WikIPedia: Unearthing a 20-year History of IPv6 Client Addressing

Erik Rye[1,2][0000−0002−8151−8252] and Dave Levin[2][0000−0003−4957−5131]

[1] Johns Hopkins University
[2] University of Maryland

Abstract. Due to their article editing policies, Wikimedia sites like Wikipedia have become inadvertent time capsules for IPv6 addresses. When Wikimedia users make edits without signing into an account, their IP addresses are used in lieu of a username. Wikimedia site dumps therefore provide researchers with over two decades worth of timestamped client IPv6 addresses to understand address assignments and how they have changed over time and space.

In this work, we extract 19M unique IPv6 addresses from Wikimedia sites like Wikipedia that were used by editors from 2003 to 2024. We use these addresses to understand the prevalence of IPv6 in countries corresponding to Wikimedia site languages, how IPv6 adoption has grown over time, and the prevalence of EUI-64 addressing on client devices like desktops, laptops, and mobile phones.

1 Introduction

In the 1993 film *Jurassic Park*, the eccentric billionaire John Hammond resurrects long-extinct fauna by extracting infinitesimal amounts of their DNA from mosquitoes trapped in amber. The DNA in these mosquitoes serves as a historical blueprint for reverse engineering myriad dinosaur species, which eventually populate his eponymous theme park.

In the field of network measurement, Wikimedia sites may well be the analog of mosquito-entrapping amber. When Wikipedians—as editors of the sites are known—make edits to pages without logging in, their public IP address is used in lieu of a username. Thus, the sites (which preserve all historical edits, even those later reversed) and its 24 year history unintentionally function as an archive for historical Internet data. The IP addresses logged-out Wikipedians used are, in essence, frozen in amber along with the timestamp when they were used.

This type of longitudinal data is rare. Few other datasets match the wide timespan over which it was collected; its earliest entries predate most social media and many other mainstays of today's web experience. And datasets that do match its prodigious timeline do not typically include its type of data. Routeviews' Border Gateway Protocol (BGP) data archives [24], for instance, match the timespan of Wikimedia's existence. However, Autonomous Systems (ASes) may advertise large swathes of unused address space, particularly in IPv6, and simply knowing which prefixes were historically advertised obscures the rich detail present in understanding real address assignments.

IPv6 research benefits significantly from knowing active IPv6 addresses. For instance, the IPv6 Hitlist [3,13] regularly provides researchers with lists of known-active addresses to use as targets of active measurement campaigns or as training data for Target Generation Algorithms (TGAs) [43,30,42,20,8,31,28,29,9,10,15]. Similarly, the IPv6 Observatory [26] releases the /48 prefixes of NTP clients that visit its servers on a weekly basis.

In this work, we use archival Wikimedia data to extract the client IP addresses of editors from 2001 through December 2024. We find almost 19 million unique IPv6 addresses, compared to 107 million IPv4 addresses, that are used in lieu of Wikipedia usernames. Our IP address data spans the gamut of sites under the Wikimedia ægis: from the popular primary "encyclopedia" sites for dozens of languages, to crowd-sourced textbooks, dictionaries, and quotes, as well.

In mining this rich vein of historical IPv6 addresses, this work makes the following primary contributions:

1. We analyze global Wikimedia IPv6 statistics, including i) differences between various Wikimedia sites and languages, ii) temporal aspects of Wikimedia IPv6 client addressing, and iii) changes in Wikimedia IPv6 data corresponding with major network events.
2. We examine Extended Unique Identifier - 64 (EUI-64) IPv6 addresses in Wikimedia data and discover that this obsolescent type of address is actually undergoing a modern revival.
3. We compare our Wikimedia IPv6 corpus with the IPv6 Hitlist to understand the information gained from this unique dataset.

2 Background and Related Work

2.1 Wikimedia

The Wikimedia Foundation [32] supports a wide variety of free knowledge projects, including Wikipedia, Wikibooks, Wiktionary, Wikiquote, and others, that allow users to freely access, contribute, and verify its content. These projects are available in a wide variety of languages; Wikipedia, for instance, is available in over 300 languages.

Most users of Wikimedia sites do not contribute information to these wikis in the form of new articles or edits to existing ones. However, some users do; these contributors are encouraged to register an account to which their edit history will be attached. In early 2025, for instance, the English Wikipedia reports over 49M registered users with about a third (15M) having committed at least one edit [41].

However, Wikimedia sites also allow users to contribute edits without logging in. When contributors submit edits without logging in, a warning (Figure 1) alerts them that their IP address will be used in lieu of a username and will be publicly visible.

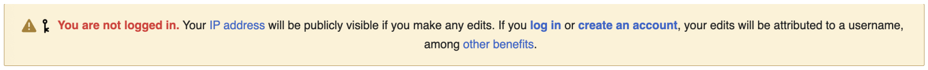


Fig. 1: Wikipedia warns logged-out users that their IP address will be publicly visible.

Wikimedia periodically publishes "dumps" of all of its constituent sites [33]. Each constituent site and language combination is available separately, so that consumers of this data can differentiate between English Wikipedia and Mandarin Chinese Wikiquote entries, for instance. Importantly, these archives contain both current and *historical* data, including edits that were made long ago and have since been overwritten, as well as edits that were later reverted.

