# Attribution Modeling Increases Efficiency of Bidding in Display Advertising

Eustache Diemert[*]
Criteo Research
e.diemert@criteo.com

Julien Meynet[*]
Criteo Research
j.meynet@criteo.com

Pierre Galland
Criteo
p.galland@criteo.com

Damien Lefortier[†]
Facebook
dlefortier@fb.com

## ABSTRACT

Predicting click and conversion probabilities when bidding on ad exchanges is at the core of the programmatic advertising industry. Two separated lines of previous works respectively address i) the prediction of user conversion probability and ii) the attribution of these conversions to advertising events (such as clicks) after the fact. We argue that attribution modeling improves the efficiency of the bidding policy in the context of performance advertising.

Firstly we explain the inefficiency of the standard bidding policy with respect to attribution. Secondly we learn and utilize an attribution model in the bidder itself and show how it modifies the average bid after a click. Finally we produce evidence of the effectiveness of the proposed method on both offline and online experiments with data spanning several weeks of real traffic from Criteo, a leader in performance advertising.

## CCS CONCEPTS

• **Information systems** → **Display advertising**; • **Theory of computation** → *Theory and algorithms for application domains*;

## KEYWORDS

Bidding; Attribution; Offline Metrics; Performance Advertising

## 1 INTRODUCTION

Performance advertising has become a very successful model of programmatic advertising where advertisers payment is based on delivered value as measured by events of interest. Two main models exist: cost-per-click (CPC) or cost-per-action (CPA). We focus on the latter where the advertiser attributes conversion events (e.g. a signup or more generally a sale) to one (or more) advertising events such as displays or clicks. The industry standard for conversion

---

attribution is to credit only the last click in the 30 days before the conversion although this is changing recently [8].

For advertising agencies and platforms the state of the art strategy is the Expected Value Bidder (EVB) that bids the expected value of the opportunity [11] [12]. EVB has been proven to be a dominant strategy in the case of non-repeated second-price auctions [7]. The expected value is computed as advertiser payment times the predicted conversion probability. Numerous previous studies focused on accurately estimating click or conversion probabilities at scale [3] [12] and open datasets were published for this task [5] [18].

On the other side, conversion attribution has been thoroughly studied in the fields of game theory and econometrics [13] [1] [6] and mechanisms have been proposed to better match the advertiser payment with the ad effectiveness, especially when multiple advertising events or channels are involved.

In order to improve the EVB policy one can research the usage of attribution prediction. A motivational example is to consider consecutive auctions for the same user. If the user has clicked after the first impression the advertising platform has good chances of getting the attribution, in the sense that a consecutive conversion would be attributed to that click - unless another advertising event captures the attribution in between. In a subsequent auction, a rational bidding policy should take into account this non-zero attribution probability and lower its bid compared to the first one as there is little chance that a second click would increase the attribution probability. The EVB policy is not able to act rationally in that respect since it doesn't keep track nor predicts the attribution probability evolution. As a first exploratory step towards improved real-time bidding policies for advertising we propose to leverage an attribution model to modify the bidding strategy.

We posit that an attribution bidding policy has the benefit of better aligning the incentives of the advertiser with the ones of the advertising platform and the user, resulting in increased efficiency for all parties involved. Intuitively this comes from a better representation of the effect of advertising over time: the advertiser should pay for ads that affect attributed conversions, the platform should reduce its costs when bidding for ads with low attribution probability and the user should see less ads once engaged with the advertiser.

The rest of the paper is organized as follows. Firstly we describe our core contribution: a suitable attribution model in Section 2.1 and the related modifications to the bidding strategy in Section 2.2. We then explain in Section 3 how to measure the improvement offline with an updated version of the Expected Utility metric [4].

We report and analyze positive results from both offline and on-line experiments in Section 4. Finally we discuss future works and connections to other related approaches in Section 5.

