The EKG Sequence

J. C. Lagarias, E. M. Rains and *N. J. A. Sloane*
Information Sciences Research Center
AT&T Shannon Lab
Florham Park, NJ 07932–0971

Email addresses: jcl@research.att.com, rains@research.att.com, njas@research.att.com

(December 11, 2001; revised March 11, 2002)

Abstract

The EKG or electrocardiogram sequence is defined by $a(1) = 1$, $a(2) = 2$ and, for $n \geq 3$, $a(n)$ is the smallest natural number not already in the sequence with the property that $\gcd\{a(n-1), a(n)\} > 1$. In spite of its erratic local behavior, which when plotted resembles an electrocardiogram, its global behavior appears quite regular. We conjecture that almost all $a(n)$ satisfy the asymptotic formula $a(n) = n(1 + 1/(3 \log n)) + o(n/\log n)$ as $n \to \infty$; and that the exceptional values $a(n) = p$ and $a(n) = 3p$, for p a prime, produce the spikes in the EKG sequence. We prove that $\{a(n) : n \geq 1\}$ is a permutation of the natural numbers and that $c_1 n \leq a(n) \leq c_2 n$ for constants c_1, c_2. There remains a large gap between what is conjectured and what is proved.

1. Introduction

Consider the sequence defined by $a(1) = 1$, $a(2) = 2$ and, for $n \geq 3$, $a(n)$ is the smallest natural number not in $\{a(k) : 1 \leq k \leq n-1\}$ with the property that $\gcd\{a(n-1), a(n)\} \geq 2$. This sequence might be called a greedy gcd sequence, but because of its striking appearance when plotted we will name it the *EKG* (or *electrocardiogram*) *sequence*—see Figures 1, 2. It was apparently first discovered by Jonathan Ayres [Ayres 2001] and appears as sequence A064413 in [Sloane 2001]. The first 30 terms are

1	2	4	6	3	9	12	8	10	5	
15	18	14	7	21	24	16	20	22	11	
33	27	30	25	35	28	26	13	39	36	$\cdots$

Although the local behavior is erratic, plots of the first 1000 or 10000 terms show considerable regularity (see Figures 3, 4 in Section 4).

The EKG seqence has a simple recursive definition, yet seems surprisingly difficult to analyze. Its definition combines both additive and multiplicative aspects of the integers, and the greedy property of its definition produces a complicated dependence on the earlier terms of the sequence. Indeed, it is not immediately obvious whether it contains all positive integers, but we show this is the case—the EKG sequence is a permutation of the positive integers.

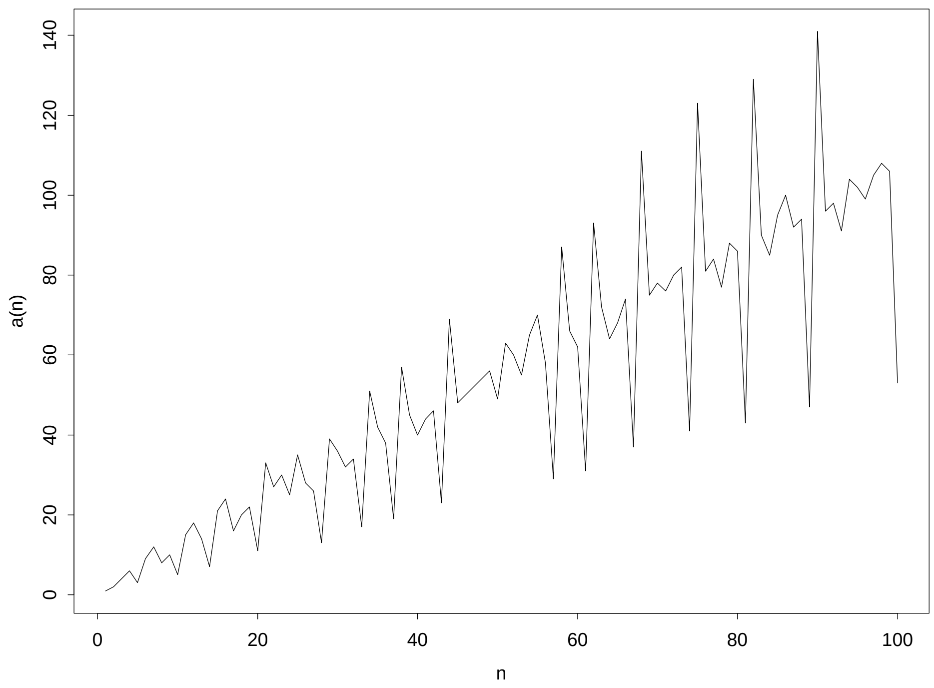

Figure 1: Plot of $a(1)$ to $a(100)$, with successive points joined by lines.

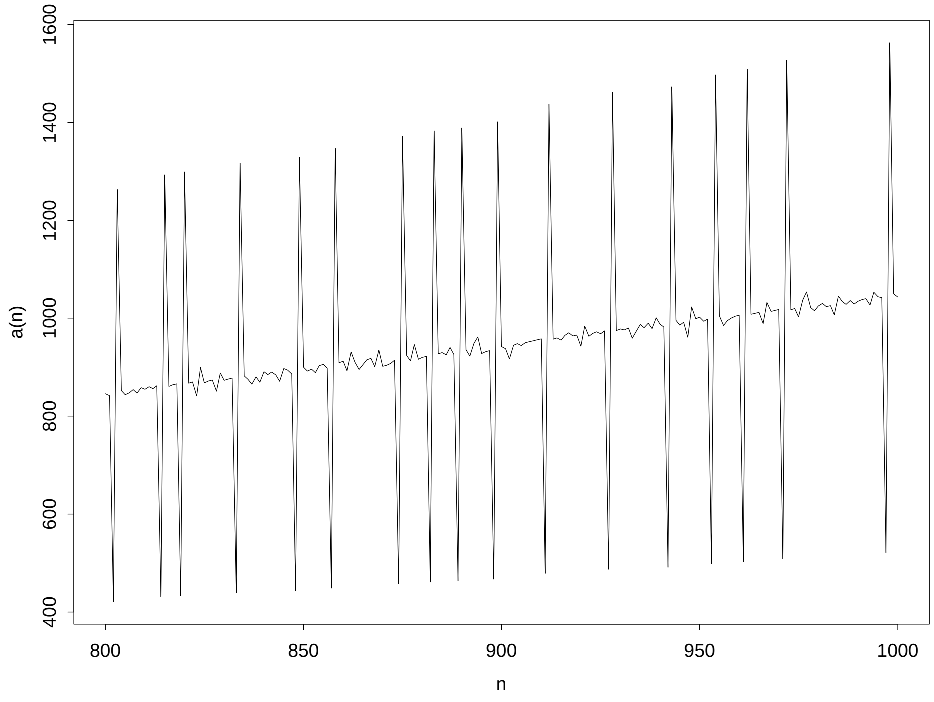

Figure 2: Terms 800 to 1000, with successive points joined by lines.

