

Improving the Speed and Accuracy of the Miller-Rabin Primality Test

Shyam Narayanan

Mentor: David Corwin

MIT PRIMES-USA

Abstract

Currently, even the fastest deterministic primality tests run slowly, with the Agrawal-Kayal-Saxena (AKS) Primality Test runtime $\tilde{O}(\log^6(n))$, and probabilistic primality tests such as the Fermat and Miller-Rabin Primality Tests are still prone to false results. In this paper, we discuss the accuracy of the Miller-Rabin Primality Test and the number of nonwitnesses for a composite odd integer n . We also extend the Miller-Rabin Theorem by determining when the number of nonwitnesses $N(n)$ equals $\frac{\varphi(n)}{4}$ and by proving that for all n , if $N(n) > \frac{5}{32} \cdot \varphi(n)$ then n must be of one of these 3 forms: $n = (2x + 1)(4x + 1)$, where x is an integer, $n = (2x + 1)(6x + 1)$, where x is an integer, n is a Carmichael number of the form pqr , where p, q, r are distinct primes congruent to 3 (mod 4). We then find witnesses to certain forms of composite numbers with high rates of nonwitnesses and find that Jacobi nonresidues and 2 are both valuable bases for the Miller-Rabin test. Finally, we investigate the frequency of strong pseudoprimes further and analyze common patterns using MATLAB. This work is expected to result in a faster and better primality test for large integers.

1 Introduction

Data is growing at an astoundingly rapid rate, and better information security is required to protect increasing quantities of data. Improved data protection requires more sophisticated cryptographic methods. Improved cryptography necessitates the use of larger semiprimes that are extremely difficult to factor, and thus requires the verification of primality of larger primes. Therefore, faster and more efficient primality tests are key to better information security.

1.1 Primality Tests

A primality test is simply an algorithm to determine whether an input number is prime. Some primality tests are deterministic; that is, they always correctly determine if a number is prime or composite. The fastest known deterministic primality test was created in 2004, when three computer scientists, Agrawal, Kayal, and Saxena, created the AKS primality test that operated in $\tilde{O}(\log(n)^6)$ time, where $\tilde{O}(f(n))$ is defined as $O(f(n) \cdot \log(f(n))^k)$ for some integer k [1]. Although a significant breakthrough, this speed is still rather slow when compared to information security needs.

Probabilistic primality tests are usually faster, but are not always accurate. These tests determine whether n satisfies one or more conditions that all primes must satisfy. So, if an input n does not satisfy these conditions, we know n is composite. If n satisfies the conditions, n is *probably prime*, but n does not have to be prime.

The Fermat Primality Test stems from Fermat's Little Theorem, which states that if n is prime, $a^{n-1} \equiv 1 \pmod{n}$. Given an input n and $a < n$, we check whether $a^{n-1} \equiv 1 \pmod{n}$. If this is not true, then n is composite. Else, n is probably prime. Unfortunately, the Fermat Primality Test has a high rate of error, with too many composites being probably prime. Instead, we use a more accurate method, the Miller-Rabin Primality Test.

1.2 The Miller-Rabin Primality Test

The Miller-Rabin Primality Test is an extension of the Fermat Primality Test. The test works as follows: Suppose we have an odd integer n , such that $n = 1 + d \cdot 2^e$ and d is odd.

We choose a positive integer $a < n$. If either $a^d \equiv 1 \pmod{n}$, or $a^{2^r \cdot d} \equiv -1 \pmod{n}$ for some $r < e$, then n is probably prime. Else, n is composite. If n is probably prime, then either n is prime, or n is composite, in which case we say that a is a *nonwitness* to n and that n is a *strong pseudoprime* to the base a . We say that a is a *witness* to n if $a < n$, and a is not a nonwitness [2, 3].

Example 1.2.1. Suppose $n = 65$. If we consider $a = 8$, we notice that $n = 1 + 1 \cdot 2^6$, $8^1 \equiv 8 \not\equiv 1$, but $8^{2^1 \cdot 1} = 64 \equiv -1$. Thus, either 65 is prime, or 65 is a composite and 8 is a nonwitness. Of course 65 is not prime, but just to check, we consider $a = 2$. Clearly, $a^1 = 2 \not\equiv 1$, $a^{2^0 \cdot 1} = 2 \not\equiv -1$, $a^{2^1 \cdot 1} = 4 \not\equiv -1$, $a^{2^2 \cdot 1} = 16 \not\equiv -1$, $a^{2^3 \cdot 1} \equiv 61 \not\equiv -1$, $a^{2^4 \cdot 1} \equiv 16 \not\equiv -1$, $a^{2^5 \cdot 1} \equiv 61 \not\equiv -1$. Since 65 fails the Miller-Rabin Primality Test in base 2, we know that 65 is composite. We also know 2 is a witness to 65, but 8 is a nonwitness to 65.

The Miller-Rabin Primality Test is significantly more accurate than the Fermat Primality Test. There exist an infinite number of composite integers known as Carmichael numbers, which satisfy the property that $\forall n$, where n is a Carmichael number, if $(a, n) = 1$, then $a^{n-1} \equiv 1 \pmod{n}$ [4]. However, Michael O. Rabin proved that for any composite odd integer n , the number of nonwitnesses of n is at most $\frac{n}{4}$, and can even be reduced to $\frac{\varphi(n)}{4}$ if $n \geq 25$ [3].

To demonstrate the improved effectiveness of the Miller-Rabin Primality Test, we check whether 91 is prime or composite.

Example 1.2.2. We test 91 with the base of 3. If we use the Fermat Primality Test, we get $3^{90} \equiv 1 \pmod{91}$. If we use the Miller-Rabin Primality Test, since $91 = 2 \cdot 45 + 1$, and since $3^{45} \equiv 27 \pmod{91}$, it is clear that 3 is a witness to 91 for the Miller-Rabin Primality Test even though 3 is a false witness for the Fermat Primality Test.

It is well known that the Miller-Rabin Primality Test has a running time of $O(\log^3(n))$. Using Fast Fourier Transforms, the running time can be reduced to $\tilde{O}(\log^2(n))$, the same time as for the Fermat Primality Test. The Miller-Rabin Primality Test is also more accurate, but some numbers have a relatively high proportion of nonwitnesses.

