Randomized MAC Address Conflict Analysis and Implications

By: Baw Chng <baw@bawman.com>
Initially authored on 2021 May 23, revised 2021 May 31
Find this paper online: https://www.bawman.com/BAWMAN/articles/RandomMAC/

EXECUTIVE SUMMARY

- Popular local area network (LAN) and personal area network (PAN) technologies (e.g., Wi-Fi, Ethernet, Bluetooth) rely on devices using unique MAC addresses to function correctly.
- As more device platforms introduce changes that make more frequent use of randomized MAC addresses, the probability of identical MAC addresses colliding in systems also increases.
- With reasonable assumptions, we estimate the probability of MAC address conflict as a function of the number of randomly generated locally administered MAC addresses as follows:

Number of randomly generated locally administered MAC addresses in a system	**Probability of MAC address conflict in the system**
Roughly 38 thousand	0.001% (OK “five 9’s” of the time)
Roughly 120 thousand	0.01%
Roughly 380 thousand	0.1%
Roughly 1.2 million	1%
Slightly below 4 million	10%
Slightly below 10 million	50%
(Beyond roughly 30 million, MAC address conflict becomes a statistical certainty)	**Estimated number of non-unique MAC addresses in the system**
50 million	36
100 million	142
250 million	888
1 billion	14211

- Conflict probability increases significantly if devices use per-OUI randomization. When limited to a single OUI, 5000 randomized MAC addresses are sufficient to yield a 50% conflict probability, 15000 randomized MAC addresses are sufficiently to statistically guarantee a conflict.
- This paper also discusses the potential impacts of randomized MAC address conflicts, some potential mitigation measures, and the attendant issues with these mitigation measures.
- Smaller, simpler systems are unlikely to be materially affected by MAC address randomization.
- Larger systems are more likely to be materially affected, though opportunities exist to mitigate the effects of randomized MAC addresses, at least for the data forwarding aspects.
- Outside of the data forwarding aspects of systems, there are few technical standards governing the use of MAC addresses, thus the impact of MAC address conflict on the non-data forwarding aspects of larger systems needs to be examined on a case-by-case basis.

TABLE OF CONTENT

INTRODUCTION: Brief Background for the Casual Readers

(Industry professionals can probably skip this section)

What are MAC addresses? MAC addresses are used by many popular digital communications technologies to identify parties in local communications networks. For example, Wi-Fi, Bluetooth, and Ethernet all use MAC addresses. A typical smartphone or tablet with Wi-Fi and Bluetooth capabilities would have at least two MAC addresses, one for Wi-Fi, one for Bluetooth. A typical laptop computer with Ethernet capabilities would also have a MAC address for Ethernet.

What do MAC addresses look like? For most popular technologies like Wi-Fi, Bluetooth, and Ethernet, each of their MAC addresses is 48 bits long, typically written in the form of `xx:xx:xx:xx:xx:xx` or `xx-xx-xx-xx-xx-xx` where each `x` can be a digit ranged from `0` to `9` or a letter ranged from `A` to `F`.

How are MAC addresses typically assigned? For most typical consumer products, manufacturers and component suppliers are assigned blocks of MAC addresses in a globally coordinated fashion. Then the manufacturers and component suppliers in turn assign MAC addresses from their respective assigned blocks to their products in a sequential fashion, thus virtually guaranteeing that no two pieces of equipment would use the same MAC address. Nonetheless, various technology standards that use MAC addresses also allow for MAC addresses to be locally administered or locally generated. Virtual machines, for example, use locally administered MAC addresses to let their virtual network interfaces communicate in local networks. Until recently, the use of locally administered MAC addresses has largely been hidden from the view of the casual consumers.

Why randomize MAC addresses for consumer devices now? What changed? In a word: Privacy. As more consumers carry their mobile devices with them and use Wi-Fi and Bluetooth wherever they go, as technologies and systems become increasingly capable of correlating information across different networks, it becomes increasing feasible to “track” a consumer across time and space if the consumer always carries around a device that always reports the same MAC address. For example, if Joe Consumer carries his smartphone with him all the time, and Joe Consumer has a routine where he stops by a particular coffee shop every morning where he also enjoys the free Wi-Fi at the coffee shop, and he goes to a particular gym every Tuesdays and Thursdays and Saturdays where he also uses the free Wi-Fi at the gym, then it would be fairly easy to piece together Joe Consumer’s routine from the coffee shop and the gym’s Wi-Fi records if Joe Consumer’s smartphone always uses the same MAC address. To give consumers a way to avoid being “tracked” in this manner, in recent years popular mobile device platforms have introduced various ways to randomize the MAC addresses used by mobile devices. If Joe Consumer’s smartphone were to use different, randomized MAC addresses with different networks, the coffee shop’s Wi-Fi network and his gym’s Wi-Fi network would “see” different MAC addresses and it would be harder for third parties to deduce from Wi-Fi usage records that Joe Consumer goes to those two places. If Joe Consumer’s smartphone were to use a different randomized MAC addresses every day, then it would also be harder for third parties to deduce from Wi-Fi usage records that Joe Consumer goes to the same coffee shop every day or that Joe Consumer goes to the same gym on Tuesdays, Thursdays, and Saturdays.

