

may be used to establish an authenticated connection to be used conventionally. The intrinsic security requirements of a public-key authentication server are easier to meet than those of a conventional one, but a complete evaluation of the system problems in implementing such a server in a real system, and the need to retain a secure record of old public keys to guarantee future correct arbitration of old signatures may minimize this advantage. We conclude that the choice of technique should be based on the economy and cryptographic strength of the encryption techniques themselves, rather than for their effects on protocol complexity.

Finally, protocols such as those developed here are prone to extremely subtle errors that are unlikely to be detected in normal operation. The need for techniques to verify the correctness of such protocols is great, and we encourage those interested in such problems to consider this area.

Acknowledgments. We are indebted to a number of people who have read drafts of this paper and made careful and helpful comments, notably: Peter Denning, Stockton Gaines, Jim Gray, Steve Kent, Gerry Popek, Ron Rivest, Jerry Saltzer, and Robin Walker.

Received September 1977; revised April 1978; final revision May 1978

References

1. Branstad, D. Security aspects of computer networks, Proc. AIAA Comptr. Network Syst. Conf., April 1973, paper 73-427.
2. Branstad, D. Encryption protection in computer data communications. Proc. Fourth Data Communications Symp., Oct. 1975, pp. 8.1-8.7 (available from ACM, New York).
3. Diffie, W., and Hellman, M. Multiuser Cryptographic Techniques, Proc AFIPS 1976 NCC, AFIPS Press, Montvale, N.J., pp. 109-112.
4. Feistel, H. Cryptographic coding for data bank privacy. Res. Rep. RC2827, IBM T.J. Watson Res. Ctr., Yorktown Heights, N.Y., March 1970.
5. Kent, S. Encryption-based protection protocols for interactive user-computer communication, M.S. Th., EECS Dept., M.I.T., 1976; also available as Tech. Rep. 162, Lab. for Comptr. Sci., M.I.T., Cambridge, Mass., 1976.
6. Kent, S. Encryption-based protection for interactive user/computer communication. Proc. Fifth Data Communication Symp., Sept. 1977, pp. 5-7-5-13 (available from ACM, New York).
7. National Bureau of Standards. Data Encryption Standard. Fed. Inform. Processing Standards Pub. 46, NBS, Washington, D.C., Jan. 1977.
8. Pohlig, S. Algebraic and combinatoric aspects of cryptography. Tech. Rep. No. 6602-1, Stanford Electron. Labs., Stanford, Calif., Oct. 1977.
9. Rivest, R.L., et al. A method for obtaining digital signatures and public-key cryptosystems. *Comm. ACM* 21, 2 (Feb. 1978), 120-126.

Programming
Techniques

S. L. Graham
Editor

A Linear Sieve Algorithm for Finding Prime Numbers

David Gries
Cornell University

Jayadev Misra
University of Texas at Austin

A new algorithm is presented for finding all primes between 2 and n . The algorithm executes in time proportional to n (assuming that multiplication of integers not larger than n can be performed in unit time). The method has the same *arithmetic complexity* as the algorithm presented by Mairson [6]; however, our version is perhaps simpler and more elegant. It is also easily extended to find the prime factorization of all integers between 2 and n in time proportional to n .

Key Words and Phrases: primes, algorithms, data structures

CR Categories: 5.25, 5.24, 5.29

1. Introduction

An algorithm is presented for finding all primes between 2 and n , for $n \geq 4$, that executes in time proportional to n . Like the sieve of Eratosthenes, it works by removing nonprimes from the set $\{2, \dots, n\}$. Unlike the sieve of Eratosthenes, no attempt is ever made to remove a nonprime that was removed earlier; this allows us to develop a linear algorithm.

The algorithm deals with sets S satisfying $S \subset \{2, \dots, n\}$. Two operations will be required on such sets:

Permission to copy without fee all or part of this material is granted provided that the copies are not made or distributed for direct commercial advantage, the ACM copyright notice and the title of the publication and its date appear, and notice is given that copying is by permission of the Association for Computing Machinery. To copy otherwise, or to republish, requires a fee and/or specific permission.

This research was partially supported by the National Science Foundation under Grants DCR75-09842 and MCS76-22360.

Authors' addresses: D. Gries, Computer Science Department, Cornell University, Ithaca, NY 14853; J. Misra, Computer Science Department, University of Texas at Austin, Austin, TX 78712.
© 1978 ACM 0001-0782/78/1200-0999 \$00.75

delete by executing $q := next(S, q)$; $k, x := 1, p \cdot q$. Similarly, Lemmas 3 and 4 state the conditions under which $y = next(S, p)^2$. We shall prove these lemmas in Section 3.