## 2 MODEL

### 2.1 Attribution Model

The industry standard for conversion attribution is to credit the advertizing platform which owns the last click in a 30 days window before the conversion. Predicting the 30 days conversion probability is thus not enough as different events may change the attribution probability until the conversion actually happens: the competition (or other channels such as search-based advertising) may capture the attribution by generating clicks on their side. This characteristic of the last-click attribution has been shown to drive more efforts on average from advertising platforms and thus be more effective than other mechanisms in the case of cost-per-mil (CPM) payment models [2]. However from the same study we observe that it is not obvious for performance advertising (CPC or CPA). It was also shown in [17] that last-click attribution might prevent advertising platform to bid proportionally to the real marginal impact of the ad on the user conversion.

In the following model we do not explicitly assume that the advertiser uses last-click attribution to reward advertising platforms, but instead we propose to learn attribution from the data and then incorporate it in the bidder itself. We consider the point of view of an advertising platform having access to advertiser conversions and their attribution label. We don't assume access to competition data such as clicks made by other platforms.

Let us introduce the notation: $C$ is a random variable representing the click, $S$ any conversion (attributed or not), $A$ an attributed conversion, all of which are Bernoulli. $X$ is the generic random variable representing all available contextual information (such as the user context, website information, advertising campaign). $\Delta$ is the delay between a conversion and a click observed prior to it. The task of attribution modeling consists in learning $P(A = 1|S = 1, X = x)$ the probability of attributed conversion given there is a conversion, in a certain context $x$.

To be used for driving the bidding strategy the attribution model needs to capture important aspects of the attribution process such as the implementation used by the advertiser and the characteristics of the competition. We propose to use mainly two sources of information:

- $\delta$: the time since the last click generated by the advertising platform that captures the uncertainty of still having the attribution in the face of competitive events
- $x$: capturing global characteristics such as per advertiser implementation, user activity level , etc.

The proposed model is an exponential decay model of $\delta$: the time between the conversion and the last click, parametrized by a decay factor $\lambda$:

$$P(A = 1|S = 1, X = x, \Delta = \delta) = e^{-\lambda(x)\delta} , \ \lambda(x) \geq 0 \quad (1)$$

The $\lambda$ decay factor represents here the speed at which the attribution probability decreases with time after the click (taking into account events such as competitors winning the attribution thanks to clicks on their ads). This model is not specific to last-click attribution. It
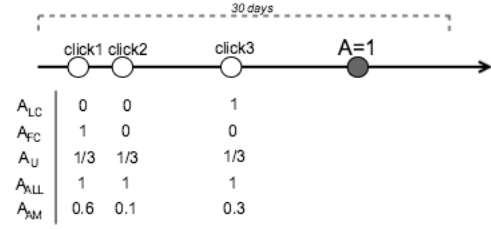


**Figure 1: Internal attribution procedure from the advertising platform point of view in a case where 3 clicks precede an attributed conversion.** 5 possibles schemes are illustrated: last-click $A_{LC}$, first click $A_{FC}$, uniform $A_U$, all-clicks $A_{ALL}$ and an attribution equivalent to our attribution model $A_{AM}$.

may work under other attribution schemes as soon as the attribution probability vanishes over time after the clicks. The choice of this exponential decay is quite common in survival analysis [9] and was also used to model delays in conversion feedback [3]. Models that better fit the data such as Kaplan-Meier or Weibull [9] could be considered, but fine tuning the attribution model is not in the scope of the paper.

*2.1.1 Model Estimation.* We estimate $\lambda$ by maximum likelihood. Let's consider a dataset built using clicks-conversion log as follows: $\{(\delta_i, a_i)\}_{i=1\ldots n}$ where $\delta_i$ is the time since last click of that particular conversion and $a_i$ the attribution: 1 if attributed, 0 if not to the advertising platform. Under the assumption $P(a_i = 1 \mid \delta_i) = e^{-\lambda \delta_i}$ we can write the conditional likelihood of the data and our convex cost function, the negative log-likelihood.

$$LH(\lambda) = \prod_{i=1}^{n} P(A_i = a_i \mid \delta_i) = \prod_{i=1}^{n} e^{-\lambda a_i \delta_i} (1 - e^{-\lambda \delta_i})^{1-a_i} \quad (2)$$

$$NLLH(\lambda) = \sum_{i=1}^{n} \lambda a_i \delta_i - (1 - a_i) log(1 - e^{-\lambda \delta_i}) \quad (3)$$