Is MAC address randomization going to be an issue? For the typical consumers, no. It is highly unlikely that randomizing MAC addresses will adversely impact individual consumers. In all likelihood, the typical consumers can continue to use Wi-Fi, Ethernet, and Bluetooth without worrying about MAC address randomization. Depending on the size of the system and what the system does with those MAC addresses (other than to facilitate basic communications for the end-users), randomized MAC addresses may introduce complications for the system operators. The matter discussed in the remainder of this paper is tailored more towards these system operators and their technology suppliers.

How to tell if a unicast MAC address is randomized or locally administered: Examine the second digit in the written MAC address x**X**:xx:xx:xx:xx:xx. If that second digit is a **2**, a **6**, an **A**, or an **E**, then the MAC address is a locally administered MAC address. A locally administered MAC address may or may not have been randomized, but a randomized MAC address must be a locally administered MAC address to be standards compliant.

ANALYSIS: How Likely Are Randomized Locally Administered MAC Addresses to Conflict?

In this context, "conflict" simply means two (or more) device interfaces that use MAC addresses end up using the same MAC address.

How "big" is the problem?

Of the 48 bits used to specify an EUI-48 MAC address, one bit is reserved to indicate whether the address is a unicast or a multicast address, one other bit is reserved to indicate whether the address is a globally unique or a locally administered MAC address. For the base analysis, we focus only on unicast addresses, and we assume that there is no other restriction. Thus, a randomly generated locally administered unicast MAC address can freely use all the remaining 46 bits, yielding an address space of 2^{46}, just above 70 trillion addresses.

The question then becomes: given an address space of 70 trillion and a variable number of devices randomly generating MAC addresses in that space, what are the chances that one would find two (or more) devices using the same MAC address?

The answer to that question obviously depends on how many devices are there generating random MAC addresses. If there are only two devices, obviously the chance of conflict is one in 70 trillion. But as the number of devices increases *linearly*, the probability of conflict increases *exponentially*. The problem is not merely "*what are the chances that any one of the multiple devices' randomized MAC addresses would be the same as one particular device's randomized MAC address*", but "*what are the chances that any one of the multiple devices' randomized MAC addresses would be the same as any other device's randomized MAC address*." It is not a "*many to one*" comparison, but a "*many to many*" comparison.

Borrowing from the Birthday Problem

To answer the randomized MAC address conflict question, we turn to the Birthday Problem, one that is commonly discussed in many introductory *Probability and Statistics* classes. Briefly, the Birthday Problem asks, in a room of N people, how likely is it to have two who have the same birthday? Obviously, if there are only two people in the room, the probability of those two people having the same birthday is one in 365 (ignoring leap year and February 29). As we put more people in the room, the probability of finding two people with the same birthday also increases. But somewhat unintuitively, to get to 50% chance of finding any two people with the same birthday, on average we only need to put 23 people in the room, not 365 ÷ 2 ≈ 182 people.

The Birthday Problem is well known and extensively studied. Assuming that a year has precisely 365 days (ignoring leap years and February 29), and assuming that the general population's birthdays are evenly distributed, the Birthday Problem has a formulaic solution that says that the probability *p* of finding shared (non-unique) birthdays among a room of *n* people is essentially $p(n) = 1 - \frac{365!}{365^n(365-n)!}$.

Borrowing from the Birthday Problem, we can map the randomized MAC address conflict problem into the Birthday Problem as follows:

- The size of the locally administered MAC address space maps to the number of days in a year (i.e., where we see "365" days for a year, we substitute with 2^{46}, the size of the address space).
- The number of randomly generated locally administered MAC addresses maps to the number of people in a room.

Thus, the Birthday Problem's formulaic solution, when applied to the randomized MAC conflict problem, becomes $p(n) = 1 - \frac{(2^{46})!}{(2^{46})^n(2^{46}-n)!}$, where *n* is the number randomly generated, locally administered MAC addresses.

Of course, just like the Birthday Problem's solution assumes that the population's birthdays are evenly distributed across all 365 days of the year, this solution also assumes that the randomly generated locally administered MAC addresses are evenly distributed across all 46 bits of the usable address space.

Approximation

The precise formular $p(n) = 1 - \frac{(2^{46})!}{(2^{46})^n(2^{46}-n)!}$ requires the evaluation of the factorial of 2^{46} ≈ 70 trillion and the n^{th} power of 2^{46} ≈ 70 trillion. For the purpose of this analysis, we want to extend *n* into the order of tens of millions. Evaluating all these factorials and high-powered exponents precisely is a computationally expensive proposition, thus we resort to approximations for the remaining analysis.

The Taylor series expansion of the exponentiation function provides a first-order approximation for $e^x \approx 1 + x$ where *x* is very small, i.e., $|x| \ll 1$. We satisfy the $|x| \ll 1$ condition by having *n* being very small compared to 2^{46}. Even when we bring the number of addresses *n* to a billion, it is still very small compared to the total usable address space of 70 trillion. With this, we approximate the probability of randomized MAC address conflict as $p(n) \approx 1 - e^{-\frac{n(n-1)}{2 \cdot 2^{46}}}$.

The remaining analysis and results are done with the approximation $p(n) \approx 1 - e^{-\frac{n(n-1)}{2 \cdot 2^{46}}}$.

