Buy it Now, Track Me Later: Attacking User Privacy via Wi-Fi AP Online Auctions

Steven Su
University of Maryland
stevensu@umd.edu

Erik Rye
University of Maryland
rye@umd.edu

Dave Levin
University of Maryland
dml@umd.edu

Robert Beverly
San Diego State University
rbeverly@sdsu.edu

Abstract

Static and hard-coded layer-two network identifiers are well known to present security vulnerabilities and endanger user privacy. In this work, we introduce a new privacy attack against Wi-Fi access points listed on secondhand marketplaces. Specifically, we demonstrate the ability to *remotely* gather a large quantity of layer-two Wi-Fi identifiers by programmatically querying the eBay marketplace and applying state-of-the-art computer vision techniques to extract IEEE 802.11 BSSIDs from the seller's posted images of the hardware. By leveraging data from a global Wi-Fi Positioning System (WPS) that geolocates BSSIDs, we obtain the physical locations of these devices both pre- and post-sale. In addition to validating the degree to which a seller's location matches the location of the device, we examine cases of device movement—once the device is sold and then subsequently re-used in a new environment. Our work highlights a previously unrecognized privacy vulnerability and suggests, yet again, the strong need to protect layer-two network identifiers.

Keywords

wi-fi, access points, geolocation, optical character recognition

1 Introduction

Online auction websites like eBay have been a popular way for users to buy and sell goods online for three decades. Sellers post pictures and descriptions of their goods, which entice and inform users who may choose to bid on them or directly buy them at a fixed price.

Sellers and potential buyers both have reasonable expectations of privacy. Other than learning about one another's identities and addresses to ship purchased goods, sellers expect that their public location information is limited to what is explicitly shared on the auction website (often just the city and state from which they are shipping). Similarly, buyers can typically expect that nobody learns their location other than the shipping address provided to the seller. Certainly, all users expect that their locations before and after the auction has taken place are not divulged to the public.

In this paper, we show that merely by examining the publicly available photos on auction pages, we are able to successfully track the locations of various network devices commonly sold on eBay: Wi-Fi Access Points (APs). The key insight behind our work is that sellers often include pictures of all sides of the devices they are selling, and APs frequently have their MAC address(es) written on them. Combining novel applications of computer vision to extract these MAC addresses along with recent attacks that geolocate APs from their MAC addresses, our attack is able to track both sellers' and buyers' locations in an automated and longitudinal manner. In addition to traditional APs, our techniques also apply to devices that act as APs as part of their device setup, such as smart-home IoT devices.

Over approximately 5 months, we gathered 788k auctions inclusive of Wi-Fi APs, Wi-Fi-enabled home routers, and a variety of Wi-Fi-enabled IoT devices, e.g., smart electrical plugs. Using modern computer vision techniques, we process Wi-Fi hardware advertisements and extract Basic Service Set Identifiers (BSSIDs)—the MAC addresses APs announce—from the seller's images from 144k auctions.

Then, by leveraging data from a global Wi-Fi Positioning System (WPS) that geolocates BSSIDs, we obtain the physical locations of 13k BSSIDs from devices both pre- and post-sale. In other words, in many scenarios, *our attack exposes both the buyer's and seller's locations*, long before and after the auction has taken place.

We perform additional experiments and data analyses to validate the degree to which a seller's location matches the location of the device. We also present several case studies of the value and danger of being able to track devices: we show, for instance, that some used network hardware has been purchased and ultimately used at U.S. military bases, and that some devices marketed as "new" have in fact been previously used.

Contributions We make the following primary contributions:

(1) Demonstration of the ability to gather Wi-Fi BSSIDs unwittingly shared in online auction images programmatically and at-scale (§3).
(2) Introduction of a novel attack that precisely geolocates the gathered BSSIDs, potentially compromising sellers' and buyers' privacy (§5 & §6).
(3) Examination of several use cases, such as identifying secondhand devices used on military sites, and identifying false claims of "new" devices (§7).
(4) A proof-of-concept demonstration that this attack generalizes to other online marketplaces, such as Craigslist and Facebook Marketplace (§8).
(5) Recommendations for attack remediation (§9).

Proceedings on Privacy Enhancing Technologies 2025(4), 912–925

https://doi.org/10.56553/popets-2025-0164

2 Background and Related Work

2.1 eBay and Online Auctions

eBay is an online auction website and marketplace founded in 1995. eBay allows users to sell new or used items in second-price ascending auctions, or to sell them for a set price (called the "Buy It Now" option). When a buyer wins an auction or purchases an item directly, their identity and address are disclosed to the seller by eBay for the purpose of shipping the purchased goods. eBay does not disclose the buyer's identity to any other user of the website. Sellers' registered identities and the city and state from which they will ship their items are made public on each auction item page so that buyers can make more informed decisions as to their trustworthiness and shipping times.

Several other prior studies have used online marketplaces to glean sensitive information about users. Minkus and Ross studied the eBay feedback system, which allows buyers and sellers to rate the quality of their transactions with each other [28]. These ratings alert future buyers and sellers to the trustworthiness of the individual they are transacting with. However, the implementation of eBay's feedback system divulges information about the buyer's purchase history, including potentially privacy-sensitive items they have purchased. In contrast, we use the photographs users post online, rather than attempting to correlate user activity over time. Additionally, rather than learn about their purchase history, our attack exposes users' precise locations.

Several studies have demonstrated that sensitive information can be obtained by purchasing goods on public auction. Roberts et al. purchased hundreds of cell phones from a police auction website to determine what types of privacy protections were implemented for the users whose (former) property was for sale [31]. They found that in many cases, the users' data was still present on the device and accessible due to poor screen lock passcodes and patterns. Garfinkel et al. purchased hard drives on secondary markets (including eBay) to analyze disk sanitization practices [20], finding that roughly a third of hard drives they purchased had recoverable sensitive information.

To the best of our knowledge, our work is the first to show that it is possible to learn private location information about both buyers and sellers without purchasing any items, based only on the pictures made publicly available on the auction website.

2.2 MAC Addresses & Wi-Fi Router Geolocation

One of the central ideas behind our attack is to extract Media Access Control (MAC) addresses from public auction photos, and use them as a persistent, trackable identifier. We review the relevant details of MAC addresses and MAC address-based tracking here.

MAC addresses are 48-bit identifiers used to indicate the source and destination of link-layer frames; they are used in 802.3 Ethernet, 802.11 Wi-Fi, and Bluetooth. MAC addresses are typically written as 12 hexadecimal characters, with each pair separated by a colon or hyphen, e.g., `a0:2b:ca:92:1c:da` or `10-29-ca-2a-be-2f`. MAC address space is managed by the IEEE, which assigns blocks of 2^{24} MAC addresses in three-byte prefixes called Organizationally Unique Identifiers (OUIs) or MAC Address-Large (MA-L) allocations[1]. The IEEE publishes and regularly updates a list of assigned OUIs [22]. The privacy risks of static MAC addresses are well-known and studied, particularly in 802.11 Wi-Fi clients [10, 12, 14, 16, 18, 19, 21, 24, 25, 27, 35, 36]. Our work diverges from prior art by developing a novel way to learn Wi-Fi MAC addresses—by recovering them from photographs from online auctions and marketplaces. This technique eliminates the requirement for an attacker to be physically proximate to a device to learn its Wi-Fi MAC address.

Network operators commonly use MAC addresses to identify and manage devices on their network. For this reason, it is common practice for manufacturers of networking hardware—especially Wi-Fi APs—to print the device's MAC address(es) on the device itself. This facilitates quickly identifying devices, but as we show in this paper, it also facilitates tracking some devices' precise locations and movements.

A Wi-Fi AP and the devices connected to a network it advertises form a Basic Service Set (BSS); the MAC address the AP uses for that Wi-Fi network is known as the Basic Service Set Identifier (BSSID). Mobile client devices will frequently use random MAC addresses while probing for nearby Wi-Fi networks and after connecting to them [16, 18, 19, 21, 25, 27, 36] for privacy reasons. However, APs' MAC addresses typically do not change, because, until recently [32], they were not considered as privacy-sensitive as client MAC addresses (in large part because most APs are generally stationary for long periods).

Our work relies on geolocation information obtained from a Wi-Fi Positioning System (WPS). WPSes are systems run by operating system vendors and others to enable devices in their ecosystems to self-geolocate without using GPS. Devices do this by scanning for nearby AP BSSIDs, which they report back to the WPS operator. These nearby BSSIDs act as landmarks to help determine their location.

Several works of prior art have geolocated Wi-Fi APs using WPSes. Rye and Beverly carried out an active measurement campaign to obtain IPv6 addresses from home routers [33]. A subset of these embed their MAC address in the lower 64 bits of the IPv6 address, which they then geolocated using Apple's WPS, the open-source project WiGLE [38], and other open-source BSSID geolocation databases. Unlike Rye and Beverly's work, our work requires no active measurements or specific types of IPv6 addresses, but rather derives BSSIDs from photographs users post of their APs online.

Rye and Levin used Apple's WPS to conduct a longitudinal measurement of Wi-Fi APs, which revealed that many move over time [32]. They used their large-scale dataset to detect the positions and movements of troops in and around Ukraine, and destruction and power outages in Gaza. In contrast, we conduct a more targeted privacy attack in this work in which the BSSID of the target is known through its disclosure in photographs posted online.

