Jason Li

Professor McNair

EE 584 WS

21 October 2023

"IPvSeeYou: Exploiting Leaked Identifiers in IPv6 for Street-Level Geolocation"

IPvSeeYou is a security vulnerability that allows remote and unauthenticated perpetrators to detect your precise location via correlating one's home MAC address with BSSID information and coverage to a specific router. Researchers Erik C. Rye from the University of Maryland and Robert Beverly of the Center of Measurement and Analysis of Network Data (CMAND) were able to source corresponding MAC addresses in various IPv6 addresses via high-speed network probing [1]. 48-bit MAC identifiers are unique layer-2 network interfaces that are seen in hardware protocols such as Ethernet, WiFi, and Bluetooth to track each product from a manufacturer. This enables MAC to be indiscoverable by remote perpetrators not on the same subnet. However, IPv6 automatic host selection is not subject to this which leads today's operating system (OS) to randomly generate host bits. Improper authentication of MAC addresses led perpetrators to access the protocol stack of WAN interfaces that control all-in-one WiFi routers. Statistical inference maps a router's WAN and WiFi MAC addresses from different manufacturers and devices to launch a large-scale data fusion attack that associates WiFi BSSSIDS with WAN MACs from geolocation (wardrive) databases. Thus, more than twelve million routers in 146 countries and territories with IPv6 are beached. Technology and deployment constraints are utilized to breach more IPv6 home routers by correlating with a

common router. Despite giving these concerns to companies, the monopolized telecommunications industry makes it so this threat can still be of concern.

High-speed active IPv6 network topology techniques allow hosts and networks to be found in the IPv6 address space. Customer Premises Equipment (CPE) devices in homes allow IPv6 periphery discovery and they utilize legacy EUI-64 addresses as they run deprecated OS and legacy embedded system software. In addition, they have System-on-a-Chip (SoC) architecture with multiple interfaces which Broadcom describes as where “each interface is assigned a MAC address predictably from a small range” [2]. This is exemplified by each network interface MAC addresses having their last digit increment by one. This pattern allows perpetrators to infer the BSSID of a WAN MAC address which enables *street-level geolocation*. If a legacy system is exposed in a WAN, non-legacy systems can also be exposed via upstreaming the router to see where non-legacy systems addresses are. Due to the reliance on legacy systems, this vulnerability is highly likely to stay. IPvSeeYou uses an algorithm to connect WAN and WiFi inference MAC addresses, validates geolocation targets within thirty-nine meters, geolocates 12 million CPE devices via IPv6 prefixes via the previous two steps, and searches through wardriving databases, captures more CPE devices via clustering and through non-legacy systems, and finally discloses this information to companies with steps to prevent this vulnerability. IPv6 only works with active responsive probes from routers, necessitating legacy systems, provable correlations between MAC addresses and their BSSIDs, and searchable BSSIDs.

In past research, Stateless Address Autoconfiguration (SLAAC) is what IPv6 addresses are autogenerated by, not from static assignment or DHCPv6. EUI-64 is where the lower 64-bits of a 128-bit address (Interface Identifier [IID]) has MAC address embedded. EUI-64 is flawed as

static unique IID is trackable by perpetrators and is globally unique with 2^{24} bits (referred to as Organizationally Unique Identifiers (OUIs) assigned to manufacturers. IP addresses and hostnames only identify the network or operator, not the actual location of a device which it may not want to be located. There have been various methods researched to see how to geolocate an IP address via advertisements, language and content customization, geo-fencing content, policy and law enforcement, anti-fraud, authentication, and forensics. Third-party solutions involve registry databases via whois and DNS, utilizing speed-of-light propagation delay to pinpoint a location, network architecture, and privileged feeds. Despite privacy concerns, these methods can provide big-picture geolocation, i.e., city. However, examples from researchers Poese et al., showed 50-90% of geolocated locations were off by 50 kilometers of error compared to actual verifiable locations [3], and researchers Komosny et al. had 50-to-567-kilometer mean errors in eight commercial databases [4]. Erik C. Rye and Robert Beverly aim to produce a method that provides accurate granularity for a location that does not rely on landmarks and geolocating targets in a high-density area due to their prominence in cloud storage services.

