Rabin's Test for Primality

David Troiano
Math 124 Project
Professor William Stein

December 13, 2002

Contents

1 Introduction

At the age of 24, Gauss wrote in his famous work *Disquisitiones Arithmeticae* (1801), "The problem of distinguishing prime numbers from composite numbers and of resolving the latter into their prime factors is known to be one of the most important and useful in arithmetic" [3]. More succinctly, primality testing and factorization are important in mathematics (we focus on the former in this paper). Now, 200 years later, these two beasts still remain important in arithmetic and have proved instrumental in other fields. Many cryptographic schemes, for example RSA, ElGamal encryption, the Diffie-Hellman key agreement, and Rabin public-key encryption, assume the availability of an efficient mechanism for randomly generating large prime numbers. This problem of prime generation immediately reduces to primality testing. Prime number generation can be done by simple trial and error due to the density of primes. If we let $\pi(x)$ be the number of primes less than or equal to x, then $\lim_{x\to\infty} \frac{\pi(x)}{\frac{x}{\ln x}} = 1$, which means that, asymptotically, $\pi(x) \approx \frac{x}{\ln x}$. Thus, if we wish to generate a prime number with hundreds of digits, then we need only test *on the order of hundreds* of random numbers for primality before we expect to stumble upon one. In this paper we discuss a probabilistic primality test, due to Rabin, heretofore referred to as Rabin's Test.[1] After introducing and explaining the algorithm, we focus largely on the probability bounds that Rabin's Test can guarantee; that is, given a prime number, the algorithm always returns PRIME, but given a composite number, there is a negligibly small probability that the algorithm incorrectly returns PRIME. We discuss the theoretical bound (proved by Rabin in [6]) as well as the bound in practice, before finally discussing a Java applet implementation and directing the reader to try out Rabin's Test on the web.

2 Notation

Most notation used in this paper is standard. We use $\mathbb{E}$ and $\mathbb{O}$ to represent the even and odd integers, respectively. Also, we assume the reader is familiar with order notation. Formally, we say a function $f(n)$ is $O(g(n))$ if there exists positive constants c and n_0 such that $f(n) \leq c \cdot g(n) \forall n \geq n_0$. We loosen the definition and say an algorithm is $O(g(n))$ if the input is of size n for a reasonable representation of the input and it requires less than $c \cdot g(n)$ "steps" for some positive c and the inequality is true for all n greater than some starting point n_0.

[1] G.L. Miller first considered a similar deterministic test in [4] which assumes the correctness of the extended Riemann hypothesis, so the primality test we consider is often called the Miller-Rabin Primality Test.

3 Rabin's Test

3.1 Background

Fermat's Little Theorem states that if p is prime and a is some number between 1 and p, then $a^{p-1} \equiv 1 \bmod p$ (see [5] or any other introductory number theory text for a proof). Unfortunately, the converse is not true, i.e. given a such that $1 < a < p$, if $a^{p-1} \equiv 1 \bmod p$ then it is not necessarily the case that p is prime. Nonetheless, the contrapositive suggests the following primality test of a number n:

STRAWMAN'S-PRIMALITY-TEST (n)
    Let a be some number between 1 and n (perhaps 2)
    If $a^{n-1} \not\equiv 1 \bmod n$
        Return COMPOSITE
    Else
        Return PRIME
    Endif
End Procedure

The above test is very efficient, as modular exponentiation can be performed in $O(\log n)$ multiplications. In addition, if the above test returns COMPOSITE, we know by Fermat's Little Theorem that the number is indeed composite, and we call a a witness to the compositeness[2] of n. For example, if $n = 9$ and $a = 2$, $2^{9-1} \equiv 4 \bmod 9 \Rightarrow 9$ is composite, as 2 is a valid witness. Unfortunately, there is an infinite number of 2-pseudoprimes (composite numbers for which 2 is not a valid witness), meaning STRAWMAN'S-PRIMALITY-TEST will return PRIME given such a composite number as input with $a = 2$. For example, $341 = 11 \cdot 31$ is the smallest 2-pseudoprime (check that $2^{341-1} \equiv 1 \bmod 341$). This leads to the idea that maybe if we try a random a, or, better yet, if we run STRAWMAN'S-PRIMALITY-TEST k times with randomly chosen bases, then our test will not incorrectly return PRIME that often. For example, if $a = 3$, STRAWMAN'S-PRIMALITY-TEST will correctly return that 341 is composite because 341 is *not* a 3-pseudoprime. A repeated randomized test like this might perform well on random inputs, but unfortunately there are infinitely many composite numbers n such that $a^{n-1} \equiv 1 \bmod n$ for all a coprime to n. Such numbers are called Carmichael numbers.[3] $561 = 3 \cdot 11 \cdot 17$ is the smallest Carmichael number. Observe the following MAGMA code which returns the number of witnesses:

```
function test_carmichael(n)
    witnesses := 0;
    for a in [1..n] do
        if (Gcd(a, n) eq 1) and (not (a^ (n - 1) mod n) eq 1) then
```

