Network Working Group
Internet-Draft
Expires: July 30, 2002

M. Bagnulo
I. Soto
A. Garcia-Martinez
A. Azcorra
UC3M
January 29, 2002

Random generation of interface identifiers
draft-soto-mobileip-random-iids-00.txt

Status of this Memo

Copyright Notice

Abstract

This document evaluates the use of random numbers to generate the interface identifier part of an IPv6 address on mobile environments, where Duplicate Address Detection (DAD) mechanisms are expensive. We have estimated the probability of having an address duplication using this mechanism and we conclude that the IPv6 addresses created in this way could be used without previously doing DAD to test the uniqueness of the address in a link.

1. Introduction

In the IPv6 aggregatable address format defined in [1] , the last 64 bits are assigned to the interface identifier. According to [1], in many cases, a link-layer address is used to create the interface identifier part of the IPv6 address, although some other mechanisms are possible when there is no MAC address available. Because there are not guarantees of having a unique link local address, after configuring an interface with an IPv6 address, DAD mechanism must be

performed to ensure that the IPv6 address is unique in the link. RFC 2462 [2] states that "Duplicate Address Detection MUST take place on all unicast addresses, regardless of whether they are obtained through stateful, stateless or manual configuration". DAD is a time consuming process, that usually do not represent a problem for a desktop computer that is initialising.

Mobility support in IPv6 networks will be provided by Mobile IPv6 [3]. Using this protocol, a Mobile Node (MN) that enters a subnet must configure an on-link address in that subnet (the Care-of Address, CoA) before being able to communicate. According to [2], before using the CoA, the MN must perform DAD for that address. But in this case, the time required for DAD is a critical matter because during this time the MN can not communicate and additionally, active communications of the MN are interrupted during this period. This is specially unsuitable for real-time communications. [3] recognizes this problem and states that a MN can decide not to perform DAD, being this is a trade-off between safety and the time needed for the DAD procedure. This document evaluates the use of random numbers to create the interface identifier of the IPv6 addresses, and assesses the risk of using these addresses without previously performing DAD.

Using mechanisms such as fast handovers [4] it is possible to perform DAD in advance, before the MN arrives to the subnet. The Access Router (AR) in the subnet is instructed to perform DAD on behalf of the MN before it enters the subnet. But then, the MN has to wait for the time needed to perform DAD before it can accomplish the handover. This can be a problem in many situations in which the handover is required because the previous layer-2 connection is experimenting difficulties. So we will again benefit for avoiding the previous DAD procedure.

Summarizing, for handover situations the importance of the time required for DAD can not be underestimated. In this document we study the possibility of using random generated interface identifiers to autoconfigure IPv6 addresses. We also reason that DAD can be performed after the node has joined the link because the probability that an address duplication happens is very low.

2. Creation of interface identifiers.

We will study the possibility that the interface identifiers are created randomly, meaning that the host will use a randomly generated 64 bit number as the interface identifier. Actually, only 62 bits of the interface identifier will be generated randomly since, as it is defined in [5], the u bit must be set to "local" and the g bit must be set to "individual". We will now evaluate the probability of collision of two or more randomly generated address identifiers.

The problem: The Address identifiers I1, I2, I3, ..., Ik are a sequence of 62 bit long random variables. Ii are randomly generated. We would like to obtain the probability that two or more Iis collide, i.e. Ii=Ij. This is a very well known mathematical problem that is called the "birthday problem". The solution is:

I1,I2,...,Ik random variables, integer and with uniform distribution between 1 and n (k<=n)

P(n,k)(at least one repeated)=1-(n!)/[(n-k)!.n^k]

(This result is explained in Appendix A)

In our case n=2^62, so the calculation of n! may be more than what a simple calculator can handle. In order to overcome this, we will try to find an upper bound to P(n,k).

