Google Marketing Platform Non-O&O/Display ARA E2E CMA Test Study

Important Note: The following results are estimates at the specific time of the experiment (March 2024) and should not be taken as representative of the impact of varied third-party cookie availability on conversion measurement reporting quality. Since March 2024, there have been additional improvements made to the Privacy Sandbox ARA, technical improvements from GMP to better utilise the ARA, and outreach efforts to optimize adtech/publisher set up. The assumptions and results of the experiment do not account for Chrome's announcement proposing an updated approach that elevates user choice rather than deprecating third-party cookies. We expect additional performance improvements compared to the results of this experiment due to the factors above.

Overview

This document is a writeup of the results of a standalone test using the Privacy Sandbox Attribution Reporting API (ARA) on Floodlight Attribution measurement for Non-O&O/Display. The efficacy of the ARA was measured using an Absolute Percentage Error (APE) metric to compare the number of Web-to-Web attributed conversions measured with the ARA to the number of attributed conversions measured with third-party cookies. Web-to-Web refers to both the ad event and conversion taking place on the Web.

APE compares the conversion volume measured by the ARA to conversion volume measured by third-party cookies, expressed as a percentage, with higher APE buckets indicating lower efficacy, and lower APE buckets indicating higher efficacy.

We recommend that the results are interpreted as early nascent indicators rather than precise estimates of varied third-party cookie availability impact on advertisers via the ARA. The experiment does not provide an accurate approximation for reporting quality in the true end-state given the following factors. These could be potential sources for significant performance improvements in the future:
- Efforts are underway to ensure that Google Marketing Platform (GMP) is prepared for varied third-party cookie availability on non-O&O/Display, and these enhancements will continue to improve reporting quality.
- GMP is in early stages of optimization and has outstanding quality and transmission loss improvements. We are working to resolve outstanding data issues to improve reporting quality.
- GMP is working with third-party publishers and adtech providers to drive adoption of the ARA and support the use of the attributionsrc parameter to improve coverage for Floodlight measurement.
- GMP is working with customers to optimize their conversion setup in the ARA to help improve conversion loss.

- Chrome [proposed](#) an updated approach that elevates user choice, as opposed to the previous plan to deprecate third-party cookies by default for Chrome users.

Experiment summary

This experiment compares Web2Web attributed conversions measured with the ARA to existing production attribution measured with third-party cookies using an [Absolute Percentage Error (APE) metric](#) based on conversion data from March 2, 2024. APE is the ratio of the error over the number of conversions measured with third-party cookies. APE buckets are predefined as 0-10%, 10-20%, 30-40%, 40-60%, and 60-100%. Buckets with higher APE indicate lower efficacy as there is a greater deviation between conversions measured with the ARA and conversions measured with third-party cookies, and buckets with lower APE indicate higher efficacy as there is a smaller deviation between conversions measured with the ARA and conversions measured with third-party cookies.

The ARA based measurement combines event level and aggregate level reports on all currently integrated web traffic. The processing of the ARA data represents an initial end-to-end version of GMP's planned processing approach assuming that third-party cookies would be fully deprecated. The test does not use Mode B traffic.

APE distributions are provided for view-through conversion (VTC) and click-through conversion (CTC) measurement for the three enterprise products using Floodlight Attribution measurement: Campaign Manager 360 (CM360), Display and Video 360 (DV360) and Search Ads 360 (SA360).

Key findings of the experiment:
- VTC measurement, which is entirely dependent on the ARA, showed high Absolute Percentage Error (APE), with the bulk of CM360 and DV360 campaigns falling in the 60% to 100% APE bucket when analyzing traffic already integrated with the ARA and analyzing all traffic. When only analyzing traffic already integrated with the ARA, VTC and CTC results show improvements. VTC results for this integrated traffic segment showed that 65% of campaigns fell within the 60-100% APE range, compared to 85% of campaigns when considering all traffic. GMP is working with third-party publishers and adtech providers to drive adoption of the ARA and support the use of the attributionsrc parameter to improve coverage for Floodlight measurement. We also observed conversion loss in customers that measure a large number of conversions. GMP is working with these customers to optimize their conversion setup for use of the ARA.
- CTCs which are only partially reliant on the ARA due to the presence of first-party cookies showed significantly better APE, particularly on SA360, where the vast majority of campaigns fell in the 0%-20% APE buckets.[1]

Significant integration and quality work remains to be completed to prepare for varied third-party cookie availability. This is consistent with the ongoing efforts referenced in the [Overview](#) section above. While ARA is recovering conversion measurement traffic, these enhancements will continue to improve the efficacy.

Experiment Design

Experiment Methodology

To determine directional estimates of the ARA's efficacy for conversion measurement, our high-level approach is as follows:

1. Invoke the ARA for all browsers/devices instances that are part of the General Availability phase where API integration has been completed.
   a. On Chrome: we expect to invoke the API on 100% of browsers where APIs are available and where we have Non-O&O/Display registration coverage.
2. Post-process the Event and Aggregatable reports from the ARA, to merge and de-bias the data set of attributed conversions.
3. Continue measuring attributed conversions using third-party cookies over all the users, as is done today.
4. Determine reporting impact metrics by comparing the third-party cookie dataset and post-processed dataset, to determine the differences.

