THE PRIME NUMBER MAZE

William Paulsen
Department of Computer Science and Mathematics
PO Box 70, Arkansas State University, State University, AR 72467
(Submitted May 2000-Final Revision October 2000)

1. INTRODUCTION TO THE MAZE

This paper introduces a fascinating maze based solely on the distribution of the prime numbers. Although it was originally designed as a simple puzzle, the maze revealed some rather startling properties of the primes. The rules are so simple and natural that traversing the maze seems more like exploring a natural cave formation than a maze of human design.

We will describe this maze using the language of graph theory. In particular, we first define an undirected graph G_0 with the set of all prime numbers as the vertex set. There will be an edge connecting two prime numbers iff their binary representations have a Hamming distance of 1. That is, two primes are connected iff their binary representations differ by exactly one digit.

The natural starting point is the smallest prime, $2 = 10_2$. Following the graph G_0 amounts to changing one binary digit at a time to form new prime numbers. The following sequence demonstrates how we can get to larger and larger prime numbers by following the edges of G_0.

$$\begin{aligned} 10_2 &= 2 \\ 11_2 &= 3 \\ 111_2 &= 7 \\ 101_2 &= 5 \\ 1101_2 &= 13 \\ 11101_2 &= 29 \\ 111101_2 &= 61 \\ 110101_2 &= 53 \\ 100101_2 &= 37 \\ 1100101_2 &= 101 \end{aligned}$$

Actually, we can get to large primes much faster, since the Hamming distance between 3 and $4099 = 1000000000011_2$ is just 1. However, the above example illustrates that we can get to 101 even if we add the restriction that the numbers increase at most one binary digit at a time. Even with this restriction, it is possible to reach 4099, but it requires a total of 46 steps.

We can include this restriction by considering a directed graph, G_1, whose vertices are again the prime numbers. There is an edge from p to q iff the Hamming distance is 1, and $3p \geq q$. Note that this always permits changing a 1 bit to a 0 bit, since $q < p$ implies $3p \geq q$. However, if a 0 bit is changed to a 1 bit, then the condition $3p \geq q$ insures that $q - p$ (which will be a power of 2) will be no more than twice the original number p.

The directed graph G_1 is easier to analyze than the graph G_0, since at any given vertex only a finite number of edges is possible. We define the *valence* of a prime number p to be the number of edges leaving the vertex p on G_1. It is not hard to have a computer map the first 70 steps (from 2) to determine which primes are attainable. A very small portion of the map is shown in Figure 1.

By a lucky coincidence, the distribution of the prime numbers is exactly what is needed to keep this graph interesting. As N increases, its number of bits grows as $\log_2 N$, so to compute

the valence of N, we will need to test $\lfloor \log_2 N \rfloor + 1$ numbers for primality. However, by the prime number theorem [4], only about $1/\ln N$ of these numbers will be prime. So, heuristically, the expected value of the valence will remain roughly constant throughout the entire graph.

Figure 1 shows all of the primes that can be reached from the prime 2 without having to go to primes larger than 1024. However, this does not show all of the primes less than 1024 that can be reached from 2. The number 353 can be reached, but not without first attaining the prime $353 + 2^{27} + 2^{392} + 2^{441}$.