Estimating the Number of Non-Unique MAC Addresses

Beyond certain threshold, the likelihood of MAC address conflict becomes a statistical certainty, and it would be more meaningful to ask "*how many non-unique MAC addresses are we likely to find in a system*" where large numbers of locally administered MAC addresses are randomly generated.

Drawing again from the Birthday Problem, it is the equivalent of asking "*how many people in the room can be expected to have a birthday that is shared with someone else in the room*."

The formular to estimate this quantity is a straightforward $N = n - n\left(\frac{2^{46}-1}{2^{46}}\right)^{n-1}$ where n is the number of randomly generated locally administered MAC addresses and 2^{46} is the usable address space.

Results

The comprehensive view on the conflict probability of randomly generated locally administered MAC addresses are shown in in Figure 1. Here, in table form, we highlight only a few notable data points:

Number of randomly generated locally administered MAC addresses in a system	Probability of MAC address conflict in the system
Roughly 38 thousand	0.001% (OK “five 9’s” of the time)
Roughly 120 thousand	0.01%
Roughly 380 thousand	0.1%
Roughly 1.2 million	1%
Slightly below 4 million	10%
Slightly below 10 million	50%
Roughly 30 million	Near 100% (statistical certainty)

Table 1: Select data points showing probability of randomized MAC address conflict

Beyond the “roughly 30 million” threshold, it becomes more meaningful to estimate the number of non-unique MAC addresses as a function of the number of randomly generated locally administered MAC addresses:

Number of randomly generated locally administered MAC addresses in a system	Estimated number of non-unique MAC addresses in the system
50 million	36
75 million	80
100 million	142
125 million	222
150 million	320
200 million	568
250 million	888
500 million	3553
1 billion	14211

Table 2: Select data points showing estimated numbers of non-unique MAC addresses

While small, the growth of the estimated number of non-unique MAC addresses relative to the number of randomly generated MAC addresses appears to be quadratic.

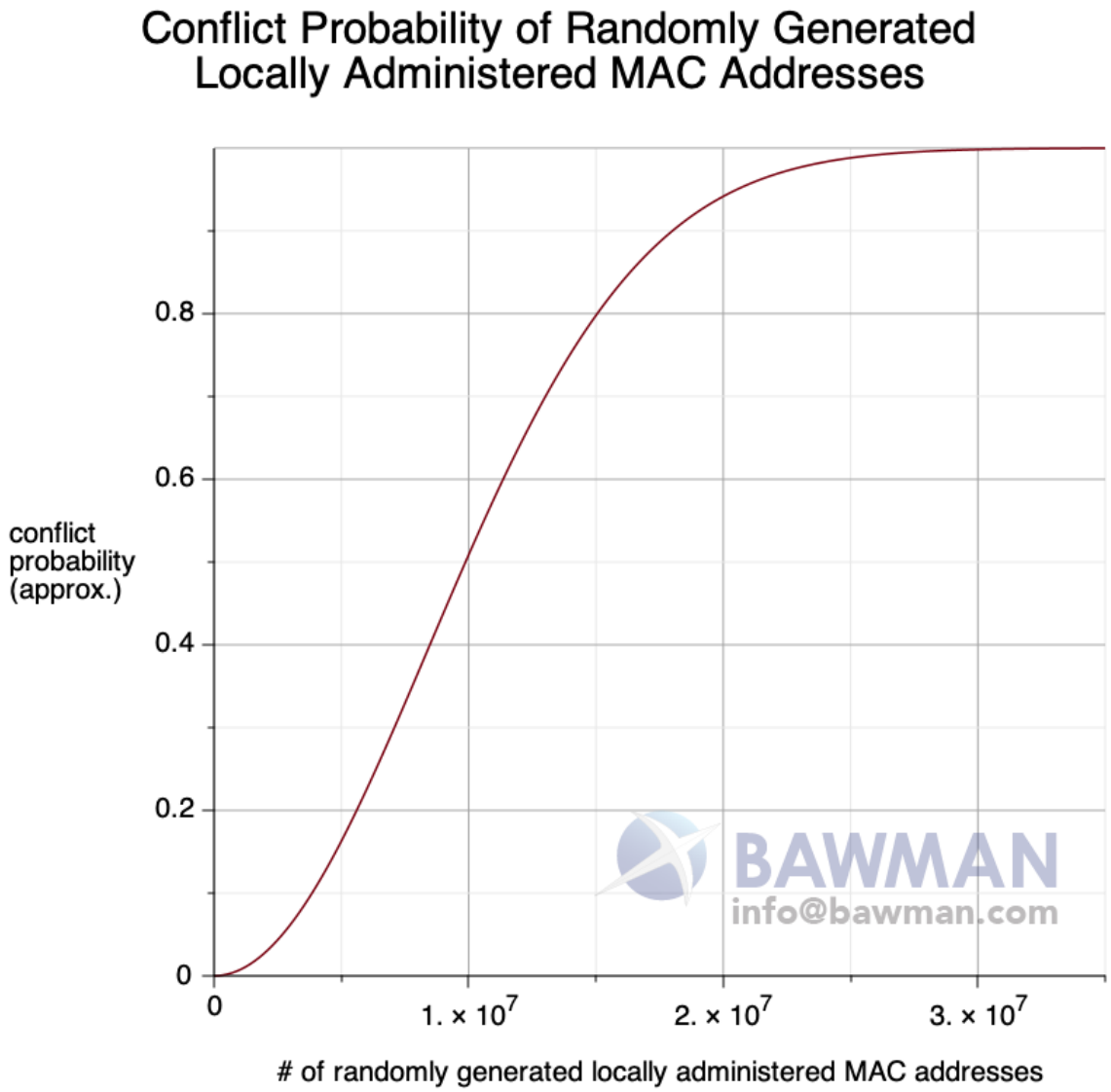

Figure 1: Approximate Conflict Probability of Randomly Generated Locally Administered MAC Addresses

Additional graphs that "zoom in" on the lower 1%, 10%, and 50% portions of the results are included in the Appendix for the reader's convenient reference.

