Internet Engineering Task Force
Internet-Draft
Intended status: Informational
Expires: November 18, 2013

T. Tsou, Ed.
Huawei Technologies (USA)
T. Murakami
IP Infusion
S. Perreault
Viagenie
May 17, 2013

Analysis of Algorithms For Deriving Port Sets
draft-tsou-softwire-port-set-algorithms-analysis-04

Abstract

This memo analyzes some port set definition algorithms used for stateless IPv4 to IPv6 transition technologies. The transition technologies using port set algorithms can be divided into two categories: fully stateless approach and binding approach. Some algorithms can work for both approaches.

Status of this Memo

This Internet-Draft will expire on November 18, 2013.

Copyright Notice

Table of Contents

1. Introduction

IPv6 transition technologies with address sharing can be divided into three categories as suggested in [I-D.softwire-unified-cpe]:

- Fully stateful approach, e.g. [RFC6333]. Stateful solutions do not make use of port sets, and are out of scope for this memo.

- Binding approach, with per-subscriber state, e.g., [I-D.softwire-lw4over6]. This type of algorithm does not embed port set information and IPv4 address in the IPv6 address when doing translation or encapsulation, so a mapping entry is required in the border router. This type of solution gives flexibility in address planning because the IPv4 address is not statically bound to the IPv6 address. To some extent, the binding approach can also be called a partially stateless approach.

- Fully stateless approach, e.g., [I-D.softwire-map]. This type of algorithm embeds port set information and an IPv4 address in the IPv6 address. For a given port number and IPv4 address, the corresponding IPv6 address can be calculated using a limited set of mapping rules rather than a mapping entry per subscriber.

Binding and stateless technologies can significantly simplify the implementation of the border router and reduce resource requirements. In these solutions, a port set is assigned to each CPE, and can be calculated from a port set identifier (PSID) in conjunction with some other parameters. For a given port number, the corresponding PSID can also be derived; that is, the mapping algorithm must be reversible.

Some port set definition algorithms have been proposed to support these technologies. It may be useful to analyze the characteristics of these algorithms for better understanding and to choose a proper algorithm for different needs.

A good port set definition algorithm must be reversible and easy to implement. It must be able to exclude the well-known ports (0-1023). It should be able to define non-continuous or random port sets for the modest gain in security against port-guessing attacks that these provide. For the fully stateless method, the restrictions imposed by the algorithm on the choice of IPv6 addresses for customer equipment should be minimized. To simplify administration, the total number of ports assigned should be roughly the same for each port set derived by the algorithm. Finally, the algorithm should be adaptable to a wide range of address sharing ratios.

This memo will analyze the following characteristics:

- Implementation: implementation complexity, performance, etc.

- Can calculate the port set identifier (PSID) from the port number at the Border Router (BR).

- Can exclude well known ports without excluding other ports.

- Port set type: continuous, non-continuous, random. Continuous port set provides common security, random port set provides good security.

- Stateless: requires per-subscriber provisioning at the BR, yes or no.

- Friendliness for NAT44: comply with NAT44 [RFC5382] or not.

- Sharing ratio: maximum, minimum sharing ratio.

2. Terminology

BR: Border Router.

CPE: Customer Premise Equipment.

GMA: Generalized Modulus Algorithm.

MAP: Map Address and Port.

PSID: Port Set Identifier, one of the key parameters used to derive the set of ports allocated to a given CPE.

3. Various Types of Algorithms

3.1. Binding Approach Algorithms

3.1.1. Mask/Value Algorithm

[RFC6431] defines an option for the PPP Internet Protocol Control Protocol (IPCP) [RFC1332] to allocate port sets to CPEs, as shown in Figure 1. The Port Range Value plays the role of a PSID. The example in [RFC6431] shows the case of a mask selecting a port number prefix, but the mask can be more general.

```
     0                   1                   2                   3
     0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    |M|          Reserved           |      Port Range Value         |
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    |      Port Range Mask          |
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 1: IPCP Option Format For Port Set Identifier (PSID)

[I-D.softwire-lw4over6] also uses this type of port set definition algorithm, for which provisioning is defined in [I-D.sun-dhc-port-set-option]. Figure 2 illustrates the DHCP option.

