An Empirical Study on Online Price Differentiation

Thomas Hupperich
University of Twente, The Netherlands

Dennis Tatang
Ruhr-University Bochum, Germany

Nicolai Wilkop
Ruhr-University Bochum, Germany

Thorsten Holz
Ruhr-University Bochum, Germany

ABSTRACT

Price differentiation describes a marketing strategy to determine the price of goods on the basis of a potential customer's attributes like location, financial status, possessions, or behavior. Several cases of online price differentiation have been revealed in recent years. For example, different pricing based on a user's location was discovered for online office supply chain stores and there were indications that offers for hotel rooms are priced higher for Apple users compared to Windows users at certain online booking websites. One potential source for relevant distinctive features are *system fingerprints*, i. e., a technique to recognize users' systems by identifying unique attributes such as the source IP address or system configuration. In this paper, we shed light on the ecosystem of pricing at online platforms and aim to detect if and how such platform providers make use of price differentiation based on digital system fingerprints. We designed and implemented an automated price scanner capable of disguising itself as an arbitrary system, leveraging real-world system fingerprints, and searched for price differences related to different features (e. g., user location, language setting, or operating system). This system allows us to explore price differentiation cases and identify those characteristic features of a system that may influence a product's price.

ACM Reference format:
Thomas Hupperich, Dennis Tatang, Nicolai Wilkop, and Thorsten Holz. 2018. An Empirical Study on Online Price Differentiation. In *Proceedings of Eighth ACM Conference on Data and Application Security and Privacy, Tempe, AZ, USA, March 19–21, 2018 (CODASPY '18),* 8 pages.
https://doi.org/10.1145/3176258.3176338

1 INTRODUCTION

Pricing policies of (online) business providers are typically not transparent to customers and are based on parameters that a customer is not aware of. This opens up a number of opportunities for so-called *price differentiation* and *price discrimination*. Price differentiation is a pricing policy in which providers demand different prices for the same asset, including special offers or discounts. In contrast, adjusting a product's price based on a customer's *personal* information (e. g., gender, wealth, home address, or other feature) is called price discrimination. In the past, suspected cases of online price discrimination captured headlines, including different pricing at Staples based on a user's location [14] and indications that offers for hotel rooms are priced higher for Apple users compared to Windows users at Orbitz [20].

From a technical point of view, an online platform can leverage many kinds of techniques to identify a user, which would be the starting point for price discrimination. Generally speaking, the term *fingerprinting* refers to the process of obtaining characteristic attributes of a system and determining attribute values that can be leveraged to recognize or identify a single system among others. In the context of online user tracking, this technique complements cookie-based recognition, which has been ubiquitously deployed for many years [4]. In practice, browser fingerprinting provides more information about a customer compared to cookie-based methods, including software attributes (i.e., the used user-agent, installed plugins, and supported mimetypes [1, 5, 15, 19]). Previous research demonstrated that browser-based system fingerprinting performs well for most types of commodity systems such as desktop computers and mobile devices [4, 11, 24].

Our assumption is that information about a user's system—obtained via browser fingerprinting—is leveraged by online providers for price discrimination as it leaks information about the system configuration and the user himself. While flight tickets have been found to be subject to too many influence factors to be able to identify methodical price discrimination [27], there has been no systematic investigation of the existence of systematic price discrimination in online commerce. In particular, hotel booking websites are often criticized for non-transparent pricing and have been suspected of price differentiation. Unfortunately, not all details about leveraged price differentiation mechanisms can be determined without detailed insight into the inner working of such platforms, and thus we need to adopt a black-box strategy to explore abnormalities.

In this paper, we apply real-world *browser fingerprints* to simulate different systems and analyze corresponding price changes. To achieve this goal, we implemented an automated price scanner capable of disguising itself as an arbitrary system leveraging real-world system fingerprints and searched for price differences related to (i) user location represented by the IP address, (ii) specific systems represented by their fingerprints, and (iii) single features of fingerprints. This enables us to expose the impact of these features on asset prices. Generally speaking, we aim to expose system configuration features that may influence prices and perform a repeatable empirical analysis to measure the effects of fingerprint changes.

In an empirical study, we examined several accommodation booking websites and a rental car provider platform to identify which parameters affect an asset's price. Our results show the existence

CODASPY '18, March 19–21, 2018, Tempe, AZ, USA

ACM ISBN 978-1-4503-5632-9/18/03…$15.00
https://doi.org/10.1145/3176258.3176338

of location-based price differentiation while price changes based on system fingerprints are found in single cases and do not reveal systematic discrimination. We also shed light on how changing single attributes in a system fingerprint affects an asset's price. Associating reproducible price changes with specific attribute values allows users to change their system fingerprint and start hunting for the best prices for hotel rooms.

In summary, we make the following contributions:

- We developed and implemented a method to find and analyze price differentiation by automatically testing different system configurations against online providers.
- We conducted an empirical study to explore price differentiation based on user location and system configuration.
- We provide insights into which specific system features influence pricing strategies and how a user can potentially affect them.

To foster additional research we present several examples of online price discrimination detected by our analysis framework at https://rawgit.com/ananonymousauthor/examples/master/index.html. A more detailed technical report about this research and results is available on arXiv [12].

2 BACKGROUND

First we introduce both *price discrimination* and *system fingerprinting* in more detail and explain why and how both concepts are related to each other.