Our work also explores the extent to which MAC addresses reveal devices' specific models. In 2016, Martin et al. developed a novel method for determining the MAC address allocation schemes by a variety of manufacturers with model-granularity [26]. To do

[1]Smaller allocations are also offered, called MAC Address-Medium (MA-M) and MAC Address-Small (MA-S) that contain 2^{20} and 2^{12} MAC addresses, respectively.

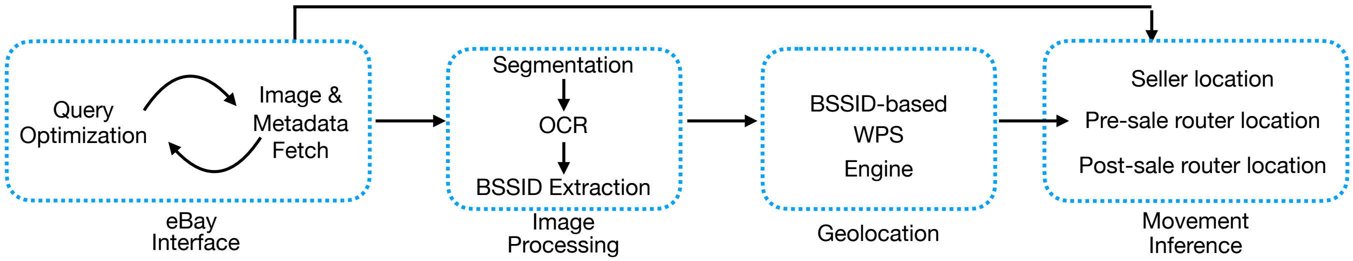


Figure 1: Methodology Overview: Images of Wi-Fi hardware on eBay are processed to find and extract BSSIDs. These BSSID are fused with precision WPS geolocation to create a novel privacy attack.

this, they captured more than 2B 802.11 frames while wardriving, and extracted model-identifying information from Wi-Fi Protected Setup information elements in 802.11 management frames and Multicast DNS (mDNS) responses. These techniques relied on devices using their global, hardware MAC address in order for the MAC address-to-model mappings ranges to have any meaning; today, client devices using modern iOS and Android operating systems use random MAC addresses before and after association to prevent long-term tracking [18]. Further, Wi-Fi Protected Setup suffers from numerous security and privacy vulnerabilities [13, 15, 29, 34, 37], and Wi-Fi Protected Setup information elements are now infrequently used by both clients and access points. In this work, we demonstrate that eBay photographs can be used to derive manufacturer model ranges too, without needing to ever be within transmission range of a device in question.

3 Methodology

To gather BSSIDs and track locations of Wi-Fi APs listed in eBay auctions, we develop a multi-stage pipeline. The pipeline ingests eBay auction data via an eBay-supplied API, downloads images of items for sale, and uses Optical Character Recognition (OCR) to extract addresses (Figure 1). We discuss the major components of this pipeline in the following sections.

3.1 eBay API

Unlike other popular online marketplaces such as Craigslist or Facebook Marketplace, eBay provides an official and documented API for developers to create integrations that augment the experience of buyers and sellers. The API provides a search endpoint intended to aid product discovery in software integrations and permits keyword queries, filters, and other search refinements [6]. Product listing query results returned via the API include a variety of information, including the title, links to images provided by the seller, price, condition, and seller among other data. Search results can be further expanded and refined by a "fieldgroups" API parameter to obtain the seller's coarse physical location, buying options (e.g. auction, "Buy it Now"), filters by brand, and gather item details specific to wireless network hardware, e.g., its highest implemented Wi-Fi standard, data rates supported, etc.

eBay images embed an implied ID, image size and format in their URLs within the path `/images/g/<ID>/s-l<size>.<format>` on the `i.ebayimg.com` domain. Manual experimentation varying the size and format allow us to obtain images between 32 to 2400 pixels horizontally and in JPEG, PNG, or WEBP formats. To balance quality and storage, we request 1600 pixel files in JPEG format. All available images for a given listing were downloaded and processed separately in our pipeline.

Auction listings use an identifier that is unique across all marketplaces and allows us to de-duplicate collected data. This identifier is returned both by the API and used as the listing URL (`www.ebay.com/itm/<ID>?<HTTP GET args>`).

3.1.1 *Exhaustive API Search.*

One notable limitation of the eBay API is a maximum paginated results set of 10,000 results. This limitation also appears in eBay's website search—while a query may report a significantly larger results set (e.g., 71,000+ results for a query of "router" on the United States eBay marketplace), a user can only view a maximum of 10,000 results before artifacts of the limitation start to appear, such as repetition of listings already shown in previous pages. Experimentation with the API reveals that eBay's 10,000 results limit applies to a given query combined with any applied filters, such as the product's brand or features. Thus, by providing unique combinations of filters in addition to the search terms of a query, we can effectively obtain more than the 10,000 results limit for a given search query text alone.

Therefore, by dividing the potential results for a given query by utilizing sets of mutually exclusive filters, we can efficiently and potentially exhaustively enumerate all possible results for a given query. In our system, for a given query, we first send a "probe" request to eBay's API to obtain a listing of possible options for the "brand" filter and the number of listings associated with each filter option. Next, to best utilize limited API calls, we use a bin packing algorithm to group filter options such that the combined listings available for each group fill the maximum 10,000 results (or as close as possible). Then, for each group of brand filter options, we generate sets of API calls that exhaustively query all listings via the API's pagination mechanism and store the data. We refer to this technique in future sections as a full or exhaustive query.

3.1.2 *International Marketplaces.*

While eBay's primary `ebay.com` domain serves its United States marketplace, the platform currently operates in a few dozen international marketplaces. The eBay API allows developers to specify the marketplace ID. To understand how our findings might generalize beyond the U.S. marketplace, we query a subset of these marketplaces by translating search terms and filters to the marketplace's primary language.

We observe diminishing returns in the number of unique auctions as we included additional marketplaces. First, some countries appear to have smaller second-hand markets on eBay for Wi-Fi APs and related devices. Second, it appears that listings are not mutually exclusive between marketplaces. eBay appears to cross-list product listings from nearby marketplaces, (e.g., Canada and United States marketplaces in North America) especially if the listing supports international shipping. Considering these factors and eBay's daily API call quota, we generally only collect eBay listings from the larger marketplaces, such as marketplaces located on their own country's top-level domain rather than a subdomain of eBay's primary .com site or those having a reasonable number of listings for Wi-Fi APs and related products after manual testing.

3.1.3 Search Queries. For each marketplace, we utilize two sets of search queries. Per eBay's documentation, queries are case-insensitive and most punctuation marks are ignored when matching keywords (i.e., no differentiation between `wifi` and `Wi-Fi`). The first set of queries are broad but still reasonably targeted terms to find products likely to contain MAC addresses in the listing images. The specific two queries in this first set, as translated for English-speaking marketplaces, were `wifi router` and `wifi access point`. This first set of "general" queries, `{wifi router, wifi access point}` was used in their language-translated form in all marketplaces.

The second set of search queries consists of brand-specific terms to retrieve listings for Wi-Fi devices manufactured by specific companies. The choice of brands to target was determined by 1) popularity of the brand in a given marketplace as indicated by the counts provided by eBay for the appropriate brand filter or 2) our observation of products that would be expected to appear in eBay's search results using the general queries but for unknown reasons do not (e.g., Verizon ISP routers). These targeted queries are especially useful to obtain listings where the seller did not include identifying terms such as `router` by instead only used the brand and model number in the title or cases where certain brands simply did not appear in search results of the more general queries due to idiosyncrasies in eBay's search algorithms. The design of this second set of queries utilizes eBay's advanced search [4] to more narrowly restrict results. The advanced query language allows clauses in the format `(term1, term2, term3, ...)`. Each of the search terms functions as a logical `OR` in the search, meaning listing titles must contain at least one of the comma-separated terms in the clause to be shown in the search results. Similarly, one may add a dash in front of the clause (i.e., `-(term1, term2, term3, ...)`) to create an exclusion list where the presence of one or more terms stops a listing from being returned.

Furthermore, when specific query filters are applied, eBay's search behavior changes. Specifically, eBay provides "automatic keyword expansion" to simple queries that do not utilize advanced query language. Per their documentation, automatic keyword expansion results in a wider variety of listings and is more forgiving in determining matching listings by, for example, also matching on synonym terms. Therefore, our usage of these advanced filters requires consideration on a balance of breadth of relevant search results and inclusion of irrelevant, "noisy" results. It is important to note that noisy results do not cause false inferences or errors in our pipeline (we simply do not find any relevant MAC addresses). However, we seek to minimize this noise in order to optimize the pipeline and limit our query volume.

