On 0x3d84a438af72f6396785eea97b32f903520e36c8

mike neuder

October 30, 2023

Abstract

The Ethereum address 0x3d84a438af72f6396785eea97b32f903520e36c8 is the first address I used in crypto. It is where (i) mikeneuder.eth resolves, (ii) milady #1696 sits, and (iii) validator #921927's withdrawal credential points. Inspired by winny dot ethereum and following her recommendation of the amazing Brooklyn-based artist, yang, this address is now permanently inked onto my upper-left forearm. This article presents the design of the tattoo and a probabilistic analysis to accompany it. We demonstrate that, when written in a three-row notation, randomly generated Ethereum addresses have five distinct matches with a probability of approximately 0.5899%. This implies that 0x3d84...36c8 has a property that only 1 out of every 170 Ethereum addresses have – pretty special!

1 Introduction

Ethereum accounts are identified with a 20-byte address derived from the public key associated with the account. With the customary 0x prefix associated with hex-encoded values, the address is a 42-character string. This string can be divided into three, 14-character rows stacked vertically, as seen in Figure 1.

```
0x3d84a438af72
f6396785eea97b
32f903520e36c8
```

Figure 1: The three, 14-character rows.

Definition 1. *A character is said to belong to a match if it has the same value as the character directly above or below it in the 14-character row notation.*

```
0x3d84a438af72
f6396785eea97b
32f903520e36c8
```

Figure 2: Row notation with matches highlighted in red.

Figure 2 shows the address in row notation with the matches highlighted in red. When written like this, it was surprising to see so many matches, which became the inspiration for the tattoo. The natural question that follows is, "What is the probability that a randomly generated address has this many matches?". Upon further inspection, it is also surprising that of the five matches in the address, they are all distinct (e.g., none of them are the same hex character). This extends our line of questioning to include, "What is the probability that the matches are all distinct?". The remainder of this article aims to answer these questions probabilistically. Section 2 calculates the probability that a random address has five *matches*. Section 3 calculates the probability that a random address has five *distinct* matches. Section 4 concludes.

2 Five matches

We start simply by asking about the probability of five matches. Because each match is an independent event, we can model the scenario with a Binomial random variable.

2.1 Analytic solution

Given the three-row notation, 26 character pairs could be matches (we exclude the `0x` prefix); 12 when comparing the first two rows and 14 when comparing the second two rows. Assuming that each character is randomly selected from the base-16 alphabet, any pair has $P(\text{match}) = 1/16$. Let X denote the number of matches given a randomly generated address, then

$$X \sim \text{Binomial}(26, 1/16). \tag{1}$$

The probability of exactly five matches is

$$\begin{aligned} P(X = 5) &= \binom{26}{5} \left(\frac{1}{16}\right)^5 \left(\frac{15}{16}\right)^{21} && (2)\\ &= \frac{8202577038923978805541992 1875}{5070602400912917605986812821504} && (3)\\ &\approx 0.0161767. && (4) \end{aligned}$$

In percentage terms, this is about 1.6% – pretty low! We can verify this value with a simple Monte Carlo simulation.

2.2 Experimental verification

To confirm this value, we can simply generate many random addresses and see how many have exactly 5 matches. Algorithm 1 shows the pseudocode for this simulation.

Algorithm 1 Approximate probability of exactly five matches

```
Require: n ≥ 0
  i ← 0
  successes ← 0
  while i ≤ n do
     addr ← random.randint(range=16, size=40)
     matches ← countMatches(addr)
     if matches == 5 then
        successes ← successes +1
     end if
  end while
  return successes / n
```

Running this simulation with different values for n, we get results as seen in Table 1.

trial #	$n = 10^5$	$n = 10^6$	$n = 10^7$
1	0.01682	0.015925	0.0161367
2	0.01565	0.016213	0.0161871
3	0.01650	0.016361	0.0161212
4	0.01591	0.016175	0.0162162
5	0.01546	0.016182	0.0161562
μ	**0.016068**	**0.0161712**	**0.01616348**

Table 1: Simulation results for five trials of $n = 10^5$, $n = 10^6$, and $n = 10^7$.

These approximations match the analytic solution well.

3 Five *distinct* matches

Beyond just matches, we can also ask about *distinct* matches.

