F codes

Marek Michalik *

Abstract

The notion of an F code is introduced as a generalization of the notion of an L code. All interrelations between ordinary codes of bounded delay, L codes of bounded delay and F codes are established. Attention is also focused on unary morphisms. Many of them are F codes.

1 Introduction

Consider a nonerasing morphism $h : A^* \to A^*$, where A is a finite alphabet. We emphasise that all morphisms discussed in this paper are nonerasing, that is, $h(a) \neq 1$ (the empty word) for every $a \in A$. A morphism h is a code if h is injective. We will denote by C the class of all codes. A morphism h is an L code if the function $\bar{h}$ given by

$$\bar{h}(a_1 a_2 ... a_n) = h(a_1)h^2(a_2)...h^n(a_n)$$

($a_i \in A$ and $h^i(a_i)$ is the i-th iterate of the morphism h) is injective. For a positive integer k and a word w, we denote by $\text{pref}_k(w)$ the prefix (initial subword) of w of length k. If a word w is shorter then k, then $\text{pref}_k(w) = w$. The first letter of the word w, we denote by $\text{first}(w)$. A morphism h is of *bounded delay* k if, for all words u and w, the equation

$$\text{pref}_k(h(u)) = \text{pref}_k(h(w))$$

implies the equation $\text{first}(u) = \text{first}(w)$. A morphism h is of *bounded delay* if it is of bounded delay k, for some k. A morphism h is of *weakly bounded delay* k if, for all words u and w, the equation

$$\text{pref}_k(\bar{h}(u)) = \text{pref}_k(\bar{h}(w))$$

implies the equation $\text{first}(u) = \text{first}(w)$. If for all $i \geq 0$, the equation

$$\text{pref}_k(h^i\bar{h}(u)) = \text{pref}_k(h^i\bar{h}(w))$$

implies the equation $\text{first}(u) = \text{first}(w)$, then h is of *strongly bounded delay* k. In general, h is of weakly or strongly bounded delay if it is so for some k. A morphism

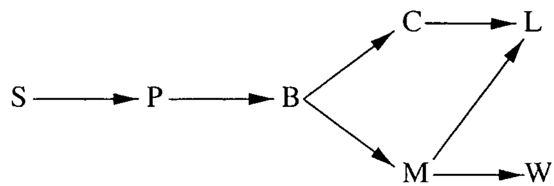
*Faculty of Applied Mathematics, University of Mining and Metallurgy, Al. Mickiwicza 30, 30-059 Cracow, Poland. e-mail: grmichal@kinga.cyf-kr.edu.pl

h is of *medium bounded delay* if, for some recursive function f and all $i \geq 0$, u and w, the equation

$$\mathrm{pref}_{f(i)}(h^i\bar{h}(u)) = \mathrm{pref}_{f(i)}(h^i\bar{h}(w))$$

implies the equation $\mathrm{first}(u) = \mathrm{first}(w)$. Morphism h is a prefix code if for every different words u, w $h(u)$ is not a prefix of $h(w)$. We will denote by L, B, W, S, M, P the corresponding classes of the morphisms.

The next diagram due to [1] shows all inclusion between the classes introduced. The arrow stands for strict inclusion.



2 F codes

From now on, f, g denote functions $\mathbb{N} \to \mathbb{N}$. We say that $f \prec g$ if there exists $n_0 \in \mathbb{N}$ such that $f(n) = g(n)$ for $n < n_0$ and $f(n_0) < g(n_0)$.

We will use the symbol $\hat{h}_f$ to denote function $\hat{h}_f : A^* \to A^*$ given by

$$\hat{h}_f(a_1a_2...a_n) = h^{f(1)}(a_1)h^{f(2)}(a_2)...h^{f(n)}(a_n)$$

($a_i \in A, h^{f(i)}(a_i)$ is the $f(i)$-th iterate of the morphism h) We call the morphism h an F code if there exists a minimal function $f : \mathbb{N} \to \mathbb{N}$ such that $\hat{h}_f$ is injective. We will denote by F the class of all F codes. It is easy to see that every L code is an F code.

