Assignment 5
Public Key Cryptography

Prof. Darrell D. E. Long
CSE 13S – Winter 2023

First `DESIGN.pdf` draft due: February 16th at 11:59 pm PST
Assignment due: February 26th at 11:59 pm PST

1 Introduction

> *Doo Jdxo lv glylghg lqwr wkuhh sduwv, rqh ri zklfk wkh Ehojdh lqkdelw, wkh Dtxlwdql dqrwkhu, wkrvh zkr lq wkhlu rzq odqjxdjh duh fdoohg Fhowv, lq rxu Jdxov, wkh wklug.*
>
> —Julius Cæsar

Cryptography, once restricted to government, spies, and the military is now pervasive in our lives. Most web sites that you visit are protected using SSL. Your SSH connections are protected in the same way.

How is this accomplished? Through a mixture of *public key* and *symmetric key* cryptography. The earliest known practical public-key cryptography algorithm is RSA, after its inventors Ronald Rivest, Adi Shamir, and Leonard Adleman (Figure 2), who published it in 1978. About five years earlier, on 20 November, 1973, Clifford Cocks (Figure 1), working for GCHQ (the British equivalent of the NSA), invented a very similar algorithm. His classified memorandum "A note on 'non-secret' encryption" was to remain secret for 24 years. In fact, when you read the Cocks memorandum, you will see that the *idea* of public key encryption was proposed by J. H. Ellis three years earlier in 1970. Unknown in the public literature, the idea was independently proposed by Ralph Merkle for public key distribution, which inspired asymmetric cryptography by Whitfield Diffie and Martin Hellman, and finally leading to RSA. RSA in turn gave rise to numerous related algorithms such as the *Schmidt-Samoa* algorithm, which will be the focus of this assignment. While Schmidt-Samoa has not gained wide adoption as compared to RSA, variety is the spice of life.

Public-key cryptography, or asymmetric cryptography, is a cryptographic system that uses pairs of keys: public keys (known to others) and private keys (known only by the owner). The generation of such key pairs depends on cryptographic algorithms that are based on mathematical objects called *one-way functions*. Security requires keeping the private key private; the public key can be distributed widely.

Any person can encrypt a message using the intended receiver's public key, but that encrypted message can only be decrypted with the receiver's private key. This allows a server to create a cryptographic key for suitable symmetric-key cryptography and then use a client's openly shared public key to encrypt the newly generated symmetric key. The server can then send this encrypted symmetric key over an insecure channel to the client; only the client can decrypt it using its private key. With the client and server both having the same symmetric key, they can safely use symmetric key encryption to communi-

cate. This scheme has the advantage of not having to pre-share symmetric keys while gaining the higher speed of symmetric-key cryptography.

Symmetric-key algorithms use the same cryptographic keys for the encryption of plaintext and the decryption of ciphertext. The keys may be identical, or there may be a simple transformation between the two keys. The keys represent a shared secret between two or more parties. The requirement that both parties have access to the secret key is one of the main disadvantages of symmetric-key encryption compared to public-key encryption.

Let's briefly look at the Cocks algorithm before moving on to the SS algorithm. We have two principals: *Alice* (A), who is the receiver, and *Bonnie* (B), who is the sender.

(a) Alice:

   i. Chooses two primes p and q such that $p \nmid (q-1)$ and $q \nmid (p-1)$. That is, p does not divide $q-1$ and q does not divide $p-1$.
   ii. Transmits the computed product $n = pq$ to the sender, which we write as $A \xrightarrow{n} B$.

(b) Bonnie:

   i. Has a message consisting of numbers $c_1, c_2, \ldots, c_r$ where $0 < c_i < n$.
   ii. Sends these encoded as d_i where $d_i = c_i^n \pmod{n}$. When written as part of a protocol, $B \xrightarrow{d_1,\ldots,d_r} A$.

(c) Alice:

   i. Computes using Euclid's Algorithm p' such that $p \times p' \equiv 1 \pmod{q-1}$, and q' such that $q \times q' \equiv 1 \pmod{p-1}$.
   ii. Decodes $c_i = d_i^{p'} \pmod{q} = d_i^{q'} \pmod{p}$.

Figure 1: Clifford Cocks

As with the SS algorithm, as you will see, the strength of the algorithm relies on the assumed difficulty of factoring large composite integers. We say *assumed difficulty* since there is, like $\mathscr{P} \stackrel{?}{=} \mathscr{NP}$, no proof of this widely held assumption. A proof that $\mathscr{P} \neq \mathscr{NP}$ would be welcome, but unsurprising, while a proof that $\mathscr{P} = \mathscr{NP}$ would probably have theoreticians jumping out of windows.

The paper published by Rivest, Shamir and Adleman in 1978,

> Ronald L. Rivest, Adi Shamir, and Leonard Adleman. "A method for obtaining digital signatures and public-key cryptosystems," *Communications of the ACM* 21.2 (1978): 120–126.

is one of the most important papers ever published. It enabled the modern Internet and changed the world. You would do well to take the time to read it.