Definition 2. *A set of matches is said to be distinct if there are no two matches with the same character.*

For example the address

```
0x3d84a438af72
f6396785eea97b
32f903520e36c8
```

has 5 matches: `3,9,e,a,7` and since none of the matches are the same character, we can say this set of matches is distinct. The probability of this scenario is much more complicated to calculate because we lose the independence of the events; one match existing changes the probability of the next match because of the distinctness property – the character must differ from the original match. We must also factor in that the `0x` prefix changes the probabilities of a match in one of the first two columns.

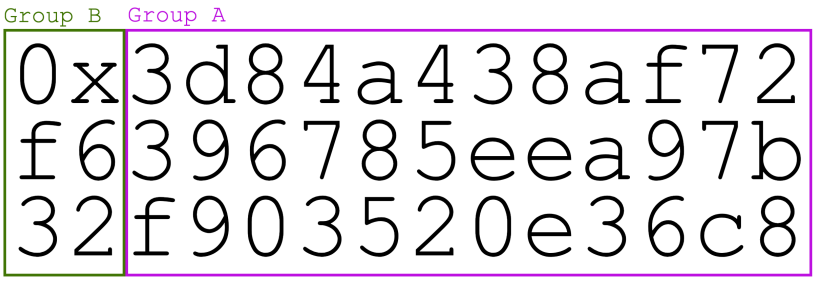

Figure 3: Groups A & B to help construct different cases for the distinct match counting.

Figure 3 partitions the characters into Groups A & B. When considering the probability of five distinct matches, we define three possible cases.

Case 1 $\implies$ All five matches are in Group A.
Case 2 $\implies$ Four matches are in Group A; one match is in Group B.
Case 3 $\implies$ Three matches are in Group A; two matches are in Group B.

Let D_5 denote the event of five distinct matches and $C_1, C_2,$ & C_3 denote the events of Cases 1, 2, & 3 respectively. Then

$$P(D_5) = P(C_1) + P(C_2) + P(C_3). \tag{5}$$

To find our desired quantity of $P(D_5)$, we will calculate the probability of each case and take the sum.

3.1 Analytic solution for Case 1

In Case 1, we are just focusing on columns 3-14 (Group A). The quantity of interest is the probability that out of these 12 columns, 5 of them are matches and the matches are distinct. We begin with a reduced version of this problem to demonstrate the technique.

Definition 3. *The two-row problem asks the same question of the 12-column construction (Group) but with 2 rows instead of 3. Let TR_5 denote the event that given two randomly generated rows of 12 hex characters, there are five distinct matches.*

Let Y denote the event that all five of the matches are distinct and X_{ii} denote the Binomial random variable representing the number of matches in the two-row problem (this is a similar binomial to what we use in Section 2, just with $n = 12$ because there are just two rows),

$$X_{ii} \sim \text{Binomial}(12,\ 1/16) \tag{6}$$

Then $P(TR_5) = P(Y \cap X_{ii} = 5)$. By the definition of conditional probability, we have

$$P(Y \cap X_{ii} = 5) = P(Y|X_{ii} = 5)P(X_{ii} = 5). \tag{7}$$

From the binomial, we have

$$P(X_{ii} = 5) = \binom{12}{5}\left(\frac{1}{16}\right)^5\left(\frac{15}{16}\right)^7 \tag{8}$$

$$= \frac{16915078125}{35184372088832} \tag{9}$$

$$\approx 0.000480755. \tag{10}$$

For $P(Y|X_{ii} = 5)$ we are looking for the probability that *given five matches*, the matches are distinct. With a 16-character alphabet, we start by selecting the first character, which is guaranteed to be distinct. The second differs from the first with a probability of 15/16. The third differs from the first two with a probability of 14/16. Continuing this pattern we have,

$$P(Y|X_{ii} = 5) = \frac{16}{16} \cdot \frac{15}{16} \cdot \frac{14}{16} \cdot \frac{13}{16} \cdot \frac{12}{16} \tag{11}$$

$$= \frac{4095}{8192} \tag{12}$$

$$\approx 0.49988779. \tag{13}$$

Note that this is the same method used to calculate probabilities in the Birthday Problem. Thus we get to our final analytical solution for the two-row problem,

