MATHEMATICS OF COMPUTATION
Volume 71, Number 238, Pages 825–832
S 0025-5718(01)01350-3
Article electronically published on May 17, 2001

DISTRIBUTION OF GENERALIZED FERMAT PRIME NUMBERS

HARVEY DUBNER AND YVES GALLOT

Abstract. Numbers of the form $F_{b,n} = b^{2^n} + 1$ are called Generalized Fermat Numbers (GFN). A computational method for testing the probable primality of a GFN is described which is as fast as testing a number of the form $2^m - 1$. The theoretical distributions of GFN primes, for fixed n, are derived and compared to the actual distributions. The predictions are surprisingly accurate and can be used to support Bateman and Horn's quantitative form of "Hypothesis H" of Schinzel and Sierpiński. A list of the current largest known GFN primes is included.

1. Introduction

In the past several years the continuous improvement in the cost-performance of computers coupled with the development of new theory and improved software packages have resulted in impressive and important advances in computational number theory. For example, on January 1, 1985, only three primes with more than 10,000 digits were known, the largest having 38,751 digits. In August 1999, there were over 5000 known primes with more than 10,000 digits, the largest with more than 2,000,000 digits.

Most of these large primes are of the form $k \cdot 2^n \pm 1$, and most were found by more than a hundred people using the "Proth" program for the PC that was developed and distributed by the second author in the past several years. The Proth program is effective working with numbers of the above form principally because the base 2 form allows multiple precision divides to be replaced by shifts. However there are other important numbers that do not have a base 2 form, and consequently prime searching took about three times longer than for base 2 numbers.

Generalized Fermat Numbers (GFN) are of the form $F_{b,n} = b^{2^n} + 1$ and are particularly interesting since they have many characteristics of the heavily studied standard Fermat numbers $F_n = F_{2,n}$. The original purpose of this paper was to report on a method that was incorporated into the Proth program so that the time to test a GFN for probable primality is faster than for a number of the form $k \cdot 2^n \pm 1$ ($k > 1$), a major technological accomplishment.

As we started using the new program, it was reasonable to try to estimate the number of primes that we could expect to find. The predictions were surprisingly

Received by the editor October 13, 1999 and, in revised form, July 10, 2000.
2000 *Mathematics Subject Classification.* Primary 11Y11; Secondary 11A41.
Key words and phrases. Prime numbers, generalized Fermat numbers, primality proving algorithms.

©2001 American Mathematical Society

License or copyright restrictions may apply to redistribution; see http://www.ams.org/journal-terms-of-use

accurate and further study indicated that our data could be used to support a conjecture of Bateman and Horn [1] which in turn supported the famous “Hypothesis H” of Schinzel and Sierpiński [11].

In this article we describe the new computational method, derive the theoretical distribution of GFN primes, compare the actual and theoretical distributions, and list the current largest known GFN primes.

2. GFN software

2.1. **History.** The “Proth” program was created in 1997 to extend the search for large factors of Fermat numbers. The fastest known method is to search systemically for all possible prime factors which have the form $k \cdot 2^n + 1$, for small k [6]. Primality can be determined quickly using the modern form of Proth’s theorem [7, Theorem 102]. Real-signal Fast Fourier Transforms were used for squaring or multiplying plus fast modular operations (utilizing the special form of the number) were also employed. The algorithm of the FFT was totally written in assembler and optimized to take into account the size of the cache-memory of the computer. At that time, the program was used to discover the largest known prime of the form $k \cdot 2^n + 1$ and in 1999 to find the largest known factor of a Fermat number: $3 \cdot 2^{382449} + 1$ divides F_{382447}.

In 1998, the program was expanded to also cover primes of the form $k \cdot 2^n - 1$. It was used with success to discover the largest known twin primes and the largest known Sophie Germain prime. But the major “theoretical” improvement was the ability to test some numbers of the form $k \cdot b^n + 1$, for small b, in about the same time as a number of the form $k \cdot 2^n + 1$ of the same size.