Yet it is known that the smallest witness of any composite integer n must be at most $2 \cdot \ln(n)^2$, assuming the Extended Riemann Hypothesis (ERH) [2]. Thus, if the ERH is true,

the Miller-Rabin Primality Test can be converted into a deterministic primality test with running time $\tilde{O}(\log^4(n))$. Although it is significantly faster than the AKS primality test, it requires the ERH to be true. Since the ERH is known to be an extremely difficult problem in mathematics, the Miller-Rabin Primality Test is not verified as a true deterministic primality test. Yet, even without proving the ERH, we can still reduce the number of nonwitnesses and improve the speed of the Miller-Rabin Primality Test.

1.3 Experimental Results

While the Miller-Rabin Primality is more accurate than the Fermat Primality Test, some numbers still have a high number of nonwitnesses with respect to the Miller-Rabin test. One of the main goals of this paper is to single out which types of integers have a lot of nonwitnesses and then use that information to build a better algorithm.

Some experimental evidence, as we describe below, hints that certain forms composite integers are significantly more likely to be strong pseudoprimes, and thus have higher proportions of nonwitnesses.

Pomerance, Selfridge, and Wagstaff determined that, for all odd composites less than $2.5 \cdot 10^{10}$, there are 13 integers which are strong pseudoprimes to the bases 2, 3, and 5 [5]. Of these 13 integers, 11 are of the form $(k+1)(ak+1)$, where a is an integer and $k+1$, $ak+1$ are prime. The other two are Carmichael numbers with three distinct prime factors. One of the numbers has prime factorization $151 \cdot 751 \cdot 28351$, and all three primes are congruent to 3 (mod 4). The other number has prime factorization $397 \cdot 4357 \cdot 8317$, and all three primes are congruent to 5 (mod 8). They also determined that among the 4842 strong pseudoprimes to the base 2 that are less than $2.5 \cdot 10^{10}$, 4200 have only 2 distinct prime factors, and 407 have 3 distinct prime factors.

Finally, when the GMP function *mpz_probab_prime_p*($N, 1$), a function that performs a Miller-Rabin Primality Test with an unknown base once on an integer N , was tested on composites, certain numbers clearly were more likely to be strong pseudoprimes. Of the composites less than $4 \cdot 10^{12}$, 3773 were strong pseudoprimes. Of them, 3523 had only 2 prime factors, 3187 were of the form $(k+1)(ak+1)$ for primes $k+1$ and $ak+1$, 1095 were of the form $(k+1)(2k+1)$ for primes $k+1$ and $2k+1$, and 856 were of the form $(k+1)(3k+1)$

for primes $k + 1$ and $3k + 1$ [7].

1.4 This Research

Despite the seemingly high rate of nonwitnesses, the Miller-Rabin Primality Test has still proven to be very useful in finding primes less than 2^{500} . For example, the Mersenne numbers M_p were tested for primality for $p \leq 500$ using repeated iterations of the Miller-Rabin Primality Test, yielding correct results in only 10 minutes [3]. Its applications to numbers of such magnitude make it especially useful for finding large semiprimes and thus for cryptography [6].

However, for certain integers, nonwitnesses appear to abound, and hundreds of iterations are necessary to remove such cases and guarantee primality. Even for some integers less than 2^{64} , when the Miller-Rabin Primality Test was iterated 10 times, two composites passed all 10 tests, and 12 tests were required to eliminate all composites [8]. When this test was iterated on larger integers, dozens of composites under 2^{80} were found to pass all 10 tests [8]. Clearly, if a probabilistic primality test is used to find more large primes with certainty, a more accurate primality test would be useful and more efficient.

Rather than finding a completely novel probabilistic or deterministic primality test, we look to improve the accuracy of the Miller-Rabin Primality Test by removing cases of composites with especially large numbers of nonwitnesses. In this paper, we analyze the number of nonwitnesses and determine, with proof, which forms of numbers have the highest proportion of nonwitnesses, as well as methods to remove such numbers by choosing bases that are always witnesses to certain forms of composites. By removing these special forms, we can make a more accurate implementation of the Miller-Rabin test and enhance our ability to find large primes and generate semiprimes for cryptographic use [6].

We note that numbers of the form $(2x + 1)(2ax + 1)$ for a small integer a are likely to pass the Miller-Rabin Primality Test, at least according to empirical results. We first devise, with proof, a general formula for the number of nonwitnesses to a composite odd integer n . Then, we use this formula to check, with proof, that numbers of the form $(2x + 1)(2ax + 1)$ for a small integer a have very high rates of nonwitnesses. We also use this formula to determine whether any other forms of composite integers have exceptionally high rates of nonwitnesses.

We also eliminate several of these cases of integers to allow for a more accurate version of the Miller-Rabin Primality Test. Some of these numbers can be easily eliminated, but others are much harder to eliminate. We therefore determine which bases are most apt to remove these more difficult composites, and empirically test the bases theoretically predicted to be strongest against other bases.

2 Finding Composites with Many Nonwitnesses

2.1 Number of Nonwitnesses to a Composite n

For each integer n , we define $N(n)$ as the number of nonwitnesses of n . Brian Higgins [9] presents a conjecture stating that $N(n) = \frac{\varphi(n)}{4}$ if n is of the form $(2x + 1)(4x + 1)$, where x is odd and $2x + 1$, $4x + 1$ are prime. We present a general formula for $N(n)$, which we will use to prove Higgins's conjecture as well as extend a theorem about the Miller-Rabin Primality Test. After this result was obtained, we became aware that a similar result was stated, though without proof, by Charles R. Greathouse IV [10], so we present an original proof to this formula.

Throughout this section, we use the same notation as explained in Theorem 2.1.1.