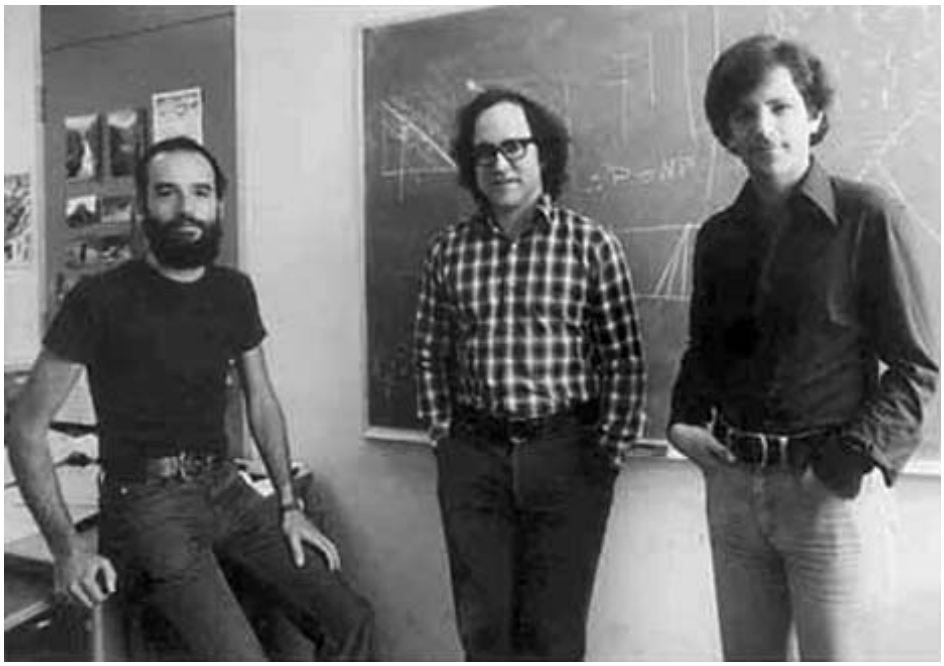


Figure 2: Adi Shamir, Ronald Rivest, and Leonard Adelman

2 Schmidt-Samoa (SS) Algorithm

> *The magic words are* Squeamish Ossifrage.
>
> —Ronald L. Rivest

The security of RSA relies on the practical difficulty of factoring the product of two large prime numbers, known as the *factoring problem*. However, doing so would allow an adversary to factor *all* ciphertexts encrypted under that key. Decrypting particular RSA ciphertexts is known as the *RSA problem*. At first look it would appear that the RSA problem may be easier than factoring, and there are some cryptographers who believe this is the case. Though there are no published, efficient methods to break RSA if a large enough key is used – short of building a *quantum computer* and running Shor's algorithm on it – this theoretical gap between the factoring problem and the RSA problem is not ideal. The Schmidt-Samoa algorithm rectifies this theoretical problem, by modifying the RSA algorithm in such a way that the SS problem is *provably equivalent* to the factoring problem. However, this property comes as the cost of efficiency, which is why today's cryptographic protocols have largely ignored SS in favor of RSA.

SS involves a public key and a private key. Everyone can know the public key, and it is used for encrypting messages. The intention is that messages encrypted with the public key can only be decrypted

by using the private key.

The public key consists of the modulus n, which is composed of two large, random primes p and q. The private key consists of the private exponent d and the private modulus $\text{lcm}(p-1, q-1)$, which must be kept secret; p, q must also be kept secret since they are used to calculate d and $\text{lcm}(p-1, q-1)$. In fact, p and q can be discarded after d has been computed.

We proceed by choosing two large random primes p and q, these numbers must be kept secret. In particular we want large random primes, where $p \nmid q-1$ and $q \nmid p-1$ and both $p-1$ and $q-1$ have large prime factors. We then publish the number $n = p^2q$. You might wonder why we can do this, and the reason is that it is *believed* to be hard to factor large composite integers into their constituent primes. The *fundamental theorem of arithmetic* tells us that every integer has a *unique* prime factorization.

We now calculate $\text{lcm}(p-1, q-1)$, the private modulus, and a unique secret integer $d \in \{0, \ldots, n-1\}$ such that $d \times n \equiv 1 \pmod{\text{lcm}(p-1, q-1)}$. How do we find this d? It turns out that we have known how to do it for more than 2300 years—we use an algorithm attributed to *Euclid*. How is it that we can easily calculate d while our adversary cannot? We know a secret that he does not: we know p and q while he only knows n. We call d our private exponent. Together pq and d make up our *private key*.

We now define two functions: $E(m) = m^n \pmod n$ and $D(c) = c^d \pmod{pq}$. We will show in §3 that $\forall m \in \{0, \ldots, pq-1\}$ that $D(E(m)) = m$.

There's a catch here. Say we want to publish the public key, so that people can encrypt messages for us. We know that our message space is $\{0, \ldots, pq-1\}$, and that messages outside of that space will not decrypt correctly, so along with n we have to publish pq, right? *Wrong!* If we did that, then anyone could calculate $p = n/pq$ and $q = pq/p$. With that information anyone can determine our private key! Instead, we observe that

$$\begin{aligned}\sqrt{n} &= p\sqrt{q} \\ &< pq,\end{aligned}$$

so instead of using our full message space, $\{0, ..., pq-1\}$, we will just allow messages in a truncated message space $\{0, \sqrt{n}-1\}$.

How do we know that this doesn't break our security like we would have had we published pq? Because, we are just computing some function of the public information n. For a system to be secure, it must be able to keep the private key secret against adversaries with polynomial resources and access to n. Because Schmidt-Samoa is secure, asking people to compute a polynomial time algorithm with n as the input can't give the adversary any advantage that he didn't already have!

This issue highlights something important: don't write your own cryptographic algorithms! (this assignment excluded) It's very easy to make mistakes and write algorithms that look secure but are trivial to break. As Bruce Schneier said, "anyone, from the most clueless amateur to the best cryptographer, can create an algorithm that he himself can't break."

3 Mathematics of SS

> *If $\mathscr{P} = \mathscr{NP}$, then all of modern cryptography collapses. On this happy thought...*
>
> Michael O. Rabin, November 1998

The mathematics of SS are based on arithmetic in a set of integers modulo n, denoted $\mathbb{Z}/n$. This is the set $\{0, \ldots, n-1\}$ and all sets that are equivalent to it. For example, $\{n, \ldots, 2n-1\} \pmod n = \{0, \ldots, n-1\}$

$(\text{mod } n)$, and there are an infinite number of such sets. Since they are *all the same* we will only concern ourselves with the one with the smallest numbers. What do we mean when we say that are the same? We mean that if $x \equiv k \pmod{n}$ then $an + x \equiv k \pmod{n}, \forall a \geq 0, a \in \mathbb{Z}$. In other words, additional integer products of n do not matter.