2.2 Related Work

Studies of Wikimedia This work uses Wikimedia sites as a source of active client IPv6 addresses over time. It builds on several studies that measure Wikipedia generally, and make use of IP addresses found in Wikipedia edits specifically.

A significant body of work has studied Wikipedia content. Voss authored an early measurement of Wikipedia, counting the number of articles, authors, and edits across multiple language versions of the site [37]. Unlike these prior studies, our work does not focus on the wikis' *content*, but rather the IP addresses of those who help curate it.

Wikipedians' IP addresses have been studied in the context of Wikipedia vandalism—that is, removing accurate information from articles, removing articles entirely, inserting inaccurate information into articles, or otherwise making "unproductive" edits [38,16]. Others have used Wikimedia address edit history to filter for Tor exit nodes [34]. By contrast, we are interested in *all* edits recorded by IPv6 addresses in the Wikimedia dumps.

Almeida et al. performed a broad study of how user behavior on Wikipedia had evolved from 2001 to 2006 [4]. As part of this larger study, they reported finding 3.8M unique IP addresses in Wikipedia edits over this nearly six-year span. By contrast, our work spans over two decades, and examines primarily IPv6 addresses, which were not widely logged by Wikimedia sites until after 2012 (§4).

Finally, Zander et al. extracted IPv4 addresses from Wikipedia edit logs over the three-year period from 2011 to 2014 [44]. They used the IPv4 addresses they obtained to improve insights into IPv4 address space exhaustion beyond the visibility achieved through active measurement campaigns. We focus exclusively on IPv6 in this work, and our study spans over two decades (2003–2024).

Gathering IPv6 addresses This work seeks to gather a large corpus of IPv6 addresses, which has been the focus of many recent studies [3,13,43,30,42,20,8,26]. These rely on a range of passive and active measurements, and have resulted in datasets orders of magnitude larger than we are able to obtain by looking only at Wikimedia sites. However, these prior efforts seek to obtain IPv6 addresses that are live and in active use *right now* (primarily to guide future scanning efforts); in contrast, our work provides a *historical* view of IPv6 addresses.

3 Data Collection Methodology

Wikipedia historical data We obtain IPv6 addresses from edits to Wikipedia and other Wikimedia sites. Wikimedia publishes "dumps" of their sites' content: every article's text, links to every picture, and—critically—all edit metadata going back to the sites' inception in 2001. We obtained the December 2024 dumps for all available content sites.

The most common site type from the December 2024 dumps is Wikipedia, the flagship encyclopedia site. There are 394 individual Wikipedia sites that span both languages (e.g., enwiki and dewiki, for the English and German versions of Wikipedia, respectively[3]) and special events (e.g., Wikimania, Wikipedia's annual conference, had its own public wiki from 2005–2018). Wikitionaries, which are user-editable dictionaries, are the second-most common site type, with 195 individual Wiktionary sites spanning many languages. Table 1 in the Appendix lists the number of individual sites per category.

Extracting IP addresses After downloading this corpus from Wikimedia, we parse the historical edit data to obtain edits that were authored by an unregistered user. Because Wikimedia's policy is to log and use the IP address of unregistered users, this is equivalent to filtering edits for those authored by an IP address. This process is highly unlikely to result in false positives, as Wikipedia specifically prohibits usernames that are (or even look like) an IP address [40]. We separate these by IP version. In this work, we are primarily interested in IPv6 addresses due to the challenges of obtaining large-scale client IPv6 addresses, but retain IPv4 addresses for comparison purposes in §4.

Contemporaneous AS data Finally, we also used Routeviews' IPv6 BGP RIB dumps [24]. In order to determine what AS a historical IP address was in at the time that it was logged, we need to be able to look up these IP addresses in BGP data contemporaneous to their appearance in a Wikimedia site. After obtaining the Routeviews data, we looked each IPv6 address in our dataset up in the chronologically closest Routeviews IPv6 RIB dump to obtain an ASN. As Routeviews produces BGP dumps every two hours, the time delta between an IP address's collection and its AS BGP lookup is rarely more than an hour.

Limitations Our Wikimedia dataset offers a unique view into the history of client IPv6 addresses, but it is not without limitations. First, it is limited to client devices; the IPv6 addresses we extract from the Wikimedia dumps are associated with the user who was submitting the edits to the wiki. It is therefore

[3] Wikimedia uses ISO-639 2-letter language codes as prefixes and the site type as suffixes for the full Wikipedia site identifier.