In comparing Figures 1 and 3, one is reminded of the contrast between the irregular plot of $\pi(x)$ (the number of primes $\le x$) for $x \le 100$ and the very smooth plot for $x \le 50000$ as shown in Don Zagier's lecture on "The first 50 million prime numbers" [Zagier 1977]. This is not a coincidence, as we will see, because (experimentally) the spikes in the EKG sequence are associated with the primes arranged in increasing order. However, the spacings between spikes are not the same as the spacings between consecutive primes.

Although the EKG sequence itself seems only to have been proposed recently, in the early 1980's Erdős, Freud and Hegyvari [Erdős et al. 1983] studied properties of integer permutations with restrictions placed on allowed values of greatest common divisors of consecutive terms.

In Section 2 we derive a number of basic properties of the EKG sequence, and prove that it is a permutation of the natural numbers $\mathbb{N}$. An efficient algorithm for computing the sequence is given in Section 3. Using this algorithm we computed 10^7 terms; this led us to conjectured asymptotic formulae given in Section 4. We give a heuristic argument why these formulae may be true, but it seems likely they will be very hard to prove. We are able to rigorously extablish linear upper and lower bounds on the EKG sequence, namely $\frac{1}{260}n \le a(n) \le 14n$ (Sections 5 and 6); the proofs use sieving ideas. Section 7 discusses experimental results concerning the cycle structure of the associated permutation of $\mathbb{N}$. The final section discusses generalizations to other sorts of integer permutations resulting from greedy constructions with restrictions on gcd's of consecutive terms.

2. The sequence is a permutation

We begin with some general remarks about the sequence.

For $n \ge 2$, let $g = \gcd\{a(n-1), a(n)\}$. For some prime p dividing $a(n-1)$, $a(n)$ is the smallest multiple of p not yet seen (otherwise the smaller multiple of p would be a better candidate for $a(n)$). We call such primes p the *controlling primes* for $a(n)$. There may be more than one, and their product divides g.

For any prime p and $n \ge 2$, let $B_p(n)$ be the smallest multiple of p that is not in $\{a(1), \dots, a(n-1)\}$. For example, the sequence $\{B_2(n) : n \ge 2\}$ begins 2, 4, 6, 8, 8, 8, 8, 10, 14, ... Clearly

$$B_p(n) \le B_p(n+1) \le pn \tag{1}$$

for all $p, n \ge 2$. Then we have $a(1) = 1$,

$$a(n) = \min\{B_p(n) : \; p \text{ divides } a(n-1)\} \; , \tag{2}$$

for $n \ge 2$, which provides an alternative definition of the sequence.

Lemma 1. *Let p be a prime > 2 that divides some term of the sequence. If p first divides $a(n)$ then $a(n) = qp$ where q is the smallest prime dividing $a(n-1)$, q is less than p, $a(n+1) = p$, and either $a(n)$ or $a(n+2)$ is equal to $2p$. The new primes that divide the terms of the sequence appear in increasing order.*

Proof. Let $a(n)$ be the first term divisible by p. The numbers pq where q is a prime dividing $a(n-1)$ are all candidates for $a(n)$, and so $a(n) = pq$ where q is the smallest such prime. Also p must be the smallest prime that has not appeared as a divisor of $\{a(1), \dots, a(n-1)\}$ (for if p' were a smaller such prime then $p'q$ would be a better candidate for $a(n)$). In particular the primes that divide the terms of the sequence must appear in increasing order, and $q < p$. Then p is a candidate for $a(n+1)$, and is less than $B_q(n+1) \ge B_q(n) = pq$, so $a(n+1) = p$. Finally, either $a(n) = 2p$ or else $2p$ is the winning candidate for $a(n+2)$. ∎

Lemma 2. *The primes that appear in the sequence occur in increasing order.*

Proof. This follows from Lemma 1, since the first time p divides a term of the sequence the next term is p itself. ∎

Lemma 3. *If infinitely many multiples of a prime p appear in the sequence then all multiples of p appear.*

Proof. We argue by contradiction, and let kp be the first multiple of p that is missed. Choose n_0 so that $a(n) > kp$ for all $n \geq n_0$. Since infinitely many multiples of p occur, there exists $n > n_0$ with $a(n) = lp$ for some l. But now we must have $a(n+1) = kp$, because $\gcd\{a(n), kp\} \geq p$ is allowed, and all smaller possible values which are ever going to appear in the sequence have already appeared. This is a contradiction. ∎

Lemma 4. *If all multiples of a prime p appear in the sequence then all positive integers appear.*

Proof. Again we argue by contradiction and let $k \geq 2$ be the first integer that is missed. Since infinitely many multiples of k occur among all the multiples of p, we get a contradiction just as in Lemma 3. Namely, there exists for $n > n_0$ a value $a(n) = klp$ for some l, and $\gcd\{a(n), k\} \geq k$ is allowed, and all smaller possible values have already been used. Thus $a(n+1) = k$, a contradiction. ∎

Theorem 1. $\{a(n) : n \geq 0\}$ *is a permutation of the natural numbers.*

Proof. No number can appear twice, by construction, so it suffices to show that every number appears. Suppose only finitely many different primes divide the terms of the sequence. Then one of them would appear infinitely many times, and Lemmas 3 and 4 would imply that all integers occur, which is a contradiction.

