Minimal Primes

Jeffrey Shallit*
Department of Computer Science
University of Waterloo
Waterloo, Ontario, Canada N2L 3G1
`shallit@graceland.uwaterloo.ca`

July 16, 2001

1 Introduction

Recreations involving the decimal digits of primes have a long history. To give just a few examples, without trying to be exhaustive, Yates [8] studied the "repunits", which are primes with base-10 representation of the form `111` $\cdots$ `1`. Caldwell and Dubner [3] studied the "near-repunits", which are primes of n decimal digits containing $n-1$ ones and 1 zero. Card [4] introduced prime numbers such as `37337999`, in which every nonempty prefix is also a prime; he called them "snowball" primes. These were later studied by Angell & Godwin [1] and Caldwell [2], who called them "right-truncatable" primes. They also studied the "left-truncatable" primes, such as `4632647`, in which every nonempty suffix is prime. Kahan and Weintraub [6] gave a list of all the left-truncatable primes. Huestis [5] introduced the "recursively laminar primes".

In this note, I discuss an apparently new problem on the decimal digits of primes — but one inspired from a classical theorem in formal language theory.

To begin with, here is some notation. I'll use the letters w, x, y, z to represent strings of digits. If w is a string of digits, then by $[w]$ I mean the integer that w represents when interpreted as a number in base 10. To distinguish a number itself from the digits of its base-10 representation, I'll use the "typewriter type" font to denote strings, like this: `352148` is the base-10 representation of the integer 352148.

Given two strings w and x, I'll call w a *subsequence* of x if I can obtain w by deleting some number of digits from x. Note that I can, if I choose, delete no digits at all, and that the digits I delete need not be consecutive. If w is a subsequence of x, then I write $w \lhd x$. For example, `514` is a subsequence of `352148`, and I write `514` $\lhd$ `352148`. Note that $\epsilon \lhd w$ for all w, where ϵ is the empty string.

Now given two strings w and x, I say they are *comparable* if either $w \lhd x$ or $x \lhd w$ (or both). They are *incomparable* if they are not comparable. For example, the strings `352148`

*Research supported in part by a grant from NSERC.

and 8217 are incomparable. A classical theorem of formal language theory states that **every set of pairwise incomparable strings is** *finite* [7, Thm. 6.1.2]. At first sight this theorem is hard to believe, since we may make a set of pairwise incomparable strings as large as we want. For example, all the strings of length 20 are pairwise incomparable, and there are more than a million of them! And nothing prevents us from replacing 20 with 30 or 100 or 1000. Nevertheless, the result is true, and in fact the proof is not that hard, although we do not give it here.

Given any set of strings S we can define the set $M(S)$ its *minimal elements*. A string w is *minimal* for S if whenever $x \in S$ and $x \lhd w$, then $x = w$. It is easy to see that the set $M(S)$ is pairwise incomparable, and hence it must always be finite, no matter what S is.

The set $M(S)$ has the following pleasant property: for every string w in S, there is some subset of the digits of w that can be removed to obtain a string in $M(S)$. Furthermore, the set $M(S)$ is the smallest subset of S with this property.

All this raises the obvious question of determining $M(S)$ for some classically interesting sets S. In this note we determine $M(S)$ when $S = \texttt{PRIMES}$, the set of prime numbers expressed in base 10.

We have the following theorem.

Theorem 1 *If* $S = \texttt{PRIMES} = \{2, 3, 5, 7, 11, 13, 17, 19, 23, \ldots\}$, *then*

$$\begin{aligned} M(S) &= \{2, 3, 5, 7, 11, 19, 41, 61, 89, 409, 449, 499, 881, 991, 6469, 6949, 9001, 9049, \\ &\quad 9649, 9949, 60649, 666649, 946669, 60000049, 66000049, 66600049\}. \end{aligned}$$

In other words, every prime number has the property that one can delete some number of its decimal digits (possibly none) to obtain some prime appearing in the list above.

Before we give the proof, we introduce a little more notation. We use juxtaposition of strings to denote concatenation, so that, for example,

$$\{1, 2\}\{3, 4\} = \{13, 14, 23, 24\}.$$

If a is a digit, we write a^n to denote the concatenation $\overbrace{aa \cdots a}^{n}$. Finally, we use the notation $*$ to denote any number of repetitions of members of a set in any order, so that, for example,

$$\{1, 2\}\{3, 4\}^* = \{1, 2, 13, 14, 23, 24, 133, 134, 143, 144, 233, 234, 243, 244, 1333, \ldots\}.$$

Now we are ready to prove Theorem 1.

Proof. First, the reader should verify that (a) all of the numbers in the statement of the theorem are indeed primes and (b) no proper subsequence of digits of any member of $M(S)$ is prime. (Note that 1 is not a prime number.)

Now let $x \in \texttt{PRIMES}$. If x is of length 1, then $x \in \{2, 3, 5, 7\}$, so $x \in M(S)$. Hence assume x is a member of PRIMES of length 2 or more. In this case the last digit of x must lie in $\{1, 3, 7, 9\}$. If this last digit is 3 or 7, then $3 \lhd x$ or $7 \lhd x$, respectively. Hence we may assume x ends in 1 or 9.

Case 1: x ends in 1.

If $x = y1$, and y contains a digit in $\{2, 3, 5, 7\}$, then (respectively) then x has a subsequence in $M(S)$. Also, if y contains 1, 4, or 6, then (respectively) $11 \lhd x$, $41 \lhd x$, or $61 \lhd x$. Hence we may assume that $y \in \{8, 9\}\{0, 8, 9\}^*$.

Case 1a: x begins with 8.