The Euler-Fermat theorem says that if $a \in \mathbb{N}$ and $n \in \mathbb{N}$ are *coprime*, that is $\gcd(a, n) = 1$, then

$$a^{\varphi(n)} \equiv 1 \pmod{n}.$$

This will allow us, for example, to take a message M and have $M^{\varphi(n)} \equiv 1 \pmod{n}$.

What is $\varphi(n)$? It is the Euler totient function, and gives the number of positive integers than that or equal to n that are relatively prime to n. For any prime number p,

$$\varphi(p) = p - 1.$$

For the SS algorithm, we choose two large primes p and q, where $p \nmid q-1$ and $q \nmid p-1$, and we make $n = p^2 q$. Then we calculate d such that $d \cdot n \equiv 1 \pmod{\text{lcm}(p-1, q-1)}$. What does this mean for us with respect to $\varphi(pq)$?

$$\begin{aligned}\varphi(pq) &= \varphi(p) \times \varphi(q) \\ &= (p-1)(q-1) \\ &= pq - (p+q) + 1.\end{aligned}$$

Now because we chose p and q such that $p \nmid q-1$ and $q \nmid p-1$, $n = p^2 q$ is relatively prime to $\varphi(pq)$, that is, $\gcd(n, \varphi(pq)) = 1$. Because of this we know that a decryption key d exists, such that $n \times d \equiv 1 \pmod{\varphi(pq)}$. Our encryption algorithm is simply $E(M) = M^n \pmod{n} = C$, and our decryption algorithm is $D(C) = C^d \pmod{pq} = M$.

Observe that,

$$M^{nd} \equiv M^{k\varphi(pq)+1} \pmod{pq}$$

for some integer $k \geq 1$. This is true because, prior to the modular reduction, nd must be one greater than some multiple of $\varphi(pq)$; applying the modulus is what makes $nd \equiv 1 \pmod{\varphi(pq)}$. We can rewrite this as

$$M^{nd} \equiv (M^{\varphi(pq)})^k \times M \pmod{pq}.$$

Here we apply Euler's theorem, which states that if a and n are coprime integers, then $a^{\varphi(n)} \equiv 1 \pmod{n}$. So, assuming that M is coprime with pq, we can simplify the above equation to

$$\begin{aligned}M^{nd} &\equiv (1)^k \times M && \pmod{pq} \\ &\equiv (1) \times M && \pmod{pq} \\ &\equiv M && \pmod{pq}\end{aligned}$$

which proves that decryption really does work on encrypted messages.

In practice, Carmichael's function can be used in place of Euler's totient in the computation of the SS key pair. Carmichael's function is denoted with $\lambda(n)$, where $\lambda(n) = \text{lcm}(\lambda(p), \lambda(q))$. We indicate the least common multiple of some a and b with $\text{lcm}(a, b)$. From the definition of the least common multiple, we see that

$$\begin{aligned}\lambda(n) &= \lambda(pq) \\ &= \text{lcm}(\lambda(p), \lambda(q)) \\ &= \frac{|\lambda(p) \times \lambda(q)|}{\gcd(\lambda(p), \lambda(q))}.\end{aligned}$$

If p and q are prime, then we know $\lambda(p) = \varphi(p) = p-1$, and $\lambda(q) = \varphi(q) = q-1$. Thus,

$$\begin{aligned}\lambda(n) &= \frac{|\lambda(p) \times \lambda(q)|}{\gcd(\lambda(p), \lambda(q))} \\ &= \frac{|\varphi(p) \times \varphi(q)|}{\gcd(p-1, q-1)} \\ &= \frac{|\varphi(n)|}{\gcd(p-1, q-1)}.\end{aligned}$$

It should be clear from this that $\varphi(pq)$ is a multiple of $\lambda(pq)$. Therefore,

$$\begin{aligned}M^{nd} &\equiv (M^{\varphi(pq)})^k \times M && \pmod{pq} \\ &\equiv (M^{j \times \lambda(pq)})^k \times M && \pmod{pq} \\ &\equiv (M^{\lambda(pq)})^{jk} \times M && \pmod{pq}\end{aligned}$$

for some multiplier $j \geq 1$.

From Carmichael's generalization of Euler's theorem, we know that $a^{\lambda(pq)} \equiv 1 \pmod{pq}$. Thus,

$$\begin{aligned}M^{nd} &\equiv (M^{\lambda(pq)})^{jk} \times M && \pmod{pq} \\ &\equiv (1)^{jk} \times M && \pmod{pq} \\ &\equiv (1) \times M && \pmod{pq} \\ &\equiv M && \pmod{pq},\end{aligned}$$

which demonstrates how $\lambda(pq)$ can be used in place of $\varphi(pq)$ for the computation of the SS key pair. For your assignment, you will be using Carmichael's function when computing your SS key pair.

4 Your Task

> *"Personally," he said, "my great ambition is to count all this,"—he waved vaguely at the treasure around him—"and possibly sort it into piles."*
>
> —John C. Gardner, *Grendel*

You will be creating three programs for this assignment:

1. A key generator: `keygen`
2. An encryptor: `encrypt`
3. A decryptor: `decrypt`

The `keygen` program will be in charge of key generation, producing SS public and private key pairs. The `encrypt` program will encrypt files using a public key, and the `decrypt` program will decrypt the encrypted files using the corresponding private key.

You will need to implement two libraries and a random state module that will be used in each of your programs. One of the libraries will be hold functions relating to the mathematics behind SS, and the other library itself will contain implementations of routines for SS. You also need to learn to *use* a library: the GNU multiple precision arithmetic library.

