Crying Wolf?
On the Price Discrimination of Online Airline Tickets

Thomas Vissers[1], Nick Nikiforakis[1], Nataliia Bielova[2], and Wouter Joosen[1]

[1] iMinds-DistriNet, KU Leuven, 3001 Leuven, Belgium
{firstname.lastname}@cs.kuleuven.be,
[2] Inria, France
nataliia.bielova@inria.fr

Abstract. Price discrimination refers to the practice of dynamically varying the prices of goods based on a customer's purchasing power and willingness to pay. In this paper, motivated by several anecdotal accounts, we report on a three-week experiment, conducted in search of price discrimination in airline tickets. Despite presenting the companies with multiple opportunities for discriminating us, and contrary to our expectations, we do not find any evidence for systematic price discrimination. At the same time, we witness the highly volatile prices of certain airlines which make it hard to establish cause and effect. Finally, we provide alternative explanations for the observed price differences.

1 Introduction

In today's web, users are accustomed to getting a variety of services without explicitly paying for them. Thus, the average user has access to mail services, maps, search, social networks, cloud storage, and online newspapers, without paying a subscription fee to any of them.

These services, in turn, usually make a profit either by directly providing ads to their users, or gathering user details that are used in house, or sold to other companies, in order to provide better targeted ads. These two are not strictly separated because the gathering of information about a user and the showing of ads can (and does) happen through a single action, such as the inclusion of a remote JavaScript library from an advertising syndicator.

An alternative monetization strategy to targeted advertising, is *price discrimination.* Price discrimination involves varying the price of any specific product or service, depending on the amount that a customer is willing to pay. The vast amounts of user-data gathered on the modern web are a natural fit for such a pricing strategy [20]. Knowing, for instance, that a user visits webpages which sell high-end goods can be used to infer the income level of that user. This knowledge can, in turn, be used to dynamically increase the price of a product on a, seemingly, unrelated website, simply because that user is likely to have the ability and willingness to pay more for that product. Prior research by Mikians et al. [17] showed that price discrimination is used to vary the prices of digital products (such as ebooks) based on the geographical location of the user. The authors later showed that crowdsourcing can be used to unearth harder-to-find cases of price discrimination, such as the ones present in regional e-shops [18].

In this paper, motivated by several anecdotal reports [7,14,23], we investigate whether airlines practice price discrimination on the tickets sold through their websites. Airline tickets are an attractive target for price discrimination due to the highly volatile nature of their prices. An airline could, for instance, recognize the same user over time, e.g., through the use of cookies, and gradually increase the price of a specific ticket. The user is likely to attribute the price increase to flight congestion (many people buying tickets for that specific flight) and buy the ticket at an increased price, out of fear that the price will increase even more.

Through a three-week long experiment, involving 25 different airline companies, we find that, at this time, airlines seem *not* to be employing any *systematic* price discrimination. While we did observe some price fluctuations that could be interpreted as due to discrimination, in most cases, these fluctuations could be explained through other means, e.g., increased tax on tickets of a regional airline, when those were purchased outside of that specific region. At the same time, we show how difficult it is to establish cause and effect for airline prices and we make available the set of prices that we collected so that it can be used on future research on the topic of price discrimination [24].

2 Background

2.1 User Tracking

Some of the strategies that could be of use in price discrimination require the tracking of users between page loads of the same site, as well as across different websites.

Browser cookies have emerged as the de facto way of tracking users in the modern web. The cookie mechanism allows a web server to store a small amount of data on the computers of visiting users, which is then sent back to the web server upon subsequent requests. Cookies can be separated in *first-party* and *third-party* depending on the way on which they are set and read.

First-party cookies are cookies set by the website that the user is consciously visiting, e.g., a user visiting the website of Delta Airlines and receiving cookies for the `delta.com` domain. First-party cookies are used to, among others, maintain state over the stateless HTTP protocol, and recognize visitors between page loads.

Third-party cookies are cookies given by third-party servers which provide content to a first-party page. If, for instance, Delta airlines uses a Facebook "Like" button, Facebook has the ability to store and read cookies when the visitor visits the `delta.com` website. Third-party cookies are more intrusive than first-party cookies, since they can be used to track users in a cross-site way [12, 22]. If for instance, the sites `travel.com` and `football.com` show ads from the same advertising syndicator, that syndicator has the ability to correlate users between these sites and build a browsing profile for each user that can be then monetized in series of ways, including targeted ads and the selling of a user's information to larger data aggregators and brokers.