In this case we can write $x = 8z1$. If $9 \lhd z$, then $89 \lhd x$. If $8 \lhd z$, then $881 \lhd x$. Hence $x \in 80^*1$. But then, since the sum of the digits of x is 9, $[x]$ is divisible by 3, so $[x]$ cannot be prime.

Case 1b: x begins with 9.

In this case we can write $x = 9z1$. If $9 \lhd z$, then $991 \lhd x$. If z contains two 0's, then $9001 \lhd x$. If z contains two 8's, then $881 \lhd x$. Thus we may assume z contains no 9's, and either zero or one 0, and either zero or one 8. The remaining possibilities are therefore $x \in \{91, 901, 981, 9081, 9801\}$, and all of these numbers are composite.

Case 2: x ends with 9.

If $x = y9$, and one of $\{2, 3, 5, 7\}$ is in y, then x contains a subsequence in $M(S)$. If $1 \lhd y$ or $8 \lhd y$, then (respectively) $19 \lhd x$ or $89 \lhd x$. Hence we may assume that $x \in \{4, 6, 9\}\{0, 4, 6, 9\}^*9$. If x contains no 4's, then $[x]$ is divisible by 3 since the sum of its digits is divisible by 3, and hence $[x]$ cannot be prime. On the other hand, if x contains at least two 4's, then $449 \lhd x$. Hence we may assume that x contains exactly one 4.

Case 2a: x starts with 4.

Then $x = 4z9$. If $9 \lhd z$, then $499 \lhd x$. If $0 \lhd z$, then $409 \lhd x$. Hence we may assume $x \in 46^*9$. But then $[x]$ is divisible by 7, since for $i \geq 0$ we have $7 \cdot [6^i\, 7] = [4\, 6^i\, 9]$.

Case 2b: x starts with 6.

Then we may write $x = 6y4z9$, where $y, z \in \{0, 6, 9\}^*$. If $6 \lhd z$, then $6469 \lhd x$. If $0 \lhd z$, then $409 \lhd x$. If $9 \lhd z$, then $499 \lhd x$. Hence we may assume z is empty.

If $9 \lhd y$, then $6949 \lhd x$. Hence we may assume $x = 6y49$, where $y \in \{0, 6\}^*$.

If $06 \lhd y$, then $60649 \lhd x$. Hence we may assume $y \in 6^*\, 0^*$.

If $666 \lhd y$, then $666649 \lhd x$. If $00000 \lhd y$, then $60000049 \lhd x$. Hence we may assume $y \in \{\epsilon, 6, 66\}\{\epsilon, 0, 00, 000, 0000\}$. There are twelve corresponding possibilities for x, and of these only 66000049 and 66600049 denote primes.

Case 2c: x starts with 9.

Then we may write $x = 9y4z9$, where $y, z \in \{0, 6, 9\}^*$.

If $0 \lhd y$, then $9049 \lhd x$. If $6 \lhd y$, then $9649 \lhd x$. If $9 \lhd y$, then $9949 \lhd x$. If $0 \lhd z$, then $409 \lhd x$. If $9 \lhd z$, then $499 \lhd x$.

Hence we may assume $x \in 946^*9$. If x contains three or more 6's, then $946669 \lhd x$. Hence we may assume x contains zero, one, or two 6's. But the numbers 949, 9469, and 94669 are all composite.

This completes the proof. ■

In the same fashion we can compute the minimal elements of

$$\texttt{COMPOSITES} = \{4, 6, 8, 9, 10, 12, 14, 15, 16, 18, \ldots\}.$$

Theorem 2 *If* $S = \texttt{COMPOSITES}$*, then*

$$M(S) = \{4, 6, 8, 9, 10, 12, 15, 20, 21, 22, 25, 27, 30, 32, 33, 35, 50, 51, 52, 55, 57, 70, 72, 75, 77, 111, 117, 171, 371, 711, 713, 731\}.$$

Proof. The proof is similar to that of Theorem 1, and is left to the reader. ■

Although we know $M(S)$ is always finite, no matter what S is, in some cases it can be very difficult to compute $M(S)$. For example, we have the following

Conjecture 3 *If* $S = \texttt{POWERS-OF-2} = \{1, 2, 4, 8, 16, 32, 64, \ldots\}$*, then*

$$M(S) = \{1, 2, 4, 8, 65536\}.$$

This conjecture seems very difficult to resolve at present. It would follow if we knew, for example, that every power of 16 greater than 65536 contains a digit in the set $\{1, 2, 4, 8\}$.

The reader may enjoy trying to compute $M(S)$ for some other classical sets, such as the squares in base 10, or computing the minimal elements of the primes expressed in bases other than 10.

References

[1] I. O. Angell and H. J. Godwin. On truncatable primes. *Math. Comp.* **31** (1977), 265–267.

[2] C. K. Caldwell. Truncatable primes. *J. Recreational Math.* **19** (1987), 30–33.

[3] C. K. Caldwell and H. Dubner. The near-repunit primes $1_{n-k-1}0_1 1_k$. *J. Recreational Math.* **27** (1995), 35–41.

[4] L. E. Card. Patterns in primes. *J. Recreational Math.* **1** (1968), 93–99.

[5] S. P. Huestis. Recursively laminar primes. *J. Recreational Math.* **25** (1993), 171–173.

[6] S. Kahan and S. Weintraub. Left truncatable primes. *J. Recreational Math.* **29** (1998), 254–264.

[7] M. Lothaire. *Combinatorics on Words*, Vol. 17 of *Encyclopedia of Mathematics and Its Applications*. Addison-Wesley, 1983.

[8] S. Yates. The mystique of repunits. *Math. Magazine* **51** (1978), 22–28.