In 1999, the transform of the program was totally rewritten and the size of the testable numbers was extended to 5,000,000 digits by using a right-angle convolution and a balanced representation [4]. Finally an efficient test of the GFN was implemented, and this algorithm speeds up the search for GFN primes by a factor of two in comparison with other numbers of the form $k \cdot b^n + 1$.

2.2. **FFT multiplication.** To multiply numbers N_1 and N_2 using FFT, first a base W is selected. The numbers are converted to polynomials P_1 and P_2, such that $P_1(W) = N_1$ and $P_2(W) = N_2$ (the cost of this step is negligible if the base representation of N_1 and N_2 is W). The polynomial product $P = P_1 \cdot P_2$ is computed using FFT [2] [3]. Finally $N = N_1 \cdot N_2$ is obtained by computing $P(W)$ (if base representation of N is W, we just need to adjust the digits of N with add-and-carry). If the transform is done with floating point arithmetic, we should be sure that

$$W^2 \cdot n \cdot \log_2 n < 2^i \tag{2.1}$$

where n is the length of the transform and i is the number of bits in the floating mantissa.

This algorithm is used because it reduces the number of word operations to $O(n \log n)$.[1] But it has another important property: the computation time of the polynomial product is independent of W. If W is a power of 2, the conversions are

[1] The complexity of FFT multiplication is $O(n \log n \log\log n)$ where n is the number of ‘bits’ of N. Only integer schemes, such as Schönhage and Strassen or Nussbaumer methods, reach this complexity.

License or copyright restrictions may apply to redistribution; see http://www.ams.org/journal-terms-of-use

computed quickly, but because the computation time is dominated by the polynomial product, *the speed of floating point FFT multiplication is virtually independent of the base W as long as the base fits in the chosen word size* (*which is limited by relation* (2.1)).

2.3. **Test of numbers of the form $k \cdot b^n \pm 1$.** If a grammar-school or a Karatsuba multiplication is used, an efficient method for computing $x \cdot y \mod N$ for a fixed N is to use a steady-state divide [3, p. 9]. The modular multiplication will take about 3 multiplication times. But with FFT multiplication and $N = k \cdot b^n \pm 1$, we can select $W = b^m$ for m such that W is the largest possible integer in (2.1). Division is replaced by shifts, adds and one multiplication and one division by k for the base W.

If we define $W_{max} = \sqrt{2^i/(n \log_2 n)}$, the best case occurs for $b^m \approx W_{max}$ and the worst case for $b^m \approx \sqrt{W_{max}}$. Then *if $b < W_{max}$ and $k > 1$, the probable primality test of a number of the form $k \cdot b^n \pm 1$ is between as fast and two times slower than the test of a number of the form $k \cdot 2^n \pm 1$ of the same size.*

2.4. **Test of GFN.** When arithmetic is to be performed modulo Fermat or Mersenne numbers, Discrete Weighted Transforms effectively reduce FFT run length by a factor 2 [4]. Negacyclic convolutions and DWT are independent of the base representation. Then if $W = b^{2^n/N}$ (N is FFT length), $x \cdot y \mod F_{b,n}$ can be computed two times faster than $x \cdot y$. We have $m = 2^n/N$ and if N is a power of 2, m is also a power of 2. But practically, except for very small b, $m = 1$ or 2, this limitation doesn't slow down the global computation. Then *if $b < W_{max}$, the probable primality test of a GFN is between two times faster and as fast as the test of a number of the form $k \cdot 2^n \pm 1$ $(k > 1)$ of the same size.*

2.5. **The implementation of the tests.** The times for performing probable prime tests on numbers having about 5000 and 80000 digits are shown in Table 1. Because the FFT used by the program is a radix-2 FFT, the computation time has a discontinuity when the number of digits (in base W) is near a power of 2, where the time suddenly doubles. To make meaningful comparisons we include tests of numbers on both sides of the discontinuity.