```
             0                             1
             0  1  2  3  4  5  6  7  8  9  0  1  2  3  4  5
            +--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+
            |   OPTION_PORT_SET     |     option-length     |
            +--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+
            |                Port Set Index                 |
            +--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+
            |                Port Set Mask                  |
            +--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+
```

Figure 2: DHCP Port Set Option Format

The bit-wise AND of port set index and mask can be encoded in an IPv6 address, which will turn it into a fully stateless solution, similar to parameter PSID in other technologies, e.g., MAP [I-D.softwire-map].

The Port Range Value corresponding to a given port can be derived by performing the bit-wise AND of the port number with the Port Range Mask.

```
       0                   1
       0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5
      +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
      |0 0 0 1 0 1 0 0 0 0 0 0 0 0 0 0| Port Range Mask
      +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
             |   |
             |   | (two significant bits)
             v   v
      +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
      |0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0| Port Range Value
      +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+

      +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
      |x x x 0 x 1 x x x x x x x x x x| Usable ports
      +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+      (x may be set to 0 or 1)
```

Figure 3: Example of Port Range Mask and Port Range Value

This algorithm can have some kind of randomization effect by setting different numbers of bits and bits at different locations in the Port Range Mask.

This algorithm may have a problem if the well known ports (0-1023) need to be excluded; it is a bit difficult to achieve that. But if the operator does not have a specific usage for the well known ports, then it is safe to allocate those port to end users, just like other common ports. Some tests have been done to confirm this.

Criterion	Result
Implementation	Easy
PSID from port number	Yes
Port exclusion	Difficult
Port set type	Continuous with prefix, non-continuous otherwise
Stateless	Requires BR to know mask, could be subscriber-independent.
NAT compliance	Care must be taken to avoid port overloading if mask varies between subscribers.
Sharing ratio	Can vary from 1 to 65536 subscribers per address.

Table 1: Evaluation of Mask/Value Algorithm

3.1.2. Cryptographic Algorithm

The cryptographic port set definition algorithm introduced in [RFC6431] can provide very good protection against port guessing attacks, but it is very difficult to derive the port set information, e.g., the starting point, from a given port number. This algorithm can only be used in binding scenarios; the BR must operate in per-subscriber state mode.

```
  0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
 +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
 |M|          Reserved           |          function             |
 +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
 |        starting point         |   number of delegated ports   |
 +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
 |                             key K                           ...
 +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
 ...                                                           ...
 +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
 ...                                                           ...
 +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
 ...                                                             |
 +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 4: Format of the Cryptographically Random Port Range Option

Criterion	Result
Implementation	Difficult
PSID from port number	No (note)
Port exclusion	Difficult
Port set type	Continuous or non-continuous
Stateless	Binding mode only.
NAT compliance	Care must be taken to avoid port overloading.
Sharing ratio	Can vary from 1 to 65536 subscribers per address.

Table 2: Evaluation of Cryptographic Algorithm

Note: it may be possible to find a cryptographic algorithm which can be reversed, e.g. define a reversible one-to-one mapping algorithm. But that is out the scope of this memo. If strong security is required, it may be worth giving this topic further study.

3.2. Fully Stateless: the Generalized Modulus Algorithm (GMA)

Currently there are three drafts supporting the GMA style algorithm: MAP-E [I-D.softwire-map], 4rd-U [I-D.softwire-4rd], and MAP-T [I-D.softwire-map-t], but they are not exactly all the same.

3.2.1. MAP-E

In MAP [I-D.softwire-map], a port set can be defined by the following parameters:

R: sharing ratio;

P: PSID;

M: maximum number of contiguous ports.