[2]We will extend the definition of witness to compositeness when we get to Rabin't Test

[3]These numbers were explored by R.D. Carmichael in [1]

```
                witnesses := witnesses + 1;
            end if;
        end for;
        return witnesses;
    end function;
```

And from the prompt, we see there are no witnesses to the compositeness of 561 that are coprime to 561:

```
> test_carmichael(561);
0
```

Clearly, if we want a robust primality test, we need to do better than the STRAWMAN'S-PRIMALITY-TEST approach. We would like a primality test that performs well regardless of the type of input. Carmichael numbers are infinite but very rare; nonetheless, a good test is one which assumes that the person feeding input numbers is trying to crack the algorithm and make it incorrectly return PRIME on a composite number. Rabin offers the crucial step that gives us peace of mind and confidence in his primality test, even if the input is a Carmichael number. What's more, it's efficient enough to be used to generate prime numbers with hundreds of decimal digits.

3.2 The Algorithm

Given an odd number $n > 2$, we wish to determine if n is prime. If $n = 2$ returning PRIME is trivial. Also, if $n > 2$ is even, then it is trivial, in fact an $O(1)$ operation, to return COMPOSITE by observing that the rightmost bit of n is $0 \Leftrightarrow n$ is even. Thus we safely assume $n > 2$ is an odd number because in an actual implementation the $n \in \mathbb{E}$ case can be handled efficiently before applying Rabin's Test. Rabin's Test works as follows. Given $n \in \mathbb{O}$, let t and u be such that $n-1 = 2^t u$ where $t \geq 1$ and u is odd. Thus the binary representation of $n-1$ is the binary representation of u left-shifted t spaces and padded with zeroes. Note that $t \geq 1$ since n is odd $\Rightarrow n-1 \in \mathbb{E}$. Now repeatedly do the following, i.e. perform the rest of the algorithm k times where k is an assurance parameter. Choose a random base $a \in \{1, 2, \ldots, n-1\}$ and compute $x_0 \equiv a^u \bmod n$, $x_1 \equiv (a^u)^2 \bmod n \equiv a^{2u} \bmod n$, $\ldots$, $x_t \equiv a^{2^t u} \bmod n$. If ever $x_i \equiv 1 \bmod n$ and $x_{i-1} \not\equiv \pm 1 \bmod n$, then we have found a non-trivial square of 1 modulo n, so we return that n is COMPOSITE. Only composite numbers n exhibit the property that there are nontrivial square roots of 1 modulo n. As an example, observe that $7^2 = 49 \equiv 1 \bmod 16$, i.e. 7 is a square root of 1 modulo 16. For a proof of why only composite numbers exhibit this property, see the Appendix. Here is Rabin's Test in pseudocode:

```
RABIN'S-TEST (n)
    for iters ← 1 to k do
        Let a ←R {1, ..., n-1}
        Let u and t be such that n - 1 = 2^t u for u odd
```

```
        Let x_0 ← a^u mod n
        for j ← 1 to t do
            Let x_j ← x_{j-1}^2 mod n
            if x_j = 1 and x_{j-1} ≠ ±1 then
                return COMPOSITE†
            End if
        End for
        if x_t ≠ 1 then
            return COMPOSITE‡ (Fermat's Little Theorem)
        End if
    End for
    return PRIME
End Procedure
```