LEMMA 1. *Invariant P implies that $next(S, q)$ is defined, $next(S, q) < n$, and $p < lp(next(S, q))$. Writing $y = \chi(p, next(S, q), 1)$, we have $x \alpha y$.*

LEMMA 2. *Suppose (P and $x > n$). Write $y = \chi(p, next(S, q), 1)$. Then no nonprime z in S satisfies $x \alpha z \alpha y$.*

LEMMA 3. *Invariant P implies that $next(S, p)$ is defined, $next(S, p) < n$, and $next(S, p)$ is prime. Writing $y = \chi(next(S, p), next(S, p), 1)$, we have $x \alpha y$.*

LEMMA 4. *Suppose (P and $x > n$ and $p \cdot next(S, q) > n$). Write $y = next(S, p)^2$. Then no nonprime z in S satisfies $x \alpha z \alpha y$.*

We write Algorithm 1 using guarded commands [1]. The arguments necessary to ascertain correctness will be discussed in Section 3. Note that variable k is used only in assignments to itself, so that all references to it may be deleted. It has been included only to clarify the relationship between p, q , and x .

ALGORITHM 1.

```

{ $n \geq 4$ }
 $p, q, k, x, S := 2, 2, 1, 4, \{2, \dots, n\}$ ;
do  $x \leq n$ 
    →  $remove(S, x)$ ;
        $k, x := k + 1, p \cdot x$ 
||  $x > n$  and  $p \cdot next(S, q) \leq n$  →  $q := next(S, q)$ ;
        $k, x := 1, p \cdot q$ 
||  $x > n$  and  $p \cdot next(S, q) > n$  and  $next(S, p)^2 \leq n$ 
    →  $p := next(S, p)$ ;
        $q, k, x := p, 1, p \cdot p$ 
od
( $S = \{y | 2 \leq y \leq n$  and  $y$  prime)

```

Algorithm 2 is essentially the same algorithm as Algorithm 1 but written more conventionally. We feel that Algorithm 1 is easier to understand and prove correct; one loop with one invariant is easier to understand in this instance than three nested loops with three invariants.

ALGORITHM 2.

```

 $p, S := 2, \{2, \dots, n\}$ ;
while  $p \cdot p \leq n$  do begin
     $q := p$ ;
    while  $p \cdot q \leq n$  do begin
         $x := p \cdot q$ ;
        while  $x \leq n$  do begin
             $remove(S, x)$ ;  $x := p \cdot x$ 
        end;
         $q := next(S, q)$ 
    end;
     $p := next(S, p)$ 
end

```

3. Showing Correctness and Linearity

In this section the proofs of Lemmas 1–4 are given. The axiomatic proof method with respect to guarded

commands [1] is also discussed and some of the details of the proof are given.

In preparation for proving Lemmas 1 and 2, we first prove the following.

LEMMA 5. *Consider any nonprime $z = \chi(\bar{p}, \bar{q}, \bar{k})$. We have (P and $x \alpha z$ and $\bar{q} \leq n$) implies $\bar{q} \in S$.*

PROOF. If $\bar{q} (\leq n)$ is prime it is in S . Suppose \bar{q} is nonprime. From $x \alpha z$ and the decompositions of x and z we deduce $lp(x) = p \leq \bar{p} < lp(\bar{q})$ so that $x \alpha \bar{q}$. From the definition of S and $x \alpha \bar{q}$ we deduce $\bar{q} \in S$. \square

Our proofs of Lemmas 1 and 2 rest on the remarkable fact that for any positive integer $i > 1$ there is a prime v satisfying $i < v < 2i$.²

PROOF OF LEMMA 1. Let v be a prime satisfying $q < v < 2 \cdot q$. From P we conclude

$$p \leq q < next(S, q) \leq v < 2 \cdot q \leq p \cdot q \leq n$$

and hence $next(S, q) < n$. Secondly, no nonprime in S has a divisor less than p , so that $p \leq lp(next(S, q))$. To show that $p \neq lp(next(S, q))$, consider the fact that any nonprime $z = \chi(p, \bar{q}, \bar{k})$ in S must have $q \leq \bar{q}$. Hence the smallest such nonprime z that may be in S is $p \cdot q$. Since $next(S, q) < p \cdot q$, $next(S, q)$ cannot be a nonprime with $p = lp(next(S, q))$.

Hence $p < lp(next(S, q))$. The relation $x \alpha y = \chi(p, next(S, q), 1)$ follows immediately.

