Interim Internet2 IPv6 Netflow Anonymization Policy, v1.0

April 16th, 2010

1. **Introduction**

   Internet2 collects and makes available anonymized IPv4 netflow data from the Internet2 Network for research and other approved purposes. See http://www.internet2.edu/observatory/archive/data-collections.html#netflow With increasing interest in IPv6 and growing production use of that protocol, it is appropriate for Internet2 to also collect and make available anonymized IPv6 netflow data for similar uses.

   At the same time, Internet2 has an obligation to respect the privacy of customer traffic, including the privacy of IPv6 traffic. Thus, any netflow data collected from the Internet2 Network and subsequently released for public analysis must be appropriately anonymized before the community is allowed to have access to it, whether that netflow data relates to IPv4 traffic or to IPv6 traffic.

   This dynamic tension between ensuring researcher access to network data while simultaneously ensuring that privacy-related issues are addressed has been affirmed in the Internet2 Strategic Plan (see http://www.internet2.edu/strategicplanning/ at Task K).

   The purpose of this interim document is to identify and explain the required "anonymization mask length" for IPv6 addresses which will be present in Internet2 IPv6 netflow data.

2. **Current Internet2 IPv4 Netflow Data Anonymization Policy**

   IPv4 netflow data is sanitized by Internet2 by having the low order 11 bits of each IPv4 address zeroed before data is released for analysis, leaving the remaining 21 bits of each 32 bit IPv4 address intact.

   For context, most sites have subnets somewhere in the /23-/25 range, which means that in general while it IS possible to use the masked IP addresses to tie a given netflow record to a particular institution, it is NOT possible to localize IPv4 data down to a unique subnet.

   That level of anonymization is designed to insure that a sufficient amount of user traffic will be inseperably "pooled" or "comingled," thereby precluding the mapping of any given netflow record to a particular user or other identifiable campus activity.

3. **IPv6 Address Allocation Practice**

   IPv6 address allocation practices strongly influence Internet2's IPv6 anonymization mask requirements, so we'll describe those allocation processes here.

   As of 2010, the ARIN Number Resource Policy Manual (see https://www.arin.net/policy/nrpm.html at 6.4.3 and 6.5.4) states that:

   - Local Internet Registries (LIRs) will receive an IPv6 /32 from ARIN
   - large sites will hierarchically receive an IPv6 /48 from their LIR's /32
   - small sites needing only a few subnets over 5 years will get an IPv6 /56 from their LIR's /32
   - if one and only one subnet is neeed, that entity will get just an IPv6 /64 (still a huge number of addresses!)

   Some entities may also get an IPv6 /48 directly from ARIN, see the microallocation policy at https://www.arin.net/knowledge/micro_allocations.html and the ARIN IPv6 direct assignment policy at https://www.arin.net/policy/nrpm.html#six58

   Thus, any given end site may have anything from one (or more) IPv6 /32s to just one /64. That represents a huge potential range of network address allocations, and an impressive "stage" upon which a site's IPv6 users may be dispersed or concentrated.

4. **Recommended IPv6 Anonymization Mask**

In keeping with the IPv4 netflow anonymization policy which allows research data to be localized to a particular site, but NOT to a specific subnet or user, one might think that an IPv6 anonymization mask which would zero somewhere between the low order 65 and 72 bits would be sufficient:

- to protect against identification of a specific subnet, we could zero the low order 65 bits (e.g., zero the low 64 to at least anonymize at the subnet level, then add at least one more bit to insure that the subnet isn't effectively unique),
- or, assuming we were concerned about protecting the anonymity of a few broadly dispersed IPv6 users at a small site which had received a /56, we might want to zero the low order (128-56) 72 bits.