Next to cookies and other techniques involving the storing of identifiers at the client-side [15], web-based device fingerprinting is another technology that has recently emerged and can also be used to track users. In fingerprinting, several attributes of a user's browsing environment, e.g., a user's browser, platform, screen dimensions, and installed plugins, can be combined to form a fingerprint that is, for all practical purposes, unique [5,9,19]. The advantage of fingerprinting is that it is stateless and, because of that, hidden from a user, i.e., no cookies to inspect or delete.

2.2 Anecdotal Evidence of Price Discrimination

Over the past years, many articles, blog posts and Internet fora have reported the use of price discrimination in online travel tickets. Generally, these claims are based on anecdotal evidence. In this section, we briefly list several noteworthy cases and discussions.

In 2011, a massively "retweeted" tweet [23] implied that Ryanair was tracking users through cookies and raising their prices based on the user's previous visits. On a blog post that followed [7], similar claims were made by people entering the discussion: TheTrainline.com, Expedia, Easyjet, Virgin, Lastminute and Eurostar were all accused of price discrimination based on previous visits or search queries. The Wall Street Journal reported a year later that Orbitz [3], an online travel agency (OTA), presents Apple-branded computers with more expensive flight and hotel search results first [14], a practice that is referred to as *search discrimination* [17].

In 2013, a column published in USA Today [16] specifically discussed the current technical possibilities and legal implications of price discrimination in the travel sector. Furthermore, they claim to have observed a price difference with a cookie-cleared browser in tests they conducted in 2007.

More recently, in January 2014, a blogger explained how he got a less expensive fare from Kayak [2], a travel metasearch engine, by using a Canadian instead of an American IP address [6].

3 Large-scale analysis

As noted in the previous section, airline companies have been accused of serving different prices according to a client's geographical location, the device they use, their consumer profile or their previous search queries. We evaluate this claim by analyzing 25 airlines continuously for 3 weeks, with dozens of unique user profiles. These profiles are constructed in light of the user-tracking methods explained in Section 2.1, and the potential discriminating factors.

To properly assess these practices, we emulate real users performing search queries for specific flight routes. We do this repeatedly for all the different user profiles on all regarded airlines. For every conducted search query, we keep track of the prices that are presented to that particular user.

In the following paragraphs, we will explain the techniques used to automate the search queries and to construct different user profiles.

3.1 Automated search querying

During 21 days, we queried 25 airlines twice a day, with 66 user profiles, from two different geographical locations simultaneously resulting to a total of over 130,000 queries. We automated this large-scale analysis by building a scraper using CasperJS [1], which is a navigation scripting and testing utility for PhantomJS [4], a headless browser. We ensured our scraper followed the exact same steps a normal user would follow when manually searching for airline tickets. The scraper ran from January 15, 2014 until February 5, 2014, querying for flights departing on March 22, 2014 and returning on March 30, 2014. The specific flight routes that were queried were different for every airline, but fixed throughout the entire analysis. The order in which the scraper iterated the 66 user profiles for every airline was randomized at each run.

3.2 Constructing user profiles

We constructed a total of 66 unique user profiles, which are the result of combining several different subprofiles, described in the following sections. The majority of these leveraged the ability of PhantomJS to easily configure the browser's internals and cookie settings. Additionally, two geographical subprofiles were created, which were both combined with the 66 user profiles.

Browser and OS profiles The browser and operating system (OS) profiles are set up to observe potential discrimination based on the user's device. From a web server's perspective, the *User-Agent string* is the most straightforward feature for identifying the client's Internet browser and OS. When making a web request, the client sends this string as an HTTP header to the web server. Moreover, the User-Agent string also resides in the browser's *navigator* Javascript-object, together with other indicative strings, such as *vendor, platform, appCodeName, appName* and *appVersion*. A web server can use this information to determine which browser and OS a particular client is using. For example, the User-Agent string of Safari Mobile running on an iPad with iOS 4.3.5 is easily recognized:

```
Mozilla/5.0 (iPad; U; CPU OS 4_3_5 like Mac OS X; en-gb)
AppleWebKit/533.17.9 (KHTML, like Gecko) Version/5.0.2
Mobile/8L1 Safari/6533.18.5
```

By spoofing the information in the User-Agent string and the navigator object, we were able to mimic different OSs and browser profiles. Our scraper was programmed to emulate several different profiles: *Internet Explorer 9 on Windows 7, Chrome 29 on Windows 7, Firefox 22 on OS X 10.7, Safari 6 on OS X 10.7* and *Safari Mobile 5 on iOS (iPad)*.