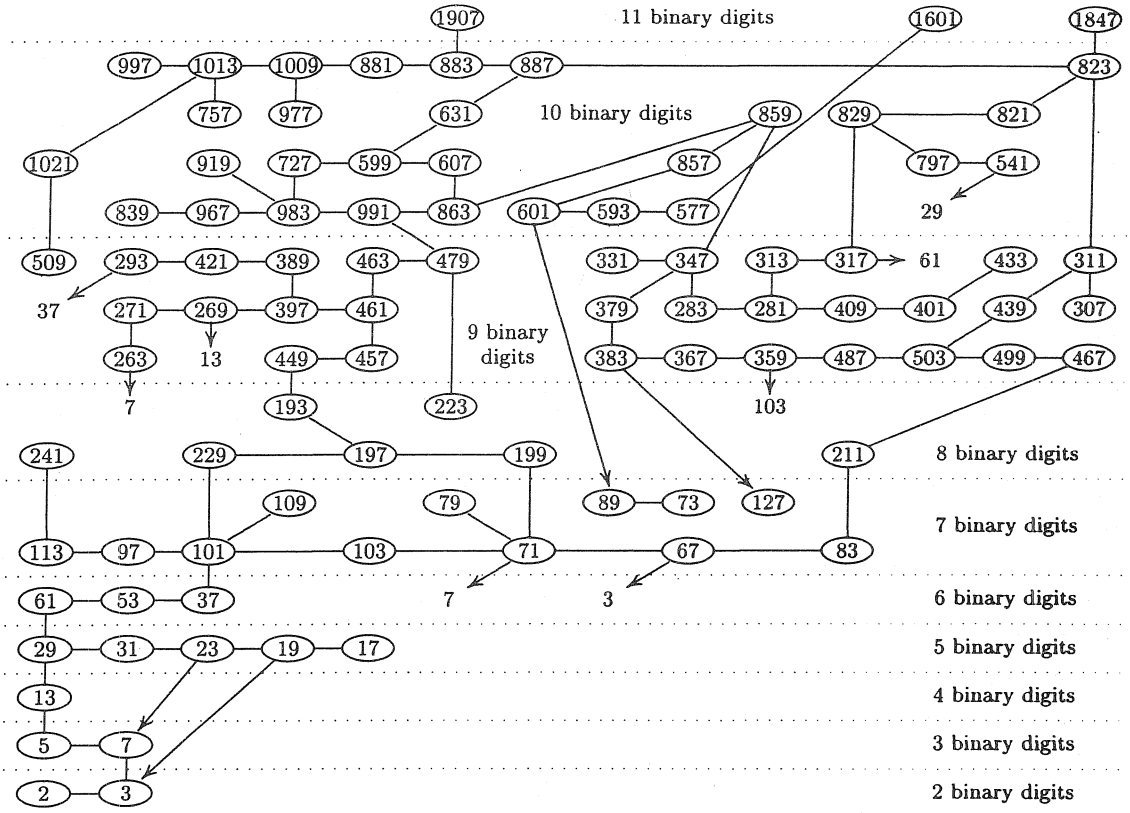

FIGURE 1. The First 9 Levels of the Directed Graph G_2

The example 353 shows how the directed graph G_1 can make back-tracking very difficult. Although only a finite number of primes can be reached from a given prime, there may in fact be an infinite number of primes from which one could get to a given prime. Some of the numbers involved will be very large, so one must be content with knowing that they are "probably prime" via the Miller-Rabin strong pseudoprime test. Since the probability of a composite number passing this test is about 4^{-100} [7], we can be fairly confident that the pseudoprimes needed to get to 353 are indeed prime.

2. THE PARTITIONING OF THE PRIMES

The prime number 11 is ominously missing in Figure 1. This begs the question as to whether one can reach the prime 11 via a much larger prime, as in the case of 353. Obviously, 11 is in the

same connected component as 2 in G_0, since there is an edge between 11 and 3. But can we get from 2 to 11 in G_1?

For each prime p, let us define G_p to be the subgraph of G_1 consisting of all vertices and edges that can be reached starting from the prime p. Note that there are many instances when G_p is a finite graph. For example, G_{73} consists of just two vertices, 73 and 89, and the bidirectional edge connecting them. The question is whether 11 is a vertex of G_2. A simple parity argument shows that it is not.

Definition: Let $p > 3$ be a prime number. We say that p is of *correct parity* if either $p \equiv 2 \mod 3$ and p has an even number of 1 bits in its binary representation, or $p \equiv 1 \mod 3$ and p has an odd number of 1 bits. We say that $p > 3$ is of *incorrect parity* if p is not of correct parity. We do not define parity for the primes 2 and 3. Note that 5 and 7 are of correct parity, but 11 is of incorrect parity.

Proposition 1: If an edge in G_0 connects two primes $p > 3$ and $q > 3$, then p and q have the same parity. In particular, all of the vertices of G_2, besides 2 and 3, are of the correct parity.

Proof: If an edge connects p and q, their binary representation differs by exactly one digit. Thus, one of the primes will have an even number of 1 bits, while the other will have an odd number.