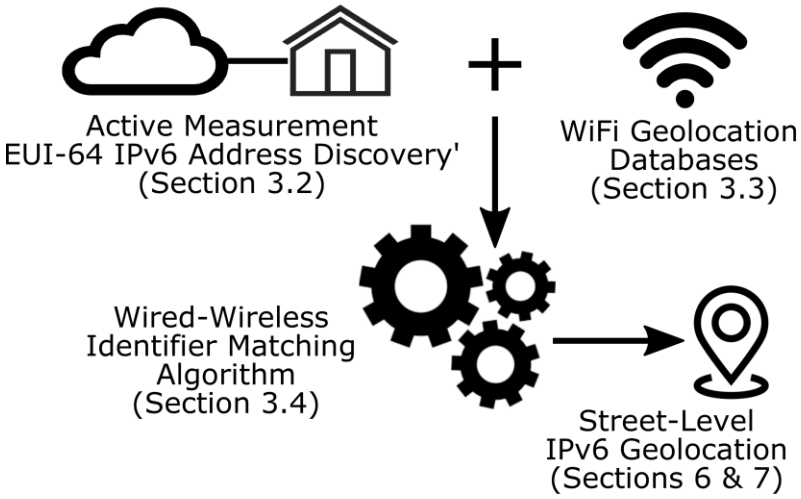


Figure 1: IPvSeeYou System Diagram. EUI-64 IPv6 MAC addresses are correlated with WiFi BSSIDs from geolocation databases.

Prior research has found that Bluetooth MAC and WiFi addresses are mostly sequential even across different link-layer protocols, so perpetrators can recognize patterns. The researchers aim to take this ideal, but it will not necessitate nearness to a target and will be primarily for CPE, not mobile devices. Network Address Translation (NAT) is eliminated due to the range of addresses IPv6 provides. CPE devices need to use a connectivity model via a routed hop whereas NAT is found in residential areas, which makes CPE discovery difficult. The researchers discovered IPv6 network devices in customer-edge network CPEs. Previous edgy algorithms had five million unique MAC addresses in sixteen million legacy IPv6 addresses but were insufficient in geolocating or identifying. DNS response semantics were studied to see if reverse zones could crack the key. Active measurements were generated via Murdock and their team. User-level IPv6 behavior and filtering, especially with dynamics, was observed in a huge online social network. The keyed hash function for the creation of IPv6 flow labels which enables device tracking even when random addresses are used has been patched in common operating systems. CPE IPv6 legacy addresses were utilized to see if prefixes and IID addresses changed. The researchers' main vision is to use the CPE exploit to geolocate unlike any of the previous methods.

CPEs have legacy EUI-64 addresses and a pattern for MAC addresses that are not refreshed. Network probes of EUI-64 addresses can be pinged to gather MAC addresses. This can then lead to exploring BSSIDs from open-source databases, even with non-EUI-64 SLAACs. Instead of perusing CPE routers, non-TTL limited ICMP6 echo request probes and scanning were utilized. The clear and reversible results have 347 million EUI-64 addresses and sixty-one million unique MAC identifiers, though some MAC addresses may be repeated due to devices in new networks or were reused. 126,730 (or 0.2%) of EUI-64 addresses may have had the same

MACs due to those addresses appearing in Autonomous Systems. The methodology can be applied to various geolocation vaults, but five sources (notably one was Apple's WiFi geolocation service) were utilized. In total, sixty-four million BSSIDs were gathered from three open-source databases but have sample bias due to their reliance on the locations of their users. 802.11 BSSIDs from Apple's geolocation service were accessible via an API along with more proximity location intel (to prevent spam requests) and can be found if the query is successful. These were also used as validation for BSSIDs that conform to EUI-64 MAC addresses which returned the BSSID coordinates and four hundred close-by BSSIDs. This was done until the offset value or location was too large resulting in 444,860,460 distinct BSSIDs from Apple's database. 450,018,123 unique BSSIDs were found in 238 countries and territories.

CPE device matching can be difficult in the rare chance that a vendor or device has a BSSID MAC address greater than its WAN MAC address. This can also be complicated as the vendors do not release how they assign MAC address assignments, even for different devices. The researchers statistically infer by having a bunch of WAN and WiFi MAC address offset correlations that are stored in a database. Sometimes this is done by having a WAN MAC be the nearest to a BSSID, but this is not wise if data is missing (no BSSID or WAN address). To curve this, the number of addresses per device was prevented from relating to addresses that differ from a particular size. Should there be a simplistic algorithm with an incredibly high offset, this can result in correlations being made for different devices. The algorithm devised infers the most probable offset of BSSID and the WAN MAC address for a given OUI. The OUI is sectioned based on how many MAC addresses are assigned to each device, with a distribution of intra-MAC distances eventually computed. The most frequent distance is then corresponded by finding the greatest common denominator which leads to a high confidence level should the ratio