With all of the above in consideration, the second set of search queries are those that follow the format: `{brand} (<inclusion terms>) -(<exclusion terms>)`. The specific inclusion terms used were `ap`, `access point`, `router`, `radio`, `wifi`, `wireless`, and `mesh`, while the specific exclusion terms were `car`, `carplay`, `mouse`, `phone`, and `gb`. One query following this format was generated for each brand targeted in each marketplace. The specific brands targeted are not exhaustively listed here for brevity, but include common consumer brands such as Ubiquiti and Netgear, enterprise-grade brands such as Ruckus and Aruba, and ISPs such as Verizon. The second sets of brand-specific queries were used only in English-speaking marketplaces (i.e., US, GB, CA, AU) with the brands tuned for each specific marketplace as previously described.

3.2 Query Schedule

A primary limitation is eBay's quota of 5,000 search API calls per day for a given credential set [5]. While it is possible to request additional API key pairs or to utilize multiple accounts, we aim to ensure efficient usage of API calls to reduce impact on eBay's servers. All marketplaces are queried fully twice a day at staggered 12 hour intervals with marketplaces grouped to spread out the total volume of queries in time. We perform full and exhaustive searches for both query sets (i.e., the general and brand-specific queries) relevant to each marketplace.

Unfortunately, we were not able to incrementally query for new results due to limitations in eBay's interface. eBay's search function and API do not provide a direct sort for listing time; only rough analogues such as "Newly Listed" and "Ending Soonest" listings are available. Manual testing reveals that these sort options do not behave as a naïve sort by time as the reported size of the results set in the site's web interface differs depending on the sort option. Nonetheless, we find that sorting by "Newly Listed" still appears to surface more recent listings and useful results.

Thus, we developed a second search methodology different from the full and exhaustive search. Compared to the methodology described for the full and exhaustive search, we utilize the "Newly Listed" sort option instead of the default "Best Match". Additionally, rather than paginating to the maximum 10,000 results for a given query and filter set, we halt pagination of results when the timestamp of the next listing exceeds the timestamp of the last stored listing in our database prior to initiation of the API querying sequence. Finally, due to the nature of this search method, we omit the initial probe request and partitioning by brands. The lightweight nature of this search method means that parallel to the scheduled twice-a-day full and exhaustive searches for each group of marketplaces, we run this second search methodology at three hour intervals for every query and marketplace combination without regard for the marketplace groupings used in the full and exhaustive searches.

3.3 OCR MAC Address Extraction

In this section, we describe the machine learning components of our pipeline used to automate extraction of MAC addresses from

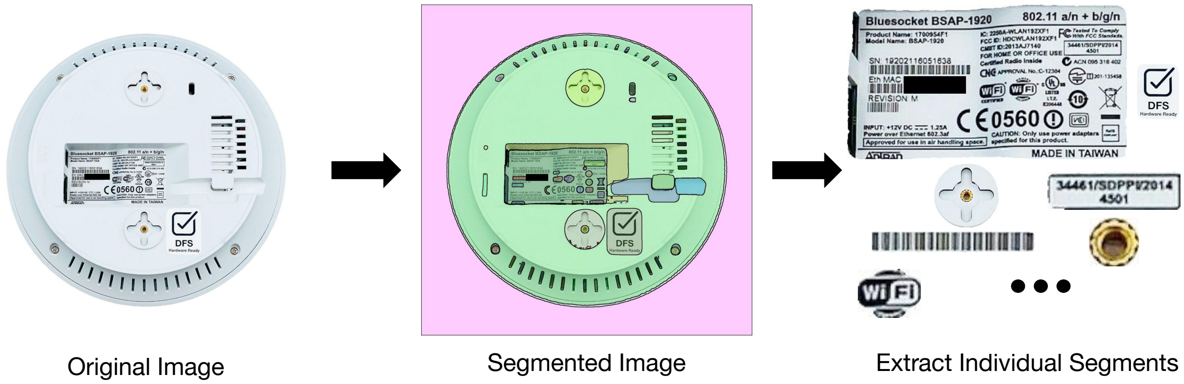


Figure 2: Segment Anything model segmentation masks and segment extraction

the collected eBay listing images and the insights that informed the choice of models. Our pipeline consists broadly of three stages: 1) image segmentation to separate an image into broad regions of interest, 2) optical character recognition (OCR) to extract all text from image segments, and 3) text processing to isolate potential MAC addresses from extracted text.

3.3.1 OCR Model Selection. The final models selected for the pipeline were chosen primarily based on observed accuracy on a small, manually labeled dataset of images and corresponding MAC addresses while also considering throughput and cost. In the initial development of the pipeline, approximately 80 images were chosen from eBay listings of Wi-Fi APs, curated such that all images had at least one MAC address present but with varying perspectives and degrees of text clarity.

We evaluated a number of popular and well-supported OCR libraries and models, as well as multi-modal large language models (LLMs) with zero-shot transfer capabilities. During the testing process, some libraries were eliminated from consideration due to bugs and difficult-to-resolve requirements and dependency issues. Others failed to include a complete pipeline and were also eliminated. Final testing revealed the PaddleOCR library [8] and an 8-bit quantization of the Qwen2-VL-2B-Instruct model of the Qwen family of LLMs [9] to be the best performing in accuracy while also balancing compute requirements on available hardware. For the Qwen model, we observed an 68.75% accuracy on the test set for finding at least one ground truth MAC address in each image compared to the PaddleOCR model which achieved a 57.5% accuracy for the same criteria. The Qwen model, however, was significantly slower in inference time when implemented with the Python transformers [39] library. Therefore, PaddleOCR was chosen as a reliable and performant library both in accuracy and inference time sufficient for use in our pipeline.

3.3.2 Pipeline Optimizations. Having selected an OCR model, we investigated additional augmentations that could boost OCR accuracy. Commonly used pre-processing techniques include background removal, binarization, and the application of filters to improve OCR performance. The below methods were primarily tested against the PaddleOCR library, though we did implement some augmentations with other models to confirm that PaddleOCR remained an optimal choice. We did not exhaustively explore combinations of models and optimization methods to reduce search space of potential pipelines.

We manually tested multiple image binarization algorithms, sharpness and blur filters, denoising algorithms, and other image processing techniques with associated algorithm hyperparameters. No tested combination of techniques and hyperparameters appeared to significantly improve the performance of PaddleOCR; in some cases, accuracy was reduced. We eventually opted out of applying any of these pre-processing techniques due to lack of impact and unknown appropriateness of filters given the wide variety of conditions for the full set of listing images (e.g., perspective, lighting, etc.).

Another method explored was the usage of image segmentation to extract areas of an image likely to contain text. From manual exploration, it was apparent that most MAC addresses on router devices are printed on sticker or labels, usually accompanied by other information such as default passwords, serial numbers, barcodes, and other device information. Processing only these regions with an OCR system could potentially improve accuracy by omitting extraneous background noise. Traditionally, image segmentation requires a task-specific model and we were unable to find any widely-utilized models appropriate for extracting objects such as labels from images. However, given the popularity of foundational models, we found Meta's Segment Anything [23] series of models to be an interesting option. These models, which are available in a variety of sizes (i.e., number of parameters), generalize the task of image segmentation to a more task-independent general case via a promptable segmentation system with zero-shot capabilities. Testing the model on a sample of images revealed strong capabilities to extract labels containing MAC addresses. The behavior and output of this segmentation system is shown in Figure 2.

Using Segment Anything to produce inputs for PaddleOCR resulted in improved MAC address extraction accuracy. Further testing informed the choice of the large, first generation Segment Anything model from an array of model sizes across two generations. In our testing, this model best balanced accuracy and throughput.

We note that the image segmentation stage adds a significant computational burden to the pipeline for an increase in accuracy.

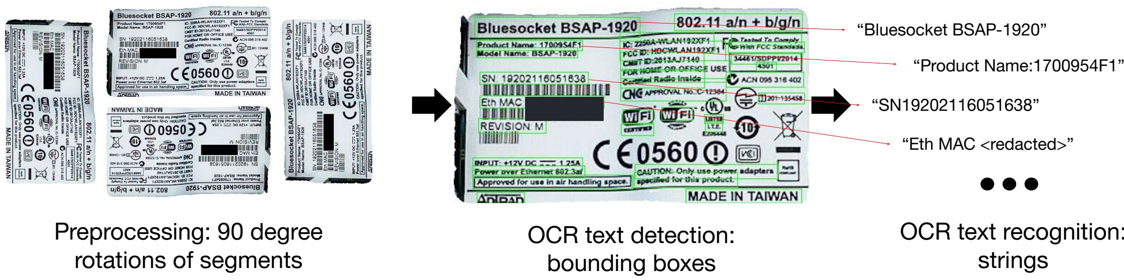


Figure 3: PaddleOCR image processing stages.

Besides the overhead of the model itself, by nature, the segmentation model also returns other segments in the image that are irrelevant to our purposes, such as backgrounds and various physical components of an AP. We are unable to filter out segments reliably without developing a classification model, so all segments are processed through the OCR model adding additional compute time. One other optimization method we found useful was attempting OCR on all 90 degree rotations of each segment. While the PaddleOCR library claims to be able to handle rotated text, our empirical testing revealed accuracy improvements when applying this technique. The stages of the PaddleOCR library with the 90 degree rotations on an extracted segment are shown in Figure 3. In order to process the backlog of historical eBay listings as well as keep up with the ingest rate of new daily listings, we utilized GPU compute nodes our institution's cluster. However, a lower resource attacker could potentially obtain acceptable results at higher throughput by forgoing the additional optimization steps we added to our pipeline. Overall, when combining the segmentation and rotation techniques with PaddleOCR as the OCR library, we observed an 82.5% accuracy on the test set for finding at least one ground truth MAC address in each image and 77.5% for finding all MAC addresses present in each image. We believed the throughput and accuracy of this pipeline to be sufficient to proceed.

