ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/242390959

Unforgeable Marker Sequences

Article · January 2008

CITATIONS
0

READS
23

2 authors, including:

Stephen Montgomery-Smith
University of Missouri
113 PUBLICATIONS **1,459** CITATIONS

SEE PROFILE

All content following this page was uploaded by Stephen Montgomery-Smith on 05 April 2014.

The user has requested enhancement of the downloaded file.

Unforgeable Marker Sequences

D.J. Greaves
University of Cambridge Computer Laboratory
New Museums Site, Pembroke Street
Cambridge CB2 3QG, United Kingdom.

S.J. Montgomery-Smith*
Department of Mathematics
University of Missouri
Columbia, MO 65211, U.S.A.

A binary number of n bits consists of an ordered sequence of n digits taken from the set $\{0, 1\}$. A sequence is said to be an unforgeable marker if all subsequences of consecutive digits starting at the left-hand end are dissimilar from the sequence of the same length which ends at the right-hand end. Unforgeable marker sequences are so called because, when misaligned in a shift-register or other view port of the correct length, there is no possibility of adjacent random digits impersonating the true sequence. Such sequences are used for frame alignment purposes in serial data communications systems [1]. In many communications systems, the unforgeable marker is also a 'comma sequence' in that it is guaranteed not to occur in any aligned or misaligned view of the data between the markers, but this relies on constraints on the data and does not impact on whether a sequence is an unforgeable marker or not

A sequence which is unforgeable with respect to right-hand misalignment is also unforgeable with respect to left-hand misalignment, and so there is a single set of unforgeable sequences of a given length.

It is useful to denote a particular sequence by a decimal number by weighting its digits with power of 2 in the conventional way. We can then concisely

*The second named author was partially supported by the National Science Foundation

tabulate the unforgeable sequences for n up to 6.

$n = 1$: $\{0, 1\}$
$n = 2$: $\{1, 2\}$
$n = 3$: $\{1, 3, 4, 6\}$
$n = 4$: $\{1, 3, 7, 8, 12, 14\}$
$n = 5$: $\{1, 3, 5, 7, 11, 15, 16, 20, 24, 26, 28, 30\}$
$n = 6$: $\{1, 3, 5, 7, 11, 13, 15, 19, 23, 31, 32, 40, 44, 48, 50, 52, 56, 58, 60, 62\}$

There is an even number of elements in each set because if a number is unforgeable, so is the sequence obtained by complementing every digit. This property is useful in NRZI (non-return to zero invert on ones) modulation since, in general, the absolute polarity of the sequence will not be known by the decoding hardware.

Number of Unforgeable Sequences

Let us write $F(n)$ for the number of unforgeable sequences of length n. We give a table of values for $F(n)$ for values of n between 1 and 20.

It is also of interest to consider what percentage of all sequences of length n are unforgeable, and we will denote this proportion by $\rho(n)$. Thus $\rho(n) = F(n)/2^n$.

n	$F(n)$	$\rho(n)$
1	2	100.00%
2	2	50.00%
3	4	50.00%
4	6	37.50%
5	12	37.50%
6	20	31.25%
7	40	31.25%
8	74	28.91%
9	148	28.91%
10	284	27.73%

n	$F(n)$	$\rho(n)$
11	568	27.73%
12	1116	27.25%
13	2232	27.25%
14	4424	27.00%
15	8848	27.00%
16	17622	26.89%
17	35244	26.89%
18	70340	26.83%
19	140680	26.83%
20	281076	26.81%

We see that the $\rho(n)$ is decreasing as n gets larger, and that it seems to converge to a non-zero limit as n tends to infinity. In fact, as we shall show, this is indeed the case, and the limit is about 26.78%. We give a very precise value for this limit in the appendix.

