Characterizing classes of regular languages using prefix codes of bounded synchronization delay

Volker Diekert and Tobias Walter

Universität Stuttgart

ICALP 2016, Rome, July 15th, 2016

Marcel-Paul Schützenberger, 1920–1996

A legacy of Schützenberger is the following program

- Consider a variety of groups $\mathbf{H}$ and the maximal variety of monoids $\overline{\mathbf{H}}$ such that all groups are in $\mathbf{H}$.
- For a language characterization of $\overline{\mathbf{H}}$, consider "$\mathbf{H}$-controlled stars" over prefix codes of bounded synchronization delay.
- Show that this leads to a language characterization of $\overline{\mathbf{H}}$.

Marcel-Paul Schützenberger, 1920–1996

A legacy of Schützenberger is the following program

- Consider a variety of groups $\mathbf{H}$ and the maximal variety of monoids $\overline{\mathbf{H}}$ such that all groups are in $\mathbf{H}$.
- For a language characterization of $\overline{\mathbf{H}}$, consider "$\mathbf{H}$-controlled stars" over prefix codes of bounded synchronization delay.
- Show that this leads to a language characterization of $\overline{\mathbf{H}}$.

For star-free languages there are no nontrivial groups, and the "$\mathbf{H}$-controlled star" is the usual Kleene-star.

Marcel-Paul Schützenberger, 1920–1996

A legacy of Schützenberger is the following program

- Consider a variety of groups $\mathbf{H}$ and the maximal variety of monoids $\overline{\mathbf{H}}$ such that all groups are in $\mathbf{H}$.
- For a language characterization of $\overline{\mathbf{H}}$, consider "$\mathbf{H}$-controlled stars" over prefix codes of bounded synchronization delay.
- Show that this leads to a language characterization of $\overline{\mathbf{H}}$.

For star-free languages there are no nontrivial groups, and the "$\mathbf{H}$-controlled star" is the usual Kleene-star.

Schützenberger (1975 and 1974): the program works for star-free languages; and it works if $\mathbf{H}$ is the class of abelian groups. One direction of the program work for all $\mathbf{H}$.

Marcel-Paul Schützenberger, 1920–1996

A legacy of Schützenberger is the following program

- Consider a variety of groups $\mathbf{H}$ and the maximal variety of monoids $\overline{\mathbf{H}}$ such that all groups are in $\mathbf{H}$.
- For a language characterization of $\overline{\mathbf{H}}$, consider "$\mathbf{H}$-controlled stars" over prefix codes of bounded synchronization delay.
- Show that this leads to a language characterization of $\overline{\mathbf{H}}$.

For star-free languages there are no nontrivial groups, and the "$\mathbf{H}$-controlled star" is the usual Kleene-star.

Schützenberger (1975 and 1974): the program works for star-free languages; and it works if $\mathbf{H}$ is the class of abelian groups. One direction of the program work for all $\mathbf{H}$.

Diekert & W. (2016): Both directions work for all varieties of groups $\mathbf{H}$.

Notation

- A = finite alphabet
- A^* = set of finite words
- M = finite monoid, G = finite group
- $h : A^* \to M$ is a homomorphism
- h recognizes $L \subseteq A^*$ if $h^{-1}(h(L)) = L$.
- If $\mathbf{V}$ is a class of finite monoids, then

$$\mathbf{V}(A^*) = \{L \subseteq A^* \mid \text{some } h : A^* \to M \in \mathbf{V} \text{ recognizes } L\}$$

Varieties of finite monoids

A variety $\mathbf{V}$ is a class of finite monoids which is closed under finite direct products and divisors.

Example

$\mathbf{1}$, $\mathbf{Ab}$, $\mathbf{G}$ are varieties of groups.

If $\mathbf{V}$ is a variety, then

$$\mathbf{V} \cap \mathbf{G} = \{G \in \mathbf{V} \mid G \text{ is a group}\}$$

is a variety of finite groups.

Varieties of finite monoids

A variety $\mathbf{V}$ is a class of finite monoids which is closed under finite direct products and divisors.

Example

$\mathbf{1}$, $\mathbf{Ab}$, $\mathbf{G}$ are varieties of groups.

If $\mathbf{V}$ is a variety, then

$$\mathbf{V} \cap \mathbf{G} = \{G \in \mathbf{V} \mid G \text{ is a group}\}$$

is a variety of finite groups.

If $\mathbf{H}$ is a variety of finite groups, then we let

$$\overline{\mathbf{H}} = \{M \in \mathbf{Mon} \mid \text{ all subgroups of } M \text{ are in } \mathbf{H}\}\,.$$

A variety $\mathbf{V}$ is a class of finite monoids which is closed under finite direct products and divisors.

Example

$\mathbf{1}$, $\mathbf{Ab}$, $\mathbf{G}$ are varieties of groups.