3.3.3 Regular Expression Extraction. The output of the OCR pipeline is lines of text and their associated positions in the input image. To extract MAC addresses from this output, we combine all text together and normalize by removing all non-alphanumeric characters and converting to lowercase. We apply a simple regular expression matching the format of a MAC address (with or without separators between octets) to find MAC address candidates. The extraction process is intentionally likely to produce false positives, matching on values such as serial numbers, as the impact of false positives is minimal when querying the WPS database. Candidate MAC addresses are further filtered for addresses whose first three octets form a valid, IEEE-assigned OUI [22].

3.4 MAC Address Extraction Validation

To validate the results of our pipeline and understand the performance of the MAC address extraction, we manually annotated a small subset of images and compared the human-extracted addresses with those identified by the system. To create the annotation subset, all images that were processed by the system were partitioned into three mutually exclusive groups: 1) images that had one or more MAC addresses (with or without valid OUIs) identified by the pipeline, 2) images without any MAC addresses extracted but having at most ten words extracted by OCR, and 3) images without any MAC addresses extracted and having more than ten words identified. The partitions represent 8.8, 33.8, and 57.4% of the entire dataset of images, respectively. For the first partition, we allow images whose extracted MAC addresses have invalid OUIs to account for cases where OCR errors result in the recognition but incorrect transcription of MAC addresses. The purpose of the latter two groups is to identify cases where a MAC address is likely present but not identified by the system. It appeared that only a relatively small percentage of images actually contain a MAC address relative to images that do not and without any intervention, we believed it would be difficult to identify failure cases. The logic behind the partition is that we generally expect images of the backs of APs to have few words while images with many words are expected to be other items such as instruction manuals or the backs of boxes with product information.

A random sample of 250 images were extracted from each partition. The collective 750 images to be annotated were then presented in mixed but consistent order to multiple trained reviewers who were asked to independently identify and transcribe all MAC addresses present in the image. Input was performed on a custom-designed web interface with basic validation on the format of MAC addresses. Reviewers were asked to note special cases such as the presence of MAC address redaction. In the case of significant uncertainty, reviewers were asked to make their best guess at individual characters or use placeholders in the worst case scenario but to nonetheless record a value indicating presence of a MAC address.

In determining ground truth MAC address annotations for each image, for consistency and internal validation, we require annotations from two reviewers and only MAC addresses that show agreement by both reviewers are accepted as ground truths; all other annotated MAC addresses are rejected. In cases where multiple reviewers annotated an image, the annotations from the first two reviewers ordered by internal reviewer ID were selected. Overall, 45 of the 750 images were considered inconclusive after annotation where the pipeline identified MAC addresses but no addresses with sufficient agreement were found in annotations. Eighty-five percent of rejected MAC addresses did not match any addresses identified in the pipeline. The cases of rejected MAC addresses likely result from

Table 1: Performance extracting MAC addresses from eBay images, using human annotators as ground truth on three different image partitions.

	Partition 1	Partition 2	Partition 3	Total
TP	150	0	0	150
FP	60	0	0	60
FN	22	2	2	26
TN	63	245	238	546

annotator data entry errors causing lack of agreement or ambiguous cases (e.g., blurry images or partial obstructions of MAC addresses) where human annotators are able to recognize the presence of but not agreement on the actual MAC address and the pipeline may or may not be able to identify an address at all depending on image quality.

For each image, unique MAC addresses from the pipeline and annotations are classified into four categories. True positives (TP) are counted each time an accepted ground truth matches an address identified from the pipeline. False positives (FP) are counted for each MAC address identified by the pipeline that does not match an accepted ground truth from annotation. Similarly, false negatives (FN) are counted for each accepted ground truth not identified by the pipeline. Finally, a true negative (TN) is counted for an image that had no accepted ground truths (but was not inconclusive as previously discussed) and no MAC addresses as identified by the system. Only filtered MAC addresses (those with valid OUIs) were considered from the pipeline for classification purposes. The overall counts are presented in Table 1. Count totals are higher than the number of images in each partition since each image may have more than one unique MAC address from the pipeline and annotations and each address (or lack thereof) contributes to a category.

The partition 1 counts show a few interesting artifacts from the design of the pipeline. The relatively high true negative count of 63, which may be unexpected for a partition of images supposedly containing MAC addresses, represent images where the pipeline extracted text resembling MAC addresses and hence met the requirements for inclusion in the first partition, but all addresses were filtered out for invalid OUIs and therefore counted as true negatives with the corresponding empty annotations. At the same time, the false positive count of 60 indicates there are many cases where even filtering for invalid OUIs was insufficient due to OCR errors in other locations of a MAC address or that the image quality resulted in sufficient ambiguity that annotators could not agree whether a pipeline extracted address is correct. The false negative count of 22 is similarly indicative of cases where OCR errors resulted in the failure to identify annotator agreed-upon addresses. Partition 1 had the highest number of inconclusive images (12.8% of images in the partition), further showing that a non-trivial portion of images containing MAC addresses were of such poor quality that even human annotators are unable to agree on ground truth addresses.

The counts for partitions 2 and 3 fell within expectations. The majority of results were true negatives which makes sense given these two partitions did not have any MAC addresses extracted by the system in the first place. Two false negative cases were situations where the image was too blurry to extract MAC addresses and human annotators were only able to recognize the presence of MAC addresses from experience but could not reasonable identify the actual characters and left placeholders. The other two false negative cases were images with extreme camera angles that human annotators could read but were missed by the pipeline. These two partitions had 13 images that were inconclusive and mostly consisted of images with MAC address but poor image quality made it difficult for human annotators to agree on ground truth values.

Overall, aggregating across all three partitions, the pipeline exhibited relatively high accuracy of 89% in identifying correct MAC addresses in images with and without MAC addresses. If the counts in each partition are weighted by the partition's representation of the entire dataset, the accuracy rises to 96.3%. As intended, the overall precision of 71% and false positive rate of 10% show the effects of permissive, regular expression-based MAC address extraction from OCR outputs with the help of OUI validation. But, as discussed previously, the impact of extracting MAC addresses that are in reality invalid are minimal as they would simply not show up in a WPS. We believe that these overall statistics show that our system is able to successfully extract MAC addresses from listing images when present with reasonable confidence in the output.

4 Ethical Considerations

Our institution's IRB reviewed our experiment design and determined that it is not human-subjects research. According to our IRB office, this determination was based on three observations: 1) Our experiments involved no interaction or intervention with human subjects, 2) All of the data we collected was public information, and 3) We did not collect information about any humans themselves, but rather about devices, and thus we were not collecting personally identifiable information. While one could certainly use the device location information that we obtain to infer information about individuals (e.g., by looking up the address where the device is located and then using that address to identify the likely owner of the device), we did not do that in any of our experimentation.

Nonetheless, given the potential sensitivities of user locations our techniques are able to infer, we took the following additional steps to respect users and protect their privacy.

We ensured that all of our data was stored on machines that only the researchers had access to. We will make our code and aggregate statistics publicly available, but no individual data—neither the scraped auctions nor the inferred locations. This data management strategy factored into our IRB's determination that the information we collected would not be used to personally identify any of the users whose devices we studied in our experiments.

We followed eBay's security researcher guidelines and did not engage in research that could directly cause damage to eBay users, systems, or applications [3]. We limited the rates at which we crawled and queried their website to minimize any potential harm to users of the site and to eBay's systems. We believe that our research to identify and help mitigate potential security and privacy risks adheres to the principle of beneficence outlined in the Menlo report [11], and outweighs the potential risk to eBay users.

Further, we have initiated the disclosure process to eBay, but as of the time of this writing have yet to receive a reply. We considered informing the individual users whose locations we were able to

Table 2: Top Wi-Fi AP manufacturers derived from eBay AP photographs containing BSSIDs.

# MAC Addresses	%	Manufacturer
31,820	20.0	Netgear
17,768	11.1	Cisco
11,451	7.2	Belkin
11,132	7.0	TP-Link
8,069	5.1	Ubiquiti
79,133	49.7	2,536 other
159,373	100	**Total**

Table 3: The number of eBay auctions per seller country in our corpus.

# Auctions	%	Country
105,489	73.1	US
20,228	14.0	GB
4,684	3.2	DE
3,426	2.4	CN
2,666	1.9	AU
7,864	5.5	142 other countries
144,357	100	**Total**

track, but ultimately decided against this to protect the researchers. Moreover, because eBay redacts expired auction pages, we believe that by the time this work is made public, the affected users' pages will no longer be available. Finally, we propose several mitigations in §9 that we believe will improve the security and privacy of future eBay users.

5 Dataset Statistics

Between September 2024 and January 2025, we retrieved 144k eBay auctions with photographs containing MAC addresses (as identified through our OCR pipeline in §3) from a total population of 788k auctions. Some auctions contained more than one string identified as a MAC address by our OCR pipeline. This resulted in slightly more (159k) unique MAC addresses than auctions, for an average of 1.1 MAC addresses per auction.

