**Measuring Ad Effectiveness**

**Note: This use case is based on conversations with multiple companies done on background. The use case is intended to illustrate a basic mechanism of measuring the effectiveness of advertising. There is no claim that any one company or set of companies does precisely the steps shown here. One key point is to show how hashing has been used in practice for the purposes of measuring online ad effectiveness.**

**Business Goals**

One challenge of online marketing is understanding whether consumers who see particular advertisements end up being more likely to purchase retail products than consumers who are not exposed to those ads.

**Privacy Concerns**

One challenge for businesses is to measure the relationship between a consumer’s ad visit or banner view to transactions occurring later in a way that minimizes privacy risks.

The linking of purchases and online advertisements in this manner can create a number of privacy concerns:

* Although the output of the process is measuring aggregate ad effectiveness, personal information of an individual consumer, as well as linking purchases to advertisements, is used to create the final report;
* Consumers may have been promised anonymity for their online activity;
* Consumers may have been promised that their personal data would not be shared;
* Local law may require consent for sharing of personal information.

Additionally, businesses that have proprietary lists of their customers do not want their customer lists to be available to other marketers.

**Data Flow and Technical De-Identification**

A range of aggregation and de-identification techniques are used in order to protect information while fulfilling business goals. An advertiser/retailer buys ads that are delivered on one or more websites (either by the websites or an ad network) that have a substantial number of registered users. Each time an ad is delivered, a flag is added to the profile the website or ad network has about the user who saw the ad. Typically, user registration information is de-identified through a hashing process with a salt. Meanwhile, the retailer’s customer list is de-identified through a hashing process with the same salt. The de-identified datasets are then matched in a data vault (some companies will involve an independent third party to facilitate the matching process; others will use a segregated database within one of their company’s servers). An aggregated report is then prepared (by the retailer or by another service provider) that shows the ads’ effectiveness. Once the matching data is compiled into an aggregated report, both the ad views dataset and the retailer’s customer list are deleted.

The following diagram provides a high-level overview of this process:

* **Website/Ad Network:** User ad views and clicks are collected
	+ Identifies ad views and clicks with registration information (name, emails, etc.)
* **Retailer/Advertiser:** Consumer purchasing behavior is recorded
* **Data Management Provider:** matches *user identifiers* with *Customer list* in segregated data vault and returns only matched group to **Retailer/Advertiser**
* **Retailer/Advertiser:** compares purchasing behavior from “matched group” to controlgroup
	+ **Retailer/Advertiser** is contractually prohibited from using data for any other purpose
* **Retailer/Advertiser:** creates high level aggregated report representing ad effectiveness
	+ High level aggregated report represents only Non-unique statistical data
* **Retailer/Advertiser:** Consumer list are hashed and sent to **Data Management provider** for matching
	+ Hashed using common key
* **Website/Ad Network:** User identifiers are hashed and sent to **Data Management provider** for matching
	+ Hashed using common key



**Administrative Controls and Legal Safeguards for Consumer Data Matching**

The administrative controls and legal safeguards surrounding this process can help ensure the integrity of the de-identification. Typically these measures will include one or all of the following:

* Obligations on all parties to delete datasets after they are used.
* Obligations not to release the report publicly or to unspecified private parties.
* Internal data segregation to ensure that data is not identified.
* Internal access control measures to ensure that only specified employees can access the data.
* Contractual protection to ensure there will be no attempt to identify consumers from the aggregated final report.
* Third party intermediary is sometimes placed in the center to segregate data