The correctness of a return value of COMPOSITE is argued as follows. If COMPOSITE is returned from ‡, then by Fermat's Little Theorem, n is composite. If COMPOSITE is returned from †, then we have found that x_{j-1} is a nontrivial square root of $x_j \equiv 1$ modulo n, and we know that 1 has nontrivial square roots modulo n only if n is composite (again, see the Appendix for a proof). Thus we have that RABIN'S-TEST will never return COMPOSITE when given a prime number as input. On the other hand, it is possible that the algorithm will return PRIME when given a composite number. For example, let $n = 17 \cdot 33 = 561$ and suppose our assurance is only $k = 1$ and $a = 50$ is selected as the base. Then you can check that $t = 4$ and $u = 35$. You can also check the following system of congruences with a big number calculator:

$$\begin{aligned} 50^{35} &\equiv -1 \bmod 561 \\ 50^{2 \cdot 35} &\equiv 1 \bmod 561 \\ 50^{2^2 \cdot 35} &\equiv 1 \bmod 561 \\ 50^{2^3 \cdot 35} &\equiv 1 \bmod 561 \\ 50^{2^4 \cdot 35} &\equiv 1 \bmod 561 \end{aligned} \tag{1}$$

Clearly, we've applied RABIN'S-TEST with $a = 50$ and Fermat's congruence holds and we haven't found any nontrivial square roots of 1 modulo 561. Naturally, we would like to bound above the probability that RABIN'S-TEST returns PRIME when the input is composite. Rabin shows in [6] that this bound is an amazing $\frac{1}{4}^k$.

4 Probability of Error

Rabin's fundamental theorem in [6] is that the number of bases in $\{1, 2, \ldots, n-1\}$ which are witnesses to the compositeness[4] of a number n is *greater than or equal to* $\frac{3(n-1)}{4}$. Since no more than $\frac{1}{4}$ of the bases are nonwitnesses to n's compositeness, this immediately leads us to the conclusion that if we try k bases independently chosen at random, if the number is prime we will definitely return PRIME, and if the number is composite, we will return COMPOSITE with probability greater than $1 - \frac{1}{4}^k$. This is an amazing result, and randomizing the process guarantees our probability bound regardless of the distribution of bases. We discuss the main theorem.

4.1 The Fundamental Theorem

In this paper we focus on a weaker but powerful result:

Theorem 1 *For a composite number n, the number of witnesses to the compositeness of n (the number of bases for which RABIN'S-TEST will return COMPOSITE) is at least* $\frac{n-1}{2}$.[5]

Proof: [6] The first piece of the proof is to show that for a composite number n, all nonwitnesses to compositeness must be elements of $\mathbb{Z}_n^*$. To see this, consider any nonwitness a. Without even considering our search for nontrivial square roots of 1 modulo n, we know that a must satisfy $a^{n-1} \equiv 1 \bmod n$, which can be written $a \cdot a^{n-2} \equiv 1 \bmod n \Rightarrow \exists$ a solution to the congruence $ax \equiv 1 \bmod n \Leftrightarrow$ [7] $\gcd(a, n) | 1 \Rightarrow \gcd(a, n) = 1$, which, by definition, means $a \in \mathbb{Z}_n^*$. We can actually characterize the nonwitnesses more tightly than this. That is, we will show that the nonwitnesses are actually *all* contained in a *proper* subgroup H of $\mathbb{Z}_n^*$. It's then a simple corollary of Lagrange that since H is proper, $|H| \leq \frac{1}{2}|\mathbb{Z}_n^*|$. So we have that since $|\mathbb{Z}_n^*| \leq n-1$, $|H| \leq \frac{1}{2}(n-1)$. Thus the number of nonwitnesses is *at most* $\frac{1}{2}(n-1)$ so the number of witnesses must be *at least* $\frac{1}{2}(n-1)$, which gives us our result. Thus we need to construct a proper subgroup H of $\mathbb{Z}_n^*$ of which every nonwitness is an element. There are 2 cases we explore depending on the pseudoprimality of n. In the first case, suppose $\exists x \in \mathbb{Z}_n^*$ such that n is composite using x and Fermat's Little Theorem. That is, $x^{n-1} \not\equiv 1 \bmod n$ for some $x \in \mathbb{Z}_n^*$. Then let $H = \{a \in \mathbb{Z}_n^* : a^{n-1} \equiv 1 \bmod n\}$. It's clear that H is a subgroup. Clearly $1 \in H$. Also, H is closed under multiplication in $\mathbb{Z}_n^*$, and the existence of inverses is simple to show. Thus H is a subgroup. Now since every nonwitness a satisfies $a^{n-1} \equiv 1 \bmod n$, all nonwitnesses are in H. Since by assumption $\exists x \in \mathbb{Z}_n^*$ such that $x^{n-1} \not\equiv 1 \bmod n$, we have $x \in \mathbb{Z}_n^* - H \Rightarrow H$ is a proper subgroup of $\mathbb{Z}_n^*$. The theorem follows. This is a pretty good result. Unless