Furthermore, we configured additional Chrome, Safari and Internet Explorer profiles with the Do Not Track (DNT) header set to 1. The DNT header is an HTTP header which can be utilized by users to request from a web application to disable their tracking [25].

Consumer profiles Inspired by the study of Mikians et al. [17], we designed 3 consumer profiles: *affluent*, *budget* and *flight comparer*. These profiles were composed by gathering cookies of websites that can be associated with them. For example, the cookies for the *affluent* profile were gathered by visiting sites associated with luxury goods and financial news. For the *budget* profile, we collected cookies from websites offering discounts and product reviews. And for the *travel comparer* profile, online travel agencies and travel metasearch engines, such as *kayak.com*, were included.

For each profile, we kept a separate *cookie jar*: all cookies gathered by a particular consumer profile were saved in an isolated file. When a consumer profile needs to be instantiated, its cookie jar can simply be loaded in PhantomJS. This way, potential third-party trackers, could access the consumer profile cookies while our crawlers visited the website of each airline.

Cookie setting profiles As discussed in Section 2.1, websites can make use of first-party and third-party cookies to track users between loads of the same page and across different websites. In our experiment, we accounted for four different cookie settings, when gathering ticket prices for each airline:

- Only consumer profile cookies (gathered in advance)
- Consumer profile cookies along with first-party cookies
- Only first-party cookies (also known as no third-party cookies)
- No cookies

In our experiments, the first-party cookies are the cookies set by the airlines themselves which can be used to track a user's previous visits and searches and potentially influence the prices shown to users over time.

Geographical profile Websites often resort to the user's IP address as an indicator of locality [6, 17]. In order to set up different geographical profiles, the scraper ran simultaneously from two different locations, one in New York, United States and one in Leuven, Belgium.

3.3 Hypothesis

In order to evaluate the presence of systematic price discrimination, we test for the following hypothesis: if airlines present unequal prices to different users according to their consumer profile, location, or browsing history, we should be able to observe this when comparing the prices presented to a set of emulated users that differ in those characteristics, over several points in time. For instance, if a company is presenting higher prices to users of Apple products, a time-series graph of ticket prices should place our emulated Apple user "above" other users.

3.4 Limitations and assumptions

The search queries were not executed in parallel, but consecutively per profile on each location. The rational for this design choice is the following: performing a large number of simultaneous search queries may disrupt the airline's web server. Furthermore, if they engage in some form of rate limiting, our requests could be rejected, prohibiting us from reliably measuring prices. Similarly, we performed all search queries only twice a day for each airline, rather than continuously for our three-week measurement.

All profiles originating from a single geographical location had the same IP address. However, because of the widespread use of NAT technology, it is commonly assumed inadequate for a web server to distinguish between users based on their IP address. Since covering every possible geographical variance is virtually impossible, we only used two geographical profiles during our measurements. In preliminary experiments, an additional location (France) was also evaluated, but no additional differences were observed.

Finally, note that, in this paper, we only considered websites provided by each airline company, rather than travel aggregators and meta search engines. As such, our results should not be generalized to characterize the latter.

4 Results

Most of conducted the search queries returned multiple flight options, differing in terms of flight routes and departure time. Therefore, after extracting the prices from all search results, the minimum, maximum and average price per query were calculated. Next, this data was plotted as a time-series for each profile. Contrary to our expectations, we were not able to find any clear evidence of price discrimination: the majority of the plots reveal no *consistent* price difference between the user profiles.

For example, Fig. 1 shows the fluctuations of Ryanair's minimum prices recorded by the scraper located in the United States. We grouped together all profiles that had the same price evolution throughout the entire experiment, resulting in two distinct subsets. Of the 42 measurements, there is only one instance (#33) where a different price is given to a subset of the profiles. When examining the exact timestamp of each profile's measurement, we find that all profiles with a higher price at #33 were recorded consecutively and prior to all measurements with a lower price. This reveals that, most likely, the observed price difference is due to the price actively dropping while conducting the measurements for #33, rather than due to price discrimination.

