4-17-2006

On Conway's generalization of the 3x + 1 problem

Robin M. Givens
University of Richmond

On Conway's Generalization of the $3x + 1$ Problem

Robin M. Givens
Honors thesis[1]
Department of Mathematics & Computer Science
University of Richmond

April 17, 2006

[1]Under the direction of Dr. Gary R. Greenfield

Abstract

This thesis considers a variation of the $3x+1$, or Collatz, Problem involving a function we call the Conway function. The Conway function is defined by letting $C_3(n) = 2k$ for $n = 3k$ and $C_3(n) = 4k \pm 1$ for $n = 3k \pm 1$, where n is an integer. The iterates of this function generate a few 'short' cycles, but the structural dynamics are otherwise unknown. We investigate properties of the Conway function and other related functions. We also discuss the possibility of using the Conway function to generate keys for cryptographic use based on a fast, efficient binary implemenation of the function. Questions related to the conjectured tree-like structure of the $3x + 1$ Problem and to other decidable tree-like structures are also considered.

The signatures below, by the thesis advisor, a departmental reader, and the honors coordinator for mathematics, certify that this thesis, prepared by Robin M. Givens, has been approved, as to style and content.

Gary R. Hufford 4/28/06

(thesis advisor)

(reader)

(honors coordinator)

1 Introduction

The $3x + 1$ Problem began as a mathematical question conceived by Lothar Collatz while he was a student at the University of Hamburg: Does the sequence

$$a_{n+1} = \begin{cases} a_n/2 & \text{if } a_n \text{ is even} \\ 3a_n + 1 & \text{if } a_n \text{ is odd} \end{cases}$$

yield a tree-like structure [7]? In other words, for any initial positive integer a_1, is there a positive integer n such that $a_n = 1$. The $3x + 1$ Conjecture is that there exists such an n for every a_1 [2, 4, 9, 11, 12]. The tree-like structure that would be produced is the directed graph satisfying $a_n \rightarrow a_{n+1}$ for each integer a_n (see Figure 1) [7]. Were the $3x+1$ Problem not to produce a tree-like structure, then some cycle other than (4 2 1) would exist. Thus a restatement of the $3x + 1$ Problem is: Does a cycle other than (4 2 1) exist?

The $3x + 1$ Problem is generally known as the Collatz Problem, but due to the many people who have studied it, it can also be found in the literature under the titles Syracuse, Kakutani, Hasse, and Ulam [5, 9]. One would be hard pressed to find an article on the topic that doesn't mention the famous comment by Paul Erdős, "Mathematics may not yet be ready for such problems" [5, 6, 7, 9]. For efficiency the $3x + 1$ Problem is often rewritten in the form

$$a_{n+1} = \begin{cases} a_n/2 & \text{if } a_n \text{ is even} \\ (3a_n + 1)/2 & \text{if } a_n \text{ is odd.} \end{cases}$$

When viewed this way, the sequence described by the $3x+1$ Problem is easily

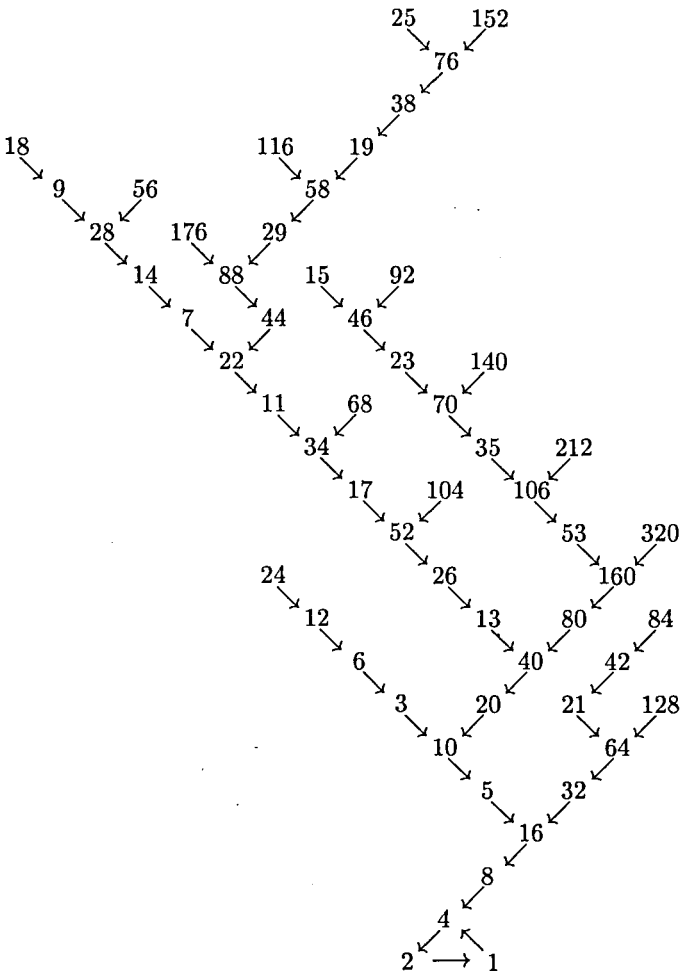

Figure 1: This directed graph shows the conjectured tree-like structure related to the $3x + 1$ Problem.

implementable on a computer [3] by observing that

$$a_{n+1} = (a_n + 1)(a_n \bmod 2) + \lfloor a_n/2 \rfloor.$$

As of February 7, 2006, the $3x + 1$ Conjecture has been verified for all $a_1 < 2^{61}$ by Oliveira e Silva [11], and has been independently verified by Roosendaal for all $a_1 < 413(2^{50})$ [12]. Proposed proofs of the conjecture have been posted online by Schorer [12], who offers money to anyone who can find a mistake in his proofs. Eliahou has shown that for $b > 0$, $c \geq 0$, and k either 301994 or 85137581, any cycle whose smallest element is larger than 2^{40} must have length $17087915b + kc$ [7]. As of September 2003, Halbeisen and Hungerbühler had taken into consideration Oliveira e Silva's verfication of the conjecture up to $3(2^{50})$ and shown that for any nontrival cycle whose minimum is greater than $3(2^{50})$, the cycle length must be at least 630 000 000 [7].

Many generalizations and related problems have arisen over the years. The '$qx + 1$' Problem asks similar questions about the sequence

$$a_{n+1} = \begin{cases} a_n/2 & \text{if } a_n \text{ is even} \\ qa_n + 1 & \text{if } a_n \text{ is odd} \end{cases}$$

for $q = 2m + 1 > 3$ [7]. Crandall has conjectured that this sequence always contains an a_1 such that 1 is not in the orbit of a_1 [7]. For instance the cycle (13 33 83) occurs for $q = 5$ [7]. Others have considered more far reaching generalizations of the sequence such as

$$a_{n+1} = \begin{cases} a_n/p & \text{if } p|a_n \\ qa_n + r & \text{otherwise} \end{cases}$$

and focused on finding p, q, and r such that the problem is solvable [7]. The $3x + 1$ Conjecture has also motivated other conjectures. For example, if the $3x + 1$ Conjecture is true, then the much more technical Weakened $3x + 1$ Conjecture is true [10], and therefore the Wild Numbers Conjecture formulated by Lagarias is true [10].

