Benchmarking Methodology for Stateful NATxy Gateways using RFC 4814 Pseudorandom Port Numbers

draft-lencse-bmwg-benchmaring-stateful

Gábor LENCSE lencse@sze.hu (Széchenyi István University) – presenter

Keiichi SHIMA keiichi@iijlab.net (IIJ Innovation Institute)

IETF 111, BMWG, July 26, 2021.

Why do we need a new measurement method?

- RFC 2544 has defined a benchmarking methodology for network interconnect devices.
- RFC 5180 addressed IPv6 specificities and it also provided a technology update, but excluded IPv6 transition technologies.
- RFC 8219 addressed IPv6 transition technologies, including stateful NAT64.
- However, none of them discussed how to apply RFC 4814 pseudorandom port numbers to any stateful NAT (NAT44, NAT64, NAT66) technologies.

What is this draft about?

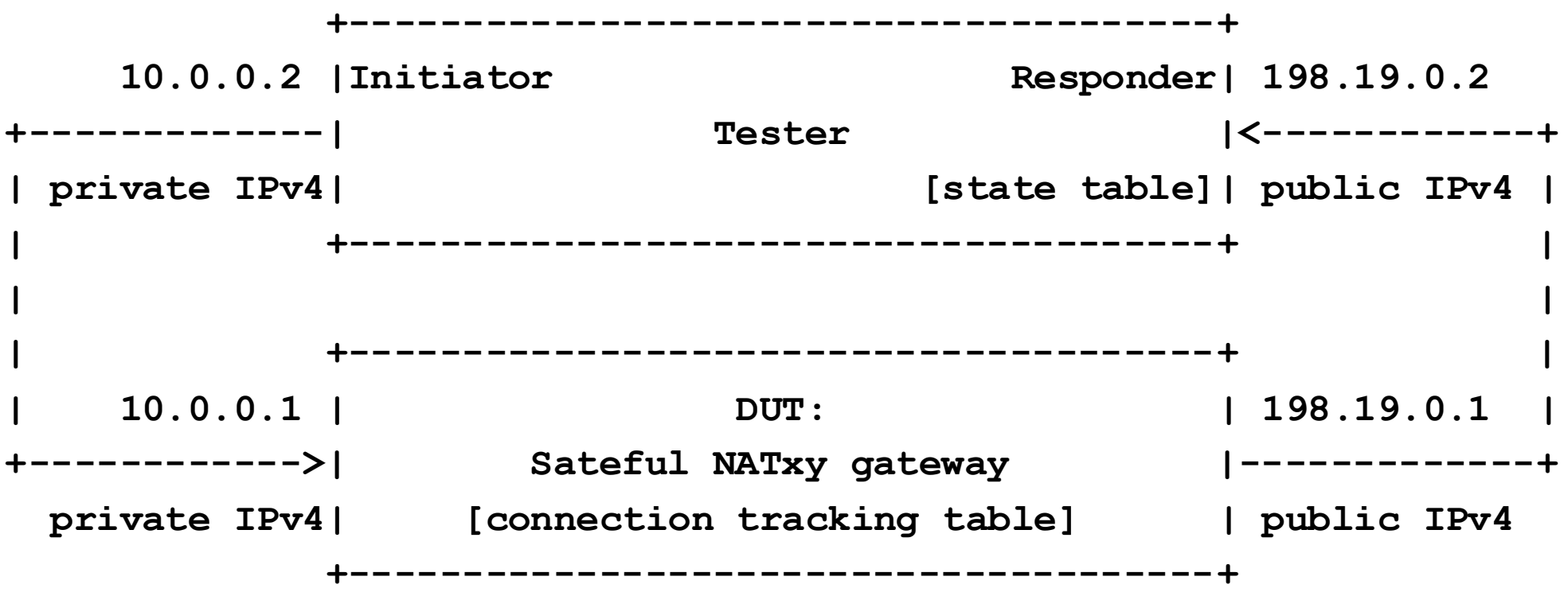
- We discuss why using pseudorandom port numbers with stateful NAT gateways is a hard problem
- We recommend a solution including
  - Test setup and terminology
  - A two-phase measurement method
  - Limitations

RFC 4814 Port Numbers and Stateful NAT

- Using RFC 4814 pseudorandom port numbers
  - In the private to public direction, it is theoretically possible, but
    - The recommended source port range is: 1024-65535, thus its size is: 64512.
    - The recommended destination port range is: 1-49151, thus its size is: 49151.
    - The number of source and destination port number combinations is: 64512*49151 = 3,170,829,312.
    - It would be a DoS (Denial of Service) attack against the connection tracking table!
  - In the public to private direction, it is not feasible (at least directly), because the stateful NAT gateway drops any packets that do not belong to an existing connection
    - But it is feasible by reusing previously observed four tuples!