Most of the auctions we identified did not result in a sale. To detect whether the auction had a winner or not, we scraped each auction page for the phrase "This listing sold," which appears on listings that resulted in sale. While eBay maintains the auction page for a period of time after an item is sold or delisted, the site eventually removes the auction listing, and at that point, we can no longer determine definitively whether an item was sold. Of the 144k unique auctions we identified, we verified that 8,879 (6.2%) resulted in a sale through the "This listing sold" banner. In §6.3, we consider the subset of items we are certain were sold in our analysis.

Using the MAC addresses extracted from our eBay Wi-Fi AP auction OCR pipeline, we turn next to the task of geolocating these devices using publicly-available geolocation data. First, we present some global and regional statistics about our extracted MAC address corpus. Then, using these extracted MAC addresses, we attempt to geolocate the devices before and after auctions to determine whether eBay listings containing MAC addresses pose a unique threat to user privacy.

5.1 Photograph-Derived MAC Address Corpus

Of the 159k MAC addresses produced by our OCR pipeline, common home Wi-Fi AP brands dominate. Using the IEEE's OUI database [22] to resolve the first three bytes of the MAC addresses we extract from the eBay photos, the most common router manufacturers are Netgear, Belkin, Cisco, Ubiquiti and TP-Link. Table 2 lists the most commonly-observed manufacturers in our corpus.

The eBay API lists the city, country, and partial postal code of the seller so that prospective buyers can estimate shipping to their location. The majority of the auctions that we enumerate through the eBay API are listed by sellers in the US. Table 3 lists the seller countries of the 144k auctions we identified.

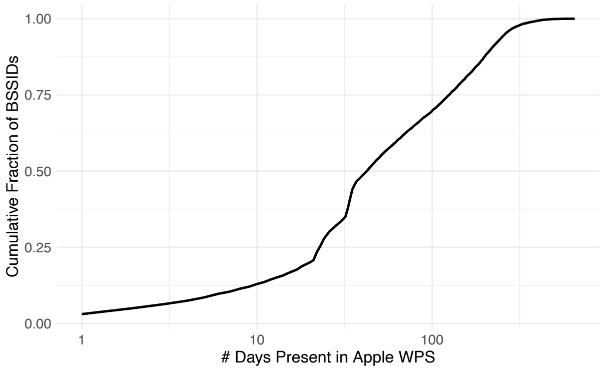
5.2 Extracted BSSID Geolocations

To geolocate MAC addresses we extracted from eBay auction photos, we relied on Apple's WPS. Apple's WPS is closed-source and the precise mechanics of the system are not publicly disclosed; however, Rye and Levin discovered over 2B BSSIDs on all seven continents through the WPS in their study [32]. Apple provides an accuracy value for geolocated BSSIDs that typically ranges between 20-30m, and in testing, we experimentally determined that it takes approximately 30 days for a previously-offline AP to become a landmark.

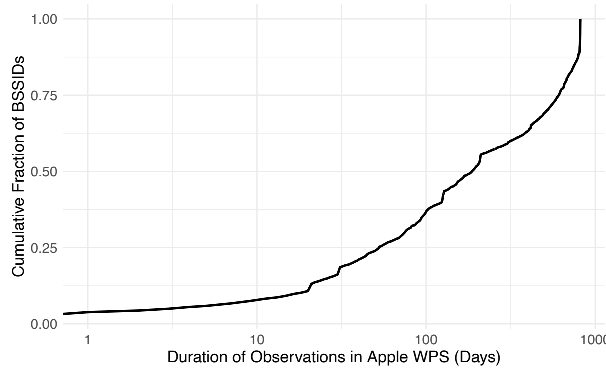
For this work, we used Apple's WPS to look up the 159k MAC addresses we extracted from eBay auction photographs. In the event that Apple's WPS does not have a record for a MAC address that we query (e.g., because it is not a BSSID on an AP or because the AP has been turned off for some time), it returns an invalid latitude and longitude (-180,-180). We continued to look these MAC addresses up every day for a month, from mid-January to mid-February 2025.

In addition, we also looked these MAC addresses up in a longitudinal database of BSSID geolocations dating to late 2022. This longitudinal data provides a historical snapshot of where items from our eBay corpus may have previously been. However, because this database was constructed without knowing the eBay MAC addresses *a priori*, there is no guarantee that eBay MAC addresses are present in the historical data even if they were in-use BSSIDs. Nonetheless, they may have been incidentally obtained and thus provide useful historical location information.

Of the 159k distinct MAC addresses we extracted from photos, 13k (8.3%) had been geolocated at some point by querying the Apple WPS. Figure 4 highlights some overall features of the subset of extracted MAC addresses that have BSSID geolocations in Apple's WPS. In particular, Figure 4(a) depicts the number of days BSSIDs were geolocated in Apple's WPS. The median number of days we observed BSSIDs in the WPS was 43; the top 10% of BSSIDs were observed in Apple's WPS on 217 or more unique days. Figure 4(b) displays cumulative fraction of BSSIDs plotted against the duration of time they were observed in Apple's WPS. That is, for each BSSID, we computed the length of time (in days) between the first and last observations in Apple's WPS. The median timespan geolocated BSSIDs were observed in the WPS was about six months at 187



(a) Number of days observed in Apple's WPS as a cumulative fraction of BSSIDs (x-axis logscale).



(b) Time between the first and last observation in the WPS as a cumulative fraction of BSSIDs (x-axis logscale).

Figure 4: Statistics for BSSID geolocations in Apple's WPS from all eBay auctions.

days. The top 10% of BSSIDs had an observation window of more than 2 years, at 810 days.

Of the 13k BSSIDs we geolocated using Apple's WPS, 1,360 (10%) were geolocated *only after* we began querying the MAC addresses extracted from our eBay photo corpus; they were not present in our longitudinal WPS dataset prior to our extraction of the MAC address from an eBay photograph. This demonstrates that using eBay as a source of Wi-Fi MAC addresses yields a significant number of new MAC addresses, despite having a 2-year longitudinal corpus of BSSID geolocation data.

6 Auction Geolocation Results

A major result of this work is the demonstration that an adversary may, with the aid of machine learning techniques and Apple's WPS, automate determining the precise locations of the buyer and seller in an online auction involving commodity Wi-Fi APs. In this section, we describe the situations in which we are able to identify the parties involved in eBay auctions, and how we validate those geolocations where possible.

Of the 13k WPS-geolocated BSSIDs from our eBay auction photograph corpus, there are three distinct categories these devices may fall into, depending on when they were geolocated relative to the time of their auction. We consider each of these categories in turn.

6.1 Geolocating eBay Sellers

The largest category of geolocated eBay auction BSSIDs is comprised of BSSIDs that were geolocated *only before* the eBay auction was listed. These 5,628 BSSIDs may have been geolocated at any time prior to their listing; our observation window of BSSIDs stretches back to late 2022.

Because these BSSIDs were geolocated *only before* the auction occurred, the geolocation obtained by Apple's WPS is *potentially* the location of the item seller. In some cases, however, an item may have changed owners before being sold at auction, or the owner may have moved locations between last having the Wi-Fi AP online (and thus discoverable via the Apple WPS) and selling it.

To validate that we are, in many cases, discovering the precise location of auction seller, due to their having used the Wi-Fi AP at their home or business, we perform a limited validation using the coarse-grained geolocation information eBay provides through its listing. For the US, this consists of the first three digits of the five digit postal code.

Of the 4,311 geolocated eBay BSSIDs whose auction listed the seller as in the US, we compared the full US postal code derived from the latitude and longitude of the Apple WPS location with the three-digit postal code prefix provided by eBay. Of these, only 4,085 had US postal codes and were thus able to be compared. The rest were geolocated outside the US. We defer a closer examination of some of these items to §7.

In approximately 50% of cases (2,031) the postal code derived from the Apple WPS geolocation matched the three-digit postal code prefix provided by eBay. Further, two-thirds (2,766, 67%) of the postal codes derived from Apple's WPS geolocations matched the first two digits of the eBay-provided postal code prefix, which serves as a coarser-grain match, and may account for moves the seller made since using the item, or selling the item through a service or dealer.

Our technique is not limited to US eBay sellers. Of the auctions in which the BSSID was only geolocated before the listing, 895 BSSIDs were found in photographs from auctions with a UK seller. Of these, 710 had geolocation coordinates from Apple's WPS that indicated they were in the UK; from these coordinates, we derived the UK postal code. When we compared the postal code from our Apple WPS geolocations, 341 (48%) of the WPS-derived postal codes matched the postal code provided by the eBay API.

This demonstrates that in many cases eBay photographs can, via Apple's WPS, leak fine-grained geolocation information about the seller of the item.

6.2 Geolocating eBay Buyers

In our corpus of photograph-derived BSSIDs, 3,123 were geolocated in Apple's WPS *only after* the eBay auction was listed. In these instances, we are likely geolocating the buyer of an item if it was sold on eBay, or the new owner of the item if it was delisted from eBay and sold or donated through another medium.

