In this work we study the class of regular strongly shuffle-closed languages and we present their description by giving a class of recognition automata.

The shuffle product operation plays an important role in the theory of formal languages, cf. [1], [2], [4]. Several properties of shuffle closed languages are studied in [3]. Among others a characterization of regular strongly shuffle-closed languages is presented by giving their expressions. Using this result, we determine a very simple class of deterministic automata accepting regular strongly shuffle-closed languages.

First of all we introduce some notions and notations. Let $X$ be a nonempty finite set and let $X^*$ denote the free monoid of words generated by $X$. We denote by $1$ the empty word of $X^*$. The *shuffle product* of two words $u, v \in X^*$ is the set

$$u \circ v = \{w : w = u_1 v_1 ... u_k v_k, u = u_1 ... u_k, v = v_1 ... v_k, u_i, v_j \in X^*\}.$$ 

A language $L \subseteq X^*$ is called *shuffle-closed* if it is closed under $\circ$, that is, if $u, v \in L$, then $u \circ v \subseteq L$. If $L$ is shuffle-closed and, for any $u \in L$, $v \in X^*$, the condition $u \circ v \cap L \neq \emptyset$ implies $v \in L$, then $L$ is called a *strongly shuffle-closed language*, or briefly, an *ssh-closed language*.

Next let $X = \{x_1, ..., x_r\}$, $r \geq 1$, be an arbitrarily fixed alphabet. For any $L \subseteq X^*$, let us denote by $\text{alph}(L)$ the set of elements of $X$ occurring in words of $L$. We shall describe those regular ssh-closed languages over $X$ for which $\text{alph}(L) = X$.

We use the Parikh mapping and its inverse which are defined as follows. Let $N = \{0, 1, 2, ...\}$. The mapping $\Psi$ of $X^*$ into the set $N^r$ defined by

$$\Psi(u) = (\mu_{x_1}(u), ..., \mu_{x_r}(u)),$$ 

is called the *Parikh mapping*, where $\mu_{x_i}(u)$ denotes the number of occurrences of $x_i$ in $u$. For a language $L \subseteq X^*$, we define $\Psi(L) = \{\Psi(u) : u \in L\}$. Moreover, if $S \subseteq N^r$, then $\Psi^{-1}(S) = \{u : u \in X^* \land \Psi(u) \in S\}$.

Now we recall a notation and a result from [3].

Let $a = (i_1, ..., i_r)$, $b = (j_1, ..., j_r) \in N^r$ and let $p_1, ..., p_r$ be positive integers. Then $a \rightarrow b \pmod{(p_1, ..., p_r)}$ means that $i_t \geq j_t$ and $i_t \equiv j_t \pmod{p_t}$, for all $t$, $t = 1, ..., r$.

---

*Department of Informatics, A. József University, Árpád tér 2, H-6720 Szeged, Hungary
†Faculty of Science, Kyoto Sangyo University, 603 Kyoto, Japan*
Theorem 1 ([3], Proposition 5.2) Let $L \subseteq X^*$ with $\text{alph}(L) = X$. Then $L$ is a regular ssh-closed language if and only if $L$ is presented as

$$L = \bigcup_{u \in F} \Psi^{-1}(\Psi(u(x_1^{P_1})^* \cdots (x_r^{P_r})^*))$$

where

(i) $p_1, \ldots, p_r$ are positive integers,

(ii) $F$ is a finite language over $X$ satisfying

(iii)-(1) for any $u \in F$, we have $0 \leq j_t < p_t$, $1 \leq t \leq r$ where $\Psi(u) = (j_1, \ldots, j_r)$,

(iii)-(2) for any $u, v \in F$, there is a $w \in F$ such that $\Psi(uv) \mapsto \Psi(u)(\text{mod} \ (p_1, \ldots, p_r))$,

(iii)-(3) for any $u, v \in F$, there is a $w \in F$ such that $\Psi(uv) \mapsto \Psi(v)(u)(\text{mod} \ (p_1, \ldots, p_r))$.

Finally, we make some further preparation. For any positive integer $p$ and $x_t \in X$, let us denote by $C(p,x_t) = (X, \{0, \ldots, p-1\}, S(p,x_t))$ the automaton defined by the following transition function. For any $j \in \{0, \ldots, p-1\}$, $x \in X$, let

$$\delta(p,x_t)(j, x) = \begin{cases}  j & \text{if } x \neq x_t, \\  j + 1(\text{mod } p) & \text{if } x = x_t \\ \end{cases}$$

where $j + 1(\text{mod } p)$ denotes the least nonnegative residue of $j + 1$ modulo $p$.