### 2.2 Attribution-aware Bidder

As mentioned before, the standard bidding strategy, in $2^{nd}$ price auctions, is to bid the expected value of an impression [11]. As last-click attribution is usually assumed for the advertiser, previously published works also use last-click attribution when learning the conversion models, meaning that when several clicks occur before a conversion in the logs, only the last one is labelled as positive while previous ones are labelled as negative samples. Figure 1 gives an illustrative example of the internal attribution procedure when 3 clicks occurred before an attributed conversion for a given advertising platform. Let us call $A_{LC}$ the random variable that measures if there was an attributed conversion in the 30 days after the click and this click was the last for the advertising platform. The bid value for an impression is estimated as:

$$bid \propto CPA \times P(A_{LC} = 1|X = x) \quad (4)$$

We can consider simple variants of this strategy by applying rule-based heuristics on the attribution process, by considering, for example, first-click ($A_{FC}$), uniform ($A_U$) or linear attribution (some of them are depicted in in Figure 1). In fact we could argue that first clicks might have contributed more to user conversions

than last ones. However, there is no evidence that fixed rules can work without making very strong hypotheses on the advertiser attribution scheme. Another alternative would be to consider all clicks leading to a conversion as positive examples ($A_{ALL}$) but, as other fixed rules, such a scheme cannot catch temporal dependencies between the auctions, which represents a key aspect in proper attribution modeling. In the example of Figure 1, how should the conversion be attributed to each click? The first click might have more impact on the conversion but as the third one is closer to the conversion so it might also be important. In all cases click2 seems to have less importance than the other two.

In line with lift-based bidding [17] we propose to bid proportionally to the marginal attribution brought by the current opportunity: if the user recently clicked on one of our ads, we already have some probability to get the attribution and showing more ads to the user might not be relevant for her given the recent interaction she had with the advertiser. To estimate the current attribution probability, we leverage the attribution model presented in Section 2.1.

At auction time, the gap between the current opportunity and the conversion $\delta_s$ if the user eventually converts is unknown. However, we can measure the time elapsed since the last observed click $\delta_c$, which gives a good proxy for estimating the uncertainty under which we can hope to get an attribution from past auctions. The attribution probability from the previous-click only is, according to Equation (1): $e^{-\lambda(x)(\delta_s+\delta_c)}$. Now, if current auction leads to another click, our attribution probability becomes $e^{-\lambda(x)\delta_s}$. The marginal contribution of the second click, normalized so that we can fully explain the conversion is then:

$$\frac{e^{-\lambda(x)\delta_s} - e^{-\lambda(x)(\delta_s+\delta_c)}}{e^{-\lambda(x)\delta_s}} = 1 - e^{-\lambda(x)\delta_c} \qquad (5)$$

The attribution-aware bidder directly modulates the attributed conversion probability by the marginal contribution of this opportunity:

$$b(x) \propto CPA \times P(A_{ALL} = 1|X = x)(1 - e^{-\lambda(x)\delta_c}) \qquad (6)$$

where $\delta_c$ is the observed time since last click, and $A_{ALL}$ refers to whether the display led to an attributed conversion or not. The first part of the model $P(A_{ALL} = 1|X = x)$ represents the post-click conversion probability without any assumption on the internal attribution process, while second term handles the attribution part.

*2.2.1 Discussion.* If there is no previous click, second term vanishes and the bidder acts as the baseline EVB bidder, with one notable difference: all clicks leading to a conversion are labelled as positive, while only the last-one would be positive in the baseline. The fact last-clicks of a given advertising platform are not favored in the learning is good from the advertiser perspective: the platform is not incentivised to "steal" the attribution when the true impact of the advertising on the conversion is low. On the other hand, right after a click, the attribution bidder will drastically decrease the bid values. The recent click already gives a good probability of getting the attribution would there be a conversion. We also think that displaying new ads after a click does not increase the user propensity to convert, so investing the money elsewhere is a better strategy. Ideally, as explained in [17], if the advertiser attribution process

credits each event according to its causal impact on the conversion a bidder should only bid the lift in the conversion probability.