Also, since p and q differ by a power of 2, they cannot be congruent mod 3. Neither can be congruent to 0 mod 3, for both p and q are primes > 3. Thus, one of the primes is congruent to 1 mod 3, while the other is congruent to 2 mod 3. By the way that we defined the parity, if either p or q is of correct parity, then the other must also be of correct parity.

Finally, we notice that in the graph of G_2, 2 only can go to 3, which can only go to 7. Thus, any other vertex in G_2 must be reached from 7 without going through 2 or 3. Since 7 has the correct parity, any prime > 3 in G_2 must also be of the correct parity. □

With this proposition and the fact that 11 has incorrect parity, one sees that 11 is not a vertex of G_2. In fact, if we delete the vertex 3 from the graph of G_0, together with all edges connecting to 3, then the resulting graph consists of 2 and at least two large disconnected subgraphs. It is highly probable that these subgraphs are both infinite. The connected components of G_0—{3} form a partition of the prime numbers. By convention, we will include 2 and 3 in the partition that contains the vertex 7.

The parity argument shows that there must be at least two partitions. We will call the partition containing the first 4 primes the *α-partition*, which would of course contain all vertices of G_2. A second partition, the *β-partition*, contains the prime 11. All primes in the β-partition would have incorrect parity.

3. ISOLATED PRIMES

In asking how many partitions there are, one must ask whether there is any prime p totally isolated from any other primes in G_0. In order for this to happen, $p + 2^n$ must always be composite whenever $2^n > p$. This is closely related to two other problems: the Polignac-Erdös problem and the Sierpiński problem.

In 1849, Polignac conjectured that every odd integer > 1 could be expressed in the form $2^n + p$ (see [10]). In 1950, Paul Erdös [3] disproved this conjecture, and in fact proved that there

is an arithmetic progression of odd numbers, no term of which is of the form $2^n + p$. In fact, no term in this sequence is of the form $2^n \pm p$, where p is a prime. If we considered negative terms in this arithmetic progression, and found a term $-k$ such that k is prime, then k would be a candidate for an isolated prime.

In 1960, Sierpiński [9] asked: for what numbers k is $2^m \cdot k + 1$ composite for all $m \geq 1$. Such numbers are called *Sierpiński numbers*. The smallest Sierpiński number is believed to be 78557, but there are several smaller candidates for which no prime of the form $2^m \cdot k + 1$ is known [1].

Sierpiński showed that, if k belongs to one of several arithmetic progressions, then any term of the sequence $k+1$, $2k+1$, $4k+1$, ..., $2^m \cdot k + 1$ is divisible by one of a set of 6 or 7 fixed primes. The set of primes is called the *covering set* for the Sierpiński number. The number 78557 has the covering set {3, 5, 7, 13, 19, 37, 73}, while the next known Sierpiński number, 271129, uses {3, 5, 7, 13, 17, 241} as its covering set [5].

The relationship between the Sierpiński numbers and the Polignac-Erdös numbers is given in [10]. Since the Polignac-Erdös numbers are in turn related to the isolated primes, there is a direct connection between the Sierpiński numbers and the isolated primes. The following proposition is taken from [10].

Proposition 2: Let k be a Sierpiński number with a covering set S. Then, for all n, $k + 2^n$ will be divisible by some prime in S.

Proof: Let N be the product of the odd primes in the set S. If we let $L = \phi(N)$, then N will divide the Mersenne number $2^L - 1$ by Euler's theorem. We then have that, for all m,

$$\gcd(2^m \cdot k + 1, N) > 1.$$

Multiplying the first part by 2^{L-m} gives

$$\gcd(2^L \cdot k + 2^{L-m}, N) > 1.$$

Since $2^L \equiv 1 \bmod N$, we can replace $2^L \cdot k$ with k and write n for $L - m$ to give us

$$\gcd(k + 2^n, N) > 1.$$

Hence, for all n, $k + 2^n$ is divisible by some prime in S. Note that this process is reversible, so any covering set which shows that $k + 2^n$ is always composite will show that k is a Sierpiński number. □

This proposition makes it clear how to search for isolated prime numbers. We need to find a Sierpiński number that is prime, and for which changing any 1 to a zero in its binary representation results in a composite number. A quick search through the known Sierpiński numbers [11] reveals tat 2131099 satisfies both the extra conditions, and so 2131099 is an isolated prime.