Theorem 2.1.1. Consider an odd composite integer n with m distinct prime factors. Suppose that $n - 1 = 2^e \cdot d$ and d is odd. Also suppose that $n = \prod_{i=1}^m p_i^{q_i}$, and each p_i can be expressed as $2^{e_i} \cdot d_i + 1$, where each d_i is odd. We prove that the number of nonwitnesses to n 's compositeness equals

$$\prod_{i=1}^m \gcd(d, d_i) \cdot \left(\frac{2^{\min(e_i) \cdot m} - 1}{2^m - 1} + 1 \right)$$

Proof. First, we determine the number of solutions to $a^d \equiv 1 \pmod{n}$. Clearly, $a^d \equiv 1 \pmod{n}$ if and only if $a^d \equiv 1 \pmod{p_i^{q_i}} \forall i$. First, we determine the number of solutions to $a^d \equiv 1 \pmod{p_i^{q_i}}$ for each i . Since $a^{p_i^{q_i-1} \cdot (p_i-1)} \equiv 1 \pmod{p_i^{q_i}}$, $a^d \equiv 1$ if and only if $a^{\gcd(d, p_i^{q_i-1} \cdot (p_i-1))} \equiv 1 \pmod{p_i^{q_i}}$. But since d is odd and is not divisible by p_i , this is equivalent to $a^{\gcd(d, d_i)} \equiv 1 \pmod{p_i^{q_i}}$. The number of solutions to this equation is known

to be $\gcd(d, d_i)$, since $U_{p_i^{q_i}}$ is cyclic. Thus, the number of solutions to $a^d \equiv 1 \pmod{p_i}$ also equals $\gcd(d, d_i)$. By the Chinese remainder theorem, the number of solutions to $a^d \equiv 1 \pmod{n}$ equals $\prod \gcd(d, d_i)$.

Now, we determine the number of solutions to $a^{2^k \cdot d} \equiv 1 \pmod{p_i^{q_i}}$. Of course, when $k = 0$, there are $\gcd(d, d_i)$ solutions. For arbitrary k , the number of solutions equals

$$\gcd(2^k \cdot d, \varphi(p_i^{q_i})) = \gcd(2^k \cdot d, p_i - 1) = \gcd(2^k \cdot d, d_i \cdot 2^{e_i}).$$

By the same method as in the previous paragraph, which equals $\gcd(2^k, 2^{e_i}) \cdot \gcd(d, d_i)$. Thus, $\gcd(2^k \cdot d, \varphi(p_i^{q_i})) = 2^{\min(k, e_i)} \cdot \gcd(d, d_i)$. The number of solutions to $a^{2^k \cdot d} \equiv -1 \pmod{p_i}$ equals the number of solutions to $(a^{2^k \cdot d})^2 \equiv 1 \pmod{p_i^{q_i}}$ minus the number of solutions to $a^{2^k \cdot d} \equiv 1 \pmod{p_i^{q_i}}$. This value equals $2^{\min(k+1, e_i)} \cdot \gcd(d, d_i) - 2^{\min(k, e_i)} \cdot \gcd(d, d_i)$, or $2^{k \cdot \gcd(d, d_i)}$ if $k < e_i$, 0 otherwise. That means the number of solutions to $a^{2^k \cdot d} \equiv -1 \pmod{n}$, by the Chinese Remainder Theorem, equals $2^{km} \cdot \prod \gcd(d, d_i)$ if $k < \min(e_i)$, and 0 otherwise.

Therefore, the total number of nonwitnesses equals

$$\left(\prod \gcd(d, d_i) \right) \cdot \left(1 + \sum_{k=0}^{\min(e_i)-1} 2^{km} \right) = \left(\sum \gcd(d, d_i) \right) \cdot \left(\frac{2^{\min(e_i) \cdot m} - 1}{2^m - 1} + 1 \right).$$

□

2.2 Miller-Rabin Extensions

It is known that for any composite odd integer n , $N(n) \leq \frac{\varphi(n)}{4}$ [3]. Here, we prove this theorem, but also show that unless n is of a few certain forms, $N(n) \leq \frac{5}{32} \cdot \varphi(n)$. We begin by proving a few lemmas to prove our extensions of the Miller-Rabin Theorem.

Lemma 2.2.1. Consider an odd composite integer n with m distinct prime factors. If $m \geq 4$, then $N(n) \leq \frac{1}{8} \cdot \varphi(n)$.

Proof. If $m \geq 4$, then

$$\begin{aligned}
\frac{\frac{2^{\min(e_i) \cdot m - 1}}{2^m - 1} + 1}{2^{\min(e_i) \cdot m}} &= \frac{1}{2^m - 1} + \left(\frac{1}{2^{\min(e_i) \cdot m}} \right) \cdot \left(1 - \frac{1}{2^m - 1} \right) \\
&\leq \frac{1}{2^m - 1} + \left(\frac{1}{2^m} \right) \left(1 - \frac{1}{2^m - 1} \right) \\
&= \frac{2(2^m - 1)}{(2^m)(2^m - 1)} \\
&= \frac{2}{2^m} \\
&\leq \frac{1}{8}.
\end{aligned}$$

Therefore,

$$\begin{aligned}
\prod \gcd(d, d_i) \cdot \left(\frac{2^{\min(e_i) \cdot m} - 1}{2^m - 1} + 1 \right) &\leq \frac{1}{8} \cdot \prod \gcd(d, d_i) \cdot 2^{\min(e_i) \cdot m} \\
&\leq \frac{1}{8} \cdot \prod (d_i \cdot 2^{e_i}) \\
&= \frac{1}{8} \cdot \prod (p_i - 1) \\
&\leq \frac{1}{8} \cdot \varphi(n).
\end{aligned}$$

□

Lemma 2.2.2. If $m = 3$, then either $N(n) \leq \frac{5}{32} \cdot \varphi(n)$ or n is a Carmichael number with all 3 factors congruent to 3 (mod 4), in which case $N(n) = \frac{1}{4} \cdot \varphi(n)$.