Now let $p_1, \ldots, p_r$ be positive integers and form the direct product of the automata $C(p_t,x_t)$, $t = 1, \ldots, r$. Let us denote by $C(p_1,\ldots,p_r)$ this direct product and by $\delta(p_1,\ldots,p_r)$ its transition function. It is easy to prove that $C(p_1,\ldots,p_r)$ has the following properties:

(a) it is a commutative automaton,

(b) if $a, b \in \prod_{t=1}^r \{0, \ldots, p_t - 1\}$, $u \in X^*$ are such that $\delta(p_1,\ldots,p_r)(a, u) = b$, then $\delta(p_1,\ldots,p_r)(a, v) = b$, for all $v \in \Psi^{-1}(\Psi(u))$,

(c) for any $u \in X^*$, $\delta(p_1,\ldots,p_r)(0, u) = \Psi(u)(\text{mod} \ (p_1, \ldots, p_r))$,

where $0$ denotes the $r$-dimensional 0-vector and $\Psi(u)(\text{mod} \ (p_1, \ldots, p_r))$ denotes the vector $(i_1(\text{mod } p_1), \ldots, i_r(\text{mod } p_r))$ with $\Psi(u) = (i_1, \ldots, i_r)$.

For each $t$, $t = 1, \ldots, r$, let us denote by $M_{p_t}$ the group defined by the addition mod $p_t$ over the set $\{0, \ldots, p_t - 1\}$. Let $M(p_1,\ldots,p_r)$ denote the direct product of the groups $M_{p_t}$, $t = 1, \ldots, r$. Then $M(p_1,\ldots,p_r)$ is also a group; let $\oplus$ denote its operation. Let us observe that the set of states of $C(p_1,\ldots,p_r)$ is equal to the set of elements of $M(p_1,\ldots,p_r)$. Therefore, for any subgroup $H$ of $M(p_1,\ldots,p_r)$, we can define the recognizer

$$R_H^{(p_1,\ldots,p_r)} = (\prod_{t=1}^r \{0, \ldots, p_t - 1\}, X, \delta(p_1,\ldots,p_r), 0, H),$$

where $0$ is the initial state and $H$ is the set of the final states.

The next property of $R_H^{(p_1,\ldots,p_r)}$ can be proved easily:

(d) if $u, v \in X^*$ are accepted by $R_H^{(p_1,\ldots,p_r)}$ with final states $a$, $b$, respectively, then $uv$ is also accepted by $R_H^{(p_1,\ldots,p_r)}$ with the final state $a \oplus b$. 
Finally, form the set of recognizers

\[ M_X = \{ R_{H}^{(p_1, \ldots, p_r)} : (p_1, \ldots, p_r) \in \mathbb{N}^r \text{ and } H \text{ is a subgroup of } M^{(p_1, \ldots, p_r)} \}. \]

Now we are ready to prove our result.

**Theorem 2** A language \( L \subseteq X^* \) with \( \text{alph}(L) = X \) is regular ssh-closed if and only if \( L \) is accepted by a recognizer from \( M_X \).

**Proof.** In order to prove the necessity, let us suppose that \( L \subseteq X^* \) is a regular ssh-closed language with \( \text{alph}(L) = X \). Then there are positive integers \( p_1, \ldots, p_r \) and \( F \subseteq X^* \) which satisfy the conditions of Theorem 1. Let us consider the automaton \( C(\pi, p_1, \ldots, p_r) \) and let us define the set \( H \) by

\[ H = \{ a : a \in \prod_{t=1}^{r} \{0, \ldots, p_t - 1\} \text{ and } \delta^{(p_1, \ldots, p_r)}(0, u) = a, \text{ for some } u \in F \}. \]

We show that \( H \) is a subgroup of \( M^{(p_1, \ldots, p_r)} \). Indeed, let \( a, b \in H \) be arbitrary elements. By the definition of \( H \), there are \( u, v \in F \) with \( \delta^{(p_1, \ldots, p_r)}(0, u) = a \) and \( \delta^{(p_1, \ldots, p_r)}(0, v) = b \). Let \( \Psi(u) = (i_1, \ldots, i_r) \) and \( \Psi(v) = (j_1, \ldots, j_r) \). Then, by \((ii) - (1)\), we have \( 0 \leq i_t, j_t < p_t \) for all \( t = 1, \ldots, r \), and hence, we obtain, by \((c)\), that \( a = (i_1, \ldots, i_r) \) and \( b = (j_1, \ldots, j_r) \). On the other hand, by \((ii) - (g)\) of Theorem 1, there exists a \( w \in F \) with \( \Psi(uw) \leftarrow \Psi(w)(\text{mod } (p_1, \ldots, p_r)) \). Let \( \Psi(w) = (k_1, \ldots, k_r) \). Then, by \((ii) - (1)\) and \((c)\), \( \delta^{(p_1, \ldots, p_r)}(0, w) = (k_1, \ldots, k_r) \). Since \( w \in F \), we have \( (k_1, \ldots, k_r) \in H \). From \( \Psi(uw) \leftarrow \Psi(w) \) it follows that \( i_t + j_t \equiv k_t(\text{mod } p_t) \), \( t = 1, \ldots, r \). But then \( a \oplus b = (k_1, \ldots, k_r) \). Therefore, \( H \) is closed under the operation \( \oplus \) implying that \( H \) is a subgroup of \( M^{(p_1, \ldots, p_r)} \). This completes the proof of the necessity.

