit assigned to ad impression i is $0.5^{5-i}/(0.5 + 0.5^2 + 0.5^3 + 0.5^4)$ for each $i \in \{1, 2, 3, 4\}$.

In practice, different ad techs might offer advertisers support different attribution rules to be used when measuring conversions. For instance, while FTA might be desirable for some advertised products, LTA might be natural in other settings. In this work, we consider all of the basic and most popular attribution rules that were mentioned above.[5]

2.2 Privacy Model

We start by describing the threat model for privacy-preserving ad conversion measurement.

2.2.1 Threat Model. To illustrate the threat model, we consider the setting where an untrusted third party (e.g., an ad tech as defined in Section 2.1.1) would like to run queries on a conversion measurement dataset pertaining to multiple publishers and advertisers. We assume that a central trusted curator is in charge of executing the query on the dataset and sending the output to the ad tech; the trusted curator could, e.g., be a web browser or mobile platform, a trusted third party, or a secure multi-party computation protocol. The goal is to protect sensitive information (e.g., related to a single impression, conversion, or user?) from being leaked to the ad tech through the output of the protocol. Examples of sensitive information in this setting include app or web browsing history (e.g., did the user recently visit a sensitive website), or the shopping history of the user. To protect the data of individual users from leaking through an output released to an untrusted third party, DP has become the gold standard. It has been suggested as a primary privacy guardrail in multiple industry proposals for privacy-preserving ad measurement systems. Therefore, in this work, we study DP as the desired privacy guarantee on the output of the ad measurement system.

Remark 1. *(Robustness to Side Information) We note that in some cases, depending on browser constraints, user sign-ins, and business arrangements, the third party could have a prior partial view of the dataset. E.g., it could know the set of all impressions (across all users but without knowing the associated user id for each impression), or the set of all conversions, or both. Nevertheless, the protection offered by DP would still be meaningful in this setting, since DP is robust to the presence of side information.*

2.2.2 Differential Privacy Ingredients.

Adjacency Relation. On a high level, DP dictates that the distributions of the output of a (randomized) algorithm on two *adjacent* conversion measurement datasets are statistically indistinguishable (see Section 3.1 for a formal definition). It is thus necessary to specify the adjacency relation (a.k.a. privacy unit) to which the DP definition applies. Due to the highly fragmented nature of conversion measurement datasets (with the conversion taking place on one of the advertisers and the impressions taking place on different publishers), it turns out there are multiple natural alternatives for defining the adjacency relation, with subtle implications on the privacy-utility trade-offs. The different possibilities are listed in Table 2. For each adjacency relation, the allowed difference between two adjacent datasets consists of all user engagements that only belong to a single tuple in the adjacency relation. For instance, for the *user* × *advertiser* relation, two adjacent datasets can differ on the set of all impressions of Alice associated with advertiser1.com (and shown on any publisher), along with all conversions of Alice on advertiser1.com; this is because all of these engagements are only associated with a single (user, advertiser) tuple, namely, (Alice, advertiser1.com). On the other hand, for the *user* × *publisher* × *advertiser* relation, the difference consists of all impressions associated with a fixed publisher and a fixed advertiser (e.g., all impressions of Alice shown on publisher1.com, and that are associated with advertiser1.com); note that the conversions associated with the advertiser (e.g., advertiser1.com) are not included here, as they can be associated with multiple other publishers (e.g., publisher2.com), and are thus also related to other (user, publisher, advertiser) tuples in the adjacency relation, e.g., (Alice, publisher2.com, advertiser1.com) is such a tuple.

Remark 2 (Intuitive Interpretation of the Different Adjacency Relations)**.** *The different adjacency relations described in Table 2 offer a spectrum of possible privacy guarantees. On one end, the conversion (respectively, impression) adjacency relation seeks to protect a single conversion (respectively, impression) from being leaked by the output of the algorithm. On the other end, the user relation protects any information related to all the impressions and conversions of any user from being leaked. The other notions can be seen as interpolating between these two ends. E.g., the user × advertiser relation protects all the user's impressions and conversions pertaining to a single advertiser, but it does not necessarily protect information that can be deduced by observing the user's impressions and conversions across multiple advertisers. In particular, under this user × advertiser relation, observing the output of the privacy-preserving ad conversion measurement the system would not substantially increase an attacker's ability to distinguish whether Alice had any attributed conversions associated with advertiser1.com. The choice of the adjacency relation depends on the privacy protection that the system designer seeks to guarantee, and the associated utility trade-offs that they are willing to accept. E.g., the user × advertiser adjacency can be natural in some settings as it would prevent the ad-tech associated with a publisher from learning that a visitor to the publisher later visited a (sensitive) advertiser site. On the other hand, a user × publisher adjacency relation can prevent an ad-tech associated with the advertiser from learning about the actions of a converting user on publisher sites.*

Remark 3 (Time Dimension in Adjacency Relations)**.** *In practice, it is common to include the time dimension in some adjacency relations, in particular, the user, user × publisher and user × advertiser ones. Moreover, DP composition can be used to handle the case where the time dimension in the adjacency relation can cover multiple releases of the output of the system on different subsets of the dataset.*

[4] We refer the reader to Section 3.2 for the formal definition.

[5] In addition to the aforementioned attribution rules, we also cover position-based, U-shaped, and impression-priority attribution; see Section 3.2 for the formal definitions.

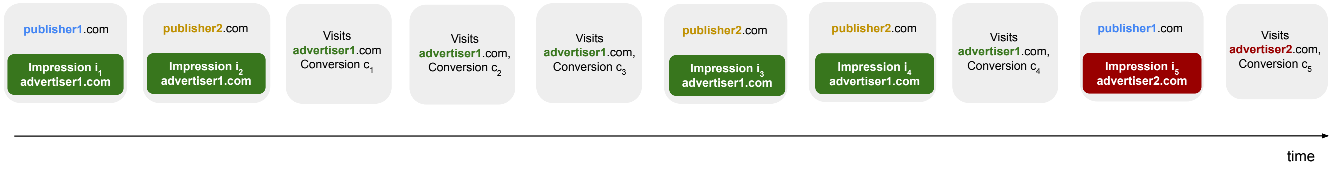
publisher1.com
Impression i_1 advertiser1.com
publisher2.com
Impression i_2 advertiser1.com
Visits advertiser1.com Conversion c_1
Visits advertiser1.com, Conversion c_2
Visits advertiser1.com, Conversion c_3
publisher2.com
Impression i_3 advertiser1.com
publisher2.com
Impression i_4 advertiser1.com
Visits advertiser1.com, Conversion c_4
publisher1.com
Impression i_5 advertiser2.com
Visits advertiser2.com, Conversion c_5
time

Figure 2: Attribution Path for Multiple Publishers and Advertisers. In this example, the user interacts with ads on two publishers, and converts on the two corresponding advertiser websites.

Contribution Bounding Scope. To ensure privacy for a given adjacency relation, a core component of DP systems is the notion of a contribution bound. Note that, in the case of ad conversion measurement, the number of interactions can naturally be unbounded. For example:

- An impression could lead to multiple conversions.
- Different impressions can be on the attribution path of the same conversion.
- The same user can be shown many impressions on the same or on different publishers, and could convert multiple times on the advertiser.

See Figure 2 for an example. To be able to guarantee a bounded privacy leakage, the contributions of the interactions to the computed function should be restricted by a certain *contribution bound*. The *contribution bounding scope* is the set of user interactions that share the same contribution bound. The scope can be any one of the options listed in Table 2 for the adjacency relation. We will consider in this work the natural setting where the contribution bounding scope is the same as the adjacency relation. It turns out that each adjacency relation / contribution bounding scope can have a different interplay with the attribution rule, and with the resulting DP guarantee. Moreover, as we will see shortly, some will be easier to operationalize than others.

Contribution Bound Enforcement. One particular choice, that turns out to be important, is whether the contribution bound should be enforced before (*pre-attribution contribution capping*) or after (*post-attribution contribution capping*) the attribution rule is applied to the dataset. In the former case, only impressions belonging to a scope with a non-zero remaining contribution bound can be considered on the attribution path of a conversion. In the latter case, a conversion could get attributed to an impression with a remaining contribution bound of 0, only to get discarded at the contribution bound enforcement step (without the attribution falling back to any other impression on the path). In other words, pre-attribution contribution bound enforcement limits the number of impressions or conversions that belong to the contribution bounding scope and that can *enter* the attribution rule. By contrast, post-attribution enforcement limits the number of post-attribution (impression, conversion) pairs that belong to the contribution bounding scope.[6] See Table 3 for the attributed dataset resulting from applying post-attribution contribution bounding with a bound of 2 to the dataset given in Figure 2 and with the LTA rule.

Contribution Bounding	Attributed Dataset
None	(i_2, c_1) (i_2, c_2) (i_2, c_3) (i_4, c_4) (i_5, c_5)
Impression (Contribution Bound = 2)	(i_2, c_1) (i_2, c_2) (i_4, c_4) (i_5, c_5)
User × Advertiser (Contribution Bound = 2)	(i_2, c_1) (i_2, c_2) (i_5, c_5)
User (Contribution Bound = 2)	(i_2, c_1) (i_2, c_2)

Table 3: Post-Attribution Contribution Bounding. The input dataset is shown in Figure 2. The LTA rule was applied.

We now explain how the attributed datasets were generated in Table 3. First, note that in the absence of any contribution bounding, each conversion should simply be attributed to the last impression occurring prior to it. So in the dataset given in Figure 2, conversions c_1, c_2 and c_3 should be attributed to impression i_2, conversion c_4 should be attributed to impression i_4, and conversion c_5 should be attributed to impression i_5. This explains the first row of Table 3. We note that post-attribution contribution bounding with a per conversion contribution bounding scope is a no-op, so it would result in the same attributed dataset as the first (i.e., "None") row in Table 3. In second row of Table 3, a post-attribution contribution bound of 2 is applied for each impression. This leads to dropping the pair (i_2, c_3) from the attributed dataset because conversions c_1 and c_2 are already attributed to impression i_2. In the third row of Table 3, a contribution bound of 2 is applied post-attribution to each (user, advertiser) pair. Since Figure 2 specifies a dataset for a single user and for two advertisers (namely, advertiser1.com and advertiser2.com), this entails capping the number of attributed conversions for of each of the two advertisers to 2. Compared to the second row, this has the additional effect of dropping the pair (i_4, c_4) from the attributed dataset, since conversions c_1 and c_2 occur on the same advertiser and already appear in the attributed dataset. Finally, in the last row of Table 3, a contribution bound of 2 is applied post-attribution for each user. For the user whose dataset is shown in Figure 2, this implies that the total number of conversions (across all advertisers) appearing in the attributed dataset should be bounded to at most 2. Compared to the third row

[6] In the case of multi-touch attribution, each (impression, conversion) in the attributed dataset is weighted, and when enforced post-attribution, the contribution bound in fact limits the *total weight* of the pairs that belong to the contribution bounding scope.

