ATTRIBUTION PRIMER

Published June 21, 2012

© 2012 Interactive Advertising Bureau

This attribution primer was developed by the IAB Attribution Working Group.
A working group of volunteers from eight IAB member companies authored the IAB Attribution Primer.

The following IAB member companies contributed to this document:

Adometry
AppNexus
AT&T AdWorks
Cars.com
CPX Interactive
eBureau
Encore
Teradata
Tru-Signal

The IAB lead on this initiative was Brendan Riordan-Butterworth

Contact adtechnology@iab.net to comment on this document.

The IAB Attribution Working Group is part of the IAB Data Council, which is dedicated to demystifying data usage and control in the interactive advertising marketplace. The IAB Data Council objective is to enable revenue growth through the establishment of quality, transparency, accountability, and consumer protection in data usage. A full list of Council member companies can be found at: www.iab.net/data_council.

This document is available on the website at: www.iab.net/attributionprimer

Table of Contents

1 Attribution Definitions 4

1.1 Attribution 4
1.2 Attribution Ecosystems 4
1.2.1 User-specific attribution 4
1.2.2 User-group attribution 5

2 Methodologies 5

2.1.1 Single Source Attribution 5
2.1.2 Multiple Source Attribution 5
2.1.2.1 Event Types 6
2.1.2.2 Factors 6
2.1.2.3 Attribution Models 7
2.1.3 Multiple Media Type Attribution 7

3 Challenges 8

3.1.1 Incomplete Data Set 8
3.1.1.1 Limitation due to coverage 8
3.1.1.2 Limitation due to time 8
3.1.2 User Identifiers 8
3.1.2.1 Multiple Screens 9
3.1.2.2 Multiple Media Types 9

4 Similar Techniques 9

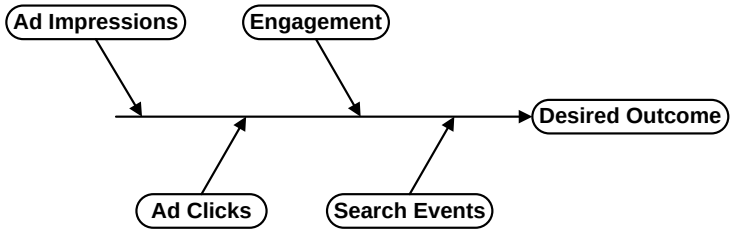
4.1.1 Comparison to Marketing Mix Modeling 9
4.1.1.1 Goals of Marketing Mix Modeling 9
4.1.1.2 Goals of online attribution 10

5 Terminology 11

1 Attribution Definitions

1.1 Attribution

Attribution is the process of identifying a set of user actions ("events") that contribute in some manner to a desired outcome, and then assigning value to each of these events.



In digital advertising, attribution is traditionally done at a user-specific level, where a consistent user identifier can be established across all analyzed events. In traditional media, attribution is generally done at the macro, user-group level, as there is no consistent user identifier available (see Media Mix Modeling).

The focus of this document is user-specific attribution, but it will also reference user-group attribution in order to highlight the challenges of integrating data across mediums.

1.2 Attribution Ecosystems

Attribution is a necessary component of advertising – without a basic understanding of who is influencing consumers to engage in desired behavior, it's very difficult to determine the ROI of one's advertising spend among multiple vendors. Attribution in digital advertising has been different from attribution in traditional media, primarily due to consistent user identifiers in the digital ecosystem, which allow for user-specific attribution.

1.2.1 User-specific attribution

Attribution of events to a specific user has always been possible in the digital advertising ecosystem, but with varying degrees of complexity. The historic implementation of single source attribution (the "last click" or "last touch" model) is a naïve attribution model that has been very widely used. Newer attribution models seek to create a mature valuation of all events in the digital advertising ecosystem.

While technology has allowed for advances in user-specific attribution (collecting and analyzing all events, not just the last event), it has also introduced new challenges. One key challenge is how to obtain a consistent user identifier across multiple devices as consumers adopt a multi-screen existence. Another key challenge is how to ensure each advertising vendor has access to consistent measure of performance of distributed outcomes so that the advertiser and each vendor can agree cost-per-performance pricing.

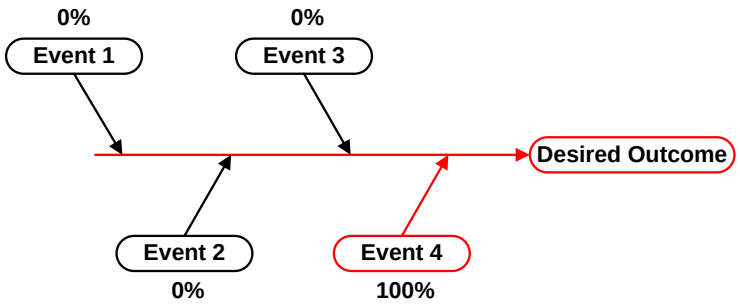
1.2.2 User-group attribution

As the digital advertising ecosystem matures, handling campaigns that span both digital and traditional media will require integrating traditional metrics and techniques with digital methods. Often this will mean providing data that is less specific – for example, aggregating click-through rates into demographic or geographic buckets, as opposed to providing user-specific ad impressions and clicks. The actual level of aggregation will vary by implementation.

