Binary Early Birds

Hagen von Eitzen

June 29, 2009

1 Notation

In the following binary representations are written using bold digits, i.e. as strings over the alphabet $\{\mathbf{0}, \mathbf{1}\}$. Let $\mathrm{bin}(n)$ denote the binary representation of n as such a string. For example $\mathrm{bin}(0) = \mathbf{0}$, $\mathrm{bin}(1) = \mathbf{1}$, $\mathrm{bin}(2) = \mathbf{10}$. Let $\mathrm{wt}(n)$ denote the binary weight of n, i.e. the number of $\mathbf{1}$'s in $\mathrm{bin}(n)$. In describing such strings, parentheses are used for grouping and powers for repetition. Thus for example $\mathbf{111}\mathbf{0}^4(\mathbf{101})^4$ is short for $\mathbf{1110000101101101101}$. Let $\mathrm{len}(\alpha)$ denote the length of string α.

2 Introduction

In June 2009, Paolo P. Lava and Giorgio Balzarotti submitted [1, sequence A161373] and its complement [1, sequence A161374] to the On-Line Encyclopedia of Integer Sequences: Write all nonnegative integers in binary and concatenate to obtain the infinite string[1] (marks between constituents are inserted for clarity only)

$$\begin{aligned} S &= \mathrm{bin}(0)\,\mathrm{bin}(1)\,\mathrm{bin}(2)\,\mathrm{bin}(3)\,\mathrm{bin}(4)\,\mathrm{bin}(5)\cdots \\ &= \mathbf{0'1'10'11'100'101'110'111'1000'1001'1010'1011'1100'1101'}\ldots \end{aligned}$$

For $i \in \mathbb{N}$ we write $\mathrm{ofs}(n, i)$ ("n occurs at offset i") if the substring of length $\mathrm{len}(\mathrm{bin}(n))$ starting at offset i in S equals $\mathrm{bin}(n)$. By the very definition of S, each $n \geq 0$ has its "natural" position in that infinite string: If we define $\nu(0) = 1, \nu(1) = 2$ and recursively $\nu(n+1) = \nu(n) + \mathrm{len}(\mathrm{bin}(n))$,[2] then we have $\mathrm{ofs}(n, \nu(n))$. If $\mathrm{ofs}(n, i)$ then there is a unique m with $\nu(m) \leq i < \nu(m+1)$, in which case we write $n \,@\, m$.

Clearly, each n occurs at infinitely many offsets, e.g. $\mathrm{ofs}(n, \nu(2^k n))$ for $k \geq 0$ if $n > 0$. For some n we observe that it occurs even before $\nu(n)$, e.g. $\mathrm{ofs}(13, 2)$, $\mathrm{ofs}(13, 13)$ and $\mathrm{ofs}(13, 26)$ whereas $\nu(13) = 39$:

$$S = \mathbf{0'}\underbrace{\mathbf{1'10'1}}_{\substack{\text{early occurance}\\ 13\,@\,1}}\mathbf{1'100'101'}\underbrace{\mathbf{110'1}}_{\substack{\text{early occurance}\\ 13\,@\,6}}\mathbf{11'1000'100}\underbrace{\mathbf{1'101}}_{\substack{\text{early occurance}\\ 13\,@\,9}}\mathbf{0'1011'1100'}\underbrace{\mathbf{1101}}_{\text{natural occurance}}\mathbf{'1110'}\ldots$$

Numbers that occur before their natural occurance at least once are called "early birds", all other numbers are called "punctual".

[1] Viewed as a number sequence this is [1, sequence A030190].

[2] For $n > 0$, $\nu(n) = \mathrm{A083652}(n-1) + 1 = 3 + n\lceil \log_2 n \rceil - 2^{\lceil \log_2 n \rceil}$, cf. [1, sequence A083652].