The generalization we wished to consider was formulated by Conway who first proposed defining a sequence using a *system* of equations of the form

$$a_{n+1} = b_i a_n + c_i, \text{ if } a_n \equiv i (\text{mod } p)$$

for $0 \leq i < p$ where $b_0, c_0, \ldots, b_{p-1}, c_{p-1}$ are *rational* constants chosen such that $a_n \in \mathbb{Z}$ for all n [4].

The main focus of this thesis is on an explicit instance of Conway's variation of the $3x + 1$ Problem that yields an iterative function, that is also a permutation on the integers, that we call the Conway Function. This instance was, according to Richard Guy, introduced by Conway at the 1972 Number Theory Conference [8], although it does not appear in the paper submitted by Conway to that conference [4]. The analagous question is how many cycles and infinite chains are produced by the Conway function

$$\begin{aligned} 3n &\leftrightarrow 2n, \\ 3n \pm 1 &\leftrightarrow 4n \pm 1. \end{aligned}$$

In the forward direction (i.e., $3n \to 2n$, $3n + 1 \to 4n + 1$, $3n - 1 \to 4n - 1$), this function can be seen as an instance of Conway's generalization using

$p = 3$ and the following choices for constants:

$$b_0 = \tfrac{2}{3}, \quad c_0 = 0$$

$$b_1 = \tfrac{4}{3}, \quad c_1 = -\tfrac{1}{3}$$

$$b_2 = \tfrac{4}{3}, \quad c_2 = \tfrac{1}{3}$$

The known cycles — starting with 1, 2, 3, 4, and 44 — are (1), (2 3), (4 5 7 9 6), and (44 59 79 105 70 93 62 83 111 74 99 66). However, starting with 8 and proceeding either forwards or backwards, no cycle seems to form. More specific questions then arise: Will the iterations of 8 ever form a cycle? If 8 forms an infinite chain, are there other integers that form other infinite chains? Are there finitely or infinitely many infinite chains? What other cycles exist? It is interesting to note that, according to Lagarias, in a journal dated July 1, 1932, Collatz investigated the problem

$$T(n) = \begin{cases} \frac{2}{3}n & \text{if } n \equiv 0 \pmod 3 \\ \frac{4}{3}n - \frac{1}{3} & \text{if } n \equiv 1 \pmod 3 \\ \frac{4}{3}n + \frac{1}{3} & \text{if } n \equiv 2 \pmod 3 \end{cases}$$

posing questions about the cycle structure of this permutation [9]. This is, of course, the same variation of the $3x + 1$ Problem introduced by Conway. The objective of [4] was to prove an undecidability result. Conway showed that there is no general purpose algorithm that given any system and given any n can determine if n belongs to a cycle.

This thesis investigates Conway's variation of the $3x + 1$ Problem and other Conway systems. It is organized as follows. In Section 2, we discuss known cycles for the Conway Problem and other related functions. Section

3 offers some proofs of elementary facts pertaining to the Conway Problem. Section 4 discusses a bit implementation of the Conway function for the purpose of developing a pseudorandom number generator (PRNG), and Secion 5 gives results from tests of randomness our proposed PRNG. Based on Collatz's original question, Section 6 investigates Conway systems that yield decidable tree-like structures.

2 The Existence of Cycles Problem

Definition 2.1. *Let $C_3(n)$ be the Conway function defined by*

$$C_3(n) = \begin{cases} 2k & \text{if } n = 3k \\ 4k \pm 1 & \text{if } n = 3k \pm 1. \end{cases}$$

If $m > 0$ is an integer, let $f^{(m)}(n)$ be the m-fold composition of f, $(f \circ \cdots \circ f)(n)$. A cycle of length m for an integer valued function f is a sequence of the form n, $f(n)$, $f^{(2)}(n)$, ... $f^{(m)}(n)$ such that $f^{(m)}(n) = n$. The cycle is written in the form $\big(n \quad f(n) \quad f^{(2)}(n) \quad \ldots \quad f^{(m-1)}(n)\big)$.

A Java program was written to search for small cycles produced by iterating $C_3(n)$. More precisely, we searched for cycles with length less than 100 whose smallest element was less than 240. It was unnecessary to consider any integer n such that $n < 0$ because $C_3(n)$ is an odd function (see Section 3). Thus if (2, 3) is a cycle, then $(-2, -3)$ is also a cycle. If at any point we encountered a t such that $C_3^{(t)}(n)$ equalled n, then the cycle containing n would be output and the program next considered $n + 1$. If after 100 iterations of $C_3(n)$ it was determined that n was not in a cycle of length less

```
if n % 3 == 0 {
     C3(n) := (n/3)*2 }
else if n % 3 == 3 {
     C3(n) := ((n-1)/3)*4 + 1 }
else {
     C3(n) := (n + 1/3)*4 - 1 }
```

Figure 2: Pseudocode implementation used to search for small cycles of $C_3(n)$.

than 100, then the program next considered $n+1$. Figure 2 gives pseudocode for our implementation of $C_3(n)$. After running our program no cycles were found other than those containing 1, 2, 4, and 44: (1), (2 3), (4 5 7 9 6), and (44 59 79 105 70 93 62 83 111 74 99 66).

Definition 2.2. *Define the 'reverse' Conway function $C_3^-(n)$ to be*

$$C_3^-(n) = \begin{cases} 2k & \text{if } n = 3k \\ 4k \mp 1 & \text{if } n = 3k \pm 1. \end{cases}$$

A similar program was written for 'reverse' Conway function as was written for the Conway function. The cycles found by that program are as follows: (1 $-$ 1), (2 5 9 6 4 3), (7), (14 21), (28 35 49 63 42), and one cycle of length 94 (see Appendix A).

Definition 2.3. *Define $C_q(n)$ to be the 'generalized' Conway function for odd $q > 1$ by letting*

$$C_q(n) = \begin{cases} \frac{q+1}{2}k & \text{if } n = qk \\ (q+1)k + \ell & \text{if } n = qk + \ell, \text{ for } 0 < |\ell| \leq \frac{q-1}{2} \end{cases}$$

Java programs were also written to search for the small cycles of the generalized Conway function for $q = 5$, 7, 9, 11, 13, and 15. For $q = 5$, the

small cycles are as follows: (1), (2), (3 4 5), (6 7 8 10), (12 14 17 20 12), and (36 43 52 62 74 89 107 128 154 185 111 133 160 96 15 69 83 100 60). For $q = 7$, the small cycles are as follows: (1), (2), (3), (4 5 6 7), (40 46 53 61 70), plus two cycles of length 21, and one cycle of length 83 (see Appendix A). For the small cycles for $q = 9$, 11, 13, and 15 see Appendix A.