Test Setup

- Methodology works with any IP versions
  - To facilitate easy understanding, we use the example of stateful NAT44



Terminology

- Connection tracking table
  - The stateful NATxy gateway uses a connection tracking table to be able to perform the stateful translation in the public to private direction.
  - Its size, policy and content is unknown for the Tester.
- Four tuple
  - The four numbers that identify a connection are: source IP address, source port number, destination IP address, destination port number.
- State table
  - The Responder of the Tester extracts the four tuple from each received test frame and stores it in its state table.

Terminology

- Initiator
  - The port of the Tester that may initiate a connection through the stateful DUT in the private to public direction.
    - If the used four tuple does not belong to an existing connection, the DUT will register a new connection into its connection tracking table.
- Responder
  - The port of the Tester that may not initiate a connection through the stateful DUT in the public to private direction.
    - It may send only frames that belong to an existing connection.
    - To that end, it uses four tuples that have been previously extracted from the received test frames and stored in its state table.

Terminology

- Preliminary test phase
  - Test frames are sent only by the Initiator to the Responder through the DUT to fill both the connection tracking table of the DUT and the state table of the Responder.
  - This is a newly introduced operation phase for stateful NATxy benchmarking.
- Real test phase
  - The actual test (e.g. throughput, latency, etc.) is performed in this phase after the completion of the preliminary test phase.
  - Test frames are sent as required (e.g. bidirectional test or unidirectional test in any of the two directions).

Benchmarking Method

- Restricted port number range
  - The Initiator SHOULD use restricted ranges for source and destination port numbers (to avoid DoS attack against the connection tracking table of the DUT)
  - The size of the source port number range SHOULD be larger
    - e.g. in the order of a few times ten thousand
  - The size of the destination port number range SHOULD be smaller
    - may vary from a few to several hundreds or thousands as needed.
  - The product of the sizes of the two ranges can be used as a parameter.
    - The performance of the stateful NATxy gateway MAY be examined as a function of this parameter.

Benchmarking Method

- Preliminary test phase
  - It serves two purposes
    - The connection tracking table of the DUT is filled.
      - It is important, because its maximum connection establishment rate may be lower than its maximum frame forwarding rate (that is throughput).
    - The state table of the Responder is filled with valid four tuples.
      - It is a precondition for the Responder to be able to transmit frames that belong to connections exist in the connection tracking table of the DUT.
  - It is always necessary before the real test phase
  - It can be used without the real test phase to measure the maximum connection establishment rate

(The details of operation can be found in Section 4.2 of the draft.)

Benchmarking Method

- Measurement of the Maximum Connection Establishment Rate
  - The maximum connection establishment rate is an important characteristic of the stateful NATxy gateway
    - Worth measuring and publishing by itself
  - Its determination is necessary for the safe execution of the preliminary test phase (without frame loss) before the real test phase.
  - Its measurement procedure is very similar to the RFC 2544 throughput measurement procedure
    - It is defined in Section 4.4 of our draft

Benchmarking Method

- Real test phase
  - It MUST be preceded by a preliminary test phase during which the frame rate is not higher (or somewhat lower) than the maximum connection establishment rate
  - The actual measurement procedure (throughput, frame loss rate, latency, PDV, IPDV) is performed as defined in RFC 8219
  - Please refer to Section 4.5 of our draft for further details

Limitations

- As for transport protocol, we recommend UDP
  - UDP is recommended in: RFC 2544, RFC 5180 and RFC 8219
  - However Stateful NATxy solutions handle TCP and UDP differently
    - E.g. **`iptables`** uses 30s timeout for UDP and 60s timeout for TCP
    - Thus benchmarking results produced using UDP do not necessarily characterize the performance of a NATxy gateway well enough, when they are used for forwarding Internet traffic
    - As for the given example, timeout values of the DUT may be adjusted, but it requires extra consideration
- As for higher layer protocols, we recommend
    - https://datatracker.ietf.org/doc/html/draft-ietf-bmwg-ngfw-performance

Benchmarking Methodology for stateful NATxy gateways using RFC 4814 Pseudorandom Port Numbers

- It has been proposed in:
  - G. Lencse, K. Shima, "Benchmarking Methodology for Stateful NATxy Gateways using RFC 4814 Pseudorandom Port Numbers", Internet Draft, May 17, 2021, draft-lencse-bmwg-benchmarking-stateful-00
    - https://datatracker.ietf.org/doc/html/draft-lencse-bmwg-benchmarking-stateful
- One partial implementation exists:
  - https://github.com/lencsegabor/siitperf/tree/stateful , documented in:
    - Lencse, G., "Design and Implementation of a Software Tester for Benchmarking Stateful NAT64 Gateways: Theory and Practice of Extending Siitperf for Stateful Tests", under review in *Computer Communications*, may be revised or removed without notice, 2021, http://www.hit.bme.hu/~lencse/publications/SFNAT64-tester-for-review.pdf