P(n,k)= 1-(n!)/[(n-k)!.n^k]

= 1-[(n).(n-1)...(n-k+1)/n^k]

= 1-[(n/n).((n-1)/n)...((n-k+1)/n)]

= 1-[1 .(1-1/n)...(1-(k-1)/n)]

It should be noted that:

i/n =< (k-1)/n when i=1...k-1

then -i/n >= -(k-1)/n when i=1...k-1

then 1-i/n >= 1-(k-1)/n when i=1...k-1

then considering that k<n so that 1-i/n >0 when i=1...k-1

(1-1/n)(1-2/n)...(1-(k-1)/n)>=(1-(k-1)/n)^(k-1)

then -(1-1/n)(1-2/n)...(1-(k-1)/n)=<-[(1-(k-1)/n)^(k-1)]

then 1-(1-1/n)(1-2/n)...(1-(k-1)/n)=<1-[(1-(k-1)/n)^(k-1)]

then P(n,k)=< 1-[(1-(k-1)/n)^(k-1)] = 1-[(n-k+1)/n]^(k-1)

then P(n,k)=< [n^(k-1)-(n-k+1)^(k-1)]/[n^(k-1)]=B

n! is not present in this bound so B is easier to calculate.

In order to quantify the result we will make a few calculations: Remembering that n is the number of possible addresses an k is the number of interfaces in the same link, we will evaluate the upper bound the following values of k

P(2^62,20)=< 7.8 e-17

P(2^62,100)=< 2.1 e-15

P(2^62,500)=< 5.4 e-14

P(2^62,1000)=< 2.2 e-13

P(2^62,5000)=< 5.4 e-12

In order to fully understand the magnitude of the probabilities above, we could compare them with other probabilities.

For instance, according to Table 1.1 of [6], the probability of being killed by a lightning (per day) is about 1.1 e-10. Then, a mobile phone user should be more worried about being killed by a lightning than to have an interface identifier repeated.

We would also like to compare the probabilities above with some issues that affect communication in a similar fashion that address

collision such as the probability of failure of the network equipment.

In case a network equipment fails, communication will be lost until it is replaced, having a similar effect to the one of having a repeated interface identifier. In order to quantify the probability of a network equipment failure, we will estimate it as:

P(NE failure)=MTTR/(MTBF+MTTR)

Being MTTR: Meat Time To Repair and MTBF: Mean Time Between Failures

Network equipment can have an MTBF of 300,000 hours and let's suppose that some backup equipment is available and that MTTR is 0,1 hour (6 minutes).

So P(NE failure)= 3,3e-7.

We can see that P(NE failure) is much more higher than P(n,k).

Besides hardware malfunctioning, network connectivity can be affected by operation errors. Usually, this type of problems are much more frequent than hardware outages, but we do not have any hard data available.

We think that it also interesting to estimate the probability of collision over a year of usage of the system. As we stated above, P(n,k) is the probability of a collision of two or more Interface identifiers when there are k interfaces in the same link. In order to quantify the probability of collision of a user during a year using the system, we will calculate the probability of one or more collision when a user joins m different networks.

The probability of NOT having a collision is Pnot(n,k)=1-P(n,k)

Then the probability of not having a collision after joining m different networks is Pnot(n,k,m)=[1-P(n,k)]^m.

Then the probability of having a collision after joining m different networks is: P(n,k,m)=1-[[1-P(n,k)]^m]

According to the bound found earlier:

P(n,k)=<B

Then -P(n,k)>=-B

Then 1-P(n,k)>=1-B

As P(n,k)and B are less than 1, (1-P(n,k))^m>=(1-B)^m

Then 1-[(1-P(n,k))^m]=<1-[(1-B)^m]

Then P(n,k,m)=<1-[(1-B)^m]

If we consider m=50.000, this means about 140 handovers per day,

P(2^62,500,50000)=< 2,7e-9

P(2^62,5000,50000)=< 2,7e-7

Considering that each time there is a collision, there are two users affected (not considering collision of 3 or more for this estimation), this means that in the case users make 140 handovers per

day in networks containing 500 interfaces, there will be 6 users out of 1.000.000.000 that will have a problem with the communication during this year. In the case users make 140 handovers per day in networks containing 5000 interfaces, there will be 6 users out of 10.000.000 that will have a problem with the communication during this year.