Proposition 2.1. *For each $q = 2m + 1$, $m \geq 1$, there are always at least $(q-1)/2$ singleton cycles for $C_q(n)$ and a cycle of length $(q+1)/2$ that contains q.*

Proof. If $n = q(0)+\ell$ for $(q+1)/2 > \ell > 0$, then n is mapped to $(q+1)(0)+\ell = n$, and there are $(q-1)/2$ possibilites for this occurence. Thus there are at least $(q-1)/2$ singleton cycles. Consider $(q+1)/2 = (2q-(q-1))/2 = q(1) - (q-1)/2$ which maps to $(q+1)(1) - (q-1)/2 = q - (q-1)/2 + 1$. This, in turn, maps to $q - (q-1)/2 + 2$. By finite induction we can now iterate to $q - (q-1)/2 + (q-1)/2 = q$. Since q maps to $(q+1)/2$, we have produced a cycle of the form $((q+1)/2, (q+1)/2+1, \ldots, q)$ which contains q and has length $(q+1)/2$. □

Definition 2.4. *The stopping time [7, 9, 11, 13] of positive integer n, $\tau(n)$, is the least positive integer such that*

$$C_3^{(\tau(n))}(n) < n.$$

If no such integer exists, $\tau(n) = \infty$.

For the Conway Function, $C_3(n)$, multiple cycles exist. The least element in each cycle has infinite stopping time, but for those positive integers, n,

where the cycle structure is unknown $\tau(n)$ may be finite or infinite. For example, the cycle or chain containing 8 can never contain a negative integer (due to the fact that C_3 is an odd function - See Section 3), and integers 0 through 7 occur in other cycles, thus, since C_3 is 1-1 (see Section 3), $C_q^{(t)}(8)$ will never be less than 8 for any t and $\tau(8) = \infty$. On the other hand, $C_3(12) = 8$ so $\tau(12) = 1$ even though 12 has the same undetermined cycle status. The next positive integer for which it is unknown whether it belongs to an infinite chain or cycle is 10 (9 is in a known cycle), and $C_3^{(3)}(8) = 10$. If the sequence containing 8 is a cycle, then 10 has finite, though large, stopping time. But if the sequence containing 8 is an infinite chain, then 10 has infinite stopping time.

Definition 2.5. *Define the 'ratio' function, starting with n, $R^{(i)}(n)$ for the Conway function $C_3(n)$ by letting $R^{(0)}(n) = n$ and*

$$R^{(i+1)}(n) = \begin{cases} (2/3)R^{(i)}(n) & \text{if } C_3^{(i)}(n) = 3k \\ (4/3)R^{(i)}(n) & \text{if } C_3^{(i)}(n) = 3k \pm 1. \end{cases}$$

Terras [13] suggests calculations using $(1/2)x$ and $(3/2)x$ in place of $(1/2)x$ and $(3/2)x+(1/2)$ can give numerical estimates of $T_n(x)$, where $T_n(x)$ is the modified version of the function used for the $3x + 1$ Problem defined by

$$T_n(x) = \begin{cases} x/2 & x \equiv 0 \pmod 2 \\ (3x+1)/2 & x \equiv 1 \pmod 2. \end{cases}$$

Because of Terras' results, it is natural to consider the corresponding ratio function $R^{(i)}$ for the Conway function C_3.

In order to determine how closely $R^{(i)}(n)$ follows $C_3^{(i)}(n)$, the ratio $R^{(i)}(n)/C_3^{(i)}(n)$ was considered. For a known cycle with length x, the values $R^{(x)}(n)/C_3^{(x)}(n)$ are the same for every n in that cycle. This is due to the fact that if x_1 is equal to the number of times elements are mapped to $2/3$, and x_2 is equal to the number of times elements are mapped to $4/3$ in the cycle, then

$$\begin{aligned} R^{(x)}(n)/C_3^{(x)}(n) &= \left(\frac{2}{3}\right)^{x_1}\left(\frac{4}{3}\right)^{x_2} n/C_3^{(x)}(n) \\ &= \left(\frac{2}{3}\right)^{x_1}\left(\frac{4}{3}\right)^{x_2} n/n \\ &= \left(\frac{2}{3}\right)^{x_1}\left(\frac{4}{3}\right)^{x_2} \end{aligned}$$

which is independent of n. If we choose an n such that $C_3^{(y)}(n)$ requires y multiplications by $2/3$, then the ratio $R^{(i)}(n)/C^{(i)}(n)$ remains equal to 1 until $i > y$.

The 'infinite chains' containing 8 and 14 were also investigated; these two values are conjectured not to be in the same cycle or chain. After 100 iterations, $R^{(i)}(8)/C_3^{(i)}(8)$ converged approximately to 0.9754, and $R^{(i)}(14)/C_3^{(i)}(14)$ converged approximately to 1.0037. We have no clear explanation why both seem to converge or why $R^{(i)}(14)/C_3^{(i)}(14)$ appeared to converge to a value closer to 1. But since $C_3^{(100)}(8) = 11908$, and $C_3^{(100)}(14) = 648077$, it is quite possible that the magnitude of the value has to do with this discrepancy, and also (since both magnitudes are large) why both appear to converge.

Proposition 2.2. *The following algebraic identities hold:*

(1) If $C_3(n) = 2k$ and $C_3(m) = 2j$, then $C_3(n+m) = C_3(n) + C_3(m)$, and

$C_3(nm) = mC_3(n) = nC_3(m)$.

(2) If $C_3(n) = 2k$ *and* $C_3(m) = 4j + 1$, *then* $C_3(n + m) = 2C_3(n) + C_3(m)$, *and* $C_3(nm) = mC_3(n)$.

(3) If $C_3(n) = 2k$ *and* $C_3(m) = 4j - 1$, *then* $C_3(n + m) = 2C_3(n) + C_3(m)$, *and* $C_3(nm) = mC_3(n)$.

(4) If $C_3(n) = 4k+1$ *and* $C_3(m) = 4j+1$, *then* $C_3(n+m) = C_3(n)+C_3(m)+1$, *and* $C_3(nm) = m(C_3(n) - 1) + C_3(m) = n(C_3(m) - 1) + C_3(n)$.

(5) If $C_3(n) = 4k + 1$ *and* $C_3(m) = 4j - 1$, *then* $C_3(n + m) = \frac{1}{2}C_3(n) + \frac{1}{2}C_3(m)$, *and* $C_3(nm) = m(C_3(n)-1)+C_3(m) = n(C_3(m)+1)-C_3(n)$.

(6) If $C_3(n) = 4k-1$ *and* $C_3(m) = 4j+1$, *then* $C_3(n+m) = C_3(n)+C_3(m)-1$, *and* $C_3(nm) = m(C_3(n) + 1) - C_3(m) = n(C_3(m) + 1) - C_3(n)$.

Proof. We give the flavor of the proof by verifying (1) and (2); the other cases follow similarly.