To derive the set of port numbers in the port set corresponding to a given PSID value, the following equation can be used:

Port = (R * M) * i + M * PSID + j

where i and j are indices which vary within limits to provide the different port numbers belonging to the port set. The range of i depends on the value (R * M) and the range of j is from 0 to (M - 1).

If (R * M) is less than or equal to 2^15, ports (e.g, the well- known ports 0-1023) can be excluded from the lower end by putting a lower limit dependent on the value (R * M) on index i. In this case, each port set defined by the algorithm consists of a series of ranges of M consecutive port numbers at intervals of (R * M).

On the other hand, if (R * M) is greater than 2^15, the first term drops out of the above equation and a lower limit dependent on the value of M has to be imposed on the value of PSID to exclude the well- known ports. In this case, each PSID is associated with a single range of M consecutive port numbers.

The GMA is easily reversible. For a given port number, the corresponding PSID is given by:

PSID = floor((Port modulo (R * M)) / M))

If R and M are powers of 2, this becomes a mask operation. The mask consists of 'a' high-order zeroes, followed by 'k' ones, followed by 'm' low-order zeroes, where:

$2^a = 65536/(R * M);$

$2^m = M;$

$k = 16 - a - m.$

See Figure 5.

MAP-E defaults to a value of 'a' equal to 6. Thus by constraining the index i to be >= 1, exactly the well-known port range is excluded. Also, each port set consists of 63 equally-sized ranges of consecutive values spaced 1024 ports apart.

```
           0                          8                       15
          +---------------+----------+------+-------------------+
          |       i       |      PSID       |         j         |
          +---------------+----------+------+-------------------+
          |<----a bits--->|<-----k bits---->|<------m bits----->|
```

Figure 5: GMA Bit Representation Of a Port Number When R and M Are Powers Of 2

For a complete explanation of the GMA, see Appendix B of [I-D.softwire-map].

MAP-E embeds the PSID in the End User IPv6 Address provisioned on the customer edge device. See Figure 6. The PSID's location within the address is determined from the Basic Mapping Rule applicable to the subscriber. A mask to extract the PSID from that address is described as follows:

- High-order zeroes in the amount of (n + 32 - r) bits, where n is the length of the IPv6 prefix in the Basic Mapping Rule and r is the length of the IPv4 prefix in that rule.

- Ones in the amount of (r + o - 32) bits, where o is the number of EA bits given by the rule.

- Zeroes for the remaining low-order portion of the address.

This operation is valid only if (r + o) is greater than 32. If not, the IPv4 address or prefix assigned to the subscriber is unshared and the customer edge device can use every port.

```
    |        32 bits           |         |    16 bits        |
    +--------------------------+         +-------------------+
    |  IPv4 endpoint address   |         |  Port in port set |
    +--------------------------+         +-------------------+
    :               :          :           ___/       :
    |    r bits     |32-r bits |          /  q bits   :  q = o - (32-r)
    +---------------+----------+         +------------+
    |  IPv4 prefix  |IPv4  sufx|         |Port Set ID |
    +---------------+----------+         +------------+
                    \          /    ____/    ________/
                      \       :  __/   _____/
                        \     : /     /
    |     n bits         |  o bits   | s bits  |   128-n-o-s bits      |
    +--------------------+-----------+---------+------------+----------+
    |  Rule IPv6 prefix  |  EA bits  |subnet ID|     interface ID      |
    +--------------------+-----------+---------+-----------------------+
    |<---  End-user IPv6 prefix  --->|
```

Figure 6: Structure of the MAP-E End User IPv6 Address

3.2.2. 4rd-U

Everything that was described in the previous section for MAP-E also applies to 4rd-U [I-D.softwire-4rd], with two differences. First, the mapping rule applicable to a particular customer site includes an indication of whether the customer edge equipment is permitted to use the well-known ports or whether they must be excluded.