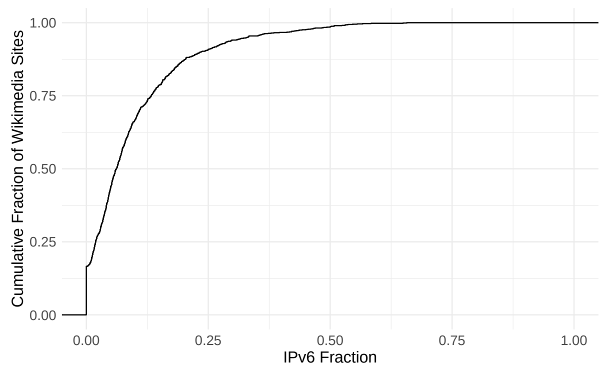


Fig. 2: Fraction of IPv6 addresses of the total number of IP addresses logged per Wikimedia site.

highly unlikely that our data comprises addresses of web servers, routers, and so on. Further, relatively few users make edits to Wikimedia pages, and even fewer make edits without logging into an account. Second, while it has broad reach, Wikipedia is not ubiquitous; several countries block access to Wikipedia (notably, China). This naturally creates bias in our dataset toward regions where Wikipedia is more accessible and popular, as populations without access or desire to visit Wikimedia sites will do less editing. Finally, it is difficult to reason about how representative the dumps' IP addresses are of the entirety of the IPv6 space. However, it is encouraging that the Wikimedia dumps contain hundreds of different language-specific sites, as they likely capture users from the regions that speak those languages.

4 Results

In this section, we analyze the IPv6 client addresses we obtained from the dumps of 1,005 Wikimedia sites, and compare this dataset to other contemporaneous data.

4.1 IPv6 Address Frequency

In parsing the Wikimedia site data dumps and extracting IPv6 addresses from them, we obtained 19,292,487 unique IPv6 addresses. This is approximately 0.18× the number of unique IPv4 addresses used as user identifiers over the same period (107,371,338). The fraction of IPv6 addresses of all IP addresses logged per Wikimedia site varies between 0 and 0.66, though the median value is 0.06. Figure 2 shows the fraction of IPv6 addresses as a CDF of Wikimedia site.

Interestingly, because some Wikimedia sites are highly regional, their proportion of IPv6 addresses indicates regional adoption rates of IPv6. In the tail

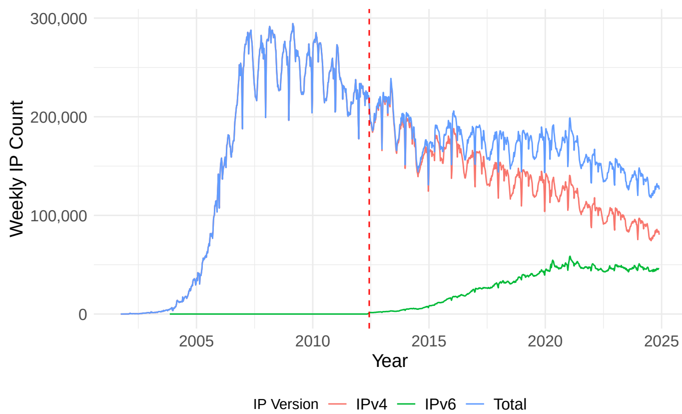


Fig. 3: Unique IP addresses per week by protocol version. Red dashed line is World IPv6 Launch day.

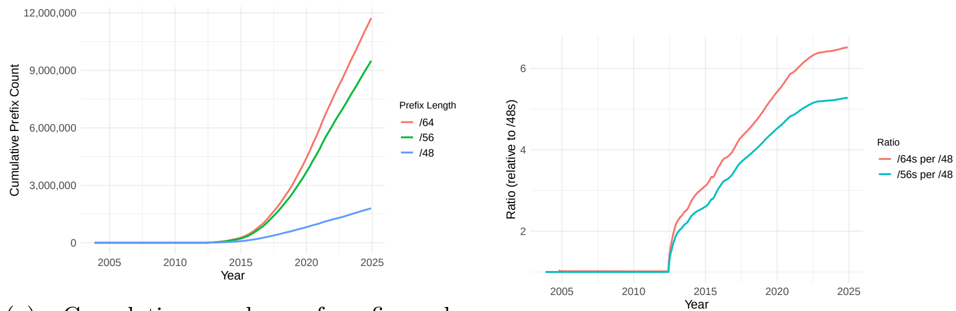
of the distribution are several sites with only a small number of total addresses, such as the Assamese Wikiquote site, which has the highest IPv6 fraction of 0.66 (12 IPv4 and 23 IPv6 addresses). However, the Hindi Wikipedia site contains both a large number of total unique IP addresses, with 297,741, and a large fraction of IPv6 addresses, with 0.57 (168,703). The Serbian Wikipedia has an approximately equal number of total IP addresses logged, with 267,377. Conversely, however, it has an extremely low fraction of IPv6 addresses to total IP addresses, at 0.03 (259,118 IPv4 to 8,259).

The significant variation in the proportion of IPv6 addresses across different Wikimedia sites is likely due to the adoption rate of IPv6 in the countries the language-specific wikis are spoken in. RIPE reports an IPv6 adoption rate of 77% in India, where Hindi is primarily spoken, compared with an IPv6 adoption rate of 7% in Serbia, where Serbian is primarily spoken [23].

Not all Wikimedia sites have logged IPv6 addresses. Of the 1,005 Wikimedia for which we have data, only 828 (82%) have at least one IPv6 address logged. There is at least one IPv4 address logged as a user identifier for 993 (99%) Wikimedia sites. The English Wikipedia site contributes half of the total number of unique logged IPv6 addresses; the German, French, Japanese, and Spanish Wikipedia sites contribute more than 5%. Table 2 in the Appendix lists the top Wikimedia sites.