Calculation of the Number of Unforgeable Sequences

We have a recurrence relation for $F(n)$ given by

$$F(n) = \begin{cases} 2 & \text{for } n = 1 \\ 2F(n-1) - F(n/2) & \text{for } n \text{ even} \\ 2F(n-1) & \text{otherwise.} \end{cases}$$

The proof is by induction on n. The cases $n = 1$ and $n = 2$ are taken from the previous tabulation. Let us assume that the recurrence relation holds for $n - 2$ and $n - 1$.

If n is odd, we can put $n = 2m + 1$. Now the sequence of length n

$$x_1, x_2, ..., x_{m-1}, x_m, x_{m+1}, ..., x_n$$

is not forgeable if and only if the sequence of length $n - 1$

$$x_1, x_2, ..., x_{m-1}, x_{m+1}, ..., x_n$$

is not forgeable (where x_m can be either 0 or 1). Therefore there are two unforgeable sequences of length n for every unforgeable sequence of length $n - 1$, that is, $F(n) = 2F(n - 2)$.

If n is even, we can put $n = 2m$. It is sufficient to show that

$$F(n) = 4F(n - 2) - F(n/2)$$

It is clear that the sequence of length n

$$x_1, x_2, ..., x_{m-1}, x_m, x_{m+1}, ..., x_n$$

is not forgeable if and only if both

1. the sequence of length $n - 2$

   $$x_1, x_2, ..., x_{m-1}, x_{m+1}, ..., x_n$$

   is not forgeable and,

2. the two sequences of length m

   $$x_1, x_2, ..., x_{m-1} \quad \text{and} \quad x_{m+1}, ..., x_n$$

   are not the same.

The number of times condition (1) is satisfied is $4F(n-2)$ since x_m and x_{m+1} may freely range over the values 0 and 1. So it only remains to show that the number of times condition (1) is satisfied while condition (2) is not, is given by $F(n/2)$. However, condition (1) is true while condition (2) is not if and only if the following two subsidiary conditions are met

1. $x_1, x_2, ..., x_{m-1}$ and $x_{m+1}, ..., x_n$ are the same, and
2. the sequence $x_1, x_2, ..., x_m$ is unforgeable.

This happens $F(m)$ times as $x_1, x_2, ..., x_m$ range over all possible values, and $F(m) = F(n/2)$.

Generating Sets of Unforgeable Sequences

A side-product of this proof is a straightforward procedure for generating unforgeable sequences.

If $X(n)$ is the set of unforgeable sequences of length n, then for odd n we can generate $X(n)$ from $X(n-1)$ as follows. There are twice as many elements in $X(n)$ as $X(n-1)$ and they can be generated by taking each sequence from $X(n-1)$, splitting it in half and then twice rebuilding it, once with a zero in the middle and once with a one.

For even n, we can generate $X(n)$ from $X(n-2)$ by splitting each sequence from $X(n-2)$ and putting a two digit sequence between the two halves. This gives four times as many elements as in $X(n-2)$, but some of the new ones are forgeable and must be removed. These are the ones where the first half of the sequence is the same as the second half of the sequence or the half sequences themselves are forgeable (not elements of $X(n/2)$).

A Convenient Test for Unforgeability

This C routine gives returns a non-zero value if the sequence held in the low order `bitwidth` bits of `x` is unforgeable.

```
int unforgeable(long x, int bitwidth)
{
  long masklo, i, lo=x, hi=x;
  masklo = (1<<bitwidth)-1;
  for(i=1;i<bitwidth;i++)
```

```
    {
      masklo >>= 1;
      lo = lo & masklo;
      hi = (hi>>1) & masklo;
      if (hi==lo)
        return 0;
    }
    return 1;
  }
```

Appendix: Derivation of the "26.78%"

