Measured Approaches to IPv6 Address Anonymization and Identity Association

David Plonka
Akamai Technologies
plonka@akamai.com

Arthur Berger
Akamai Technologies
Massachusetts Institute of Technology
arthur@akamai.com

ABSTRACT

IPv6-based attacks in the Internet today pose challenges that differ substantially from IPv4-based attacks in two facets of attack response: *(1)* sharing IP address-related information to inform coordinated efforts, while still protecting the privacy of victims and *possible* attackers, and *(2)* mitigating abuse by altering treatment, *e.g.,* dropping or rate-limiting, of only victim's and/or attacker's packets. Meeting these challenges depends on knowledge or assumptions about IP address identities, typically in the form of a public, globally-routed IP address *prefix* – the *Identity Associations* (or IAs) – of the victimized or attacking parties. IA discovery, especially remotely, is complicated by the ephemeral nature of many active IPv6 addresses and the freedom operators have in associating identities given the unconstrained IPv6 address resource.

Recent research reports introduce IPv6-specific approaches to address *anonymization* and address *association* identification. We propose these methods as preferred practices in coordinated attack response and invite community feedback.

1. MOTIVATION & INTRODUCTION

Both protecting personally identifiable information (PII) in the form of IP addresses and identifying IP address associations, *e.g.,* with operators, users, or network elements, in the face of attack, warrants special attention with IPv6 due *(a)* to nascent privacy concerns and mandates, *e.g.,* in the European Union, and *(b)* to increased IPv6 use, worldwide. Given today's significant IPv6 deployment and dual-stack operation, the IPv6 address may be the identifier most likely to be unique to a client or server on the World-Wide Web (WWW). While individual IPv4 addresses are increasingly shared due to address exhaustion, such sharing is neither intended nor commonplace with IPv6 which offers unique, globally-routed addresses end-to-end. This note involves two recent research results that introduce IPv6-specific approaches to address *anonymization* and, conversely, address *association.* While these have quite different applications, we wish to highlight how the two are interrelated and how they are pertinent to coordinated response to network abuse or attacks.

We propose the reader join in considering these questions: First, how can passive and active Internet measurements inform decisions about address anonymization and identity association? Second, is there reason to believe that any one IP prefix length would perform satisfactorily for either? Third, in the face of attack, when, where, and how should IP addresses be deaggregated or coalesced to effectively associate them with victims or attackers?

1.1 Address Anonymization

As a privacy measure, such as *k*IP presented by Plonka and Berger [6], anonymization by address *truncation* means simply to delete a set of contiguous low (rightmost) bits, *i.e.,* to remove a suffix from an input address. Typically the suffix' bits are replaced with zeroes so that the anonymized output is an address-sized value. While more complex anonymization techniques have been implemented and are well-studied, *e.g.,* [8], they anonymize addresses in a way that prevents the result from being used for standard security, operations, and research tasks. Specifically, they prevent correlation with network topology, routing, service providers, and locations. For these purposes, truncation-based anonymization is ideal *if, and only if,* it can be guaranteed to improve privacy.

Such anonymization is typically performed by truncating input addresses to one fixed length. Consider, for instance, a WWW analytic system employing truncation-based IP address anonymization; *e.g.,* zeroing the last 8 bits of a user's IPv4 IP address and the last 80 bits of an IPv6 address [3]. Essentially, this is equivalent to masking or aggregating to /24 and /48 prefixes, respectively, perhaps combining information about as many as 256 IPv4 addresses or 64K IPv6 /64 prefixes. Of course, the utilization of the IPv4 and IPv6 address spaces differ dramatically.

A central problem is how to decide at what prefix (bit) length(s) real addresses should be cleaved into a "public," suitably anonymous prefix to be reported as is and a private suffix to be discarded or obscured, except when necessary in network operations and abuse mitigation. To tackle the problem of determining whether truncated prefixes or aggregates might effectively provide anonymity, *i.e.,* to make an individual appear indistinguishable amongst a set of individuals (see Section 6.1.1. [2]), kIP (passively) counts active addresses to determine how many they actually aggregate. Then, it uses such counting as the basis for anonymization by *variable length* truncation or aggregation, resulting in different lengths to anonymize different areas of the address space. In the paper, the authors evaluate this method of IPv6 address anonymization and demonstrate that truncation to a single prefix length of 48 bits (one existing practice, used Internet-wide), fails to anonymize information associated with individuals' IP address identities, *e.g.,* /64 prefixes. Their results demonstrate how kIP anonymization, having anonymous aggregates of adjustable prefix lengths (varying across the active IPv6 address space) outperforms IPv6 address anonymization by 80-bit truncation.