As noted above, there is a small yet important difference between price discrimination and price differentiation: while price differentiation describes a strategy to determine a product's or service's price based on a potential customer's needs, it does not depend on a customer's characteristics. In price discrimination, however, the price is determined on the basis of a potential customer's *attributes*, such as location, financial status, possessions, gender, or behavior. According to Varian [26], price discrimination is defined as specific pricing for specific groups and has been a common technique since 1920. Traditionally, price discrimination and differentiation can be subdivided into three different degrees [26]:

First degree: Involves individualization of prices for all customers. Second degree: Prices differ based on additional services. It is possible to distinguish between service-related, quantitative, and price-pack forms. Third degree: Involves individual prices for groups of people. They can be individual, location, or time-related.

Online commerce has widely been resistant to price discrimination as customers typically decide to buy a product for the lowest price possible. Furthermore, few customer characteristics were customarily revealed during an online purchase (like residential area) and there are usually no negotiations (at least for standard products). Today, however, a client's computer system reveals more information about its user [1, 4, 11, 24]. This presents new opportunities for online shop operators to personalize their content for each individual customer [16, 18]. From their perspective, price discrimination is a way to maximize their profits and thus they have an incentive to utilize such techniques.

To implement such a strategy, they can use system fingerprinting methods to identify user groups that are likely willing to pay more than other user groups.

Fingerprinting is a technique to obtain characteristic attributes of a given system, enabling the recognition or identification of a single system among others. While this is a general method and can be applied to different kinds of systems, including servers, mobile devices, or websites, we focus in this work on client-side systems, especially browsers on commodity systems like desktop computers and smartphones. This approach enables Web platform providers to fingerprint—and consequently recognize or identify—a user's system and improves on classical cookie-based user tracking to enhance the reliability of tracking techniques [24].

In practice, the attributes of a system are examined and analyzed if they are unique compared to the attributes of other systems. Such characteristic attributes serve as so-called *features* that can be used to create a fingerprint that is as unique as possible. Consequently, every system is assigned a fingerprint which describes the system's characteristic attributes (e. g., configuration items like a browser's settings, display size, or the IP address). As our work is in the context of online shopping, we focus on attributes accessible from the Web and hence use browser attributes as our browser fingerprints. Common browsers reveal adequate information to generate this kind of fingerprint [24], and web-based fingerprinting of personal computers and mobile devices is a common technique that has been investigated by other researchers [4, 11, 17, 24, 28].

3 SEARCHING FOR PRICE DISCRIMINATION

Below, we outline goals, workflow, and functionality of our method for searching the Web for potential cases of price discrimination. For more details of our approach for searching for price discrimination, we kindly refer to our technical report [12].

3.1 Design Goals

We want to conduct a systematic study as well as an objective analysis to clarify the existence of online price discrimination based either on location information or on system configuration. Therefore, we define the following goals for our implementation of systematic, non-offensive scans: (i) fingerprint variety, (ii) simulation of user behavior, (iii) robustness, and (iv) deterministic behavior.

Besides these design goals, we also follow three additional principles. First, as we aim to include multiple platforms in our study, the implementation needs to be modular. For every scan, the platforms, search parameters, fingerprints, etc. can be chosen freely, which also enables us to extend the system with additional scrapers so more websites and product categories may be scanned for fingerprint-based price discrimination in future work. Second, we strive for minimal invasiveness and avoid to produce too many requests to a given website at once. As we certainly do not want to disturb legitimate services, we apply a time delay to our low-traffic implementation and hence ensure that our scans will be tolerable to platform providers and do not interfere with their daily business. Third, we want to be transparent about our work and thus plan to publish the code and data obtained by our scanning practice.

3.2 High-level Overview of Workflow

We begin by providing a high-level overview of the system's workflow. We have two data sources (system fingerprints and provider websites), three data processors (scanner, scraper, and price analysis), and result data (cases of price discrimination).

First, we build system profiles, each including four components: (i) a real-world fingerprint, (ii) a proxy server to be used, (iii) search parameters, such as the dates of arrival and departure for hotels, and (iv) the providers and websites to be examined. Bundles of such profiles are loaded by the scanner.

The scanner's duty is to automatically browse the website of a given provider to end up on certain product result pages. Our scraper implementations then extract the relevant price information from these pages. Finally, we analyze the extracted price information; this analysis of the collected data can point to cases of price discrimination.

In the following sections, we describe each of these steps in more detail and provide information about implementation aspects.

3.3 System Fingerprints

The real-world systems fingerprints that we use for our study are derived from two data sources: First, a previous study [11] providing 385 fingerprints, primarily from mobile devices, and second a project partner that has provided 15,000 fingerprints to a large browser gaming platform.

We re-grouped these fingerprints in order to identify the most and fewest common feature values (see Sec. 3.1). This set of most common and uncommon system fingerprints is suitable for our purpose: we need to include in our study those systems that are frequently found in the wild, but we also need to include special systems with unusual appearances in order to test how such rare fingerprints may influence a product's price. We also reduced the set, since many features' values were identical across several fingerprints. Following this re-grouping and reduction, our set includes a total of 332 real-world fingerprints for scanning Web platforms.

As noted above, a fingerprint may encompass manifold features of a system. However, we include only the following features, **AvailHeight**, **AvalWidth**, **ColorDepth**, **CookieEnabled**, **Height**, **Language**, **Languages**, **MimeTypes**, **PixelDepth**, **Platform**, **Plugins**, **ProductSub**, **UserAgent**, **Vendor**, and **Width**. All features were gathered either from the Browser Object Model (BOM) or the HTTP header, as these have been proven to be common features used for browser fingerprinting [4, 24].