4.2 Temporal Characteristics

IPv6 addresses used as edit identifiers in Wikimedia sites appear between November 2003 to December 2024, the month that we obtained the Wikimedia data. However, IPv6 addresses as edit identifiers remained relatively sparse from Wikimedia's inception through the mid-2010s.



(a) Cumulative numbers of prefixes observed in the Wikimedia corpus ($N = 19,292,487$ /128s).

(b) /56 and /64 ratio per /48 over time.

Fig. 4: Prefix observations in the Wikimedia corpus.

Figure 3 displays the number of unique weekly IP addresses logged by the 1,005 Wikimedia sites since 2001 across both versions of IP. While IPv6 addresses first appear in 2003, IPv4 addresses are first logged in 2001 and dominate throughout the first decade of Wikimedia's existence. IPv6 addresses start to increase shortly after World IPv6 Launch day, which occurred on 6 June 2012 and Wikimedia participated in [11], and is annotated in Figure 3 as a red, vertical dashed line. At the end of 2024, roughly twice as many IPv4 addresses appeared as IPv6 addresses.

Figure 4(a) depicts the number of cumulative /48, /56, and /64 prefixes observed in the Wikimedia IPv6 corpus over time. While best practice recommends /48 and /56 as prefix delegation sizes to customer end sites, in practice, ISPs may assign /48, /52, /56, /60 and /64 [22]. This helps us estimate the number of unique clients with IPv6 addresses logged in the Wikimedia corpus. While the number of total /48s is relatively low (∼1.8M), the number of /64s is approximately 60% of the total number of unique /128s. This indicates that 40% of Wikimedia IPv6 addresses come from the same /64 networks, which strongly suggests edits were made by users that are part of the same home or campus network, if not the same individuals themselves. Figure 4(b) shows that the number of /56s and /64s per /48 increased rapidly through 2022. However, the number of observed subnets per /48 has flattened from 2023 onward, indicating relatively stable number of observed /56 and /64 subnets per /48.

Finally, Figure 5 displays the length of time between the first and last observations of each IP address in the Wikimedia dataset. Most addresses (∼52% of IPv4 and ∼64% of IPv6) are only observed once in the Wikimedia dump, which manifests as a lifetime of 0. Because best practice recommends that IPv6 client addresses be both random and ephemeral [21], IPv6 address lifetimes are significantly shorter than IPv4 address lifetimes. Further, IPv4 addresses may have *many* clients behind a device performing Network Address Translation (NAT), whereas NAT is extraordinarily rare in IPv6, and most IPv6 addresses correspond to a single host.

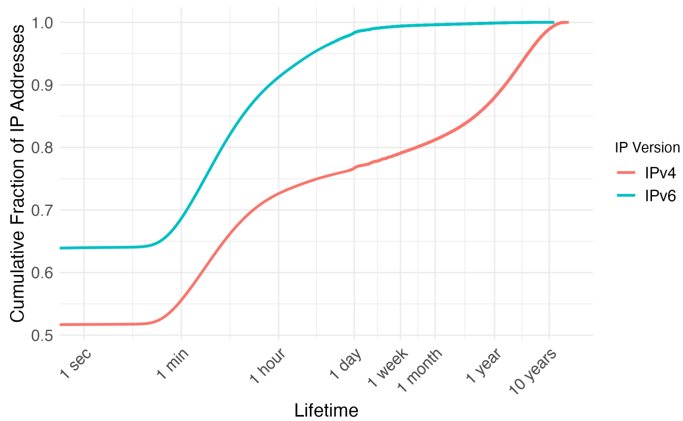


Fig. 5: Lifetimes of the IP addresses logged in the Wikimedia dataset by IP protocol version.

4.3 AS Contributions

Next, we examine the number of IPv6 addresses from Wikimedia edits according to their AS. Because the AS to which an IPv6 address belongs may change over time as address blocks are reallocated and reassigned, we use contemporaneous BGP RIB data from Routeviews to look up the AS of an IP address in the chronologically-closest RIB dump.

Figure 6 displays the number of unique weekly IPv6 addresses per week observed in the Wikimedia site dumps for each of the top five ASes. While Comcast (AS7922), AT&T (AS7018), and Deutsche Telekom (AS3320) have had IPv6 addresses in the Wikimedia corpus since 2012, other ASes did not deploy IPv6 until later. This is reflected in Figure 6. For instance, Reliance Jio (AS55836), an Indian telecom provider, did not deploy IPv6 widely until September 2016 [19]. This deployment date tracks with AS55836's curve in Figure 6, and correlates with contemporaneous IPv6 availability measurements, e.g. from APNIC [6].

Finally, note that the Figure 6 displays sharp decreases for many of the ASes. For instance, Comcast (AS7922) exhibited a large decrease in IPv6 addresses logged in early 2015, while Reliance Jio (AS55836) had an extended dip in IPv6 edits over a three-month period in 2023. While we were unable to definitively link these to publicized network events, we speculate that these decreases in IPv6 edits occurred due to changes in IPv6 address assignment policies or routing by these networks.