Adjacency Relation	Difference between Adjacent Datasets
Impression	A single impression
Conversion	A single conversion
User × Publisher	All impressions shown to a user on a given publisher
User × Advertiser	All impressions shown to a user and for a given advertiser, and all conversions by the same user on the same advertiser
User × Publisher × Advertiser	All impressions shown to a user on a given publisher and corresponding to a given advertiser
User	All impressions shown on all publishers, and all conversions occurring on all advertisers, for a given user

Table 2: DP Adjacency Relations.

of Table 3, this has the further effect of dropping the pair (i_5, c_5) from the attributed dataset, because the same user already has two conversions, c_1 and c_2, appearing in the attributed dataset (despite the fact that these conversions occur on a different advertiser).

For the results for pre-attribution contribution bounding, see Table 4.

Contribution Bounding	Attributed Dataset
User × Advertiser (Contribution Bound = 2)	(i_5, c_5)
User (Contribution Bound = 2)	Empty

Table 4: Pre-Attribution Contribution Bounding. The input dataset is shown in Figure 2. The LTA rule was applied.

We next explain how the attributed datasets were generated in Table 4. In the first row, a contribution bound of 2 is applied *pre-attribution* for each (user, advertiser) pair. For the single user corresponding to Figure 2, the first advertiser will have its contribution bound exhausted after impressions i_1 and i_2 are processed; hence, no conversion occurring on the first advertiser will appear in the attributed dataset. By contrast, for the second advertiser, impression i_5 and and conversion c_5 will be processed, and the resulting pair (i_5, c_5) will be added to the attributed dataset (and at that point the contribution bound for the second advertiser would be exhausted). For the last row of Table 4, the contribution bound of 2 is enforced pre-attribution at the user level. For the user corresponding to Figure 2, the contribution bound would be exhausted after impressions i_1 and i_2 are processed, and thus the attributed dataset would be empty. We point out that pre-attribution contribution bounding with an per impression contribution bounding scope is a no-op; hence it results in the same attributed dataset as the first (i.e., "None") row in Table 3.

As is the case in Tables 3 and 4, pre-attribution contribution bounding in general results in a larger signal loss in the attributed dataset compared to post-attribution contribution bounding. As we will discuss shortly, the design choice of whether to enforce the contribution bound pre- or post-attribution significantly affects the end-to-end privacy of the ad conversion measurement system.

2.3 Valid Configurations

The DP aspects of a private conversion measurement system are mostly captured by the choice of (i) the attribution rule, (ii) the adjacency relation, (iii) the contribution bounding scope, and (iv) the contribution bound enforcement point. We refer to a setting of each of these choices as a *configuration.* It is natural to consider a configuration to be operationally *valid* if for every positive integer r, enforcing a contribution bound of r at the required point ensures that any two adjacent datasets always result in two post-attribution post-enforcement datasets that differ on at most $C_0 \cdot r$ many (impression, conversion) pairs, where C_0 is an absolute constant independent of the numbers of publishers and advertisers.[7] If this property does not hold, then a change of, e.g., a single impression's contributions, within the the contribution bound of r, could result in a change in the attributed dataset of magnitude growing with the (unbounded and potentially very large) number of publishers and advertisers showing ads to a given user. It turns out that this condition is not only sufficient to ensure the DP of the ad measurement system, but also in a sense necessary, unless the noise is increased with the total number of publishers or advertisers—which is practically unwieldy as this number is not fixed and can vary from user to user (note that the subset of publishers and advertisers showing ads to a given user as they browse the Web cannot be fixed ahead of time). In other words, a configuration is deemed *invalid* if the sensitivity increases as the number of advertisers or publishers increases. For more details, we refer the reader to Lemma 1 and the paragraph following it.

2.4 Our Contributions

In addition to formally defining the framework for ad conversion measurement and defining the notion of an operationally valid configuration, we provide a complete characterization of the validity of the configurations that most commonly arise in practice.

Classification. We provide a complete classification of all the configurations of attribution rule, adjacency relation and contribution bound enforcement point, that are operationally valid; see Table 5. We discuss the obtained classification next.

We first note that pre-attribution contribution bound enforcement results in valid configurations for all considered attribution

[7] To cover multi-touch attribution, we in fact bound the ℓ_1-distance between the two post-attribution post-enforcement datasets of weighted (impression, conversion) pairs. See Definitions 2 and 4 for more details.

rules and adjacency relations. A possible challenge to such an enforcement point is that an impression or publisher can incur a deduction from their contribution bound whenever they are part of the *input* to the attribution rule, and even if they are not selected for attribution. This can result in situations where a publisher can see their contribution bound totally exhausted due to conversions that got attributed to other publishers. This is in fact the main motivation for considering post-attribution contribution bound enforcement, which we discuss next.

It turns out that the adjacency relations that are valid for all attribution rules in the case of post-attribution enforcement are the *conversion*, *user* × *advertiser*, and *user* options. A limitation of the conversion adjacency relation is that the privacy would degrade as a user converts more than once. Given that conversion events are in practice not restricted to purchases (e.g., page views, email signups, and add-to-carts can also qualify as conversions), the privacy leakage could increase noticeably. On the other hand, the *user* adjacency relation could be operationally challenging to enforce as all the publishers and advertisers would have to share the same contribution bound, which could end up being dominated by certain publishers and/or advertisers. For the other relations of *impression*, *user* × *publisher*, and *user* × *publisher* × *advertiser*, we demonstrate in Table 5 that the situation is much more delicate, as the validity of the different attribution rules turns out to depend on the adjacency relation. For instance, we show that, surprisingly, while the *impression*, and *user* × *publisher* × *advertiser* adjacency relations and contribution bounding scopes admit valid configurations for post-attribution contribution bound enforcement, the *user* × *publisher* adjacency relation does not.

Our results suggest that if all the considered attribution rules are to be supported, then either pre-attribution enforcement, or a *user*, *user* × *advertiser*, or *conversion* adjacency relation should be used. If, however, post-attribution enforcement is desired and a middle ground is sought between the *conversion* and *user* contribution bounding scopes, then only a subset of the attribution rules can be supported (as in Table 5).

We emphasize that the dimensions considered in our classification are fundamental to *any* conversion measurement system. Specifically, any such system has to select an attribution rule. Moreover, any DP implementation has to choose a privacy unit. It also has to bound contributions, and keep track of a remaining contribution bound.

For a high-level overview of the proofs, we refer the reader to Section 6.4.

We next give an example illustrating the idea captured by the notion of *invalid* configurations.

Example of an Invalid Configuration. Consider the configuration where LTA is selected as the attribution rule, the adjacency relation (and the contribution bounding scope) is set to *user* × *publisher*, and contribution bounding is performed post-attribution. Moreover, consider the typical setting where an advertiser runs a campaign displaying ads on multiple publishers. The advertiser's goal is to estimate the number of conversions attributed to ad impressions shown on each publisher. Since the adjacency relation is *user* × *publisher*, and since summation has sensitivity $O(1)$, one would hope for an ϵ-DP algorithm with error $O(1/\epsilon)$. On a high level, our results suggest that surprisingly, this is not possible to achieve since adding last-touch interactions on a publisher can remove attributed conversions for *all* other publishers. Hence, the sensitivity in fact grows with the (practically unbounded) number of publishers, which would result in poor measurements even if a publisher has thousands of attributed conversions.

2.5 Additional Related Work

There have been several previous works on (non-private) conversion measurement, e.g., [5, 27, 28, 33]. We point out that it is common in the literature on DP to consider notions between protecting a single contribution and protecting all the user's contributions; see, e.g., [31] and the references therein. We also note that some recent ad conversion measurement systems rely on ad-hoc privacy notions; see, e.g., [9]. To the best of our knowledge, our work is the first study of the end-to-end (differential) privacy of ad conversion measurement systems.

The very recent works of [13] and [6] provide an empirical evaluation of a differentially private ad conversion measurement system similar to the one studied in this work. Their focus is the Privacy Sandbox Attribution Reporting API (ARA) on Chrome and Android. They consider last touch attribution and an *impression* privacy unit. Both of these work consider the linear queries functionality (e.g., conversion counts and values). The former focuses on hierarchical queries whereas the latter studies the non-hierarchical setting. These works provide a concrete instantiation of a DP ad conversion measurement system similar to the one studied in this work, and they empirically evaluate the error, on real ad conversion datasets, for different values of the differential privacy parameter ϵ. We refer the reader to these two papers for more details. Note that, in our terminology, ARA assumes that the attributed dataset is the input, and that post-attribution capping and noising is performed. Thus, our work complements these previous works: the valid configurations (for the *impression* adjacency relation) in our work imply that ARA satisfies an end-to-end DP guarantee for the corresponding attribution rules. Meanwhile, the invalid configurations imply that the end-to-end DP guarantee may not hold for those attribution rules.

Organization of Rest of the Paper. We start Section 3 with some notation that will be used in the rest of the paper. We recall the formal definition of DP in Section 3.1. In Section 3.2, we formally define the various attribution rules that will be studied in this paper. In Section 4, we present the notion of an operationally valid configuration, along with its connection to the design of a DP ad conversion measurement system. The pre- and post-attribution contribution bound enforcement algorithms are described in Section 5. We present our main validity and invalidity results in Section 6. Some of the proofs are given in Section 6.5 (with the rest deferred to the Appendix). Our work opens up several interesting areas of exploration; we describe some of these in Section 7 where we also discuss our results in the context of related practical applications.

	LTA	FTA	UNI	EXP	U-S	POS	IPA
	Post-Attribution						
Impression	Thm. 8(+)	Thm. 2(+)	Thm. 9(−)	Cor. 1(−)	Thm. 10(−)	(±)	(±)
User	Thm. 6(+)						
User × Publisher	Thm. 1(−)						
User × Advertiser	Thm. 7(+)						
User × Pub × Adv	Thm. 3(−)	Thm. 11(+)	Cor. 2(−)			(±)	(±)
	Pre-Attribution						
Impression User User × Publisher User × Advertiser User × Pub × Adv	Thm. 5(+)						
Conversion*	Thm. 4(+)						

Table 5: Validity of post- and pre-attribution enforcement configurations. A 🟩(+) cell indicates a valid configuration, a 🟥(−) cell indicates an invalid configuration, and a 🟧(±) cell means means that there are attribution rules in that family that result in a valid configuration and an invalid configuration. Specifically, both the IPA class and the POS class contain FTA and UNI; in the *impression* and *user* × *publisher* × *advertiser* adjacency relations, the former is valid but the latter is invalid. (*) For the *conversion* adjacency relation, no contribution bound enforcement is applied as the conversion is already only used once in the attribution rule.

3 PRELIMINARIES

Notation. For any positive integer n, we denote by $[n]$ the set $\{1, \ldots, n\}$. For any finite set S, we denote by S^* the set of all finite-length non-empty sequences of elements of S. For any positive integer m, the m-dimensional probability simplex, denoted by Δ_m, is defined as the set of all vectors in $[0,1]^{m+1}$ whose coordinates add up to 1. The ℓ_1-norm of a vector $v \in \mathbb{R}^d$ is defined as $\|v\|_1 = \sum_{i=1}^{d} |v_i|$. The Laplace distribution with zero mean and scale parameter $b > 0$ is the continuous probability distribution whose probability density function is given by $f(x; b) = \frac{1}{2b} \cdot e^{-\frac{|x|}{b}}$, for any real number x.

3.1 Differential Privacy

We denote two adjacent datasets $\mathbf{D}$ and $\mathbf{D}'$ by $\mathbf{D} \sim \mathbf{D}'$. The adjacency notions considered in this work are listed in Table 2, and will be further discussed in Section 4.1, but DP can be defined generally for any such relation.