(1) If $C_3(n) = 2k$ and $C_3(m) = 2j$, write $n = 3k$ and $m = 3j$. We have $n + m = 3(k + j)$ and $nm = 3(3kj)$, so $C_3(n + m) = 2(k + j) = 2k + 2j = C_3(n) + C_3(m)$, and $C_3(nm) = 2(3kj) = (3k)(2j) = (3j)(2k) = mC_3(n) = nC_3(m)$.

(2) If $C_3(n) = 2k$ and $C_3(m) = 4j + 1$ write $n = 3k$ and $m = 3j + 1$. From $n + m = 3(k + j) + 1$ and $nm = 3k(3j + 1)$, we obtain $C_3(n + m) =$

$4(k+j)+1 = 4k+4j+1 = 2C_3(n)+C_3(m)$ and $C_3(nm) = 2k(3j+1) = mC_3(n)$.

□

3 Some Proofs of Elementary Facts

Theorem 3.1. *$C_q(n)$ is an odd function.*

Proof. For $n \in \mathbb{Z}$ let $n = qk + \ell$, where $0 \leq |\ell| \leq \frac{q-1}{2}$:

Case 1. If $\ell = 0$, then $-n = q(-k)$. Thus

$$\begin{aligned} C_q(-n) &= C_q(q(-k)) \\ &= \frac{q+1}{2} - k \\ &= -C_q(n). \end{aligned}$$

Case 2. If $\ell \neq 0$, then $-n = q(-k) - \ell$. Thus

$$\begin{aligned} C_q(-n) &= C_q(q(-k) - \ell) \\ &= (q+1)(-k) - \ell \\ &= -((q+1)k + \ell) \\ &= -C_q(n). \end{aligned}$$

□

Theorem 3.2. *The function $C_q(n)$ is one-to-one.*

Proof. For $n_1, n_2 \in \mathbb{Z}$ d suppose $n_1 = qk_1 + \ell_1$, $n_2 = qk_2 + \ell_2$, where $k_1, k_2 \in \mathbb{Z}$ and $0 \leq |\ell_1, \ell_2| \leq \frac{q-1}{2}$. There are three cases to consider.

Case 1. Suppose $\ell_1 = \ell_2 = 0$, and $C_q(n_1) = C_q(n_2)$. Then $C_q(qk_1) = C_q(qk_2)$ implies $((q+1)/2)k_1 = ((q+1)/2)k_2$. Thus $k_1 = k_2$ and $n_1 = n_2$.

Case 2. Suppose $\ell_1 \neq 0$, $\ell_2 \neq 0$, and $C_q(n_1) = C_q(n_2)$. Then $C_q(qk_1 + \ell_1) = C_q(qk_2 + \ell_2)$ implies $(q+1)k_1 + \ell_1 = (q+1)k_2 + \ell_2$. Thus $\ell_1 = \ell_2 \bmod (q+1)$ which implies $\ell_1 = \ell_2$ since $0 < |\ell_1, \ell_2| < (q+1)$. Therefore $(q+1)k_1 = (q+1)k_2$ implies $k_1 = k_2$, and thus $n_1 = n_2$.

Case 3. Without loss of generality suppose $\ell_1 \neq 0$, $\ell_2 = 0$, and $C_q(n_1) = C_q(n_2)$. Then $C_q(qk_1 + \ell_1) = C_q(qk_2)$ implies $(q+1)k_1 + \ell_1 = ((q+1)/2)k_2$. Thus $\ell_1 = 0 \bmod ((q+1)/2)$. But, by assumption, $0 < |\ell_1| < (q+1)/2$. This contradiction completes the proof of this case.

□

Theorem 3.3. *The function $C_q(n)$ is onto.*

Proof. For every $n \in \mathbb{Z}$, $n \equiv 0, 1, \ldots$ or $q \bmod (q+1)$. Since q is odd, $2|(q+1)$. Again, there are three cases to consider

Case 1. If $n \equiv 0 \bmod (q+1)$ or $n \equiv (q+1)/2 \bmod (q+1)$ then $n = ((q+1)/2)k$ for some $k \in \mathbb{Z}$. This shows $n = C_q(qk)$.

Case 2. If $n \equiv \ell \mod (q+1)$ for $0 < \ell < (q+1)/2$, then $n = (q+1)k + \ell$ for some $k \in \mathbb{Z}$. Therefore $n = C_q(qk + \ell)$.

Case 3. If $n \equiv \ell \mod (q+1)$ for $(q+1)/2 < \ell < (q+1)$ then $n = (q+1)k + \ell$ for some $k \in \mathbb{Z}$. Thus $n = (q+1)(k+1) + \ell - (q+1)$ with $-(q+1)/2 < \ell - (q+1) < 0$. This gives $n = C_q(q(k+1) + \ell - (q+1))$.

□

Theorem 3.4. *For $q > 3$, $C_q(x)$ does not yield a 2-cycle.*

Proof. Suppose n produces the 2-cycle $n \mapsto C_q(n) \mapsto C_q(C_q(n)) = n$.

Case 1. If $q \nmid n$ and $q \nmid C_q(n)$ then $n \leq C_q(n) \leq C_q(C_q(n))$ so either $n \neq C_q(C_q(n))$ or $n = C_q(n) = C_q(C_q(n))$ which is a 1-cycle.

Case 2. If $q|n$ and $q|C_q(n)$ then $n > C_q(n) > C_q(C_q(n))$ so $n \neq C_q(C_q(n))$.

Case 3. Since a 2 cycle can be written as $(a\ b)$ or $(b\ a)$, without loss of generality suppose $q \nmid n$ and $q|C_q(n)$. We compute

$$C_q(C_q(n)) = C_q\left(q \cdot \frac{C_q(n)}{q}\right) = \left(\frac{q+1}{2}\right)\frac{C_q(n)}{q}$$

and consider n represented as

$$n = qk + \ell, \text{ with } 0 < |\ell| \leq \frac{q-1}{2}.$$

If $n = C_q(C_q(n))$, then

$$\begin{aligned} qk+\ell &= \left(\frac{q+1}{2}\right)\frac{C_q(n)}{q} \\ &= \left(\frac{q+1}{2}\right)\frac{C_q(qk+\ell)}{q} \\ &= \left(\frac{q+1}{2}\right)\frac{(q+1)k+\ell}{q} \\ &= \frac{(q+1)^2k+(q+1)\ell}{2q}. \end{aligned}$$

Thus

$$\begin{aligned} 2q^2k+2q\ell &= (q+1)^2k+(q+1)\ell \\ &= (q^2+2q+1)k+(q+1)\ell \\ &= q^2k+2qk+k+q\ell+\ell. \end{aligned}$$

This implies $q^2k-2qk-k = \ell-q\ell$ and $k(q^2-2q-1) = (1-q)\ell$. Thus

$$|k||q^2-2q-1| = |1-q||\ell| < |1-q|\frac{(q+1)}{2}.$$

Notice that since $q \geq 5$, $q+1 = |q+1|$ and $|q^2-11| > \frac{|q^2-1|}{2}$. Therefore

$$\begin{aligned} 2|k| &< \frac{|1-q||q+1|}{|q^2-2q-1|} \\ &= \frac{|1-q^2|}{|q^2-2q-1|} \\ &= \frac{|q^2-1|}{q^2-2q-1|} \\ &\leq \frac{|q^2-1|}{|q^2-2(5)-1|} \\ &= \frac{|q^2-1|}{|q^2-11|} \end{aligned}$$

$$< \frac{|q^2-1|}{\left|\frac{q^2-1}{2}\right|}$$
$$= 2$$

This shows $|k| < 1$. Since $k \in \mathbb{Z}$, $k = 0$. This implies that $n = \ell$ which sasys $C_q(n) = \ell$, where $0 < \ell < q$. The fact that $q|C_q(n)$ provides a contradiction and completes the proof.