If $\mathbf{V}$ is a variety, then

$$\mathbf{V} \cap \mathbf{G} = \{G \in \mathbf{V} \mid G \text{ is a group}\}$$

is a variety of finite groups.

If $\mathbf{H}$ is a variety of finite groups, then we let

$$\overline{\mathbf{H}} = \{M \in \mathbf{Mon} \mid \text{ all subgroups of } M \text{ are in } \mathbf{H}\} .$$

Example

- $\overline{\mathbf{1}} = \mathbf{AP}$
- $\overline{\mathbf{G}} = \mathbf{Mon}$

Examples of language characterizations

- Regular languages: finite subsets & closure under finite union, concatenation, and Kleene-star
  $=$ recognizable by a finite monoid
  $= \overline{\mathbf{G}}(A^*)$.

Examples of language characterizations

- Regular languages: finite subsets & closure under finite union, concatenation, and Kleene-star
  $=$ recognizable by a finite monoid
  $= \overline{\mathbf{G}}(A^*)$.
- Star-free languages: finite subsets & closure under finite union, concatenation, complementation, but no Kleene-star
  $=$ recognizable by a finite aperiodic monoid
  $= \overline{\mathbf{1}}(A^*) = \mathbf{AP}(A^*)$.

Prefix codes of bounded synchronization delay

$K \subseteq A^+$ is called prefix code if it is prefix-free. That is: $u, uv \in K$ implies $u = uv$.

A prefix-free language K is a code since every word $u \in K^*$ admits a unique factorization $u = u_1 \cdots u_k$ with $k \geq 0$ and $u_i \in K$.

A prefix code K has bounded synchronization delay if for some $d \in \mathbb{N}$ and for all $u, v, w \in A^*$ we have:
if $uvw \in K^*$ and $v \in K^d$, then $uv \in K^*$.

Example

- $B \subseteq A$ yields a prefix code with synchronization delay 0.
- If $c \in A \setminus B$, then B^*c is a prefix code with delay 1.
- A^2 has unbounded synchronization delay.

$\mathbf{H}$-controlled star

Let $\mathbf{H}$ be a variety of groups and $G \in \mathbf{H}$. Let $K \subseteq A^+$ be a prefix code of bounded synchronization delay. Consider any mapping $\gamma : K \to G$ and define $K_g = \gamma^{-1}(g)$. Assume further that $K_g \in \overline{\mathbf{H}}(A^*)$ for all $g \in G$.

With these data the $\mathbf{H}$-controlled star $K^{\star\downarrow\gamma}$ is defined as:

$$K^{\star\downarrow\gamma} = \{u_{g_1} \cdots u_{g_k} \in K^* \mid u_{g_i} \in K_{g_i} \wedge g_1 \cdots g_k = 1 \in G\}\,.$$

$\mathbf{H}$-controlled star

Let $\mathbf{H}$ be a variety of groups and $G \in \mathbf{H}$. Let $K \subseteq A^+$ be a prefix code of bounded synchronization delay. Consider any mapping $\gamma : K \to G$ and define $K_g = \gamma^{-1}(g)$. Assume further that $K_g \in \overline{\mathbf{H}}(A^*)$ for all $g \in G$.

With these data the $\mathbf{H}$-controlled star $K^{\star\downarrow\gamma}$ is defined as:

$$K^{\star\downarrow\gamma} = \{u_{g_1} \cdots u_{g_k} \in K^* \mid u_{g_i} \in K_{g_i} \wedge g_1 \cdots g_k = 1 \in G\}.$$

Example

If G is the trivial group $\{1\}$, then $K^{\star\downarrow\gamma} = K^*$ is the usual star.

$\mathbf{H}$-controlled star

Let $\mathbf{H}$ be a variety of groups and $G \in \mathbf{H}$. Let $K \subseteq A^+$ be a prefix code of bounded synchronization delay. Consider any mapping $\gamma : K \to G$ and define $K_g = \gamma^{-1}(g)$. Assume further that $K_g \in \overline{\mathbf{H}}(A^*)$ for all $g \in G$.

With these data the $\mathbf{H}$-controlled star $K^{\star\downarrow\gamma}$ is defined as:

$$K^{\star\downarrow\gamma} = \{u_{g_1} \cdots u_{g_k} \in K^* \mid u_{g_i} \in K_{g_i} \wedge g_1 \cdots g_k = 1 \in G\}.$$

Example

If G is the trivial group $\{1\}$, then $K^{\star\downarrow\gamma} = K^*$ is the usual star.

Proposition (Schützenberger, **RAIRO**, 8:55–61, 1974.)

$\overline{\mathbf{H}}(A^*)$ *is closed under the* $\mathbf{H}$*-controlled star.*

Schützenberger's $\mathrm{SD}_{\mathbf{H}}$ classes

By $\mathrm{SD}_{\mathbf{H}}(A^*)$ we denote the set of regular languages which is inductively defined as follows.