DISCUSSION: Impacts and Potential Mitigating Measures

As the analysis shows, the conflict probability of randomly generated locally administered MAC addresses in a system differ materially depending on the number of such addresses in the system. As such, discussion of impact needs to be grounded on the size of the system.

That said, whenever two devices try to operate in the same local area network (LAN) with the same MAC address, one or both of them will have issues with basic communications. A simple mitigation measure such as having the devices with conflicting MAC addresses reconnect with different MAC addresses will in most cases be enough to restore basic communication functions to these devices. Even then, there are secondary issues and considerations with such a simple mitigation measure that will be discussed in a later part of this paper.

Impact on Smaller Systems

For the basic functions of forwarding data to and from typical end-user devices, randomly generated locally administered MAC addresses are very unlikely to be an issue. Using first order estimates, it would take roughly 38 thousand such MAC addresses to yield a basically negligible 0.001% probability of MAC address conflict in a system. The typical independently operated home and small business local area networks scale in the order of tens or a few hundreds of MAC addresses at most. These independently operated small systems are two to three orders of magnitudes away from having to worry about being materially impacted by randomized MAC addresses.

Impact on Larger Systems

Larger systems are, in one form or another, aggregations of multiple smaller systems. While a physical LAN segment can rarely accommodate more than a few hundred devices, it is conceivable that an operator who operates multiple such LAN segments may logically aggregate multiple such LAN segments into one larger system. Even then, from a data forwarding perspective it is highly unlikely that one can or would aggregate more than few thousand hosts into one switching domain without having additional information that the system can use to disambiguate hosts attached through different physical LAN segments. Such “additional information” would in theory afford the aggregating systems additional latitude to deal with randomized MAC address conflicts. (A particular implementation may or may not do any disambiguation based on such “additional information” today, but the opportunity to do so exists and may be leveraged in future software/firmware upgrades for that system.) Thus, basic traffic forwarding for the end-users remains unlikely to be affected by randomly generated MAC addresses.

Impact Beyond Data Forwarding

Beyond basic data forwarding, the impact of randomized MAC address conflict on operators is harder to gauge as it depends on what else the operators and their systems do with MAC addresses. There are few technical standards governing what can or cannot be done with MAC addresses outside of data forwarding. Security threat analytic software, traffic pattern analytic software, marketing-oriented user behavior analytic software, and metadata aggregators that aggregate MAC addresses collected from multiple LANs may use MAC addresses in varied, creative, and proprietary ways. Hence the impact of

randomized MAC address conflict on aspects of systems outside of data forwarding has to be separately considered on a case-by-case basis.

Potential Mitigating Measures, Their Limitations, and Secondary Impacts

In this subsection we delve a little deeper into mitigation ideas and their attendant issues.

Local Mitigation

We mentioned before that whenever there is a MAC address conflict in the same LAN, one or all of the devices using the same MAC address will have issues with basic communications. We also mentioned that one conceptually simple mitigation measure is for every host that was using the same, conflicted MAC address to each rejoin the network with a different, newly generated MAC address.

Note that with the mitigation method mentioned above, the power and responsibility to execute the mitigating procedure rests with the client devices (and by extension the client device vendors and platform providers), not with the network (and by extension not with the network's system operators or their technology vendors). Thus, even though at the system level MAC address conflicts are more likely to impact the system operators, paradoxically the power to mitigate this issue appear to lie more with the device vendors.

Even so, there are still technical limitations and secondary impacts that need to be considered.

One limitation has to do with how a host device can "discover" that there is a MAC address conflict. In theory a host can just monitor its LAN and look for frames carrying its source MAC address that it did not send. One can conceivably design a protocol extension that let hosts actively check for MAC address conflicts on the LAN. When limited in this fashion, the mitigation measure should work well at least for the LAN segment.

Coordination Beyond a Single LAN

What if a larger system wants to coordinate the discovery of MAC address conflict across multiple LAN segments? Imagine an aggregator or controller of some sort that oversees multiple LAN segments. One with incentives to avoid MAC address conflict across all LAN segments in an aggregator's domain may be tempted to device a way to use that aggregator to help the two hosts on two different LAN segments "discover" that their MAC addresses have conflicted.

Doing so implies that the hosts will be given knowledge that they would normally not have; two separate LAN segments that are normally considered segregated now have information "leaked" between them. The security and privacy considerations of doing so must be carefully examined before implementing such a mitigation measure.