4.4 EUI-64 Addresses

Next, we turn to an analysis of the devices manufacturers of the logged Wikimedia IPv6 addresses. To do this, we filter for the subset of IPv6 addresses that are EUI-64. EUI-64 addresses embed the Media Access Control (MAC) address

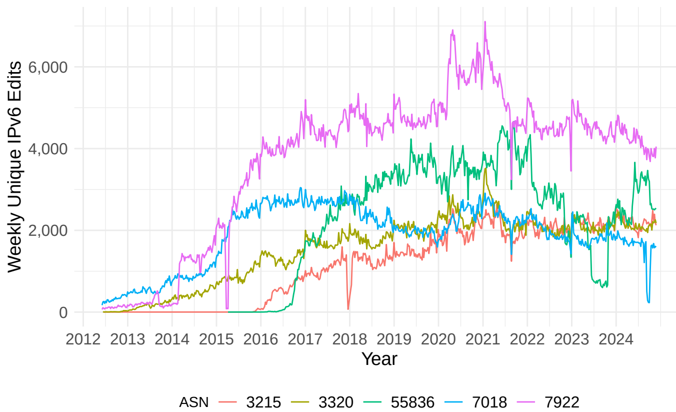
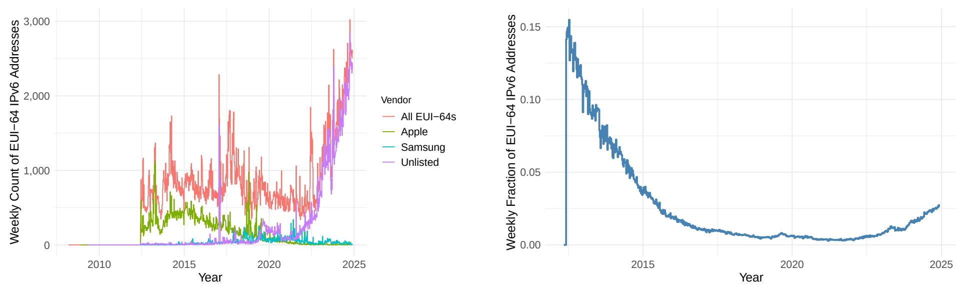


Fig. 6: Number of unique IPv6 addresses per week for the top 5 ASes.



(a) Number of EUI-64 IPv6 addresses per week.

(b) The fraction of EUI-64 IPv6 addresses of all IPv6 addresses per week.

Fig. 7: EUI-64 addresses in the Wikimedia corpus.

of the device's interface into the lower 64 bits of the IPv6 address[4]; because MAC addresses frequently encode the manufacturer of the device in the MAC address's upper three bytes, we can determine the type of devices unregistered Wikimedia users were using for this subset.

Of the 19M total IPv6 addresses logged, 167,417 (0.87%) of are EUI-64. These EUI-64 IPv6 addresses contain 145,832 unique MAC addresses. Table 3 in the Appendix lists the number of unique MAC addresses by the manufacturer derived from looking up the upper three bytes (the Organizationally Unique Identifier (OUI)) of each MAC address in the public list of IEEE OUI assignments [1]. While prior work has found large numbers of EUI-64 addresses assigned to routers [25] and IoT devices [27,26], EUI-64 addresses in the Wikimedia

[4] The MAC address is extended to 64 bits by inserting the bytes `0xfffe` between its third and fourth bytes; the second-least significant bit of the most significant byte of the MAC address (the Universal/Local (U/L) bit) is typically also inverted.

dataset largely belong to mobile, laptop, and desktop clients. For instance, Apple is the most commonly resolved OUI vendor, along with HP, Intel, ASUS, and Samsung. However, more than half (56%) of the EUI-64-derived MAC addresses did not resolve to an IEEE-assigned OUI, a phenomenon also observed by other recent work [26]. This suggests that some devices may be using EUI-64 IPv6 addresses in conjunction with randomized MAC addresses, a common privacy protection employed by mobile devices [36,18,17,12,35].

EUI-64 addresses are well-known to present a privacy risk to users, as they allow long-term tracking of a static identifier [21]. Surprisingly, the number of edits coming from EUI-64 addresses has *increased* over time. Figure 7(a) depicts the number of EUI-64 IPv6 addresses making edits per week. This is due in large part to the "Unlisted" MAC addresses, whose OUIs are not in the IEEE OUI database. Figure 7(a) also shows that while Apple historically employed EUI-64 addressing in the mid-2010s, it has largely phased out these types of addresses. However, the prevalence of "Unlisted" MAC addresses has dramatically increased since 2021. The privacy ramifications of using random MAC addresses to form EUI-64 IPv6 addresses are less severe than embedding a device's true hardware MAC address, as random MAC addresses obscure the device manufacturer. Nonetheless, if the random MAC address used in the EUI-64 IPv6 address is sufficiently long-lived (e.g., Android and iOS use a stable, random MAC address on a per-network basis for most Wi-Fi networks[7,5]) the longitudinal tracking threat of EUI-64 IPv6 addresses remains.