The implementation of the test of the numbers of the form $k \cdot b^n + 1$ is different from the test of the GFN. For $k \cdot b^n + 1$, the modular reduction step uses some signed 32-bit integers and b^{2m} as base representation: it speeds up the reduction by a factor 2 but b should be smaller than $46340 < \sqrt{2^{31}}$. In Table 1, the values 215 and 216 were chosen to show a discontinuity due to internal base representation: $215^2 = 46225 < 46340$ then base 46225 is used, but $216^2 = 46656 > 46340$ then base 216 should be used. With the GFN, the size of b is just limited by relation (2.1) and the modular reduction uses directly base b. With this implementation, the test of the numbers of the form $k \cdot b^n + 1$ is between 12% and 25% slower than the expected theoretical result and the test of the GFN is between 20% and 40% slower. With the Proth program, the probable primality test of a GFN can be up to 70% faster than the test of a number of the form $k \cdot 2^n \pm 1$ $(k > 1)$ of the same size.

License or copyright restrictions may apply to redistribution; see http://www.ams.org/journal-terms-of-use

Table 1. Probable prime test times on a Pentium II 400

N	digits	W	FFT size	time
$3 \cdot 2^{16300} + 1$	4908	65536	1024	13 sec
$3 \cdot 2^{16400} + 1$	4938	65536	2048	29 sec
$6 \cdot 215^{2000} + 1$	4666	46225	1024	16 sec
$6 \cdot 215^{2100} + 1$	4899	46225	2048	36 sec
$6 \cdot 216^{2000} + 1$	4670	216	2048	36 sec
$6 \cdot 216^{2100} + 1$	4904	216	4096	76 sec
$65538^{1024} + 1$	4933	65538	512	9 sec
$258^{2048} + 1$	4939	258	1024	19 sec
$3 \cdot 2^{262000} + 1$	78871	65536	16384	83 mn
$3 \cdot 2^{263000} + 1$	79172	65536	32768	185 mn
$6 \cdot 215^{32000} + 1$	74639	46225	16384	93 mn
$6 \cdot 215^{33000} + 1$	76972	46225	32768	205 mn
$6 \cdot 216^{32000} + 1$	74704	216	32768	212 mn
$6 \cdot 216^{33000} + 1$	77038	216	65536	475 mn
$65538^{16384} + 1$	78914	65538	8192	50 mn
$258^{32768} + 1$	79024	258	16384	108 mn

2.6. **Remarks about the implementation.** To obtain some efficient tests, it is important to optimize the conversion routines in assembler: today, high-level languages don't have efficient operators to compute modular operations with floating point numbers and to convert them to integers.

Initially, the Proth program used a real-signal FFT with zero-padding. When the GFN test was implemented, a right-angle convolution [4] and a complex FFT was employed. Both implementations were tested on numbers of the form $k \cdot b^n \pm 1$ and the right-angle convolution is 20% faster. Theoretically, both methods are equivalent but today the second one is more adapted to the memory constraints of modern computers.

The probable primality of the GFN is tested by evaluating $3^{F_{b,n}-1} \pmod{F_{b,n}}$ and comparing the result to 1, a Fermat test. The primality is proved and verified by applying the Pocklington theorem [8, p. 52] for two different a.

3. Estimated distribution of GFN primes

The form of a Generalized Fermat Number is

$$F_{b,n} = b^{2^n} + 1 \; .$$

It is generally expected that there are an infinite number of primes of this form for each n. In fact, this is a consequence of the famous "Hypothesis H" in 1958 of Sierpiński and Schinzel. In 1962, Bateman and Horn indicated a quantitative form of "Hypothesis H" which could be used to predict the number of primes for given polynomials [1]. Since then there have been various studies that have heuristically confirmed the predictions for selected functions. See [8] for interesting details and history.

Because of the special characteristics of $F_{b,n}$ we can derive one set of equations for the expected number of primes for all n. Our approach will be to derive the

License or copyright restrictions may apply to redistribution; see http://www.ams.org/journal-terms-of-use

prime probability for each GFN, then add the appropriate GFN probabilities to determine the prime distribution.