[4] Here, a base is a witness if RABIN'S-TEST returns COMPOSITE with the base.

[5] With more work, it can be shown that the lower bound is actually $\frac{3}{4}(n-1)$. We direct the reader to Rabin's paper for the improved bound.

[6] This proof is adapted from [2].

[7] See [2] page 869

n is a Carmichael number (so $\nexists$ such an x), we have our theorem, but if we want to guarantee the robustness of our primality test, we need to prove our theorem even for Carmichael numbers despite the fact that they are rare. So assume n is a Carmichael number. Then by the argument Carmichael sets forth in [1], n does not have a repeated prime factor and n is not the product of two prime factors. Now since we only apply RABIN'S-TEST to odd numbers (recall that returning COMPOSITE given an even number $\neq 2$ is a constant time operation which we don't even consider), let $n = n_1 \cdot n_2$ where n_1 and n_2 are odd numbers relatively prime to each other. So in our algorithm we let t and u be such that $n - 1 = 2^t u$ with u odd and we looked at the sequence $X = (a^u, a^{2u}, \ldots, a^{2^t u} = a^{n-1})$ and checked for nontrivial square roots of 1 modulo n. Now call a pair of integers (v, j) acceptable if $v^{2^j u} \equiv -1 \bmod n$ for $v \in \mathbb{Z}_n^*$ and $j \in \{0, 1, \ldots, t\}$. For example, $(n-1, 0)$ is acceptable since $(n-1)$ raised to an odd power modulo n is congruent to -1. Now let j be the largest possible number such that there exists an acceptable pair (v, j), and fix v. Now define $H = \left\{x \in \mathbb{Z}_n^* : x^{2^j u} \equiv \pm 1 \bmod n\right\}$. It is a simple exercise to show that H is a subgroup of $\mathbb{Z}_n^*$. It's also easy to argue that all nonwitnesses are in H. The sequence X produced by a nonwitness is either all 1's or has a -1 no later than the j^{th} position since we defined j to be maximal. Now from v we can construct a $w \in \mathbb{Z}_n^* - H$ from which the theorem will follow. By definition of acceptable, we have $v^{2^j u} \equiv -1 \bmod n \Rightarrow v^{2^j u} \equiv -1 \bmod n_1$. Since n_1 and n_2 are coprime, then it follows from the Chinese Remainder Theorem that $\exists w$ satisfying:

$$w \equiv v \bmod n_1$$

$$w \equiv 1 \bmod n_2$$

Raising both sides to the $(2^j u)^{\text{th}}$ power, we have

$$w^{2^j u} \equiv -1 \bmod n_1$$

$$w^{2^j u} \equiv 1 \bmod n_2$$

We have that $w^{2^j u} \not\equiv 1 \bmod n_1 \Rightarrow w^{2^j u} \not\equiv 1 \bmod n$, and, similarly, $w^{2^j u} \not\equiv -1 \bmod n_2 \Rightarrow w^{2^j u} \not\equiv -1 \bmod n$. Thus $w^{2^j u} \not\equiv \pm 1 \bmod n$ so we have that $w \notin H$. All that remains is justifying that $w \in \mathbb{Z}_n^*$. Since $v \in \mathbb{Z}_n^*$, $\gcd(v, n) = 1 \Rightarrow \gcd(v, n_1) = 1$. Now $w \equiv v \bmod n_1$ so we have $\gcd(w, n_1) = 1$. Similarly, $w \equiv 1 \bmod n_2 \Rightarrow \gcd(w, n_2) = 1$, so we can combine these to get that $\gcd(w, n) = 1$ which means $w \in \mathbb{Z}_n^*$. This concludes the proof. Whether or not n is a Carmichael number, the number of witnesses to the compositeness of n is at least $\frac{1}{2}(n-1)$. ■

4.2 A Note on the Bound

It appears that $\frac{3}{4}$ is the best bound possible. Rabin cites experimental work done by Oren where a computer search of the fraction of witnesses to the compositeness of $2000436751 = 487 \cdot 1531 \cdot 2683$ was found to be 0.7507. Thus Rabin seems to have found to tightet bound possible. It should be noted that in practice the bound is usually much better than $\frac{3}{4}$, although this statement is difficult to quantify.