If the well-known ports are to be excluded, the default value of 'a' (recall Figure 5) is 4 rather than 6. That means that the port set consists of 15 rather than 63 ranges, spaced 4096 values apart. It also means that ports 0-4095 rather than ports 0-1023 are excluded. At an earlier point in time MAP-E had the same default, for which the 4rd-U document provides arguments. However, it was decided that the waste of ports entailed (which implies a 6% reduction in the number of subscribers sharing the same IPv4 address) was a sufficient reason to change. However, see Section 4 for new evidence on this point.

If the well-known ports can be used, the default value of 'a' is zero. That is, the PSID is positioned at the beginning of the port number. As mentioned in the previous section, this implies that subscribers assigned this mapping rule are assigned a single range of consecutive ports. The subscribers assigned the lowest PSID values receive port sets consisting partly or completely of well-known port number values.

3.2.3. MAP-T

MAP-T [I-D.softwire-map-t] uses the same algorithm to assign port sets to customer sites, this time with just one difference. The default value of the offset 'a' is always 4. The consequences in terms of wasted ports were spelled out in the previous section.

3.2.4. Evaluation

This section provides an evaluation of the GMA against our comparison criteria.

Criterion	Result
Implementation	Easy
PSID from port number	Yes
Port exclusion	Easy, but using a value of the offset 'a' between 1 and 5 wastes ports and hence reduces the maximum practical sharing ratio.
Port set type	Continuous for 'a' = 0, non-continuous otherwise
Stateless	No subscriber-specific data required.
NAT compliance	Port sets are guaranteed to be non-overlapping.
Sharing ratio	Equal to 65536/(M * 2^a), where M is the range size for all subscribers sharing the same address. See note.

Table 3: Evaluation of Cryptographic Algorithm

Note: a practical value of the total number of ports in the port set is in the order of 400. Suppose one wants to guarantee each subscriber at least this number of ports. Recall that the number of equal ranges into which the port allocation is divided is equal to 1 for a = 0, 15 for a = 4, and 63 for a = 6. Because of the assumption of equal range sizes, the number of ports M in each range has to be rounded up in the general case to give a total number of ports at least equal to 400. Table 4 shows the consequent impact on sharing ratio. The rounding effect very much dominates the results. If the target were 305 ports instead, the sharing ratio would be the same for all three values of a, since 305 is a multiple of 15 and 63.

a	2^a	# Ranges	Range Size M	Tot. Ports	Ratio R
0	1	1	400	400	163
4	16	15	27	405	151
6	64	63	7	441	146

Table 4: Port Allocations and Range Size For Different Values Of Offset a

In Table 4, the value M is rounded up from the ratio 400/N, where N is the number of separate ranges in the port set. The total number of ports in the port set is this result multiplied by the number of ranges. The sharing ratio is then the stated 65536/(M * 2^a), rounded down to ensure every subscriber sharing the address gets the same number of ports. For a = 0, this ratio would be reduced by 3 to exclude the three ranges containing well-known ports.

4. Conclusion

The Generalized Modulus Algorithm (GMA) clearly comes the closest to satisfying all of our criteria. As the example calculation in Table 4 shows, the sharing ratio is sensitive to the rounding necessary to guarantee at least a certain total number of ports to each subscriber. In this regard, sensitivity will be higher for larger values of the offset parameter 'a', leading to the surprising result that for some ranges of values of the target total number of ports, the sharing ratio will be less for a = 6 than for a = 4 even though the latter wastefully excludes an extra 3072 ports.

The sensitivity of this result to the target total number of ports per subscriber is shown if one assumes that that number is 441 ports. Then the sharing ratio for a = 6 remains at 146, but that for a = 4 drops to 136.

The mask/value algorithm is really a generalization of the GMA. One has the GMA if the one-bits of the mask are constrained to be consecutive. The difference between the binding and fully stateless approaches lies not in the algorithm itself, but in how the algorithm parameters are conveyed to the border router. Binding uses per-subscriber rules. The fully stateless approaches reviewed in this document use a combination of shared mapping rules and information embedded in specially-constructed addresses.