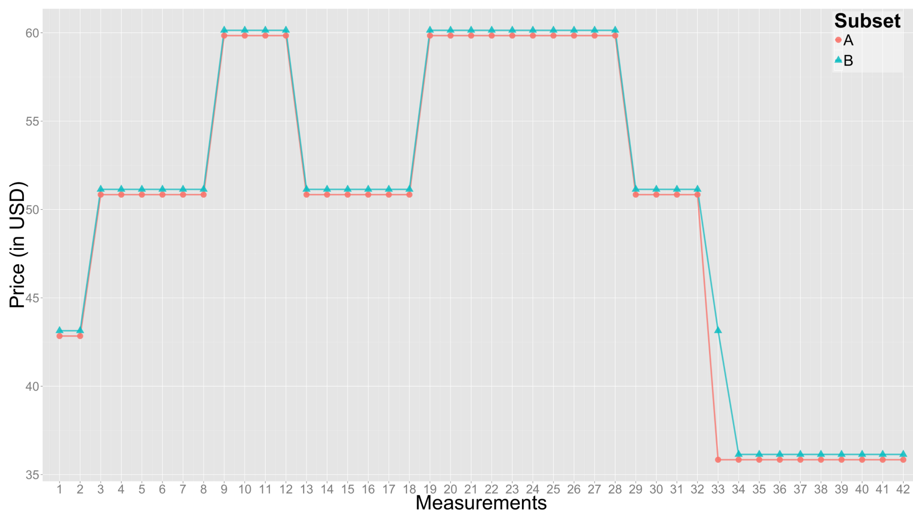

Fig. 1. Minimum price offered by Ryanair for all user profiles (US) over 21 days, for the CRL - LPA route. A small amount of vertical jitter was applied to the data points, in order to prevent the plotted lines from overlapping.

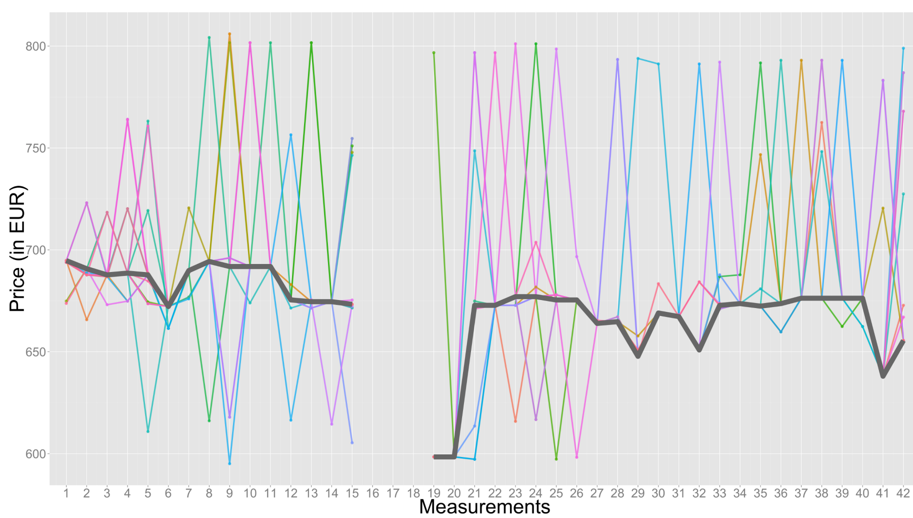

Fig. 2. Mean prices offered by KLM for all user profiles (BE) over 21 days, for the JFK - AMS route. Note, that there are no measurements between #16-19, due to a network problem. The thick black line represents the trend of all other profiles.

Some airlines showed rapid variations in the availability of tickets. In Fig. 2, the mean price of a KLM route is plotted for each profile. For every measurement, the vast majority of profiles has the same mean price, which results in a clear and steady trend over time. However, numerous steep peaks and valleys occur, affecting different profiles every time. When looking into these outliers, it appears that they are caused by a constantly changing availability of flights. The fluctuations are so dense that no two profiles shared the same price evolution throughout the experiment.