Not only have the raw number of EUI-64 IPv6 addresses increased over time, but the *fraction* of IPv6 addresses that are EUI-64 has simultaneously increased. Figure 7(b) shows that while EUI-64 IPv6 addresses represented less than 1% of the total number of IPv6 addresses seen weekly between 2017 and 2023 (down from a high of ∼15%), they have become increasingly prevalent once again, comprising nearly 3% of IPv6 addresses in December 2024.

4.5 Comparison with the IPv6 Hitlist

Finally, we compare the IPv6 addresses in the Wikimedia corpus to the IPv6 Hitlist's [3,13,14] ICMPv6-responsive addresses over the lifetime of the Hitlist. Figure 8 displays the number of unique /48 prefixes contained in the Wikimedia dumps, as well as overlapping /48s with the IPv6 Hitlist (IPv6 Hitlist-only /48s omitted for clarity) binned by month. The number of /48s seen monthly in Wikimedia data reaches ∼100k by about 2020. The overlap with the IPv6 Hitlist remains relatively low (∼5-10%) until December 2023, when the IPv6 Hitlist began incorporating addresses from IPinfo [2]. From this point forward, the amount of overlap with the IPv6 Hitlist is approximately one-third, indicating that the /48s coming from IPinfo are likely from customer networks.

This demonstrates that the Wikimedia IPv6 corpus provides a unique source of IPv6 address data that is difficult to obtain from either active measurements or without the broad reach than Wikimedia provides. For applications that rely on obtaining active IPv6 addresses (e.g., IPv6 TGAs), the Wikimedia corpus provides another data source currently missing from state-of-the-art hitlists.

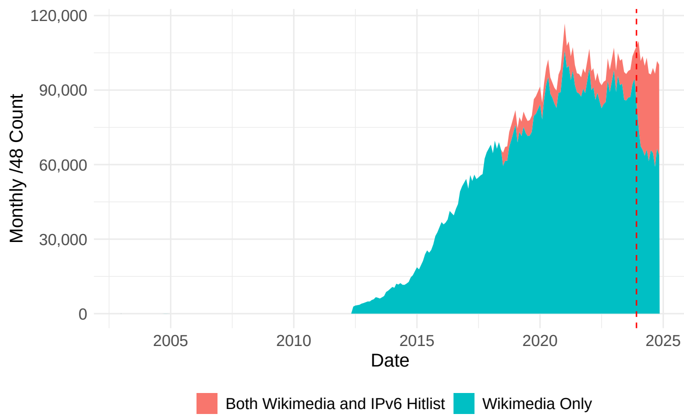


Fig. 8: Number of unique /48s in the Wikimedia dumps and overlap with the IPv6 Hitlist. The red dashed line at December 2023 is when the IPv6 Hitlist began incorporating data from IPinfo.

5 Conclusion

Currently, the network measurement community is focused on collecting as many active, in-use IPv6 addresses as possible [3,8,26], so as to enable the future of Internet-wide scanning. In this paper, we have instead opted to peer into the past. By using IPv6 addresses encased in the "amber" of decades of edits to Wikipedia and other Wikimedia websites, we showed it is possible to reason about significant changes in IPv6 adoption and policy over decades.

Toward that end, we extracted 19M unique IPv6 addresses over 2003–2024, spanning multiple languages, and ASes around the world. Our historical view of IPv6 shows how critical the World IPv6 Launch Day truly was for IPv6 adoption: prior to it, IPv6 was an almost nonexistent novelty among client devices.

We find that IPv6 addresses tend to have shorter lifetimes than their IPv4 peers. The majority ($\sim$64%) of the IPv6 addresses that are logged in Wikimedia edits appear only once. This is likely an artifact of the IPv6 client addressing best practice to choose client addresses randomly and to change them periodically.

We observe IPv6 roll-outs occurring by ASes. When IPv6 service is deployed to customers by ASes, steep upticks in logged IPv6 addresses from those ASes occur. This allows us to retroactively detect when IPv6 is *deployed*, rather than simply *announced* by the AS using BGP.

Finally, our analyses showed how EUI-64 IPv6 addresses—largely considered a privacy violation [26]—had been getting phased out, but have recently started to see a resurgence. This points to randomized MAC addresses being used to build EUI-64 addresses, and confirms trends detected in other recent work.

Acknowledgments. This work was supported by NSF grant CNS-2323193.