2 Methodologies

2.1.1 Single Source Attribution

The most basic type of attribution is "last touch" attribution, which is a simple model of single source attribution that gives 100% of contribution to the last meaningful event, generally the last ad impression (sometimes called ad view), last click, or last engagement.



Single source attribution has fallen out of favor with buyers as it does not accurately reflect nor credit all of the contributors to the desired outcome. With this methodology, performance-based-pricing media sellers maximize their revenue by finding prospects furthest down the marketing funnel to "win" the last-touch race.

2.1.2 Multiple Source Attribution

Effectively collecting and analyzing more than one contributing event requires an understanding of the events that occur and of the factors that influence their values.

The set of event types, factors, and models described in this document is not exhaustive, and does not intend to be prescriptive.

2.1.2.1 Event Types

Sources that may be taken into account include:

Clicks

A click is the measurement of navigating from one page to another by activating a hyperlink. Clicks may be broken down into ad clicks, search clicks, affiliate clicks, and possibly other sub-categories.

Engagements

- **Internal Engagements**
  User activity that occurs on the advertiser's directly owned content, but is not the desired outcome. For example, browsing product feature pages without completing a purchase at that time.
- **External Engagements with 3rd party content**
  User activity that occurs on 3rd party owned content, yet promotes the desired outcome in some way. This could include viewing promotional content on an affiliate site, engaging the advertiser's brand on a social media site, or through other methods.
- **External Engagements with 1st-party content**
  User activity that occurs in an ad that leverages rich media capabilities highlighted in the IAB Rising Stars program. The user interacts with the ad, but this interaction doesn't result in navigation off the web site.

Ad Impressions

An ad impression is the measure of the delivery of an paid content to a web client.

Ad Views

An ad view has often been used interchangeably with ad impression, but is now used to refer to ads that are shown in an area of the web page that is determined to be viewable during that client page view.

Direct Navigations

Direct navigation is the act of typing an address into the URL bar directly, or using a bookmark, or otherwise accessing a web site without having clicked on a hyperlink.

Searches

An event where the user has provided keywords to a form in order to locate something of interest.

2.1.2.2 Factors

Additionally, each source may be modified by one or more factors:

Recency

How recently an event occurred, whether measured by time or by number of intervening events.

Frequency

How often a specific event occurs.

Sequence position

Whether an event was the first, last, or n^{th} in sequence. Determining the "first" event is not perfectly reliable in attribution efforts, as cookie churn and scope issues may mask the true first event.

Path
The serial list of events experienced by a user.

Engagement Depth or Duration
Specific events can indicate different levels of interest by a specific user. When a formal hierarchy is defined for understanding this, the engagement depth is measured.

2.1.2.3 Attribution Models

An attribution model is the set of rules by which the value of each event is determined. For example, in single source attribution, the simplest model is the "last touch" model. Here are some example valuation rule sets:

"Last Touch"

- The event has a value of 100 if it was the last event recorded, else it is worth 0.

"Simple"

- Ad clicks are worth 500 points, ad impressions are worth 100 points, ad views are worth 250 points.

"Complex"

- Ad clicks start with 500 points, ad impressions start with 100 points, ad views start with 250 points.
- Reduce the value of any event type by 10 points for every 24 hours before the desired outcome it occurred, down to a minimum of 0.
- Reduce the value of any event type by 25 points for each time it happened before, down to a minimum of 0.
- Increase the value of any event type by 50 points if it happened in the 30 minutes before the desired outcome, and was the last event recorded.

These rules may be determined exclusively by hand, or may adapt automatically over time.

"Algorithmic"

- The value of any specific event is determined by understanding its relationship with other events, including the desired outcome.
- The value calculation of any event can take into account the value of any other event, even if that other event didn't lead to the desired outcome.

2.1.3 Multiple Media Type Attribution

When a consistent user identifier is not available, user-specific attribution is not possible. However, by drawing on aggregation techniques, estimates can be provided that generate meaningful user modeling data.

Signals may be drawn from known multi-user identifiers like unified WiFi or IP addresses, or from geographic regions. By using IP to Geo technology, which can establish a ZIP code or other geographical coordinate from an IP address, the use of additional, aggregate signals is enabled. However, this is done at a loss of accuracy.