Definition 1 (Differential Privacy [18]). *Let $\epsilon \geq 0$. A randomized mechanism $\mathcal{M}$ is ϵ-differentially private (denoted by ϵ-DP) if for each pair $\mathbf{D} \sim \mathbf{D}'$ of adjacent datasets and each subset $\mathcal{S}$ of outputs of $\mathcal{M}$, it holds that $\Pr[\mathcal{M}(\mathbf{D}) \in \mathcal{S}] \leq e^{\epsilon} \cdot \Pr[\mathcal{M}(\mathbf{D}') \in \mathcal{S}]$, where the probabilities are over the randomness in $\mathcal{M}$.*

Intuitively, the DP definition guarantees that the outputs of two adjacent datasets are approximately statistically indistinguishable. The degree of indistinguishibility is dictated by the ϵ parameter. The smaller ϵ is, the more private the algorithm would be. In our setting, the dataset $\mathbf{D}$ consists of the impressions and conversions (pre-attribution), across all users, advertisers and publishers. The output $\mathcal{M}(\mathbf{D})$ is the output of the privacy-preserving ad conversion measurement system.

DP satisfies several useful mathematical properties that have made it an appealing measure of privacy. These include robustness to post-processing, composition, and group privacy. For a comprehensive overview of the area, we refer the reader to the monographs [19, 39].

3.2 Attribution Rule

The *attribution rule* function (see, e.g., [15] for background) takes as input a sequence of m impressions and a conversion, all corresponding to the same user and advertiser, and returns a fraction in $[0,1]$ for each of the m impressions. We denote this function by $\mathsf{a} : \mathcal{I}^* \times C \longrightarrow [0,1]^*$, where $\mathcal{I}$ is the set of all possible impressions, and C is the set of all possible conversions. We assume that for any input $((i_1, \ldots, i_m), c)$ to the attribution function a, it is the case that the impressions $i_1, \ldots, i_m$ have been sorted from least to most recent according to their timestamps and c occurs later than i_m. Moreover, it is assumed that $\mathsf{a}((i_1, \ldots, i_m), c) \in \Delta_{m-1}$.

3.2.1 Single-Touch. In *single-touch* attribution, only a single coordinate in the output $\mathsf{a}((i_1, \ldots, i_m), c)$ is equal to 1 and all the other $m-1$ coordinates are equal to 0. We next describe some notable special cases of single-touch attribution.

Last-Touch Attribution (LTA). $\mathsf{a}((i_1, \ldots, i_m), c) = (0, \ldots, 0, 1)$, i.e., the last impression in the sequence is selected.

First-Touch Attribution (FTA). $\mathsf{a}((i_1, \ldots, i_m), c) = (1, 0, \ldots, 0)$, i.e., the first impression in the sequence is selected.

3.2.2 Multi-Touch. While the single touch attribution rules assign all the credit to a single impression, *multi-touch* attribution allows spreading the credit over more than one impression. The simplest multi-touch attribution rule is uniform (aka linear).

Uniform (UNI). $\mathsf{a}((i_1, \ldots, i_m), c) = (1/m, \ldots, 1/m)$, i.e., the credit is split equally among all the impressions.

Exponential Time Decay (EXP). In the EXP rule with half-life parameter $t_{1/2}$, the credit assigned to a given impression is proportional to $(0.5)^{\frac{t}{t_{1/2}}}$, where t is the difference between the timestamp of the impression and that of the conversion.

U-Shaped (U-S). If there are at least three impressions, 40% of the credit goes to the first touch, 40% of the credit goes to the last touch, and the remaining credit is divided uniformly over all the intermediate impressions (i.e., those that are neither first nor last). If there are two impressions, we assume that the credit is split equally between them.

Positional (POS). In positional (aka position based) attribution, the credit assigned to each impression is based on the total number of impressions and the order in which this impression occurs, i.e., the credit does not depend on the user, publisher, or advertiser IDs, or on the metadata. More precisely, a positional attribution is parameterized by a class $\mathcal{F} = \{v_m\}_{m\in\mathbb{N}}$ of vectors, where $v_m \in \Delta_{m-1}$. The attribution function is defined as $\mathsf{a}((i_1, \ldots, i_m), c) = v_m$.

It is worth noting that POS is a class of attribution rules, one for each choice of $\mathcal{F}$. The POS class contains FTA ($v_m = (1, 0, \ldots, 0)$), LTA ($v_m = (0, \ldots, 0, 1)$), UNI ($v_m = (1/m, \ldots, 1/m)$) and U-S ($v_m = (0.4, 0.2/(m-2), \ldots, 0.2/(m-2), 0.4)$), but it does not contain EXP.

Impression-Priority Attribution (IPA). One of the impressions is selected by applying a prioritization function $\mathsf{pr} : \mathcal{I}^* \to [0, 1]^*$ that depends only on the impressions and not on the conversion, i.e., $\mathsf{a}((i_1, \ldots, i_m), c) = \mathsf{pr}(i_1, \ldots, i_m) \in \Delta_{m-1}$.

Similar to POS, IPA is a class of attribution rules; it contains FTA, LTA, UNI, EXP, and U-S.

4 DIFFERENTIALLY PRIVATE CONVERSION MEASUREMENT SYSTEMS

To describe our framework for DP conversion measurement, we first discuss in Section 4.1 the adjacency relations and contribution bounding scopes that we consider, and then describe attribution systems and how to privatize their outputs in Section 4.2

4.1 Adjacency Relations and Contribution Bounding Scopes

As we saw in Definition 1, any application of DP should specify a notion of when two datasets are considered adjacent. In the conversion measurement setting, there are several options; the most natural of them are summarized in Table 2.

For any relation in the first column of Table 2, we can then define two datasets to be *adjacent* if one can be obtained from the other by adding or removing impressions and/or conversions as listed in the second column of the table.

Any application of DP should, at some level, limit the individual contributions; otherwise, the finite amount of noise that is injected would not be sufficient to ensure DP when the individual contributions become too large. In the conversion measurement use case, there are multiple contribution bounding scopes in which the contributions could be limited. These include the same choices listed in Table 2 for the adjacency relation. We consider henceforth the most natural setting where the contribution bounding scope matches the adjacency relation. E.g., for the *user* × *publisher* adjacency relation, all the contributions of a given user on a given publisher share the same contribution bound.

4.2 Attribution Systems

An *attribution system* is an algorithm that takes as input impressions and conversions sequences and outputs the attributions, represented by a set of weighted pairs of impressions and conversions (defined formally as an *attributed dataset* below).

Definition 2 (Attributed Dataset)**.** *An* attributed dataset $\mathcal{D}_{attr}$ *is a set of triplets* $(i, c, w) \in \mathcal{I} \times C \times \mathbb{R}_{\geq 0}$*, where for each* (i, c) *there is a unique* w*. We may represent an attributed dataset as a function* $w_{\mathcal{D}_{attr}} : \mathcal{I} \times C \to \mathbb{R}_{\geq 0}$*, where* $w_{\mathcal{D}_{attr}}(i, c)$ *represents*[8] *the total weight attributed to impression* i *by conversion* c*. The* ℓ_1-distance *between two attributed datasets* $\mathcal{D}_{attr}, \mathcal{D}'_{attr}$ *is given by*

$$\|\mathcal{D}_{attr} - \mathcal{D}'_{attr}\|_1 := \sum_{i\in\mathcal{I}, c\in C} |w_{\mathcal{D}_{attr}}(i, c) - w_{\mathcal{D}'_{attr}}(i, c)|.$$

Given an attribution system, we can build a *conversion measurement system* by applying a function f that maps the attributed dataset to a vector in $\mathbb{R}^d$; the vector measures the statistics that an ad tech would like to estimate. For example, if the ad tech wants to know the total attributions for each slice of (campaign × time-of-day), then each of the d dimensions can represent a valid (campaign ID, time-of-day) pair, and the value that f assigns to that dimension would be the total attribution for that campaign ID and time-of-day.

Of course, as described above, the system is not (differentially) private: the ad tech can allocate a dimension for a particular user and then count exactly, e.g., the number of impressions that user sees. Since this is not desirable, we employ two methods to ensure privacy. First, we apply contribution bound enforcement *within the attribution system*, which will be discussed below. Second, we add (appropriately scaled) Laplace noise to each of the d coordinates of the values of f; these noisy estimates are then sent to the ad tech. See Figure 3 for an illustration of such a conversion measurement system. Note that we consider the *central* DP setting, where a (trusted) curator runs the attribution rule, computes the function f, and adds the noise; the output of this curator is required to be DP.

It turns out that there are two important properties needed to ensure DP of the output. The first is that the sensitivity of f is small, i.e., that a small change (in the ℓ_1-distance) to the attributed dataset does not change the value of f (again, in the ℓ_1-distance) by much. This is formalized below.

Definition 3 (Sensitivity of f)**.** *For a function* f *that maps an attributed dataset to a vector of real numbers in* $\mathbb{R}^d$*, we define its* (ℓ_1-)sensitivity *to be*

$$\Delta(f) := \max_{\substack{\mathcal{D}_{attr}, \mathcal{D}'_{attr} \\ \|\mathcal{D}_{attr} - \mathcal{D}'_{attr}\|_1 \leq 1}} \|f(\mathcal{D}_{attr}) - f(\mathcal{D}'_{attr})\|_1.$$

For many natural functions, such as the "sum by slices" example above, the sensitivity is bounded (e.g., by 1 in the example).

[8] In this notation, $w_{\mathcal{D}_{\mathrm{attr}}}(i, c) = 0$ could correspond to different situations: one is that impression i was considered when attributing conversion c but was not selected by the attribution rule; the other is that impression i and conversion c are incompatible, e.g., impression i takes place after conversion c, or they correspond to different users or advertisers.

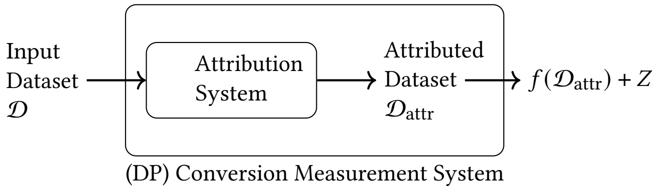
Input
Dataset
$\mathcal{D}$
Attribution System
Attributed
Dataset
$\mathcal{D}_{\text{attr}}$
$f(\mathcal{D}_{\text{attr}}) + Z$
(DP) Conversion Measurement System

Figure 3: Illustration of a (DP) Conversion Measurement System. Each coordinate of the noise Z is drawn from the Laplace distribution with an appropriate scale (see Lemma 1). We note that the attribution system can include a contribution bound enforcement component (this is the case in Algorithms 1 and 2).

The second property we need is with regards to the attribution system itself. Although we have not defined the contribution bound enforcement yet, it takes in a positive integer parameter r, considered as the "contribution bound".[9] To ensure DP, we need this parameter r to be an upper bound on the possible change (in the ℓ_1 sense) in the resulting attributed dataset.

More specifically, an attribution system—which can be specified by a "configuration" of adjacency relation, contribution bound enforcement point, and attribution rule—is "valid" if two adjacent datasets get mapped to attributed datasets that are at most $O(r)$ apart, as stated more precisely below.