$$P(TR_5) = P(Y \cap X_{ii} = 5) \tag{14}$$

$$= P(Y|X_{ii} = 5)P(X_{ii} = 5) \tag{15}$$

$$= \frac{4095}{8192} \cdot \frac{16915078125}{35184372088832} \tag{16}$$

$$= \frac{69267244921875}{288230376151711744} \tag{17}$$

$$\approx 0.000240319. \tag{18}$$

We went through this extended side quest because it makes the calculation of $P(C_1)$ much easier to follow. Recall that for Case 1, we have 12 columns of 3 rows of hex characters. Let $r_1, r_2, \& \ r_3$ denote one column of characters (r_1 is the top row, r_2 is the middle row, r_3 is the bottom row). For that column to qualify as one of the five distinct matches, we need either $r_1 = r_2 \ \wedge \ r_2 \neq r_3$ or $r_1 \neq r_2 \ \wedge \ r_2 = r_3$. In other words, the middle row character needs to match either the top row or the bottom row, but not both (exclusive or). The exclusivity property is needed because if $r_1 = r_2 = r_3$, then by definition the two matches cannot be distinct. Let M denote the event of a randomly generated column having this "match" property. We have

$$P(M) = 2 \cdot \frac{1}{16} \cdot \frac{15}{16} \tag{19}$$

$$= \frac{15}{128}. \tag{20}$$

If you fix r_2, then 1/16 is the probability that r_1 matches, and 15/16 represents the probability that r_3 differs. We multiply by 2 to account for the other case where r_3 is the match and r_1 differs. Now we can define our new binomial, which represents the number of columns in Group A that match,

$$X_{iii} \sim \text{Binomial}(12,\ 15/128). \tag{21}$$

We use *iii* as the subscript because we are back in the three-row regime, but $n = 12$ still because we treat each column as an independent Bernoulli trial. Now we have

$$P(X_{iii} = 5) = \binom{12}{5}\left(\frac{15}{128}\right)^5\left(\frac{113}{128}\right)^7 \tag{22}$$

$$= \frac{17686446888481758028125}{24178516392292583494123 52} \tag{23}$$

$$\approx 0.00731494. \tag{24}$$

Recall that Y still denotes the event of the matches being distinct. Thus the conditional probability of the three-row construction is the same as the two-row,

$$P(Y|X_{iii} = 5) = P(Y|X_{ii} = 5) \tag{25}$$
$$\approx 0.49988779. \tag{26}$$

The full probability of Case 1 directly follows,

$$P(C_1) = P(Y \cap X_{iii} = 5) \tag{27}$$
$$= P(Y|X_{iii} = 5)P(X_{iii} = 5) \tag{28}$$
$$= \frac{4095}{8192} \cdot \frac{17686446888481758028125}{24178516392292583494123 52} \tag{29}$$
$$= \frac{72426000008332799125171875}{19807040628566084398385987584} \tag{30}$$
$$\approx 0.0036565786. \tag{31}$$

Nice! Ok, I promise that was the hard part. Cases 2 and 3 are just small modifications of Case 1.

3.2 Simulation results for Case 1

To sanity check the probability of each case, we will run simulations as done in Table 1. We use five trials for each $n \in \{10^4, 10^5\}$.

trial #	$n = 10^4$	$n = 10^5$
1	0.0041	0.00393
2	0.0030	0.00388
3	0.0045	0.00351
4	0.0023	0.00367
5	0.0035	0.00357
μ	**0.00348**	**0.003712**

Table 2: Simulation results for five trials of $n = 10^4$ and $n = 10^5$ of Case 1.

3.3 Analytic solution for Case 2

Case 2 occurs if there is one match in Group B and four matches in Group A. Since the columns in Group B are only two rows (because the first two characters of row 1 are the `0x` prefix), the probability of exactly one match between rows 2 and 3 in those two columns is $2 \cdot 1/16 \cdot 15/16 = 15/128$ (a familiar quantity by this point :-). Now we can define the probability of Case 2 as

$$P(C_2) = \frac{15}{128} \cdot P(Y \cap X_{iii} = 4) \tag{32}$$
$$= \frac{15}{128} \cdot P(Y|X_{iii} = 4) \cdot P(X_{iii} = 4). \tag{33}$$

In other words, the probability of exactly one match in the first two columns is multiplied by the probability that there are exactly four distinct matches in the remaining 12 columns (Group A). Using the binomial in Equation 21, we have

$$P(X_{iii} = 4) = \binom{12}{4} \left(\frac{15}{128}\right)^4 \left(\frac{113}{128}\right)^8 \tag{34}$$
$$= \frac{66618949946614621905937 5}{19342813113834066795298816} \tag{35}$$
$$\approx 0.03444119. \tag{36}$$

Using the same logic as Equation 11, we can calculate $P(Y|X_{iii} = 4)$, except the four matches must differ from the match in Group B to maintain distinctness (so we start at 15/16 instead of 16/16).