Based on user feedback to an earlier draft of this policy, however, anonymizing with even a 72 bit mask might or might not have been sufficient to adequately address potential privacy concerns. To understand why, recognize that IPv6 adoption on our campuses may currently be sparse, and in fact, there may only be one IPv6 user (or department or lab or other identifiable campus entity) per IPv6 /56. In that case, even anonymizing with a 72 bit mask might not adequately comingle multiple users' IPv6 traffic for traffic anonymization purposes. The possibility of that sort of worst case IPv6 user distribution implies that we need, out of an abundance of caution, to initially use a broader netmask than we otherwise might.

Given the current lack of hard data, the community consensus appears to be that only a (128-48) 80 bit mask (zeroing the low order 80 bits and leaving only the remaining 48 bits of the IPv6 addresses for analysis) can be relied on to adequately protect the privacy of IPv6 user traffic while additional empirical data is collected.

5. **Further Research**

Because we currently lack empirical information about the actual addressing practices of IPv6-connected sites, the anonymization recommendations contained in this document shall be of an interim nature only, lasting no longer than 24 months.

While this interim policy is in effect, Internet2 measurement staff shall empirically assess whether the interim 80 bit IPv6 anonymization mask is necessary, or whether a lesser degree of anonymization would still be sufficient to fully protect user privacy. For example, based on empirically observed data, could the anonymization mask be safely relaxed from 80 bits to 72 bits without negatively impacting user privacy?

For the purpose of conducting this review, and only for that pupose, IPv6 netflow traffic anonymized with only a 64 bit mask may be collected. This data may be referred to as "lightly anonymized IPv6 netflow data."

Access to lightly anonymized IPv6 netflow data shall be limited solely to the Internet2 staff member (or staff members) collecting and analyzing that data, and that data shall be used solely for the purpose of evaluating data anonymization/user privacy requirements.

Lightly anonymized IPv6 netflow data shall be carefully protected from unauthorized disclosure, and the retention of lightly anonymized IPv6 netflow data shall be minimized to the fullest extent possible. When the lightly anonymized data collected for this study are no longer required, and no later than the end of the review period, all lightly anonymized IPv6 traffic data shall be deleted or reanonymized in accordance with the normal (80 bit mask) provisions established in this original interim policy.

Other data relevant to the subject of this study may also be collected. For example, IPv6-enabled sites may be contacted for information related to their addressing plans, or to get community input on anonymization-related requirements.

Acknowledged experts in network measurement, such as staff members from CAIDA, may be

invited by Internet2 to participate in this work, subject to the execution of a binding non-disclosure agreement by those external experts.

Upon completion of the review, or in any event after no more than 24 months, the staff's recommendations for a production (non-interim) anonymization mask length shall be forwarded to Internet2 management for review, and to appropriate Internet2 advisory groups and councils for their consideration and approval.

6. **Special Case IPv6 Addresses: 6to4 Addresses in the 2002::/16 Prefix**

   We also want to note one special case involving IPv6 address anonymization, and that would be the anonymization of IPv6 6to4 addresses (see http://www.ietf.org/rfc/rfc3056.txt).

   6to4 works by using addresses in 2002::/16 with IPv4 source addresses embedded within the 6to4 IPv6 addresses. Protecting those embedded IPv4 addresses will require the use of a 91 bit anonymization mask to get the equivalent of normal 11 bit anonymization for those "embedded" IPv4 addresses.

7. **Special Case IPv6 Addresses: IPv6 Multicast Addresses**

   IPv6 multicast addresses are defined in http://www.ietf.org/rfc/rfc4291.txt at Section 2.7, and as mentioned in that section, IPv6 multicast addresses will always begin with the hexadecimal digits "FF" (eight bits of leading 1's).

   IPv6 multicast addresses do not require anonymization, and for the purpose of this policy, IPv6 multicast addresses shall not be anonymized.

8. **Netflow v9 MAC Address Fields**

   Netflow Version 9, as defined by http://www.ietf.org/rfc/rfc3954.txt, collects an expanded range of flow-related fields, including SRC_MAC and DST_MAC ("source MAC address" and "destination MAC address"). See section 8 of RFC3954.