Definition 4 (Valid Configurations)**.** *An adjacency relation along with a contribution bound enforcement point is said to be* C_0-valid *for a given attribution rule if, for every positive integer r, applying a contribution bound of r at the required enforcement point ensures that any two adjacent datasets always result in two attributed datasets that are at an ℓ_1-distance of at most $C_0 \cdot r$, where C_0 is an absolute constant independent of the numbers of publishers and advertisers. We call a configuration* valid *if it is C_0-valid for some absolute constant $C_0 > 0$.*

Assuming the above two properties, if we appropriately scale the Laplace-distributed noise injected in Figure 3, then we can guarantee that the system is DP:

Lemma 1. *If the attribution system is instantiated with a C_0-valid configuration and each coordinate of the noise Z is sampled according to the Laplace distribution with scale parameter $C_0 \cdot r \cdot \Delta(f)/\varepsilon$, then the conversion measurement system is ε-DP.*

We remark that the "converse" of Lemma 1 is also true in the following sense: if we let f be the identity function, i.e., the range of f is associated with $\mathbb{R}^{\mathcal{I} \times \mathcal{C}}$ and let $f(\mathcal{D}_{\text{attr}})_{(i,c)} = w_{\mathcal{D}_{\text{attr}}}(i, c)$, then the conversion measurement system with Z sampled according to the Laplace distribution with scale $C_0 \cdot r/\varepsilon$ is ε-DP iff the attribution system is instantiated with a C_0-valid configuration. In other words, the validity of the configuration characterizes the DP property of the conversion measurement system in this sense.

Proof of Lemma 1. Consider any two adjacent datasets $\mathcal{D}, \mathcal{D}'$; let $\mathcal{D}_{\text{attr}}$ and $\mathcal{D}'_{\text{attr}}$ be the results of the attribution system on these two datasets, respectively. Then, since the configuration is C_0-valid, we have $\|\mathcal{D}_{\text{attr}} - \mathcal{D}'_{\text{attr}}\|_1 \leq C_0 \cdot r$. Therefore, from the definition of $\Delta(f)$, we can conclude that $\|f(\mathcal{D}_{\text{attr}}) - f(\mathcal{D}'_{\text{attr}})\|_1 \leq C_0 \cdot r \cdot \Delta(f)$. In other words, the entire measurement system before adding noise is simply a function with ℓ_1-sensitivity at most $C_0 \cdot r \cdot \Delta(f)$. Thus, the standard DP guarantee of the Laplace mechanism [18] implies that the system is ε-DP as desired. □

Remark 4. *While the treatment above considers the evaluation of a* single *function on the input dataset, it can be readily extended to the setting where one would like to compute* multiple *(possibly adaptively chosen) functions on the same dataset; this can be done using the standard* composition *properties of DP [19].*

Remark 5. *For typical functions f that are of interest in ad conversion measurement, their sensitivity $\Delta(f)$ can be computed explicitly. For example, if f computes the total (attributed) conversion count, then* $\Delta(f) = 1$. *Similarly, if f computes the number of distinct users with attributed conversions, then* $\Delta(f) = 1$ *too. On the other hand, if f computes the sum of capped conversion values, where each value is capped to a positive real number V, then the sensitivity of f is given by $\Delta(f) = V$.*

Queries of interests are also often "sliced" by certain attributes. For example, one might be interested in histogram of the total (attributed) conversion count for each publisher or each geographic location (which is e.g. determined by the impression's metadata). In this case, the sensitivity $\Delta(f)$ remains 1.

5 CONTRIBUTION BOUND ENFORCEMENT

As stated earlier, a major part of the attribution system is contribution bound enforcement algorithm. This algorithm takes a positive integer r, the "contribution bound", and tries to "enforce" this contribution bound. We consider two types of enforcement in this paper, *pre-attribution* and *post-attribution*, which we will explain next. (It is an interesting future direction to understand if there are other contribution bound enforcement strategies that may be more privacy safe and/or practical than the ones considered here.)

First, let us note that for the *conversion* adjacency relation (Table 2), no contribution bound enforcement is applied. The reason is that each conversion is used only once in the attribution rule. Therefore, the enforcement strategies described below will anyway not affect it.

To describe the enforcement strategies for the other adjacency relations, it is key to define the scope of a contribution bound.

Definition 5 (Contribution Bounding Scope)**.** *The* contribution bounding scope *for an adjacency relation is the unit of that relation.*

For example, for the user adjacency relation, a contribution bounding scope would be each user.

5.1 Post-Attribution Contribution Bound Enforcement

For post-attribution contribution bound enforcement, we simply run the attribution algorithm as usual. However, we only add each weighted (impression, conversion) pair to the attributed dataset if it does not exceed the (remaining) contribution bound of that scope. The pseudo-code is given in Algorithm 1.

[9] We note that while in post-attribution enforcement, the contribution bound r could be set to any positive real number without any modification to our treatment, we choose to keep it integral for consistency with the case of pre-attribution enforcement.

ALGORITHM 1: Attribution with Post-Attribution Contribution Bound Enforcement.

```
Parameters: Attribution rule $\mathsf{a}$, contribution bound $r$.
Input: Dataset $\mathcal{D} = (i_1, \ldots, i_n), (c_1, \ldots, c_m)$ ordered from oldest to newest.
for each contribution bounding scope $s$ do
    $b_s \leftarrow r$.        {Set remaining contribution bound of this scope to $r$.}
$\mathcal{D}_{\mathrm{attr}} \leftarrow \emptyset$.
for $j = 1, \ldots, m$ do
    $i'_1, \ldots, i'_\ell \leftarrow$ impressions that come before $c_j$ in time and are associated with the same advertiser and the same user as $c_j$.
    $(w_1, \ldots, w_\ell) \leftarrow \mathsf{a}((i'_1, \ldots, i'_\ell), c_j)$.        {Standard attribution alg.}
    for $k = 1, \ldots, \ell$ do
        $s \leftarrow$ contribution bounding scope corresponding to $(i'_k, c_j, w_k)$.
        if $b_s \geq w_k$ then
            $b_s \leftarrow b_s - w_k$.        {Subtract the spent contribution bound.}
            Add $(i'_k, c_j, w_k)$ to $\mathcal{D}_{\mathrm{attr}}$.
return $\mathcal{D}_{\mathrm{attr}}$
```

5.2 Pre-Attribution Contribution Bound Enforcement

Pre-attribution enforcement is much more pessimistic than the post-attribution approach. Specifically, we charge one unit from the contribution bound of *every scope involved with the input impressions* to the attribution rule. If any scope does not have enough contribution bound left, we remove all impressions associated with that scope. As we will show below, such a pessimistic approach is—perhaps not too surprisingly—more privacy-safe than the post-attribution approach in certain settings. The full pseudo-code of the pre-attribution contribution bound enforcement is given in Algorithm 2.

ALGORITHM 2: Attribution with Pre-Attribution Contribution Bound Enforcement.

```
Parameters: Attribution rule $\mathsf{a}$, contribution bound $r$.
Input: Dataset $\mathcal{D} = (i_1, \ldots, i_n), (c_1, \ldots, c_m)$ ordered from oldest to newest.
for each contribution bounding scope $s$ do
    $b_s \leftarrow r$.        {Set remaining contribution bound of this scope to $s$.}
$\mathcal{D}_{\mathrm{attr}} \leftarrow \emptyset$.
for $j = 1, \ldots, m$ do
    $I \leftarrow$ set of impressions that come before $c_j$ in time and are associated with the same advertiser and the same user as $c_j$
    $S \leftarrow$ set of contribution bounding scopes corresponding to at least one impression in $I$.
    for $s \in S$ do
        if $b_s \geq 1$ then
            $b_s \leftarrow b_s - 1$.        {Subtract the spent contribution bound.}
        else
            Remove $s$ from $S$. {Not enough contribution bound remaining; discard from the final output.}
    $i'_1, \ldots, i'_\ell \leftarrow$ impressions in $I$ corresponding to scopes in $S$.
    $(w_1, \ldots, w_\ell) \leftarrow \mathsf{a}((i'_1, \ldots, i'_\ell), c_j)$.
    Add $(i'_1, c_j, w_1), \ldots, (i'_\ell, c_j, w_\ell)$ to $\mathcal{D}_{\mathrm{attr}}$.
return $\mathcal{D}_{\mathrm{attr}}$
```

6 CLASSIFICATION RESULTS

In this section we present our results on the validity of different adjacency relations; they are summarized in Table 5. For ease of presentation, we organize our results into two categories: when the validity holds independent on the attribution rule and otherwise.

We only state in this section the theorems that are proved in Section 6.5. For the other theorems, we provide forward pointers to their formal statements (along with their proofs) in the Appendix.

6.1 Attribution Rule-Independent Validity

We start by discussing the validity of different adjacency relations, which turn out to be independent of the attribution rule. The *conversion* adjacency relation (which, as stated earlier, involves no contribution bound enforcement) turns out to be valid for every adjacency relation and every attribution rule (Theorem 4).

For pre-attribution enforcement, it also turns out that all adjacency relations and all attribution rules result in valid configurations (Theorem 5).

We next turn our attention to post-attribution enforcement. In this case, we show that the validity of the *user* × *advertiser* and *user* adjacency relations are independent of the attribution rule (Theorems 6 and 7). By contrast, the following theorem (which we will prove in Section 6.5) shows the invalidity of the *user* × *publisher* adjacency relation for any attribution rule.

Theorem 1. *For any attribution rule, the user × publisher adjacency relation with contribution bound enforced post-attribution constitutes an invalid configuration.*

6.2 Post-Attribution Enforcement and *Impression* Adjacency

In this section we consider the *impression* adjacency under post-attribution enforcement of the contribution bound. Our first result proves the validity of FTA.[10] Following Lemma 1, this means that these configurations require adding twice as much noise compared to those in previous validity results (under the same contribution bound).

Theorem 2. *For the FTA rule, the impression adjacency relation with contribution bound enforced post-attribution constitutes a* 2*-valid configuration.*

We also prove a similar result for LTA (Theorem 8). By contrast, we prove that the UNI, EXP, and U-S attribution rules are all valid in this case (Theorem 9, Corollary 1, and Theorem 10 respectively).

6.3 Post-Attribution Enforcement and *User* × *Publisher* × *Advertiser* Adjacency

We next consider the *user* × *publisher* × *advertiser* adjacency relation. In this case, and under post-attribution enforcement, it turns out that only FTA results in a valid configuration (Theorem 11) whereas the LTA, UNI, EXP, and U-S attribution rules result in invalid configurations (Theorem 3 and Corollary 2).

[10] Note that it has a slightly worse absolute constant of $C_0 = 2$ compared to $C_0 = 1$ as in Theorem 4.