5 GNU Multiple Precision Arithmetic

> *One reason you should not use web applications to do your computing is that you lose control. It's just as bad as using a proprietary program. Do your own computing on your own computer with your copy of a freedom-respecting program. If you use a proprietary program or somebody else's web server, you're defenceless.*
>
> —Richard Stallman

As you should know by now, **C**, unlike languages like **Python**, does not natively support arbitrary precision integers. The security of SS, however, relies on large integers. So, we elect to use the GNU multiple precision arithmetic library, usually referred to as GMP. You can find the manual and documentation for the library here: https://gmplib.org/manual.
Take some time to look through the manual, taking note of which functions may be useful.

You will need to install both `gmp` and `pkg-config`. The latter is a utility used to assist in finding and linking libraries, instead of having the program hard-code where to find specific headers and libraries during program compilation. To install these packages on Ubuntu 20.04, run the following:

```
$ sudo apt install pkg-config libgmp-dev
```

Get started on this *as soon as possible*. Make sure to attend section for assistance on using `pkg-config` in a `Makefile` to direct the compilation process for your programs.

You may notice that GMP already provides number theoretic functions, several of which *could* be used in SS. You may not use any GMP-implemented number theoretic functions. You *must* implement those functions yourself.

The following two sections (§6 and §7) will present the functions that you have to implement, but they both will require the use of random, arbitrary-precision integers.

GMP requires us to explicitly initialize a random state variable and pass it to any of the random integer functions in GMP. Not only that, we also need to call a dedicated function to clean up any memory used by the initialized random state. To remedy this, you will implement a small random state module, which contains a single `extern` declaration to a global random state variable called `state`, and two functions: one to initialize the state, and one to clear it. The interface for the module will be given in `randstate.h` and the implementation must go in `randstate.c`.

`void randstate_init(uint64_t seed)`

Initializes the global random state named `state` with a Mersenne Twister algorithm, using `seed` as the random seed. You should call `srandom()` using this seed as well. This function will also entail calls to `gmp_randinit_mt()` and to `gmp_randseed_ui()`.

`void randstate_clear(void)`

Clears and frees all memory used by the initialized global random state named `state`. This should just be a single call to `gmp_randclear()`.

6 Number Theoretic Functions

> *No one has yet discovered any warlike purpose to be served by the theory of numbers or relativity, and it seems unlikely that anyone will do so for many years.*
>
> —G. H. Hardy

Number Theory is the branch of mathematics that studies the nature and properties of numbers. Though many have made important contributions to the field, including Gauß in his *Disquisitiones Arithmeticae* (which he completed when he was 21 years old), the most important for public-key cryptography are Fermat and Euler (Figure 3).

You will first need to implement the functions that drive the mathematics behind SS before you can tackle your SS library. The interface for these functions will be given in `numtheory.h` and should be defined in corresponding **C** file. Read each of the subsections carefully to understand, on some level, the theory behind each of the number theoretic functions. Pseudocode is provided to assist you.

Figure 3: Leonhard Euler (1707–1783) and Pierre de Fermat (1607–1665)

6.1 Modular Exponentiation

As shown in §2, we must compute a^n where both $a, n \in \mathbb{N}$ for SS . We could simply multiply:

$$a^n = \overbrace{a \times a \times \cdots \times a \times a}^{n}.$$

The number of multiplications is $n-1$, which is $O(n)$. While correct, this approach is naïve and *extremely inefficient*. Since we are working with very large numbers in SS, we must be able to compute modular exponentiation quickly. So the question is, can we do better? We can in fact do much better, computing a^n in $O(\log_2(n))$ steps.

Recall that we can write any integer as a polynomial

$$n = c_m 2^m + c_{m-1} 2^{m-1} + \cdots + c_1 2^1 + c_0 2^0 = \sum_{0 \le i \le m} c_i 2^i,$$

where $n \ge 2^m$ and $c_i \in \{0, 1\}$. And so,

$$a^n = a^{c_m 2^m + c_{m-1} 2^{m-1} + \cdots + c_1 2^1 + c_0 2^0}.$$

Since $a^{b+c} = a^b \times a^c$, then we can rewrite the formula as

$$a^n = a^{c_m 2^m} \times a^{c_{m-1} 2^{m-1}} \times \cdots \times a^{c_1 2^1} \times a^{c_0 2^0} = \prod_{0 \le i \le m} a^{c_i 2^i}.$$

As an example, consider $a^{13} = a^{2^3+2^2+2^0} = a^{8+4+1} = a^8 \times a^4 \times a^1$. You will want to try a few more to get a feeling for it before you attempt to write code.

This leaves us with the problem of computing the a^{2^i} terms. We start with $a = a^1$ and if we square it then $(a^1)^2 = a^2$. Each time we square, $(a^2)^2 = a^4, (a^4)^2 = a^8, \ldots$ the exponents are a power of 2. We only have to square our previous result $\log_2 n$ times at most.

You will notice that the numbers get *very large, very fast.* Although we want enormous numbers for cryptography, we do not want numbers that would be impossible to even write down if we used every atom in the universe. Recall that 10^k is k digits long. That means that if $k = 10^{1000}$ then there are that many digits (there are approximately 10^{82} atoms in the observable universe). Consequently, we will usually do such computations (mod n) for some modulus n, meaning that all numbers are in $\{0, \ldots, n-1\}$.

To implement modular exponentiation, you should simply follow the steps to perform exponentiation by squaring as shown above and reduce your results modulo n after each operation that is likely to yield a large result (you do not need to do it, if for example, you just add a small constant). The following pseudocode shows the repeated squaring and modular reduction at each step.

```
POWER-MOD(a, d, n)
1  v ← 1
2  p ← a
3  while d > 0
4      if ODD(d)
5          v ← (v × p) mod n
6      p ← (p × p) mod n
7      d ← ⌊d/2⌋
8  return v
```

The function that you are expected to implement to perform modular exponentiation should be declared as follows:

```
void pow_mod(mpz_t out, mpz_t base, mpz_t exponent, mpz_t modulus)
```

Performs fast modular exponentiation, computing `base` raised to the `exponent` power modulo `modulus`, and storing the computed result in `out`.