However, 2131099 may not be the smallest isolated prime. The prime 19249 is still a candidate for being Sierpiński. If a covering set is discovered for this number, it will be the smallest isolated prime.

A natural question that arises is whether there is an infinite number of isolated primes. To answer this question, we introduce two more sets of numbers related to the Sierpiński numbers, the Riesel numbers, and the Brier numbers.

Definition: A *Riesel number* is a number k for which $2^n \cdot k - 1$ is composite for all $n > 0$. A *Brier number* is a number that is both Sierpiński and Riesel.

We can use an argument similar to that in Proposition 2 to show that, if k is a Riesel number with a covering set S, then $k - 2^n$ will always have a divisor in the set S.

In 1998, Eric Brier [2] discovered the 41-digit number,

$$293646956601235432781150254051144529108 89,$$

and suggested that it might be the smallest such number. However, this record for the smallest known Brier number has been beaten numerous times by Keller and Nash [6] and by Gallot in [8]. The current record is the 27-digit Brier number,

$$B = 878503122374924101526292469,$$

using the covering set

$$S = \{3, 5, 7, 11, 13, 17, 19, 31, 37, 41, 61, 73, 97, 109, 151, 241, 257, 331, 61681\}.$$

Just one Brier number is sufficient to prove the following proposition.

Proposition 3: There is an infinite number of isolated primes.

Proof: Let B be the above Brier number, and let $N = 2^{17}$ times the product of the primes in S. Since B and N are coprime, by Dirichlet's theorem [7] there is an infinite number of primes of the form $aN + B$ with a a positive integer. All that needs to be shown is that these primes are all isolated. In fact, we can prove that $aN + B \pm 2^n$ is composite for all $a \geq 0$ and $n \geq 0$. Note that

if $n \equiv 0 \pmod 2$,	$3 \mid aN + B - 2^n$;	if $n \equiv 1 \pmod 2$,	$3 \mid aN + B + 2^n$;
if $n \equiv 2 \pmod 3$,	$7 \mid aN + B - 2^n$;	if $n \equiv 0 \pmod 4$,	$5 \mid aN + B + 2^n$;
if $n \equiv 7 \pmod{12}$,	$13 \mid aN + B - 2^n$;	if $n \equiv 6 \pmod 8$,	$17 \mid aN + B + 2^n$;
if $n \equiv 13 \pmod{24}$,	$241 \mid aN + B - 2^n$;	if $n \equiv 1 \pmod 5$,	$31 \mid aN + B + 2^n$;
if $n \equiv 1 \pmod{48}$,	$97 \mid aN + B - 2^n$;	if $n \equiv 0 \pmod{10}$,	$11 \mid aN + B + 2^n$;
if $n \equiv 9 \pmod{16}$,	$257 \mid aN + B - 2^n$;	if $n \equiv 18 \pmod{20}$,	$41 \mid aN + B + 2^n$;
if $n \equiv 0 \pmod 9$,	$73 \mid aN + B - 2^n$;	if $n \equiv 34 \pmod{40}$,	$61681 \mid aN + B + 2^n$;
if $n \equiv 15 \pmod{18}$,	$19 \mid aN + B - 2^n$;	if $n \equiv 12 \pmod{15}$,	$151 \mid aN + B + 2^n$;
if $n \equiv 3 \pmod{36}$,	$37 \mid aN + B - 2^n$;	if $n \equiv 22 \pmod{30}$,	$331 \mid aN + B + 2^n$;
if $n \equiv 21 \pmod{36}$,	$109 \mid aN + B - 2^n$;	if $n \equiv 2 \pmod{60}$,	$61 \mid aN + B + 2^n$;

so the only case left to consider is if $aN + B - 2^n$ happens to be one of the primes in the set S. If $n < 17$, we have $aN + B - 2^n > B - 2^{17}$, which is of course greater than all the primes in S. If, on the other hand, $n \geq 17$, then

$$aN + B - 2^n \equiv B \equiv 67573 \pmod{2^{17}},$$

which is again greater than all of the primes in S. Thus, $aN + B \pm 2^n$ is always composite, and so there is an infinite number of isolated primes.

