The Drawbridge Cross Device Matching Technology

Dr. Kamakshi Sivaramakrishnan, PhD
Founder and CEO, Drawbridge Inc.
PhD, Information Theory, Statistics, Probability Theory, Stanford University
Papers and patents by Dr. Sivaramakrishnan

Introduction

What Drawbridge does:

- Matched over 200MM desktop cookies to related mobile devices (belonging to the same user) probabilistically using *no PII*
- Uses cross device audience behavior to serve display ads over *open mobile exchanges and ad networks*

How it works:

The Drawbridge pairing technology is a fully anonymized, data safe algorithmic procedure that matches users between mobile and desktop devices. This pairing technology does not use any personally identifiable information (PII) such as email handle, social network handle, full legal name, address book, phone number etc. It uses fully anonymized simple http cookies[1] in mobile and desktop browsers to identify and match users across devices.

Drawbridge cookies users on a desktop browser through its partnership with a variety of Data Management Platforms (DMPs) and Ad Exchanges. On the mobile browser Drawbridge operates through its demand side partnership on mobile supply side platforms/ exchanges/ ad networks. In doing so, Drawbridge cookies users upon serving an ad impression on a mobile device. In other words, the only information that is exposed to Drawbridge is a simple display http (web service) ad request on mobile and desktop.

Drawbridge's technology ("The Bridging Algorithm") [5] uses a statistical triangulation technique based on information theoretic concepts to match a desktop and mobile cookie-- creating a bridge between the desktop and mobile cookie.

The Bridging Algorithm

The Bridging Algorithm estimates the probability that an arbitrary desktop cookie (from a browser on a desktop or laptop) and a mobile cookie (from a mobile browser or in app) belong

[1] See Appendix A for a more detailed explanation of the term

to the same user. A couple of different events comprise signals, ~~or~~ data or observations for the cross-device matching:

- An ad request from a mobile or desktop device
- A desktop cookie synch with a (desktop) DMP (like Exelate, Lotame, etc) to avail 3rd party audience data payload

Concretely, let's define a couple quantities to understand how the algorithm works.

MC denote an arbitrary Mobile Cookie
DC, an arbitrary Desktop Cookie.
Given M_t observations of *MC* and N_t observations of *DC* at some time *t*, the algorithm estimates the following probability

$$p_t = P(MC \ \& \ DC \ belong \ to \ the \ same \ user \mid M_t \ observations \ of \ MC, N_t \ observations \ of \ DC)$$

This probability is estimated using features and signals from the observations (in other words, nothing but ad requests and cookie synchs for 3rd party audience data from the DMP). The algorithm further applies Bayesian techniques [3] to update the probability, p_t , at the next instant of time, *t+1* resulting in something like this:

$$p_{t+1} = g(p_t, M_{t+1}, N_{t+1})$$

where $g(\cdot)$is a function that updates the probability of DC and MC belonging to the same user based on additional observations of *MC* and *DC* and a past estimate. With the entropy [2] of the observations not increasing with time, we are able to get increasingly confident about the probability estimates as time passes. At any time instant, if we are confident enough of the probability estimate, as measured by a threshold ($\underline{Sc}(0.6)$) necessary to maintain an overall accuracy level of cross-device matching at 60%, the pair *{MC,DC}* graduates to a matched user pair.

	Bridging Algorithm

	At time t_0, initialize p_0
	for $t = 1$ *to* N do
	for each *mc* in the set of Mobile cookies *{MC}* and *dc* in the set of desktop cookies, *{DC}* do
	Input: $M_t, N_t, p_{t-1}, \underline{Sc}(0.6)$
	Update step: $p_t = g(p_{t-1}, M_t, N_t)$
	Measurement step: Construct confidence scores, Sc_t *on* p_t
	Promotion step: $If\ Sc_t >$

```
Sc(0.6), promote mc, dc to a matched user pair
end
end
```

Precision and Recall

The Drawbridge cross device pairing platform constantly calibrates the precision and recall of the learning model. This is done by constructing a confusion matrix[2] on the training samples of the **same** user handles on mobile and desktop devices.

Precision measures the accuracy of cross device matches and Recall measures the coverage of the same. Together, this comprises a comprehensive measurement of the performance of the core technology. Definitions of the above measures are available in any information retrieval literature [1] and are described in Appendix B. There is a precision/recall (PR) [2] tradeoff in any learning system. The latest calibrated PR tradeoff in the Drawbridge pairing technology is shown in Fig.1 below:

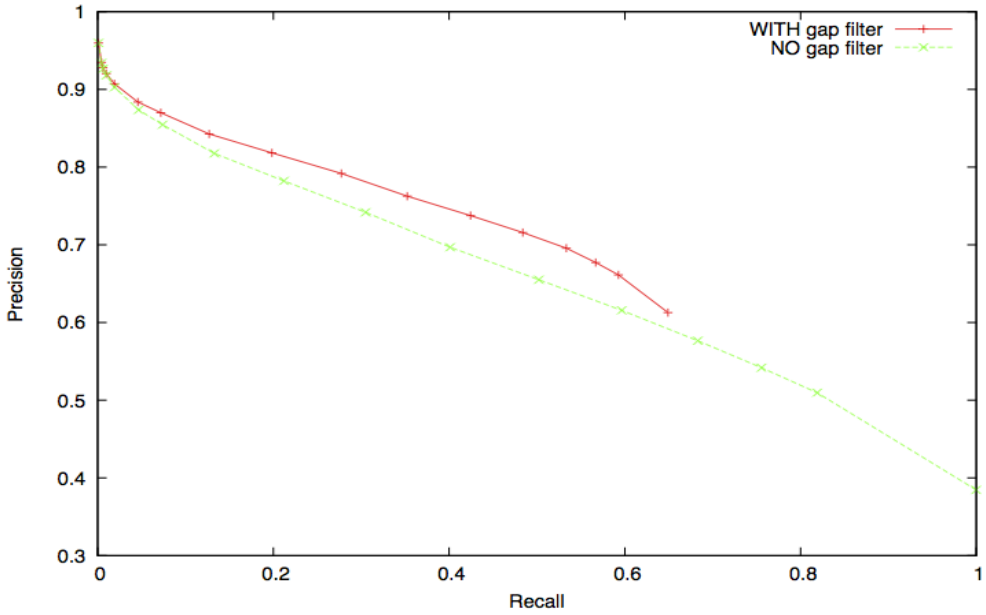

Fig.1 Precision Recall tradeoff for the Drawbridge coss device matching technology

The two graphs in the above chart are a performance depiction of two different techniques for the degree of confidence calibration.

System Overview

[2] See Appendix B for definitions and further details

Fig. 2 below illustrates the flow of information in the Drawbridge ad serving and data management pipeline. User data is ingested, processed, stored, and leveraged in a fully anonymized (non-PII) fashion. As shown, nowhere in the flow is the insertion or transaction of any PII information.

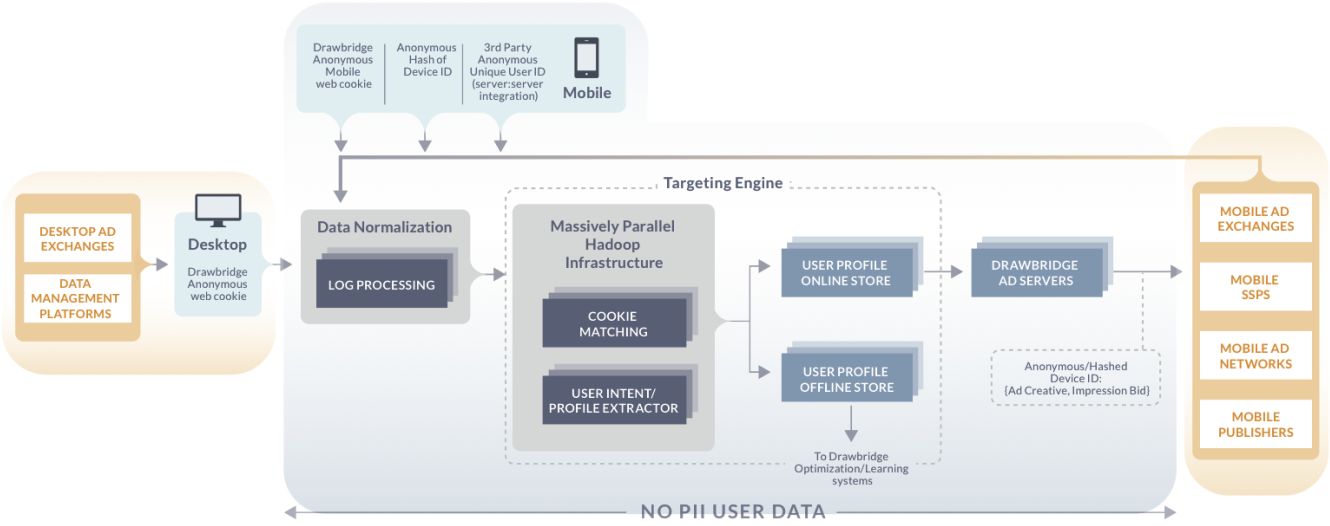

Fig 2. System overview describing the end-to-end flow of information in the Drawbridge ad serving and data management pipeline

Data storage and dissemination through the Drawbridge system

Storage and dissemination of the matched cookies is an integral and crucial aspect of the Drawbridge ad serving systems. As with previous steps, there is no PII information ingested anywhere in the system nor a location trail stored against the matched cross device cookies.

Storage

Data is stored in the Drawbridge system as a table map of targeting data of the following form:

$$\{\text{Drawbridge Anonymous Web cookie}_A\ ,\ \text{Drawbridge Anonymous Mobile Cookie}_B\} \rightarrow \{\text{Audience Segment}_1,\ \ldots,\ \text{Audience Segment Label}_N\}$$

where Drawbridge Anonymous Web cookie_A is the desktop browser cookie and Drawbridge Anonymous Mobile cookie_B is the mobile cookie of what is estimated to be the same user by Drawbridge. The list of Audience segments denotes audience profiles based upon demographic,

behavioral and intent data inferred from the desktop and/or mobile cookie. Here is an example of the data stored on the cookie pair:

```
{3D9efb55e39df30077629fefd7731d6a93, 68753A44-4D6F-1226-9C60-0050E4C00067} →
{Gender: Female, Age: 25-34, Behavior Type: Style and Fashion, Behavior Type:
Frequent Traveler, Intent: In market for Auto, Make- BMW Model- 528i Year-
2012}
```

There is no other information including location coordinates or IP addresses stored against this map.