If function $\bar{h}$ is injective then there exists a minimal function f such that $\hat{h} := \hat{h}_f$ is injective. Thus we conclude that every L code is an F code. We show that $F - L \neq \emptyset$.

Lemma 2.1 *Let* $A = \{a_1, \ldots, a_n\}, h : A^* \to A^*$,
$\mathrm{min}_k := \min\{||h^k(a_i)| - |h^k(a_j)||, |h^k(a_i)| : i \neq j; i,j \in \{1, \ldots, n\}\}$,
$\mathrm{max}_k := \max\{|h^k(a_i)| : i \in \{1, \ldots, n\}\}$.
If for each $n \in N$ *there exists* k, *such that* $\mathrm{min}_k > n$ *then we can define the function* f *as follows*
$f(1) = \min\{k : \mathrm{min}_k > 0\}$, $d_1 := \mathrm{max}_{f(1)}$,
$\forall i \in N$ $f(i+1) := \min\{k : \mathrm{min}_k > d_i\}$, $d_{i+1} := \mathrm{max}_{f(i+1)} + d_i$
The function $\hat{h}_f$ *is injective.*

Proof. It suffices to prove that different words have different length. The proof of this is by induction on word length. By the definition of $f(1)$ we have $||\hat{h}_f(a_i)| -$

$|\hat{h}_f(a_j)|| > 0$ for all $a_i, a_j \in A, i \neq j$. Let $w = a_1 \dots a_k, u = a'_1 \dots a'_k a'_{k+1}$. It is clear that $|\hat{h}_f(u)| > |h^{f(k+1)}(a'_{k+1})| > d_k \geq |\hat{h}_f(w)|$. The proof is completed by showing that for all $u, w \in A, |u| = |w| = k+1$, it holds that $||\hat{h}_f(u)| - |\hat{h}_f(w)|| > 0$. Consider $w = a_1 \dots a_{k+1}, u = a'_1 \dots a'_{k+1}$ and $a_{k+1} \neq a'_{k+1}$. We see at once that $||h^{f(k+1)}(a_{k+1})| - |h^{f(k+1)}(a'_{k+1})|| > d_k, |\hat{h}_f(a_1 \dots a_k)| \leq d_k, |\hat{h}_f(a'_1 \dots a'_k)| \leq d_k$, thus $||\hat{h}_f(a_1 \dots a_{k+1})| - |\hat{h}_f(a'_1 \dots a'_{k+1})|| > 0$ and finally $||\hat{h}_f(u)| - |\hat{h}_f(w)|| > 0$. □

Lemma 2.2 *Let* $A = \{a_1, \dots, a_n\}, h : A^* \to A^*$,
$\min_k := \min\{||h^k(a_i)| - |h^k(a_j)||, |h^k(a_i)| : i \neq j; i, j \in \{1, \dots, n\}\}$.
If the sequence $\min_k$ *is not bounded then morphism* h *is an* F *code.*

Proof. Let f be defined as in lemma 2.1. There exists a minimal function g such that $\hat{h}_g$ is injective, thus $h \in F$. □

Theorem 2.3 *The class* L *is strictly included in the class* F.

Proof. Every L code is an F code. Let $h: \{a, b\}^* \to \{a, b\}^*$ be given by $h(a) = a^2$, $h(b) = a^6$. Morphism h is not an L code ($\bar{h}(aa) = \bar{h}(b)$). We have $\min_k = 2^k$, hence by lemma 2.2, $h \in F$. □

Remark 2.4 *For the morphism* $h: \{a, b\}^* \to \{a, b\}^*$ *given by* $h(a) = a^2$, $h(b) = a^6$ *the function* $f(i) = 2i - 1$ *is a minimal function such that* $\hat{h}_f$ *is injective.*