turn out high. Each EUI-64 MAC address in ascending sorted order for each OUI with greater than 100 WAN MAC and 100 BSSID instances is looked through. The absolute value offset is determined between a BSSID and EUI-64 MAC address, constrained by OUI section size. The offset for a device is then justified to be the most common offset among all the successes. The algorithm succeeded and failed, but the data can result in one device having both of its addresses more likely than two devices having adjacent addresses. OUIs with a substantial number of devices can throw things off, but the algorithm is close to tangible results. For MAC addresses in an EUI-64 IPv6 address, the database investigates the BSSID offset of a particular OUI then correlated with information from geolocation databases.

The researchers note that IPvSeeYou rely on EUI-64 legacy addresses that ping back when triggered and are most optimal for CPEs with WiFi integrated inside unlike modems with forward-facing legacy addresses that go from Ethernet to an Access Point (AP). The offset mathematics is dependent on a single integrated system on chip (SoC) CPEs with MACs address interfaces deriving from one OUI. Should one OUI be split into different device models where there are lots of distinct WAN MAC addresses, the algorithm will not work. Also, the databases with all geolocation information may have changed or be deprecated due to changes in the values resulting in deviations in the original data. Validation, the prominence of CPE devices with the right specifications, and the inference to CPE devices that do not fit the mold show that this algorithm is a robust one. Researchers are cognizant of the privacy MAC addresses bear and as such the research overview and instructions were sent to the team's Institution Review Board (IRB) which confirmed the study as no confidential data is inferenced to that of people and due to its potential beneficial impact. Geolocated data and MAC addresses are private and data from the study is from the whole aggregation via statistics and does not compromise anyone's

information. Perpetrators could analyze the results, but the researchers see that security experts can look at the vulnerability of IPv6 addresses to network modem manufacturers and Internet Service Providers (ISPs). One manufacturer and residential service are patching their network architectures to prevent this vulnerability from occurring on their attack surfaces.

Verification was performed based on crowd-sourcing statistics, certain CPE devices, and collaborating with a large residential ISP. Volunteers were directed to a web app that assesses their IPv6 connection, whether it was operating, tracks the user's address, and pings for the user's geolocation via an HTML5 API. Approving users had their IPv6 address tuples and accurate geolocation recorded. In addition, active probing (edgy) and IPvSeeYou inference were performed utilizing the CPE router's IPv6 address. Eighty-four percent of 50 participants in Asia, Europe, and North America did not utilize EUI-64 addresses. Five out of eight with legacy EUI-64 addresses were successfully located with IPvSeeYou. Four of those geolocations via IPvSeeYou had 50-meter precision which was better than the competitor's geolocation services, whereas the last one was .68 kilometers from the verified HTML5 location compared to 300 km from competitors. The three that did not work were due to one device's inability to find the offset as the BSSIDs and WAN MAC addresses had limited observations for inferencing. IPvSeeYou worked for the other two devices but geolocation data from established sources was not able to be found and correlated with IPvSeeYou. Through these means, besides the one device, IPvSeeYou is effective but due to a small sample size, the researchers collaborated with a big North American ISP. 1,350 random samples were provided, and bulk reverse geocoding was used to gather necessary data. 486 CPE or 36% of samples were not sufficient due to lack of data or overprediction. Eighty percent of inferences were accurate based on the null hypothesis while information regarding individual error or error bounds was restricted for confidentiality reasons.

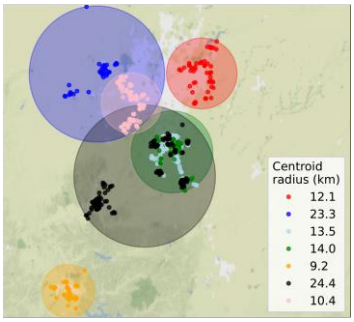
Eighty percent of the results were valid if ZIP code geo-granularity is provided overprediction is minimized and results are not lagging, as well as accurate information regarding the ground-truth ZIP. To further verify, eighteen used CPEs with relevant OUIs were studied and shown to have no matching BSSID. This is due to different OUIs being prevalent. Due to this incompatibility, ten devices resulted in negatives with IPvSeeYou, five had correct predictions, and three were not right due to a lack of nearby MAC observations meaning they did not have EUI-64 compatibility.