Therefore infinitely many different primes p divide the terms of the sequence. Then by Lemma 1 infinitely many even numbers $2p$ occur, by Lemma 3 all even numbers occur and by Lemma 4 all positive integers occur. ∎

Remark. As will be discussed in Section 8, the principle of this proof generalizes to a wide variety of other integer sequences defined by restrictions on the gcd's of consecutive terms.

3. Numerical investigations

To compute the EKG sequence it is better not to use the original definition but to use (2) and to store the current values of $B_p(n)$ for primes p. An efficient way to arrange the computation is to maintain four tables:

$$\begin{aligned}
\mathrm{hit}(m) &= 0 \text{ if } m \text{ has not yet appeared, otherwise } 1;\\
\mathrm{gap}(m) &= \text{current value of } B_m(n) \text{ if } m \text{ is a prime, otherwise } m;\\
\mathrm{small}(m) &= \text{smallest prime factor of } m;\\
\mathrm{quot}(m) &= \text{largest factor of } m \text{ not divisible by } \mathrm{small}(m).
\end{aligned}$$

Combining these tables in a C "struct" minimizes memory access.

Suppose we wish to compute the sequence until $a(n)$ reaches or exceeds N. The first step is to precompute small(m) and quot(m) for $m \leq N$. Since it is only necessary to consider primes $\leq \sqrt{N}$, this takes about $\Sigma_{p \leq \sqrt{N}} N/p = O(N \log\log N)$ steps.

In the main loop, let $a(n)$ be the current value. Set $k = a(n)$, $B = N$ and repeat until k reaches 1:

$$\begin{aligned} p &= \mathrm{small}(k)\,, \\ B &= \min\{B, \mathrm{gap}(p)\}\,, \\ k &= \mathrm{quot}(k)\,. \end{aligned}$$

Then we set $a(n+1) = B$, $\mathrm{hit}(B) = 1$, and update $\mathrm{gap}(q)$ for primes q dividing B.

We have not analyzed the complexity of the main loop in detail, but it also appears to take roughly $O(N \log\log N)$ steps, comparable to and not much greater than the number of steps needed for the precomputation part of the calculation. The program computed 10^7 terms of the sequence in less than a minute. For example $a(10954982) = 11184814$.

4. A conjectured asymptotic formula

The results from the experimental data suggest that whenever a prime p occurs in the sequence it is preceded by $2p$ and (consequently) followed by $3p$. Although it is theoretically possible that some other multiple of p occurs before p, for example, we might have seen $\ldots$, $3p$, p, $2p, \ldots$, this does not happen in the first 10^7 terms.

Conjecture 1. *Whenever a prime p occurs in the sequence it is immediately preceded by $2p$ (and hence followed by $3p$).*

The numerical results also strongly suggest that the terms of the sequence fall close to three lines (see Figs. 3, 4).

- if $a(n) = m$ and m is neither a prime nor three times a prime, then $a(n) \approx n$;
- if $a(n) = p$, p prime, then $a(n) \approx n/2$;
- if $a(n) = 3p$, p prime, then $a(n) \approx 3n/2$.

This was also observed by Ayres [Ayres 2001]. In fact, if we smooth the sequence by replacing every term $a(n) = p$ or $3p$, p prime > 2, by $a(n) = 2p$, the terms of the sequence lie close to a single line (see Fig. 5).

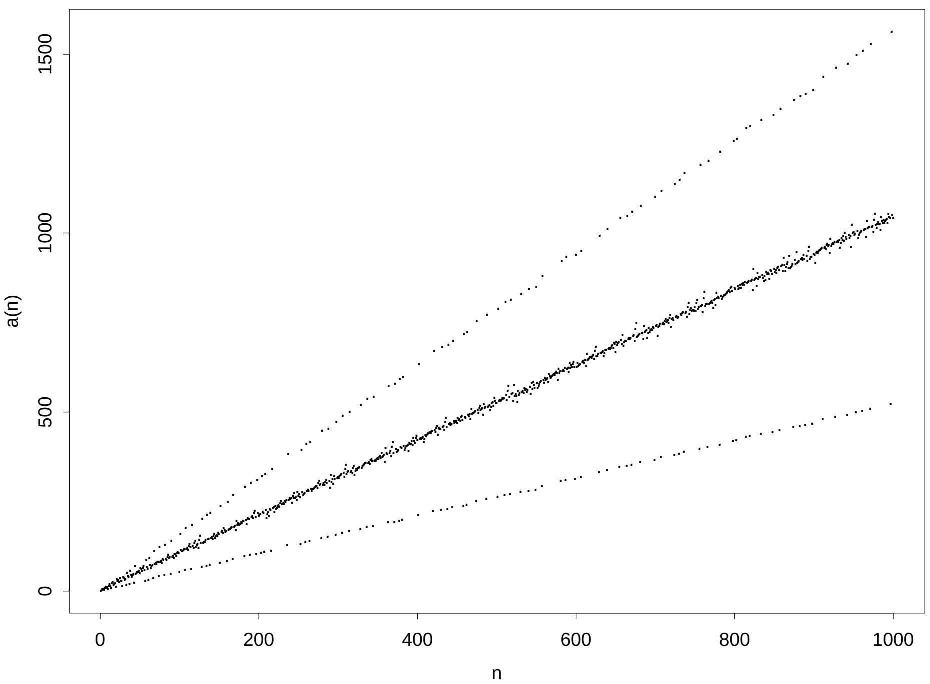

Figure 3: The first 1000 terms (represented by dots), successive points not joined.