□

In view of the preceding *proof* we are led to make the following conjecture. One reason this conjecture is important is because it tells us that an *algebraic* proof about the nonexistence of cycles of length $t > 2$ must consider q^t cases!

The Remainder Conjecture. *For any sequence of remainders* $r_0, r_1, \cdots r_m$ *such that* $0 \leq |r_i| < (q+1)/2$, *there exists an* n *such that* $n = qk_0 + r_0$ *for some* k_0 *and* $k_1, \ldots k_m$ *such that* $C_q^i(n) = qk_i + r_i$ *for* $0 < i \leq m$.

We prove the following special cases of the Remainder Conjecture, $r_i \equiv 0$ for all i and $r_i \equiv \ell$, $0 < |\ell| \leq (q-1)/2$, for all i.

Proposition 3.1. *For* $r_i \equiv 0$ *for all* i, *the Remainder Conjecture is true.*

Proof. Let $r_0, r_1, \ldots r_{t-1} = 0$. Consider $n = q^t$. Since $n \equiv 0 \pmod q$, $r_0 = 0$. Now,

$$\begin{aligned}
C_q^1(n) &= \tfrac{q+1}{2q} q^t = \left(\tfrac{q+1}{2}\right) q^{t-1} \equiv 0 \pmod q \quad \Rightarrow r_1 = 0, \\
C_q^2(n) &= \left(\tfrac{q+1}{2}\right)^2 q^{t-2} \equiv 0 \pmod q \quad \Rightarrow r_2 = 0, \\
&\vdots \\
C_q^{t-1}(n) &= \left(\tfrac{q+1}{2}\right)^{t-1} q \equiv 0 \pmod q \quad \Rightarrow r_{t-1} = 0.
\end{aligned}$$

□

Proposition 3.2. *For $r_i \equiv \ell$, $0 < |\ell| \leq (q-1)/2$, for all i, the Remainder Conjecture is true.*

Proof. Let $r_0, r_1, \ldots r_{t-1} = \ell$ for some ℓ $0 < |\ell| \leq (q-1)/2$. Consider $n = q^t + \ell$. Since $n \equiv \ell \pmod{q}$, $r_0 = \ell$. Now,

$$\begin{aligned} C_q^1(n) &= (q+1)q^{t-1} + \ell \equiv \ell \pmod{q} \quad \Rightarrow r_1 = \ell, \\ C_q^2(n) &= (q+1)^2 q^{t-2} + \ell \equiv \ell \pmod{q} \quad \Rightarrow r_2 = \ell, \\ &\vdots \\ C_q^{t-1}(n) &= (q+1)^{t-1} q + \ell \equiv \ell \pmod{q} \quad \Rightarrow r_{t-1} = \ell. \end{aligned}$$

□

4 A Fast, Efficient Implementation of $C_3(n)$ using Bit Operations

Many cryptographic algorithms require a key consisting of a short sequence of random bits [14]. A simple example is the one-time pad cryptographic system, where a random bit sequence is first generated and then added bit by bit to the binary message meant to be sent [14]. The Blum-Blum-Shub (BBS) pseudorandom bit generator is the gold standard for pseudorandom bit generators for cryptographic purposes, but the calculations involved are slow [14]. We are interested in determining if the Conway function $C_3(n)$ can provide an equally good but faster, more efficient bit generator.

A binary implementation of $C_3(n)$ was developed to examine the possibility that this function could serve as a short, or "burst," binary pseudorandom number generator (PRNG) for purposes such as cryptographic key generation. Bostwick [3] divised a binary implementation of the Collatz function

using a cellular automaton. In order for the Conway function, $C_3(n)$, to be implemented using efficient binary operators, a method for dividing by 3 in binary must be found. Artzy et al. [1] have proposed a method for fast division of binary numbers by constant divisors. Their general purpose method was altered to specifically handle the case of division by 3, and their proof was adapted to show that the division by 3 algorithm was correct.

Algorithm 4.1. A Division by 3 Algorithm Using Bit Operations [1]

Using two ℓ-bit registers R and T, and setting $k_\ell = max\{\lceil \log_2(\ell) - 1 \rceil, 0\}$, if R is divisible by 3, then $R/3$ can be obtained by the algorithm in Figure 3.

```
for j = 1 to k_ℓ {
    T := R << 2^j
    R := R + T }
R := twos complement of R
```

Figure 3: Bit operation division by 3 algorithm.

The following proof of the Division by 3 Algorithm was modelled after the proof in [1].

For bit length $\ell \in \mathbb{N}$ and $x \in \mathbb{Z}_{2^\ell}$ let

$$k_\ell = \max\{\lceil \log_2(\ell) - 1 \rceil, 0\},$$

$$T_\ell = \prod_{i=1}^{k_\ell} \left(2^{2^i} + 1\right), \text{ and}$$

$$f_\ell(x) = -T_\ell \pmod{2^\ell}.$$

Observe that $f_\ell(x)$ computes the result of the division by 3 algorithm. The proof requires the following lemmas.

Lemma 4.1. $2^{k_\ell+1} \geq \ell$.

Proof. Let $x \in \mathbb{N}$ satisfy $2^{x-1} < \ell \leq 2^x$. Since $k_\ell = x-1$, $2^{k_\ell+1} = 2^x \geq \ell$. □

Lemma 4.2. $3T_\ell = 2^{2^{k_\ell+1}} - 1$.