As a click is a strong engagement signal from the user we expect the law of diminishing return to apply to the impact of subsequent displays. However, baseline EVB models tend to bid higher after a click which increase both advertising costs and user exposure. The proposed bidder is thus closer to a lift-based bidder.

## 3 METRICS

### 3.1 Offline Metrics in Online Advertising

Offline evaluation of bidding policies is especially important in advertising as online experiments imply to invest real money in ad auctions. Therefore offline evaluation consists in replaying logged auctions that were won by a previous production policy [18]. Lost auctions are not used as the bid needed to win the auction is unknown (which impairs counter-factual evaluation due to selection bias). Still, as the state-of-the-art EVB strategy is to bid one's expected value i.e. payoff × conversion_probability, classical offline metrics focus on the prediction problem as the payoff is usually fixed for a given campaign. Weighted Mean Squared Error [15] combine event payoff and prediction quality in a single metric. However such metric may not accurately predict online results if the competition changes. Recently [4] proposed Utility to tackle this problem by injecting noise in the competing bids distribution.

We now briefly describe the Utility metric before exploring its limits with respect to attribution in the following section.

Consider $v_i$, the value for the advertising platform of an action on an ad (e.g. a conversion value), $p_i$ the estimated probability of a user performing that action. In this setup the value of the impression can be estimated as $v_i \times p_i$. Let $c_i$ be the observed price, which is the highest competing bid in a second price auction, and $a_i$ a binary variable that represent if the action is attributed to the display or not. Then the payoff of the auction can be written as:

$$u_i = (a_i \times v_i - c_i)\mathbb{1}_{[p_i \times v_i > c_i]} \qquad (7)$$

The Utility of a new prediction model $p'$ is estimated on dataset $D$ as:

$$U(p') = \sum_{i \in D}(a_i \times v_i - c_i)\mathbb{1}_{[p'_i v_i > c_i]} \qquad (8)$$

Expected utility was proposed to model potential changes in the competing bids using a cost distribution $\Pr(c|c_i)$ instead of the actual observed cost:

$$EU(p') = \sum_{i \in D}\int_0^{p'_i \times v_i}(a_i \times v_i - c)\Pr(c|c_i)dc \qquad (9)$$

A good choice for $\Pr(c|c_i)$ is a Gamma distribution, parametrized by $\alpha, \beta$: It interpolates between two extreme cases: empirical utility and WMSE [4]:

$$\Pr(c|c_i) \sim \Gamma(\alpha = \beta c_i + 1, \beta) \qquad (10)$$

As in [4] we use: $\alpha = \beta c + 1$.

### 3.2 Attribution-aware Expected Utility

The expected utility basically supposes the conversions are attributed implicitly on a single display, usually the last-clicked display before the conversions. Evaluating other bidders on this metric will

naturally favor last-click based bidder. Therefore, we propose to extend the metric by taking into account the attribution evolution. For this purpose we weigh the value of each display to make it depend on the attribution function $a : x \in \mathbb{R}^N \mapsto \mathbb{R}$:

$$AEU(p', a) = \sum_{i \in D} \int_0^{p'_i \times v_i} (a(x_i) \times v_i - c) \Pr(c|c_i) dc \qquad (11)$$

If the attribution rule is last-click, then AEU is exactly equivalent to Expected Utility:

$$a_{lc}(x) = \begin{cases} 1 & \text{if } x \text{ is last-click} \\ 0 & \text{otherwise} \end{cases} \qquad (12)$$

A natural choice for the attribution function is to re-use the attribution model proposed in Section 2.1:

$$a_{am}(x) = \begin{cases} 1 - e^{-\lambda(x)\delta_c} & \text{if } x \text{ was clicked} \\ 0 & \text{otherwise} \end{cases} \qquad (13)$$

As this attribution scheme is not normalized, each conversion may not contribute equally to the metric (attributions of all displays preceding a conversion might not sum to one). This tends to penalize conversions for which the last click occurred long time ago: we believe it is a desired property for a metric to limit the importance of conversions that occurred days or even weeks after clicks, as the real causal impact of these clicks is expected to be small. Anyway we also report in the experiments a normalized version (where each conversion contributes the same amount, as shown by $A_{AM}$ in Figure 1) in addition to the raw one.