6.2 Primality Testing

> *There are many methods—none of them as good as the randomized primality test.*
>
> —Michael O. Rabin, October 1997

The simplest primality test is trial division: given an input number, n, check whether it is evenly divisible

by any prime number between 2 and $\sqrt{n}$. Thus, this simple algorithm[1] is $O(\sqrt{n})$, but can we do better? The answer is subtle. To be certain, we must try all of the primes from up to $\sqrt{n}$; there is no way to escape it. But we can do much better if we are willing to accept an answer of *probably*.

Since it is infeasible to use a *deterministic* algorithm, we can solve many problems with *high probability* by using a *randomized algorithm*. Such algorithms explore random parts of the problem space so that we have high (but not perfect) confidence that they have solved the problem.

Probabilistic tests are more rigorous than heuristic tests in that they provide provable bounds on the probability of being fooled by a composite number. All practical primality tests are probabilistic tests. These tests use, apart from the tested number n, some other number a (called a *witness*) that is chosen at random from some sample space. The usual randomized primality tests never report a prime number as composite, but a composite number may be reported as prime.

The simplest probabilistic primality test is the Fermat primality test. It works as follows: Given an integer n, choose some integer a coprime to n and calculate $a^{n-1} \pmod n$. If the result is different from 1, then n is composite. If it is 1, then n may be prime. If $a^{n-1} \equiv 1 \pmod n$ n is not prime, then n is called a pseudoprime to base a. In practice, we observe that, if $a^{n-1} \equiv 1 \pmod n$, then n is usually prime.

Better yet is the Solovay-Strassen probabilistic primality test, developed by Robert M. Solovay and Volker Strassen in 1977. It is of particular importance since it made practical public-key algorithms such as SS.

Figure 4: Gary L. Miller and Michael O. Rabin

The Miller–Rabin primality test, invented by Gary Miller and Michael O. Rabin (Figure 4), is an even more sophisticated probabilistic test, which detect all composites (once again, this means: for every composite number n, at least $\frac{3}{4}$ of numbers a are witnesses of compositeness of n). The accuracy of these tests is compared in Figure 5.

The Miller–Rabin primality test works as follows: Given an integer n, choose some positive integer $a < n$. Let $2^s d = n-1$, where d is odd. If $a^d \not\equiv \pm 1 \pmod n$ and $a^{2^r d} \not\equiv \pm 1 \pmod n$ for all $0 \le r \le s-1$, then n is composite and a is a witness for the compositeness. Otherwise, n *might* be prime. That is, we might be wrong $\frac{1}{4}$ of the time. If we repeat the test 100 times then our chance of being wrong is

[1] How big is $\sqrt{n}$? A typical encryption key has more than 600 decimal digits. Thus, $\sqrt{10^{600}} = 10^{300}$. Suppose we can do one trial division every *nanosecond*, then that's $10^{300-9} = 10^{291}$ seconds. There are 22,896,000 or about 10^7 seconds per year, so it will take about 10^{284} years (the Big Bang was about 13.7×10^9 years ago).

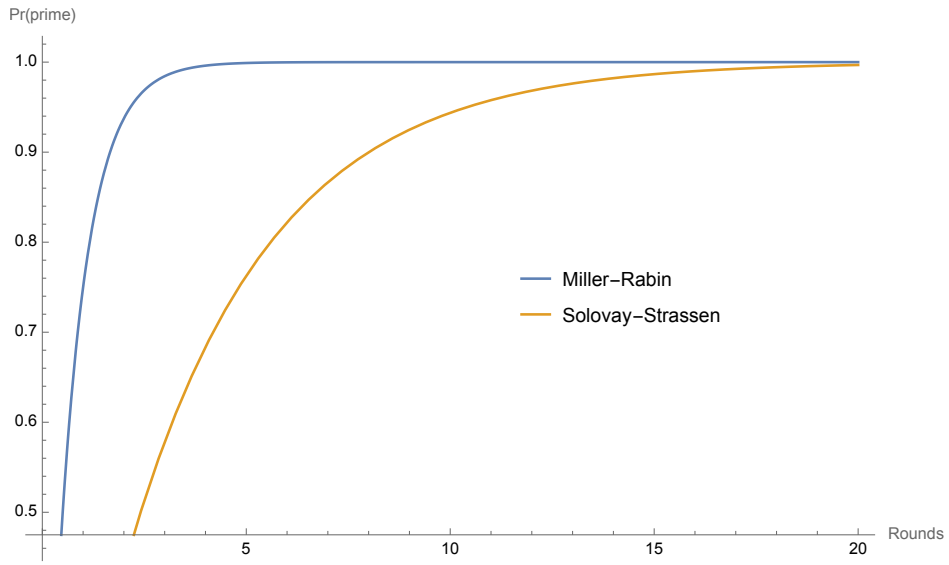


Figure 5: Pr[prime(p)] after successfully passing a given number of rounds.

$(\frac{1}{4})^{100} = 2^{-200}$ and that is usually more than good enough. The Miller–Rabin test is considered a strong pseudoprime test, where a pseudoprime is a number that is determined to be *probably prime* by a probabilistic test, but not actually prime. Deterministic primality tests, such as the AKS primality test, do not give false positives.

MILLER-RABIN(n, k)

```
1  write n − 1 = 2^s r such that r is odd
2  for i ← 1 to k
3      choose random a ∈ {2, 3, ..., n − 2}
4      y = POWER-MOD(a, r, n)
5      if y ≠ 1 and y ≠ n − 1
6          j ← 1
7          while j ≤ s − 1 and y ≠ n − 1
8              y ← POWER-MOD(y, 2, n)
9              if y == 1
10                 return FALSE
11             j ← j + 1
12         if y ≠ n − 1
13             return FALSE
14 return TRUE
```

The function that you are expected to implement to perform primality testing should be declared as follows:

```
bool is_prime(mpz_t n, uint64_t iters)
```

Conducts the Miller-Rabin primality test to indicate whether or not `n` is prime using `iters` number of

Miller-Rabin iterations. This function is needed when creating the two large primes p and q in SS, verifying if a large integer is a prime.