PROOF OF LEMMA 2. From $x = \chi(p, q, k) > n$ and $y = \chi(p, next(S, q), 1)$, we see that a z in S satisfying $x \alpha z \alpha y$ would have a decomposition $z = \chi(p, \bar{q}, \bar{k})$ with $q < \bar{q} < next(S, q)$. But \bar{q} would not be in S , contradicting Lemma 5.

PROOF OF LEMMA 3. Let v be a prime satisfying $p < v < 2 \cdot p$. From P we have

$$p < next(S, p) \leq v < 2 \cdot p \leq p^2 \leq n$$

and $next(S, p) < n$. Secondly, no nonprime in S has a divisor less than p , so that for all nonprimes $z \in S$ we have $p^2 \leq z$. Since $next(S, p) < p^2$, $next(S, p)$ must be prime. The fact $x \alpha y = \chi(next(S, p), next(S, p), 1)$ follows immediately.

PROOF OF LEMMA 4. This is similar to the proof of Lemma 2 and is left to the reader.

We now discuss the proof method and give some details. The main part of Algorithm 1 is a loop of the form

do $B1 \rightarrow SL1$ || $B2 \rightarrow SL2$ || $B3 \rightarrow SL3$ od

Showing correctness involves exhibiting an invariant P (ours is given in Definition 2) and an integer function t , and showing that the following hold:

² As might be imagined, this is difficult to prove. J. Bertrand conjectured this fact in 1845, after showing empirically that it was true for $i \leq 10^6$. Chebyshev proved the conjecture in 1850 (see [5] for details). Our first draft of a proof and algorithm did not rely on this fact at all. It used weaker lemmas with more complicated proofs and required the additional element $n + 1$ to be in S so that $next(S, i)$ would be sure to be defined. For example, the original Lemma 1 read: *Let $y = p \cdot next(S, q)$. Suppose P and $x > n$. Then either $next(S, q) = n + 1$; or $y = \chi(p, next(S, q), 1)$ and $x \alpha y$.*

- (1) P is true before execution of the loop;
- (2) P and not ($B1$ or $B2$ or $B3$) implies the desired result;
- (3) $\{P \text{ and } Bi\} SLi \{P\}$ for $i = 1, 2, 3$;
- (4) Execution of the loop terminates:
 - (a) $(P \text{ and } (B1 \text{ or } B2 \text{ or } B3)) \Rightarrow t \geq 0$
 - (b) Execution of SLi reduces t (for $1 \leq i \leq 3$). Using an extra variable T , this means that

$$\{P \text{ and } Bi\} T := t; SLi \{t \leq T - 1\}.$$

Point (1) is obvious; point (2) we leave to the reader, since it can be shown quite easily with the help of Lemma 4. Point (3) concerns the invariance of P under execution of each guarded command SLi . The only difficulty concerns the generation of new values for q , p , and x to satisfy P . Lemmas 1–4 yield the necessary facts.

To see this a bit more formally in at least one case, consider determining the precondition Q in $\{Q\} SL3 \{P\}$ where $SL3$ is the third guarded command list of the loop of Algorithm 1 and P is in Definition 2. $SL3$ is a sequence of assignments, so we apply the normal assignment and concatenation rules to arrive at:

- $$Q = \begin{aligned} &(1) \text{ next}(S, p) \text{ prime, } 4 \leq \text{next}(S, p)^2 \leq n; \\ &(2) \text{ next}(S, p) = \text{next}(S, p) \text{ or } (\text{next}(S, p) < lp(\text{next}(S, p)); \\ &\quad \text{next}(S, p)^2 \leq n; \\ &(3) 1 \leq i; \\ &(4) \text{ next}(S, p)^2 = \chi(\text{next}(S, p), \text{next}(S, p), 1); \\ &(5) S = \{2, \dots, n\} - \{y | y \text{ nonprime and } y \alpha \text{next}(S, p)^2\}. \end{aligned}$$

It is then a simple matter to prove that $(P \text{ and } B3)$ implies Q , with the help of Lemmas 3 and 4. To show termination, we use the function t : the number of nonprimes $z \in S$ satisfying $(x = z \text{ or } x \alpha z)$ plus the number of nonprimes $z \in S$ satisfying $x \alpha z$.

Note that $t \geq 0$. Execution of the first guarded command $SL1$ reduces the first term of t by at least one, since it removes x from S . Execution of the second or third guarded command begins with $x = x0$ (say) and $x \notin S$ and finishes with $x0 \alpha x$ and $x \in S$; hence they reduce the second term by one. Hence we conclude that the algorithm terminates.