A risk of the *AEU* is to choose a bidder that overfits the learned attribution model. In that case we advise to consider a regularized version that, for example, would not be strictly dependent on the lambda parameter of the model. It is however not in the scope of this paper and will be left for future research.

In the remaining of the paper, for sake of simplicity, we call $U_A$ the *AEU* using normalized proposed attribution model, $U_{A*}$ its un-normalized counterpart and $U_{LC}$ the variant with last-click attribution function.

## 4  EXPERIMENTS

### 4.1  Criteo Attribution Dataset

Attribution information (the difference between all sales observed on an advertiser site and attributed ones) is mandatory for our study and not present in previously published conversion datasets [5] [18]. Therefore for all offline experiments we use a log sampled from 30 days of Criteo production traffic with displays, clicks, conversions and attribution data. It will be released publicly at http://research.criteo.com. The data has been sub-sampled and anonymized so as not to disclose proprietary elements but the exact method is irrelevant here. Each line represents an ad impression with following information: timestamp, price paid, categorical features of the user, ad and publisher, click*, position of the click*, conversion*, conversion value*, attribution*.

Elements marked with a * are present if applicable. Conversion and their value are credited by the advertiser and made comparable across the dataset. If multiple clicks occurred and there was a conversion the click position is provided, which is necessary for implementing different attribution schemes.
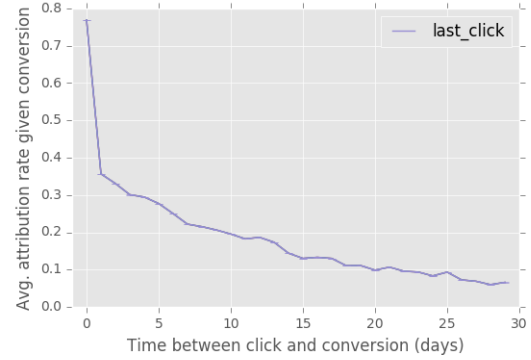


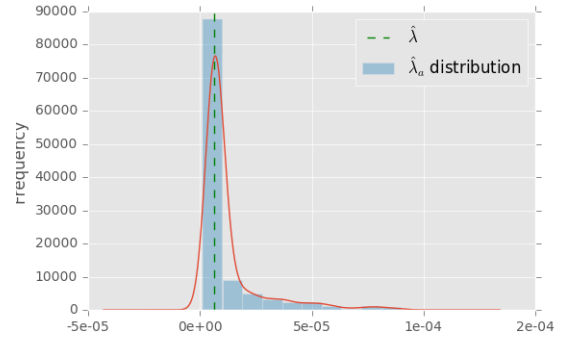**Figure 2: Average attribution given conversion and time since last click**



**Figure 3: Distribution of per advertiser $\hat{\lambda}_a$ versus global $\hat{\lambda}$**

### 4.2  Attribution Study

The model presented in Section 2.1 encodes the attribution probability as a function of the time between a click and the conversion. The empirical attribution rate as function of this delay is illustrated on Figure 2 and shows a sharp decline in attribution rate after a click followed by a slower decrease tending to 0 after 30 days.

We estimated a global $\hat{\lambda} \approx 6.25 \times 10^{-6}$ on the log. It is remarkably stable over time with a daily relative variation $< 0.1\%$. As seen on Figure 3 the $\lambda$ parameter is also stable across advertisers. We posit that fitting one parameter per advertiser would only improve the training objective marginally and thus prefer to use a global one.