4.3 Complexity Discussion

RABIN'S-TEST has many appealing properties. First of all, it's easy to implement. Second, the probability of error approaches zero very quickly as we increase our assurance parameter k. RABIN'S-TEST can test a number for primality in seconds on a machine of reasonable power and the probability of error will be smaller than the probability that a particle in the universe selected at random happens to be part of your body. Let's break down the complexity of algorithm. Finding t and u can be done in $O(\log n)$ time by observing the binary representation of n. The first modular exponentiation requires $O(\log n)$ multiplications, and then from 1 to t (so $O(\log n)$ times) we repeatedly square in a single multiplication and do $O(1)$ work to check for a nontrivial square root of 1. It should be clear that, taking our assurance parameter into consideration, $O(k \cdot \log n)$ multiplications are required. Modular multiplication is an $O(\log^2 n)$ operation. Thus each iteration of RABIN'S-TEST can be completed in polylogarithmic time, and k only has to be about 50 to guarantee correct output with overwhelming probability.

5 Rabin's Test on the Web

As a supplement to this paper, I developed a Java applet implementation of Rabin's Test that you can run from my webpage at
`http://www.people.fas.harvard.edu/~troiano/math124`. The assurance parameter is set at $k = 50$. The applet only works for 64-bit numbers, so it's not useful for the prime numbers needed in cryptographic applications, but feel free to download my code and extend the applet to work with BigIntegers.

6 Appendix

There are a couple basic number theory results I've separated from the main text:

Theorem 1 *If there is a nontrivial square root of 1 modulo n, then n is composite.*

Proof: This theorem follows from the contrapositive of the next theorem. That is, if $x^2 \equiv 1 \bmod p$ has any solutions other than ± 1, then p is not an odd prime. Finally, the $p = 2$ case is ruled out immediately by the fact that if $x^2 \equiv 1 \bmod 2$, then $x \equiv 1 \bmod 2$ so all square roots of 1 modulo 2 are trivial (we needn't even consider this case though since we rule out the $p = 2$ case in our algorithm immediately). ■

Theorem 2 *If p is an odd prime, then the equation $x^2 \equiv 1 \bmod p$ has only two solutions, namely $x = 1$ and $x = -1$.*

Proof: One can find Euler's criterion in any elementary number theory textbook (see, for example, [5]). We simply state it here. If p is an odd prime and $\gcd(a, p) = 1$, then $x^2 \equiv a \bmod p$ has two solutions or no solution according as $a^{\frac{(p-1)}{2}} \equiv 1$ or $\equiv -1 \bmod p$. Letting $a = 1$, we see that is p is an odd prime, then $x^2 \equiv 1 \bmod p$ has two solutions, and from inspection we can see these are ± 1. ■

References

[1] R.D. Carmichael. *On Composite Numbers p Which Satisfy the Fermat Congruence* $a^{p-1} \equiv 1 \bmod p$. Amer. Math. Monthly 19 (1912), 22-27.

[2] Cormem, Leiserson, Rivest, and Stein. *Introduction to Algorithms*. The MIT Press. Cambridge, Massachusetts, 2001.

[3] C.F. Gauss. *Disquisitiones Arithmeticae (A.A. Clarke, Transl.)*. Yale Univ. Press. New Haven, Conn. London, 1966.

[4] G.L. Miller. *Riemann's Hypothesis and a Test for Primality*. J. Comput. and System Sci. 13 (1976), 300-317.

[5] Ivan Niven. *An Introduction to the Theory of Numbers*. John Wiley and Sons, Inc. New York, 1991.

[6] M.O. Rabin. *Probabilistic Algorithm for Testing Primality*. Journal of Number Theory 12 (1980), 128-138.