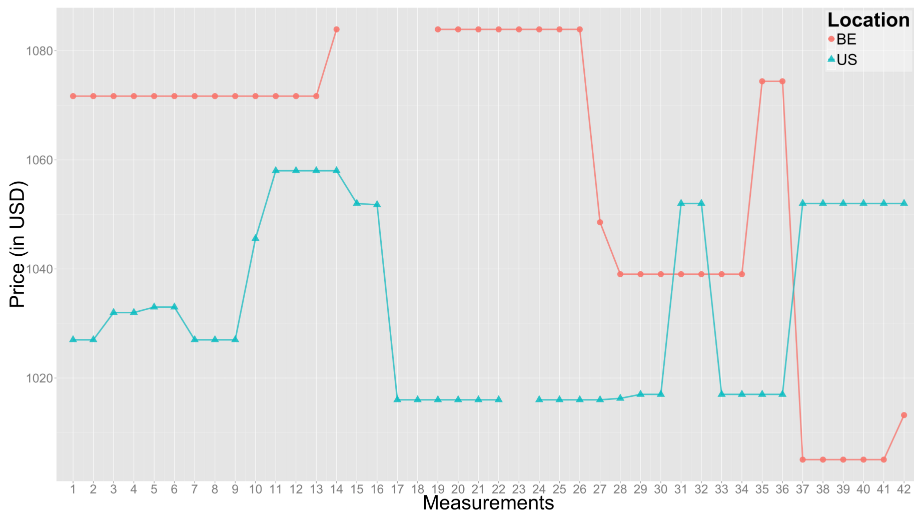

Fig. 3. Minimum price offered by United Airlines in Belgium and the United States over 21 days, for the SFO - BRU route. For a meaningful comparison, the listed Belgian prices (in EUR) were converted to USD using the applicable exchange rate for that time.

To evaluate price discrimination based on location, we examined the calculated statistics of the geographical profiles, e.g., we compared the absolute minima of all American profiles to the absolute minima of Belgian profiles. Again, no clear evidence of price discrimination was found: Nearly all airlines show similar prices for both locations. However, as seen in Fig. 3, there was one exception with United Airlines: for the first 15 days, the Belgian price was consistently higher than the American price. At times, the Belgian price was listed 67.92 USD above the American. Remarkably, though, near the end of the experiment, the Belgian minimum price dropped below the American price several times.

In a preliminary experiment, we looked into price discrimination in multiregional and multilingual websites. We compared the native region or language (with respect to the airline company's origin) with a non-native alternative. A particularly interesting find was the Argentinian website of American Airlines. As shown in Fig. 4, the price was consistently about 19% higher than the price from the American website. After looking into this phenomenon, we discovered that, as of March 18, 2013, the Argentinian government has imposed a tax of

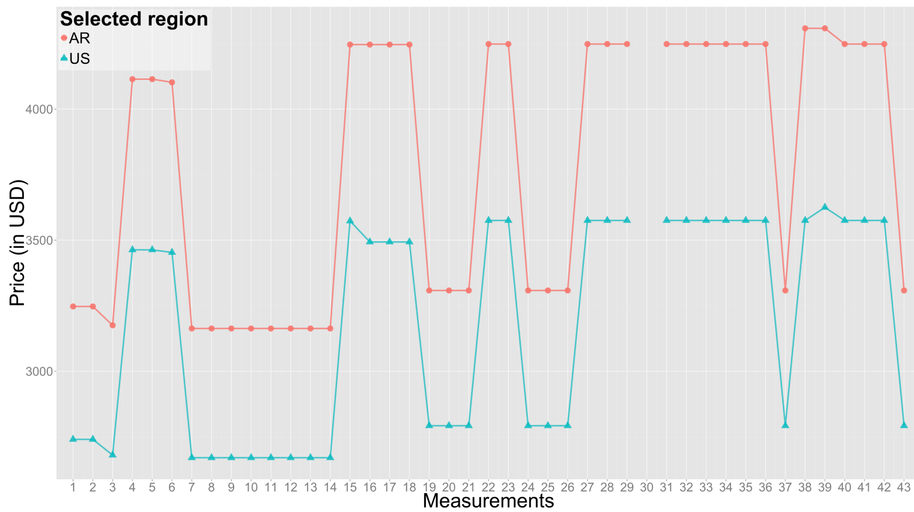

Fig. 4. Price for a particular SFO - BRU flight offered by American Airlines for the non-native (Argentina, in red) and native (US, in blue) profile over 21 days.