Proof. This proof is by induction on k_ℓ. For $k_\ell = 1$,

$$\begin{aligned} 3T_\ell &= 3\left(2^{2^1}+1\right) \\ &= 15 \\ &= 2^{2^2}-1 \\ &= 2^{2^{k_\ell+1}}-1. \end{aligned}$$

Suppose $3T_\ell = 2^{2^{k_\ell+1}} - 1$ for $k_\ell = k$. For $k_\ell = k+1$,

$$\begin{aligned} 3T_\ell &= \prod_{i=1}^{k+1}\left(2^{2^i}+1\right) \\ &= \prod_{i=1}^{k}\left(2^{2^i}+1\right)\left(2^{2^{k+1}}+1\right) \\ &= \left(2^{2^{k+1}}-1\right)\left(2^{2^{k+1}}+1\right) \text{ by hypothesis} \\ &= 2^{2^{k+2}}-1 \\ &= 2^{2^{k_\ell+1}}-1. \end{aligned}$$

Thus by induction $3T_\ell = 2^{2^{k_\ell+1}} - 1$. □

Theorem 4.1. *If $x = 3q \in \mathbb{Z}_{2^\ell}$, then $f_\ell(x) = q$, i.e., the division by 3 algorithm correctly divides by 3 when x is divisible by 3.*

Proof.

$$\begin{aligned} f_\ell(3q) &= -T_\ell 3q \pmod{2^\ell} \\ &= -\left(2^{2k_\ell+1}-1\right) q \pmod{2^\ell} \\ &= -(-1)q \pmod{2^\ell} \\ &= q \end{aligned}$$

The second equality follows from Lemma 4.2 and the third from Lemma 4.1 since $2^{2k_\ell+1} \equiv 0 \pmod{2^\ell}$. □

Once a method for dividing by 3 was finalized, a method to determine if a binary number was congruent to 0, 1, or 2 (mod 3) was implemented. Pseudocode describing this method for a binary integer R with ℓ bits as input can be found in Figure 4. The reason this method works is based on the fact that $2^k \equiv 1 (\text{mod } 3)$ if $k \geq 0$ is even, and $2^k \equiv 2 (\text{mod } 3)$ if $k > 0$ is odd.

```
int divisibility(R) {
     int p := -1;
     int sum := 0;
     for i := 1 to ℓ {
          p := 0 - p
          sum := (sum + (p * i^th LSB of R) % 3 }
     return sum }
```

Figure 4: Method used to determine the congruence class mod 3 of a positive integer represented in binary. [Note: LSB stands for least significant bit.]

When methods for dividing by 3 and determining the divisibility of a binary number were completed, the method describing $C_3(n)$ could be imple-

mented. For an ℓ-bit binary number R, the pseudocode in Figure 5 outlines the $C_3(n)$ calculation.

```
int d = divisibility(R)
if d == 1 {
     R := R - 1 }
else if d == 2 {
     R := R + 1 }
divideBy3(R)
if d == 1 {
     R << 2 //multiply by 4
     R := R + 1 }
else if d == 2 {
     R << 2 //multiply by 4
     R := R - 1 }
else {
     R << 1 } //multiply by 2
```

Figure 5: $C_3(n)$ pseudocode for an integer n represented in binary form.

5 Tests of Randomness

Tests of randomness were performed based on the values of the least significant bit (LSB) of the binary integers obtained from a sequence of iterations of our $C_3(n)$ binary implementation. We note that in order for the binary implemenation to actually be used as a PRNG key generator, the Remainder Conjecture is assumed true, and we assume that infinite chains exist. Our generator would be especially unique and unusual because the probability of obtaining a zero is 1/3 and the probability of obtaining a one is 2/3 as we now prove.

Theorem 5.1. *With assumptions as above, in any arbitrary pseudorandomly generated sequence using the fucntion $C_3(n)$ as a generator, there is a 1/3 probability that a 0 will occur in the least significant bit of the output, and a 2/3 probability that a 1 will occur in the least significant bit of the output.*

Proof. The probability that a randomly selected integer has remainder 0 modulo 3 is 1/3. The probability that a randomly selected integer has remainder 1 or 2 modulo 3 is 2/3. Thus the probability that a randomly selected integer n maps to $2k \equiv 0(\text{mod } 2)$ for some k is 1/3, and the probability that n maps to $4k \pm 1 \equiv 1(\text{mod } 2)$ for some k is 2/3. In other words, the expectation is that in any "run" determined by iterating the $C_3(n)$ function, the empirical probability of a LSB being 0 is 1/3, and the empirical probability of a resulting LSB being 1 is 2/3. □

Using a randomly generated 128-bit binary seed, x iterations of the $C_3(n)$ binary implementation were used to create an equivalent length bit string, $b_1 b_2 \ldots b_x$, by extracting the least significant bit after each iteration. These bit strings were then tested using the following tests:

- Frequencies of Zeros and Ones Distribution Test: From sequences of $b_1 b_2 \ldots b_n$ of length $n = 600$, the frequencies of zeros and ones were calculated. From three samples of size 600 we obtained 401, 404, and 405 ones. The respective tests for the proportion of ones gave observed levels of significance of .928, .723, and .6672 ($z = 0.087, 0.346, 0.433$). Therefore, we did not find any statistical evidence to indicate

that "short" randomly generated sequences would not fit the expected proportioned frequencies.

- Bit Sequences of Length Two Distribution Test: From sequences of $b_1 b_2 \ldots b_{2n}$ of length $2n$, by taking bits two at a time we obtained samples of the form $x_1, x_2, \ldots, x_n$ where each x_i was either 00, 01, 10, 11. A Chi-Square Goodness of Fit Test using probabilities 1/9, 2/9, 2/9, and 4/9 respectively for these four possibilities was performed. From three samples of size $n = 540$ we obtained observed levels of significance .116, .329, and .549 ($\chi^2 = 5.913, 3.438, 2.113$). Thus, we did not find any statistical evidence to indicate that "short" randomly generated sequences would not fit our theoretical distribution.

- Overlapping Sequences of Length Four Distribution Test: Similarly, from a bit sequence $b_1 b_2 \ldots b_{n+3}$ of length $n+3$ we obtained a sample of n *overlapping* bit sequences of length four by setting $x_i = b_i b_{i+1} b_{i+2} b_{i+3}$. A Chi-Square Goodness of Fit Test was performed using proportions 1/81, 2/9, 4/9, 8/9, and 16/81 for those sequences with zero, one, two, three, and four ones respectively. Three samples of size $n = 1543$ produced observed levels of significance .921, .209, and .015 ($\chi^2 = 8.063, 19.114, 29.254$). In only one instance did a "short" randomly generated sequence fail match the hypothesized distribution. For that sample the sequences 1001 and 1010 were *under-represented.*

- Runs of Zeros Distribution Test: From a bit string $b_1 b_2 \ldots b_n$ we let x_1, x_2, x_3, x_4 be the number of distinct *isolated* substrings with 1, 2, 3, and 4 consecutive zeros. Since the probability of obtaining a run of zeros with $k+1$ consecutive zeros is one-third as likely as obtaining a run with k consecutive zeros, we performed a Chi-Square Goodness of Fit Test using the ratios $27:9:3:1$. Using x_1, x_2, x_3, x_4 as the observed cell counts, from three random sequences of length $n = 1024$ we obtained samples of size 207, 222, and 228 which yielded observed levels of significance .896, .909, and .596 ($\chi^2 = .601, .543, 1.890$). In no case was there statistical evidence to indicate that runs of zeros from "short" randomly generated sequences did not fit these predicted ratios.

6 Tree Structures for Conway Systems

One of the questions posed by Collatz about the $3x+1$ Problem was whether or not it formed a tree-like structure (see Figure 1). The answer to this question rests on the validity of the $3x+1$ Conjecture.