Unlike with sellers, eBay provides no location metadata about the *buyer* of an item, and it is unaware of the locations of new

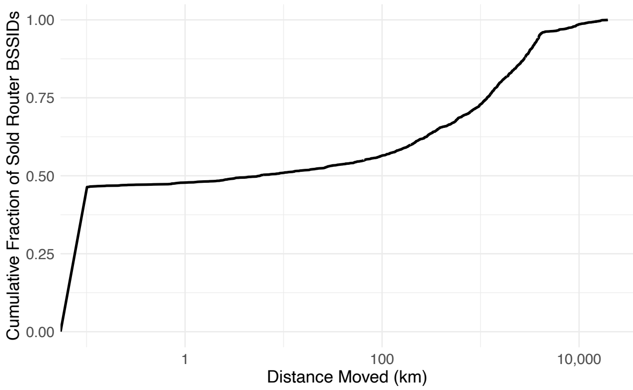


Figure 5: Distance moved by Wi-Fi AP BSSIDs from auctions in which the AP sold (x-axis logscale).

owners who may have obtained the item outside of its platform. As such, we have nothing to compare our post-auction geolocations to, other than to the seller's location.

Intuitively, items sold on eBay are relatively unlikely to be sold to a buyer in the same postal code as the seller, and the fraction of the Apple WPS geolocations of eBay-derived postal codes for items geolocated *after* the auction should be much lower than the geolocations of items *before* the auction was listed.

In contrast to the items geolocated in Apple's WPS prior to their listing on eBay, BSSIDs on items geolocated *after* their listing on eBay were rarely in the same postal code as the seller. When we considered the three-digit postal code prefix eBay reports for US sellers, only 139 of 2,222 (6.3%) BSSIDs geolocated to the US via Apple's WPS matched the eBay-listed postal code prefix. This rate of WPS-derived BSSID postal code matching is approximately eight times less than the rate at which postal codes matched devices geolocated *before* the auction started.

This phenomenon is not limited to the US, either. Among auctions with a BSSID geolocated to the UK (403 total), only 17 (4.2%) BSSIDs geolocated to the same postal code listed by eBay. This validates our assumption that when devices are geolocatable only after they appear in an eBay auction, they have likely been sold (whether on eBay or not), and are frequently sold outside of the seller's postal code.

6.3 Geolocating eBay Buyers *and* Sellers

The worst outcome from a privacy perspective is that eBay photographs for an auction listing leak both the seller's location, from past geolocations, *as well as* the buyer's location after it is purchased and installed at their location. We attempt quantify this phenomenon by identifying the eBay auctions we learned about that resulted in the sale of the device.

Of the 8,879 auctions that we determined were sold (§5) 458 (5.2%) auctions contained 439 distinct BSSIDs that were geolocatable in Apple's WPS both before and after the auction occurred. In order to validate that these devices were indeed sold (and thus likely moved), we calculate the distance between the first and last geolocations for each of these BSSIDs.

Figure 5 plots the distances moved by cumulative fraction of sold AP BSSIDs. The plot is bimodal; there is a first mode between 0 and 1 kilometers moved, indicating that the AP, which was ostensibly sold on eBay, did not actually move locations at all. While initially surprising, we discovered two reasons for this phenomenon. First, some auctions list photos of a device that is not actually the one sold. For instance, many bulk sales of Wi-Fi APs may list a close-up photograph of a single exemplar device, while many individual devices may be sold under the same listing. Second, some vendors appear to reuse MAC addresses. Thus, while one device with a specific BSSID may be online and appear stationary in Apple's WPS before and after the auction, there is another device with an identical BSSID being sold during the same time period.

Roughly half of the BSSIDs we extract from eBay photos do not move significantly. The remainder move nontrivial distances between their initial and final observations in our WPS corpus. This demonstrates that in hundreds of cases, we are able to geolocate both the seller and buyer of eBay APs via BSSIDs extracted from the auction photos.

7 Case Studies

In this section we detail several case studies that highlight the additional information that can be gleaned from eBay auctions that include visible MAC addresses in their photos.

7.1 Devices from Conflict Areas

The eBay listings we gathered included some devices that had recently been in conflict zones. Our listings include two Kasa Smart Plugs, Wi-Fi-enabled electrical plugs that can be turned on and off through the companion smartphone application [7], that were present in Gaza at the time of the October 7th, 2023 Hamas attacks and subsequent Israel-Hamas War. While the smart plugs were in Gaza, they were approximately 10 kilometers apart.

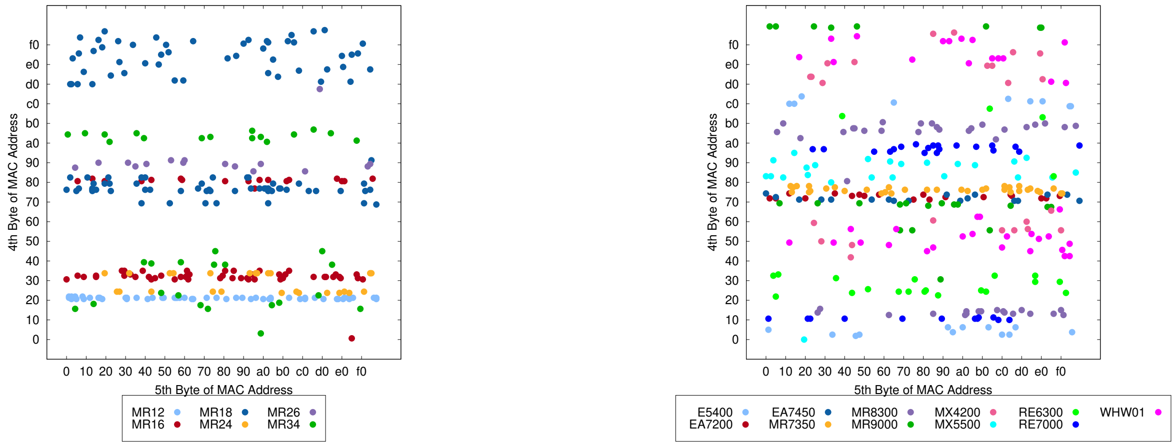
On eBay, these devices were listed by two different accounts on opposite sides of the US. How these devices came to be listed by these two sellers is unknown, as both of the devices are stationary within Gaza and have no other geolocations in the WPS.

7.2 Auctioned Devices in Sensitive Locations

Some of the eBay auction photographs we collected contained BSSIDs that were later geolocated at particularly sensitive locations, such as military installations. Using US military base location data [30], we determined whether any of the geolocated BSSIDs derived from eBay photos were within their boundaries.

Thirteen of the BSSIDs from eBay photos that we geolocated with Apple's WPS were on US military bases. Of these, 5 were geolocated *after* the eBay auction occurred, indicating that they were likely purchased by residents or organizations located on the military base.

This case study presents a potential attack vector for attackers—sell vulnerable Wi-Fi access points in online forums, such as eBay, and wait for them to appear as landmarks in a WPS. If they are at a potentially sensitive location, such as a military base, university, or hospital, the attacker then knows about a vulnerable network device that can be used to gain a foothold in the buyer's network.



(a) Model observations from a Cisco Meraki OUI (`00:18:0a`).

(b) Model observations from a Belkin OUI (`e8:9f:80`).

Figure 6: Plots of device models for two different manufacturer's OUIs. Different color points indicate different model types.

7.3 Used Items Sold as New

We investigated whether eBay sellers ever miscategorize the condition of their auctioned goods by indicating that devices are new when they have in fact been used previously. eBay has different categories for different types of items (e.g., clothes have a "new with tags" category, while computer equipment does not); we consider both "new" and "open box"—which eBay describes as "excellent, new condition with no wear" [17]—labels and check our longitudinal WPS data for evidence of the device operating before the auction took place.

Our data shows that previously-used Wi-Fi APs are regularly mischaracterized as "new". Of the 13k eBay photo-derived BSSIDs, 1,089 (8.1%) were observed in our Apple WPS corpus *before* the auction was listed, and the auction described the Wi-Fi AP's device type as "new" or "open box". The majority (684, 63% of the mischaracterized BSSIDs) of these mislabeled auctions listed their device as "open box", rather than an unqualified "new". We speculate that the sellers may have chosen "open box", rather than "new", to provide a plausible explanation for any wear or damage to the devices incurred while it was actually in operation.

As an example, in late 2024 a seller from the US listed a lot of eight Aruba enterprise wireless APs with the condition "open box." However, a MAC address displayed in one of the photographs (a MAC address on only one of the eight routers was visible) was geolocated to a university campus in the Middle East via Apple's WPS for most of 2023. This seller's auction feedback, which is publicly visible, lists several negative comments from eBay users complaining that the seller auctions used electronics equipment that they have salvaged.

7.4 Granular Model Inferences

In our final case study, we demonstrate another attack that our corpus has made possible: inferring the specific model of devices from knowing only their MAC address. Such *model inference attacks* can be valuable to attackers in scenarios where the attacker is able to learn a target's MAC address, and wishes to use a model-specific exploit. As discussed in §2, prior work on model inference required communicating directly with devices using obsolescent protocols. Here, we show that Martin et al.'s results [26] can be replicated remotely by extracting MAC addresses using OCR.