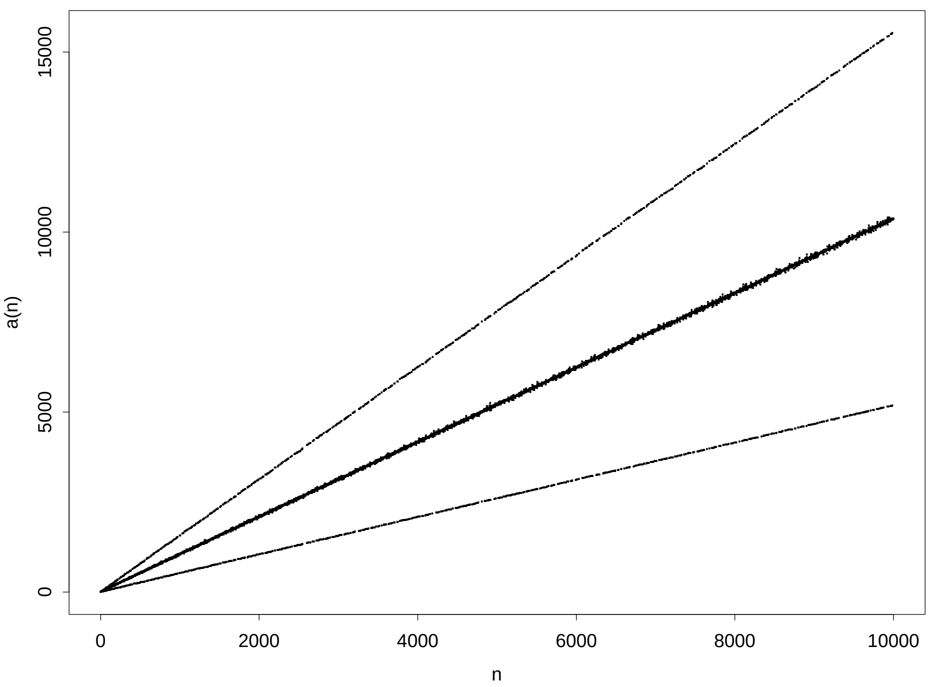

Figure 4: The first 10000 terms (represented by dots), successive points not joined.

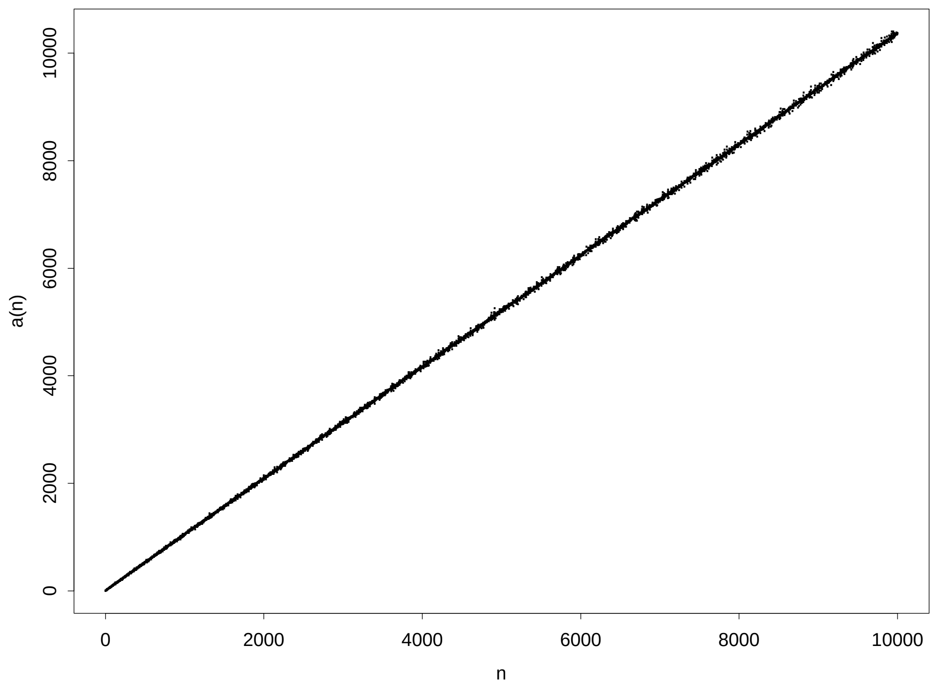

Figure 5: The sequence smoothed by replacing $a(n) = p$ or $3p$, p prime > 2, by $a(n) = 2p$.

A plausible but nonrigorous argument suggests a more precise conjecture.

Conjecture 2. *Let $f(n) \sim g(n)$ mean that the ratio of the two sides approaches 1 as $n \to \infty$.*
(1) *If $a(n) = m$, $m \neq p$ or $3p$ for p prime, then*

$$a(n) \sim n\left(1 + \frac{1}{3\log n}\right) ; \tag{3}$$

(2) *If $a(n) = p$, p prime, then*

$$a(n) \sim \frac{1}{2}n\left(1 + \frac{1}{3\log n}\right) ; \tag{4}$$

(3) *If $a(n) = 3p$, p prime, then*

$$a(n) \sim \frac{3}{2}n\left(1 + \frac{1}{3\log n}\right) . \tag{5}$$

To see why this conjecture might be true, consider a term $a(n) = m$ of the smoothed sequence. Examination of Fig. 5 suggests that the smoothed sequence has hit all the numbers from 1 to m at least once (the numbers occur a little out of order, but never by much). However, we have smoothed away the numbers p and $3p$ that are $\leq m$, while picking up the primes p and $3p$ for $2p \leq m$. Therefore

$$n \sim m - \pi(m) - \pi\left(\frac{m}{3}\right) + 2\pi\left(\frac{m}{2}\right) , \tag{6}$$