### 4.3  Offline Evaluation

*4.3.1  Baselines.* The tested bidding strategies include two expected value bidders. LCB is the traditional baseline defined in Eq. 4. First-click bidder (FCB) is its first-click counterpart. Both actually use the same features but different labels for learning the conversion model. To understand the difference we observe the attribution rate with respect to an important variable: time since last click, as illustrated on Figure 4. We observe a high proportion of positive examples for LCB within a short delay after a click and that proportion decreases with the delay. Conversely, for FCB the initial proportion is low and increases slowly over time.
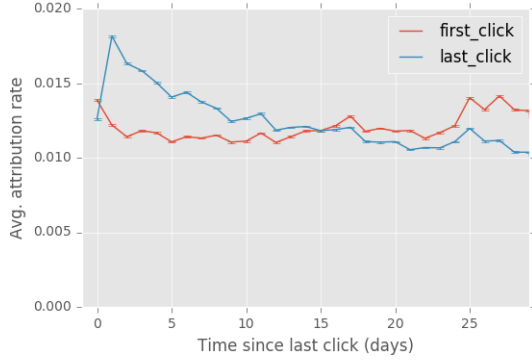
Figure 4: Average attribution rate per display, given time since last click for two attribution schemes

|  | $U_A$ | $U_{LC}$ |
|---|---|---|
| AB vs LCB | +12.32% | -14.15% |

Table 1: Offline comparison of Attribution (AB) versus Last-Click (LCB) bidders. $U_A$ and $U_{LC}$ refer to variants of *AEU* with resp. last-click or proposed attribution model and cost perturbation $\beta = 1000$. Result is significant at the .05 quantile of 100 bootstraps resamples.

Our proposed attribution bidder (AB) defined per Equation 6 is composed of two multiplicative parts. As described earlier the bid values will drastically decrease right after clicks as can be deduced from Figure 2.

*4.3.2    Evaluation Setup.* Our evaluation protocol is as follows: the last 7 days of log are considered for testing and for each test day we use the 21 previous days for learning. For learning conversion models, all the variables are mapped into a sparse binary feature vector via the hashing trick and an $L^2$-penalized logistic regression model is trained using the L-BFGS algorithm (similar to [14]). The evaluation metrics are variants of the Attribution-aware Expected Utility: $U_A$, $U_A*$, and $U_{LC}$ as presented in Section 4.2. In order to have comparable values between bidders for a given metric, predictions are calibrated so as to predict the same value on average. In a real system this calibration property is usually controlled by a feedback loop mechanism to ensure constant spend. More details are given in Section 4.4. For each variant of Utility, we use different $\beta \in (1000/inf)$ which correspond to medium/no variations of the competing bids.

*4.3.3    Results & Discussion.* Our first result is summarized in Table 1: as expected the proposed AB is favored by the Utility when the attribution is given by the same attribution model. Conversely, LCB is favored by the last-click Utility. We notice in this result how LCB becomes sub-optimal when the metric focuses on earlier clicks. At this point we posit that the Utility that uses an attribution model $U_A$ best reflects the real attribution system.

Our second result is summarized in Table 2: when measured across different variants of the $U_A$, the proposed AB policy performs the best among the three. The reported win rate tells us that

|  | LCB | FCB | AB |
|---|---|---|---|
| Win Rate | 0.94 | 0.90 | 0.89 |
| $U_A, \beta_{1000}$ | 1679 ± 31 | 1843 ± 37 | 1886 ± 39 |
| $U_A, \beta_{inf}$ | 3383 ± 46 | 3471 ± 52 | 3549 ± 58 |
| $U_A*, \beta_{1000}$ | 2852 ± 43 | 2888 ± 43 | **3396 ± 53** |
| $U_A*, \beta_{inf}$ | 5105 ± 56 | 5083 ± 57 | **5408 ± 61** |

Table 2: Offline comparison of 3 bidders (Last-click LCB, first-click (FCB) and attribution bidder (AB)). $U_A/U_A*$ refers to Eq. 11 with cost perturbation $\beta$ and normed/un-normed attribution model.

LCB wins 5% more auctions than AB, auctions that are probably less valued by the metric, representing excessive post-click user exposure. Event though *FCB* performs quite on par with *AB* on the normed Utility $U_A$ this is not the case with the un-normed $U_A*$ version and the gap grows larger with smaller $\beta$ ($\beta_{inf}$: +6% / $\beta_{1000}$:+17%), i.e when there is more uncertainty on the competing bids. This let us think AB is more robust in real systems. This confirms the comments in Section 3.2. Thoroughly evaluating $U_A$ deserves to study the correlation of offline versus online results. We deliberately skip such analyses in the current work as it is not the main focus of this paper.