$$P(Y|X_{iii} = 4) = \frac{15}{16} \cdot \frac{14}{16} \cdot \frac{13}{16} \cdot \frac{12}{16} \tag{37}$$

$$= \frac{4095}{8192} \tag{38}$$

$$\approx 0.4998779. \tag{39}$$

Thus we conclude

$$P(C_2) = \frac{15}{128} \cdot P(Y|X_{iii} = 4) \cdot P(X_{iii} = 4) \tag{40}$$

$$= \frac{15}{128} \cdot \frac{4095}{8192} \cdot \frac{666189499466146219059375}{193428131133834066795298816} \tag{41}$$

$$= \frac{4092069000470803150572210937 5}{2028240960365167042394725128 6016} \tag{42}$$

$$= 0.00201754578. \tag{43}$$

3.4 Simulation results for Case 2

We use five trials for each $n \in \{10^4, 10^5\}$.

trial #	$n = 10^4$	$n = 10^5$
1	0.0020	0.00211
2	0.0017	0.00219
3	0.0023	0.00181
4	0.0028	0.00203
5	0.0015	0.00184
μ	**0.00206**	**0.001996**

Table 3: Simulation results for five trials of $n = 10^4$ and $n = 10^5$ of Case 2.

3.5 Analytic solution for Case 3

Case 3 occurs if there are two matches in Group B and three matches in Group A. Since Group B is just a two-row version (because the first two characters of row 1 are the `0x` prefix), the probability of both characters being matches and those matches being distinct is $1/16 \cdot 1/16 \cdot 15/16 = 15/4096$. Now we can define the probability of Case 3 as

$$P(C_3) = \frac{15}{4096} \cdot P(Y \cap X_{iii} = 3) \tag{44}$$

$$= \frac{15}{4096} \cdot P(Y|X_{iii} = 3) \cdot P(X_{iii} = 3). \tag{45}$$

Using the binomial in Equation 21, we have

$$P(X_{iii} = 3) = \binom{12}{3} \left(\frac{15}{128}\right)^3 \left(\frac{113}{128}\right)^9 \tag{46}$$

$$= \frac{5576252847383297981756 25}{4835703278458516698824704} \tag{47}$$

$$\approx 0.11531421. \tag{48}$$

Using the same logic as Equation 11, we can calculate $P(Y|X_{iii} = 4)$, except we must differ from both matches in Group B.

$$P(Y|X_{iii} = 3) = \frac{14}{16} \cdot \frac{13}{16} \cdot \frac{12}{16} \tag{49}$$

$$= \frac{273}{512} \tag{50}$$

$$\approx 0.533203. \tag{51}$$

Thus we conclude

$$P(C_3) = \frac{15}{4096} \cdot P(Y|X_{iii} = 3) \cdot P(X_{iii} = 3) \tag{52}$$

$$= \frac{15}{4096} \cdot \frac{273}{512} \cdot \frac{557625284738329798175625}{483570327845851669882470 4} \tag{53}$$

$$= \frac{228347554100346052352918 4375}{10141204801825835211973625643008} \tag{54}$$

$$= 0.000225168. \tag{55}$$

3.6 Simulation results for Case 3

We use five trials for each $n \in \{10^4, 10^5\}$.

trial #	$n = 10^4$	$n = 10^5$
1	0.0002	0.00017
2	0.0003	0.00030
3	0.0002	0.00026
4	0.0002	0.00018
5	0.0003	0.00023
μ	**0.00024**	**0.000228**

Table 4: Simulation results for five trials of $n = 10^4$ and $n = 10^5$ for Case 3.

4 Conclusion

From Equation 5, we have

$$P(D_5) = P(C_1) + P(C_2) + P(C_3) \tag{56}$$

$$\approx 0.0036565786 + 0.00201754578 + 0.000225168 \tag{57}$$

$$\approx 0.00589929238. \tag{58}$$

To conclude, given a random Ethereum address, the probability of having five distinct matches is about 0.5899%. `0x3d84a438af72f6396785eea97b32f903520e36c8` may not be "1 in a million", but it is 1 in 170, which is special enough for me.

thanks for reading :-) !

p.s. You may be wondering, "That is just the probability of five distinct matches... what about the probability of five **or more** distinct matches? Wouldn't that better quantify the 'uniqueness' of this address?" Well, dear astute reader, the answer is yes. However, given the length of the "exactly five" analysis, we have omitted the "five or more" calculation for the sake of brevity; it's left as an exercise for the reader.