In addition to all of these device-level features, we also need to consider the network location (i. e., IP address), as this represents an important feature for location analyses. We opted to use free proxy servers and rent VPN gateways to enable a flexible routing of requests. As a result, we can issues queries from different network locations and observe changes in responses.

3.4 Scanner and Scraper

Figure 1 depicts the components of our scanner implementation.

The real-world fingerprints, the proxies, the provider websites, and the search parameters serve as input data for the scanner, which uses Selenium to communicate with the custom PhantomJS browser via its extended GhostDriver implementation.

The scraper, in general, extracts product information from selected websites.

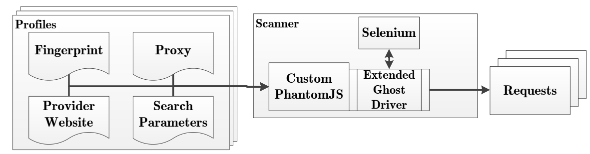

Figure 1: Scanner components operation chart

When extracting price information from a website, one has to handle different price presentation formats, currencies, and the meaning of the displayed prices. Therefore, this data must be converted to a common format for use in subsequent data analysis.

4 EVALUATION

Based on the implementation of the scanning infrastructure, we performed several empirical tests. We focus on two specific types of business: hotel booking platforms and rental car suppliers.

4.1 Price Analyses

We scanned different providers for hotels and rental cars, namely *Booking.com*, *Hotels.com*, *Hrs.com*, *Orbitz.com*, and *Avis.com* and conduct three kinds of analyses: (i) location-based, (ii) fingerprint-based, and (iii) fingerprint-feature-based price differentiation analyses.

First, we investigate location-based price differentiation. We consider several countries (including France, Germany, the United States, Russia, Pakistan, and the Netherlands) to determine how realistic it is that a higher or lower price for the same asset will be obtained when requesting it from a different country. For these countries, we obtained proxy servers or VPN gateways and re-routed our search requests through these servers. The target websites will treat these as search requests coming from the corresponding country. Furthermore, we randomly picked six fingerprints from our set to repeat these scans with different system configurations. Note that we focus in this analysis on hotel providers.

Second, we shed light on price differentiation based on system configurations. This analysis is normalized to France, the United Kingdom, Germany, and the United States because we aim to highlight the systems' fingerprints instead of different originating countries and because we obtained complete result sets for our scans for these countries. While we generally do not consider single fingerprints for location-based analyses, we do so in this step. We used our set of 332 representative system fingerprints for the following analyses and utilized them to disguise our scanner.

Third, these fingerprints are leveraged to create pairs in which one fingerprint yields a high price and the other yields a low price for the same asset with significant frequency. Intermediate fingerprints are then forged, simulating single feature changes. By re-scanning the providers' platforms, we harvest insights on which specific system attributes affect online pricing policies.

Note that we are always searching for one person and one single night in the case of hotel booking websites, hence, the search parameters described in Sec 3.4 are kept constant in the following analyses. After sending a search request, we scrape the top offer prices per hotel for every provider as our ground data for analysis. Finally, we repeat search requests and confirm that using the same configuration reproduces the same prices, so that we can exclude randomness and consider only reproducible price changes.

4.2 Location-based Price Differentiation

We sent search requests for different parameters, e. g., dates of arrival and departure, to all accommodation providers, querying assets in four major cities, namely Los Angeles (USA), London (United Kingdom), Berlin (Germany), and Tokyo (Japan). Each scan lasted about one hour in order to not overwhelm a given site with queries. As a result of these scans, we obtained over 455,500 data records, including an accommodation's name, its provider, and the normalized price in Euro.

Figure 2 shows boxplots for all providers, including the countries we re-routed the search requests through, on the X-axis and the prices in Euro on the Y-axis. Each box depicts the median, quartiles as well as minimum and maximum values of prices for the corresponding country. Note that the prices for each country refer to the same set of hotels in all cities, while there may be differences when comparing providers, as some of them may not cooperate with specific accommodations. This set is used for all location-based analyses and contains only hotels that were found in all single scans for all configurations. We omitted results with fewer than 1,000 responses per provider to avoid bias and keep the results representative; therefore the number of countries varies in Figure 2.

Summary. The result of our price differentiation analysis regarding location is mixed: Not all providers seem to leverage price adjustments based on a user's location. On Orbitz.com, all examined countries were treated the same in our study, giving no indication that this platform performs systematic price differentiation. In contrast, we see for the other accommodation search providers a medium variance of prices for the same assets. The USA received privileged prices at Booking.com and Hotels.com, while the Netherlands and Pakistan were given rather high prices at Booking.com, as was Germany at Hotels.com. At Hrs.com, prices tend to be higher for requests from the Georgian Republic, whereas requests from Germany and Russia likely achieve lower prices. Finally, we can confirm the existence of price adjustment based on a user's location, though prices seem to vary within a limited range only.

4.3 Fingerprint-based Price Differentiation

We scanned the providers mentioned above instrumenting our fingerprint set containing 332 system fingerprints. As a result, we obtained over 4,370,000 data records, including an asset's name, its provider, the used fingerprint, and the normalized price in Euro within about 19 hours total. In this iteration the request country has been set to a fixed parameter, as are the destination and dates of travel. In particular, we tested how much prices vary for every single hotel when the fingerprint of a request changes.