In the search for isolated primes, a few primes were discovered that were *almost* isolated, meaning that there was only one edge in G_1 directed away from the prime p rather than toward it. The prime 36652489 is a Sierpiński number, so we can tell that the only edge in G_0 is one that connects to the prime 3098057. Yet this is a directed edge in G_1, so there are no edges that connect a prime number to the prime 36652489. Hence, for $p \neq 36652489$, the vertex 36652489 is

not in G_p. Ironically, $G_{36652489}$ not only contains 36652489, it also contains the vertex 2; hence, G_2 is a strict subgraph of $G_{36652489}$.

One could also ask whether there are any finite partitions of G_0 other than the isolated primes. We may never be able to answer this question, since such a partition would have to contain a prime p that is "almost Sierpiński," that is, $p+2^n$ would be composite for all n with one exception, that being another member of the partition. The one exception would preclude the possibility for a covering set for p. Without a covering set, proving $p+2^n$ is composite for all other n would be at least as difficult as proving that there are exactly 4 Fermat primes. A computer search will likely produce some "candidates" for finite partitions, but no amount of computation would be able to prove that the partition is really finite.

4. SOME CONJECTURES ABOUT THE MAZE

Conjecture 1: All Fermat primes are vertices in G_2.

This is a very safe conjecture, for it is almost certain that the only Fermat primes are 3, 5, 17, 257, and 65537, which can be verified to be in G_2. Furthermore, any Fermat prime will have the correct parity. The first three primes show up quickly in Figure 1, but getting to 257 requires as many as 627 steps in the maze, since one first must reach the number $2^{91}+2^{26}+769$. The prime 65537 requires first getting to $2^{268}+2^{100}+2^{98}+2^{83}+3$. Finding the shortest path to these primes remains an unsolved problem.

Conjecture 2: All Mersenne primes are vertices in G_2.

The binary representation of the Mersennes makes them the natural goal for this maze of primes, and by a fortunate coincidence all Mersenne primes have the correct parity. Besides the easy ones found in Figure 1, 8191 requires exactly 38 steps, 131071 requires 48 steps, and $2^{19}-1$ requires 62 steps. The shortest path to $2^{31}-1$ is unknown, since one must first reach $2^{74}+2^{31}-1$. Getting to $2^{61}-1$ and $2^{89}-1$ are straightforward; however, getting to $2^{107}-1$ requires first going to $2^{135}+2^{107}-2^{47}-2^{33}-4097$. Reaching $2^{127}-1$ requires first getting to $2^{182}+2^{127}-1$. By backtracking, a computer has verified that $2^{521}-1$ is in G_2, but the smallest neighbor to $2^{607}-1$ is $2^{1160}+2^{607}-1$, which is currently too large for the computer to handle.

Conjecture 3: There are four infinite partitions of $G_0-\{3\}$ that contain primes less than 1000.

Proving this conjecture if the fundamental unsolved problem of this maze. We have already seen using parity that there are at least two main partitions, the α-partition and the β-partition. But as we explore G_0, two more partitions seem to crop up. Although there is no proof that these extra partitions do not connect in some way to the α-partition or the β-partition, there is very strong evidence that no such connection is possible, hence the conjecture. A table of the four partitions that seem to exist is shown below.

The conjectured partitions			
	Lowest prime	Starting point	Comments
α-partition	2	2	Main maze
β-partition	11	547	Can go from $\beta \to \alpha$ via 3
γ-partition	277	4957	Incorrect parity
δ-partition	683	35759	Correct parity

The primes less than 16000 in the δ-partition are {683, 2699, 2729, 2731, 6827, 8363, 8747, 8867, 10427, 1067, 10799, 10859, 10883, 10889, 10891, 10937, 10939, 10979, 10987, 11003, 11171, 11177, 11243, 11939, 12011, 12203, 14891, 15017, 15083, ...}. All other primes <16000 of correct parity are in the α-partition.