In order to prove the sufficiency, let us suppose that \( L \subseteq X^* \) with \( \text{alph}(L) = X \) and there exists a recognizer \( R_{H}^{(p_1, \ldots, p_r)} \in M_X \) accepting \( L \). We show that \( L \) is a regular ssh-closed language.

The regularity of \( L \) is obvious. Now let \( u, v \in L \) and let \( w \) be an arbitrary element of the set \( u \circ v \). Since \( L \) is accepted by \( R_{H}^{(p_1, \ldots, p_r)} \), there are \( a, b \in H \) such that \( \delta^{(p_1, \ldots, p_r)}(0, u) = a \) and \( \delta^{(p_1, \ldots, p_r)}(0, v) = b \). Therefore, by \((d)\), we obtain that \( uv \) is accepted by \( R_{H}^{(p_1, \ldots, p_r)} \) with the final state \( a \oplus b \). From this, by \((b)\), we get that \( w \in L \), and so, \( L \) is shuffle-closed.

Finally, let \( u \in L, v \in X^* \) and let us assume that \( u \circ v \cap L \neq \emptyset \). If \( v = 1 \), then \( \delta^{(p_1, \ldots, p_r)}(0, u) = 0 \in H \), and so, \( v \subseteq L \). Now let us suppose that \( v \neq 1 \). Let \( \delta^{(p_1, \ldots, p_r)}(0, u) = a, \delta^{(p_1, \ldots, p_r)}(0, v) = b \) and let \( \Psi(u) = (i'_1, \ldots, i'_r), \Psi(v) = (j'_1, \ldots, j'_r) \). Then there exist nonnegative integers \( i_t < p_t, j_t < p_t, i_t, j_t, t = 1, \ldots, r \), such that \( i'_t = i_t + l_t p_t, j'_t = j_t + k_t p_t, t = 1, \ldots, r \). Let us denote by \( u' \) and \( v' \) the words \( x_1^{i_1+l_1p_1} \ldots x_r^{i_r+l_r^r} \) and \( x_1^{j_1+k_1p_1} \ldots x_r^{j_r+k_r^r} \), respectively. Using \((b)\) and \((c)\), we obtain that \( \delta^{(p_1, \ldots, p_r)}(0, u') = a, \delta^{(p_1, \ldots, p_r)}(0, v') = b \), where \( a = (i_1, \ldots, i_r), b = (j_1, \ldots, j_r) \). By our assumption on \( u \circ v \), there exists a word \( w \in u \circ v \cap L \). Let

\[ w' = x_1^{i_1+j_1+(l_1+k_1)p_1} \ldots x_r^{i_r+j_r+(l_r+k_r)p_r}. \]
Since \( w \in u \circ v \cap L \) and \( \Psi(w') = \Psi(u'v') = \Psi(uv) = \Psi(w) \), (b) implies \( w' \in L \). On the other hand, by (c), we have
\[
\delta^{(p_1, \ldots, p_r)}(0, w') = (i_1 + j_1 \text{mod } p_1, \ldots, i_r + j_r \text{mod } p_r).
\]
Now let us observe that \( (i_1 + j_1 \text{mod } p_1, \ldots, i_r + j_r \text{mod } p_r) = a \oplus b \). Since \( w' \in L \), we have \( a \oplus b \in H \). But \( H \) is a subgroup of \( \mathcal{M}^{(p_1, \ldots, p_r)} \), thus \( a \in H \) and \( a \oplus b \in H \) imply \( b \in H \). Therefore, by \( \delta^{(p_1, \ldots, p_r)}(0, v) = b \), we obtain that \( v \in L \), and so, \( L \) is an ssh-closed language. This completes the proof of the theorem.

References


Received September 1, 1994