20% on credit card and holiday package purchases [21]. These taxes are countermeasures for Argentina's high inflation rate and perfectly explain the systematic price difference.

5 Alternative explanations for the anecdotal evidence

In this section we discuss several possible phenomena that may appear to a user as evidence of price discrimination, but actually have other causes. These alternative explanations are based on the results of our analysis and additional information concerning regulations and business specifics.

As shown earlier in Fig. 2, flight routes and their prices can be extremely volatile. When we look at consecutive measurements for KLM, it is not uncommon for a flight to suddenly become unavailable, only to reappear in the next measurement. Taking into account that our measurements are roughly 1 minute apart, it is likely that users will encounter different prices when performing two consecutive searches.

When performing a search from several geographical locations, a user may encounter different prices due to regulations and agreements specific to that region. For example, differences may occur simply due to currency conversions [13]. Also, a country may impose additional taxes, e.g., as seen with Argentina. In contrast, in Peru, some airlines have cheaper fares solely for Peruvian residents [11]. Specific agreements and partnerships may differ between regions as well, e.g., Southwest Airlines prohibit third-parties to use their price information. This has, in the past, lead to an injunction against Orbitz [8] and withdrawal of its flight data from the Airline Tariff Publishing Co. (ATPCO). This is a possible explanation for the less expensive flights on Kayak while searching from Canada [6], mentioned in Section 2.2. As it turns out, the Canadian results contained the lower priced Southwest Airlines flights, while the American results did not.

Another phenomenon that can be perceived as price discrimination, is the caching of prices: as explained in [10], travel companies do not display real time price information. Instead, for the sake of cost and time efficiency, prices are cached. This can possibly result in discrepancies between prices that were initially displayed and the price at the final step of the booking process. Furthermore, subsequent search queries may display the updated price which can be higher or lower than the previously reported price.

6 Conclusion

In this paper, motivated by several anecdotal reports, we investigated the presence of price discrimination in airline ticket prices. Contrary to our expectations, our three-week long experiment, spanning 25 different airlines, tens of user profiles, and two geographical locations, did not reveal the use of any systematic price discrimination. Based on our findings, we argue that at least some of the anecdotally reported price differences can be explained through alternate means, such as government policies and variable taxation. We hope that our study can serve as a reference for the current state of practice, and inspire longitudinal studies that may uncover the future adoption of price discrimination by airline companies. We make the data from our research publicly available so that it can be used in future research [24].

Acknowledgments: We want to thank the anonymous reviewers for the valuable comments. This research was performed with the financial support of the Prevention against Crime Programme of the European Union (B-CCENTRE), the Research Fund KU Leuven, the IWT project SPION, the EU FP7 projects NESSoS, WebSand and the French ANR-10-LABX-07-01 Laboratoire dexcellence CominLabs.

References

1. Casperjs - navigation scripting and testing utility for phantomjs. `http://casperjs.org`
2. Kayak - flight, hotel and car rental search `http://www.kayak.com`
3. Orbitz travel `http://www.orbitz.com`
4. Phantomjs - a headless webkit scriptable with a javascript api. `http://phantomjs.org`
5. Acar, G., Juarez, M., Nikiforakis, N., Diaz, C., Gürses, S., Piessens, F., Preneel, B.: FPDetective: Dusting the Web for fingerprinters. In: Proceedings of the ACM Conference on Computer and Communications Security (CCS) (2013)