There is a slight ambiguity in the words "occur before". For example, the number $n = 22$ has an early occurance that overlaps with its natural occurance:

$$\ldots{}^{\shortmid}\mathbf{10010}{}^{\shortmid}\mathbf{10011}{}^{\shortmid}\mathbf{10100}{}^{\shortmid}\mathbf{10}\underbrace{\mathbf{101}{}^{\shortmid}\mathbf{10}}_{\substack{\text{overlapping occurance}\\ 22\,@\,21}}\mathbf{110}{}^{\shortmid}\ldots$$

One can choose to accept or not accept such overlapping occurances as early.[3] In the following we will investigate both options to define the set E resp. $\hat{E}$ of early birds:

$$\begin{aligned} E &= \{n \in \mathbb{N}_0 \mid \exists i \leq \nu(n) - \mathrm{len}(\mathrm{bin}(n))\colon\ \mathrm{ofs}(n,i)\} \\ \hat{E} &= \{n \in \mathbb{N}_0 \mid \exists i < \nu(n)\colon\ \mathrm{ofs}(n,i)\}. \end{aligned}$$

Clearly, $\hat{E} \subseteq E$ and as indicated above one can verify $22 \in E \setminus \hat{E}$. We will see below that in fact $\hat{E} \setminus E = \{22\}$.

3 Preliminary Results

Lemma 1. *If $n\,@m$ with $m < n$, say $\mathrm{ofs}(n,i)$ with $\nu(m) \leq i < \nu(m+1)$, then $\nu(m+1) < i + \mathrm{len}(\mathrm{bin}(n))$. We have $n \in \hat{E}$ if and only if $n\,@m$ for some $0 \leq m < n$. We have $n \in E$ if and only if $n\,@m$ for some $0 \leq m < n-1$.*

Proof. Assume $\mathrm{ofs}(n,i)$ with $\nu(m) \leq i < \nu(m+1)$. If $\nu(m+1) \geq i + \mathrm{len}(\mathrm{bin}(n))$, then $\mathrm{bin}(n)$ is a substring of $\mathrm{bin}(m)$ and hence $m \geq n$. This is the first claim. Since ν is strictly increasing, we see that $i < \nu(n) \iff m + 1 \leq n$, hence the second claim follows. For the last claim observe that $\nu(n-1) \leq i < \nu(n)$ implies $i + \mathrm{len}(\mathrm{bin}(n)) \geq i + \mathrm{len}(\mathrm{bin}(n-1)) \geq \nu(n-1) + \mathrm{len}(\mathrm{bin}(n-1)) = \nu(n)$. □

Corollary 1. *If $k \geq 0$ then $2^k \notin \hat{E}$. If $k \geq 1$ then $2^{2k+1} + 2^k \notin \hat{E}$.*

Proof. By lemma 1, $n \in \hat{E}$ implies $i < \nu(m+1) < i + \mathrm{len}(\mathrm{bin}(n))$ for some $0 \leq m < n$ and i with $\mathrm{ofs}(n,i)$. That is, the leading digit $\mathbf{1}$ of $\mathrm{bin}(m+1)$ must correspond to a non-leading digit of $\mathrm{bin}(n)$. If $n = 2^k$, all non-leading digits are $\mathbf{0}$, hence the first claim.

If on the other hand $n = 2^{2k+1} + 2^k$, i.e. $\mathrm{bin}(n) = \mathbf{10}^k\mathbf{10}^k$, then the first half of that string must be the least significant $k+1$ digits of $\mathrm{bin}(m)$ and hence the least significant $k+1$ digits of $\mathrm{bin}(m+1)$ are $\mathbf{10}^{k-1}\mathbf{1}$. Since there is no third $\mathbf{1}$ in $\mathrm{bin}(n)$, the second half must be the most significant $k+1$ digits of $\mathrm{bin}(m+1)$. The strings $\mathbf{10}^k$ and $\mathbf{10}^{k-1}\mathbf{1}$ cannot overlap, hence $\mathrm{bin}(m+1) = \mathbf{10}^k\alpha\mathbf{10}^{k-1}\mathbf{1}$ for some string α, which implies $m + 1 > n$ in contradiction to $m < n$. Hence the second claim. □