Our third result presented in Figure 5 illustrates the average post-click bid evolution for the different bidders in the first 24 hours. As expected, LCB overbids strongly just after a click and then slowly converges to the nominal pre-click bid. Conversely, FCB underbids after a click (although less sharply than LCB) and overbids increasingly compared to the pre-click level. We observe that our proposed AB policy bids at an extremely low level just after a click and then steadily increases its bids.
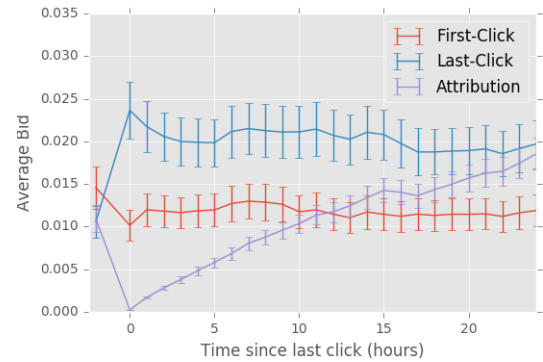


Figure 5: Average bid given time since last click for different bidding strategies

## 4.4    Online Experiments

*4.4.1    Experimental Setting.* We fit an attribution model with a global $\lambda$ using the procedure described in Section 2.1.1. We learn an attribution model encompassing the business standard which means looking back up to 30 days before the conversion for potential clicks.

Our online experimental policy is a simple multiplier applied to all bids:

$$bid_{test} \propto bid_{ref} \, A \, (1 - B \, e^{-\lambda \delta}) \tag{14}$$

This implementation is straightforward and flexible for different practical use cases. Firstly, it allows for more or less aggressive bid decreases thanks to the $B$ factor: small $B$ will only slightly change the reference bidding policy. Secondly, the proposed bidding policy consists only in reducing bids and hence should result in underspending with respect to the reference population. To compare policies one should be able to measure their performance in a setting where both spend the entire available advertiser budget. The $A$ factor can be used to equalize spent budget between the control and test populations for this purpose.

Different values for the $A$ and $B$ parameters have been A/B tested by user split on the Criteo production platform.

| $\Delta OEC$ (long term) | Revenue (short term) | Advertiser ROI | User ad exposure |
|---|---|---|---|
| +5.5% | negative | positive | lower |

**Table 3: Online results summary.** OEC is a long-term performance metric. We report the uplift of the proposed bidder wrt. the baseline production bidder. All reported results are significant at the .05 quantile of 100 bootstraps resamples.

*4.4.2 Results & Discussion.* Overall results are summarized in Table 3. We found that the proposed bidding policy showed positive long-term results in all settings. The Overall Evaluation Criterion (OEC) [10] reported for this experiment captures the value generated by Criteo for the advertisers: $OEC \propto (AdvertiserValue - Cost)$ and is not necessarily correlated to short-term revenue. The consistent improvement of the OEC by +5.5% across all production traffic is remarkable since it is quite rare given the difficulty to move such long term criteria.

The OEC was moved mainly through reducing bids and hence budget spend in subsequent auctions. The accompanying negative impact in revenue shall be explained as follows. A better efficiency of the proposed bidding policy delivers more value for less cost, hence reducing the spend of advertisers (and Criteo revenue) to achieve a target CPA. Of course rational advertisers should then increase their target CPA to benefit from additional value.

From a user perspective such policies reduced drastically the exposition to ads from a given advertiser after a click on an ad of that advertiser. Incidentally, and beyond overall reduced exposure, it brought more diversity on ads seen as other advertisers on the Criteo platform had now better chances of winning bids.

For advertisers we noticed an increased return on investment (ROI) as measured by the number of conversions (or conversions values) generated for a given budget.

## 5 CONCLUSION

We proposed a novel, effective yet simple bidding policy leveraging attribution modeling. Our model can be trained efficiently and requires minimal changes to be implemented in an advertising platform. We proposed a modification of the Utility offline metric.