Likewise, {277, 337, 349, 373, 853, 1093, 1109, 1117, 1237, 1297, 1301, 1303, 1362, 1367, 1373, 1381, 1399, 1429, 1489, 1493, 1621, 1861, 1873, 1877, 1879, 2389, 3413, 3541, 4177, 4357, 4373, 4421, 4423, 4441, 4447, 4549, 4561, 4567, 4597, 4933, 4951, 4957, 5077, 5189, 5197, 5209, 5233, 5237, 5333, 5381, 5393, 5399, 5407, 5413, 5431, 5437, 5441, 5443, 5449, 5471, 5477, 5479, 5501, 5503, 5521, 5527, 5557, 5569, 5573, 5581, 5591, 5623, 5653, 5701, 5717, 5749, 5953, 5981, 6007, 6037, 6101, 6133, 6229, 6421, 6469, 6481, 6997, 7237, 7253, 7477, 7489, 7507, 7517, 7537, 7541, 7549, 7573, 7621, 7639, 7669, 8017, 8053, 10069, 12373, 12613, 12637, 12757, 13381, 13397, 13399, 13591, 13597, 13633, 13649, 13669, 13681, 13687, 13693, 13781, 13789, 14149, 14173, 14197, 14293, 15733, ...} are in the γ-partition. All other primes <16000 of incorrect parity, with the possible exception of 6379, are in the β-partition. (Analyzing 6379 requires working with numbers larger than 2^{1396}, which takes too long to determine which of these two sectors it is in.)

This table includes the *starting point* for each partition. The starting point is the smallest prime s in the partition for which G_s apparently contains an infinite number of the vertices of the partition. In other words, for all smaller values of p in the partition, G_p produces a finite graph. (For the primes in the β-partition, we would delete the vertex 3 before computing G_p.) It would be tempting to think that G_s would contain all of the vertices of the partition, but the almost isolated" primes in the partition, such as 36652489, would be excluded. Hence, the most we could say is that G_s contains almost all of the vertices of the partition. In fact, all primes less than 16000, with the possible exception of 6379, are in either G_2, G_{547}, G_{4957}, or G_{35759}. Furthermore, for all primes less than 50000, G_p is either finite or contains one of the four graphs. Thus, if there were a fifth infinite partition, the starting point would have to be larger than 50000. So the four partitions in the above table are the first four partitions in every sense.

5. CONCLUSION

It is amazing that the simple rules of the prime maze can raise so many theoretical questions. What started out as a simple puzzle turned into a fountain of problems, some of them solvable, while others may never be solved. It is ironic that the solution to some of the problems, such as finding an infinite number of isolated primes, turns out not involving the binary number system but rather just the powers of two. Therefore, the results of the prime number maze is likely to have significance in other areas of number theory.

REFERENCES

1. Robert Baillie, G. Cormack, & H. C. Williams. "The Problem of Sierpiński Concerning $k \cdot 2^n + 1$." *Math. Comp.* **37.155** (1981):229-31.
2. Eric Brier. "A Smaller 'Brier Number': 41 Digits." Personal correspondence to Primes-L@autm.edu, September 28, 1998.
3. Paul Erdös. "On Integers of the Form $2^k + p$ and Some Related Problems." *Summa Brasil. Math.* **2** (1950):113-23.

4. Richard K. Guy. *Unsolved Problems in Number Theory*. New York: Springer-Verlag, 1994.
5. G. Jaeschke. "On the Smallest k Such That All $k \cdot 2^N + 1$ Are Composite." *Math. Comp.* **40.161** (1983):381-84.
6. Wilfrid Keller & Chris Nash. http://www.primepuzzles.net/problems/WKCN.zip.
7. Victor Klee & Stan Wagon. *Old and New Unsolved Problems in Plane Geometry and Number Theory*. Washington, D.C.: The Mathematical Association of America, 1991.
8. Carlos Rivera. http://www.primepuzzles.net/problems/prob_031.htm.
9. W. Sierpiński. "Sur un probleme concernant les numberes $k \cdot 2^n + 1$." *Elem. Math.* **15** (1960):73-74.
10. R. G. Stanton & H. C. Williams. "Computation of Some Number-Theoretic Coverings." In *Combinatorial Math. VIII, Lecture Notes in Math.* **884**:8-13. Berlin-New York: Springer-Verlag, 1980.
11. R. G. Stanton. "Further Results on Covering Integers of the Form $1+k2^n$ by Primes." In *Combinatorial Math. VIII, Lecture Notes in Math.* **884**:107-14. Berlin-New York: Springer-Verlag, 1980.

AMS Classification Numbers: 11A41, 11N05

❖❖❖