We emphasize that, for devices listed on eBay, this attack would have little utility, as the specific model could be inferred from the description or photo. Rather, this attack is for scenarios where the attacker learns a target's MAC address through some other means (e.g., inclusion in an IPv6 EUI-64 address).

In addition to any MAC addresses extracted as part of our OCR pipeline described in §3, we also extract manufacturer and model information, when present, from text on the listing title. When both pieces of information are present, we create a linkage between the MAC address, manufacturer, and model listed on the device.

Figure 6 depicts the model observations across two OUIs from two different manufacturers, and highlights vendor MAC address allocation strategy differences between them. Figure 6(a) plots the fourth and fifth bytes of the MAC addresses of devices with the `00:18:0a` OUI; each color represents a different model as indicated by the legend. Most models have distinct banding patterns, as evidenced by points of the same color being grouped in layers across fourth-byte values. Figure 6(b) also exhibits banding, albeit with more than twice as many models observed within the OUI, and with tighter bands.

8 Limitations

While the preceding results demonstrate the real-world feasibility and broad applicability of our attack, there are several current and future potential limitations that bear discussion.

First, a potential source of error lies in the fallibility of the OCR MAC address extraction pipeline. When a MAC address is present in the user-supplied photograph, the pipeline may fail to detect its presence, or may infer an incorrect address, for one of several reasons. First, the image may be blurry, contain the text of the MAC address at an indirect angle causing occlusion or distortions, or have a resolution insufficient to distinguish individual characters.

For instance, the valid hexadecimal characters "B" and "8" are often difficult—even for a human—to differentiate in low resolution photos or when the image is taken at an angle. In other cases, the MAC address may be truncated due to the framing of the photo or otherwise obscured. Last, it can sometimes be difficult to differentiate MAC addresses from other numeric identifiers such as serial numbers due to the variety of representation formats manufacturers employ, e.g., use of a colon, dash, space, or no separator at all to mark digits, octets or pairs. To quantify the extent to which such errors affect our pipeline in practice, we validate our pipeline's MAC address extraction results against human annotation in §3.4.

A second potential confounding factor involves cases in which the equipment sold is not the same as the equipment depicted in the auction image. For instance, some eBay auctions are lots—an auction in which many of the same item is for sale. The image depicted in a lot listing may thus contain a valid MAC address, but for a single unit of the multiple-item lot. Similarly, it is possible that for an auction of a single unit, the seller may include a picture of a device different from the one that they ship to the eventual buyer, for instance by using a "stock" image.

Naturally, our attack depends on the presence of MAC addresses in the auction listing image. We are unaware of a legal or regulatory requirement for addresses to be printed on a wireless access point or its packaging, and presumably some manufacturers omit this information. In this case, our OCR MAC address extraction pipeline will fail to produce a valid MAC address because none exists. Alternatively, it is possible that the seller of the device may neglect to take a picture of that part of the device in the event that a MAC address *is* listed. For some rare cases, we have observed sellers redacting MAC addresses in their listing photos along with other sensitive information. In either of these cases, it is impossible for our pipeline to extract the correct MAC address for an auction—indeed, we advocate for automatically redacting addresses as a potential mitigation mechanism in §9.

A fourth source of potential error stems broadly from correlation with the WPS data. A MAC address may not be present in Apple's WPS database if it is unstable, moving, or short-lived (and, hence, does not serve as a valuable landmark for their service). Further, Apple's data is dynamic and even if the MAC address were present at one point in time, it may not be within the data we query or the available historical data. We utilize historic data and quantify the extent to which we successfully find addresses in Apple's database in §5.2. Of course, the correlation attack depends trivially on the availability of the WPS database itself. Should Apple elect to restrict access to, or protect the privacy of, the content in the database—as recommended by [32]—our location tracking will cease to be effective.

Related to the prior limitation, the attack depends on persistent MAC addresses. While most client devices employ MAC randomization [18], this behavior has not been widely adopted in Wi-Fi APs. An AP could choose random short-lived BSSIDs when sending beacon frames advertising their SSID, yet, in practice, few do. However, as awareness of the sensitivity of MAC addresses with respect to privacy and tracking increases, we may reasonably expect vendors to adopt randomization.

8.1 Generalizability

This study focused on the eBay marketplace due to its available and well-documented API. To understand how generalizable our attack is to other platforms, we performed a small-scale experiment to examine the efficacy of our MAC address extraction pipeline on Craigslist and Facebook Marketplace.

Both Craigslist and Facebook Marketplace localize their results to the user's location with the intention of facilitating in-person sales and transactions. For consistency, we localized results near our academic institution on both sites. We used the search term "wifi router"—one of the general search queries we used in our eBay data collection—and downloaded all available images for approximately the top 100 listings from each website. For Facebook Marketplace, which permits sellers to upload a video, we retrieved the video thumbnail.

On Craigslist, our OCR-based extraction identified 17 of 113 listings (15%) as containing images with MAC addresses. Note that Craigslist permits listings without photos; 13 of the 113 listings (11.5%) had no associated images. On Facebook Marketplace, our OCR extraction identified 29 of 135 listings (21%) with images containing MAC addresses. Manual inspection of the extracted MAC addresses from both platforms indicated an accuracy comparable to our more in-depth validation performed with the eBay images in §3.4.

These MAC address extraction percentages (15% and 21%) are roughly in line with our eBay MAC address extraction rate, in which approximately 18% ($\sim\frac{188{,}000}{788{,}000}$) of listings had images with MAC addresses present. The lower hit rate of Craigslist is in part attributable to the fact that listing images are optional.

Similar to eBay auctions, we also observed some Craigslist and Facebook Marketplace listings that contained multiple MAC addresses. On Craigslist, 19 unique MAC addresses were extracted from 17 listings; on Facebook Marketplace, 43 MAC addresses were extracted from 29 listings.

The ethical considerations of this validation study largely mirror those described in §4. In addition, we collected our limited validation photos from both sites by manually browsing to retrieve listing images, rather than programatically scraping or spidering the sites. While Craigslist's Terms of Use [1] prohibit copying of Craigslist data "by hand", our small-scale validation study is indistinguishable from normal browsing. Finally, we have initiated the responsible disclosure process with Craigslist and Facebook to inform them of our findings.

In sum, while our study exclusively focuses on eBay, our preliminary results show that when users list Wi-Fi APs for sale, they frequently include photos containing MAC addresses irrespective of platform. Further, our experimentation with Craigslist and Facebook Marketplace demonstrate that that our OCR pipeline is able to extract these MAC addresses, as well.

9 Recommendations

We propose several solutions that will mitigate the threat posed by posting photos online of Wi-Fi AP device identifiers.

Policy Modifications eBay and other online forums can advise users *not* to post photographs of sensitive device information. At the time of this writing, eBay encourages sellers to post many, detailed photographs of their item for sale [2], without warning them that particular information might compromise their location privacy. A warning explaining the privacy threat posed by uploading images with sensitive device data may deter users from doing so.

Among our three recommendations, this is the easiest to deploy—it would only involve eBay changing some text on their website—but likely the least effective, because it relies completely on users taking the appropriate action.

Image Manipulations If users do upload images containing device identifiers, eBay should manipulate those photos to obscure sensitive information, such as the BSSID or other MAC addresses, Service Set Identifiers (SSIDs), and passwords, that are commonly printed on the backs of Wi-Fi APs. Though used for different purposes, mapping websites like Google Maps blur sensitive information like license plates and peoples' faces. eBay and other online forums should employ a similar method of redacting sensitive information using an OCR pipeline similar to the one we employed; after identifying areas that contain sensitive information, a blur or opaque box could obscure the data below.

Compared to our other recommendations, image manipulations would be moderately difficult to deploy, as it requires deploying additional code at a site like eBay. Additionally, the computational burden of processing the volume of images eBay's platform receives through a vision pipeline similar to the one we develop in this paper may present a challenge. However, further refinements for efficiency such as pre-filtering of images with less intensive methods can alleviate this barrier. It would be more effective than policy changes, but would be limited by the OCR pipeline's accuracy—which, while effective in the typical conditions we explored (clear images of a single AP) may be less effective in conditions we did not test for (e.g., if the image contained multiple APs). One of the benefits of image manipulations is that it would apply also to legacy APs that cannot deploy their own defenses.

Wi-Fi AP Protections Wi-Fi AP manufacturers can protect users from location privacy threats by refraining from using the BSSID printed on the device. Rye and Levin [32] recommend the use of *random BSSIDs* to prevent longitudinal tracking via a WPS; using a random MAC address prevents targeted localization of a particular device through a known BSSID extracted from an eBay photo.

Among our three recommendations, AP-based protections are by far the most effective, as they preclude the possibility of attributing any given MAC address to any picture of hardware. However, it is also by far the most difficult to deploy; it would likely require changes to the wireless specifications to ensure that AP manufacturers adopt MAC randomization. Worse yet, it is unclear whether MAC randomization is even desirable amongst the entire networking community; network administrators commonly use static MAC addresses to assist them in managing and tracking the devices on their network. Future work is necessary to find the balance between privacy preservation and facilitated network management.