Mitigation Upon Mitigation and Its Attendant Impact

Assuming one resorts to "leaking" MAC address information across two LAN segments anyway, then one potential tactic to mitigate the attendant privacy and security concerns may be to introduce false positives into the system, e.g., by having the aggregator randomly choose to signal a cross-LAN MAC address conflict even when there is not one. This can certainly serve to confuse and mislead any malicious agents trying to glimpse exploitable information on the LAN segments but doing so may also unnecessarily disrupt a host's connection because the host being notified of a MAC address conflict has to switch to a different MAC address and rejoin the network. Thus, the user experience impact also needs to be considered before one deploys such a mitigation measure.

Global Administratively Coordinated Mitigation, Single-OUI Randomization

Instead of having all devices that randomly generate locally administered MAC addresses do so in the entire 46-bit usable address space, it is conceivable that one can administratively divide the address space and have different groups of devices generate randomized MAC addresses only in the smaller address spaces reserved for each respective group.

For example, the 48-bit MAC address is generally divided into the 24-bit Organizationally Unique Identifier (OUI) portion (this portion includes the Multicast/Unicast bit and the Globally Unique/Locally Administered bit) and the 24-bit Network Interface Controller (NIC) Specific portion. Consider, for the sake of discussion, a scheme that requires hosts that want to use a randomized locally administered MAC address to only randomize the 24-bit NIC Specific portion of the MAC address. Even though a scheme like this separates devices by OUI beforehand, it also reduces the randomizable address space drastically, from $2^{46} \approx 70$ trillion to $2^{24} \approx$ a little under 17 million.

Figure 2 shows single-OUI conflict probability of randomly generated MAC addresses among devices with the same OUI if they are limited to using only a 24-bit randomizable address space.

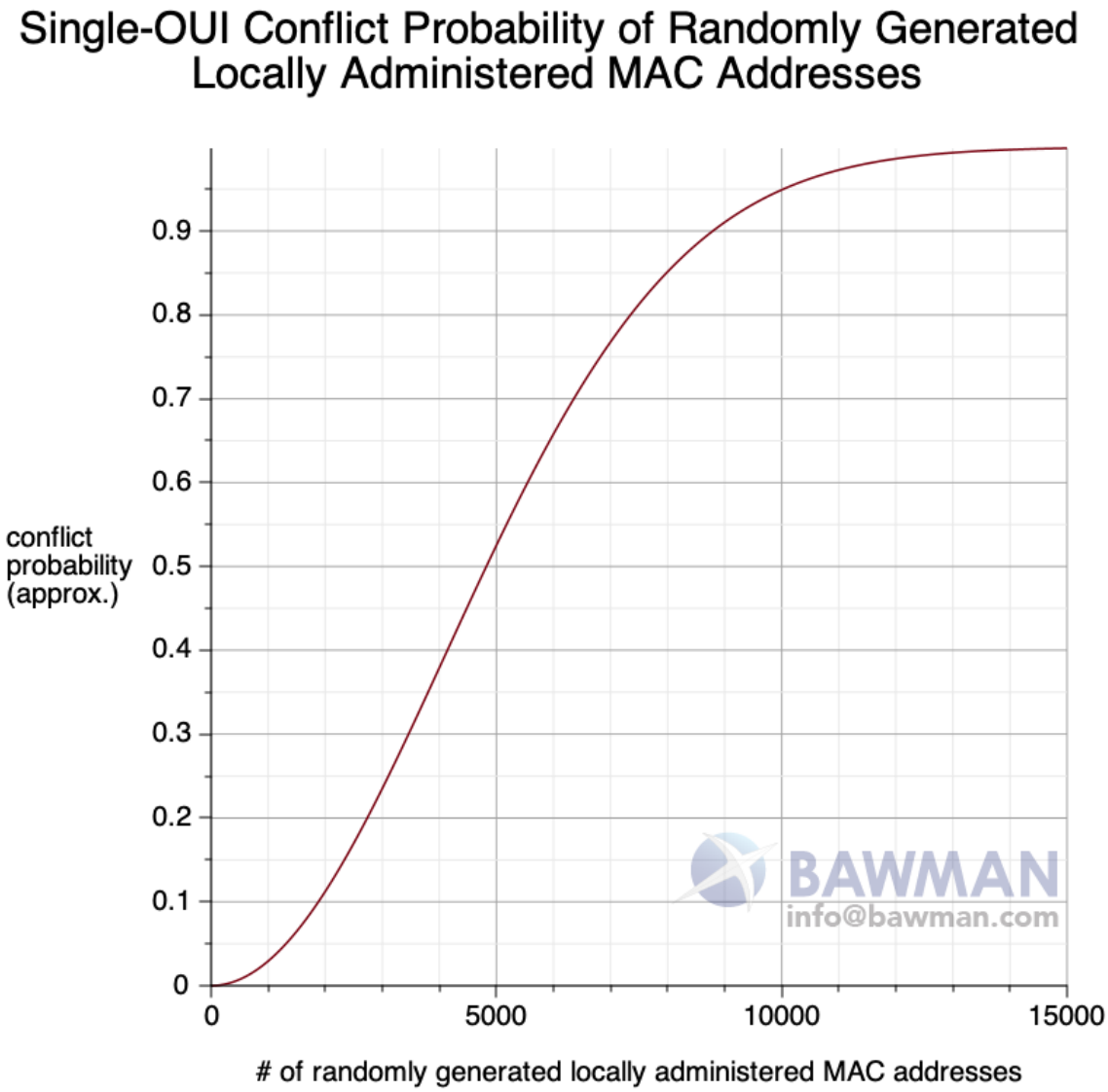

Figure 2: Single-OUI Conflict Probability of Randomly Generated Locally Administered MAC Addresses

Since OUIs are assigned by organizations and some organization's devices are bound to be more popular than others, one should expect that in general the number of devices will not be evenly distributed across all OUIs.

