ON STORING AND ANALYZING LARGE STRINGS OF PRIMES

SOL WEINTRAUB*
Queens College, CUNY, Flushing, New York

By the prime number theorem, the number of primes less than x is asymptotic to x/log x. A short table of actual counts follows.

RANGE	NO. OF PRIMES
$0 - 2.5 \times 10^6$	183,072
$10^8 - 10^8 + 2.5 \times 10^6$	135,775
$10^{10} - 10^{10} + 2.5 \times 10^6$	108,527
$10^{12} - 10^{12} + 2.5 \times 10^6$	90,509
$10^{14} - 10^{14} + 2.5 \times 10^6$	77,254
$10^{16} - 10^{16} + 2.5 \times 10^6$	68,081

Computer runs for finding the larger numbers are very time-consuming and it is often desirable to store the primes on magnetic tape or punched cards for use in certain statistical routines. Many users also store the lower primes for computing the higher ones, applying some variation of the sieve of Eratosthenes.

Assume we want to store the 68,081 primes in the interval from 10^{16} to $10^{16} + 2.5 \times 10^6$ on punched cards. How many cards are required? The first prime is 10 000 000 000 000 061 (17 digits) and if all digits are used, we would require $68{,}081 \times 17/80$ (a card can hold 80 alphanumeric characters) or 14,468 cards.

Obviously, we don't need to record the value of 10^{16} for every prime. We can store only the last seven digits (since we have an interval of 2,500,000) and keep in mind that every number is to be augmented by 10^{16}. Using only the last 7 digits requires $68{,}081 \times 7/80$ or 5958 cards.

Now, we don't have to store the actual primes. If we record the first one we need simply store the difference to the next one. For example, the second prime in this interval is 10 000 000 000 000 069 and so we just record the number 8. The next one is $10^{16} + 79$ and we record the number 10. Allowing for a 3-digit maximum difference (the actual maximum

*Research supported in part by AEC Grant AT(30-1)37 64.

difference is 432, i.e., 431 consecutive composites) we now need $68{,}080 \times 3/80$ or 2553 cards.

It is desirable to cut this number down stil further. There are 68,081 primes in this interval of 2,500,000 numbers. Thus the average difference is about 36. Indeed, a computer count reveals that 52,273 of the 68,081 primes have gaps (differences) less than or equal to 52. Moreover, since all primes are odd (except for the number 2) all differences are even; and we need to store only half the difference (keeping in mind that when reconstructing the primes from the differences, we will double the gap). Thus for most of the gaps we could use a number from 1 to 26 or a single letter from A to Z.

What about a gap of 54? This would be stored as 1A. The numeric 1 signifying 52 and A a difference of 2. A gap of 104 would appear as 1Z and 106 would be 2A. This method allows for a difference up to 572 using the ten numeric digits and 26 alphabetics. (It could be extended in an obvious fashion by having two numerics precede the alphabetic, etc.) A numeric digit is present only if it precedes an alphabetic, never by itself.

As an example, consider the first three cards for the primes after 10^{16}. The first prime $(10^{16} + 61)$ is recorded elsewhere and the first letter (D) gives the increment to the next prime, and so on.

DEJ6UN2FHMQ1PTMURFKIDQS1JN2C1AIE1BW1A1JAH1SA1DBDFBLIVT1G1KRBO1A1G1G
F1MU1JCSOK1EF

1RSTGNMOLIB1PF1A2FLML1LVCTAFNLJTRDC1DIRHYXILI1IU1BTL1G2RE1EHMHG1GEL
LUFJHA2JLJEY1

DHYBF1E1VUKACLT1QFXUTRJ1ILC1TB2FNMN1SCRDCCRI1LC2Q1GIA1DH1PCO1AL2COE
M1SC1D1AE1NQA

These three cards* translate to the 190 primes:

$$(61),\ 69,\ 79,\ 453,\ \cdots,\ 7357,\ 7359\ .$$

The last A on the third card indicating the twin prime (a gap of 2). With this system the number of cards needed to store the primes between 10^{16} and $10^{16} + 2.5$ million reduces to 1048. (About half of a box.) Of course, cards are only an illustration. The same economy is effected using magnetic tape, terminal display, or any other device.

Based on the above rules, a computer program could easily construct and reconstruct the primes in any given interval. (It is desirable to store the last prime, as well as the first, for a check.)

*Comment. Two lines represent one card. Our margins required putting each card in two lines.

For many applications, however, it is not necessary to reconstruct the primes. For example, if one wishes to find the number of twin primes in an interval one simply looks for isolated A's (A's not preceded by a numeric character). Or one could have the program search for the combination ABA signifying a quadruple of primes within a span of eight integers; this occurs for example at $10^{16} + 2,470,321$, 323, 327, 329; as indicated in the following line:

ONM1V1FAXQA1ATR1SY1C<u>ABA</u>2GOABJRICOLQILDU1VI1V2EWJIFQFSHRAFONAQMHPRH

M1F2TVOK1AFJOE

Similarly, one can search for any permissible combination of letters. Certain sequences are obviously forbidden; such as AA which would mean that p, p + 2, p + 4 are all primes and evidently one of these is divisible by 3. FIBONACCI, for example, is also forbidden. An interesting problem is: what is the probability that a random sequence of N letters is permissible? Is Shakespeare's Macbeth, word for permissible word, somewhere amongst the primes? After all, as x goes to infinity, so does $x/\log x$.

Finally, is there a way of storing primes (or any similar string of numbers) using fewer characters? How close can one come to using only <u>one</u> binary bit (0 or 1) for each prime?

REFERENCES

1. D. Shanks, <u>Solved and Unsolved Problems in Number Theory</u>, Spartan, Washington, 1962.

2. D. Shanks, "On Maximal Gaps between Successive Primes," <u>Mathematics of Computation</u>, Vol. 18, No. 88, October 1964, pp. 646-651.

3. D. H. Lehmer, "Tables Concerning the Distribution of Primes up to 37 Millions," 1957, copy deposited in the UMT File and reviewed in MTAC, Vol. 13, 1959, pp. 56-57.

4. S. Weintraub, "Primes between 10^{14} and $10^{14} + 10^8$," 1971, copy deposited in the UMT File of <u>Mathematics of Computation</u> and to be reviewed in a future issue.