We observed in the table above that $\rho(n)$ converges to a limit, which we will denote ρ. In fact, ρ can be calculated to a very high precision using the following, rapidly converging, infinite series:

$$\begin{aligned}
\rho &= \frac{2}{7} - \frac{2\cdot 8}{7\cdot 127} + \frac{2\cdot 8\cdot 128}{7\cdot 127\cdot 32767} - \frac{2\cdot 8\cdot 128\cdot 32768}{7\cdot 127\cdot 32767\cdot (2^{31}-1)}\cdots \\
&= \sum_{n=1}^{\infty}(-1)^{n-1}\left(\frac{2}{2^{2^{n+1}-1}-1}\right)\left(\prod_{m=2}^{n}\frac{2^{2^m-1}}{2^{2^m-1}-1}\right) \\
&\approx 0.267786840217889112376671403584302552555055989799348
\end{aligned}$$

We will prove this formula. First, an easy argument shows that the sequence $\rho(n)$ is a decreasing sequence, bounded below by zero. Hence the limit ρ exists. Now define a generating function

$$h(x) = \sum_{n=1}^{\infty} F(n)x^n.$$

It is clear that $h(x)$ has radius of convergence at least $1/2$.

Next, we show that

$$\rho = \lim_{x\nearrow 1/2}(1-2x)h(x).$$

To see this, multiply out to obtain

$$(1-2x)h(x) = F(1)x + \sum_{n=2}^{\infty}(F(n)-2F(n-1))x^n,$$

and thus

$$\lim_{x \nearrow 1/2} (1-2x)h(x) = \rho(1) + \sum_{n=2}^{\infty} (\rho(n) - \rho(n-1)),$$

which may be seen to be a telescoping series that converges to ρ.

Applying the recurrence relation for the sequence $F(n)$, we see that

$$\begin{aligned} h(x) &= 2x + \sum_{n=2}^{\infty} 2F(n-1)x^n - \sum_{\substack{n=2 \\ n \text{ even}}}^{\infty} F(n/2)x^n \\ &= 2x + \sum_{n=1}^{\infty} 2F(n)x^{n+1} - \sum_{n=1}^{\infty} F(n)x^{2n} \\ &= 2x + 2xh(x) - h(x^2), \end{aligned}$$

that is

$$(1-2x)h(x) = 2x - h(x^2).$$

We see immediately that

$$\rho = \lim_{x \nearrow 1/2} (1-2x)h(x) = 1 - h(\tfrac{1}{4}).$$

Applying this relation again, we get that

$$h(\tfrac{1}{4}) = 1 - 2h(\tfrac{1}{16}).$$

Again, we see that

$$\begin{aligned} h(\tfrac{1}{16}) &= \tfrac{8}{7}(\tfrac{1}{8} - h(\tfrac{1}{256})) \\ &= \tfrac{8}{7}(\tfrac{1}{8} - \tfrac{128}{127}(\tfrac{1}{128} - h(\tfrac{1}{65536}))) \\ &= \tfrac{8}{7}(\tfrac{1}{8} - \tfrac{128}{127}(\tfrac{1}{128} - \tfrac{32768}{32767}(\tfrac{1}{3276} - h(2^{-32})))) \end{aligned}$$

Continuing in this way, and multiplying out, we see that

$$\begin{aligned} \rho &= \sum_{n=1}^{N} (-1)^{n-1} \left(\frac{2}{2^{2^{n+1}-1} - 1} \right) \left(\prod_{m=2}^{n} \frac{2^{2^m-1}}{2^{2^m-1} - 1} \right) \\ &+ (-1)^N 2h(2^{-2^{N+2}}) \left(\prod_{m=2}^{N+1} \frac{2^{2^m-1}}{2^{2^m-1} - 1} \right). \end{aligned}$$

Letting $N \to \infty$, and noting that $h(2^{-2^{N+2}}) \to h(0) = 0$, it is readily seen that the last term in the above expression converges to zero, and hence we obtain the formula for ρ that we stated above.

References

[1] P. Bylanski and D.G.W. Ingram, Digital transmission systems, Stevenage Eng.: P. Peregrinus on behalf of the Institution of Electrical Engineers, 1980