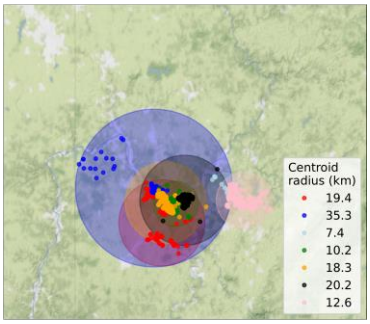
Inferred offsets and their given geolocation were analyzed to see if IPvSeeYou was effective in its tasks. Filtering was done to smooth the data and not permit any wild-off factors most prominently with the lowest confidence offset OUI as some OUI may have multiple device models with differing offsets or one medium (wireless BSSIDs or wired MACs) has a lot more observation than another. Three percent of 32,345 (1,008) unique OUIs came from 31,720,611 distinct legacy MAC addresses with 52% discovered from probing. Approximately twelve million have a BSSID in geolocation databases in a respectable offset, with about 38% of EUI-64-based MAC addresses from 1,008 OUIs and about 20% from legacy MAC address space. The inferred offset ranged from -16 to 15 with a statistical mode which is fine as sequential MAC address assignment results in small offsets which corresponds well. Greater than one-fourth of OUIs had an inferred offset of zero between wireless BSSID and wired MAC. Zero mode may be determined from no link-layer ID for an EUI-64 IPv6 address for specific interfaces, i.e., cellular interface in a hotspot, which has the MAC address of another interface (BSSID) generates EUI-64 IPv6 address that avoids address collision. MAC address reuse could exist between wired and wireless interfaces which means noncompliance with IEEE standards. Results also confirm at least one IPv6 address from 1,114 unique autonomous system numbers (ASNs) with a

geolocated BSSID, accounting for about 5% of approximately twenty-seven thousand IPv6 ASNs in the universal Border Gateway Protocol (BGP) routing table. About 34% (118,429,034) of 347 million EUI-64 addresses had a WAN MAC address correlated with a location-tracked BSSID. Filtering prefix cycling and address overprediction, some EUI-64 addresses paired to the same BSSID. Interestingly, Germany had the most pinpoint correlations with about one-fourth of total matches as an immensely popular German router model was found in the data.

IPvSeeYou currently applies to CPE devices with EUI-64 addresses and associated prefixes, but it can be extended to any CPE device via geolocation correlation. Despite it resulting in less accuracy, it allows for CPE devices to be geolocated without knowing its information via associations in geolocation drives. Short routes due to protocol necessities and physical requirements between an upstream router and a CPE in a network full of CPEs allow upstream provider infrastructure distance, unreachable EUI-64 CPE devices, and even non-EUI-64 devices to be detected. In addition, a wrong CPE association can be caught if there is a huge inaccuracy between the result and the given truth, which may come up in examples such as virtualized network topology. Each EUI-64 router in the data was probed via a high-speed randomized IP topology prober called Yarrp that shows the same layering of router interfaces and data plane as the Unix command traceroute would. The penultimate hop of a particular EUI-64 CPE is considered the upstream provider router. The centroid of CPEs in each area as well as a minimum radius showcases a probable coverage area for each ISP. Any routers not traceable with IPvSeeYou have infrastructure router clustering which is not as pinpoint location-wise as IPvSeeYou. Figure 3 showcases Pittsburgh and shows the radii of different ISPs that are all within the blue circle which indicates a diverse ISP architecture that is robust in deploying in dense metropolitan areas. No matter what EUI-64 address, any one device can be a huge

vulnerability as random IPV6 addresses in CPE6 do not result in protection due to network propagation, even for legacy devices as they still can be geolocated like in IPvSeeYou.





Figures 2 & 3: Geolocations of Indianapolis, Indiana (left) and Pittsburgh, Pennsylvania (right)

Volunteers were gathered to see if this pursuit had any merit by revealing their LAN IPv6 subnet which allowed Yarrp to gather WAN IPv6 addresses and find nearby CPE devices in other subnets. IPvSeeYou then kicked in and found an offset inference to find more accurate location pins. An Olympia Washington router was used as a ground truth with one non-legacy CPE device being used as ground zero to find other CPE devices in the area. This resulted in a 4.75km distance from neutral ground meaning that IPvSeeYou was validated for that datapoint. Other volunteers' devices resulted in 550 meters to 9 kilometers off from the true centroid. IPvSeeYou was also validated by RIPE Atlas IPv6 probes, which are light measurement nodes installed in homes and networks. Nonowners can gather this location data with an error of less than or equal to one kilometer, which is assumed to be accurate. Only querying to home networks resulted in 3,500 probes which were analyzed via Yarrp to find close CPE routers. The geolocations reveal a median distance of about ten kilometers showing even with a wide spread of RIPE probes in an area, IPvSeeYou was fully accurate. IPvSeeYou could even detect that a probe had been moved since it took its original geolocation position.