Theorem 3. *For the LTA attribution rule, the user × publisher × advertiser adjacency relation with contribution bound enforced post-attribution constitutes an invalid configuration.*

6.4 Intuition and Proof Overview

Before we proceed to the formal proofs of these classification results, let us briefly (and informally) discuss the high-level ideas behind them. For the invalidity results in the post-attribution case, there are (roughly speaking) two root causes behind them:

- **Cascading Effect within Attribution Rule.** A single impression can be an input of multiple executions of the attribution algorithm (Line 8 of Algorithm 1). When removing such an impression from the dataset, the attribution rule also changes the weights assigned to other input impressions. Below we construct datasets which make sure that these changes affect different privacy units, implying that it can exceed the contribution bound. This is the idea behind our constructions for the impression adjacency relation (Theorem 9, Corollary 1, and Theorem 10).
- **Multiple Impressions Affecting Multiple Privacy Units.** In some scenarios (see LTA, FTA discussion below), changing a single impression does not result in a cascading effect. In this case, the high-level idea is to construct multiple impressions—corresponding to the same privacy unit—in such a way that removing each one affects some other different privacy unit. When all of these impressions are removed simultaneously, the effect occurs across different privacy units and therefore bypasses the contribution bounding. This is the gist of our constructions for the user × publisher adjacency relation (Theorem 1) and the user × publisher × advertiser adjacency relation (Theorem 3).

We next discuss the validity results. We remark that the attribution rule-independent results (e.g., for the user-level adjacency relation) are relatively straightforward to prove, so we will focus our discussion here on the exceptions: LTA for the impression-level adjacency relation (Theorem 8), and FTA for the impression-level adjacency (Theorem 2) and the user × publisher × advertiser adjacency relations (Theorem 11).

- **LTA.** When we remove an impression, LTA essentially "routes" all the attributions of this impression to the previous one with the same advertiser. Thus, in the impression-level adjacency relation, contribution bounding will upper-bound the change. On the other hand, this reasoning fails for the user × publisher × advertiser adjacency relation because it is possible to remove multiple impressions that affect different privacy units (i.e., different publishers); this is indeed the second root cause described above for invalidity.
- **FTA.** When we remove an impression, if this impression is not the first one of this advertiser, then no change occurs in attribution. Otherwise, FTA "routes" all the attributions of this impression to the second impression of this advertiser. Similar to LTA, this implies validity for the impression-level adjacency relation. However, in contrast to LTA, this argument remains true in the user × publisher × advertiser adjacency relation; this is because, even after removing multiple impressions of the same advertiser, all attributions are routed to the same impression—the first one of this advertiser after the removal. Therefore, post-attribution enforcement successfully bounds the change.

This concludes our summary of the proof ideas. We will next formalize these by providing the proofs of Theorems 1, 2, and 3. (The remaining proofs are deferred to the Appendix.)

6.5 Selected Proofs

FTA, Impression Adjacency. We now prove the validity of the FTA rule for the *impression* adjacency relation. The proof follows the outline from the previous subsection.

Theorem 2. *For the FTA rule, the impression adjacency relation with contribution bound enforced post-attribution constitutes a 2-valid configuration.*

Proof of Theorem 2. Consider two adjacent datasets $\mathcal{D}, \mathcal{D}'$ such that $\mathcal{D}'$ results from removing an impression $\tilde{i}$. Let A be $\tilde{i}$'s advertiser, U be $\tilde{i}$'s user and C_A denote the set of conversions from the advertiser and the user. We may assume w.l.o.g. that all conversions in C_A occur *after* the first impression w.r.t. the advertiser A and the user U. (As other conversions remain unattributed in both $\mathcal{D}$ and $\mathcal{D}'$.) We consider two cases, based on whether $\tilde{i}$ is the first impression w.r.t. its advertiser A and its user U (in $\mathcal{D}$).

- Case I: $\tilde{i}$ is *not* the first impression w.r.t. the advertiser A and the user U. In this case, the two attributed datasets $\mathcal{D}_{\text{attr}}, \mathcal{D}'_{\text{attr}}$ are exactly the same, because all conversions in C_A are attributed to the first impression w.r.t. the advertiser A (which is not $\tilde{i}$).
- Case II: $\tilde{i}$ is the first impression w.r.t. the advertiser A and the user U. In this case, all conversions $c_j \in C_A$ are attributed to $\tilde{i}$. These are the only conversions whose attributions change between $\mathcal{D}$ and $\mathcal{D}'$.

  To analyze this change, consider two subcases, whether $\tilde{i}$ is the only impression in $\mathcal{D}$ from A.
  - Case IIa: $\tilde{i}$ is the only impression in $\mathcal{D}$ from A and the user U. In this case, all conversions in C_A become unattributed in $\mathcal{D}'$. Therefore, $\|\mathcal{D}_{\text{attr}} - \mathcal{D}'_{\text{attr}}\|_1 = \sum_{c_j} w_{\mathcal{D}_{\text{attr}}}(\tilde{i}, c_j) \leq r$, where the inequality follows from the post-attribution contribution bound enforcement for $\tilde{i}$.
  - Case IIb: $\tilde{i}$ is not the only impression in $\mathcal{D}$ from A and the user U. Let $\tilde{i}'$ be the second impression in $\mathcal{D}$ from A. In this case, every conversion in C_A is either attributed to $\tilde{i}'$ or unattributed in $\mathcal{D}'$. Furthermore, no conversions are attributed to $\tilde{i}'$ in $\mathcal{D}$ (because $\tilde{i}$ comes before $\tilde{i}'$ in the same advertiser A and the same user U). Therefore, we have
  $$\|\mathcal{D}_{\text{attr}} - \mathcal{D}'_{\text{attr}}\|_1 = \sum_{c_j} w_{\mathcal{D}_{\text{attr}}}(\tilde{i}, c_j) + \sum_{c_j} w_{\mathcal{D}'_{\text{attr}}}(\tilde{i}', c_j) \leq 2r,$$
  where the inequality follows from the post-attribution contribution bound enforcement for impressions $\tilde{i}$ and $\tilde{i}'$.

In all cases, we can conclude that $\|\mathcal{D}_{\text{attr}} - \mathcal{D}'_{\text{attr}}\|_1 \leq 2r$. □

LTA, User × Publisher × Advertiser Adjacency. Next, we prove the invalidity of the LTA rule under the *user × publisher × advertiser* adjacency relation. This is due to the fact that we may arrange the impressions/conversions in such a way that a single publisher gets a large amount of attribution weight (before contribution bound enforcement) and that, once this publisher is removed, this weight is re-attributed to multiple publishers. The latter ensures that the change grows with the number of publishers. (This is also the main difference between LTA and FTA, since we cannot ensure such a condition for FTA.) Such an example is given together with a formal argument below.

Theorem 3. *For the LTA attribution rule, the user × publisher × advertiser adjacency relation with contribution bound enforced post-attribution constitutes an invalid configuration.*

Proof of Theorem 3. Let $r = 1$ and let $p > 1$ be any integer. We construct the dataset $\mathcal{D}$ as follows:

- Let there be a single user, a single advertiser, and p publishers $P_1, \ldots, P_p$.
- Let $i_1, \ldots, i_{2p-2}$ be impressions such that impression i_{2k-1} is associated with publisher P_k and impression i_{2k} is associated with publisher P_p for all $k \in [p-1]$.
- Let $c_1, \ldots, c_{p-1}$ be conversions such that c_k appears after i_{2k} and before i_{2k+1}, for all $k \in [p-1]$.

Finally, let $\mathcal{D}'$ be the dataset resulting from removing publisher P_p's impressions (i.e., $i_2, \ldots, i_{2p-2}$) from $\mathcal{D}$.

In $\mathcal{D}$, publishers $P_1, \ldots, P_{p-1}$'s impressions get attributed with zero weight. On the other hand, in $\mathcal{D}'$, each of these publishers get attribution weight of exactly one. Therefore, $\|\mathcal{D}_{\text{attr}} - \mathcal{D}'_{\text{attr}}\|_1 \geq p - 1$, invalidating the attribution system for this configuration. □

Any Attribution Rule, User × Publisher Adjacency. Finally, we prove the invalidity of the *user × publisher* adjacency for any attribution rule. We remark that, if we were looking for an invalidity proof of a specific attribution rule, then the construction could have been simplified. For example, the construction in Theorem 3 above also shows the invalidity of LTA in this setting. However, we would like our proof to generalize to *all* attribution rules. Our construction below accomplishes this by first creating another "dummy" dataset $\tilde{\mathcal{D}}$ (with multiple publishers) to understand how the attribution weights are distributed across different publishers. We then create the datasets $\mathcal{D}, \mathcal{D}'$ that differ on the highest weighted publisher to ensure that there is a large–unbounded–change between the two attributed datasets.

Theorem 1. *For any attribution rule, the user × publisher adjacency relation with contribution bound enforced post-attribution constitutes an invalid configuration.*

Proof of Theorem 1. Let a be any attribution rule. To create our datasets $\mathcal{D}, \mathcal{D}'$, let us start by constructing another dataset $\tilde{\mathcal{D}}$ as follows.

- Let there be $\binom{p}{2}$ advertisers $A_{\{1,2\}}, A_{\{1,3\}}, \ldots, A_{\{p-1,p\}}$, and p publishers $P_1, \ldots, P_p$.
- For each advertiser $A_{\{j,k\}}$, let there be impressions $i_j^{\{j,k\}}, i_k^{\{j,k\}}$ and conversion $c^{\{j,k\}}$, coming after both impressions. Furthermore, let $i_j^{\{j,k\}}$ and $i_k^{\{j,k\}}$ be associated with publishers P_j and P_k, respectively.

Now, suppose we run the attribution system—without any contribution bound enforcement—on $\tilde{\mathcal{D}}$. Let P_ℓ denote the publisher that gets the largest total attribution weight (with ties broken arbitrarily). Note that the total weight it receives must be at least $\binom{p}{2}/p = 0.5(p-1)$.

We now construct $\mathcal{D}$ by keeping only advertisers $A_{\{\ell,j\}}$ for $j \in [p] \setminus \{\ell\}$ in $\tilde{\mathcal{D}}$ (and discard the rest of advertisers together with all impressions and conversions associated to them). Let the contribution bound r be 1. Furthermore, let $\mathcal{D}'$ denote the dataset resulting from removing all impressions corresponding to publisher P_ℓ from $\mathcal{D}$. We will now show that $\|\mathcal{D}_{\text{attr}} - \mathcal{D}'_{\text{attr}}\|_1 \geq 0.5(p-1)$, which implies that the attribution system is an invalid one.

To show this, first observe that, in $\mathcal{D}'$, each $i_j^{\{\ell,j\}}$ is the only impression that gets fed into the attribution rule for $c^{\{\ell,j\}}$; therefore, it gets attributed with weight one. Furthermore, since we use a contribution bound $r = 1$ for each (user, publisher) pair and each $i_j^{\{\ell,j\}}$ corresponds to a different publisher, the contribution bound enforcement leaves these weights unchanged. In summary, we have

$$w_{\mathcal{D}'_{\text{attr}}}(i_j^{\{\ell,j\}}, c^{\{\ell,j\}}) = 1.$$

On the other hand, by our choice of the publisher P_ℓ, we have

$$\sum_{j \in [p] \setminus \{\ell\}} w_{\mathcal{D}_{\text{attr}}}(i_j^{\{\ell,j\}}, c^{\{\ell,j\}}) \leq 0.5(p-1).$$

Combining the above two inequalities, we get that

$$\begin{aligned} &\|\mathcal{D}_{\text{attr}} - \mathcal{D}'_{\text{attr}}\|_1 \\ &\geq \sum_{j \in [p] \setminus \{\ell\}} |w_{\mathcal{D}'_{\text{attr}}}(i_j^{\{\ell,j\}}, c^{\{\ell,j\}}) - w_{\mathcal{D}_{\text{attr}}}(i_j^{\{\ell,j\}}, c^{\{\ell,j\}})| \\ &\geq 0.5(p-1). \end{aligned}$$

□

7 DISCUSSION AND FUTURE DIRECTIONS

In this paper, we presented a formal framework for DP ad conversion measurement setting. We also demonstrated a delicate interplay between attribution and privacy. We defined the notion of operationally valid configurations, and provided a complete classification of the validity of the configurations based on the most popular attribution rules, adjacency relations, contribution bounding scopes, and contribution bound enforcement points. We hope that our end-to-end differential privacy framework can lead to a solid foundation for practical privacy-preserving ad conversion measurement systems.