We were motivated to consider other tree structures by a mistakenly written reverse Conway function that we will call the 'Reverse 3' function or $V(n)$. It is defined as

$$V(n) = \begin{cases} 2k & \text{if } n = 3k \\ 3k \mp 1 & \text{if } n = 3k \pm 1. \end{cases}$$

This function creates cycles of the form $(3k-1 \quad 3k+1)$ and finite chains of

the form

$$3^m k \mapsto 2(3^{m-1})k \mapsto 2^2 3^{m-2} k \mapsto \ldots \mapsto 2^m k$$

so every even number belongs to a chain (see Figure 6).

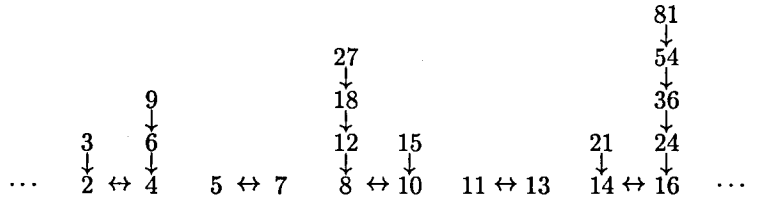

Figure 6: The 'tree' structure of the reverse 3 function $V(n)$.

Since the structure of the Reverse 3 function is decidable, we were interested in finding other examples of functions in which the cycle dynamics could be determined. First, other rearrangements of the the terms $3n$, $4n+1$, $4n-1$ of the Conway function were considered. An example of one of the five possible variants of the Conway function $C_3(n)$ we considered is described by

$$\begin{aligned} 3n &\mapsto 4n-1 \\ 3n+1 &\mapsto 2n \\ 3n-1 &\mapsto 4n+1. \end{aligned}$$

Every one of the five rearrangements of the Conway function showed similar dynamics to the original Conway function: a few short cycles and seqeuences with presumed chain-like properties.

The next system of equations we considered was defined by letting $p > 2$ be prime and letting

$$U(n) = \begin{cases} n/p & \text{if } p|n \\ in & \text{if } n \equiv i(\text{mod } p). \end{cases}$$

This system obeys the rules for Conway's generalization of the $3x+1$ Problem by letting $b_0 = 1/p$, $b_i = i$ for $0 < i < p$, and $c_i = 0$ for $0 \leq i < p$. The dynamics of this system seems to be decidable, at least for small p. For $p = 3$ or 5, the system is structured as an infinite number of finite chains that each iterate to an integer congruent to 1 modulo p which then cycles with itself (see Figures 7 and 8).

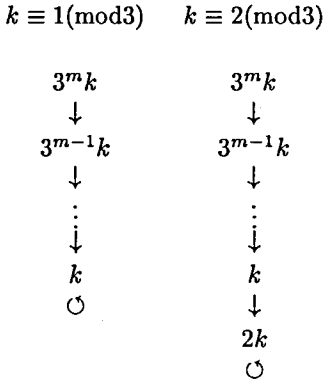

Figure 7: 'Tree' structure for $U(n)$ with $p = 3$, showing the two types of finite chains that descend to cycles of length one, i.e., loops.

When $p = 7$, however, the only finite chains are those that eventually reach 1 or $p - 1(\text{mod } p)$ (see Figure 9). It appears that infinite chains also occur whenever $p > 7$.

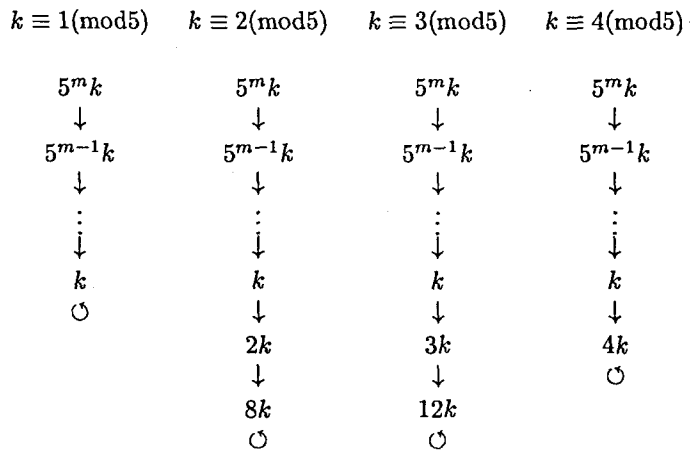

Figure 8: 'Tree' structure for $U(n)$ with $p = 5$.

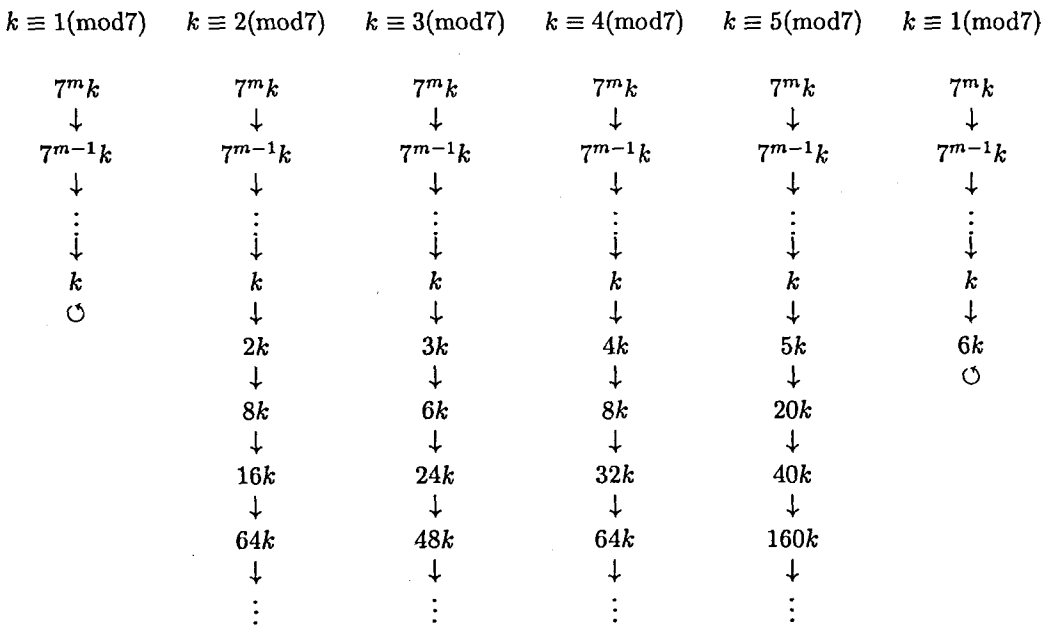

Figure 9: 'Tree' structure for $U(n)$ with $p = 7$.

7 Conclusions

In view of our work, Erdős would still probably say mathematics may not be ready for problems like these. We were unable to find any new cycles for the Conway Function $C_3(n)$, but we were able to show that for $C_q(n)$ with $q > 3$, no cycles of length 2 exist. If the Remainder Conjecture is true, then determining how many cycles Conway's function $C_3(n)$ provides will probably be just as hard as settling the $3x+1$ Conjecture, since determining whether $C_q(n)$ has cycles of length t requires q^t cases.