Lemma 2. *If $n = 2^j + 2^k$ with $j > k \geq 0$ and $j \neq 2k+1$, then $n \in E$.*

[3] The author first considered 22 an early bird because calling it punctual felt strange. However, one of the original submitters has clarified in [2] that 22 is not an early bird in their intended interpretation, i.e. only an early occurance of n in A that does not even overlap the natural one should count.

There were also two other discrepancies in the original submission: It contained 20 as punctual, but that is definitely an early bird in both interpretations; and it started with an offset of 0, even though the OEIS usually uses offset 1 for sequences like this that describe a list or a subset of $\mathbb{N}_0$ (as opposed to more "function-like" sequences), see [1, FAQ "What does the 'offset' mean?"].

Proof. Let $r = j - k - 1$. Then by assumption, $r \geq 0$ and $r \neq k$. The binary representation of n is $\mathrm{bin}(n) = \mathbf{10}^r\mathbf{10}^k$. We distinguish three cases:

- $k < r$: Then $r - k - 1 \geq 0$ and $n \,@\, m$ with $m = 2^r = \frac{n-2^k}{2^{k+1}} < n - 1$:

$$\dots|\underbrace{\mathbf{10}^r|\mathbf{10}^k}_{n}\mathbf{0}^{r-k-1}\mathbf{1}|\dots$$

- $r = 0$: Then $k \geq 1$ and $n \,@\, m$ with $m = 2^k - 1 = n - 2^j - 1 < n - 1$:

$$\dots|\,\mathbf{1}^{k-1}\underbrace{\mathbf{1}|\mathbf{10}^k}_{n}|\dots$$

- $k > r > 0$: Then $n \,@\, m$ with $m = 2^j + 2^r = n - 2^r(2^{k-r} - 1) < n - 1$:

$$\dots|\,\mathbf{10}^k\underbrace{\mathbf{10}^r|\mathbf{10}^k}_{n}\mathbf{10}^{r-1}\mathbf{1}|\dots$$

In all cases we thus find $n \in E$ by lemma 1. $\square$

Lemma 3. *If* $\mathrm{wt}(n) \geq 3$ *and* $n \neq 22$ *then* $n \in E$.

Proof. Assume $\mathrm{wt}(n) \geq 3$. Then the binary representation of n can be written as $\mathrm{bin}(n) = \mathbf{10}^a\mathbf{1}\beta\mathbf{10}^c$ with $a \geq 0$, $c \geq 0$ and a possibly empty string β. Let $b = \mathrm{len}(\beta) \geq 0$. Clearly $\mathrm{wt}(n) \leq b + 3$. To show the existence of $m < n - 1$ with $n \,@\, m$ we distinguish several cases:

- $\mathrm{wt}(n) < b + 3$: Then β contains at least one $\mathbf{0}$. Write $\beta = \gamma\mathbf{01}^t$ with $t \geq 0$ and a string γ and let $\delta = \gamma\mathbf{10}^t$. Then $\mathrm{bin}(m) = \alpha\beta$ implies $\mathrm{bin}(m + 1) = \alpha\delta$. We distinguish two subcases:

  - $c > a$: Let m be the number with $\mathrm{bin}(m) = \mathbf{10}^c\mathbf{10}^a\mathbf{1}\beta$. Then $\mathrm{bin}(m+1) = \mathbf{10}^c\mathbf{10}^a\mathbf{1}\delta$. Since $\mathrm{len}(\mathrm{bin}(m + 1)) = \mathrm{len}(\mathrm{bin}(n))$, comparing the c leading digits yields $m + 1 < n$. Then $n \,@\, m$:

$$\dots|\,\mathbf{10}^c\underbrace{\mathbf{10}^a\mathbf{1}\beta|\mathbf{10}^c}_{n}\mathbf{10}^a\mathbf{1}\delta|\dots$$