For every product (hotel or car) we obtained two lists: (i) fingerprint(s) which yield a maximum price for this asset, and (ii) fingerprint(s) which achieve a minimum price for it. This results in almost 50,000 cases showing price differences, which is only about 1.12 % of all scanning results.

For Booking.com, we recorded 20,868 cases, representing a share of 0.48 %. Hrs.com and Orbitz.com show almost the same amount of cases with 9,786 and 9,600 both being a share of 0.22 % of all scanning results. Hotels.com produced 9,174 cases, meaning a share of 0.21 %. Finally, for Avis.com, we found 181 cases which are negligible as their share is below 0.01 %. Hence, we see that fingerprint-based pricing is applied to different extents. While we found the majority of suspected price variation based on fingerprints at Booking.com, the other three providers seem to deploy price differentiation at about the same intensity. However, the share of suspicious cases that exhibit a high price variance is rather small compared to the over 4 million scanned prices. We speculate that these are individual cases, as a systematic price differentiation—or even price discrimination—usually has a greater impact and is not limited to a small share of cases.

Building on these initial findings, we perform a statistical significance analysis to further investigate how changing a system's fingerprint affects prices. For this purpose, we conduct the Friedman test [6, 7]. We used the Friedman test because it is a parameter-free alternative to classical analysis of variance (ANOVA). The result of both analysis variants is equivalent. An ANOVA requires data in a normal distribution which we do not have. The Friedman test does not necessarily need it and is therefore suitable for our significance analysis. We assembled nearly 600 hotels and a selection of 130 fingerprints that yield price results for all of the assembled hotels, so that there is a scanned price for every combination of fingerprint, hotel, request country, and provider. The Friedman test calculates the significance of price changes resulting from these fingerprints. By reducing the number of fingerprints to only those which occur in all records of our data gathering, we guarantee the comparability between the various characteristics.

However, before the Friedman test can be performed, additional cleaning of the input data is necessary. Hotels with no free rooms must be removed. This keeps the sample size (number of hotels) identical for each fingerprint, which is important for statistical analysis. Altogether we use a data matrix including the numeric hotel prices of the fingerprints as our input data. Each record has 130 columns for 130 fingerprints and a certain number of lines for hotels. We made sure that the hotels used for comparison occur in all records. Due to proxy availability, we scanned Hotels.com from France, Germany, and Romania, adding the United States for HRS.com and Orbitz.com. Unfortunately, we could not include Booking.com, as we did in the previous tests, since the Web application changed during our research, making scraping hotel prices impossible. In total we conducted eleven Friedman tests—one for each combination of provider and country. In almost all cases, the p-value was lower than 0.05, representing a significant difference between at least two fingerprints in the corresponding subset. Only one test (Hotels.com from Romania) produced a p greater than 0.05, presumably because the median values are all equal. We calculated the median of medians directly for this single case instead of the post-hoc tests we conducted for all other cases. Using a post-hoc test in this case could possibly lead to false positives. Table 1 shows an excerpt of the Friedman test results, showing the median of each fingerprint for all combinations of provider and country. More results of the Friedman test can be found in the appendix of our technical report [12]. Note that only intra-column comparisons

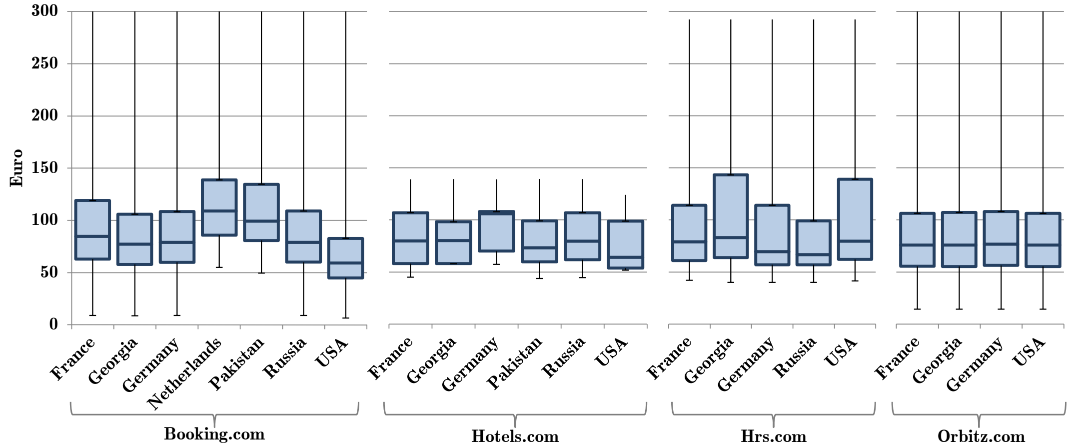

Figure 2: Location-based price discrimination by provider

are allowed as the sample sizes, i.e., the number of hotels, varies between 397 and 594.

In these results, we see isolated price changes for Hotels.com regardless of the requesting country. In fact, only a few fingerprints were found to be disadvantaged. With France as request country, only one fingerprint (FP 171) deviates by €6, while all other fingerprints yield a median price value of €74. For Germany, there are three fingerprints (FP 105, FP 169, and FP 183) which deviate by €5.50 and €8, and for Romania all fingerprints yield the same median price of €74. While these fingerprints resulted in reproducible and significant price changes, the majority of prices remained the same or showed only little variation for all other fingerprints. More significant price variations among fingerprints can be found at HRS.com. Generally, there are many different prices in the median for every request country, which means that the provider's website responded with different prices for different fingerprints. However, almost all of these significant price differences are less than one Euro, so currency conversions cannot be excluded as the cause. Only two fingerprints (FP 35 and FP 95) deviated by about €2.70 and €2.80. Again, these price differences are significant according to the Friedman test, but as such deviations occur only twice, it is questionable whether a price differentiation system exists.