If $F_{b,n}$ was a random number, the probability of it being prime would be approximately

$$Q(b,N) = \frac{1}{\log F_{b,n}} \sim \frac{1}{2^n \log b} \ .$$

But a GFN is not a random number. The probability of being prime is higher than this because b must be even and it is well known that the prime factors of $F_{b,n}$ must be of the form

$$P(k,n) = k \cdot 2^{n+1} + 1 \qquad \text{for } k = 1, 2, 3, \dots \quad .$$

In a recent paper concerning factors of generalized Fermat numbers, the first author and Wilfrid Keller determined that the probability of $P(k,n)$ dividing a particular $F_{b,n}$ is $2^n/P(k,n)$ [6].

Considering the possible divisors up to a limit $k = K$, the probability of $F_{b,n}$ being prime is decreased and must be multiplied by

$$t(K,n) = \prod_{k=1}^{K} \left(1 - \frac{2^n}{k \cdot 2^{n+1} + 1}\right) \quad \text{for } k \text{ such that } k \cdot 2^{n+1} + 1 \text{ is prime.} \tag{3.1}$$

After removing the above factors there are no factors of $F_{b,n}$, b even, up to $P(K,n)$. Then by Mertens' theorem [7, p. 351], the probability of being prime is increased and must be multiplied by $1/u(K,n)$ where

$$u(K,n) = 2 \prod_{p \text{ prime}}^{P(K,n)} \left(1 - \frac{1}{p}\right) \sim \frac{2}{e^\gamma \log(K \cdot 2^{n+1} + 1)} \ . \tag{3.2}$$

The extra factor of 2 is needed to account for the composite $F_{b,n}$ for b odd.

The new estimated probability for $F_{b,n}$ being prime becomes

$$C_n \cdot Q(b,n) \sim C_n \cdot \frac{1}{2^n \log b} \tag{3.3}$$

where

$$C_n = \lim_{K \to \infty} \frac{t(K,n)}{u(K,n)} \ . \tag{3.4}$$

It is shown in [1] that this limit exists.

C_n is easy to compute from equations (3.1) and (3.2) for moderate or even large values of K. The value remains virtually unchanged when K changes from 10^4 to 10^6 (see Table 2).

For each n the expected number of primes from $b = 2$ to B is the sum of the probabilities shown in (3.3),

$$E(B,n) = \frac{C_n}{2^n} \sum_{b=2}^{B} \frac{1}{\log b} \sim \frac{C_n}{2^n} \int_2^B \frac{dt}{\log t} \ . \tag{3.5}$$

License or copyright restrictions may apply to redistribution; see http://www.ams.org/journal-terms-of-use

TABLE 2. C_n as a function of K

n	2^n	$K = 10^4$	$K = 10^6$
1	2	1.3719	1.3727
2	4	2.6762	2.6788
3	8	2.0907	2.0927
4	16	3.6699	3.6712
5	32	3.6137	3.6129
6	64	3.9424	3.9424
7	128	3.1064	3.1085
8	256	7.4391	7.4347
9	512	7.4892	7.4888
10	1024	8.0157	8.0187
11	2048	7.2142	7.2256
12	4096	8.4193	8.4259
13	8192	8.4552	8.4676
14	16384	8.0030	8.0102
15	32768	5.7958	5.8026
16	65536	11.192	11.196

4. RESULTS

The derivation for the estimated number of GFN primes was developed at the same time as the actual results were being compiled. We were hoping that we might obtain reasonably good prime estimates at least for combinations of large values of n and B. We were pleasantly surprised to find that we were getting excellent results for all values of n and B.

The estimated number of primes is computed using equation (3.5). The estimated and actual values are shown in Table 3.

Bateman and Horn's quantitative form of "Hypothesis H" was previously numerically verified for some polynomials of degree 2, 3 or 4 [8, Ch. 6]. The actual distribution of GFN primes is in significant agreement with the values predicted by the conjecture for some polynomials of degree as large as 8192.