Proof. We first note that

$$\frac{\frac{2^{\min(e_i) \cdot m - 1}}{2^m - 1} + 1}{2^{\min(e_i) \cdot m}} = \frac{1}{2^m - 1} + \left(\frac{1}{2^{\min(e_i) \cdot m}} \right) \left(1 - \frac{1}{2^m - 1} \right).$$

If $\min(e_i) \geq 2$, then

$$\frac{1}{2^m - 1} + \left(\frac{1}{2^{\min(e_i) \cdot m}} \right) \left(1 - \frac{1}{2^m - 1} \right) \leq \frac{1}{7} + \frac{1}{64} \cdot \frac{6}{7} = \frac{5}{32}.$$

Similar to our proof of the first lemma, we have

$$\begin{aligned}
\prod \gcd(d, d_i) \cdot \left(\frac{2^{\min(e_i) \cdot m} - 1}{2^m - 1} + 1 \right) &\leq \frac{5}{32} \cdot \left(\prod \gcd(d, d_i) \right) \cdot 2^{\min(e_i) \cdot m} \\
&\leq \frac{5}{32} \cdot \prod (d_i \cdot 2^{e_i}) \\
&= \frac{5}{32} \cdot \prod (p_i - 1) \\
&\leq \frac{5}{32} \cdot \varphi(n).
\end{aligned}$$

Else, if $\min(e_i) = 1$, then $\frac{2^{\min(e_i) \cdot m} - 1}{2^{\min(e_i) \cdot m}} + 1 = \frac{1}{2^m} = \frac{1}{4}$. We also have that $\frac{1}{4} \cdot \left(\prod \gcd(d, d_i) \right) \cdot 2^{(\min(e_i) \cdot m)} = \frac{1}{4} \prod (\gcd(d, d_i) \cdot 2^{e_i})$ if and only if $e_1 = e_2 = e_3 = 1$. Else, $\frac{1}{4} \cdot \left(\prod \gcd(d, d_i) \right) \cdot 2^{(\min(e_i) \cdot m)} \leq \frac{1}{8} \cdot \prod (\gcd(d, d_i) \cdot 2^{e_i})$.

If $e_1 = e_2 = e_3 = 1$, then $\frac{1}{4} \cdot \prod (\gcd(d, d_i) \cdot 2^{e_i}) = \frac{1}{4} \cdot \prod (d_i \cdot 2^{e_i}) = \frac{1}{4} \cdot \prod (p_i - 1)$ if and only if $d_i | d \forall d$. Else, $\frac{1}{4} \cdot \prod (\gcd(d, d_i) \cdot 2^{e_i}) \leq \frac{\frac{1}{4} \cdot \prod (d_i \cdot 2^{e_i})}{2} = \frac{1}{8} \cdot \prod (p_i - 1) \leq \frac{\varphi(n)}{8}$. Either way, if n is not squarefree, then $N(n) \leq \frac{1}{4} \cdot \prod (p_i - 1) \leq \frac{1}{12} \cdot \varphi(n)$. If n is squarefree, then $d_i | d \forall d$ is equivalent to $(p_i - 1) | (n - 1) \forall i$ and $p_i \equiv 3 \pmod{4} \forall i$. In other words, if $m = 3$, then $N(n) = \frac{\varphi(n)}{4}$ if and only if n is squarefree and $(p_i - 1) | (n - 1) \forall i$, which is equivalent to n is a Carmichael number with three distinct prime factors all congruent to 3 (mod 4). Else, $N(n) \leq \frac{\varphi(n)}{8}$. \square

Lemma 2.2.3. Suppose $m = 2$ and $e_1 \neq e_2$. Then $N(n) = \frac{1}{4} \cdot \varphi(n)$ if and only if $n = (2x + 1)(4x + 1)$, where x is odd and $2x + 1, 4x + 1$ are prime. $\frac{1}{6} \cdot \varphi(n) < N(n) < \frac{1}{4} \cdot \varphi(n)$ if and only if $n = (2x + 1)(4x + 1)$, where x is even and $2x + 1, 4x + 1$ are prime. Also, if n is not of the form $(2x + 1)(4x + 1)$, then $N(n) \leq \frac{\varphi(n)}{8}$.

Proof. Suppose $n = (1 + 2^{e_1} \cdot d_1)(1 \cdot 2^{e_2} \cdot d_2)$. We have $\frac{2^{\min(e_i) \cdot m} - 1}{2^{\min(e_i) \cdot m}} + 1 \leq \frac{2}{2^m} = \frac{1}{2}$, so $N(n) = \left(\prod \gcd(d, d_i) \right) \left(\frac{2^{\min(e_i) \cdot m} - 1}{2^m - 1} + 1 \right) \leq \frac{1}{2} \cdot \left(\prod \gcd(d, d_i) \right) \cdot 2^{2 \cdot \min(e_i)}$.

If $e_1 \neq e_2$,

$$\begin{aligned} \frac{1}{2} \cdot \prod \gcd(d, d_i) \cdot 2^{2 \cdot \min(e_i)} &\leq \frac{1}{2} \cdot \left(\prod d_i \right) \cdot 2^{2 \cdot \min(e_i)} \\ &\leq \frac{1}{4} \cdot \prod (d_i \cdot 2^{e_i}) \\ &\leq \frac{1}{4} \cdot \varphi(n). \end{aligned}$$

There is equality if and only if $\frac{2^{\min(e_i) \cdot m - 1} + 1}{2^{\min(e_i) \cdot m}} = \frac{1}{2}$, $d_i | d \forall i$, n is squarefree, and $|e_1 - e_2| = 1$.

The last three conditions require the following: n can be written as $(1 + d_1 \cdot 2^{e_1})(1 + d_2 \cdot 2^{e_1+1}) = 1 + 2^{e_1}(d_1 + 2d_2)$. Clearly, $d = d_1 + 2d_2$, so if $d_1 | d$, then $d_1 | d_1 + 2d_2$, so $d_1 | d_2$. Similarly, if $d_2 | d$, then $d_2 | d_1 + 2d_2$, so $d_2 | d_1$, so $d_1 = d_2$ according to the last three conditions. If any of these are false, it is evident that $N(n) \leq \frac{1}{8} \cdot \varphi(n)$.