In addition to the `is_prime()` function, you are also required to implement the following function:

```
void make_prime(mpz_t p, uint64_t bits, uint64_t iters)
```

Generates a new prime number stored in p. The generated prime should be at least `bits` number of bits long. The primality of the generated number should be tested using `is_prime()` using `iters` number of iterations.

6.3 Modular Inverses

The Euclidean algorithm, also called Euclid's algorithm, is an efficient method for computing the greatest common divisor (gcd) of two integers, the largest number that divides them both with a zero remainder. The Euclidean algorithm is based on the principle that the greatest common divisor of two numbers does not change if their difference replaces the larger number with the smaller number. Since this replacement reduces the larger of the two numbers, repeating this process gives successively smaller pairs of numbers until the two numbers become equal. We can accomplish this much faster if we compute the remainder, which is equivalent to subtracting the smaller number from the larger until it is no longer larger. You will first want to implement the following function to compute the greatest common divisor of two integers, which should be defined as follows:

```
void gcd(mpz_t d, mpz_t a, mpz_t b)
```

Computes the greatest common divisor of a and b, storing the value of the computed divisor in d.

```
GCD(a, b)
1  while b ≠ 0
2      t ← b
3      b ← a mod b
4      a ← t
5  return a
```

The extended Euclidean algorithm is an extension to the Euclidean algorithm, and computes, in addition to the greatest common divisor (gcd) of integers a and b, also the coefficients of Bézout's identity, which are integers x and y such that

$$ax + by = \gcd(a, b).$$

The extended Euclidean algorithm is particularly useful when a and b are coprime. With that provision, x is the modular multiplicative inverse of $a \pmod{b}$, and y is the modular multiplicative inverse of $b \pmod{a}$.

Bézout's identity asserts that a and n are coprime if and only if there exist integers s and t such that

$$ns + at = 1.$$

Reducing this identity modulo n gives

$$at \equiv 1 \pmod{n}.$$

To adapt the extended Euclidean algorithm to the problem of computing the multiplicative inverse, note that the Bézout coefficient of n is not needed and so does not need to be computed. Also, for getting a positive and result that is less than n, use the fact that the integer t provided by the algorithm satisfies $|t| < n$. That is, if $t < 0$, add n to it at the end.

MOD-INVERSE(a, n)

```
1  (r, r') ← (n, a)
2  (t, t') ← (0, 1)
3  while r' ≠ 0
4      q ← ⌊r/r'⌋
5      (r, r') ← (r', r − q × r')
6      (t, t') ← (t', t − q × t')
7  if r > 1
8      return no inverse
9  if t < 0
10     t ← t + n
11 return t
```

The function that you are expected to implement to compute modular inverses should be declared as follows:

`void mod_inverse(mpz_t i, mpz_t a, mpz_t n)`

Computes the inverse `i` of `a` modulo `n`. In the event that a modular inverse cannot be found, set `i` to 0. Note that this pseudocode uses parallel assignments, which **C** *does not* support. Thus, you will need to use auxiliary variables to fake the parallel assignments.

7 An SS Library

> *If you think cryptography is the answer to your problem, then you don't know what your problem is.*
>
> —Peter G. Neumann

`void ss_make_pub(mpz_t p, mpz_t q, mpz_t n, uint64_t nbits, uint64_t iters)`

Creates parts of a new SS public key: two large primes `p` and `q`, and `n` computed as `p * p * q`. Begin by creating primes `p` and `q` using `make_prime()`. We first need to decide the number of bits that go to each prime respectively such that $\log_2(n) \geq$ `nbits`. Let the number of bits for `p` be a random number in the range $[$`nbits`$/5, (2 \times$ `nbits`$)/5)$. Recall that $n = p^2 \times q$: the bits from `p` will be contributed to `n` twice, the remaining bits will go to `q`. The number of Miller-Rabin iterations is specified by `iters`. You should obtain this random number using `random()` and check that $p \nmid q-1$ and $q \nmid p-1$.

`void ss_write_pub(mpz_t n, char username[], FILE *pbfile)`

Writes a public SS key to `pbfile`. The format of a public key should be `n`, then the `username`, each of which are written with a trailing newline. The value `n` should be written as a *hexstring*. See the GMP functions for formatted output for help with writing hexstrings.

`void ss_read_pub(mpz_t n, char username[], FILE *pbfile)`

Reads a public SS key from `pbfile`. The format of a public key should be `n`, then the `username`, each of which should have been written with a trailing newline. The value `n` should have been written as a *hexstring*. See the GMP functions for formatted input for help with reading hexstrings.

`void ss_make_priv(mpz_t d, mpz_t pq, mpz_t p, mpz_t q)`

Creates a new SS private key `d` given primes `p` and `q` and the public key `n`. To compute `d`, simply compute the inverse of `n` modulo $\lambda(pq) = \text{lcm}(p-1, q-1)$.

`void ss_write_priv(mpz_t pq, mpz_t d, FILE *pvfile)`

Writes a private SS key to `pvfile`. The format of a private key should be `pq` then `d`, both of which are written with a trailing newline. Both these values should be written as hexstrings.

`void SS_read_priv(mpz_t pq, mpz_t d, FILE *pvfile)`

Reads a private SS key from `pvfile`. The format of a private key should be `pq` then `d`, both of which should have been written with a trailing newline. Both these values should have been written as *hexstrings*.

`void ss_encrypt(mpz_t c, mpz_t m, mpz_t n)`

Performs SS encryption, computing the ciphertext `c` by encrypting message `m` using the public key `n`. Remember, encryption with SS is defined as $E(m) = c = m^n \pmod n$.