5. IANA Considerations

This memo includes no request to IANA.

6. Security Considerations

The major security consideration related to the subject matter of this document is the vulnerability of port allocation to a port guessing attack. See [RFC6056] for details. The most important factor in countering such an attack is to allocate ports randomly from the assigned port set as they are required by different applications. However, allocating port sets as non-continuous or random entities requires the attacker to go to some extra effort in order to determine the complete port set allocated to a subscriber. Thus resistance to port guessing attacks is improved to a certain degree by allocating non-continuous port sets. For the GMA, this means that non-zero values of the offset value 'a' are to be preferred.

7. References

7.1. Normative References

[RFC5382] Guha, S., Biswas, K., Ford, B., Sivakumar, S., and P. Srisuresh, "NAT Behavioral Requirements for TCP", BCP 142, RFC 5382, October 2008.

[RFC6056] Larsen, M. and F. Gont, "Recommendations for Transport-Protocol Port Randomization", BCP 156, RFC 6056, January 2011.

[RFC6431] Boucadair, M., Levis, P., Bajko, G., Savolainen, T., and T. Tsou, "Huawei Port Range Configuration Options for PPP IP Control Protocol (IPCP)", RFC 6431, November 2011.

7.2. Informative References

[I-D.bsd-softwire-stateless-port-index-analysis]
Boucadair, M., Skoberne, N., and W. Dec, "Analysis of Port Indexing Algorithms", September 2011.

[I-D.softwire-4rd]
Jiang, S., Despres, R., Penno, R., Lee, Y., Chen, G., and M. Chen, "IPv4 Residual Deployment Via IPv6 - A Unified Stateless Solution (4rd) (Work in progress)", April 2013.

[I-D.softwire-lw4over6]
Cui, Y., Sun, Q., Boucadair, M., Tsou, T., Li, Y., and I. Farrer, "Lightweight 4over6: An Extension to the DS-Lite Architecture (Work in progress)", April 2013.

[I-D.softwire-map]
Troan, O., Dec, W., Li, X., Bao, C., Matsushima, S., Murakami, T., and T. Taylor, "Mapping of Address and Port (MAP) (Work in progress)", May 2013.

[I-D.softwire-map-t]
Li, X., Bao, C., Dec, W., Troan, O., Matsushima, S., and T. Murakami, "Mapping of Address and Port using Translation (MAP-T)", February 2013.

[I-D.softwire-unified-cpe]
Boucadair, M. and I. Farrer, "Unified IPv4-in-IPv6 Softwire CPE (Work in progress)", March 2013.

[I-D.sun-dhc-port-set-option]
Sun, Q., Li, Y., Sun, Q., Bajko, G., and M. Boucadair, "Dynamic Host Configuration Protocol (DHCP) Option for Port Set Assignment (Work in progress)", April 2013.

[RFC1332] McGregor, G., "The PPP Internet Protocol Control Protocol (IPCP)", RFC 1332, May 1992.

[RFC6333] Durand, A., Droms, R., Woodyatt, J., and Y. Lee, "Dual-Stack Lite Broadband Deployments Following IPv4 Exhaustion", RFC 6333, August 2011.

Authors' Addresses

Tina Tsou (editor)
Huawei Technologies (USA)
2330 Central Expressway
Santa Clara CA 95050
USA

Phone: +1 408 330 4424
Email: tina.tsou.zouting@huawei.com

Tetsuya Murakami
IP Infusion
1188 East Arques Avenue
Sunnyvale
USA

Email: tetsuya.murakami@ipinfusion.com

Simon Perreault
Viagenie
246 Aberdeen
Quebec, QC G1R 2E1
Canada

Phone: +1 418 656 9254
Email: simon.perreault@viagenie.ca
URI: http://viagenie.ca