Schützenberger's $\mathrm{SD}_{\mathbf{H}}$ classes

By $\mathrm{SD}_{\mathbf{H}}(A^*)$ we denote the set of regular languages which is inductively defined as follows.

1. Finite subsets of A^* are in $\mathrm{SD}_{\mathbf{H}}(A^*)$.

Schützenberger's $\mathrm{SD}_{\mathbf{H}}$ classes

By $\mathrm{SD}_{\mathbf{H}}(A^*)$ we denote the set of regular languages which is inductively defined as follows.

1. Finite subsets of A^* are in $\mathrm{SD}_{\mathbf{H}}(A^*)$.
2. If $L, K \in \mathrm{SD}_{\mathbf{H}}(A^*)$, then $L \cup K, L \cdot K \in \mathrm{SD}_{\mathbf{H}}(A^*)$.

Schützenberger's $\mathrm{SD}_{\mathbf{H}}$ classes

By $\mathrm{SD}_{\mathbf{H}}(A^*)$ we denote the set of regular languages which is inductively defined as follows.

1. Finite subsets of A^* are in $\mathrm{SD}_{\mathbf{H}}(A^*)$.
2. If $L, K \in \mathrm{SD}_{\mathbf{H}}(A^*)$, then $L \cup K, L \cdot K \in \mathrm{SD}_{\mathbf{H}}(A^*)$.
3. Let $K \subseteq A^+$ be a prefix code of bounded synchronization delay, $\gamma : K \to G \in \mathbf{H}$, and $\gamma^{-1}(g) \in \mathrm{SD}_{\mathbf{H}}(A^*)$ for all g. Then the $\mathbf{H}$-controlled star $K^{\star\downarrow\gamma}$ is in $\mathrm{SD}_{\mathbf{H}}(A^*)$.

Note: the definition doesn't involve any complementation!

Schützenberger's $\mathrm{SD}_{\mathbf{H}}$ classes

By $\mathrm{SD}_{\mathbf{H}}(A^*)$ we denote the set of regular languages which is inductively defined as follows.

1. Finite subsets of A^* are in $\mathrm{SD}_{\mathbf{H}}(A^*)$.
2. If $L, K \in \mathrm{SD}_{\mathbf{H}}(A^*)$, then $L \cup K, L \cdot K \in \mathrm{SD}_{\mathbf{H}}(A^*)$.
3. Let $K \subseteq A^+$ be a prefix code of bounded synchronization delay, $\gamma : K \to G \in \mathbf{H}$, and $\gamma^{-1}(g) \in \mathrm{SD}_{\mathbf{H}}(A^*)$ for all g. Then the $\mathbf{H}$-controlled star $K^{\star\downarrow\gamma}$ is in $\mathrm{SD}_{\mathbf{H}}(A^*)$.

Note: the definition doesn't involve any complementation!

Proposition (Schützenberger (1974) reformulated)

$$\mathrm{SD}_{\mathbf{H}}(A^*) \subseteq \overline{\mathbf{H}}(A^*)$$

Schützenberger's result holds for all varieties.

Theorem (Schützenberger (1975) and (1974))

$\mathrm{SD}_{\mathbf{1}}(A^*) = \overline{\mathbf{1}}(A^*) = \mathbf{AP}(A^*)$ *and* $\mathrm{SD}_{\mathbf{Ab}}(A^*) = \overline{\mathbf{Ab}}(A^*)$

Schützenberger's result holds for all varieties.

Theorem (Schützenberger (1975) and (1974))

$$\mathrm{SD}_{\mathbf{1}}(A^*) = \overline{\mathbf{1}}(A^*) = \mathbf{AP}(A^*) \text{ and } \mathrm{SD}_{\mathbf{Ab}}(A^*) = \overline{\mathbf{Ab}}(A^*)$$

Theorem (Diekert, W., 2016)

Let $\mathbf{H}$ be any variety of finite groups. Then we have

$$\mathrm{SD}_{\mathbf{H}}(A^*) = \overline{\mathbf{H}}(A^*).$$

Local divisor technique

- The local divisor technique is the "algebraic child" of an induction in Thomas Wilke's habilitation for a simplified proof of $\mathrm{FO} = \mathrm{LTL}$ over finite words, see STACS 1999 and this ICALP.

Local divisor technique

- The local divisor technique is the "algebraic child" of an induction in Thomas Wilke's habilitation for a simplified proof of $\mathrm{FO} = \mathrm{LTL}$ over finite words, see STACS 1999 and this ICALP.
- Let M be a monoid and $c \in M$. Consider the set $cM \cap Mc$ and define a new multiplication

$$xc \circ cy = xcy.$$

  Then $M_c = (cM \cap Mc, \circ, c)$ is monoid: the local divisor at c.