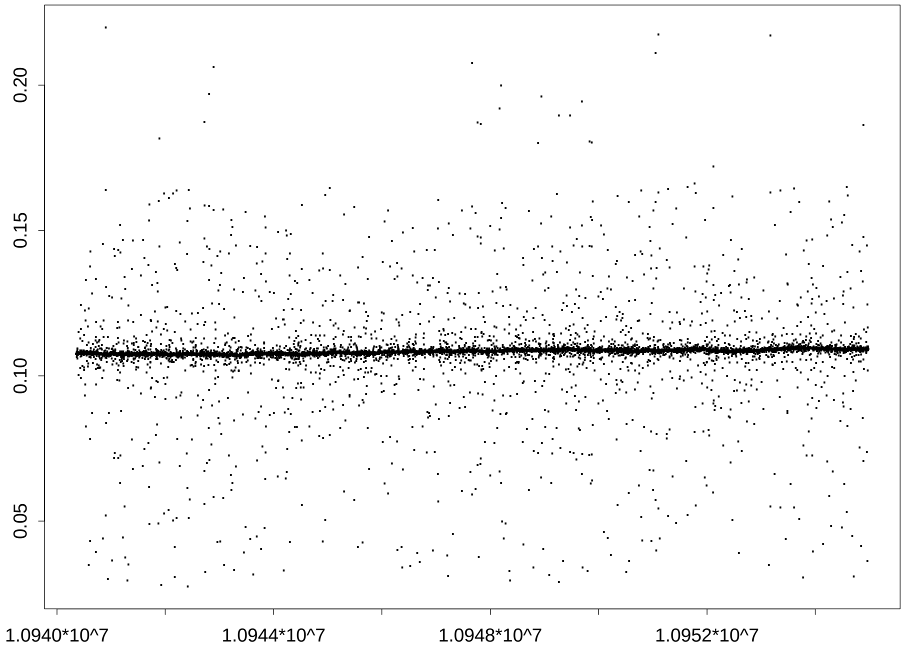

Figure 6: Values of c in (7) for n near 10^7.

where $\pi(x)$ = number of primes $\leq x$. Then (2) follows at once from the asymptotic formula $\pi(x) \sim x/\log x$. Equations (4) and (5) are based on (2) and the observations made at the beginning of this section.

Although we are unable to prove this conjecture, it is an excellent fit to the data.

If we try to write

$$a(n) \approx n\left(1 + \frac{1}{3\log n} + \frac{c}{(\log n)^2}\right) \quad (?) \tag{7}$$

then the values of c do not appear to converge to a single value (see Fig. 6), although c is very often close to 0.11. It seems conceivable that c might converge in distribution to a limiting distribution. Using two terms of the asymptotic expansion of $\pi(x)$ would give

$$a(n) \sim n\left(1 + \frac{1}{3\log n} + \frac{c'}{(\log n)^2}\right) \quad (?) \tag{8}$$

where $c' = 4/9 + (\log 3)/3 - \log 2 = 0.1175\ldots$.

Conjectures 1 and 2 predict that the k-th prime p_k will occur in the pattern $a(n) = 2p_k$, $a(n+1) = p_k$, $a(n+2) = 3p_k$, where

$$n \sim \frac{2p_k}{1 + \frac{1}{3\log(2p_k)}}.$$

These conjectures may be hard to settle, because the permutation $a(n)$ encodes an intricate interaction between additive and multiplicative properties of integers, which by the "greedy" property of the definition depends on all the earlier terms of the sequence.

In the next two sections we establish linear upper and lower bounds for the sequence, namely

$$\frac{1}{260}n \le a(n) \le 14n.$$

The numerical evidence supports the following conjectural bounds.

Conjecture 3. *The sequence $a(n)$ satisfies $a(n) \ge \frac{13}{28}n$, with equality if and only if $n = 28$, and $a(n) \le \frac{12}{7}n$, with equality if and only if $n = 7$.*

For large n the asymptotic lower and upper bounds would be $n/2$ and $3n/2$, by Conjecture 2.

5. A linear upper bound

In this section we show:

Theorem 2. *$a(n) \le 14n$, for $n \ge 1$.*

Remark. The proof is by contradiction, and the basic idea of the proof uses the fact that the function values $\{a(k) : 1 \le k \le n\}$ have a sandpile-like structure, in which each element can be built only at the top of a ladder of all smaller multiples of a controlling prime p dividing it. To reach a number at height exceeding $14n$ via a ladder of multiples of a controlling prime p, one repeatedly falls off this ladder while building it, at various smaller multiples kp of p of size between $2n$ and $14n$ (see property (P4) below). To get back on this ladder one must use ladders of other controlling primes q which reach to some currently omitted multiple of p at the top of their ladder. The total number of elements in such ladders is shown to be large using a combinatorial sieve argument (compare [Hooley 1976], pp. 4-5); a contradiction results by showing that the sandpile contains more than n elements.

We begin with a preliminary lemma.

Lemma 5. *If $a(n)$ is divisible by a prime p, then $p \le n$. If $p \ne 2$, $p < n$.*

Proof. The result is true if $p = 2$ or $n \le 3$, so we may assume $p \ge 3$ and $n \ge 4$. We may also assume $a(n)$ is the first term divisible by p. By Lemma 1, $a(n-1) = pq$ where q is the controlling prime for $a(n)$. Then $pq = a(n) = B_q(n) \le q(n-1)$ by (1). ∎

Proof of Theorem 2. We argue by contradiction, and let n be the smallest number such that $a(n) > 14n$. By direct verification we know n is large (in fact $n > 10^7$, although we will not use that in the proof). Let p be the smallest controlling prime for $a(n)$, say $a(n) = lp > 14n$, so $l \ge 15$. Also $a(n) = B_p(n)$, so from (1) and Lemma 5, $17 \le p < n$.

Let $l' = \lceil l/2 \rceil \ge 8$, and consider the "window"

$$\mathcal{W} = \{l'p, (l'+1)p, \ldots, (l-1)p\}.$$

Since p is a controlling prime for $a(n)$, every element of $\mathcal{W}$ has already appeared in the sequence: $a(t) \in \mathcal{W}$ implies $t \le n-1$. Call $a(t) \in \mathcal{W}$ an *entry point* if $a(t-1) \notin \mathcal{W}$, an *exit point* if $a(t+1) \notin \mathcal{W}$. The following properties hold.

(P1) The number of entry points is equal to the number of exit points.

(P2) At most one entry point $a(t)$ has p as a controlling prime. This can only happen if $a(t-1) = (l'-1)p$, $a(t) = l'p$, for some $t \leq n-1$.
(P3) At most one exit point $a(t)$ has p as a controlling prime for $a(t+1)$. This happens just if $a(n-1) \in \mathcal{W}$. Furthermore, if this is not the case then $a(n-1) = ip$ with $i < l'$, and hence there is no entry point with p as a controlling prime.
(P4) Let $a(t) = \alpha p \in \mathcal{W}$, $a \in [l', l-1]$. If α is a multiple of a prime $q \leq 7$, then $a(t)$ is an exit point. (For $a(t+1) \leq B_q(t+1) \leq 7t \leq 7(n-1) < l'p$.)