These findings also apply for Orbitz, as there are also many price variances for this site. But again, the differences among the prices is about one Euro or less, and not a single fingerprint delivered a significant price difference of several Euro. In fact, the price differences were found to be significant, but the reasons for these differences may lie in rounding errors rather than being an indication of systematic price discrimination.

4.4 Price-influencing Features

To investigate the individual cases of price changes due to system fingerprints, we dissected those fingerprints that we suspected of price changes in the previous section. Although these are rare and individual cases, we aim to learn which of these features are involved in price changes. We therefore created pairs combining a fingerprint that resulted in a low price with a fingerprint that resulted in a high price. Then we built intermediate fingerprints for all these pairs, so-called *morphprints*, fading from one fingerprint to another by successively changing their attribute values. The morphprints are naturally not real-world fingerprints, they are only intended to compare single feature changes. Combining these morphprints (M_x) with the two original fingerprints (O1, O2) results in a pack of feature changes. This matched-pairs design enables a precise analysis of which feature values influence an asset's price and in what way.

To find the correct order for feature replacement, we applied the *information gain* algorithm, instrumenting the Kullback-Leibler divergence [9], to our data set, revealing every feature's importance to distinguish all data records. It provides an order of how important and descriptive each feature is in relation to our data. We instrument this output to set the order for successive feature value replacement. In total, we created 111 morphprints and re-scanned accommodation websites, resulting in over 14,000 records. These additional scans took about six hours each. To test for reproducibility, every fingerprint and morphprint has been re-scanned twice.

First, we examine which features affect an asset's price most often. Second, we shed light on how these features' values influence online pricing.

Features. While previous research identified a system's user agent string to be the top feature for fingerprinting (see Sec. 6), we see that a system's language is the most frequently occurring price changing feature in our empirical data set. About one third of all discovered cases in our study include a language feature. However, we confirm `navigator.userAgent` to be of particular importance, occurring in about 8 % of all cases in our data set. The screen resolution as well as the property `navigator.vendor` were found to be involved in about 6 % of cases. This indicates that these attributes might only play a minor role in pricing policies. Surprisingly, plugins and mime types are not often involved in price changes, as they occurred in fewer than 4 % of all price changes. Usually these attributes are considered to be highly personalized and should therefore have a greater affect on price customization. This, however, cannot be confirmed on the basis of our data. Table 2 lists each feature's share in price changes.

Table 1: Excerpt of Median Hotel Prices as Result of the Friedman Test

	Hotels			HRS				Orbitz			
FP	Fr	De	Ro	Fr	De	Ro	USA	Fr	De	Ro	USA
1	74	74	74	70	69.9	70	70.2	62.93	62.93	62.93	62.93
3	74	74	74	70	69.9	70	70.2	63.24	63.24	64.19	64.19
5	74	74	74	70.83	70.73	70.83	70.2	63.25	63.25	64.2	64.2
...	...	...	...	...	...	...	...	...	...	...	...
165	74	74	74	70.4	70.24	70.4	70.65	63.24	63.24	64.19	64.19
167	74	74	74	70.34	70.19	70.4	70.41	63.25	63.25	64.2	64.2
169	74	79.5	74	70.53	70.3	70.4	70.41	62.93	62.93	63.87	63.87
171	80	74	74	70	69.9	70	70.2	63.24	63.24	64.19	64.19
173	74	74	74	70	69.9	70	70.2	63.25	63.25	64.2	64.2
175	74	74	74	70.53	70.3	70.4	70.41	62.93	62.93	63.87	63.87
...	...	...	...	...	...	...	...	...	...	...	...
295	74	74	74	70	69.9	70	70.2	62.93	62.93	63.87	63.87
297	74	74	74	70.4	70.24	70.4	70.65	63.24	63.24	64.19	64.19

Table 2: Features share (price change cases)

Feature	**Share**
httpHeader.acceptLanguage	14.57 %
navigator.languages	9.73 %
navigator.language	9.05 %
navigator.userAgent	7.95 %
screen.availHeight	6.90 %
navigator.vendor	6.77 %
screen.height	6.50 %
navigator.platform	6.31 %
screen.availWidth	6.17 %
screen.width	5.37 %
screen.colorDepth	4.63 %
navigator.productSub	4.26 %
screen.pixelDepth	4.04 %
navigator.plugins	3.97 %
navigator.mimeTypes	3.79 %

Feature Values. Given these findings, we now investigate which feature changes result in a price difference. For the following analysis, we only consider reproducible cases with just one single feature changing its value. Due to irregular website responses more than one feature may have changed before scraping these websites, but we eliminated these cases beforehand. Table 3 presents the feature changes, their occurrences, and average price changes.

Summary. Our results show that language settings and user agent strings are the most influential of all features. Changing these features to specific values may increase the chance of receiving a lower price for online hotel bookings. Adjusting other attributes, like vendor and screen resolution, may also affect online pricing policies, but only to a small degree and in specific cases.

Although we cannot make a general claim about how certain feature values should be set to optimize a search for the best price, our results indicate that features which are closer to the user (like language settings, operating system, and browser) have a greater impact when it comes to fingerprint-based pricing policies.

