Can you have a DFA that only accepts strings of prime length?

Ajitesh Dasaratha

September 2025

We want to see whether the language

$$L = \{1^p \mid p \text{ is prime}\}$$

is regular or not.

I genuinely have no idea how I'd construct a DFA, NFA or regular expression for this. If there were only a finite number of primes, then we could just hardcode all those strings in a regular expression. Except you can prove there are infinite primes by contradiction (it's a very standard proof, any of the top results from a Google search have the right proof)

It's hard to think of a fooling set as well, but we can think of some qualities a fooling set F would have. It needs to be an infinite set of strings such that for any strings $x = 1^i$ and $y = 1^j$ in F, you can add some string 1^k, so that exactly of 1^{i+k} and i^{j+k}, has prime length.

Let's go through some examples:

$x = 11, y = 111$: With suffix $z = 1$, $xz = 111 \in L, yz = 1111 \notin L$. What about $x = 1^{13}, y = 1^{19}$? Then all strings $z = \epsilon, 1, 11, 111, 1111, 11111$ won't work. $z = 1^6$, works since $xz = 1^{19} \in L$ and $yz = 1^{25} \notin L$.

There isn't an obvious way to distinguish any two arbitrary strings. We can rephrase our problem slightly: given two integers, i, j (the lengths of our strings), what integer k can we add to both, such that exactly one of $i + k$

and $j+k$ is prime? This k will be the length of our distinguishing suffix. We can also restrict which i and j values we want to be in our fooling set if needed.

Great, we now have the problem phrased as something else we have no idea how to solve!

There's a theorem called Dirichlet's theorem that states:

Let a and d be positive integers with $\gcd(a,d)=1$ (This means they have no common factors other than 1, so for example $gcd(3,5)=1$) Another way of saying this is that a and d are co-prime. Then the arithmetic progression

$$a,\ a+d,\ a+2d,\ a+3d,\ \ldots$$

contains infinitely many prime numbers.

How is this useful? Say we have two strings $x=1^i$ and $y=1^j$, $i<j$, both i and j are composite, and $gcd(i,j)=1$. Let $d=j-i$. Now, d and i must be co-prime (since if they had a common factor more than 1, then $j=i+d$ would also have that same factor, causing $gcd(i,j)>1$, but we know $gcd(i,j)=1$). By Dirichlet, the arithmetic progression

$$i,\ i+d,\ i+2d,\ i+3d,\ \ldots$$

has infinitely many prime numbers. Let's call the *smallest* of these numbers $p=i+nd$. Now, $n\geq 2$ since $j=i+d$ was taken to be composite, and $i+nd$ is prime, while $i+(n-1)d$ isn't. Now, consider the strings 1^i and 1^j, and the suffix $z=1^{(n-1)d}$. If we do this, then $xz=1^i1^{(n-1)d}=1^{i+(n-1)d}\notin L$, and $yz=1^j1^{(n-1)d}=1^{i+d}1^{(n-1)d}=1^{i+nd}\in L$.

This proves that if two strings x and y have composite lengths that are co-prime, then they are distinguishable since we can come up with some suffix z that gets exactly one of xz and yz accepted.

Now we need to form an infinite fooling set of strings $F=1^{l_n}$, where all l_n are composite, and every pair $l_a, l_b\in F$ is co-prime. How do we pick out such lengths?

If we ditch the requirement of being composite, and just focus on co-primeness then $P=\{n : n\ is\ prime\}$ works. The easiest way to satisfy the requirement of being composite is by squaring all those prime numbers. This way, we don't lose the property of any two elements l_a and l_b having $gcd(l_a, l_b) = 1$, but we can also make all elements composite.

So, our final fooling set is:

$$F = \{1^{p^2} \mid p\ is\ prime\}.$$

Note that F is infinite because there are infinitely many primes.
And by the argument earlier in the text, since every pair of strings will have co-prime lengths, and composite havelength, this fooling set is valid.