For a set of primes $\mathcal{P}$, let $D_{\mathcal{P}}(a, b)$ denote the number of integers $a \leq i \leq b$ such that i is a multiple of some element of $\mathcal{P}$. Let

$$\theta = D_{\{2,3,5,7\}}(l', l-1) \, .$$

By (P4), the number of exit points is at least θ, while if there exists an entry point controlled by p then there are in fact at least $\theta + 1$ exit points. By (P1) we conclude that the number of entry points not controlled by p is at least θ.

Let $\mathcal{S} = \{q_1, q_2, \ldots\}$ (with $q_1 < q_2 < \cdots$) denote the set of controlling primes for entry points not controlled by p. Note that $2, 3, 5, 7, p$ are not in $\mathcal{S}$ (by the same argument as in (P4)), and $|\mathcal{S}| \leq \pi(l-1) - 4$. Then

$$\begin{aligned} \theta &= D_{\{2,3,5,7\}}(l', l-1) \\ &\leq \text{number of entry points not controlled by } p \\ &\leq D_{\mathcal{S}}(l', l-1) \\ &\leq \sum_{q \in \mathcal{S}} \left\lceil \frac{l-l'}{q} \right\rceil \\ &\leq (l - l' - 1) \sum_{q \in \mathcal{S}} \frac{1}{q} + |\mathcal{S}| \, . \end{aligned} \tag{9}$$

Setting $\phi = \sum_{q \in \mathcal{S}} \frac{1}{q}$, we have

$$\phi \geq \frac{\theta - \pi(l-1) + 4}{\lfloor \frac{l}{2} \rfloor - 1} \, . \tag{10}$$

The right-hand side of (10) is a function of the single variable l, and is $\geq 2/9$ for all $l \geq 15$, with equality if and only if l is 20 or 21. (This is easily verified by computer for small l, say $l \leq 1000$, and analytically for larger l.) In other words,

$$\sum_{i \geq 1} \frac{1}{q_i} \geq \frac{2}{9} \, .$$

Define k by

$$\sum_{i=1}^{k-1} \frac{1}{q_i} < \frac{2}{9} \leq \sum_{i=1}^{k} \frac{1}{q_i} \, , \tag{11}$$

and let $\mathcal{S}' = \{q_1, \ldots, q_k\} \subseteq \mathcal{S}$. Note that $q_k \geq 17$, since $1/11 + 1/13 + 1/17 \geq 2/9$, but no proper subset of $\{1/11, 1/13, 1/17\}$ has this property.

Every multiple of any element $q \in \mathcal{S}$ that is $\leq l'p$ must have already occurred in the sequence, by definition. We obtain a contradiction by showing that there are more than n different multiples of elements of $\mathcal{S}$ that are $\leq l'p$. By inclusion-exclusion we have

$$D_{\mathcal{S}}(1, l'p) \geq D_{\mathcal{S}'}(1, l'p) \geq \sum_{i=1}^{k} \left\lfloor \frac{l'p}{q_i} \right\rfloor - \sum_{1 \leq i < j \leq k} \left\lfloor \frac{l'p}{q_i q_j} \right\rfloor \, . \tag{12}$$

To bound the first term in (12), observe that for $q_i \in \mathcal{S}$ there is a multiple of $q_i p$ in $\mathcal{W}$, so

$$q_i < \frac{14n}{17} < \frac{lp}{17},$$

$$\begin{aligned}
\left\lfloor \frac{l'p}{q_i} \right\rfloor &\geq 8, \\
\left\lfloor \frac{l'p}{q_i} \right\rfloor &\geq \frac{8}{9}\frac{l'p}{q_i}, \\
\sum_{i=1}^{k} \left\lfloor \frac{l'p}{q_i} \right\rfloor &\geq \frac{8}{9}\sum_{i=1}^{k}\frac{l'p}{q_i}.
\end{aligned}$$

To bound the second term in (12) we use

$$\begin{aligned}
&\sum_{i \leq i < j \leq k} \frac{1}{q_i q_j} \\
&= \frac{1}{2}\left(\sum_{i=1}^{k-1}\frac{1}{q_i}\right)^2 - \frac{1}{2}\sum_{i=1}^{k-1}\frac{1}{q_i^2} + \frac{1}{q_k}\sum_{i=1}^{k-1}\frac{1}{q_i} \\
&< \frac{1}{2}\left(\sum_{i=1}^{k-1}\frac{1}{q_i}\right)^2 + \frac{1}{2q_k}\sum_{i=1}^{k-1}\frac{1}{q_i}.
\end{aligned}$$

Then, since $q_k \geq 17$, we have

$$\begin{aligned}
D_{\mathcal{S}}(1, l'p) > l'p &\left\{ \frac{8}{9}\left(\sum_{i=1}^{k}\frac{1}{q_i}\right) \right. \\
&\left. -\frac{1}{2}\left(\sum_{i=1}^{k-1}\frac{1}{q_i}\right)^2 - \frac{1}{2q_k}\sum_{i=1}^{k-1}\frac{1}{q_i} \right\} \\
&\geq l'p\left\{\frac{8}{9}\cdot\frac{2}{9} - \frac{1}{2}\left(\frac{2}{9}\right)^2 - \frac{1}{34}\cdot\frac{2}{9}\right\} \\
&> 7n\frac{229}{1377} = \frac{1603}{1377}n > n,
\end{aligned}$$

which is the desired contradiction. ∎

6. A linear lower bound

In this section we show:

Theorem 3. *$a(n) \geq \lceil \frac{1}{260} n \rceil$, for $n \geq 1$.*

Remark. The proof is a modification of that of Theorem 1. It aims to show that if some number less than $n/260$ is missed in $\{a(k) : 1 \leq k \leq n\}$ then there are at least $n/65$ numbers in this set that are even numbers, and Lemma 7 below provides the mechanism to get a contradiction. The method of Theorem 1 seems inherently weaker when used for a lower bound, so we have not attempted to streamline the proof. It would certainly be possible to reduce the constant 260, but not to anything close to 14.