The initial value for t is a bound on the number of times the loop will iterate. The initial value is bounded by $2 \cdot (\text{number of nonprimes in } S) < 2 \cdot n$. Hence the algorithm is linear if we can satisfactorily implement S and the operations on it. The implementation is the subject of Section 4.

4. Implementing the Set S

We discuss three approaches to implement $S \subset \{2, \dots, n\}$, all dealing with forms of linked lists. We will actually implement sets $S \cup \{n + 1\}$. The purpose is to provide an “anchor” for one end of the linked list. The integer 2 serves the same purpose at the other end of the list since it is never deleted.

Approach 1. If we implement S as a doubly linked list, then $remove(S, i)$ and $next(S, i)$ each can be performed in constant time. Thus we use

var s : array $(2:n + 1)$ of record $(pred, succ:integer)$

where the various parts of s are used as follows:

$$\begin{aligned} s(i).succ &= \text{next}(S \cup \{n + 1\}, i) \text{ for } i \in S; \\ s(i).pred &= \text{unique integer } j \text{ such that } \text{next}(s, j) = i, \text{ for } i \in S \cup \\ &\quad \{n + 1\}, i \neq 2. \end{aligned}$$

At any time, the elements of S can be found by following the successor chain beginning at $s(2)$ and ending just before $s(n + 1)$. This approach requires roughly $2n$ locations (each of $\ln n$ bits).³ The three operations on S are:

$$\begin{aligned} S := \{2, \dots, n\} &:: i := 1; \\ &\quad \text{do } i < n \rightarrow i := i + 1; s(i).succ := i + 1; \\ &\quad \quad \quad s(i + 1).pred := i \text{ od}; \\ \text{remove}(S, i) &:: s(s(i).pred).succ := s(i).succ; \\ &\quad \quad \quad s(s(i).succ).pred := s(i).pred; \\ \text{next}(S, i) &:: s(i).succ \end{aligned}$$

*Approach 2.*⁴ Under the assumptions that 2 is not removed from S and $remove(S, i)$ is only executed if i is in S , a singly linked list can be used to implement $S \in \{2, \dots, n\}$. We use an array s :

var s : array $(2..n)$ of integer.

If $i \in S$, then $s(i)$ will contain its successor $\text{next}(S \cup \{n + 1\}, i)$. It will also be necessary to determine i 's predecessor. Clearly, if $i \in S$ and $(i - 1) \in S$, then $i - 1$ is i 's predecessor and $s(i - 1) = i$. On the other hand, if $i \in S$ and $(i - 1) \notin S$, the element $s(i - 1)$ can be used to contain i 's predecessor. Note then that for $i > 2$, $i \in S$, i 's predecessor is given by $\min(i - 1, s(i - 1))$. Thus the array s above is used to describe S as follows:

$$\begin{aligned} 2 \in S; \\ \text{for } i \in S, s(i) &= \text{next}(S \cup \{n + 1\}, i); \\ \text{for } i \in S, i \neq 2, &\text{ the predecessor } i \text{ is } \min(i - 1, s(i - 1)). \end{aligned}$$

At any time S is the set of values 2, $s(2)$, $s(s(2))$, ... , up to but not including the value $n + 1$.

The three operations $S := \{2, \dots, n\}$, $remove(S, i)$, and $next(S, i)$ are then

$$\begin{aligned} S := \{2, \dots, n\} &:: i := 1; \\ &\quad \text{do } i < n \rightarrow i := i + 1; s(i) := i + 1 \text{ od}; \\ \text{remove}(S, i) &:: \text{pred} := \min(i - 1, s(i - 1)); [\text{pred is local}] \\ &\quad \quad \quad s(\text{pred}) := s(i); \\ &\quad \quad \quad s(s(i) - 1) := \text{pred}; \\ \text{next}(S, i) &:: s(i) \end{aligned}$$

Approach 3. In Approach 2, the value $s(i)$ always contains an integer representing a successor or predecessor j of value i in the set. Instead, one could store the increment needed to get from i to j ; that is, the value $i - j$. Since the increment will in general be much smaller

³ Throughout, \ln refers to the base 2 logarithm.

⁴ In our initial formulation of Approach 2, we used an extra bit vector in , where $in(i)$ denoted whether or not $i \in S$. $remove(S, i)$ simply set $in(i)$ to false. However, $next(S, i)$ was a complicated operation that in order to achieve a linear algorithm, had the “benevolent side effect” of changing the representation of S without changing the values of $next(S, i)$ for any i . During a presentation of the algorithm at Harvard, Norman Cohen proposed a technique that was further refined by the authors to yield the present Approach 2. The development of $next(S, i)$ in the case that $remove(S, i)$ simply sets $in(i)$ to false might be an interesting exercise for the reader.

than n , we may be able to reduce the number of bits needed for each $s(i)$. Asymptotically speaking, there are $n/\ln n$ or more primes in $\{2, \dots, n\}$, and if they were evenly distributed the increment to get from one to another would not be greater than $\ln n$. Hence we would need only $\text{ceil}(\ln \ln n)$ bits for each $s(i)$ instead of $\text{ceil}(\ln n)$.