  - $c \leq a$: Let m be the number with $\mathrm{bin}(m) = \mathbf{10}^a\mathbf{1}\beta$. Then $\mathrm{bin}(m+1) = \mathbf{10}^a\mathbf{1}\delta$. Since this has less digits than n, we see that $m + 1 < n$. Then $n \,@\, m$:

$$\dots|\underbrace{\mathbf{10}^a\mathbf{1}\beta|\mathbf{10}^c}_{n}\mathbf{0}^{a-c}\mathbf{1}\delta|\dots$$

- $\mathrm{wt}(n) = b+3$. Then $\beta = \mathbf{1}^b$ and $\mathrm{bin}(n) = \mathbf{10}^a\mathbf{1}^{b+2}\mathbf{0}^c$. Again we distinguish subcases:

  - $c \geq a > 0$: Let m be the number with $\mathrm{bin}(m) = \mathbf{10}^c\mathbf{10}^a\mathbf{1}^{b+1}$. Then $\mathrm{bin}(m + 1) = \mathbf{10}^c\mathbf{10}^{a-1}\mathbf{10}^{b+1}$ and $n \,@\, m$:

$$\dots|\,\mathbf{10}^c\underbrace{\mathbf{10}^a\mathbf{1}\beta|\mathbf{10}^c}_{n}\mathbf{10}^{a-1}\mathbf{10}^{b+1}|\dots$$

    Note that $\mathrm{len}(\mathrm{bin}(m + 1)) = \mathrm{len}(\mathrm{bin}(n))$. If $c > a$ or $a > 1$ or $b > 0$, then one finds $m + 1 < n$ by comparing digitwise from left to right. Hence the only case where this fails to prove $n \in E$ is $c = a = 1$ and $b = 0$, i.e. $n = 22$.

– $c \geq a = 0$: Then $\mathrm{bin}(n) = \mathbf{1}^{b+3}\mathbf{0}^c$, i.e. $n = 2^{b+c+3} - 2^c$.

  ∗ $b + 2 \geq c$: Then $n\,@\,m$ with $m = 2^{b+2} - 1 = \frac{n-2^c}{2^{c+1}} < n - 1$:

$$\ldots\underbrace{\mathbf{1}^{b+2}\mathbf{1}\mathbf{0}^c}_{n}\mathbf{0}^{b+2-c}\ldots$$

  ∗ $b + 2 < c$: Then $n\,@\,m$ with $m = 2^c - 1 < n - 1$:

$$\ldots\mathbf{1}^{c-b-2}\underbrace{\mathbf{1}^{b+2}\mathbf{1}\mathbf{0}^c}_{n}\ldots$$

– $c < a$: Then $n\,@\,m$ with $m = 2^{a+b+1} + 2^{b+1} - 1 = \frac{n-2^c}{2^{c+1}} < n - 1$:

$$\ldots\underbrace{\mathbf{1}\mathbf{0}^a\mathbf{1}^{b+1}\mathbf{1}\mathbf{0}^c}_{n}\mathbf{0}^{a-c-1}\mathbf{1}\mathbf{0}^{b+1}\ldots$$

□

4 Main results

We can summarize the above results as

Theorem 1.

$$\begin{aligned} \hat{E} &= \{n \in \mathbb{N} \mid \mathrm{wt}(n) \geq 2\} \setminus \{2^{2k+1} + 2^k \mid k \geq 1\} \\ E &= \hat{E} \setminus \{22\} \end{aligned}$$

Proof. For $n \in \{0, 3, 22\}$ one directly verifies $0 \notin \hat{E}$, $3 \in E$, $22 \notin E$, $22 \in \hat{E}$. All other cases are treated by lemmas 1, 2 and 3 together with the fact $E \subseteq \hat{E}$. □