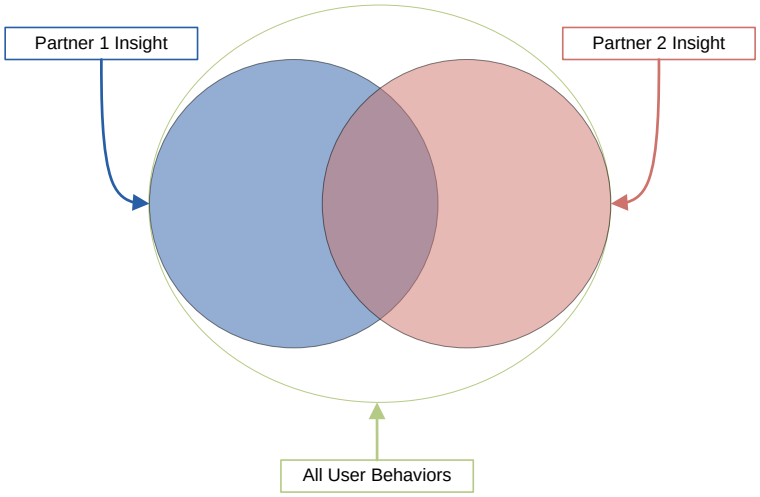
3 Challenges

3.1.1 Incomplete Data Set

It is expected that the whole set of user behaviors will not be available during attribution. There are two primary limitations – the coverage of data collection systems, and time.

3.1.1.1 Limitation due to coverage

Many participants in the effort to drive a consumer to a specific action will not have first party access to the complete set of user events.



There are various techniques available to merge the sets of data, including cookie mapping, extra web beacons, and 3rd party mapping services.

3.1.1.2 Limitation due to time

Storage and processing limits may restrict the look back period for attribution models. Depending on the attribution technology provider, the inclusion of events that occurred more than a certain amount of time before the desired outcome can be limited.

Time also plays a factor in the recognition of the user, because the longer the interval between event and desired behavior, the higher the probability of "user identifier churn." User identifier churn occurs when existing identifiers, such as cookies, are deleted and replaced by new values.

3.1.2 User Identifiers

In the short history of digital advertising, the ready availability of user identifiers has often been taken for granted. As the ecosystem matures, digital advertising will need to address situations where user identifiers are no longer ubiquitous.

The essential areas of growth that best highlight the challenge of reduced user identifiers are the multiple screen scenario, and dealing with online/offline campaigns.

3.1.2.1 Multiple Screens

A user who accesses the same marketing message across multiple devices will generally have a different user identifier for each device. This means that their engagement with the mobile version of the web site can't be counted in the attribution model for a purchase made via their laptop.

The loss of coverage due to multiple screens is mitigated when the user is asked to authenticate with the same credentials across all devices, provided that an identifier based on this authentication is made available to the advertising systems.

3.1.2.2 Multiple Media Types

A user who receives a supporting message from traditional media will likely behave differently than a user who does not. However, the traditional media ecosystem doesn't maintain a unique identifier that can be mapped to the IDs used by digital advertising.

An understanding of the effect of digital advertising in the traditional media ecosystem can be established by mapping user behaviors into compatible buckets (geographic and demographic), and then performing marketing mix modeling.

4 Similar Techniques

4.1.1 Comparison to Marketing Mix Modeling

Like marketing mix modeling, online attribution aims to provide a framework for understanding the valuation of the various consumer-directed messages.

Marketing Mix Modeling	Online Attribution
Aggregate Data Modeling	User-specific Data Modeling, requiring consistent user IDs
Affects Future Spend	Affects Current Spend

4.1.1.1 Goals of Marketing Mix Modeling

Marketing mix modeling is the use of statistical analysis to optimize future media mix and promotional tactics with respect to sales revenue or profit. It works by modeling large aggregate datasets, and does not require uniquely identified individuals.

4.1.1.2 Goals of online attribution

Online attribution uses the use of the techniques described in this document for several purposes:

- To optimize ongoing campaign delivery based on user-specific data.
- To measure the ROI of campaigns.
- To determine the amount of payment owed to performance-based pricing vendors.

The techniques described are the modeling of multiple signals to understand motivating events leading toward a desired outcome with the specific intent of allocating active revenue based on this understanding.

5 Terminology

- Ad
  A commercial message targeted to an advertiser's customer or prospect.
- Ad Impression
  The delivery of an ad to a web client.
- Ad View
  Historically used interchangeably with ad impression, emerging techniques are being evaluated for determining whether an ad is in the viewable area of a web page.
- Click
  The measurement of a navigation from one document to another by activating a hyperlink.
- Direct Navigation
  Direct navigation is the act of typing in a URL into the address bar of a web browser.
- Engagement
  Content-specific metric that evaluates the interest level of a user.
- Engagement Depth
  Specific events can indicate different levels of interest by a specific user. When a formal hierarchy is defined for understanding this, the engagement depth is measured.
- Event
  Any user action that is measured.
- Factor
  A supplementary data point that modifies the impact of the event with which it is associated.
- Frequency
  How often a specific event occurs.
- IP Address
  Generally, the IPv4 address of an internet-connected device.
- IP to Geo
  The process of establishing a ZIP code or other geographic coordinate from an IP address.
- Path
  The serial list of events experienced by a user.
- Recency
  How recently an event occurred, whether measured by time or by number of intervening events.
- Scope
  The set of data available to a participant.
- Search
  An event where the user has provided keywords to a form in order to locate something of interest.
- Sequence position
  Whether an event was the first, last, or n^{th} in sequence. Determining the "first" event is not perfectly reliable in attribution efforts, since cookie churn and scope issues may mask the true first event.