`void ss_encrypt_file(FILE *infile, FILE *outfile, mpz_t n)`

Encrypts the contents of `infile`, writing the encrypted contents to `outfile`. The data in `infile` should be in encrypted in *blocks*. Why not encrypt the entire file? Because of `n`. We are working modulo `n`, which means that the value of the block of data we are encrypting must be strictly less than `n`. We have two additional restrictions on the values of the blocks we encrypt:

1. The value of a block cannot be 0: $E(0) \equiv 0 \equiv 0^n \pmod n$.
2. The value of a block cannot be 1. $E(1) \equiv 1 \equiv 1^n \pmod n$.

A solution to these additional restrictions is to simply *prepend* a single byte to the front of the block we want to encrypt. The value of the prepended byte will be `0xFF`. This solution is not unlike the padding schemes such as PKCS and OAEP used in modern constructions of RSA. To encrypt a file, follow these steps:

1. Calculate the block size k. This should be $k = \lfloor (\log_2(\sqrt{n}) - 1)/8 \rfloor$.

2. Dynamically allocate an array that can hold k bytes. This array should be of type `(uint8_t *)` and will serve as the block.

3. Set the zeroth byte of the block to `0xFF`. This effectively prepends the workaround byte that we need.

4. While there are still unprocessed bytes in `infile`:

   (a) Read at most $k-1$ bytes in from `infile`, and let j be the number of bytes actually read. Place the read bytes into the allocated block starting from index 1 so as to not overwrite the `0xFF`.

   (b) Using `mpz_import()`, convert the read bytes, including the prepended `0xFF` into an `mpz_t` `m`. You will want to set the `order` parameter of `mpz_import()` to `1` for most significant word first, `1` for the `endian` parameter, and `0` for the `nails` parameter.

   (c) Encrypt `m` with `ss_encrypt()`, then write the encrypted number to `outfile` as a *hexstring* followed by a trailing newline.

`void ss_decrypt(mpz_t m, mpz_t c, mpz_t d, mpz_t pq)`

Performs SS decryption, computing message `m` by decrypting ciphertext `c` using private key `d` and public modulus `n`. Remember, decryption with SS is defined as $D(c) = m = c^d \pmod{pq}$.

`void ss_decrypt_file(FILE *infile, FILE *outfile, mpz_t pq, mpz_t d)`

Decrypts the contents of `infile`, writing the decrypted contents to `outfile`. The data in `infile` should be decrypted in *blocks* to mirror how `ss_encrypt_file()` encrypts in blocks. To decrypt a file, follow these steps:

1. Dynamically allocate an array that can hold $k = \lfloor (\log_2(pq) - 1)/8 \rfloor$ bytes. This array should be of type `(uint8_t *)` and will serve as the block.

   - We need to ensure that our buffer is able to hold at least the number of bits that were used during the enryption process. In this context we don't know the value of `n`, but we can overestimate the number of bits in $\sqrt{n}$ using `pq`.

   $$\log_2(\sqrt{n}) = \log_2(\sqrt{p^2 \times q}) = \log_2(p \times \sqrt{q}) < \log_2(pq)$$

2. Iterating over the lines in `infile`:

   (a) Scan in a hexstring, saving the hexstring as a `mpz_t c`. Remember, each block is written as a hexstring with a trailing newline when encrypting a file.

   (b) First decrypt `c` back into its original value `m`. Then using `mpz_export()`, convert `m` back into bytes, storing them in the allocated block. Let j be the number of bytes actually converted. You will want to set the `order` parameter of `mpz_export()` to `1` for most significant word first, `1` for the `endian` parameter, and `0` for the `nails` parameter.

   (c) Write out $j-1$ bytes starting from index 1 of the block to `outfile`. This is because index 0 must be prepended `0xFF`. Do not output the `0xFF`.

8 Key Generator

> *This method, seemingly very clever, actually played into our hands! And so it often happens that an apparently ingenious idea is in fact a weakness which the scientific cryptographer seizes on for his solution.*
>
> —Herbert Yardley, *The American Black Chamber*

Your key generator program should accept the following command-line options:

- `-b`: specifies the minimum bits needed for the public modulus `n`.
- `-i`: specifies the number of Miller-Rabin iterations for testing primes (default: 50).
- `-n pbfile`: specifies the public key file (default: ss.pub).
- `-d pvfile`: specifies the private key file (default: ss.priv).
- `-s`: specifies the random seed for the random state initialization (default: the seconds since the UNIX epoch, given by `time(NULL)`).
- `-v`: enables verbose output.
- `-h`: displays program synopsis and usage.

The program should follow these steps:

1. Parse command-line options using `getopt()` and handle them accordingly.
2. Open the public and private key files using `fopen()`. Print a helpful error and exit the program in the event of failure.
3. Using `fchmod()` and `fileno()`, make sure that the private key file permissions are set to `0600`, indicating read and write permissions for the user, and no permissions for anyone else.
4. Initialize the random state using `randstate_init()`, using the set seed.
5. Make the public and private keys using `ss_make_pub()` and `ss_make_priv()`, respectively.
6. Get the current user's name as a string. You will want to use `getenv()`.
7. Write the computed public and private key to their respective files.
8. If verbose output is enabled print the following, each with a trailing newline, in order:
   (a) username
   (b) the first large prime `p`
   (c) the second large prime `q`
   (d) the public key `n`
   (e) the private exponent `d`
   (f) the private modulus `pq`

All of the `mpz_t` values should be printed with information about the number of bits that constitute them, along with their respective values in *decimal*. See the reference key generator program for an example.

9. Close the public and private key files, clear the random state with `randstate_clear()`, and clear any `mpz_t` variables you may have used.

9 Encryptor

> *The laws of mathematics are very commendable, but the only law that applies in Australia is the law of Australia.*
>
> —Malcolm Turnbull, Australian Prime Minister

Your encryptor program should accept the following command-line options:

- `-i` : specifies the input file to encrypt (default: `stdin`).
- `-o` : specifies the output file to encrypt (default: `stdout`).
- `-n` : specifies the file containing the public key (default: `ss.pub`).
- `-v` : enables verbose output.
- `-h` : displays program synopsis and usage.