Statistically, if the objective is to minimize MAC address conflict, it would be more advantageous to let all devices randomize into one large address space than to pre-sort devices into separate groups and then only allow devices in each group to randomize into smaller address spaces reserved for their respective groups.

MAC Address Randomization Across Time and Space

While this paper largely discusses MAC address randomization in terms of “number of hosts” or “number of devices,” the reader should keep in mind that various systems also cache or otherwise “remember” MAC addresses for various lengths of time for various reasons. Within a caching window of a system, it may make little difference to that system whether two different MAC addresses have been generated by two different devices or by the same device at different times.

From a data forwarding perspective, ARP cache timeouts are typically in the order of “minutes” or “hours”. Randomized MAC address conflict induced data forwarding issues would thus be unlikely to last beyond “minutes” or “hours.” Outside of data forwarding however, time frame for how long the effects of randomized MAC address conflict would last needs to be examined on a case-by-case basis.

SUMMARY and CONCLUSION

Prompted by the proliferation of randomly generated locally administered MAC addresses used by popular consumer devices, this paper looks at the likelihood of MAC address conflict when large number randomized MAC addresses are introduced into a system.

With reasonable assumptions, we find that it is possible to model the randomized MAC address conflict problem after the Birthday Problem. Through reasonable approximations, we estimate that in the base case it would take roughly 1.2 million randomly generated MAC addresses to yield a 1% chance of conflict, a little under 4 million randomly generated MAC addresses to yield a 10% change of conflict, and a little under 10 million randomly generated MAC addresses to yield a 50% chance of conflict.

Beyond roughly 30 million randomly generated MAC addresses, conflict becomes a statistical certainty, and it becomes more meaningful to estimate the number of non-unique MAC addresses. We estimate that generating 50 million MAC addresses randomly will yield roughly 36 non-unique addresses, generating 250 million MAC addresses randomly will yield roughly 888 non-unique MAC addresses, and generating one billion MAC addresses random will yield roughly 14 thousand non-unique MAC addresses.

We find also that the probability of MAC address conflict can increase significantly if randomization is restricted by OUI. If devices were limited to randomize only into the non-OUI part of the address space, then it would take only around 5000 randomized MAC addresses to yield a 50% chance of conflict, and only around 15000 randomized MAC addresses to statistically guarantee a conflict.

For basic data forwarding functions, we find that it is highly unlikely that randomly generated MAC addresses will materially impact the smaller systems (these include the typical home and small business networks). While randomly generated MAC addresses have a higher likelihood of materially impacting larger systems, the impact on the data forwarding aspects of these larger systems will likely be limited.

A few potential mitigation measures are also examined and discussed.

Outside of data forwarding, there are few technical standards that govern the use of MAC addresses. As such, outside of data forwarding, most issues that stem from MAC address conflict need to be separately considered on a case-by-case basis.

CONTACT

For any inquiry, please email:

BAWMAN <info@bawman.com>

Find this paper on the web:

https://www.bawman.com/BAWMAN/articles/RandomMAC/

APPENDIX

For the reader's convenient reference, we provide below graphs that "zoom in" on the lower 50%, 10%, and 1% portions of the results of our approximations for the conflict probability of randomly generated locally administered MAC addresses.