While we have focused for simplicity on pure-DP (Definition 1), ℓ_1-sensitivity (Definition 2, 3, and 4), and Laplace mechanism (Lemma 1), our formalism extends readily to the case of approximate-DP [17], other type of sensitivities, and other DP mechanisms. For example, if the set of measurements is large, we could replace the Laplace mechanism with the partition selection algorithm (see, e.g., [14] and the references therein) if we relax to approximate DP. Similarly, we can extend our framework to ℓ_2-sensitivity[11] and Gaussian mechanism, which could for instance be used to train DP predicted conversion rate models based on DP stochastic gradient descent [2]. (For prior work on non-private conversion models, see, e.g., [5, 27, 28, 33].)

We describe next some interesting future research directions.

[11] Note that changing the sensitivity notion may change the set of valid configurations.

Adjacency Relation ≠ Contribution Bounding Scope. We focused in this work on the most natural setting where the contribution bound scope is the same as the adjacency relation. In principle, this is not necessary: e.g., one might consider a *user* contribution bounding scope with a *user* × *advertiser* adjacency relation. It might be interesting to give a characterization in such cases, as it will lead to an even more fine-grained understanding of the privacy provided by the conversion measurement system.

Contribution Capping: Beyond Pre- and Post-Attribution? While we focus on pre-attribution and post-attribution contribution capping, it remains an interesting open question whether there are other (general) capping procedures that can further improve the utility-privacy trade-off.

To illustrate the challenge, note that an intuitive capping strategy is to do it "at the query evaluation time". Although such a strategy makes sense for certain query functions f and adjacency relations, it is not completely well-defined for all functions f. For example, let f be the number of distinct users with attributed conversions from Remark 5 and suppose we are interested in the *impression* adjacency relation. If two impressions share the same user ID, then it is not clear what their contributions are; on one hand, removing each of them alone does not cause any change to the value of f. Meanwhile, removing them both may decrease the value. Such a situation is only exacerbated when we have a more complicated function f. We also remark that it is preferable if a single capping procedure is used for all functions f since it allows more flexibility for the measurements that can be made on the platform.

Privacy of the Computation. Our work has focused on guaranteeing privacy against an adversary that has access to "what" is being computed, but not to "how" it is being computed. Concretely, our results capture the setting where a single entity has access to the all the raw impression and conversion data, and seeks to release DP estimates to some requested conversion measurement queries. Studying the privacy of "how" this is computed is an important direction for future work. For instance, one could naturally extend this formalism to distributed settings where the trust in a single entity is relaxed by relying on methods such as secure multiparty computing [21], and on-device noise addition as in local DP [18, 22, 26] or shuffle DP [10, 11, 20]. Moreover, we studied the case of a static dataset of impressions and conversions; it would be of interest to study the online variant of the problem where privacy needs to be ensured at any time as the impressions and conversions take place.

Enhanced Attribution. Some attribution systems offer a conversion lookback window option (which limits how far back in time from a conversion are impressions eligible for attribution), and an impression expiry option (which limits how far in the future would the impression be eligible for attribution). It would be interesting to investigate the interplay of these enhancements with privacy and their impact on the validity of the configurations.

In our classification, we considered the simplest and most commonly used attribution rules (listed in Section 3.2), which operate on a single user's data. It would be interesting to investigate the interplay between DP and more advanced alternatives such as those based on the Shapley value (e.g., [35]) as well as data-driven attribution (DDA), which by contrast is a class of attribution rules that operate on the entire dataset (across users).

Incentives. While there has been interesting prior work at the intersection of privacy and economics, e.g., [4, 23], understanding privacy and incentives in the conversion measurement setting would greatly benefit from further investigation. For instance, our study captures the case where a single ad tech company would like to query the DP conversion measurement system across multiple publishers. In reality, multiple ad techs, which are often competing but could in principle collude, would want to issue DP queries on overlapping impressions and conversions taking place on the same set of publishers and advertisers.

Finally, while a *user* contribution bounding scope can admit valid configurations, it is vulnerable to "crowding out attacks", where, e.g., one publisher can exhaust the contribution bound of a user (by showing them a large number of impressions). Incorporating the economic incentives of different entities into the analysis of privacy and utility of conversion measurement systems seems worthwhile.

Privacy-Utility Trade-offs of Various Tasks. The classification in this work is in terms of sensitivity, which is closely related to additive noise mechanisms as these naturally calibrate the noise scale to the sensitivity. While most proposed DP conversion measurement systems follow this sensitivity and additive noise paradigm, it would be valuable to consider other families of mechanisms, and to quantify the privacy-utility trade-offs of various estimation tasks.

Correlation across Users. There are settings where different users' data can be correlated. In ad measurement, this can arise if multiple users watch the same ad (e.g., on a TV). Then, their impressions are correlated, and extra care is needed when applying DP [38]. We leave the exploration of this interesting setting for future work.

DP Advertising. Applying DP in practical advertising systems has been notoriously difficult for the reasons considered in this work, namely: How to define adjacent datasets? Given the correlation between a given user's behavior across (a practically unbounded number of) websites (and/or apps), can DP be applied without adding a disproportionately large amount of noise that would preclude the measurement of simple statistics? We hope that our work provides a stepping stone for tackling these questions in the setting of attribution measurement—the cornerstone of digital advertising—and leads to solid deployments of DP in practical advertising systems.

REFERENCES

[1] 2023. *SKAdNetwork*. https://developer.apple.com/documentation/storekit/skadnetwork/.

[2] Martin Abadi, Andy Chu, Ian Goodfellow, H Brendan McMahan, Ilya Mironov, Kunal Talwar, and Li Zhang. 2016. Deep learning with differential privacy. In *CCS*. 308–318.

[3] John M Abowd. 2018. The US Census Bureau adopts differential privacy. In *KDD*. 2867–2867.

[4] John M Abowd and Ian M Schmutte. 2019. An economic analysis of privacy protection and statistical accuracy as social choices. *American Economic Review* 109, 1 (2019), 171–202.

[5] Deepak Agarwal, Rahul Agrawal, Rajiv Khanna, and Nagaraj Kota. 2010. Estimating rates of rare events with multiple hierarchies through scalable log-linear models. In *KDD*. 213–222.

[6] Hidayet Aksu, Badih Ghazi, Pritish Kamath, Ravi Kumar, Pasin Manurangsi, Adam Sealfon, and Avinash V Varadarajan. 2023. Summary Reports Optimization in the Privacy Sandbox Attribution Reporting API. *arXiv preprint arXiv:2311.13586* (2023).

[7] Android. 2022. Attribution reporting. https://developer.android.com/design-for-safety/privacy-sandbox/attribution.

[8] Apple Differential Privacy Team. 2017. Learning with privacy at scale. *Apple Machine Learning Journal* (2017).

[9] Frederick Ayala-Gómez, Ismo Horppu, Erlin Gülbenkoğlu, Vesa Siivola, and Balázs Pejó. 2022. Show me the Money: Measuring Marketing Performance in F2P Games using Apple's App Tracking Transparency Framework. (2022).

[10] Andrea Bittau, Úlfar Erlingsson, Petros Maniatis, Ilya Mironov, Ananth Raghunathan, David Lie, Mitch Rudominer, Ushasree Kode, Julien Tinnes, and Bernhard Seefeld. 2017. Prochlo: Strong privacy for analytics in the crowd. In *SOSP*. 441–459.

[11] Albert Cheu, Adam D. Smith, Jonathan Ullman, David Zeber, and Maxim Zhilyaev. 2019. Distributed Differential Privacy via Shuffling. In *EUROCRYPT*. 375–403.

[12] Hana Choi, Carl F Mela, Santiago R Balseiro, and Adam Leary. 2020. Online display advertising markets: A literature review and future directions. *Information Systems Research* 31, 2 (2020), 556–575.

[13] Matthew Dawson, Badih Ghazi, Pritish Kamath, Kapil Kumar, Ravi Kumar, Bo Luan, Pasin Manurangsi, Nishanth Mundru, Harikesh Nair, Adam Sealfon, and Shengyu Zhu. 2023. Optimizing Hierarchical Queries for the Attribution Reporting API. In *AdKDD*. arXiv:2308.13510 https://arxiv.org/abs/2308.13510

[14] Damien Desfontaines, James Voss, Bryant Gipson, and Chinmoy Mandayam. 2022. Differentially private partition selection. *Proceedings on Privacy Enhancing Technologies* 1 (2022), 339–352.

[15] Benjamin Dick. 2016. Digital Attribution Primer 2.0. https://www.iab.com/wp-content/uploads/2016/10/Digital-Attribution-Primer-2-0-FINAL.pdf.

[16] Bolin Ding, Janardhan Kulkarni, and Sergey Yekhanin. 2017. Collecting telemetry data privately. In *NeurIPS*. 3571–3580.

[17] Cynthia Dwork, Krishnaram Kenthapadi, Frank McSherry, Ilya Mironov, and Moni Naor. 2006. Our data, ourselves: Privacy via distributed noise generation. In *EUROCRYPT*. 486–503.

[18] Cynthia Dwork, Frank McSherry, Kobbi Nissim, and Adam D. Smith. 2006. Calibrating Noise to Sensitivity in Private Data Analysis. In *TCC*. 265–284.

[19] Cynthia Dwork and Aaron Roth. 2014. The Algorithmic Foundations of Differential Privacy. *Found. Trends Theor. Comput. Sci.* 9, 3-4 (2014), 211–407.

[20] Úlfar Erlingsson, Vitaly Feldman, Ilya Mironov, Ananth Raghunathan, Kunal Talwar, and Abhradeep Thakurta. 2019. Amplification by shuffling: From local to central differential privacy via anonymity. In *SODA*. 2468–2479.

[21] David Evans, Vladimir Kolesnikov, and Mike Rosulek. 2017. A pragmatic introduction to secure multi-party computation. *Foundations and Trends® in Privacy and Security* 2, 2-3 (2017).

[22] Alexandre Evfimievski, Johannes Gehrke, and Ramakrishnan Srikant. 2003. Limiting privacy breaches in privacy preserving data mining. In *PODS*. 211–222.

[23] Arpita Ghosh and Aaron Roth. 2011. Selling privacy at auction. In *EC*. 199–208.

[24] Nils Homer, Szabolcs Szelinger, Margot Redman, David Duggan, Waibhav Tembe, Jill Muehling, John V Pearson, Dietrich A Stephan, Stanley F Nelson, and David W Craig. 2008. Resolving individuals contributing trace amounts of DNA to highly complex mixtures using high-density SNP genotyping microarrays. *PLoS genetics* 4, 8 (2008), e1000167.