   One commentor expressed concern that these MAC addresses might actually be the MAC addresses of the originating and terminating systems, such as the MAC address of a user's PC or the MAC address of a site's web server.

   We don't believe that MAC address data for originating and terminating systems would be available for flows written from Internet2 layer three backbone routing equipment, but, if present at all, would rather most likely be the MAC addresses of immediately adjacent routers, switches, or other directly connected network devices.

   See also fields 56, 57, 80 and 81 ("IN_SRC_MAC", "OUT_DST_MAC", "IN_DST_MAC" and "OUT_SRC_MAC") as described in table 6 of
   http://www.cisco.com/en/US/technologies/tk648/tk362/technologies_white_paper09186a00800a3db9

   Because the RFC is not definitive on this point, however, and because the interpretation of those fields may vary from vendor to vendor, or be impacted by any given site's architecture, again, from an abundance of caution, for the purposes of this interim policy any values in the Netflow v9 MAC address fields shall be zeroed before IPv6 flow data is released for use by community researchers.

   In addition to the 80 bit mask study described in section 5 of this interim policy, Internet2 staff shall also conduct a study of unanonymized Netflow v9 MAC address data for the purpose of definitively understanding how those fields are populated on Internet2 equipment generating Netflow v9 records.

9. **Summary of Interim Anonymization Policy Recommendation**

   Zero the low order 80 bits of all IP address in all IPv6 netflow records before publicly releasing

those records for research use or other approved purposes.

Exceptions to that general rule are:

- In the case of netflow records containing 6to4 traffic associated with the 2002::/16 netblock, zero the low order 91 bits in order to anonymize embedded IPv4 addresses in accordance with existing IPv4 netflow anonymization policies before publicly releasing those records for research use or other approved purposes.
- IPv6 multicast addresses (e.g., IPv6 addresses beginning with the hex digits "FF") shall be disclosed without anonymization.

Internet2 measurement staff shall do a study of the need for an eighty bit IPv6 anonymization mask, and the extent to which Netflow v9 MAC address fields may be disclosed without impacting user privacy.

10. **Alternative Access**

    A researcher whose data access needs are not met by this interim policy may ask Internet2 to consider releasing data in an alternative format. Requests for exceptions to this default policy should be sent to dataaccess@internet2.edu

11. **Disclaimers**

    All Internet2 activities are governed by the Internet2 Intellectual Property Framework (see http://www.internet2.edu/membership/ip.html).

12. **Acknowledgements**

    Thanks to Jay Ford, Michael Lambert, Bill Cerveny, David Farmer, Richard Machida, kc claffy, the Internet2 RAC and AOAC advisory councils, and all the other individuals who commented and offered suggestions related to draft versions of this document, although all errors or other problems remain solely the responsibility of the author.

13. **Policy Approvals/Endorsements**

    RAC endorsed this policy on March 5th, 2010.
    AOAC endorsed this policy on March 24th, 2010.

14. **Revision History**

    v0.1 Initial draft recommending a 72 bit mask
    v0.2 Revised draft revised to use an 80 bit mask for most IPv6 traffic, and a 91 bit mask for 6to4 traffic
    v0.3 Revised draft emphasizing interim nature of the policy, and adding a research effort to study whether the 80 bit mask can safely be relaxed.
    v0.4 Revised draft to include Netflow v9 MAC address field-related concerns
    v0.5 IPv6 multicast address section added.
    v0.6 Alternative access advisory paragraph added per the request of the Internet2 Network Research Review Committee (NRRC).
    v1.0 Policy approvals/endorsements added, draft tags removed.

15. **Feedback and Comments**

    Feedback or comments relating to this policy are welcome, and can be sent to Joe St Sauver, (joe@oregon.uoregon.edu or joe@internet2.edu)

The URL for this document is http://www.internet2.edu/policies/ipv6-mask.html and it was last modified on April 16th, 2009.