Nevertheless, our findings—especially regarding single features and their values—refer to individual cases in our data set. Although we have shown the statistical significance of these cases, we cannot claim a systematic third-degree price differentiation or price discrimination. Small price changes of a few Eurocent may be related to currency conversions, and price changes of more than one Euro are rare and cannot be proven to be based on system fingerprinting.

5 THREATS TO VALIDITY

Although we handled both the data collection and analysis phases thoroughly, there are limitations and threats to validity.

First, there are various sources that can influence prices, which is why we cannot be completely sure to produce deterministic results with our method. However, in the gathered data the same input parameters, e. g., fingerprint, destination and travel date, produced the same price in all corresponding scans. Hence, we may consider deterministic behaviour concerning our analysis.

Our findings are not omni-valid as we examined only a subset of all available accommodation booking platforms and one rental car provider. Our results and conclusions are in general only valid for our data set, and investigating other providers, product categories, countries, or fingerprints may verify or refute them. However, our data and results derive from realistic search requests and their valid responses, including real-world prices. To foster research on this topic, we plan to publish all data collected during this study.

Our analysis regarding location-based price differentiation sheds light on differences in pricing on a per-country basis determined by the geolocation of IP addresses. Such differences might also exist intra-nationally, i. e., between regions and cities. This type of fine-grained analysis is not within the scope of this work.

Probably the greatest threat to validity are special offers, hidden price boosters or discounts and other secret price-fixing agreements. In a worst case scenario, a discount is offered during only parts of our scan, so that fingerprints which are applied early in the scanning order, for example, would get a lower special offer price than all fingerprints later on receive. To remedy this threat, we applied a filter to catch these cases and to ensure that only nonlinear price changes are taken into account. For instance, if a hotel cost €100 per night for fingerprints 1 to i, but only €80 per night for fingerprints $i + 1$ to n, it is possible that this price change is due to a special offer. In contrast, if a hotel cost €100 per night for fingerprints 1 to i, but €140 per night for fingerprints $i + 1$ to n, we cannot exclude the possibility that the price has risen just because of our scanning, since the first fingerprints simulate a high demand for this asset: the price could have been increased as a reaction, meeting supply and demand. The exceeding of a room quota may be another cause for such artifacts. All these ambiguous cases are omitted in our analyses. However, we cannot guarantee that we caught all potential external influence factors.

Another possible source of distortion may be the hotel providers' booking conditions. During the scraping process, we obtain the price offered at first sight per accommodation regardless of room type and amenities, e. g., breakfast. It is reasonable to assume that this is the best price for an offer as a lower price attracts more customers than would a price for a premium suite including amenities. Hence, we assume that a provider's platform would always list this best price for all search requests. In practice, if a hotel offered

Table 3: Most influencing features value changes

Feature	Old Value	New Value	Occurrence	Change
language	en-US	de	11.87 %	8.88 %
language	ru	de-de	14.16 %	1.27 %
language	ru-RU	en-US	9.32 %	0.83 %
language	en-US	it-IT	8.48 %	0.77 %
language	ko-KR	en-US	9.10 %	0.30 %
language	de	es-ES	4.01 %	0.06 %
navigator.productSub	20030107	None	4.01 %	0.06 %
navigator.userAgent	Android 4.4.2 Android Browser	Windows 7 Firefox	0.10 %	17.33 %
navigator.userAgent	Mac OS X 10.9.4 Safari	iPad OS 7.0.4 Safari	0.18 %	14.69 %
navigator.userAgent	Android 5.0.1 Chrome	iPad OS 8.1 Safari	0.35 %	10.81 %
navigator.userAgent	Android 4.1.2 Android Browser	Linux Iceweasel	0.34 %	10.67 %
navigator.userAgent	Windows 10 Chrome	Android 4.1.2. Android Browser	0.95 %	8.89 %
navigator.userAgent	Android 4.4.2 Chrome	Windows Phone 8.1 IE Mobile	0.26 %	0.06 %
navigator.userAgent	Android 4.4.4 Android Browser	Windows Phone 8.1 IE Mobile	4.01 %	0.06 %
navigator.vendor	Google Inc.	null	13.10 %	0.06 %
screen.availHeight	588	942	4.43 %	0.06 %
screen.availWidth	384	338	2.17 %	0.06 %

For better readability we present only operating system and browser instead of the complete user agent string. The column *Change* represents the average price change in percent.

standard rooms and premium rooms at different prices, and the standard room price is advertised for the first search request, we presume that the prices shown in response to other requests by our scan are also the advertised standard room price. This does not apply to providers of rental cars, as there are fewer car types than there are possible room types. Although there are typically several room types available, it is possible that during a scan, standard rooms are fully booked and only premium suites are offered at a higher price. Such incidents are also detected by our filter described above and excluded from our data set.

Although we normalized the accommodation prices to compensate changes in currency exchange rates, there may be external factors we cannot consider without insider knowledge. For instance, additional transaction fees for providers may differ based on their bank or foreign exchange company.

With respect to our analyses of the ability of single features to increase or decrease a price depending on their specific values, we have analyzed the most striking fingerprints and created artificial morphprints. Due to the huge amount of data, a complete analysis of all possible feature changes considering all possible values in all possible combinations is not feasible. However, our findings are derived from real-world data, though additional feature values may be seen in the wild, meaning that additional value changes may occur, influencing online pricing policies.

In this study, we instrumented browser fingerprints as well as proxy connections/VPN gateways to create profiles. While unlikely, it might be possible for a cross-layer fingerprinting mechanism to discover a profile, e. g., if a user agent shows a Windows machine, but a TTL (Time To Live) value in the IP header analysis reveals a Linux system. Note that our results show clear price variations based on browser fingerprints, regardless of whether or not such a complex mechanism was in place.

