



A DSP view on attribution API

Alois Bissuel, Thomas Prieur, Maxime Vono

Criteo



Objective

- Provide our view on ARA
- Discuss what to improve with Attribution API to ensure good adoption

Measurement, Machine Learning & Budgeting

• Attribution to a specific impression

- Required for machine learning (learn “good” impression features)



Attribution to a channel

- Not linked to a specific impression
- Requires consent on advertiser website
- Used to prove value of advertising
- Advertiser are ROI centred
- Advertiser use Measurement tools aggregating data from all channels (Google Analytics) to assign budget based on performance

Source de la session	Auteurs	Sessions	Sessions avec engagement	Durée d'engagement moyenne par session	Sessions avec engagement par utilisateur	Événements par session	Taux d'engagement	Nombre d'événements	Conversions	Ré
	Tous les événements	Tous les événements	Tous les événements	Moyenne	Moyenne	Tous les événements	Tous les événements	Tous les événements	Tous les événements	Tous les événements
1 (direct)	29 127 100 % du total	42 657 100 % du total	29 732 100 % du total	1 min 13 s Égal à la moyenne	1,02 Égal à la moyenne	16,44 Égal à la moyenne	69,7 % Égal à la moyenne	70 1201 100 % du total	54 518,00 100 % du total	80 999,6 100 % du total
2 sites.google.com	16 812	24 758	16 980	1 min 17 s	1,01	15,83	68,58 %	391 964	28 872,00	49 195,2
3 bing	8 307	11 396	8 454	48 s	1,02	11,82	74,18 %	134 698	12 965,00	10 319,4
4 newsletter_february_2024	1 795	2 875	2 506	1 min 21 s	1,40	16,78	87,17 %	48 241	5 873,00	8 256,4
5 sites.google.com	2 373	2 134	2	1 min 58 s	<0,01	40,94	0,09 %	87 366	3 269,00	5 866,6
6 newsletter_november_2023	420	578	494	44 s	1,18	12,32	85,47 %	7 120	1 077,00	916,8
7 newsletter_january_2023	199	342	254	2 min 15 s	1,28	24,37	74,27 %	8 334	761,00	1 256,1
8 newsletter_analyticsappspot_com	173	368	275	1 min 36 s	1,59	17,88	74,73 %	6 580	541,00	2 153,6
9 newsletter_analyticsappspot_com	123	203	156	1 min 54 s	1,27	19,84	76,85 %	4 028	317,00	825,6
10 art-analytics.appspot.com	143	220	171	1 min 40 s	1,20	20,43	77,73 %	4 495	315,00	296,8
	180	223	156	1 min 02 s	0,87	11,56	69,96 %	2 578	161,00	0,0

Summary of ARA testing findings



- Basic Reporting, PV / PC (campaign level)
- API dedicated for ML (event level)
- Cross-env attribution (Android App <> Chrome)
- Requeryring / reprocessing



- Advanced Reporting (fine level)
- Omnichannel, cross device, etc.



- Performance of ML optimization ?



For more details, check our
blog posts :



Attribution API : pros & cons

What works

- Simple primitives
- Flexible primitives
 - No constraints on event types
- Clear API structure
- Flexibility on the aggregation method (TEE or MPC)

Attribution API: pros & cons

What doesn't work for adoption at scale

- Privacy budget depletion by other adtechs
 - Clear coordination issue
- No ML
 - Though summation could be an acceptable primitive
- No standard for noise management
- No cross-advertiser aggregation (yet!)
 - Long tail of small clients where aggregation per advertiser is not possible
- Limited way of integrating data from the conversion in the histogram

We don't know

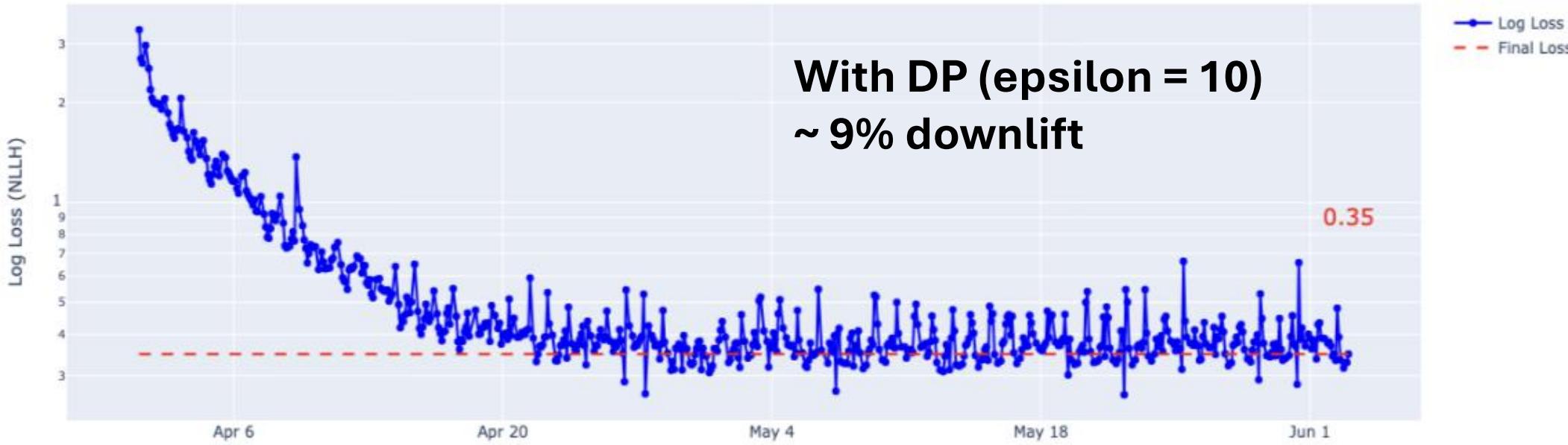
- Impact of the “new” user-level privacy granularity on the utility

ML on aggregated data : what we need

We need to be able, at least, to train GLMs (incl. logistic regression):

- Two main approaches we considered:
 1. "Online" learning: gradient is expressed through sums, so sum is a required primitive
 2. Learning from noisy label proportions (LP): requires to build noisy LP first and then standard training with relevant loss function
- We built a prod-ready solution for 1 under Privacy Sandbox Aggregation Service constraints. Not ideal but acceptable performance with epsilon greater than 10
- Need: DP budget constraints do not result in a higher noise, compared to what was possible with Privacy Sandbox Aggregation Service

ML on aggregated data : what we need



Online gradient
descent training
loss