Corollary 2. *For $n \in \mathbb{N}$ let $a(n)$ denote the nth element of $\mathbb{N}_0 \setminus E$ in order and let $\hat{a}(n)$ denote the nth element of $\mathbb{N}_0 \setminus \hat{E}$ in order. Write $n = 3q + r$ with integers q, r and $0 \leq r < 3$. Then for $n > 8$ we have*

$$a(n) = \begin{cases} 2^{2q-1} & \text{if } r = 0, \\ 2^{2q-1} + 2^{q-1} & \text{if } r = 1, \\ 2^{2q} & \text{if } r = 2. \end{cases} \tag{1}$$

and for $n > 3$ we have

$$\hat{a}(n) = \begin{cases} 2^{2q-1} + 2^{q-1} & \text{if } r = 0, \\ 2^{2q} & \text{if } r = 1, \\ 2^{2q+1} & \text{if } r = 2. \end{cases} \tag{2}$$

Proof. For $s \geq 1$, one quickly checks that the numbers of the form $n = 2^k$ with $2^{2q-1} \leq n < 2^{2(q+1)-1}$ are exactly 2^{2q-1} and 2^{2q} and the only number $n = 2^{2k+1} + 2^k$ with $2^{2q-1} \leq n < 2^{2(q+1)-1}$ is exactly $2^{2q-1} + 2^{q-1}$. Since $2^{2q-1} < 2^{2q-1} + 2^{q-1} < 2^{2q}$, we conclude from theorem 1 that the right hand side of (1) for $n \geq 9$ indeed runs through the punctual numbers $\geq 2^{2\cdot 3-1} = 32$ in order.

Using theorem 1 again, we see that the first terms of the sequence are $a(1) = 0$, $a(2) = 1$, $a(3) = 2$, $a(4) = 4$, $a(5) = 8$, $a(6) = 10$, $a(7) = 16$, $a(8) = 22$. Since this is followed by $a(9) = 32$, the first claim of the corollary follows.

Sine $\hat{E} = E \cup \{22\}$ we have $\hat{a}(n) = a(n+1)$ for $n > 7$. Therefore (2) holds at least for $n > 7$. It is also directly verified for $n \in \{4, 5, 6, 7\}$. □

Corollary 3. *The generating functions of the sequences* $a, \hat{a}$ *are given by*

$$\begin{aligned} A(x) &= \sum_{n=1}^{\infty} a(n)x^n \\ &= \frac{1+x+2x^2}{2(1-4x^3)} + \frac{x}{2(1-2x^3)} - \frac{1}{2} - x + x^4 + 4x^5 + 2x^6 + 6x^7 + 6x^8, \\ \hat{A}(x) &= \sum_{n=1}^{\infty} \hat{a}(n)x^n = \frac{1+2x+4x^2}{2(1-4x^3)} + \frac{1}{2(1-2x^3)} - \frac{1-x^4}{1-x}. \end{aligned}$$

Proof. One can use corollary 2 to investigate each residue class modulo 3 separately. Per residue class, the generating function is essentially constructed from

$$(1-4x^3)^{-1} = \sum_{k=0}^{\infty} 4^k x^{3k}$$

for the weight-one numbers and additionally

$$(1-2x^3)^{-1} = \sum_{k=0}^{\infty} 2^k x^{3k}$$

for the weight-two numbers. By combining these subresults and a few exceptional terms at the beginnig of the sequence, one straightforwardly obtains the claims. □

References

[1] N. J. A. Sloane, Ed. (2008), *The On-Line Encyclopedia of Integer Sequences,* `http://www.research.att.com/~njas/sequences/`

[2] P. Lava (2009), *[seqfan] Re: SImplified description of A161373 and A161374,* `http://list.seqfan.eu/pipermail/seqfan/2009-June/001673.html`