Discovering other Conway systems that have decidable tree-like structures may provide insight into the structure and dynamics of other systems including the Conway function. Both the Reverse 3 function $V(n)$ and the function $U(n)$ with $p = 3, 5, 7$ had deciable tree structures. Future research could look into other functions with decidable structure as well as continuing to explore the structure of Conway's function $C_3(n)$.

With further research and testing, our binary implementation of Conway's function may provide an easy to implement and efficient key generator. Our generator has already passed several statistical tests, and with additional testing may prove to be of further interest. Given these possibilities, we believe iterated sequence problems are likely to continue to be of interest for many years to come.

References

[1] E. Artzy, J. Hinds, and H. Saal, A fast division technique for constant divisors, *Communications of the ACM*, Vol. 19, No. 2, 1976, 98 – 101.

[2] J. Borwein and D. Bailey, *Mathematics by Experiemnt: Plausible Reasoning in the* 21^{st} *Century*, A. K. Peters, Natick, MA, 2004, 75.

[3] B. Bostwick, On two mechanisims related to the "$3n+1$" problem, *Midstate Conference for Undergraduate Research in Computer Science and Mathematics*, 2003. http://www.denison.edu/mathsci/mcurcsm2003/papers/3nplus1_v2.pdf.

[4] J. Conway, Unpredictable iterations, *Procedures of the 1972 Number Theory Conference*, University of Colorado, 1972, 49 – 52.

[5] R. Guy, Don't try to solve these problems!, *American Mathematical Monthly*, Vol. 90, No. 1, 1983, 35 – 41.

[6] R. Guy, Nothing's new in number theory, *American Mathematical Monthly*, Vol. 105, No. 10, 1998, 951 – 954.

[7] R. Guy, *Unsolved Problems in Number Theory*, Springer-verlag, 2004, 330 – 336.

[8] R. Guy, private communication.

[9] J. Lagarias, The $3x+1$ problem and its generalizations, *American Mathematical Monthly*, Vol. 92, No. 1, 1985, 3 – 23.

[10] J. Lagarias, Wild and wooley numbers, *American Mathematical Monthly*, Vol. 113, No. 2, 2006, 97 – 108.

[11] T. Oliveira e Silva, 3x+1 conjecture verification and results, Instituto de Egenharia Electronica e Telematica de Aveiro (IEETA), 2006. http://www.ieeta.pt/ tos/3x+1.html.

[12] E. Roosendaal, On the $3x + 1$ problem, 2006. http://www.ericr.nl/wondrous/index.html.

[13] R. Terras, A stopping time problem on the positive integers, *Acta Arithmetica XXX*, 1976, 241 – 252.

[14] W. Trappe and L. C. Washington, *Introduction to Cryptography with Coding Theory*, Prentice Hall, 2002, 38 – 42.

Appendix A: Some Additional Examples of Small Cycles of $C_3^-(n)$ and $C_q(n)$

$C_3^-(n)$

(142 187 247 327 218 293 393 262 347 465 310 411 274 363 242 325
431 577 767 1025 1369 1823 2433 1622 2165 2889 1926 1284 856
1139 1521 1014 676 899 1201 1599 1066 1419 946 1259 1681 2239
2983 3975 2650 3531 2354 3141 2094 1396 1859 2481 1654 2203
2935 3911 5217 3478 4635 3090 2060 2749 3663 2442 1628
2173 2895 1930 2571 1714 2283 1522 2027 2705 3609 2406 1604
2141 2857 3807 2538 1692 1128 752 1005 670 891 594 396 264
176 237 158 213)

$C_7(n)$

(20 23 26 30 34 39 45 51 58 66 75 86 98 56 32 37 42 24 27 31 35)
(8 9 10 11 13 15 17 19 22 25 29 33 38 43 49 28 16 18 21 12 14)
(68 78 89 102 117 134 153 175 100 114 130 149 170 194 222 254 290
331 378 216 247 282 322 184 210 120 137 157 179 205 234 267 305
349 399 228 261 298 341 390 446 510 583 666 761 870 994 568 649
742 424 485 554 633 723 826 472 539 308 176 201 230 263 301 172
197 225 257 294 168 96 110 126 72 82 94 107 122 139 159 182 104
119)

$C_9(n)$

(1) (2) (3) (4) (5 6 7 8 9) (10 11 12 13 14 16 18)
(15 17 19 21 23 26 29 32 36 20 22 24 27)
(60 67 74 82 91 101 112 124 138 153 85 94 104 116 129 143 159 177 197
219 243 135 75 83 92 102 113 126 70 78 87 97 108)
(110 122 136 151 168 187 208 231 257 286 318 353 392 436 484 538 598
664 738 410 456 507 563 626 696 773 859 954 530 589 654 727 808
898 998 1109 1232 1369 1521 845 939 1043 1159 1288 1431 795 883
981 545 606 673 748 831 923 1026 570 633 703 781 868 964 1071 595
661 734 816 907 1008 560 622 691 768 853 948 1053 585 325 361 401
446 496 551 612 340 378 210 233 259 288 160 178 198)

$C_{11}(n)$

(1) (2) (3) (4) (5) (6 7 8 9 10 11) (24 26 28 31 34 37 40 44)
(48 52 57 62 68 74 81 88) (54 59 64 70 76 83 91 99)
(96 105 115 125 136 148 161 176)
(114 124 135 147 160 175 191 208 227 248 271 296 323 352 192 209)
(126 137 149 163 178 194 212 231)
(180 196 214 233 254 277 302 329 359 392 428 467 509 555 605 330)
(186 203 221 241 263 287 313 341)
(203 221 241 263 287 313 341 186)

$C_{13}(n)$

(1) (2) (3) (4) (5) (6) (7 8 9 10 11 12 13)
(14 15 16 17 18 19 20 22 24 26) (21 23 25 27 29 31 33 36 39)
(28 30 32 34 37 40 43 46 50 54 58 62 67 72 78 42 45 48 52)
(35 38 41 44 47 51 55 59 64 69 74 80 86 93 100 108 116 125 135 145
156 84 90 97 104 56 60 65)
(238 256 276 297 320 345 372 401 432 465 501 540 582 627 675 727 783
843 908 978 1053 567 611 329 354 381 410 442)

$C_{15}(n)$

(1) (2) (3) (4) (5) (6) (7) (8 9 10 11 12 13 14 15)
(16 17 18 19 20 21 22 23 25 27 29 31 33 35 37 39 42 45 24 26 28 30)
(32 34 36 38 41 44 47 50 53 57 61 65 69 74 79 84 90 48 51 54 58 62 66
70 75 40 43 46 49 52 55 59 63 67 71 76 81 86 92 98 105 56 60)
(176 188 201 214 228 243 259 276 294 314 335 357 381 406 433 462 493
526 561 598 638 681 726 774 826 881 940 1003 1070 1141 1217 1298
1385 1477 1575 840 448 478 510 272 290 309 330)