Future enhancements could take into account additional providers, as well as more fingerprints, in order to enlarge the data set and gain additional insights. In addition, a longitudinal analysis of possible price differentiation behavior by several providers is another possible direction for future work. Including different product categories also seems promising.

6 RELATED WORK

Several studies have revealed that online price discrimination is a common technique for online shop operators [2, 10, 22, 23, 27].

Hannak et al. recently analyzed several e-business websites which personalize their content. They found that while personalization on e-business websites can provide their users with advantages, aspects such as price customization, for example, can also create disadvantages for those users [10]. Their results provide evidence of price steering and discrimination practices in 9 of 16 analyzed websites. Vissers et al. analyzed price discrimination in online airline tickets. Their results, however, demonstrate that it was not possible to find any evidence for systematic price discrimination on such platforms. This result may be due to the fact that airlines utilize highly volatile pricing algorithms for their tickets [27]. Another empirical study was performed by Mikians et al.; they were among the first to empirically demonstrate the existence of price discrimination [22]. With this knowledge, they started another large-scale crowd-source study and they were able to confirm that there are price differences in e-business based on location [23]. One more recent study by Chen et al. takes a closer look at the algorithmic pricing on Amazon Marketplace [2]. Our work concentrates on price discrimination on hotel booking and car rental websites. In addition, we make use of system fingerprints and analyze which fingerprinting features are the main attributes causing price changes.

Web personalization work continues to improve the quality of Web search requests and their personalized site content [16, 18]. Personalization is important for our work because we analyze the levels on which system fingerprinting methods are used for personalization. To the best of our knowledge, we are the first to extract specific fingerprinting attributes which cause price changes.

Finally, system fingerprinting of clients is a conventional method wielded for user tracking and identification, among other objectives [4, 8, 11, 17, 24, 28]. In this work, we discuss our assumption that client fingerprinting methods are also utilized for price discrimination. The economic fundamentals are extensively discussed by several economists [25, 26].

Iordanou et al. presented a system to detect e-commerce price discrimination [13]. Although the authors faced a similar challenge, they did not inspect fingerprint-based pricing policies explicitly. Additionally, our approach does not require user interaction as we automatically scan provider websites and scrape their contents.

Datta et al. found that user profile information is instrumented for gender discrimination in the context of advertising [3]. Although this indicates the existence of discrimination on the Internet, this study does not include price differentiation.

Melicher et al. have shown that users are uncomfortable especially with invisible methods of user-tracking, such as price discrimination [21]. In contrast, noticeable effects (e. g., advertising) are experienced as tolerable. This shows the importance of secret price differentiation based on user behavior or system fingerprints.

7 CONCLUSION

In this paper, we proposed a method to search for online price differentiation in a systematic way. To this end, we implemented a system capable of disguising itself as different systems based on real-world fingerprints. Utilizing this system, we sent search requests from several locations and systems to four accommodation booking websites and one rental car provider. The returned prices of all found assets (hotel rooms and cars) were examined regarding systematic price differentiation behavior. We ensured that only reproducible cases of online pricing were considered to exclude randomness and external factors.

Despite recent articles about possible price discrimination based on a user's system, we could not prove the existence of such a system for the examined providers. Getting a lower (or higher) price for an asset based on a digital system fingerprint is probably limited to individual cases. Our data show that such cases are rare or may be the result of currency conversions. Nevertheless, it is possible that price differentiation based on other attributes and factors is applied in the wild, such as regional price discrimination.

Furthermore, we investigated single attributes to find which values will provoke a reproducible price change. We found that a user's language settings and user agent (containing information about the operating system and browser) to be the most promising attributes to manipulate when searching for an asset's best price. In contrast to other attributes like screen resolution, these features represent a user's choice and may, therefore, be more frequently instrumented for fingerprint-based price discrimination. Though price discrimination does exist, we found price fluctuations based on changed feature values to be individualized, specific cases. Our study shows that systematic price differentiation is applied by booking providers for locations while system fingerprints do not affect pricing of online accomodation bookings in our setup.

REFERENCES

[1] Gunes Acar, Marc Juarez, Nick Nikiforakis, Claudia Diaz, Seda Gürses, Frank Piessens, and Bart Preneel. 2013. FPDetective: Dusting the Web for fingerprinters. In *ACM Conference on Computer and Communications Security (CCS)*.

[2] Le Chen, Alan Mislove, and Christo Wilson. 2016. An Empirical Analysis of Algorithmic Pricing on Amazon Marketplace. In *World Wide Web Conference (WWW)*. 11. https://doi.org/10.1145/2872427.2883089

[3] Amit Datta, Michael Carl Tschantz, and Anupam Datta. 2015. Automated experiments on ad privacy settings.

[4] Peter Eckersley. 2010. How Unique is Your Web Browser?. In *Proceedings on Privacy Enhancing Technologies (PETS)*. 18.

[5] Christian Eubank, Marcela Melara, Diego Perez-botero, and Arvind Narayanan. 2013. Shining the Floodlights on Mobile Web Tracking – A Privacy Survey. In *Web 2.0 Security & Privacy Conference (W2SP)*.

[6] Milton Friedman. 1937. The use of ranks to avoid the assumption of normality implicit in the analysis of variance. *J. Amer. Statist. Assoc.* 32, 200 (1937), 675–701.