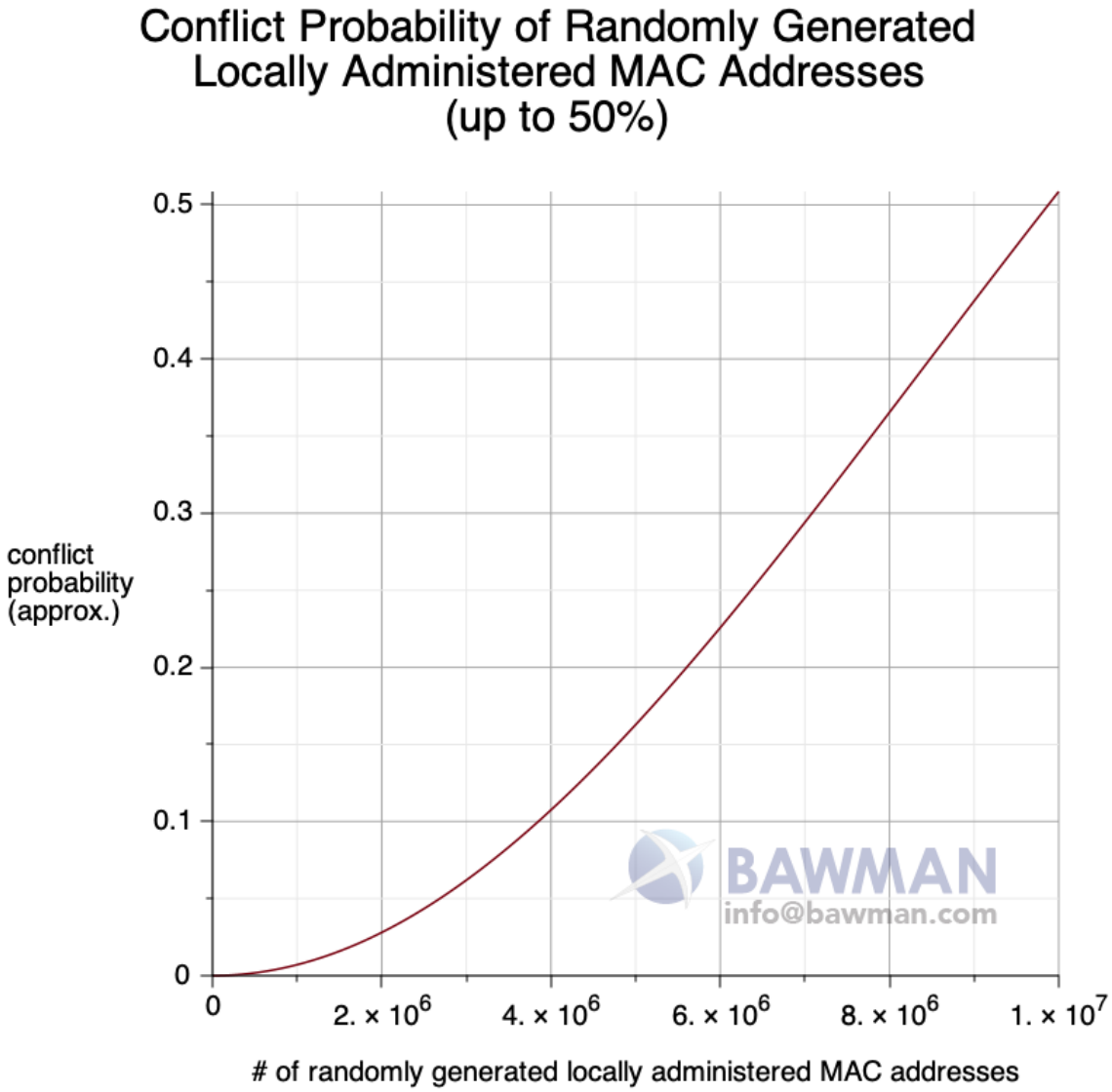

Figure 3: Lower 50% conflict probability of randomly generated locally administered MAC addresses