3. Random generation of Interface identifiers..

Another relevant issue is how to generate a 62 bit long random number. In some cases, such as laptops, a random number generator can be available on the system, but in other cases, such as mobile phones, it may not. In such cases, it should be noted that because of the randomness of the identifier, it is not necessary to create the identifier in real time, i.e. when the node joins the network, but the identifier can be already created (such as the day of birth in the birthday problem). What is more, this random number can be pre-configured in the interface driver and the node can use always the same number without changing the probabilities calculations made above. This reduces the needed complexity in the nodes.

4. Conclusions.

In this draft, we have studied the possibility of using random generated numbers for the interface identifier part of the IPv6 addresses. We have also calculated the probability of address collision when interface identifiers are generated this way. After evaluating this probability in several different scenarios, we have found that collision probability with random generated Interface Identifiers is lower enough to consider this mechanism as a possible option. Considering that DAD mechanism is time consuming when time is critical i.e. mobile communications, we think random generation of Interface identifier part of IPv6 addresses is an attractive option because it can be used without previously doing DAD. Whether

DAD should be performed after or it can be avoided, needs further study. For now, we think that performing DAD after joining the link is needed.

5. Security Considerations.

This draft does no introduce new security risks. What's more, random generation of interface identifiers can enable improved anonymity features by allowing interface identifier generation each time a node joins a new link, what can prevents tracing.

6. Acknowledgments.

This work has been done under IST projects LONG and Moby Dick

7. Appendix A: The Birthday Problem

The problem: we want to calculate the probability that in a group of k people, al least two of them have the same birthday.

We model the birthday as a integer random variable, with uniform distribution between 1 and 365 (we forget about the 29th of February, sorry about that :-(

The solution: Let's find out the number N of ways that we can have k values with no duplicates. We can choose 365 values the first time, 364 the second time,....

So, N=365*364*...*(365-k+1)

If we remove the restriction that there are no duplicates, the number of possibilities is 365^k.

So the probability of not having a collision is:

Pnot= 365!/[(365-k)!*365^k

Then the probability of having at least one duplicate is:

P=1-Pnot=1-365!/[(365-k)!*365^k

The general case it would be:

P(n,k)= 1-n!/[(n-k)!*n^k with n>k

References

[1] Hinden, R., O´Dell, M. and S. Deering, "An IPv6 Aggregatable Global Unicast Address Format", RFC 2374, July 1998.

[2] Thomson, S. and T. Narten, "IPv6 Stateless Address Autoconfiguration", RFC 2462, December 1998.

[3] Johnson, D. and C. Perkins, "Mobility Support in IPv6", Internet draft Work in progress, July 2001.

[4] Dommety, G., "Fast Handovers for Mobile IPv6", Internet draft Work in progress, July 2001.

[5] Hinden, R. and S. Deering, "IP Version 6 Addressing Architecture", RFC 1998, 1998.

[6] Schneier, B., "Applied cryptography", Wiley ISBN 0-471-12845-7, 1996.

Authors' Addresses

Marcelo Bagnulo
Universidad Carlos III de Madrid
Av. Universidad 30
Leganes, Madrid 28911
SPAIN

Phone: 34 91 6249500
EMail: marcelo@it.uc3m.es
URI: http://www.it.uc3m.es

Ignacio Soto
Universidad Carlos III de Madrid
Av. Universidad 30
Leganes, Madrid 28911
SPAIN

Phone: 34 91 6249500
EMail: isoto@it.uc3m.es
URI: http://www.it.uc3m.es

Alberto Garcia-Martinez
Universidad Carlos III de Madrid
Av. Universidad 30
Leganes, Madrid 28911
SPAIN

Phone: 34 91 6249500
EMail: alberto@it.uc3m.es
URI: http://www.it.uc3m.es

Arturo Azcorra
Universidad Carlos III de Madrid
Av. Universidad 30
Leganes, Madrid 28911
SPAIN

Phone: 34 91 6249500
EMail: azcorra@it.uc3m.es
URI: http://www.it.uc3m.es

Full Copyright Statement

Acknowledgement

Funding for the RFC Editor function is currently provided by the Internet Society.