We favor this approach over Mode B style experimentation for the ARA standalone testing because it allows us to evaluate all traffic that has been integrated with the ARA. This is possible for the standalone ARA because conversions can be measured in two different ways simultaneously (whereas it is not possible to target users in two different ways simultaneously - counterfactuals in targeting are very difficult to achieve).

Experiment Scope

- Products: CM360 (Campaign Manager 360), DV360 (Display & Video 360), and SA360 (Search Ads 360)
- Traffic type: VTC and CTC
- Traffic: web-to-web (W2W) only (Privacy Sandbox on Android integration is still ongoing)
- Simulations: Post-processed ARA data only, no simulations
- Time frame: Conversions on March 2, 2024 with 30 day lookback window for views/clicks
  - Results are stable for conversions from March 1 - 7, 2024
- Conversions: All conversions tracked by Floodlight (biddable and non biddable)
- Handling of traffic categories:
  - Unaffected: Traffic that is not reliant on third-party cookies
  - Mitigation ready: Slice of traffic that is impacted by third party cookie deprecation and is available via API registrations, and post-processing support is available
  - Mitigation not ready: Such traffic could result from various circumstances: 1) Registration or post-processing steps aren't fully implemented, 2) API does not support certain use cases by design

First we're providing reporting quality data for VTC and CTC, broken down by product across *Unaffected and Mitigation ready* traffic categories. The *Mitigation not ready traffic* is excluded to allow for more direct comparison between Test and Control arm on the same set of traffic. We also provide a supplementary analysis across *Unaffected, Mitigation ready* and *Mitigation not ready* traffic categories to show the impact of not having ARA registration. This supplementary analysis includes a breakdown by both product and third party conversion reliance for CTC.

Metrics

APE

In this experiment, we are reporting on APE(T) as the metric to measure ARA's efficacy for conversion measurement reporting, and a general evaluation strategy of showing percent of campaigns passing a certain quality bar, i.e. *APE(T) < a chosen threshold e.*

APE(T) is the absolute percentage error, adjusted for campaigns with very small conversion counts. The methodology is described [here](#):

$$APE(T) \; = \; \frac{|\,error\,|}{max\,\{T,\,true\;\#\;of\;conversions\}}$$

The metric parameters (T, threshold e, time horizon) specified for this report are:
- T threshold: 5 conversions (to adjust for campaigns with very small conversion counts)
- Time horizon: Conversions on March 2, 2024 with 30 day lookback window for views/clicks

Distribution of ARA's efficacy for conversion reporting

We summarize the distribution of reporting quality over all campaigns weighted by certain weighting schemes for each product. When computing the percent, we will weight each campaign by:
- CM360: conversion weighted (as a reporting/measurement product)
- DV360: revenue weighted (as a media buying platform)
- SA360: revenue weighted (as a media buying platform)

Towards that end, we use *APE(T)* of each campaign directly, and present numbers and percentages of campaigns in pre-defined buckets of quality scores (*see table below*), with buckets with higher APE indicating lower efficacy as there is a greater deviation between conversions measured with the ARA and conversions measured with third-party cookies, and buckets with lower APE indicate higher efficacy as there is a smaller deviation between conversions measured with the ARA and conversions measured with third-party cookies.

All numbers reported are percentages (i.e. percent of conversions or revenue weighted campaigns falling into the APE bucket). Additional granular results are also provided for the following data splits.

Third-Party Conversion Reliance

In the supplementary analysis for CTC, we analyze data by third-party conversion reliance: percent of third-party cookie conversions in the Control arm of a given campaign that is not attributable via first-party cookies. We present reporting quality in three buckets of third-party conversion reliance: 0%-20%, 20%-50%, and 50%-100%, broken down by product. As all Floodlight VTCs are reliant on third-party conversions, this analysis is only applicable to CTC.

Analysis

ARA Efficacy for VTC Reporting

This analysis of ARA's efficacy for VTC reporting contains:
- Test: Conversion data based ARA attribution from integrated W2W traffic for Chrome M118+[1]
- Control: Conversion data based on third-party cookie attribution from the same traffic as test arm

First, we report ARA's Efficacy for VTC reporting by product. Distribution of efficacy is weighted by conversions for CM360 and revenue for DV360 across all breakdowns. *Mitigation not ready* traffic is excluded, allowing for more direct comparison between Test and Control arm on the same set of traffic.

ARA Efficacy's for VTC Reporting distribution by product (excluding Mitigation not ready traffic)
Numbers reported are percents

	APE Buckets					
	0%-10%	10%-20%	20%-30%	30%-40%	40%-60%	60%-100%
CM360, conv weighted	4.0	5.1	4.2	5.0	16.7	65.0
DV360, rev weighted	7.7	10.4	6.9	10.3	19.9	44.8

Supplementary analysis: Analyzing all traffic

The analysis for VTC reporting contains:
- Test: Conversion data based ARA attribution from Chrome versions M115+
- Control: Conversion data based on third-party cookie attribution from all Chrome versions.