Else, if $m = 2$ and $e_1 \neq e_2$, then $n = (1 + 2^{e_1} \cdot d_1)(1 + 2^{e_1+1} \cdot d_1)$ (we can assume without loss of generality that $e_2 > e_1$). If $e_1 \geq 2$, then $\min(e_1, e_2) \geq 2$, so

$$\begin{aligned} \frac{1}{2^m - 1} &< \frac{\frac{2^{\min(e_i) \cdot m - 1} + 1}{2^m - 1} + 1}{2^{\min(e_i) \cdot m}} \\ &= \frac{1}{2^m - 1} + \left(\frac{1}{2^{\min(e_i) \cdot m}} \right) \left(1 - \frac{1}{2^m - 1} \right) \\ &\leq \frac{1}{2^m - 1} + \left(\frac{1}{2^{2 \cdot m}} \right) \left(1 - \frac{1}{2^m - 1} \right), \end{aligned}$$

and since $m = 2$, $\frac{1}{3} < \frac{\frac{2^{\min(e_i) \cdot m - 1} + 1}{2^m - 1} + 1}{2^{\min(e_i) \cdot m}} \leq \frac{3}{8}$. This means $\frac{1}{6} \cdot 2^{e_1+e_2} < \frac{2^{\min(e_i) \cdot m} - 1}{2^m - 1} + 1 \leq \frac{3}{16} \cdot 2^{e_1+e_2}$.

Thus, we have the following, if n is not of the form $(2x+1)(4x+1)$, where $2x+1, 4x+1$ are primes, then $N(n) \leq \frac{1}{8} \cdot \varphi(n)$. If n is of the form $(2x+1)(4x+1)$, where x is odd, $2x+1, 4x+1$ are primes, then $n = 1 + 6x + 8x^2 = 1 + 2(3x + 4x^2)$, so $N(n) = \gcd(x, 3x + 4x^2) \cdot \gcd(x, 3x + 4x^2) \cdot (2) = 2x^2 = \frac{1}{4} \cdot \varphi(n)$. Finally, if n is of the form $(2x+1)(4x+1)$, where x is even, $2x+1, 4x+1$ are primes, then $\frac{1}{6} \cdot \varphi(n) < N(n) \leq \frac{3}{16} \cdot \varphi(n)$. \square

Lemma 2.2.4. If $m = 2$ and $e_1 = e_2$, then $N(n) = \frac{1}{6} \cdot \varphi(n)$ if and only if $n = (2x+1)(6x+1)$, where x is odd and $2x+1, 6x+1$ are prime. Else, $N(n) \leq \frac{1}{8} \cdot \varphi(n)$.

Proof. We again have $N(n) = \left(\prod \gcd(d, d_i) \right) \left(\frac{2^{\min(e_i) \cdot m} - 1}{2^m - 1} + 1 \right) \leq \frac{1}{2} \cdot \left(\prod \gcd(d, d_i) \right) \cdot 2^{2 \cdot \min(e_i)}$.

First, suppose that n is squarefree. Then we have $n = (1 + 2^{e_1} \cdot d_1)(1 + 2^{e_1} \cdot d_2) = 1 + 2^{e_1}(d_1 + d_2 + 2^{e_1} \cdot d_1 \cdot d_2)$. Now, we have $d = d_1 + d_2 + 2^{e_1} \cdot d_1 \cdot d_2$, so if $d_1|d$, then $d_1|d_2$. Since $d_1 \neq d_2$, either $\gcd(d_1, d) \neq d_1$ or $\gcd(d_2, d) \neq d_1$, and since d_1, d_2, d are all odd, $\gcd(d_1, d) \cdot \gcd(d_2, d) \leq \frac{d_1 \cdot d_2}{3}$. If $\gcd(d_1, d) \cdot \gcd(d_2, d) = \frac{d_1 \cdot d_2}{3}$, then we have two options: $\gcd(d, d_1) = \frac{d_1}{3}, \gcd(d, d_2) = d_2$, or vice versa. Without loss of generality, we assume the first. Note that $\gcd(d_1, d_1 + d_2 + 2^{e_1} \cdot d_1 \cdot d_2) = \gcd(d_1, d_2) = \gcd(d_2, d_1 + d_2 + 2^{e_1} \cdot d_1 \cdot d_2)$, so $\frac{d_1}{3} = \gcd(d, d_1) = \gcd(d, d_2) = d_2$. Thus, $d_1 = 3d_2$. Also, $\frac{2^{\min(e_i) \cdot m} - 1}{2^{\min(e_i) \cdot m}} + 1 = \frac{1}{2}$ only if $\min(e_i) = 1$, so if n 's prime factorization can be written as $(2d_1 + 1)(6d_1 + 1)$, then $N(n) = \frac{d_1 \cdot d_2}{3} \cdot \frac{1}{2} = \frac{d_1^2}{2} = \frac{1}{6} \cdot \varphi(n)$. Else, either $\gcd(d, d_1) \cdot \gcd(d, d_2) \leq \frac{d_1 d_2}{5}$ or $\frac{2^{\min(e_i) \cdot m} - 1}{2^{\min(e_i) \cdot m}} + 1 \leq \frac{3}{8}$. Either way, if n does not have a prime factorization of the form $(2x + 1)(6x + 1)$, clearly, $N(n) \leq \frac{1}{8} \cdot \varphi(n)$.

Now, assume n is not squarefree. Then if $n = p_1^{q_1} \cdot p_2^{q_2}$, then since $N(n) \leq (p_1 - 1)(p_2 - 1)$, $N(n) \leq \frac{1}{2} \cdot \frac{\varphi(n)}{p_1^{q_1-1} \cdot p_2^{q_2-1}} = \frac{\varphi(n)}{6}$, with the last equality if and only if n is of the form $3^2 \cdot p_2$ where $p_2 \neq 3$. Else, $N(n) \leq \frac{\varphi(n)}{10}$. Assume $p_2 = 1 + 2^{e_2} \cdot d_2$ and $9p = 1 + 2^e \cdot d$. Thus,

$$\begin{aligned} \prod_{i=1}^m \gcd(d, d_i) \cdot \left(\frac{2^{\min(e_i) \cdot m} - 1}{2^m - 1} + 1 \right) &= \gcd(1, d) \cdot \gcd(d_2, d) \cdot \left(\frac{2^{1 \cdot 2} - 1}{2^2 - 1} + 1 \right) \\ &= 2 \cdot \gcd(d_2, d). \end{aligned}$$

As $\gcd(p_2 - 1, 9p_2 - 1) \leq \gcd(9p_2 - 9, 9p_2 - 1) = \gcd(8, 9p_2 - 1) | 8$, and since d, d_2 are odd, $\gcd(d_2, d) = 1$, so $N(n) = 2 \cdot 1 = 2$. But as $p_2 \geq 5$ and $\varphi(n) = 6(p_2 - 1) \geq 24$, $N(n) \leq \frac{\varphi(n)}{12}$.