[7] Milton Friedman. 1940. A comparison of alternative tests of significance for the problem of m rankings. *The Annals of Mathematical Statistics* 11, 1 (1940), 86–92.

[8] Gábor György Gulyás, Gergely Acs, and Claude Castelluccia. 2016. Near-Optimal Fingerprinting with Constraints. *Proceedings on Privacy Enhancing Technologies (PoPETs)* 4 (2016), 1–17.

[9] Mark Hall, Eibe Frank, Geoffrey Holmes, Bernhard Pfahringer, Peter Reutemann, and Ian H. Witten. 2009. The WEKA Data Mining Software: An Update. *SIGKDD Explor. Newsl.* 11, 1 (Nov. 2009), 10–18. https://doi.org/10.1145/1656274.1656278

[10] Aniko Hannak, Gary Soeller, David Lazer, Alan Mislove, and Christo Wilson. 2014. Measuring Price Discrimination and Steering on E-commerce Web Sites. In *Internet Measurement Conference (IMC)*. 14. https://doi.org/10.1145/2663716.2663744

[11] Thomas Hupperich, Davide Maiorca, Marc Kührer, Thorsten Holz, and Giorgio Giacinto. 2015. On the Robustness of Mobile Device Fingerprinting: Can Mobile Users Escape Modern Web-Tracking Mechanisms?. In *Anual Computer Security Applications Conference (ACSAC)*.

[12] Thomas Hupperich, Dennis Tatang, Nicolai Wilkop, and Thorsten Holz. 2017. An Empirical Study on Price Differentiation Based on System Fingerprints. *arXiv preprint arXiv:1712.03031* (2017).

[13] Costas Iordanou, Claudio Soriente, Michael Sirivianos, and Nikolaos Laoutaris. 2017. Who is Fiddling with Prices?: Building and Deploying a Watchdog Service for E-commerce. In *Proceedings of the Conference of the ACM Special Interest Group on Data Communication (SIGCOMM '17)*.

[14] Jeremy Singer-Vine Jennifer Valentino-Devries and Ashkan Soltani. 2017. Websites Vary Prices, Deals Based on Users Information. (2017). http://www.wsj.com/articles/SB10001424127887323777204578189391813881534.

[15] Samy Kamkar. 2010. Evercookie – never forget. (2010). http://samy.pl/evercookie/.

[16] Chloe Kliman-Silver, Aniko Hannak, David Lazer, Christo Wilson, and Alan Mislove. 2015. Location, Location, Location: The Impact of Geolocation on Web Search Personalization. In *Internet Measurement Conference (IMC)*.

[17] Andreas Kurtz, Hugo Gascon, Tobias Becker, Konrad Rieck, and Felix C. Freiling. 2016. Fingerprinting Mobile Devices Using Personalized Configurations. *Proceedings on Privacy Enhancing Technologies (PoPETs)* (2016).

[18] Mathias Lecuyer, Riley Spahn, Yannis Spiliopolous, Augustin Chaintreau, Roxana Geambasu, and Daniel Hsu. 2015. Sunlight: Fine-grained targeting detection at scale with statistical confidence. In *ACM CCS*.

[19] Bin Liang, Wei You, Liangkun Liu, Wenchang Shi, and M. Heiderich. 2014. Scriptless Timing Attacks on Web Browser Privacy. In *Annual IEEE/IFIP International Conference on Dependable Systems and Networks (DSN)*.

[20] Dana Mattioli. 2017. On Orbitz, Mac Users Steered to Pricier Hotels. (2017). http://www.wsj.com/articles/SB10001424052702304458604577488822667325882.

[21] William Melicher, Mahmood Sharif, Joshua Tan, Lujo Bauer, Mihai Christodorescu, and Pedro Giovanni Leon. 2016. (Do Not) Track Me Sometimes: Users' Contextual Preferences for Web Tracking.

[22] Jakub Mikians, László Gyarmati, Vijay Erramilli, and Nikolaos Laoutaris. 2012. Detecting Price and Search Discrimination on the Internet. In *ACM Workshop on Hot Topics in Networks*. 6. https://doi.org/10.1145/2390231.2390245

[23] Jakub Mikians, László Gyarmati, Vijay Erramilli, and Nikolaos Laoutaris. 2013. Crowd-assisted Search for Price Discrimination in E-Commerce: First results. *CoRR* abs/1307.4531 (2013). http://arxiv.org/abs/1307.4531

[24] Nick Nikiforakis, Alexandros Kapravelos, Wouter Joosen, Christopher Kruegel, Frank Piessens, and Giovanni Vigna. 2013. Cookieless monster: Exploring the ecosystem of web-based device fingerprinting. In *IEEE S&P*.

[25] Benjamin Reed Shiller et al. 2014. First degree price discrimination using big data. *Presented at The Federal Trade Commission* (2014).

[26] Hal R Varian. 1989. Price discrimination. *Handbook of industrial organization* 1 (1989).

[27] Thomas Vissers, Nick Nikiforakis, Nataliia Bielova, and Wouter Joosen. 2014. Crying wolf? On the price discrimination of online airline tickets. In *Workshop on Hot Topics in Privacy Enhancing Technologies (HotPETs)*.

[28] Ting-Fang Yen, Yinglian Xie, Fang Yu, Roger Peng Yu, and Martin Abadi. 2012. Host Fingerprinting and Tracking on the Web: Privacy and Security Implications.. In *Symposium on Network and Distributed System Security (NDSS)*.