1.2 Identity Association

Conversely, in IP address identity association the goal is to determine what address or prefix best matches some entity, such that it might effectively be used in configuring appropriate packet treatments, such as access or rate limits when under attack. Typically, an associated entity may be a local area network, an Internet service subscriber, or a piece of user equipment.

Lets consider attacks on anonymity and situations that might call kIP's claimed privacy guarantee into question. As described in [6], kIP treats an address' /64 prefix and anything more specific, *e.g.,* the IID, as private. While it's common for ISPs to provide a /64 prefix to a customer, some ISPs will honor requests for a larger prefix, *e.g.,* a /60 or /56 [4, 5, 7]. Then, the customer's router might advertise multiple SLAAC prefix(es) for their local hosts. In this case, it is possible for an individual customer to have a set of simultaneously-assigned /64 prefixes, resulting in an anonymous aggregate where the number of distinct *customers* therein could number fewer than the k expected). To combat this, an anonymizer wants to know the customer's prefix length, or Identity Association (IA), so that it might increase k accordingly. Efforts to discern the IA and prefix length automatically (*e.g.,* via kIP's passive activity matrix) is ongoing work. Similarly, if a malicious party generates traffic, perhaps via forgery, from what would otherwise be quiescent source addresses in many /64 prefixes, they might cause kIP-anonymization to report more specific anonymous aggregates allowing them to determine what their neighbors' nearest active prefixes might be. For this reason, it may be important to keep time series of simultaneously assigned address counts, so that anomalous counts, *e.g.,* during flash crowds, or attacks, can be identified and/or ignored.

The discerning of an entity's associated IP prefix, and therefore an identity, might be an attack on privacy in one setting, discerning it is a necessary feature in other settings, such as in coordinated mitigation of denial of service attacks. In another recent work, Beverly *et al.* [1] report results worldwide of `traceroute`-like measurement campaigns. In the process, they discover that it is often possible to remotely determine the prefix, or identity, associated with target hosts, based on the IPv6 addresses of routers near the periphery of the Internet. In the paper's "Subnet Discovery" section, they call this the "Identity Association (IA) Hack" and describe how the method works. In coordinated attack investigation and response, having a reasonable notion of the prefix associated with a candidate victim or attacker is useful and likely necessary (*i*) to implement effective mitigations and (*ii*) to avoid causing collateral damage, *e.g.,* by unnecessarily affecting unrelated parties legitimately using nearby IPv6 prefixes.

2. CONCLUSION

There are challenges involving Internet data sharing and abuse mitigation that differ with IPv6, especially in performing IPv6 address anonymization and in discerning IPv6 identity associations. Given these are key operations in coordinated attack response, we offer these nascent methods for community discussion.

3. REFERENCES

[1] R. Beverly, R. Durairajan, D. Plonka, and J. P. Rohrer. In the IP of the Beholder: Strategies for Active IPv6 Topology Discovery. In *Proceedings of the 2018 ACM Conference on Internet Measurement Conference*. ACM, 2018.

[2] A. Cooper, H. Tschofenig, B. Aboba, J. Peterson, J. Morris, M. Hansen, and R. Smith. Privacy Considerations for Internet Protocols. IETF RFC 6973, July 2013.

[3] Google. IP Anonymization in Analytics. https://support.google.com/analytics/answer/2763052, 2017.

[4] John Brzozowski. Personal conversation, 2017.

[5] Lee Howard. Personal conversation, 2017.

[6] David Plonka and Arthur W. Berger. kIP: a Measured Approach to IPv6 Address Anonymization. *CoRR*, abs/1707.03900, 2017.

[7] O. Troan and B. Volz. Issues and Recommendations with Multiple Stateful DHCPv6 Options. IETF RFC 7550, May 2015.

[8] Jun (Jim) Xu, Jinliang Fan, Mostafa H. Ammar, and Sue B. Moon. On the design and performance of prefix-preserving IP traffic trace anonymization. In Vern Paxson, editor, *Proceedings of the 1st ACM SIGCOMM Internet Measurement Workshop, IMW 2001, San Francisco, California, USA, November 1-2, 2001*, pages 263–266. ACM, 2001.