EUI-64 IPv6 addresses can be patched by a dynamic generation of IPv6 addresses which is not modern technology. If left unchecked by random IPv6 addresses along with sequential MAC address assignments, a correlation can form in geolocation from potential attackers. Vendors were alerted of 7 million EUI-64 MAC addresses that can be compromised which one vendor will fix with state-of-the-art randomized IDDS when WAN IPv6 addresses are formed. Another vendor is critical of the study stating that their devices were exposing MAC addresses via EUI-64, which can be disproved. Other mitigations involve scrambling MAC address allocation bytes and keeping track of duplicate MAC addresses, allowing for BSSIDs to be uncorrelatable. This solution is still rooted in EUI-64 SLAAC addressing and thus is not the best solution. The random MAC address can also only apply to WAN MAC or the 802.11 BSSID on each power cycle or a new EUI-64 address generation, to prevent the correlation. This is tough to implement due to its complexity and redundancy from just randomizing both and due to similar time changes can result in an Evil-Twin attack. This involves creating a fake router that unsuspecting users connect to resulting in said user being compromised [5]. ICMPv6 pings can also be deprecated to obtain a CPE WAN address, but this will interfere with troubleshooting network problems via this feature and some cheap CPEs cannot disable ICMPv6. The best solution is to randomize all IPv6 addresses on all CPE devices otherwise this vulnerability will stay still.

Potential applications lie with the Internet Service Providers and router manufacturers. Internet Service Providers can see how in detail the compromise propagates. They may be inclined to go in and re-evaluate their network architecture, both on a metropolitan scale and on a more micro scale in line with home networks. Hurdles can be made to potential troublemakers by implementing traffic check algorithms to see if the packets being sent across the network are

legitimate. The actual vendors can utilize the IPvSeeYou algorithm in a small or big test network that can determine how the CPE MAC addresses become compromised. Through this study, they will be able to determine how best to change their addressing methods to prevent the propagation of geolocation agents.

Future opportunities for this technological solution involve both the white hat and red hat perspectives. The white hat division would be able to evaluate this on many other CPE devices and see whether the protocol would compromise other devices and locations. It can be evaluated to see where in the network, or one device's network interfaces can be improved upon. Design considerations involve implementing data-checking algorithms to make sure nothing fishy is flying through, confidentiality checks to identify if a user is meant to access the information, or a tighter network interface redesign that emphasizes random MAC address assignment without relying on existing patterns. Security architects on the hardware level may also utilize encryption algorithms and theory to determine how to make a random sequence that will be unable to be decrypted by the red hat division. The red hat division will have full reign with this study in determining how further to break networks. Should they gather the source code, whether, through illegal means of stealing or through legal means such as reverse engineering, they will be able to utilize this program to find vulnerable devices and start attacking these servers. The source code is not given in this paper, but gathering or reverse engineering can lead CPE devices to be vulnerable.

References

[1] E. C. Rye and R. Beverly, "IPvSeeYou: Exploiting Leaked Identifiers in IPv6 for Street-Level Geolocation," 2023 IEEE Symposium on Security and Privacy (SP), San Francisco, CA, USA, 2023, pp. 3129-3145, doi: 10.1109/SP46215.2023.10179376.

[2] Broadcom, “Broadcom BCM3390 DOCSIS3.1 modem/gateway SoC,” https://www.broadcom.com/products/broadband/cable/modems/bcm3390.

[3] I. Poese, S. Uhlig, M. A. Kaafar, B. Donnet, and B. Gueye, “IP Geolocation Databases: Unreliable?” ACM SIGCOMM Computer Communication Review, vol. 41, no. 2, pp. 53–56, 2011.

[4] D. Komosn`y, M. Voz˘n´ak, and S. U. Rehman, “Location Accuracy of Commercial IP Address Geolocation Databases,” 2017.

[5] “Evil twin attack: What it is and how to prevent it.” NordVPN. https://nordvpn.com/blog/evil-twin-attack/ (accessed Oct. 19, 2023).