We also reported highly positive experiments both offline and online at scale on Criteo production platform.

We plan to release the dataset used for offline experiments under the same conditions as previous Criteo datasets so as to foster additional research in attribution modeling. An immediate perspective would be to study alternative attribution models and their impact. Future works could also include researching this problem in i) the reinforcement learning setting where the state would incorporate the attribution evolution and ii) the online learning framework where attributed value is learnt across repeated auctions in the spirit of [16].

## REFERENCES

[1] Vibhanshu Abhishek, Peter Fader, and Kartik Hosanagar. 2012. The Long Road to Online Conversion: A Model of Multi-Channel Attribution. *SSRN Electronic Journal* (aug 2012).
[2] Ron Berman. 2015. Beyond the Last Touch : Attribution in Online Advertising. *Preliminary Version* (2015).
[3] Olivier Chapelle. 2014. Modeling delayed feedback in display advertising. In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining - KDD '14*. ACM Press, New York, New York, USA, 1097–1105.
[4] Olivier Chapelle. 2015. Offline Evaluation of Response Prediction in Online Advertising Auctions Categories and Subject Descriptors. In *Proceedings of the 24th International Conference on World Wide Web*. 919–922.
[5] Criteo. 2014. Criteo Conversion Logs Dataset. (2014).
[6] Brian Dalessandro, Claudia Perlich, Ori Stitelman, and Foster Provost. 2012. Causally motivated attribution for online advertising. *Proceedings of the Sixth ...* (2012), 7–9.
[7] David Easley and Jon Kleinberg. 2010. *Networks, crowds, and markets: Reasoning about a highly connected world.* Cambridge University Press.
[8] Wendi Ji, Xiaoling Wang, and Dell Zhang. 2016. A Probabilistic Multi-Touch Attribution Model for Online Advertising. In *CIKM*. Association for Computing Machinery, 1373–1382.
[9] J. D. Kalbfleisch and Ross L. Prentice. 2002. *The statistical analysis of failure time data.* J. Wiley. 439 pages.
[10] Ron Kohavi, Randal M. Rm Henne, and Dan Sommerfield. 2007. Practical Guide to Controlled Experiments on the Web: Listen to Your Customers not to the HiPPO. *KDD 2007* (2007), 1–9.
[11] Claudia Perlich, Brian Dalessandro, Rod Hook, Ori Stitelman, Troy Raeder, and Foster Provost. 2012. Bid optimizing and inventory scoring in targeted online advertising. In *Proceedings of the ACM SIGKDD 2012*. ACM Press, New York, New York, USA, 804–812.
[12] Rómer Rosales, Haibin Cheng, and Eren Manavoglu. 2012. Post-click conversion modeling and analysis for non-guaranteed delivery display advertising. *Proceedings of the fifth ACM international conference on Web search and data mining - WSDM '12* (2012), 293.
[13] Xuhui Shao and Lexin Li. 2011. Data-driven multi-touch attribution models. In *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining - KDD '11*. ACM Press, New York, New York, USA, 258.
[14] Munindar P. Singh. 2015. Norms as a basis for governing sociotechnical systems. In *IJCAI International Joint Conference on Artificial Intelligence*, Vol. 2015-Janua. ACM, 4207–4211.
[15] Flavian Vasile, Damien Lefortier, and Olivier Chapelle. 2017. Cost-sensitive Learning for Utility Optimization in Online Advertising Auctions. In *KDD 2017 Workshop: AdKDD and TargetAd*.
[16] Jonathan Weed, Vianney Perchet, and Philippe Rigollet. 2015. Online learning in repeated auctions. *JMLR* 49 (2015), 1–31.
[17] Jian Xu, Xuhui Shao, Jianjie Ma, Kuang-chih Lee, Hang Qi, and Quan Lu. 2016. Lift-Based Bidding in Ad Selection. *Proceedings of the 30th Conference on Artificial Intelligence (AAAI 2016)* (2016), 651–657.
[18] Weinan Zhang, Shuai Yuan, Jun Wang, and Xuehua Shen. 2004. Real-Time Bidding Benchmarking with iPinYou Dataset. (2004).