Unfortunately, the primes are not evenly distributed, so we cannot assume $s(i)$ will be so small. Knuth [4, p. 402] gives a table of "record breaking" gaps between prime numbers. For primes less than or equal to 20831533 we see that the largest gap is 210, so that eight bits will suffice for $n \leq 20831533$.

5. Discussion

Mairson's [6] paper, which tied for second place in the annual George E. Forsythe Student Paper Competition, also presents a linear sieve algorithm. To delete from S all composite integers whose lowest prime factor is p , Mairson's algorithm first uses the set S to compile a list of these integers, then sequences through this list to delete them from S . The use of an auxiliary list is unnecessary in our algorithm because of Theorem 2.1. Mairson does an excellent job in analyzing the efficiency of his algorithm, and most of his analyses will carry over to our algorithm.

Algorithm 1 also works for $n = 2$ and $n = 3$; the condition $n \geq 4$ is used only to simplify the proof.

Dexter Kozen has discovered an easy way to extend the algorithm to find the complete factorization of an integer n in no worse than linear time. This extension is not surprising, since Shank's [8] algorithm finds factors of n in time $O(n^{(1/4+\epsilon)})$ for $\epsilon > 0$. However, Kozen's technique actually can be used to build a table in time n that yields the complete factorization of *all* integers between 2 and $n!$ It is quite simple. Assume a singly linked list is used to implement set S , say using Approach 2, and use three new arrays $xp, xk, xq(2, \dots, n)$. Initially, let $xp(x) = x$, for all x . When a nonprime $x = p^k \cdot q$ is about to be deleted in Algorithm 1 the values p, k , and q are available, so just record them in $xp(x), xk(x)$, and $xq(x)$. Upon termination, for each $i, 2 \leq i \leq n$, if $xp(i) < i$ then i is not prime, i 's lowest prime factor is in $xp(i)$ and its multiplicity in $xk(i)$, while the other factors can be determined from $xq(i)$ in a similar fashion.

The development of this algorithm emphasizes several points. First, it could not have been developed without recognition of an important property of nonprimes—their unique decomposition given in Theorem 2.1. Efficient algorithms come less from clever tricks than from a good understanding of properties of the values being manipulated. Secondly, the correctness of the algorithm rests on some nontrivial mathematical theorems (Lemmas 1–4). Once these theorems are understood, the algorithm itself seems quite simple. We see here a distinction between the complexity of an algo-

rithm and the complexity of its mathematical underpinnings, two quite different things. Finally, the algorithm provides a good basis for a discussion of control structures and programming style—we showed two ways of writing the algorithm—and for a discussion of the selection of data structures.

Acknowledgments. We note that Gale and Pratt [2] have discovered a different linear sieve algorithm. The authors would like to thank Jim Donahue, Don Knuth, and Gary Levin for carefully reading earlier drafts of this paper and for providing many constructive criticisms; and Norman Cohen and Dexter Kozen for their suggestions.

Received June 1977; revised April 1978

References

1. Dijkstra, E.W. Guarded commands, nondeterminacy and formal derivation of programs. *Comm. ACM* 18, 8 (Aug. 1975), 453–457.
2. Gale, R., and V. Pratt. CGOL—an algebraic notation for MACLISP users. Working paper, M.I.T. AI Lab., Cambridge, Mass., January 1977.
3. Gries, D. The multiple assignment statement. *IEEE Trans. Software Eng. SE-2* (March 1978), 89–93.
4. Knuth, D. *The Art of Computer Programming, Vol. 3: Sorting and Searching*. Addison-Wesley, Reading, Mass., 1973.
5. LeVeque, W.J. *Topics in Number Theory, Volume I*. Addison-Wesley, Reading, Mass., 1956.
6. Mairson, H.G. Some new upper bounds on the generation of prime numbers. *Comm. ACM* 20, 9 (Sept. 1977), 664–669.
7. Miller, G.L. Riemann's hypothesis and tests for primality. Proc. Seventh Annual ACM Symp. Theory of Computing, 1975, 234–239.
8. Shank, D. Class number, a theory of factorization and Genera. Proc. Symp. in Pure Mathematics 20, 1969, Amer. Math. Soc., Providence, R.I., 1971, 415–440.