TABLE 3. Comparison between the estimates of the number of generalized Fermat primes and the actual number of primes found

$B =$	10^3			10^4			10^5			10^6		
2^n	*Est.*	Act.	Err.	*Est.*	Act.	Err.	*Est.*	Act.	Err.	*Est.*	Act.	Err.
2	*121*	111	−0.9	*855*	840	−0.5	*6609*	6655	+0.6	*53970*	54109	+0.6
4	*118*	110	−0.8	*834*	789	−1.6	*6449*	6395	−0.7	*52659*	52610	−0.2
8	*46*	40	−0.9	*326*	335	+0.5	*2519*	2498	−0.4	*20568*	20886	+2.2
16	*41*	48	+1.2	*286*	291	+0.3	*2209*	2194	−0.3	*18041*	17907	−1.0
32	*20*	22	+0.5	*141*	146	+0.5	*1087*	1062	−0.8	*8877*	8963	+0.9
64	*11*	8	−0.9	*77*	92	+1.7	*593*	606	+0.5	*4843*	4835	−0.1
128	*4*	7	+1.3	*30*	25	−1.0	*234*	242	+0.5	*1909*	1933	+0.5
256	*5*	4	−0.5	*36*	30	−1.0	*280*	272	−0.5	*2283*	2322	+0.8
512	*3*	1	−1.0	*18*	28	+2.3	*141*	160	+1.6	*1150*	1247	+2.9
1024	*1*	1	−0.3	*10*	14	+1.4	*75*	81	+0.6	*616*	578	−1.5
2048	*1*	1	+0.5	*4*	4	−0.2	*34*	40	+1.0	*277*	276	−0.1
4096	*0*	0	−0.6	*3*	2	−0.4	*20*	16	−0.9	*162*	170	+0.6
8192	*0*	0	−0.4	*1*	0	−1.1	*10*	3	−2.2	*81*	–	–
16384	*0*	0	−0.3	*1*	0	−0.8	*5*	1	−1.7	*38*	–	–

License or copyright restrictions may apply to redistribution; see http://www.ams.org/journal-terms-of-use

Table 4. The five largest primes found during the GFN search

b	2^n	digits
48594	65536	307140
167176	32768	171153
509622	16384	93508
506664	16384	93467
498904	16384	93357

Note that we are dealing with 56 separate prime distributions simultaneously. Fortunately there is a well-known statistical tool, the Chi Square Test, which gives a measure of how good a set of predictions matches the actual data as in our problem [5]. We eliminated some of the cases with little or no data, performed a χ^2 test on the remaining 51 cases and found that there was less than a 5% probability that the match between our estimates and the actual values happened by chance, an excellent result.

Next, we wanted to present the error between the estimated and actual prime counts in a meaningful way. Although we do not know the true distribution of GFN primes we can make the reasonable assumption that this distribution can be approximated by a Poisson distribution since this is true for almost all distributions of rare phenomena. We can then present the error between the estimated and actual number of primes as the number of standard deviations, which effectively normalizes the error. If N_p is the actual number of primes found (for fixed B and n), the error is defined by

$$\text{error} = \frac{N_p - E(B,n)}{\sqrt{E(B,n)}} \ .$$

These normalized errors are shown in Table 3. We would expect that about 68% of the errors would be within one standard deviation. This is actually true for 75% of the errors. The largest error is 2.9 standard deviations. Both of these are statistically reasonable results.

Only 3 primes were found for $n = 13$ and $B \leq 10^5$, where 10 were expected. The search was extended for this value of n. For $B \leq 2.5 \cdot 10^5$, 19 primes were found and 23 were expected.

During the extension of the search, Steve Scott discovered the largest known GFN prime $48594^{2^{16}} + 1$. Table 4 is a list of the 5 largest primes that were found.