Thus, if $m = 2$ and $e_1 = e_2$, we either have $n = (2x + 1)(6x + 1)$ and $N(n) = \frac{1}{6} \cdot \varphi(n)$ or $N(n) = \frac{1}{8} \cdot \varphi(n)$. \square

The last set of composites we must work with are those that are prime powers. Luckily, if $n = p^q$, it is clear from Theorem 1 that $N(n) \leq p$. If $n \geq 81$, it is easy to prove that $\frac{N(n)}{\varphi(n)} \leq \frac{p}{p^{q-1} \cdot (p-1)} < \frac{5}{32}$.

The following theorem is proved as a direct result of our lemmas in the previous section:

Theorem 2.2.1. Suppose n is an odd composite integer ≥ 81 . Then $N(n) = \frac{\varphi(n)}{4}$ if and only if n is of the form $(2k+1)(4k+1)$, where k is odd and $2k+1, 4k+1$ are prime or n is a Carmichael Number of the form pqr , where p, q, r are distinct primes $\equiv 3 \pmod{4}$. Also, $\frac{\varphi(n)}{6} < N(n) < \frac{\varphi(n)}{4}$ if and only if $n = (2k+1)(4k+1)$, where k is even and $2k+1, 4k+1$ are primes. Finally, $N(n) = \frac{\varphi(n)}{6}$ if and only if n is of the form $(2k+1)(6k+1)$, where k is odd and $2k+1, 6k+1$ are prime. Else, $N(n) \leq \frac{5}{32} \cdot \varphi(n)$.

2.3 New Primality Test

We now explain how to use the results of the previous section to create a primality testing algorithm.

The new primality test proposed consists of 4 steps:

1. Remove integers n of the form $(2x+1)(4x+1)$.
2. Remove integers n of the form $(2x+1)(6x+1)$.
3. Remove integers n that are Carmichael numbers of the form pqr , where $p, q, r \equiv 3 \pmod{4}$.
4. Perform the Miller-Rabin Primality Test with increased accuracy, since we have gotten rid of the composites with high rates of nonwitnesses.

We note that Theorem 2.2.1 and our primality test is consistent with the experimental results. As composites of the form $(2x+1)(4x+1)$ and $(2x+1)(6x+1)$ made the majority of the pseudoprimes in [7], our new primality test should increase the accuracy of the Miller-Rabin Primality Test. The Carmichael form is not as common, but still makes up a significant portion of the common pseudoprimes.

However, this primality test will still result in high rates of error. For integers greater than $4 \cdot 10^{12}$, integers of our three forms become less and less frequent. Also, other forms such as $(2x+1)(8x+1)$ or $(2x+1)(2ax+1)$ for larger integers a still are prevalent. Finally,

there is no known polynomial time algorithm that determines whether n is a Carmichael number of the form pqr , where $p, q, r \equiv 3 \pmod{4}$. Thus, we obtain methods to remove other forms of integers with many nonwitnesses by performing Miller-Rabin Primality Tests with special bases.

3 Eliminating Integers with Many Nonwitnesses

3.1 Choosing Bases to Remove Composites

From the experimental results, we note that our three forms are not necessarily the only forms that we can eliminate. Being able to eliminate other forms $(2x + 1)(2kx + 1)$ can also remove extra composites and increase the accuracy of our primality test.

To remove the first two cases, we just have to check if $8n + 1$ or $3n + 1$ is a perfect square. However, this strategy will not cover $(2x + 1)(2kx + 1)$ for large values of k . Instead, we attempt to remove such values of k with Miller-Rabin tests for specific bases.

We prove the following:

Theorem 3.1.1. If $n = \prod p_i$, and $v_2(p_i - 1) = v_2(p_j - 1) \forall i, j$, then if x is a nonwitness to n , $\left(\frac{x}{p_i}\right) = \left(\frac{x}{p_j}\right) \forall i, j$.

Proof. If x is a nonwitness, $x^{\prod p_i - 1} \equiv 1 \pmod{\prod p_i}$. If we suppose p_{i_0} is the remainder when $\prod p_i - 1$ is divided by $p_i - 1$, then $x^{p_{i_0}} \equiv 1 \pmod{p_i}$.

Suppose $v_2(\prod p_i - 1) - v_2(p_i - 1) = k$ for some nonnegative integer k . Since $v_2(p_i - 1) \geq 1$, we can define p_{i_j} for any j from 0 to $k + 1$ as p_{i_j} being $\frac{\prod p_i - 1}{2^j}$ taken mod $p_i - 1$.

We prove that $x^{p_{i_j}} \equiv 1 \pmod{p_i} \forall j, 1 \leq j \leq k$. To show this, suppose $x^{p_{i_{j-1}}} \equiv 1 \pmod{p_i}$. Then $p_{i_{j-1}} \equiv 2 \cdot p_{i_j} \pmod{p_i - 1}$, so $x^{p_{i_{j-1}}} \equiv (x^{p_{i_j}})^2$, or $x^{p_{i_j}} \equiv \pm 1 \pmod{p_i}$. But since $v_2(p_{i_j}) \geq v_2(p_i - 1)$, $\exists p'$ such that $p' \cdot p_{i_{j-1}} \equiv p_{i_j} \pmod{p_i - 1}$. Thus, $x^{p_{i_j}} \equiv (x^{p_{i_{j-1}}})^{p'} \equiv 1 \pmod{p_i}$. This also means $x^{\frac{\prod p_i - 1}{2^j}} \equiv 1 \pmod{n} \forall j, 1 \leq j \leq k$.

Now, since x is a nonwitness, we know that $x^{\frac{\prod p_i - 1}{2^{k+1}}} \equiv \pm 1 \pmod{n}$. If $x^{\frac{\prod p_i - 1}{2^{k+1}}} \equiv 1 \pmod{n}$, then for each prime p_i , $x^{\frac{\prod p_i - 1}{2^{k+1}}} \equiv 1 \pmod{p_i}$. Since $v_2\left(\frac{\prod p_i - 1}{2^{k+1}}\right) = v_2(p_i - 1) - 1$ and since U_{p_i} is cyclic, we know that $x^{\frac{\prod p_i - 1}{2^{k+1}}} \equiv 1 \pmod{p_i}$ is the same as saying $\left(\frac{x}{p_i}\right) = 1$.

