A DIRECT PROOF OF A RESULT OF THUE

James D. Currie

In an undergraduate course in automata theory, the following result, left as an exercise in [1], (p. 56), was given as a challenge problem:

> There are arbitrarily long sequences of 1's and 0's in which no subsequence appears three times consecutively.

The earliest proof of the result was by Thue in [2]. A brief survey of the many generalizations of this problem is found in [3]. Applications exist in number theory, (via Burnside's problem and diophantine equations), surface flow theory, and in the theory of semigroups.

The following proof uses recursive properties of Thue's sequence. It is thus more direct than the original, which used properties of sequences of three symbols. It was discovered in response to the problem in [1].

Notation: To represent a 1 or 0, we will use the symbols a, b, a_i, u_i, v_i. For particular sequences, we will use w, u, v, z. For the sequence consisting of v, repeated n times consecutively, we use v^n. Write S for the set of sequences of 1's and 0's of finite length and write T for the set of sequences of 1's and 0's of finite, even length.

Define the substitution $f:S \rightarrow T$ by

$$f(0) = 01$$
$$f(1) = 10$$
$$f(w) = f(a_1)f(a_2)\dots f(a_m)$$

for $w = a_1a_2\dots a_m$.

Then the definition

(1) $$w_0 = 0$$
$$w_n = f(w_{n-1}), \quad n > 0.$$

gives rise to a sequence of sequences 0, 01, 0110, 01101001, ... known as Thue's sequence on two symbols. We will show that for all n, no subsequence of w_n is repeated three times in a row.

Define $g:T \to S$, a left inverse of f, by

$$g(ab) = a$$
$$g(w) = g(a_1a_2)g(a_3a_4)\ldots g(a_{2m-1}a_{2m})$$

for $w = a_1a_2\ldots a_{2m}$.

Since $g(f(w)) = w$, for all w in S, we have

(2) $$w_{n-1} = g(w_n), \quad n > 0.$$

Another recursive relation on the w_n arises as follows: Define the function $h: T \to S$ by

$$h(ab) = b$$

$$h(w) = a_1h(a_2a_3)h(a_4a_5)\ldots h(a_{2m-2}a_{2m-1}) \text{ for } w = a_1a_2\ldots a_{2m} .$$

For example,

$$h(01101001) = 0h(11)h(01)h(00) = 0110$$
$$h(0110) = 0h(11) = 01$$
$$h(01) = 0$$

In fact, an easy induction shows that

(3) $$w_{n-1} = h(w_n), \quad n > 0 .$$

Once we have disposed of these preliminaries, the result follows quickly from the following lemma.

LEMMA. *If* w_n *can be written as* uv^3z, *where* u,v,z *are words in* S, *then the length of* v *is even.*

Proof of Lemma. Suppose $w_n = uv^3z$ where v has odd length k. Certainly then, $n > 2$. From (1) w_n is of even length, so write

$$w_m = (a_1a_2)(a_3a_4)\ldots(a_{2m-1}a_{2m}) \; . \tag{4}$$

But by definition, $w_n = f(w_{n-1})$, and each pair $(a_{2i-1}a_{2i})$ is either $f(0) = 01$, or $f(1) = 10$. In either case, we know that $a_{2i-1} + a_{2i} = 1$, $0 < i < m+1$.

Let $v = v_1v_2\ldots v_k$. Then $w_n = uv_1v_2\ldots v_kv_1v_2\ldots v_kv_1v_2\ldots v_kz$. As v has odd length, a section v^2 of w_n aligns with the divisions of (4).

We thus write v^2 as

$$(v_1v_2)(v_3v_4)\ldots(v_kv_1)(v_2v_3)(v_4v_5)\ldots(v_{k-1}v_k)$$

such that the sum of each bracketed pair is one.

But then

$$2(v_1+v_2+\ldots+v_k) = (v_1+v_2) + (v_3+v_4) + \ldots + (v_k+v_1)$$

$$+ (v_2+v_3) + \ldots + (v_{k-1}+v_k) = 1 + 1 + \ldots + 1 \text{ (k times)} = k.$$

This is a contradiction, for k is odd.

Thus if $w_n = uv^3z$, we require that v be of even length.

But now the result follows easily:

Suppose $w_n = uv^3z$, with length v greater than zero. Pick the least such n.

Let $u = u_1u_2\ldots u_s$, $v = v_1v_2\ldots v_{2m}$.

Case 1) s odd. Then

$$w_{n-1} = h(w_n) = h(uv^3z) = u_1h(u_2u_3)h(u_4u_5)\ldots h(u_{s-1}u_s)[h(v_1v_2)h(v_3v_4)$$

$$\ldots h(v_{2m-1}v_{2m})h(v_1v_2)\ldots h(v_{2m-1}v_{2m})h(v_1v_2)\ldots h(v_{2m-1}v_{2m})]h(z), \text{ by (3)}.$$

Then w_{n-1} contains a triple repetition, contradicting our choice of n.

Case 2) s even. By (2)

$$w_{n-1} = g(w_n) = g(u_1u_2)g(u_3u_4)\ldots g(u_{s-1}u_s)g(v)g(v)g(v)g(z) ,$$

again contradicting our choice of n.

REFERENCES

[1] J. E. Hopcraft and J. D. Ullman, *Introduction to Automata Theory, Languages and Computation*, Addison-Wesley, Don Mills (1979).

[2] A. Thue, *'Über unendliche Zeichenreihen'*, Norske Vid Selske. Skr., I Mat. Nat. Kl., Christiana, *7* (1906), 1-22.

[3] Dwight R. Bean, Andrzej Ehrenfeucht and George F. McNulty, *Avoidable Patterns in Strings of Symbols*, Pacific J. Math. Vol. *85*, No. 2, (1979), 261-293.

Dept. of Mathematics and Statistics
Carleton University
Ottawa, Ontario.

Received September 14, 1983; Revised January 8, 1984.