We begin with three lemmas.

Lemma 6. *For a prime p, if $a(n) = kp$ for some k then $a(j) = k$ for some $j \le n+1$.*

Proof. We argue by induction on k. The result is true for $k = 1$ since $a(1) = 1$.

Let q be a controlling prime for $a(n)$, so $q|a(n-1)$ and $q|a(n) = kp$. Case (i): $q \ne p$. Then $q|k$, say $k = mq$, $a(n) = mqp$. Hence all multiples iq with $i < mp$ have already appeared, and in particular $a(j) = mq = k$ for some $j < n$.

Case (ii): $q = p$. All multiples ip with $i < k$ have already occurred. By the induction hypothesis, i has occurred for all $i < k$. If k has occurred then $a(j) = k$ with $j < n$, and otherwise $a(n+1) = k$. ■

Lemma 7. *If at least $4k$ even numbers occur in $\{a(1), \ldots, a(n)\}$ then all numbers $\{1, \ldots, k\}$ occur in $\{a(1), \ldots, a(n+1)\}$.*

Proof. In view of Lemma 6 (taking $p = 2$) it is enough to show that $\{2, 4, \ldots, 2k\}$ are in $\{a(1), \ldots, a(n)\}$.

Suppose not, and let $2m$ be the largest even number $\le 2k$ not in $\{a(1), \ldots, a(n)\}$. Every even number $a(i) > 2m$ with $i \le n$ will be followed by $a(i+1) \le 2k$ (since $2m$ is always available). But in $\{a(1), \ldots, a(n)\}$ we have at least $4k$ even numbers, and so at least $2k$ even numbers $> 2k$. Therefore in $\{a(1), \ldots, a(n+1)\}$ we see all the numbers from 1 to $2k$, including $2m$, a contradiction. ■

Lemma 8. *If $a(n) = p$, a prime, then all numbers $\{1, \ldots, p-1\}$ occur in $\{a(1), \ldots, a(n-1)\}$.*

Proof. By Lemma 1, $a(n-1) = qp$, q prime, $q < p$; so $q, 2q, \ldots, (p-1)q$ have already appeared in $\{a(1), \ldots, a(n-2)\}$. The result now follows by Lemma 6. ■

Proof of Theorem 3. By direct verification we may assume $n \ge 260^2$. Let $m = \lceil n/260 \rceil$, and suppose, seeking a contradiction, that $a(n) < m$. Note that the lower bound on n implies that $n/m > 259$.

We will show that at least $4m$ even numbers have occurred in $\{a(1), \ldots, a(n-2)\}$, which gives a contradiction by Lemma 7. No primes greater than m can occur in this interval, or we get a contradiction by Lemma 8.

Some number $\ge n$ must occur among $\{a(1), \ldots, a(n-1)\}$, since there are $n-1$ numbers and $a(n) < m$ is missing. Suppose $a(j) \ge n$, with controlling prime $p \le m$, say $a(j) = lp$, and $j \le n-1$. Since $lp \ge n$, $l \ge 260$. Let $R \le m$ be the smallest missing number among $\{a(1), \ldots, a(n-1)\}$. Then $l \le R$, for if $l > R$ then we have seen Rp at time $j-1$, and by Lemma 6 we have seen R by time j, a contradiction. Therefore $l \le m$ and so $p \ge n/l > 259$. Both l and p are in the range $[260, m]$.

We consider the "window"

$$\mathcal{W} = \left\{ \left\lfloor \frac{lp}{4} \right\rfloor, \left\lfloor \frac{lp}{4} \right\rfloor + p, \ldots, lp \right\},$$

and define "entry" and "exit" point as in the proof of Theorem 2.

There are at least $\lfloor 3l/4 \rfloor$ multiples of p in $\mathcal{W}$, and at least $\lfloor 3l/8 \rfloor - 1$ of them are even. Any such even multiple of p is an exit point (from Lemma 7). There must therefore be at least $\lfloor 3l/8 \rfloor - 2$

entry points which are not controlled by p. Let $\mathcal{S} = \{q_1, q_2, \ldots\}$ (with $q_1 < q_2 < \cdots \le l$) be the set of controlling primes for these entry points. Then

$$D_{\mathcal{S}}\left(\left\lfloor \frac{l}{4} \right\rfloor, l\right) \ge \lfloor 3l/8 \rfloor - 2\,.$$

As in (9), (10), we get

$$\sum_{i \ge 1} \frac{1}{q_i} \ge \frac{\lfloor \frac{3l}{8} \rfloor - 2 - \pi(l)}{l - \lfloor \frac{l}{4} \rfloor} \tag{13}$$

which is $\ge 43/214$ for $l \ge 260$. We define k by

$$\sum_{i=1}^{k-1} \frac{1}{q_i} < \frac{43}{214} \le \sum_{i=1}^{k} \frac{1}{q_i}\,. \tag{14}$$

Every multiple of any element $q \in \mathcal{S}$ that is $\le \lfloor lp/4 \rfloor$ must have already occurred in $\{a(1), \ldots, a(n-2)\}$. Of these, at least $\lfloor lp/8q \rfloor$ are even. Therefore the number of distinct even numbers in the range $1, \ldots, \lfloor lp/4 \rfloor$ that have occurred is at least

$$\sum_{i \ge 1}^{k} \left\lfloor \frac{X}{2q_i} \right\rfloor - \sum_{1 \le i < j \le k} \left\lfloor \frac{X}{2q_i q_j} \right\rfloor$$

where $X = lp/4$. Proceeding as in the proof of Theorem 2, and using $q_1 \ge 2$, we find that this is at least $4.024m$. This exceeds $4m$ and provides the desired contradiction. ∎

7. Cycle structure

Since the sequence is a permutation of $\mathbb{N}$, it is also natural to investigate the cycle structure. The experimental evidence suggests that there are infinitely many finite cycles and infinitely many infinite cycles. It seems very likely to be hard to prove either of these observations, or even to prove that there is at least one infinite cycle.