By Product

ARA's Efficacy for VTC Reporting distribution by product
Numbers reported are percents

---

[1] As the ARA launched in general availability in M118, excluding all prior versions with partial API support for the purpose of this analysis.

	APE Buckets					
	0%-10%	10%-20%	20%-30%	30%-40%	40%-60%	60%-100%
CM360, conv weighted	1.4	1.7	1.8	2.8	7.4	85.0
DV360, rev weighted	4.2	7.1	4.2	7.7	13.5	63.2

ARA Efficacy for CTC Reporting

This analysis of ARA's efficacy for CTC reporting contains:
- Test: Conversion data based on first-party cookie (1PC) from Chrome versions M118+ and conversion data based ARA attribution from integrated W2W traffic for Chrome M118+[2]
- Control: Conversion data based on first-party cookie (1PC) from Chrome versions M118+ and conversion data based on third-party cookie attribution from the same traffic as test arm

Distribution of ARA's efficacy for conversion reporting is weighted by conversions for CM360 and revenue for DV360 and SA360 across all breakdowns.

ARA's Efficacy for CTC Reporting distribution by product (excluding Mitigation not ready traffic)
Numbers reported are percents

	APE buckets					
	0%-10%	10%-20%	20%-30%	30%-40%	40%-60%	60%-100%
CM360, conv weighted	30.0	13.8	4.8	4.9	25.0	21.5
DV360, rev weighted	18.6	15.3	5.7	10.4	13.0	37.1
SA360, rev weighted	59.3	12.7	7.1	4.6	4.9	11.4

Supplementary analysis: Analyzing all traffic

The analysis for CTC reporting contains:
- Test: Conversion data based on first-party cookies (1PC) from all Chrome versions and the ARA attribution from Chrome versions M115+
- Control: Conversion data based on first-party cookie and third-party cookie attribution from all Chrome versions.

---

[2] As the ARA launched in general availability in M118, excluding all prior versions with partial API support for the purpose of this analysis.

First we report ARA's efficacy for CTC reporting by product. Lastly, we analyze reporting by product and third-party cookie reliance buckets.

By Product

ARA's Efficacy for CTC Reporting distribution by product
Numbers reported are percents

	APE buckets					
	0%-10%	**10%-20%**	**20%-30%**	**30%-40%**	**40%-60%**	**60%-100%**
CM360, conv weighted	21.3	10.5	12.0	6.5	14.3	35.5
DV360, rev weighted	18.9	12.6	5.6	9.8	12.3	40.8
SA360, rev weighted	57.9	12.7	7.8	5.5	5.6	10.4

The differences are caused by low adoption of specific features in CM360 and DV360 vs SA360. When analyzing CTC reporting broken down by product and third-party reliance buckets, results are more comparable across products:

By Product and Third-Party Conversion Reliance

ARA's Efficacy for CTC Reporting distribution by product and third-party cookie conversion reliance
Numbers reported are percents

Product	Third-party cookie conversion reliance	APE buckets					
		0%-10%	**10%-20%**	**20%-30%**	**30%-40%**	**40%-60%**	**60%-100%**
CM360, conv weighted	**0%-20%**	67.1	32.8	0.0	0.0	0.0	0.0
	20%-50%	0.2	1.3	46.6	25.2	26.7	0.0
	50%-100%	0.0	0.0	0.0	0.0	17.3	82.6
DV360, rev weighted	**0%-20%**	74.1	23.4	0.5	1.0	0.4	0.6
	20%-50%	2.5	13.8	31.4	30.1	22.0	0.2
	50%-100%	0.8	7.7	0.3	7.6	14.5	69.0
SA360,	**0%-20%**	83.5	16.0	0.1	0.2	0.2	0.0

| rev weighted | 20%-50% | 0.8 | 7.9 | 45.1 | 29.1 | 17.1 | 0.0 |
| | 50%-100% | 0.4 | 2.7 | 0.1 | 2.8 | 18.6 | 75.3 |

Conclusion

There are a few factors that contributed to the results and could serve as potential sources of significant future improvements. First at the time of the experiment, the Floodlight pipeline that supports Google Non-O&O/ third-party web Display attribution measurement had finished basic end-to-end integration, but was still early in quality optimization and integration outreach to third-party Publishers and AdTechs. The VTC and CTC results were highly dependent on the ecosystem enabling ARA registration (by adding attributionsrc param). GMP is actively collaborating with third-party publishers and adtech providers to encourage the adoption of the ARA and facilitate Floodlight measurement on their inventory.

Secondly, API configuration limits were used to manage the ARA contribution budget and resulted in conversion losses for customers that measure a large number of conversions. GMP's project to optimize the contribution budget in the Aggregate API per customer is underway. This will adjust the limit for a customer within the constraints of the API, balancing the number of conversions recorded against added noise. Additionally GMP is working with customers to optimize their conversion setup in the ARA to help improve conversion loss.

Most importantly, these results do not account for Chrome's proposal for an updated approach that elevates user choice rather than deprecating third-party cookies. We expect that the conversion measurement quality will improve as we continue these efforts and as changes are incorporated.