[25] PK Kannan, Werner Reinartz, and Peter C Verhoef. 2016. The path to purchase and attribution modeling: Introduction to special section. *International Journal of Research in Marketing* 33, 3 (2016), 449–456.

[26] Shiva Prasad Kasiviswanathan, Homin K Lee, Kobbi Nissim, Sofya Raskhodnikova, and Adam Smith. 2011. What can we learn privately? *SIAM J. Comput.* 40, 3 (2011), 793–826.

[27] Kuang-Chih Lee, Burkay Orten, Ali Dasdan, and Wentong Li. 2012. Estimating conversion rate in display advertising from past performance data. In *KDD*. 768–776.

[28] Aditya Krishna Menon, Krishna-Prasad Chitrapura, Sachin Garg, Deepak Agarwal, and Nagaraj Kota. 2011. Response prediction using collaborative filtering with hierarchies and side-information. In *KDD*. 141–149.

[29] Maud Nalpas, Sam Dutton, and Alexandra White. 2021. Chrome Attribution Reporting. https://developer.chrome.com/docs/privacy-sandbox/attribution-reporting/.

[30] Arvind Narayanan and Vitaly Shmatikov. 2008. Robust de-anonymization of large sparse datasets. In *S & P*. 111–125.

[31] Balázs Pejó and Damien Desfontaines. 2022. Neighborhood Definition (N). In *Guide to Differential Privacy Modifications: A Taxonomy of Variants and Extensions*. Springer, 19–28.

[32] Joseph J Pfeiffer III, Denis Charles, Davis Gilton, Young Hun Jung, Mehul Parsana, and Erik Anderson. 2021. Masked LARk: Masked Learning, Aggregation and Reporting worKflow. *arXiv preprint arXiv:2110.14794* (2021).

[33] Rómer Rosales, Haibin Cheng, and Eren Manavoglu. 2012. Post-click conversion modeling and analysis for non-guaranteed delivery display advertising. In *WSDM*. 293–302.

[34] Justin Schuh. 2020. Building a more private web: A path towards making third party cookies obsolete. https://blog.chromium.org/2020/01/building-more-private-web-path-towards.html.

[35] Raghav Singal, Omar Besbes, Antoine Desir, Vineet Goyal, and Garud Iyengar. 2022. Shapley meets uniform: An axiomatic framework for attribution in online advertising. *Management Science* (2022).

[36] Latanya Sweeney. 2015. Only you, your doctor, and many others may know. *Technology Science* 2015092903, 9 (2015), 29.

[37] Martin Thomson. 2022. Privacy Preserving Attribution for Advertising. https://blog.mozilla.org/en/mozilla/privacy-preserving-attribution-for-advertising/.

[38] Michael Carl Tschantz, Shayak Sen, and Anupam Datta. 2020. SoK: Differential privacy as a causal property. In *2020 IEEE Symposium on Security and Privacy (SP)*. IEEE, 354–371.

[39] Salil Vadhan. 2017. The complexity of differential privacy. In *Tutorials on the Foundations of Cryptography*. Springer, 347–450.

[40] John Wilander. 2020. Full Third-Party Cookie Blocking and More. https://webkit.org/blog/10218/full-third-party-cookie-blocking-and-more/.

[41] John Wilander. 2021. Introducing Private Click Measurement, PCM. https://webkit.org/blog/11529/introducing-private-click-measurement-pcm/.

[42] Luke Winstrom. 2023. A proposal for privacy preserving ad attribution measurement using Prio-like architecture. https://github.com/patcg/proposals/issues/17.

[43] Marissa Wood. 2019. Today's Firefox Blocks Third-Party Tracking Cookies and Cryptomining by Default. https://blog.mozilla.org/en/products/firefox/todays-firefox-blocks-third-party-tracking-cookies-and-cryptomining-by-default/.

A DEFERRED STATEMENTS AND PROOFS

In this section, we provided all the remaining statements and proofs for our classification results from Section 6. We remark that, in all the datasets that we construct below for the invalidity results, all the impressions/conversions belong to a single user, and we will henceforth not specify this explicitly. Throughout, let $\mathcal{D}_{\text{attr}}$ and $\mathcal{D}'_{\text{attr}}$ denote the attributed datasets resulting from $\mathcal{D}, \mathcal{D}'$, respectively.

A.1 *Conversion* Adjacency Relation

We start with the simple proof for the validity of the *conversion* adjacency relation (Theorem 4).

Theorem 4. *For any attribution rule, the conversion adjacency relation (without contribution bound enforcement) constitutes a valid configuration with $r = C_0 = 1$.*

Proof. Consider two adjacent datasets $\mathcal{D}, \mathcal{D}'$ such that $\mathcal{D}'$ results from adding a conversion c to $\mathcal{D}$. Notice that $\mathcal{D}'_{\text{attr}}$ is exactly equal to $\mathcal{D}_{\text{attr}}$ together with $(i'_1, c, w_1), \dots, (i'_\ell, c, w_\ell)$ where $i'_1, \dots, i'_\ell$ are the impressions that come before c and correspond to the same advertiser as c, and $(w_1, \dots, w_\ell) = \mathsf{a}((i'_1, \dots, i'_\ell), c)$. Therefore, we have $\|\mathcal{D}_{\text{attr}} - \mathcal{D}'_{\text{attr}}\|_1 \le \sum_{j\in[\ell]} w_j = 1$, where the equality follows from the definition of an attribution rule. Thus, the *conversion* adjacency relation (without contribution bound enforcement) constitutes a valid configuration with $r = C_0 = 1$ as desired. □

A.2 Pre-Attribution Enforcement

Once again, pre-attribution enforcement results in valid configurations for all adjacency relations and attribution rules.

Theorem 5. *For any attribution rule, any adjacency relation with contribution bound enforced pre-attribution constitutes a valid configuration with $C_0 = 1$.*

Proof. Consider adjacent datasets $\mathcal{D}, \mathcal{D}'$ such that $\mathcal{D}'$ results from adding impressions $\tilde{i}_1, \dots, \tilde{i}_a$ and conversions $\tilde{c}_1, \dots, \tilde{c}_b$ all associated with a single contribution bounding scope s. Let $\tilde{I} := \{\tilde{i}_1, \dots, \tilde{i}_a\}$ and $\tilde{C} := \{\tilde{c}_1, \dots, \tilde{c}_b\}$.

Observe that, in all privacy notions considered, the set of impressions considered for attribution to $c \in \tilde{C}$ (on Line 7 of Algorithm 2) must come from $\tilde{I}$. This means that the attribution rule applied to $\tilde{C}$ does not affect any scope apart from s. This in turn implies that, in both $\mathcal{D}, \mathcal{D}'$, the impressions fed into the attribution rule for each $c_j \notin \tilde{C}$ (on Line 14) are the same apart from those in $\tilde{I}$.

Thus, $\|\mathcal{D}_{\text{attr}} - \mathcal{D}'_{\text{attr}}\|_1$ can be at most the number of conversions c_j for which at least one impression from $\tilde{I}$ is included in its attribution rule (on Line 14). However, when this happens, we subtract one from the contribution bound of scope s; therefore, the number of such c_j's can be at most r. Hence, we have $\|\mathcal{D}_{\text{attr}} - \mathcal{D}'_{\text{attr}}\|_1 \le r$. □

A.3 Post-Attribution Enforcement

We will now move on to the second—and more subtle—privacy enforcement: post-attribution.

A.3.1 Validity of User and User × Advertiser Relations.

We start by considering the *user* adjacency relation.

Theorem 6. *For any attribution rule, the user adjacency relation with contribution bound enforced post-attribution constitutes a valid configuration with $C_0 = 1$.*

Proof. Consider adjacent datasets $\mathcal{D}, \mathcal{D}'$ such that $\mathcal{D}'$ results from adding impressions $\tilde{i}_1, \dots, \tilde{i}_a$ and conversions $\tilde{c}_1, \dots, \tilde{c}_b$ all associated with a single user U. Let $\tilde{I} := \{\tilde{i}_1, \dots, \tilde{i}_a\}$ and $\tilde{C} := \{\tilde{c}_1, \dots, \tilde{c}_b\}$.

Notice that outside of the conversions in $\tilde{C}$, the attribution system produces exactly the same result for both $\mathcal{D}, \mathcal{D}'$. Therefore, we have $\|\mathcal{D}_{\text{attr}} - \mathcal{D}'_{\text{attr}}\|_1 = \sum_{\tilde{i}\in\tilde{I},\tilde{c}\in\tilde{C}} w_{\mathcal{D}'_{\text{attr}}}(\tilde{i}, \tilde{c})$. Moreover, the post-attribution enforcement exactly ensures that the quantity on the right-hand side is at most r. □

We next consider the *user* × *advertiser* adjacency relation.

Theorem 7. *For any attribution rule, the user × advertiser adjacency relation with contribution bound enforced post-attribution constitutes a valid configuration with $C_0 = 1$.*

Proof of Theorem 7. Similar to the proof of Theorem 6, consider two adjacent datasets $\mathcal{D}, \mathcal{D}'$ such that $\mathcal{D}'$ results from adding impressions $\tilde{i}_1, \dots, \tilde{i}_a$ and conversions $\tilde{c}_1, \dots, \tilde{c}_b$ all associated with a single user U and a single advertiser A. Let $\tilde{I} := \{\tilde{i}_1, \dots, \tilde{i}_a\}$ and $\tilde{C} := \{\tilde{c}_1, \dots, \tilde{c}_b\}$.

Outside of the conversions in $\tilde{C}$, the attribution system produces exactly the same result for both $\mathcal{D}, \mathcal{D}'$. This implies that $\|\mathcal{D}_{\text{attr}} - \mathcal{D}'_{\text{attr}}\|_1 = \sum_{\tilde{i}\in\tilde{I},\tilde{c}\in\tilde{C}} w_{\mathcal{D}'_{\text{attr}}}(\tilde{i}, \tilde{c})$, which is at most r due to post-attribution enforcement. □

A.3.2 Impression Relation.

We next consider the *impression* adjacency relation. As stated earlier, the validity in this case does depend on the attribution rule. We note that the validity for FTA was already shown in the main body (Theorem 2).

Last-Touch Attribution. Again this results in a valid configuration (Theorem 8). The idea is similar to before, except that instead of re-attribution to the second impression, the re-attribution happens to the second-to-last impression before the conversion.

Theorem 8. *For last-touch attribution, the impression adjacency relation with contribution bound enforced post-attribution constitutes a valid configuration with $C_0 = 2$.*

Proof. Consider adjacency datasets $\mathcal{D}, \mathcal{D}'$ such that $\mathcal{D}'$ results from removing an impression $\tilde{i}$. Let A be $\tilde{i}$'s advertiser, U be its user, and C_A denote the set of conversions attributed to $\tilde{i}$ in $\mathcal{D}$. We consider two cases, based on whether $\tilde{i}$ is the first impression w.r.t. its advertiser A and its user U (in $\mathcal{D}$).