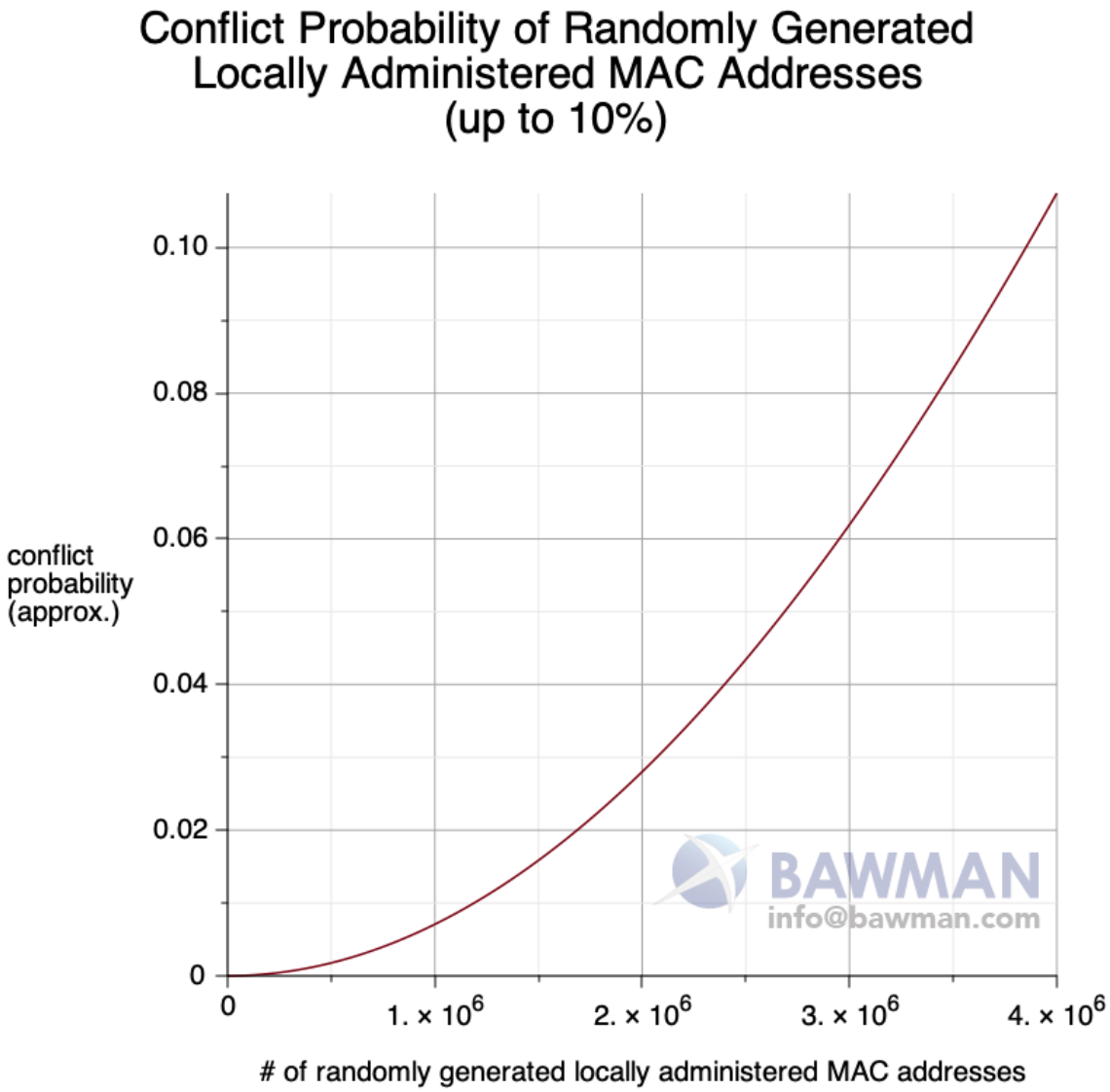

Figure 4: Lower 10% conflict probability of randomly generated locally administered MAC addresses

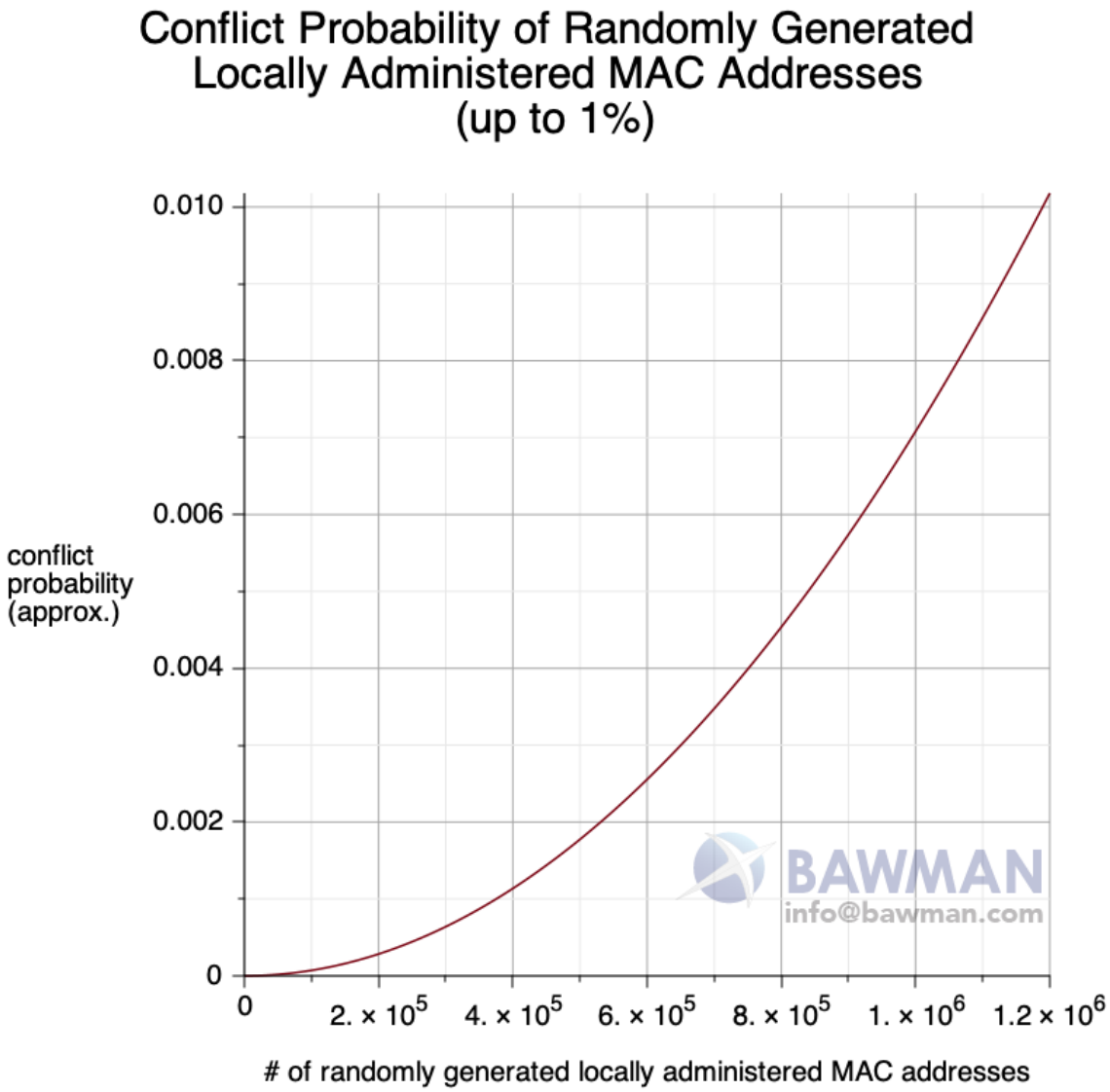

Figure 5: Lower 1% conflict probability of randomly generated locally administered MAC addresses