5. Future studies

Historically, the largest known prime has almost always been a Mersenne prime. Since the DWT (discrete weighted transform) has been used to perform fast arithmetic modulo numbers of the form $2^n \pm 1$, the Mersenne numbers have been considered the best and "only" candidates for the largest known prime. But now, testing GFN's takes about the same time as testing Mersenne numbers, and because there are many more GFN candidates in a fixed range, a properly organized search could soon change the status of the largest known primes.

It is an open question whether the number of Fermat numbers that are prime is finite. As more factors of Fermat numbers are found it becomes more likely that

License or copyright restrictions may apply to redistribution; see http://www.ams.org/journal-terms-of-use

this is true. It appears worthwhile to try to extend the GFN theory and expand Table 2 in an attempt to make the probability argument more precise.

In the Bateman and Horn paper [1], the authors heuristically test an important conjecture by counting the number of primes of the form $p^2 + p + 1$ with p prime less than 113,000. In 1961 it took 400 minutes on the ILLIAC computer to find 776 primes. We repeated the computation in less than one second on a PII 400 PC! In 1960 Shanks did two related studies; the first concerned the number of primes of the form $n^2 + a$ [9], and the second primes of the form $n^4 + 1$ [10]. Modern resources should be used to extend the results of these excellent papers and of the GFN. We now have world-wide computing power which can generate very large amounts of data. There are many reliable statistical programs available for analyzing the data. There is every reason to expect that properly designed projects could result in more confidence in existing conjectures as well as the possibility of developing reliable estimates of second order effects. This might even contribute to a rigorous proof of such conjectures.

6. Acknowledgments

The authors are indebted to Kimmo Herranen for having tested the primality of the GFN for $n = 11$, 12 and $10^5 \leq b \leq 10^6$, and for $n = 13$ and $10^5 \leq b \leq 2.5 \cdot 10^5$.

References

1. P. T. Bateman and R. A. Horn, *A Heuristic Asymptotic Formula Concerning the Distribution of Prime Numbers*, Math. Comp. **16** (1962), 363–367. MR **26:**6139
2. R. E. Crandall, *Projects in Scientific Computation*, Springer TELOS, 1994. MR **95d:**65001
3. R. E. Crandall, *Topics in Advanced Scientific Computation*, Springer TELOS, 1996. MR **97g:**65005
4. R. Crandall and B. Fagin, *Discrete Weighted Transforms and Large-Integer Arithmetic*, Math. Comp. **62** (1994), 305–324. MR **94c:**11123
5. W. J. Dixon and F. J. Massey, Jr., *Introduction to Statistical Analysis*, 3rd ed., McGraw-Hill, 1969. MR **20:**1376
6. H. Dubner and W. Keller, *Factors of generalized Fermat numbers*, Math. Comp. **64** (1995), 397–405. MR **95c:**11010
7. G. H. Hardy and E. M. Wright, *An Introduction to the Theory of Numbers*, 5th ed., Oxford University Press, 1979. MR **81i:**10002
8. P. Ribenboim, *The New Book of Prime Number Records*, 3rd ed., Springer-Verlag, New York, 1995. MR **96k:**11112
9. D. Shanks, *On the Conjecture of Hardy & Littlewood concerning the Number of Primes of the Form $n^2 + a$*, Math. Comp. **14** (1960), 321–332. MR **22:**10960
10. D. Shanks, *On Numbers of the Form $n^4 + 1$*, Math. Comp. **15** (1961), 186–189. MR **22:**10941
11. A. Schinzel and W. Sierpiński, *Sur certaines hypothèses concernant les nombres premiers*, Acta Arith. **4** (1958), 185–208, Erratum **5** (1959), 259. MR **21:**4936

449 Beverly Road, Ridgewood, New Jersey 07450
E-mail address: `hdubner1@compuserve.com`

12 bis rue Perrey, 31400 Toulouse, France
E-mail address: `galloty@wanadoo.fr`

License or copyright restrictions may apply to redistribution; see http://www.ams.org/journal-terms-of-use