References

1. IEEE OUI database (2025), http://standards-oui.ieee.org/oui.txt
2. IPInfo.io (2025), https://www.ipinfo.io
3. IPv6 Hitlist Service (2025), https://ipv6hitlist.github.io/
4. Almeida, R.B., Mozafari, B., Cho, J.: On the Evolution of Wikipedia (2007)
5. Android: MAC randomization behavior (2025), https://source.android.com/docs/core/connect/wifi-mac-randomization-behavior
6. APNIC: IPv6 Per-Country Deployment for AS55836: RELIANCEJIO-IN Reliance Jio Infocomm Limited, India (IN) (2025), https://stats.labs.apnic.net/ipv6/AS55836?c=IN&p=1&v=1&w=30&x=1
7. Apple: Use private Wi-Fi addresses on Apple devices (2025), https://support.apple.com/en-us/102509
8. Beverly, R., Durairajan, R., Plonka, D., Rohrer, J.P.: In the IP of the Beholder: Strategies for Active IPv6 Topology Discovery. In: ACM Internet Measurement Conference (IMC) (2018)
9. Cui, T., Gou, G., Xiong, G.: 6GCVAE: Gated Convolutional Variational Autoencoder for IPv6 Target Generation. In: Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD) (2020)
10. Cui, T., Xiong, G., Gou, G., Shi, J., Xia, W.: 6VecLM: Language Modeling in Vector Space for IPv6 Target Generation. In: European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML PKDD) (2020)
11. Erik Möller: IPv6 initiative/2012 IPv6 Day announcement (2012), https://meta.wikimedia.org/wiki/IPv6_initiative/2012{_}IPv6{_}Day{_}announcement
12. Fenske, E., Brown, D., Martin, J., Mayberry, T., Ryan, P., Rye, E.C.: Three Years Later: A Study of MAC Address Randomization In Mobile Devices And When It Succeeds. In: Privacy Enhancing Technologies Symposium (PETS) (2021)
13. Gasser, O., Scheitle, Q., Foremski, P., Lone, Q., Korczyński, M., Strowes, S.D., Hendriks, L., Carle, G.: Clusters in the Expanse: Understanding and Unbiasing IPv6 Hitlists. In: ACM Internet Measurement Conference (IMC) (2018)
14. Gasser, O., Scheitle, Q., Gebhard, S., Carle, G.: Scanning the IPv6 Internet: Towards a Comprehensive Hitlist. CoRR **abs/1607.05179** (2016), http://arxiv.org/abs/1607.05179
15. Hou, B., Cai, Z., Wu, K., Yang, T., Zhou, T.: 6Scan: A High-Efficiency Dynamic Internet-Wide IPv6 Scanner With Regional Encoding. Networking, IEEE/ACM Transactions on (2023)
16. Kiesel, J., Potthast, M., Hagen, M., Stein, B.: Spatio-Temporal Analysis of Reverted Wikipedia Edits (2017)
17. Martin, J., Mayberry, T., Donahue, C., Foppe, L., Brown, L., Riggins, C., Rye, E.C., Brown, D.: A Study of MAC Address Randomization in Mobile Devices and When it Fails (2017)
18. Matte, C., Cunche, M., Rousseau, F., Vanhoef, M.: Defeating MAC Address Randomization through Timing Attacks. In: ACM Conference on Security and Privacy in Wireless and Mobile Networks (WiSec) (2016)
19. Molay Ghosh: Reliance Jio boosts India past 20% IPv6 capability (2017), https://blog.apnic.net/2017/02/07/reliance-jio-boosts-india-past-20-ipv6-capability/
20. Murdock, A., Li, F., Bramsen, P., Durumeric, Z., Paxson, V.: Target Generation for Internet-wide IPv6 Scanning. In: ACM Internet Measurement Conference (IMC) (2017)

21. Narten, T., Draves, R., Krishnan, S.: Privacy Extensions for Stateless Address Autoconfiguration in IPv6. RFC 4941 (Draft Standard) (Sep 2007). https://doi.org/10.17487/RFC4941, https://www.rfc-editor.org/rfc/rfc4941.txt
22. RIPE: Best Current Operational Practice for Operators: IPv6 Prefix Assignment for End-Users - Persistent vs Non-Persistent, and What Size to Choose (2017), https://www.ripe.net/publications/docs/ripe-690
23. RIPE: Use of IPv6 for World (2025), https://stats.labs.apnic.net/ipv6/XA
24. Routeviews: University of Oregon Route Views Project (2025), https://www.routeviews.org/routeviews/
25. Rye, E., Beverly, R., claffy, k.: Follow the Scent: Defeating IPv6 Prefix Rotation Privacy. In: ACM Internet Measurement Conference (IMC) (2021)
26. Rye, E., Levin, D.: IPv6 Hitlists at Scale: Be Careful What You Wish For. In: ACM SIGCOMM (2023)
27. Saidi, S.J., Gasser, O., Smaragdakis, G.: One Bad Apple Can Spoil Your IPv6 Privacy (2022)
28. Shen, Z., Chen, P., Xie, Y., Chen, C., Zhang, Y., Yang, G.: 6Trace: An Effective Method for Active IPv6 Topology Discovery. MDPI Electronics (2025)
29. Song, G., Yang, J., Wang, Z., He, L., Lin, J., Pan, L., Duan, C., Quan, X.: DET: Enabling Efficient Probing of IPv6 Active Addresses. Networking, IEEE/ACM Transactions on (2022)
30. Steger, L., Kuang, L., Zirngibl, J., Carle, G., Gasser, O.: Target Acquired? Evaluating Target Generation Algorithms for IPv6 (2023)
31. Sun, X., Dang, F., Yang, Z., Jin, X., Li, J., Liu, Y.: 6Loda: Pattern Filtering and Ensemble Learning for IPv6 Target Generation and Scanning. IEEE Conference on Computer Communications (INFOCOM) (2025)
32. The Wikimedia Foundation: Wikimedia (2025), https://www.wikimedia.org/
33. The Wikimedia Foundation: Wikimedia Downloads (2025), https://dumps.wikimedia.org/
34. Tran, C., Champion, K., Forte, A., Hill, B.M., Greenstadt, R.: Are anonymity-seekers just like everybody else? An analysis of contributions to Wikipedia from Tor. In: IEEE Symposium on Security and Privacy (2020)
35. Uras, M., Ferrara, E., Cossu, R., Liotta, A., Atzori, L.: MAC Address De-Randomization for WiFi Device Counting: Combining Temporal-and Content-Based Fingerprints. Computer Networks (2022)
36. Vanhoef, M., Matte, C., Cunche, M., Cardoso, L.S., Piessens, F.: Why MAC Address Randomization is not Enough: An Analysis of Wi-Fi Network Discovery Mechanisms (2016)
37. Voss, J.: Measuring Wikipedia. In: International Conference of the International Society for Scientometrics and Informetrics (ISSI) (2005)
38. West, A.G., Kannan, S., Lee, I.: Detecting Wikipedia Vandalism via Spatio-Temporal Analysis of Revision Metadata? In: European Workshop on Systems Security (2010)
39. Wikipedia: Trust and Safety Product/Temporary Accounts (2025), https://www.mediawiki.org/wiki/Trust_and_Safety_Product/Temporary_Accounts
40. Wikipedia: Username policy (2025), https://en.wikipedia.org/wiki/Wikipedia:Username_policy
41. Wikipedia: The Free Encyclopedia: Wikipedia:Wikipedians (2025), https://en.wikipedia.org/wiki/Wikipedia:Wikipedians
42. Williams, G., Erdemir, M., Hsu, A., Bhat, S., Bhaskar, A., Li, F., Pearce, P.: 6Sense:Internet-Wide IPv6 Scanning and its Security Applications. In: USENIX Security Symposium (2024)