Dissemination

The map described earlier in this section is not disseminated outside of the Drawbridge systems. Drawbridge is the primary and only consumer of the map and targeting data for the purposes of behaviorally relevant ad serving. In other words, the Drawbridge ad server makes an ad serving decision in response to an incoming ad request by looking up the map for audience data and matching its relevance against available its available ad inventory.

Drawbridge Technology v/s Fingerprinting

Drawbridge Cross Device Matching is not a Fingerprinting [6] technology nor a derivative thereof. There is no intrusive client-side presence through an SDK or a Javascript tag covertly sniffing attributes of the device or traffic. The Drawbridge technology does not use the HTML5 local storage or flash to store any client side cookie on the device.

The Drawbridge technology is strictly privacy compliant in its ability to honor the DNT (Do Not Track) on the mobile cookie if its corresponding desktop counterpart has opted out. This approach on cross device DNT is currently under evaluation with the Network Advertising Initiative (NAI).

References

1.Introduction to Information Retrieval. C.D. Manning, P. Raghavan, H. Schütze. Cambridge UP, 2008. *Classical and web information retrieval systems: algorithms, mathematical foundations and practical issues.*
2. Elements of Information Theory, by T. Cover and J. Thomas, Wiley 2006 (**2nd edition**)
3. Bayesian Data Analysis, Second Edition by Andrew Gelman, John B. Carlin, Hal S. Stern, and Donald B. Rubin
4. Glossary of Terms, Editorial for the Special Issue on Applications of Machine Learning and the Knowledge Discovery Process, Machine Learning (1998), 30(2-3), Ron Kohavi, Foster Provost
5. System and Method for Determining Related Digital Identities, Kamakshi Sivaramakrishnan, Devin Guan, Tin H Kyaw, Jerry Ye, Ravi Menon, Yang Yu, inventors; 2012 May 10. United States provisional patent filed US 61,645,549
6. Wikipedia page on Device Fingerprint.

Appendix A

What is a cookie ?

A cookie is a piece of data issued in an HTTP response (for example, an ad response) for future use by the HTTP client (for example, a web browser). The client then re-supplies the cookie in subsequent requests to the same server. This mechanism allow the server to store user preferences and identify individual users.

How is the Drawbridge cookie set?

Drawbridge Ad/Data Servers supply cookies by populating the **set-cookie** response header with the following details:

Name	U
Value	3D9efb55e39df30077629fefd7731d6a93
Expires	Thu Apr 14 14:59:17 2022
Path	/
Domain	.adsymptotic.com

Here is the sample Drawbridge HTTP response

HTTP/1.1 200 OK
Content-Type: text/html; charset=utf-8
Set-Cookie: U=3D9efb55e39df30077629fefd7731d6a93; expire=Thu Apr 14 14:59:17 2022; domain=.adsymptotic.com; path=/

What does this cookie look like in a mobile application?

In the case of a mobile application, a cookie can be set using the above standard HTTP cookie approach. Additional options for a cookie are one of the following:

- An anonymous one-way hash of a randomly generated fully anonymous 40 character string. This has the limitation of being available only inside an application and does not carry across applications. For example
    - Name: `NSUUID`, Value: `68753A44-4D6F-1226-9C60-0050E4C00067`
- An anonymous one way hash of a device ID
    - In case of Android, the device ID is the Android ID. For example,
        - Name: `ANDROID_ID, Value:643ba7cf4165bf8b`

- In case of iOS, the device ID is the identiferForAdvertising, also known as the IFA. For example,
  - Name: `advertisingIdentifier`, Value:`68753A44-4D6F-1226-9C60-0050E4C00067`

Appendix B

A confusion matrix [4] contains information about actual and predicted classifications done by a classification system. Performance of such systems is commonly evaluated using the data in the matrix. The following table shows the confusion matrix for a two class classifier.

The entries in the confusion matrix have the following meaning in the context of our study:

- *a* is the number of **correct** predictions that an instance is **negative**,
- *b* is the number of **incorrect** predictions that an instance is **positive**,
- *c* is the number of **incorrect** of predictions that an instance **negative**, and
- *d* is the number of **correct** predictions that an instance is **positive**.

	Predicted	
Actual	Negative	Positive
Negative	a	b
Positive	c	d

Several standard terms have been defined for the 2 class matrix:

- The *recall* or *true positive rate* (*TP*) is the proportion of positive cases that were correctly identified, as calculated using the equation:

$$TP = \frac{d}{c+d}$$

- Finally, *precision* (*P*) is the proportion of the predicted positive cases that were correct, as calculated using the equation:

$$P = \frac{d}{b+d}$$