- Case I: $\tilde{i}$ is the first impression w.r.t. the advertiser A and the user U. In this case, all conversions in C_A become unattributed. As a result,
$$\|\mathcal{D}_{\text{attr}} - \mathcal{D}'_{\text{attr}}\|_1 = \sum_{c_j} w_{\mathcal{D}_{\text{attr}}}(\tilde{i}, c_j) \le r,$$
where the inequality follows from the post-attribution contribution bound enforcement for the impression $\tilde{i}$.
- Case II: $\tilde{i}$ is *not* the first impression w.r.t. the advertiser A and the user U. Let $\tilde{i}'$ denote the impression that comes right before $\tilde{i}$ w.r.t. the advertiser A and the user U. $\tilde{i}$ and $\tilde{i}'$ are the

only two impressions whose attributions change between the two datasets. Furthermore, all attributions to $\tilde{i}'$ in $\mathcal{D}$ remains unchanged in $\mathcal{D}'$, because the corresponding conversions must come before $\tilde{i}$ and therefore before all conversions in C_A. In other words, the only changes to $\tilde{i}'$ are the additional attributions from conversions in C_A. As a result, we have

$$\|\mathcal{D}_{\text{attr}} - \mathcal{D}'_{\text{attr}}\|_1 = \sum_{c_j} w_{\mathcal{D}_{\text{attr}}}(\tilde{i}, c_j) + \sum_{c_j \in C_A} w_{\mathcal{D}'_{\text{attr}}}(\tilde{i}', c_j) \leq 2r,$$

where the inequality again follows from the post-attribution contribution bound enforcement for impressions $\tilde{i}$ and $\tilde{i}'$.

In all cases, we can conclude that $\|\mathcal{D}_{\text{attr}} - \mathcal{D}'_{\text{attr}}\|_1 \leq 2r$, which means that this is a valid configuration with $C_0 = 2$ as desired. □

Uniform Multi-Touch and Exponential Time Decay Attributions. Next, we move on the prove the invalidity for uniform multi-touch and exponential time decay attribution rules.

Theorem 9. *For uniform multi-touch attribution, the impression adjacency relation with contribution bound enforced post-attribution constitutes an invalid configuration.*

Since exponential time decay can be used to implement uniform multi-touch rule when assuming that the time of every impression is the same, Corollary 1 is immediate.

Corollary 1. *For exponential time decay attribution, the impression adjacency relation with contribution bound enforced post-attribution constitutes an invalid configuration.*

The idea of the proof of Theorem 9 is to construct an impression that is involved in many conversions' attribution rule. Due to the uniform attribution, such an impression naturally "draws" large weights from the overall attribution. Therefore, when removing it, such weights get re-attributed back to the other impressions, resulting in a large change in the attributed dataset.

Proof. Let $r = 1$ and let p be any positive integer. Consider $\mathcal{D}$ such that there are p impressions $i_1, \dots, i_p$ and $\binom{p+1}{2}$ conversions $c_{1,1}, c_{2,1}, c_{2,2}, \dots, c_{p,p}$ such that conversion $c_{k,1}, \dots, c_{k,k}$ appears after i_k (and before i_{k+1} if it exists); all impressions and conversions are associated with the same advertiser.

Then, let $\mathcal{D}'$ denote $\mathcal{D}$ but with i_1 removed.

Under the uniform attribution rule, it is simple to see that, for all $j = 2, \dots, p$, we have $w_{\mathcal{D}_{\text{attr}}}(i_j, c_{j,k}) = \frac{1}{j}$ for all $k = 1, \dots, j$ and $w_{\mathcal{D}'_{\text{attr}}}(i_j, c_{j,k}) = \frac{1}{j-1}$ for all $k = 1, \dots, j-1$. Therefore,

$$\begin{aligned}\|\mathcal{D}_{\text{attr}} - \mathcal{D}'_{\text{attr}}\|_1 &\geq \sum_{j=2}^{p} \sum_{k \in [j]} |w_{\mathcal{D}_{\text{attr}}}(i_j, c_{j,k}) - w_{\mathcal{D}'_{\text{attr}}}(i_j, c_{j,k})| \\ &= \sum_{j=2}^{p} \frac{2}{j} \geq 2\ln(p/2).\end{aligned}$$

By taking $p \longrightarrow \infty$, we can see that this is not bounded above by $C_0 \cdot r$ for any constant C_0. □

U-Shaped Attribution. U-shaped attribution rule also results in an invalid configuration (Theorem 10). The construction is an adaptation of the above construction for the uniform attribution. The rough idea is that the "middle" part of U-shaped attribution rule is essentially the uniform attribution (scaled by a factor of 0.2). This allows us to use the U-shaped attribution rule to "implement" the uniform distribution rule.

Theorem 10. *For U-shaped attribution, the impression adjacency relation with contribution bound enforced post-attribution constitutes an invalid configuration.*

Proof of Theorem 10. Let $r = 1$ and let p be any positive integer greater than three. Consider $\mathcal{D}$ such that there are p impressions $i_1, \dots, i_p$ and $\binom{p-1}{2} - 1$ conversions, constructed as follows: For every $j = 4, \dots, p$, there are $j - 2$ impressions $c_{j,1}, \dots, c_{j,j-2}$ after i_j (and before i_{j+1} if it exists). All impressions and conversions are associated with the same advertiser.

Then, let $\mathcal{D}'$ denote $\mathcal{D}$ but with i_1 removed.

Under the U-shaped attribution rule, for all $j = 4, \dots, p-1$, we have $w_{\mathcal{D}_{\text{attr}}}(i_j, c_{j+1,k}) = \frac{0.2}{j-1}$ for all $k = 1, \dots, j-1$ and $w_{\mathcal{D}'_{\text{attr}}}(i_j, c_{j,k}) = \frac{0.2}{j-2}$ for all $k = 1, \dots, j-2$. Therefore,

$$\begin{aligned}\|\mathcal{D}_{\text{attr}} - \mathcal{D}'_{\text{attr}}\|_1 &\geq \sum_{j=4}^{p-1} \sum_{k \in [j-2]} |w_{\mathcal{D}_{\text{attr}}}(i_j, c_{j,k}) - w_{\mathcal{D}'_{\text{attr}}}(i_j, c_{j,k})| \\ &= \sum_{j=4}^{p-1} \sum_{k \in [j-2]} \left(\frac{0.2}{j-2} - \frac{0.2}{j-1}\right) \\ &\geq \sum_{j=4}^{p-1} \frac{0.2}{j-1} \quad \geq \quad 0.2 \ln\left(\frac{p-1}{4}\right).\end{aligned}$$

By taking $p \longrightarrow \infty$, we can see that this is not bounded above by $C_0 \cdot r$ for any constant C_0. □

A.3.3 User × Publisher × Advertiser Relation. We now consider the last adjacency relation: *User × Publisher × Advertiser.*

First-Touch Attribution. FTA is the only valid configuration in this case. The key observation here is that, although removing impressions for a scope may result in re-attribution of multiple conversions, these re-attribution is only to the first impression left for that advertiser. Thereby, an argument analogous to that in the proof of Theorem 8 can show the validity of the configuration.

Theorem 11. *For first-touch attribution, the user × publisher × advertiser adjacency relation with contribution bound enforced post-attribution constitutes a valid configuration with $C_0 = 2$.*

Proof of Theorem 11. Consider two adjacent datasets $\mathcal{D}, \mathcal{D}'$ such that $\mathcal{D}'$ results from removing impressions $\tilde{i}_1, \dots, \tilde{i}_a$ (ordered from oldest to newest) all associated to a single user U, a single publisher P and a single advertiser A. Let C_A denote the set of conversions at the advertiser A and the user U.

We consider two cases, based on whether $\tilde{i}_1$ is the first impression w.r.t. its advertiser A and the user U.

- Case I: $\tilde{i}_1$ is *not* the first impression w.r.t. the advertiser A and the user U. In this case, the two attributed datasets $\mathcal{D}_{\text{attr}}, \mathcal{D}'_{\text{attr}}$ are exactly the same, because all conversions in C_A are attributed to the first impression w.r.t. the advertiser A and the user U (which is not among $\tilde{i}_1, \dots, \tilde{i}_a$).

- Case II: $\tilde{i}_1$ is the first impression w.r.t. the advertiser A and the user U. In this case, all conversions $c_j \in C_A$ are attributed to $\tilde{i}_1$. These are the only conversions whose attributions change between $\mathcal{D}$ and $\mathcal{D}'$.

  To analyze this change, consider two subcases, whether $\tilde{i}_1, \ldots, \tilde{i}_a$ are the only impressions in $\mathcal{D}$ from the advertiser A and the user U.

  - Case IIa: $\tilde{i}_1, \ldots, \tilde{i}_a$ are the only impressions in $\mathcal{D}$ from the advertiser A and the user U. In this case, all conversions in C_A become unattributed in $\mathcal{D}'$. Thus, $\|\mathcal{D}_{\text{attr}} - \mathcal{D}'_{\text{attr}}\|_1 = \sum_{k\in[a]} \sum_{c_j} w_{\mathcal{D}_{\text{attr}}}(\tilde{i}_k, c_j) \leq r$, where the inequality follows from the post-attribution contribution bound enforcement for the contribution bounding scope (U, P, A).
  - Case IIb: $\tilde{i}_1, \ldots, \tilde{i}_a$ are not the only impressions in $\mathcal{D}$ from the advertiser A and the user U. Let $\tilde{i}'$ be the first impression w.r.t. the advertiser A and the user U that is not among $\tilde{i}_1, \ldots, \tilde{i}_a$. In this case, all conversions in C_A are attributed to $\tilde{i}'$ in $\mathcal{D}'$. Furthermore, no conversions are attributed to $\tilde{i}'$ (or other impressions in $\tilde{i}'$'s contribution bounding scope) in $\mathcal{D}$ because $\tilde{i}_1$ comes first for the same advertiser A and the same user U. Therefore, we have

$$\|\mathcal{D}_{\text{attr}} - \mathcal{D}'_{\text{attr}}\|_1 = \sum_{k\in[a]} \sum_{c_j} w_{\mathcal{D}_{\text{attr}}}(\tilde{i}_k, c_j) + \sum_{c_j} w_{\mathcal{D}'_{\text{attr}}}(\tilde{i}', c_j) \leq 2r,$$

  where the inequality again follows from the post-attribution contribution bound enforcement for the contribution bounding scope (U, P, A) and the contribution bounding scope of $\tilde{i}'$.

In all cases, we can conclude that $\|\mathcal{D}_{\text{attr}} - \mathcal{D}'_{\text{attr}}\|_1 \leq 2r$, which means that this is a valid configuration with $C_0 = 2$ as desired. □

Uniform Multi-Touch, Exponential Time Decay and U-Shaped Attributions. The invalidity of these attribution rules (Corollary 2) follows directly from that of the *impression* case.

Corollary 2. *For uniform multi-touch attribution, exponential time decay attribution and U-shaped attribution, the user × publisher × advertiser adjacency relation with contribution bound enforced post-attribution constitutes an invalid configuration.*

Proof of Corollary 2. We may use the same constructions as in the proof of Theorem 9, Corollary 1, and Theorem 10 respectively, except we assign each $i_1, \ldots, i_p$ to p different publishers. This way $\mathcal{D}, \mathcal{D}'$ are adjacent under the *user* × *publisher* × *advertiser* relation. The remainder of the proof then ensures that $\|\mathcal{D}_{\text{attr}} - \mathcal{D}'_{\text{attr}}\|_1$ is not bounded above by $C_0 \cdot r$ for any constant C_0. □