Proof. To prove this, we observe that for every function $p(i) = n_i$ such that $\hat{h}_p$ is an injective function we have $\forall i \neq j\ (n_i \neq n_j) \wedge (|n_i - n_j| \neq 1)$ (as a consequence of $h^n(ab) = h^n(ba)$ and $h^n(a)h^{n+1}(a) = h^n(b)$), hence p is minimal. □

It is easy to check that if $h : A^* \to A^*$ is an F code then $h_{|A} : A \to A^*$ is injective. The reverse implication is not true. Let $A = \{a, b, c, d\}$ and h be given by $h(a) = b, h(b) = a, h(c) = bd, h(d) = d$. Function $h_{|A}$ is injective. For every $f : N \to N$, $\hat{h}_f(ad) = \hat{h}_f(c)$, hence $h \notin F$.

Similarly to L codes we have

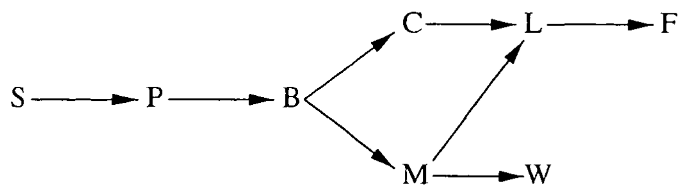
Remark 2.5 *The composition of two* F *codes is not necessarily an* F *code.*

Proof. Consider the morphisms g and h defined by $g(a) = ab, g(b) = ba, h(a) = a^2, h(b) = a$. Clearly g is a code and, hence, an F code. h is an L code. However, the composition $h \circ g \equiv a^3$ is not an F code. □

Theorem 2.6 *Classes* F *and* W *are incomparable.*

Proof. It suffices to show that $W \not\subseteq F$. Consider a morphism h given by $h(a) = edb, h(b) = b^2, h(c) = deb, h(d) = a, h(e) = a^3$, then $h \in W$ (see [1]). For all $i \geq 2$, $h^i(a) = h^i(c)$ and $h(ddd) = h(e)$, hence $h \notin F$. From this we conclude that W and F are incomparable. □

Now we can redraw the diagram as follows:



If we restrict our considerations to the class of morphisms $h : \{a, b\}^* \to \{a, b\}^*$ we obtain

Theorem 2.7 *$B = C = M = W$ and*

$$S \longrightarrow P \longrightarrow C \longrightarrow L \longrightarrow F$$

Proof. Let $h(a) = ba$, $h(b) = b^2$. $h \in P \setminus S$ (see [1]). Morphism $h(a) = a$, $h(b) = ab$ is a code but not a prefix code. Morphism $h(a) = a^2$, $h(b) = a$ is an element of $L - C$. From theorem 2.3 we obtain $F \setminus L \neq \emptyset$. To complete the proof it suffices to show that $C \subset B$ and $W \subset C$.
($C \subset B$) Every prefix code is of bounded delay. Let $h(a) = x$, $h(b) = xy$, $x, y \in A^+, x \neq y$. Morphism h is of bounded delay $k = 2|x| + |y|$.
($W \subset C$) Let $h \notin C$, then there exist $n, m \in \mathbb{N}$ such that $h(a) = x^m$, $h(b) = x^n$ thus $h \notin W$. □

3 F codes and the unary morphism

Theorem 3.1 *Let $A = \{a_1, \ldots, a_n\}$, $a \in A$, $h : A^* \to \{a\}^*$,*
$\min_k := \min\{||h^k(a_i)| - |h^k(a_j)||, |h^k(a_i)| : i \neq j; i, j \in \{1, \ldots, n\}\}$.
The unary morphism h is an F code if and only if the sequence $\min_k$ is not bounded.