43. Williams, G., Pearce, P.: Seeds of Scanning: Exploring the Effects of Datasets, Methods, and Metrics on IPv6 Internet Scanning. In: ACM Internet Measurement Conference (IMC) (2024)
44. Zander, S., Andrew, L.L., Armitage, G.: Capturing Ghosts: Predicting the Used IPv4 Space by Inferring Unobserved Addresses. In: ACM Internet Measurement Conference (IMC) (2014)

Appendix A: Ethical Considerations

Users that wish to make an edit to a Wikimedia site page, but do not wish to create or log into an account, are warned with text that reads “You are not logged in. Your IP address will be publicly visible if you make any edits” (see Figure 1). Continuing to submit an edit constitutes the consent of the user. Consistent with this statement, Wikimedia site dumps that contain user addresses are publicly available [32].

Due to these facts, the collection of the IP addresses logged by Wikimedia sites does not itself raise ethical issues. As of July 2025, Wikimedia is phasing out the collection of IP addresses from logged-out users due to GDPR and other privacy concerns [39]; as such, we will not release the set of IP addresses we extracted from these dumps separately.

Appendix B: Wikimedia Sites

Table 1: Number of Wikimedia sites by wiki type

Wiki Type	**Site Count**	**Wiki Type**	**Site Count**
Wikipedia	394	Wikimedia	37
Wiktionary	195	Wikinews	36
Wikibooks	121	Wikivoyage	27
Wikiquote	97	Wikiversity	18
Wikisource	80	**Total**	**1,005**

Table 1 lists the number of Wikimedia sites by category; within each category, most of the individual sites are language-specific variations of the site type, although others are for special events or uses (e.g., wikis for the Wikipedia “Wikimania” event.)

Appendix C: IPv6 Addresses by Wikimedia Site

Table 2 lists the top Wikimedia sites by number of unique IPv6 addresses logged in their site’s dump.

Table 2: Number of unique logged IPv6 addresses per Wikimedia site; some addresses appear in the logs of multiple sites.

Wiki Site	**#IPv6 Addresses**	**% of All IPv6**
English Wikipedia	9,638,421	50
German Wikipedia	1,518,286	7.9
French Wikipedia	1,457,954	7.6
Japanese Wikipedia	1,173,841	6.1
Spanish Wikipedia	1,061,633	5.5
1,000 other	5,123,867	26.6
Total	**19,292,487**	**100**

Table 3: Number of distinct devices per manufacturer as determined from EUI-64 IPv6 address-embedded MAC addresses.

Manufacturer	**Count**	**Manufacturer**	**Count**
Unlisted	82,244	ASUSTek	2,962
Apple	23,168	HP	1,662
Samsung	5,367	Hon Hai Precision	1,246
Intel	4,393	1,141 other	21,631
Dell	3,159	**Total**	**145,832**

Appendix D: EUI-64 IPv6 Addresses

Table 3 lists the number of distinct MAC addresses observed in EUI-64 IPv6 addresses in the Wikimedia data. We use the IEEE OUI database [1] to resolve the MAC addresses to manufacturers. Surprisingly, the most commonly observed manufacturer is “unlisted”, meaning that the OUIs did not resolve to any manufacturer.