Similarly, $x^{\frac{\prod p_i - 1}{2^{k+1}}} \equiv -1 \pmod{p_i}$ is the same as saying $\left(\frac{x}{p_i}\right) = -1$. Thus, $\left(\frac{x}{p_i}\right) = 1 \forall i$ or $\left(\frac{x}{p_i}\right) = -1 \forall i$. \square

This theorem can lead us to a few lemmas:

Lemma 3.1.1. If n is a Carmichael number of the form pqr , where $p, q, r \equiv 3 \pmod{4}$, and $1 \leq x < n$, then x is a nonwitness if and only if $\left(\frac{x}{p}\right) = \left(\frac{x}{q}\right) = \left(\frac{x}{r}\right)$.

Proof. Note that $v_2(p-1) = v_2(q-1) = v_2(r-1) = 1$. By our theorem, the lemma is clearly true. \square

Lemma 3.1.2. If n is of the form $(2k+1)(2ak+1)$, where a is odd and $2k+1, 2ak+1$ are prime, then if x is a nonwitness to n , $\left(\frac{x}{n}\right) = 1$.

This lemma is clearly true from our theorem. Unfortunately, there does not exist a method to find an integer x such that $\left(\frac{x}{n}\right) = -1$. However, this can be easily done if $a \equiv 3 \pmod{4}$ and k is odd (x can equal 2). It is also known that assuming the Extended Riemann Hypothesis, $\forall n, \exists x \leq 2 \cdot \ln(n)^2$ such that $\left(\frac{x}{n}\right) = -1$ [2, 9].

We can also eliminate composites of the form $n = (2k+1)(4ak+1)$, where a, k are odd and $2k+1, 4ak+1$ are prime. To do so, we first prove the following theorem.

Theorem 3.1.2. If $n = (2k+1)(4ak+1)$, where $2k+1, 4ak+1$ are prime, if $\left(\frac{x}{4ak+1}\right) = -1$, x is a witness to n .

Proof. Note $x^{n-1} = x^{8ak^2+(4a+2)k} \equiv x^{2k} \pmod{4ak+1}$. If x were a nonwitness, then $1 \equiv x^{2k} \equiv x^{2ak} \equiv \left(\frac{x}{4ak+1}\right) \pmod{4ak+1}$. The contrapositive of this statement is equivalent to the theorem. \square

This theorem is an extension of a theorem published by F. Arnault, who proves a similar theorem except assuming that $a = 1$ [12].

Now, we prove the following:

Lemma 3.1.3. If n is of the form $(2k+1)(4ak+1)$, where a, k are odd and $2k+1, 4ak+1$ are prime, 2 is a witness to n .

Proof. Note that $\left(\frac{2}{4ak+1}\right) = -1$ since $4ak+1 \equiv 5 \pmod{8}$, so 2 is a quadratic nonresidue.

From our previous theorem, since $\left(\frac{2}{4ak+1}\right) = -1$, $2^{n-1} \not\equiv 1 \pmod{n}$, so 2 is a witness. □

3.2 Refined Primality Test

From our results here, we can refine our original test in Section 2.3 as follows: [5]

1. Remove integers n of the form $(2x+1)(4x+1)$.
2. Remove integers n that are Carmichael numbers of the form pqr , where $p, q, r \equiv 3 \pmod{4}$.
3. Perform the Miller-Rabin Primality Test for base 2.
4. For each remaining n , determine the smallest integer x such that $x > 2$, $\left(\frac{x}{n}\right) = -1$ and perform the Miller-Rabin Primality Test for that base.
5. If required, perform additional Miller-Rabin Primality Tests.

Determining whether n is of the form $(2x+1)(4x+1)$ can be done in $O(\log(n))$ time, as square root computation can be done in $O(\log(n))$ time. Also, it is known that calculating $\left(\frac{a}{b}\right)$ can be done in $O(\log(a) \cdot \log(b))$ time, so assuming the Extended Riemann Hypothesis, finding such an x can be done in $(2 \cdot \ln(n)^2) \cdot O(\log(n) \cdot \log(2 \cdot \ln(n)^2)) = O(\ln(n)^3 \cdot \ln(\ln(n)))$ time [13].

Apart from Step 2, which does not have a known minimum running time algorithm, all of the steps which are not direct Miller-Rabin implementations take at most $\tilde{O}(\log(n)^3)$ time, which means this primality test's running time, at least if Step 2 is excluded, is not significantly slower than the Miller-Rabin Primality Test. However, as it removes the forms of the most likely pseudoprimes, it is nevertheless significantly more accurate.

4 Implementations

To determine the approximate frequency of strong pseudoprimes to 2 or to its Jacobi nonresidues, we used MATLAB to determine the number of strong pseudoprimes to 2, 3, and 5 less than 2^{23} , as well the number of strong pseudoprimes to its smallest Jacobi nonresidue less than 2^{23} .

Base	2	3	5	Smallest Jacobi Nonresidue
2^{16}	11	17	15	8
2^{17}	18	26	21	13
2^{18}	24	35	29	22
2^{19}	34	56	45	33
2^{20}	49	75	65	44
2^{21}	75	110	93	69
2^{22}	104	150	133	89
2^{23}	147	189	185	115

Table 4.1: Number of Strong Pseudoprimes less than 2^n , $16 \leq n \leq 23$

From the table shown above, it appears as if the smallest proportion of strong pseudoprimes are strong pseudoprimes to its smallest Jacobi nonresidue. This is not very surprising, since using the Jacobi nonresidue removes any composites of the form $(x+1)(kx+1)$, where k is odd and $(x+1)(kx+1)$ is prime. As shown in countless empirical studies, these are extremely common forms of strong pseudoprimes.

Strong pseudoprimes to base 2 seem rarer than strong pseudoprimes to bases 3 or 5, which supports our theoretical justification for 2 being a strong base.