Proof. If sequence $\min_k$ is not bounded then from lemma 2.1 $h \in F$.
Consider $h \in F$. Suppose that there exists $M \in \mathbb{N}$ such that for all $k \in \mathbb{N}$ $\min_k \leq M$. The morphism h is a nonerasing morphism, thus for all $a_i \in A$ and $n \in \mathbb{N}$ we have $|h^n(a_i)| \leq |h^{n+1}(a_i)|$. If $\min_k \leq M$ then $\exists n_0 \in \mathbb{N} \forall t \geq n_0 \forall i \in \{1, \ldots, n\}$ $|h^t(a_i)| = |h^{t+1}(a_i)|$. We have $a^p = h^t(a_i) = h^{t+1}(a_i) = h(h^t(a_i)) = h(a^p)$ for some $p \in \mathbb{N}$. This implies $h(a) = a$ and finally $\hat{h}(a_i) = \hat{h}(a^{|\hat{h}(a_i)|})$ which contradicts $h \in F$. □

Remark 3.2 *The last theorem is not true for arbitrary morphism.*

Proof. Let $h(a) = a$, $h(b) = ab$, $h(c) = b$. For every $w, u_1, u_2, u_3 \in \{a, b, c\}^*$ $\bar{h}(wbu_1) \neq \bar{h}(wcu_2)$, $\bar{h}(wbu_1) \neq \bar{h}(wau_3)$, $\bar{h}(wcu_2) \neq \bar{h}(wau_3)$. To show this we observe that $\forall u_3, v, w \in A^*$ $\forall k \in \mathbb{N}$
$\mathrm{pref}_k\{ah^{|w|+1}(\bar{h}(u_3))\} \notin \{a^{|w|+1}bv, a^{|w|}bv\}$. Thus h is an L code, but $\min_1 = 0$ and $\forall k > 1$ $\min_k = 1$. □

Corollary 3.3 *It is decidable whether the unary morphism is an F code.*

The morphism h is an almost L code if and only if h is not an L code and $\exists t \in \mathbb{N}\, \forall w, u \in A^*, \text{first}(w) \neq \text{first}(u)\ \left(\bar{h}(w) = \bar{h}(u) \Rightarrow (|w| = t \vee |u| = t)\right)$

Remark 3.4 $F \setminus \text{almost L} \neq \emptyset$.

Proof. Let $h(a) = a^3$, $h(b) = a^6$, $h(c) = a^2$, then $h^n(a) = a^{3^n}$, $h^n(b) = a^{6 \cdot 3^{n-1}}$, $h^n(c) = a^{2 \cdot 3^{n-1}}$, $\min_n = 3^{n-1}$. From lemma 2.2 we obtain $h \in F$. For every $w \in A^*$ $\bar{h}(aaw) = \bar{h}(bcw)$, thus $h \notin \text{almost}L$. □

Let U be the class of all unary morphisms $h : \{a, b\} \to \{a\}^*$ such that $h(a) = a^n$, $h(b) = a^r$, $n \neq r$, $n \geq 2$.

Theorem 3.5 $F \cap U = (L \cup \text{almost L}) \cap \text{U}$

Proof. It is clear that if $n = r$ or $n = 1$ then h is neither an F code nor an almost L code. If $n \neq r$ and $n \geq 2$ then the sequence $\min_k = \min\{|n^k - r \cdot n^{k-1}|, n^k, r \cdot n^{k-1}\}$ is not bounded. This implies $h \in F$. Every unary morphism such that $h(a) = a^n$, $h(b) = a^r$, $n \neq r$, $n \geq 2$ is either an L code or an almost L code (see [3]). □

References

[1] H.A. Maurer, A. Salomaa, D. Wood, *Bounded delay L codes* (Theoretical Computer Science 84 (1991) 265-279)

[2] J. Honkala and A. Salomaa, *Characterization results about L codes* (Informatique theorique et Applications vol. 26, $n°3$, (1992) 287-301)

[3] H.A. Maurer, A. Salomaa, D. Wood, *L codes and number systems* (Theoretical Computer Science 22 (1983) 331-346)

Received January, 1999