6. Casanova, J.: How i hacked kayak and booked a cheaper flight (2014), http://www.josecasanova.com/blog/how-i-hacked-kayak-and-booked-a-cheaper-flight/
7. Dawson, A.: Save money on ryanair fares by deleting your cookies? (2011), http://www.bitterwallet.com/save-money-on-ryanair-fares-by-deleting-your-cookies/42133
8. Disabatino, J.: Southwest pulls flight data to hinder orbitz (2001), http://edition.cnn.com/2001/TECH/industry/07/11/southwest.pulls.data.idg/
9. Eckersley, P.: How Unique Is Your Browser? In: Proceedings of the 10th Privacy Enhancing Technologies Symposium (PETS). pp. 1–17 (2010)
10. Elliott, C.: The low airfare that vanishes in a click (2010), http://elliott.org/the-navigator/the-low-airfare-that-vanishes-in-a-click/
11. Jonjak, B.: Lan surprises tourists with $178.50 fee (2009), http://archive.peruthisweek.com/features-891-art-culture-lifestyle-lan-surprises-tourists-with-178-50-fee
12. Krishnamurthy, B., Wills, C.: Privacy diffusion on the web: A longitudinal perspective. In: Proceedings of the 18th International Conference on World Wide Web. pp. 541–550. WWW '09, ACM, New York, NY, USA (2009)
13. Matt, N.: How to be a travel hacker (2011), http://www.nomadicmatt.com/travel-tips/travel-hacking-guide/
14. Mattioli, D.: On orbitz, mac users steered to pricier hotels (2012), http://online.wsj.com/article/SB10001424052702304458604577488822667325882.html
15. Mayer, J.R., Mitchell, J.C.: Third-party web tracking: Policy and technology. In: Proceedings of the IEEE Symposium on Security and Privacy. pp. 413–427 (2012)
16. McGee, B.: Do travel deals change based on your browsing history (2013), http://www.usatoday.com/story/travel/columnist/mcgee/2013/04/03/do-travel-deals-change-based-on-your-browsing-history/2021993/
17. Mikians, J., Gyarmati, L., Erramilli, V., Laoutaris, N.: Detecting price and search discrimination on the internet. In: Proceedings of the 11th ACM Workshop on Hot Topics in Networks. pp. 79–84. HotNets-XI (2012)
18. Mikians, J., Gyarmati, L., Erramilli, V., Laoutaris, N.: Crowd-assisted Search for Price Discrimination in E-Commerce: First results. In: Proceedings of the 9th International Conference on emerging Networking EXperiments and Technologies (CoNEXT) (2013)
19. Nikiforakis, N., Kapravelos, A., Joosen, W., Kruegel, C., Piessens, F., Vigna, G.: Cookieless monster: Exploring the ecosystem of web-based device fingerprinting. In: Proceedings of the IEEE Symposium on Security and Privacy. pp. 541–555 (2013)
20. Odlyzko, A.: Privacy, economics, and price discrimination on the internet. In: Proceedings of the 5th International Conference on Electronic Commerce. pp. 355–366. ICEC '03, ACM, New York, NY, USA (2003), http://doi.acm.org/10.1145/948005.948051
21. Popper, H.: Argentina hikes levy on credit card purchases abroad (2013), http://www.reuters.com/article/2013/03/18/argentina-dollar-idUSL1N0CA3OI20130318
22. Roesner, F., Kohno, T., Wetherall, D.: Detecting and defending against third-party tracking on the web. In: NSDI'12: Proceedings of the 9th USENIX conference on Networked Systems Design and Implementation. pp. 12–12. USENIX Association, Berkeley, CA, USA (2012)
23. Sampson, A.: Ryanair exhibit a. (2011), https://twitter.com/sampsonian/status/50199798099357696

24. Vissers, T., Nikiforakis, N., Bielova, N., Joosen, W.: Online airline price discrimination dataset (2014), http://people.cs.kuleuven.be/~thomas.vissers/data/price_discrimination.zip
25. W3C: Tracking preference expression (dnt) (2014), http://www.w3.org/TR/tracking-dnt/

A Airlines

The list of airlines included in our analysis is as follows:

- American Airlines (aa.com)
- Aegean Airlines (aegeanair.com)
- Alaska Airlines (alaskaair.com)
- Alitalia (alitalia.com)
- British Airways (britishairways.com)
- Brussels Airlines (brusselsairlines.com)
- Cathay Pacific (cathaypacific.com)
- Delta (delta.com)
- EasyJet (easyjet.com)
- Emirates (emirates.com)
- HOP! (hop.fr)
- KLM (klm.com)
- LAN (lan.com)
- Lufthansa (lufthansa.com)
- Qantas (qantas.com.au)
- Qatar Airways (qatarairways.com)
- Ryanair (ryanair.com)
- Saudi Airlines (saudiairlines.com)
- Southwest Airlines (southwest.com)
- Turkish Airlines (turkishairlines.com)
- United Airlines (united.com)
- US Airways (usairways.com)
- Virgin Atlantic (virgin-atlantic.com)
- Widerøe (wideroe.no)
- Wizz Air (wizzair.com)