10 Conclusion

In this work, we demonstrated that users' reasonable expectations of location privacy before and after an online auction can be violated merely through the MAC addresses included in the auction's public photos. Our primary contributions include a data analysis pipeline for identifying and extracting MAC addresses from photos in an automated manner, and then cross-referencing these MAC addresses against a WPS dataset known to expose device locations. We were able to extract over 144k MAC addresses, identify the locations of 5,628 sellers, 3,123 buyers, and an additional 458 buyer-seller pairs. Our validation experiments show the efficacy of our techniques, and our case studies demonstrate some of the potential severity of the attack, such as the presence of secondhand network devices being used on US military bases.

In sum, our work shows that a seemingly innocuous photo of a MAC address displayed on the side of a networking device can have surprisingly large privacy implications. We provide several recommendations for users, auction services, and network device manufacturers to take to mitigate this attack.

Acknowledgments

We are grateful for the anonymous reviewers and shepherd for their valuable feedback and suggestions during the revision process. This work was supported in part by NSF grants CNS-1943420 and CNS-2323193.

References

[1] 2025. Craigslist – Terms of Use. https://www.craigslist.org/about/terms.of.use/en.
[2] 2025. eBay – Adding pictures to your listing. https://www.ebay.com/help/selling/listings/adding-pictures-listings?id=4148.
[3] 2025. eBay – Security Researchers. https://pages.ebay.com/securitycenter/security_researchers.html.
[4] 2025. eBay Advanced Search Documentation. https://web.archive.org/web/20241210060542/https://developer.ebay.com/api-docs/user-guides/static/finding-user-guide/finding-searching-by-keywords.html.
[5] 2025. eBay Call Limit Documentation. https://developer.ebay.com/develop/get-started/api-call-limits.
[6] 2025. eBay search API Documentation. https://developer.ebay.com/api-docs/buy/browse/resources/item_summary/methods/search.
[7] 2025. Kasa: Creating a true smart home experience all in one app. https://www.kasasmart.com.
[8] 2025. PaddleOCR. https://github.com/PaddlePaddle/PaddleOCR.
[9] 2025. Qwen-VL Github. https://github.com/QwenLM/Qwen-VL.
[10] Naeim Abedi, Ashish Bhaskar, and Edward Chung. 2013. Bluetooth and Wi-Fi MAC Address Based Crowd Collection and Monitoring: Benefits, Challenges and Enhancement. In *Australasian Transport Research Forum*.
[11] Michael Bailey, David Dittrich, Erin Kenneally, and Doug Maughan. 2012. The Menlo Report. *IEEE Symposium on Security and Privacy* (2012).
[12] Johannes K Becker, David Li, and David Starobinski. 2019. Tracking Anonymized Bluetooth Devices. *Privacy Enhancing Technologies Symposium (PETS)* (2019).
[13] Dominique Bongard. 2014. Offline Bruteforce Attack on WiFi Protected Setup. *Passwordscon* (2014).
[14] Guillaume Celosia and Mathieu Cunche. 2020. Discontinued Privacy: Personal Data Leaks in Apple Bluetooth-Low-Energy Continuity Protocols. *Privacy Enhancing Technologies Symposium (PETS)* (2020).
[15] Daniel Costantin, Krishnun Sansurooah, and Patricia AH Williams. 2017. Vulnerabilities Associated with Wi-fi Protected Setup in a Medical Environment. In *Australasian Computer Science Week Multiconference*.
[16] Mathieu Cunche. 2014. I Know Your MAC Address: Targeted Tracking of Individuals using Wi-Fi. *Computer Virology and Hacking Techniques* (2014).
[17] eBay. 2025. Item Condition by Category. https://www.ebay.com/help/selling/listings/creating-managing-listings/item-conditions-category.
[18] Ellis Fenske, Dane Brown, Jeremy Martin, Travis Mayberry, Peter Ryan, and Erik C Rye. 2021. Three Years Later: A Study of MAC Address Randomization In Mobile Devices And When It Succeeds. *Privacy Enhancing Technologies Symposium (PETS)*.
[19] Julien Freudiger. 2015. How Talkative is Your Mobile Device? An Experimental Study of Wi-Fi Probe Requests. In *ACM Conference on Security and Privacy in Wireless and Mobile Networks (WiSec)*.
[20] Simson L Garfinkel and Abhi Shelat. 2003. Remembrance of Data Passed: A Study of Disk Sanitization Practices. *IEEE Symposium on Security and Privacy* (2003).
[21] Dan Goodin. 2013. No, this isn't a scene from Minority Report. This trash can is stalking you. *Ars Technica* (2013).

[22] IEEE. 2025. MAC Address Block Large (MA-L). https://standards-oui.ieee.org/oui/oui.txt.

[23] Alexander Kirillov, Eric Mintun, Nikhila Ravi, Hanzi Mao, Chloe Rolland, Laura Gustafson, Tete Xiao, Spencer Whitehead, Alexander C Berg, Wan-Yen Lo, et al. 2023. Segment Anything. In *IEEE Conference on Computer Vision*.

[24] Jeremy Martin, Douglas Alpuche, Kristina Bodeman, Lamont Brown, Ellis Fenske, Lucas Foppe, Travis Mayberry, Erik C Rye, Brandon Sipes, and Sam Teplov. 2019. Handoff All Your Privacy: A Review of Apple's Bluetooth Low Energy Continuity Protocol. *Privacy Enhancing Technologies Symposium (PETS)* (2019).

[25] Jeremy Martin, Travis Mayberry, Collin Donahue, Lucas Foppe, Lamont Brown, Chadwick Riggins, Erik C Rye, and Dane Brown. 2017. A Study of MAC Address Randomization in Mobile Devices and When it Fails. *Privacy Enhancing Technologies Symposium (PETS)*.

[26] Jeremy Martin, Erik Rye, and Robert Beverly. 2016. Decomposition of MAC Address Structure for Granular Device Inference. In *Annual Computer Security Applications Conference (ACSAC)*.

[27] Célestin Matte, Mathieu Cunche, Franck Rousseau, and Mathy Vanhoef. 2016. Defeating MAC Address Randomization through Timing Attacks. In *ACM Conference on Security and Privacy in Wireless and Mobile Networks (WiSec)*.

[28] Tehila Minkus and Keith W Ross. 2014. I Know What You're Buying: Privacy Breaches on eBay. In *Privacy Enhancing Technologies Symposium (PETS)*.

[29] Hamed Mohtadi and Alireza Rahimi. 2015. New Attacks on Wi-Fi Protected Setup. *Advances in Computer Science* (2015).

[30] U.S. Department of Transportation: ArcGIS Online. 2025. Military Bases. https://geodata.bts.gov/datasets/usdot::military-bases/explore.

[31] Richard Roberts, Julio Poveda, Raley Roberts, and Dave Levin. 2023. Blue Is the New Black (Market): Privacy Leaks and Re-Victimization from Police-Auctioned Cellphones. In *IEEE Symposium on Security and Privacy*.

[32] Erik Rye and Dave Levin. 2024. Surveilling the Masses with Wi-Fi-Based Positioning Systems. *IEEE Symposium on Security and Privacy* (2024).

[33] Erik C Rye and Robert Beverly. 2023. IPvSeeYou: Exploiting Leaked Identifiers in IPv6 for Street-Level Geolocation. In *IEEE Symposium on Security and Privacy*.

[34] Amirali Sanatinia, Sashank Narain, and Guevara Noubir. 2013. Wireless Spreading of WiFi APs Infections Using WPS Flaws: An Epidemiological and Experimental Study. In *IEEE Conference on Communications and Network Security (CNS)*.

[35] Marco Uras, Enrico Ferrara, Raimondo Cossu, Antonio Liotta, and Luigi Atzori. 2022. MAC Address De-Randomization for WiFi Device Counting: Combining Temporal-and Content-Based Fingerprints. *Computer Networks* (2022).

[36] Mathy Vanhoef, Célestin Matte, Mathieu Cunche, Leonardo S. Cardoso, and Frank Piessens. 2016. Why MAC Address Randomization is not Enough: An Analysis of Wi-Fi Network Discovery Mechanisms. In *Asia Conference on Computer and Communications Security (ASIACCS))*.

[37] Stefan Viehböck. 2011. Brute Forcing Wi-Fi Protected Setup. https://sec-consult.com/fileadmin/user_upload/sec-consult/Dynamisch/Blogartikel/Old_Blogposts/sec-consult-kcodes-netsub-viehboeck.pdf.

[38] WiGLE. 2025. WiGLE – All the Networks. Found by Everyone. https://wigle.net.

[39] Thomas Wolf, Lysandre Debut, Victor Sanh, Julien Chaumond, Clement Delangue, Anthony Moi, Pierric Cistac, Tim Rault, Rémi Louf, Morgan Funtowicz, Joe Davison, Sam Shleifer, Patrick von Platen, Clara Ma, Yacine Jernite, Julien Plu, Canwen Xu, Teven Le Scao, Sylvain Gugger, Mariama Drame, Quentin Lhoest, and Alexander M. Rush. 2020. Transformers: State-of-the-Art Natural Language Processing. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*. Association for Computational Linguistics, Online, 38–45. https://www.aclweb.org/anthology/2020.emnlp-demos.6