The first few finite cycles start at the points

$$1, 2, 3, 8, 40, 64, 121, 149, 359, 2879, 5563, 28571, 251677,$$

and have lengths

$$1, 1, 6, 1, 1, 1, 2, 12, 11, 25, 8, 22, 11$$

respectively. There are also a large number of apparently infinite cycles, of which the first two are

$$\cdots 229310, 117833, \ldots, 22, 27, 26, 28, 13, 14, \mathbf{7}, 12, 18, 20, 11, 15, 21, \ldots, 636551, 652766, \ldots$$

and

$$\ldots, 502008, 257519, \ldots, 248, 253, 131, 138, \mathbf{73}, 82, 129, 201, 212, \ldots, 645906, 662330, \ldots$$

with minimal representatives 7 and 73 respectively. The first fifteen of these apparently distinct cycles have not coalesced in the first 700000 terms of the sequence. However, although it seems unlikely, it is theoretically possible that they could coalesce at some later point. It would be nice to know more!

8. Generalizations

The EKG sequence can be generalized in various ways, while retaining the basic construction of a greedy sequence with a condition on gcd's of consecutive terms. For fixed $M \geq 2$, let $b(n) = n$ for $1 \leq n \leq M$, and for $n \geq M+1$ let $b(n)$ be the smallest natural number not already in the sequence with the property that $\gcd\{b(n-1), b(n)\} \geq M$. The proof of Theorem 1 easily extends to show that $(b(n) : n \geq 1)$ is also a permutation of $\mathbb{N}$. For the cases $M = 3, 4, 5$ see sequences A064417, A064418, A064419 in [Sloane 2001].

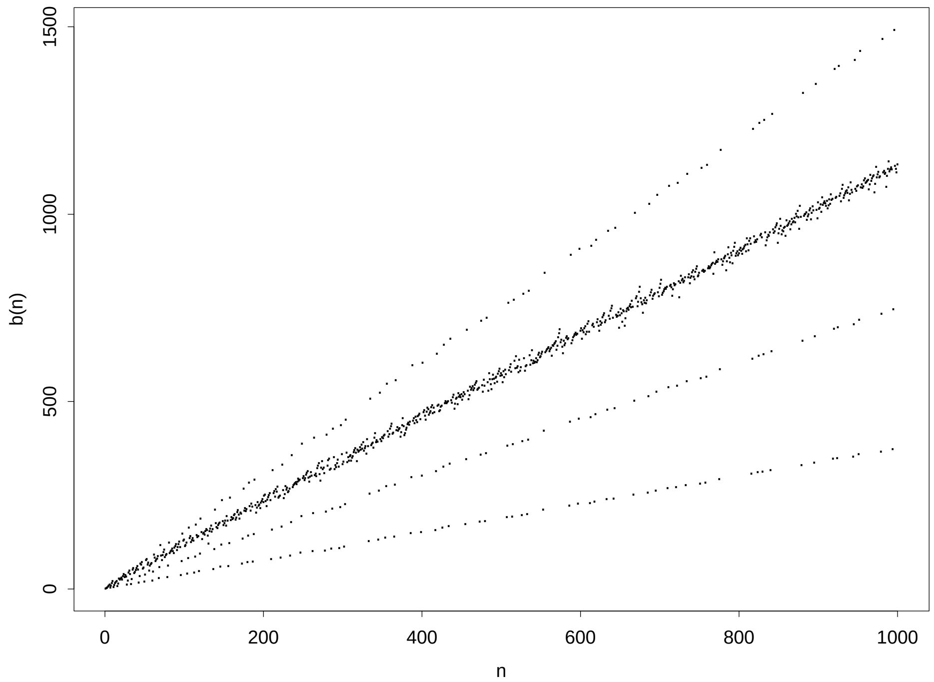

Figure 7: Plot of $b(1)$ to $b(1000)$ for the case $M = 3$.

The first 1000 terms for the case $M = 3$ are shown in Figure 7. This sequence appears to behave in a similar way to the EKG sequence: the spikes in the sequence are associated with the primes, occurring in the order $3p, p, 2p, 4p$. All points of the sequence seem to lie near lines of slope $1/3, 2/3, 1$ and $4/3$.

More generally, we can allow an arbitrary finite prefix at the beginning of such sequences. Consider a sequence $c(n)$ with the generating rule that $\gcd\{c(n-1), c(n)\} \geq M$, for $n \geq N$, but with a finite prefix $c(n) = a_n : 1 \leq n \leq N$, such that all terms a_n are distinct natural numbers, and that the prefix includes values $a(n) = k$ for of $1 \leq k \leq M$. With a little work the proof of Theorem 1 can be modified to apply to these sequences as well to show they are permutations.

If we use the greedy gcd rule with $M = 2$, starting with an arbitrary finite prefix, one obtains an infinite number of different EKG-like permutations. Many of these sequences, perhaps all, eventually coalesce with the original EKG sequence. In any case, the qualitative properties of the resulting permutations appear similar to the EKG sequence: the primes provably appear in consecutive order (excluding those in the prefix terms); the general plot of the permutations remains similar to Figure 3.

It appears that linear upper and lower bounds hold for all the permutations $c(n)$ of these types.

Very likely such bounds can be established rigorously in any particular case using methods similar to those used in proving Theorems 2 and 3.

Acknowledgements

We thank Jonathan Ayres for discovering this wonderful sequence. We also thank a referee for helpful comments.

References

[**Ayres 2001**] J. Ayres, personal communication, Sept. 30, 2001.

[**Erdős et al. 1983**] P. Erdős, R. Freud and N. Hegyvari, Arithmetical properties of partitions of integers, *Acta Math. Acad. Sci. Hungar.*, **41** (1983), 169–176.

[**Hooley 1976**] C. Hooley, *Applications of Sieve Methods to the Theory of Numbers*, Cambridge Tracts in Math. No. 70, Cambridge Univ. Press: Cambridge 1976.

[**Sloane 2001**] N. J. A. Sloane, *The On-Line Encyclopedia of Integer Sequences*, published electronically at www.research.att.com/∼njas/sequences/.

[**Zagier 1977**] D. Zagier, The first 50 million prime numbers, *Math. Intelligencer*, **0** (1977), 7–19.