The program should follow these steps:

1. Parse command-line options using `getopt()` and handle them accordingly.
2. Open the public key file using `fopen()`. Print a helpful error and exit the program in the event of failure.
3. Read the public key from the opened public key file.
4. If verbose output is enabled print the following, each with a trailing newline, in order:
   (a) username
   (b) the public key `n`

   All of the `mpz_t` values should be printed with information about the number of bits that constitute them, along with their respective values in *decimal*. See the reference encryptor program for an example.
5. Encrypt the file using `ss_encrypt_file()`.
6. Close the public key file and clear any `mpz_t` variables you have used.

10 Decryptor

So much technology, so little talent.

—Vernor Vinge, *Rainbow's End*

Your decryptor program should accept the following command-line options:

- `-i`: specifies the input file to decrypt (default: `stdin`).
- `-o`: specifies the output file to decrypt (default: `stdout`).
- `-n`: specifies the file containing the private key (default: `ss.priv`).
- `-v`: enables verbose output.
- `-h`: displays program synopsis and usage.

The program should follow these steps:

1. Parse command-line options using `getopt()` and handle them accordingly.
2. Open the private key file using `fopen()`. Print a helpful error and exit the program in the event of failure.
3. Read the private key from the opened private key file.
4. If verbose output is enabled print the following, each with a trailing newline, in order:
   (a) the private modulus `pq`
   (b) the private key `d`

   Both these values should be printed with information about the number of bits that constitute them, along with their respective values in *decimal*. See the reference decryptor program for an example.
5. Decrypt the file using `ss_decrypt_file()`.
6. Close the private key file and clear any `mpz_t` variables you have used.

11 Deliverables

People are frugal in guarding their personal property; but as soon as it comes to squandering time they are most wasteful of the one thing in which it is right to be stingy.

—Seneca, On the Shortness of Life

You will need to turn in the following source code and header files:

1. `decrypt.c`: This contains the implementation and `main()` function for the `decrypt` program.

2. `encrypt.c`: This contains the implementation and `main()` function for the `encrypt` program.
3. `keygen.c`: This contains the implementation and `main()` function for the `keygen` program.
4. `numtheory.c`: This contains the implementations of the number theory functions.
5. `numtheory.h`: This specifies the interface for the number theory functions.
6. `randstate.c`: This contains the implementation of the random state interface for the SS library and number theory functions.
7. `randstate.h`: This specifies the interface for initializing and clearing the random state.
8. `ss.c`: This contains the implementation of the SS library.
9. `ss.h`: This specifies the interface for the SS library.

Your code must pass `scan-build` *cleanly*. If there are any bugs or errors that are false positives, document them and explain why they are false positives in your `README.md`. Improper explanations will *not* be considered. You will also need to turn in the following:

1. `Makefile`:
    - `CC = clang` must be specified.
    - `CFLAGS = -Wall -Wextra -Werror -Wpedantic` must be specified.
    - `pkg-config` to locate compilation and include flags for the GMP library must be used.
    - `make` must build the `encrypt`, `decrypt`, and `keygen` executables, as should `make all`.
    - `make decrypt` should build only the `decrypt` program.
    - `make encrypt` should build only the `encrypt` program.
    - `make keygen` should build only the `keygen` program.
    - `make clean` must remove all files that are compiler generated.
    - `make format` should format all your source code, including the header files.
2. `README.md`: This must use proper Markdown syntax and describe how to use your program and `Makefile`. It should also list and explain any command-line options that your program accepts. Any false positives reported by `scan-build` should be documented and explained here as well. Note down any known bugs or errors in this file as well for the graders.
3. `DESIGN.pdf`: This document *must* be a proper PDF. This design document must describe your design and design process for your program with enough detail such that a sufficiently knowledgeable programmer would be able to replicate your implementation. This does not mean copying your entire program in verbatim. You should instead describe how your program works with supporting pseudocode.
4. `WRITEUP.pdf`:This document must be a proper PDF done using LaTeX. You must use the *fullpage* and *fourier* packages. This write-up document should include everything you learned from this assignment. You should also discuss the applications of public-private cryptography and how it influences the world today. Be sure to describe at least one way in which you personally take advantage of public-private cryptography on a day-to-day basis.

12 Submission

> *"Nevertheless, something will come of all this," I said.*
> *"Nothing," he said. "A brief pulsation in the black hole of eternity. My advice to you—"*
> *"Wait and see," I said.*
> *He shook his head. "My advice to you, my violent friend, is to seek out gold and sit on it."*
>
> —John C. Gardner, *Grendel*

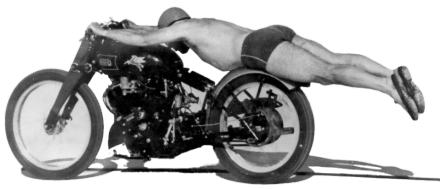
Refer back assignment 0 for the instructions on how to properly submit your assignment through `git`. Remember: *add*, *commit*, and *push*!

Your assignment is turned in *only* after you have pushed and submitted the commit ID you want graded on Canvas. "I forgot to push" and "I forgot to submit my commit ID" are not valid excuses. It is *highly* recommended to commit and push your changes *often*.

13 Supplemental Readings

> *The more that you read, the more things you will know. The more that you learn, the more places you'll go.*
>
> —Dr. Seuss

- *The C Programming Language* by Kernighan & Ritchie
  - Chapter 7
  - Appendix B
- *Introduction to Algorithms* by T. Cormen, C. Leiserson, R. Rivest, & C. Stein
  - Chapter 31 §31.2, §31.3, §31.6, §31.7, §31.8



et ecce simia pallidus et qui nomen illi ***C*** *et inferus sequebatur eum*