Yet for every value from 2^{16} to 2^{23} , strong pseudoprimes to base 5 are less frequent than strong pseudoprimes to base 3. It is unclear whether this is mere coincidence, or whether 5 is in fact a stronger base than 3. But this leads to a conjecture we propose:

Conjecture 4.1. For squarefree integers x greater than 1, define $P_x(n)$ as the number of strong pseudoprimes to base x less than n . Then the distributions of strong pseudoprimes to x follow $P_x(n) = (1 + o(1)) \cdot c_x \cdot f(n)$, where c_x is a distinct constant for each x , and f_n is

a function of n irrelevant of the base. In other words, $\lim_{x \rightarrow +\infty} \frac{P_x(n)}{P_y(n)}$ is a constant $\forall x, y$, not necessarily equal to 1.

5 Conclusion and Future Research

In conclusion, we have proven that $N(n) = \frac{\varphi(n)}{4}$ if and only if $n = (2x + 1)(4x + 1)$, where x is odd and $2x + 1, 4x + 1$ are prime or $n = pqr$, where n is a Carmichael number and p, q, r are primes congruent to 3 (mod 4). Also, we have proven that if n cannot be written as $(2x + 1)(4x + 1)$, where x is an integer, as $(2x + 1)(6x + 1)$, where x is an integer, or as $n = pqr$, where n is a Carmichael number and p, q, r are primes congruent to 3 (mod 4), then $N(n) \leq \frac{5}{32} \cdot \varphi(n)$. We also found a way to choose bases that would remove certain forms of composites with high rates of witnesses. Finally, we analyzed the frequency of pseudoprimes in MATLAB and noticed our predictions regarding 2 and the smallest Jacobi nonwitness as strong bases were clearly supported.

For further research, we hope to implement our new primality test and test for speed as well as accuracy.

Also, further research could be performed to help answer our previous conjecture along with the following conjectures we propose:

Conjecture 5.1. If a composite n satisfies $N(n) \leq \frac{5}{32} \cdot \varphi(n)$ (i.e. n is not one of the three special forms), \exists a witness to n less than $(2 \cdot \ln(n)^2)$, and this can be proved without the need of the Extended Riemann Hypothesis.

Conjecture 5.2. There exists a method that determines whether an integer n is a Carmichael number of the form pqr , where p, q , and $r \equiv 3 \pmod{4}$, that operates in $\tilde{O}(\log(n)^4)$ running time.

Stronger version: For any real $\epsilon > 0$, there exists a method which determines whether an integer n satisfies $N(n) > \epsilon * \varphi(n)$, which operates in $\tilde{O}(\log(n)^4)$ running time.

Remark: We choose $\tilde{O}(\log(n)^4)$ running time since this is the running time of the deterministic variant of the Miller-Rabin primality test, assuming the Extended Riemann Hypothesis.

As the Miller-Rabin test is usually iterated hundreds of times for large primes (and occasional pseudoprimes), increasing the maximum probability of a nonwitness for a composite means fewer iterations will be needed. This clearly has vast applications in not only primality testing and number theory, but also in developing methods to create large semiprimes quickly. This will improve cryptography, and will also have practical implications in many other aspects of computer science and number theory.

6 Acknowledgments

I would like to thank the MIT PRIMES-USA program for sponsoring this research. I am deeply indebted to my mentor, David Corwin (MIT), for all the help he has provided me, as well as to the lead mentor, Dr. Tanya Khovanova (MIT). I would like to acknowledge Dr. Stefan Wehmeier and Dr. Ben Hinkle from MathWorks, who proposed this project idea. I would also like to acknowledge the PRIMES-USA faculty, especially Dr. Slava Gerovitch and Dr. Pavel Etingof for making this research possible. Finally, I would like to thank my parents for all the love and support they have provided me throughout my life.

7 References

- [1] Agrawal, M., Kayal, N., and Saxena, N. (2004). PRIMES is in P. *Annals of Mathematics*, 160(2), 781-793.
- [2] Miller, G. (1976), Riemann's Hypothesis and Tests for Primality. *Journal of Computer and System Sciences* 13(3): 300–317.
- [3] Rabin, M. (1980). Probabilistic algorithm for testing primality. *Journal of Number Theory*, 12(1), 128-138.
- [4] Alford, W. R., Granville, A., and Pomerance, C. (1994). There are Infinitely Many Carmichael Numbers. *Annals of Mathematics*, 139, 703–722.
- [5] Pomerance, C., Selfridge, J., and Wagstaff, S. (1980). The pseudoprimes to $25 \cdot 10^9$. *Mathematics of Computation*, 35(151), 1003–1026.
- [6] Karaarslan, E. Primality testing techniques and the importance of prime numbers in security protocols. Retrieved September 4, 2014, from <http://www.karaarslan.net/bildiri/PrimeTestingAndSecurity.pdf>.
- [7] GNU GMP 5.0.1 mpz_probab_prime_p Pseudoprimes. (2011, March 5). Retrieved January 5, 2014, from <http://www.hoegge.dk/gmp/gmp501.htm>.
- [8] Nicely, T. (2004). GNU GMP mpz_probab_prime_p Pseudoprimes. Retrieved January 5, 2014, from <http://www.trnicely.net/misc/mpzpsp.html>.
- [9] Higgins, B., and Burchard, C. (1997, February). The Rabin- Miller Probabilistic Primality Test: Some Results on the Number of Non-Witnesses to Compositeness. Retrieved January 24, 2014, from http://www.plouffe.fr/simon/OEIS/archive_in_pdf/higgins.pdf.
- [10] Takusagawa, K. (n.d.). Odd composites with increasing proportion of nontrivial non-witnesses of compositeness by the Miller-Rabin primality test (C. Greathouse, Ed.). Retrieved April 25, 2014, from <http://oeis.org/A090659>.
- [11] E. Bach (1985). *Analytic Methods in the Analysis and Design of Number-Theoretic Algorithms*. MIT Press, Cambridge, Mass.
- [12] Arnault, F (1995). Rabin-Miller Primality Test: Composite Numbers Which Pass It. *Mathematics of Computation* 64 (209), 355-361.

- [13] Cohen, H. A Course in Computational Algebraic Number Theory (1993). Springer, Berlin.
- [14] Crandall, R. and Pomerance, C. (2001). Prime Numbers: A Computational Perspective. Springer, NY.