Local divisor technique

- The local divisor technique is the "algebraic child" of an induction in Thomas Wilke's habilitation for a simplified proof of $\mathrm{FO} = \mathrm{LTL}$ over finite words, see STACS 1999 and this ICALP.
- Let M be a monoid and $c \in M$. Consider the set $cM \cap Mc$ and define a new multiplication

$$xc \circ cy = xcy.$$

Then $M_c = (cM \cap Mc, \circ, c)$ is monoid: the local divisor at c.

Facts

- $\lambda_c : \{x \in M \mid cx \in Mc\} \to M_c$ given by $\lambda_c(x) = cx$ is a surjective homomorphism. Hence, M_c is a divisor of M.

Local divisor technique

- The local divisor technique is the "algebraic child" of an induction in Thomas Wilke's habilitation for a simplified proof of $\mathrm{FO} = \mathrm{LTL}$ over finite words, see STACS 1999 and this ICALP.
- Let M be a monoid and $c \in M$. Consider the set $cM \cap Mc$ and define a new multiplication

$$xc \circ cy = xcy.$$

Then $M_c = (cM \cap Mc, \circ, c)$ is monoid: the local divisor at c.

Facts

- $\lambda_c : \{x \in M \mid cx \in Mc\} \to M_c$ given by $\lambda_c(x) = cx$ is a surjective homomorphism. Hence, M_c is a divisor of M.
- If c is a unit, then M_c is isomorphic to M.

Local divisor technique

- The local divisor technique is the "algebraic child" of an induction in Thomas Wilke's habilitation for a simplified proof of $\mathrm{FO} = \mathrm{LTL}$ over finite words, see STACS 1999 and this ICALP.
- Let M be a monoid and $c \in M$. Consider the set $cM \cap Mc$ and define a new multiplication

$$xc \circ cy = xcy.$$

Then $M_c = (cM \cap Mc, \circ, c)$ is monoid: the local divisor at c.

Facts

- $\lambda_c : \{x \in M \mid cx \in Mc\} \to M_c$ given by $\lambda_c(x) = cx$ is a surjective homomorphism. Hence, M_c is a divisor of M.
- If c is a unit, then M_c is isomorphic to M.
- If $c = c^2$, then M_c is the standard "local monoid".

Local divisor technique

- The local divisor technique is the "algebraic child" of an induction in Thomas Wilke's habilitation for a simplified proof of $\mathrm{FO} = \mathrm{LTL}$ over finite words, see STACS 1999 and this ICALP.
- Let M be a monoid and $c \in M$. Consider the set $cM \cap Mc$ and define a new multiplication

$$xc \circ cy = xcy.$$

Then $M_c = (cM \cap Mc, \circ, c)$ is monoid: the local divisor at c.

Facts

- $\lambda_c : \{x \in M \mid cx \in Mc\} \to M_c$ given by $\lambda_c(x) = cx$ is a surjective homomorphism. Hence, M_c is a divisor of M.
- If c is a unit, then M_c is isomorphic to M.
- If $c = c^2$, then M_c is the standard "local monoid".
- If c is not a unit, then $1 \notin M_c$. Hence, if c is not a unit and if M is finite, then $|M_c| < |M|$.

Main steps in showing $\overline{\mathbf{H}}(A^*) \subseteq \mathrm{SD}_{\mathbf{H}}(A^*)$

- Starting point: L recognized by $\varphi : A^* \to M \in \overline{\mathbf{H}}$.
- Wlog. M is not a group: Choose $c \in A$ s.t. $\varphi(c)$ is not a unit.

Main steps in showing $\overline{\mathbf{H}}(A^*) \subseteq \mathrm{SD}_{\mathbf{H}}(A^*)$

- Starting point: L recognized by $\varphi : A^* \to M \in \overline{\mathbf{H}}$.
- Wlog. M is not a group: Choose $c \in A$ s.t. $\varphi(c)$ is not a unit.
- We write $B = A \setminus \{c\}$, $[w] = \varphi^{-1}(w)$ and $T = \{[u] \mid u \in B^*\}$.

Main steps in showing $\overline{\mathbf{H}}(A^*) \subseteq \mathrm{SD}_{\mathbf{H}}(A^*)$

- Starting point: L recognized by $\varphi : A^* \to M \in \overline{\mathbf{H}}$.
- Wlog. M is not a group: Choose $c \in A$ s.t. $\varphi(c)$ is not a unit.
- We write $B = A \setminus \{c\}$, $[w] = \varphi^{-1}(w)$ and $T = \{[u] \mid u \in B^*\}$.
- Wlog. $w = u_0 c u_1 \dots c u_{n+1}$.
- New homomorphism: $\psi : T^* \to M_c, \quad [u] \mapsto \varphi(cuc)$.

Main steps in showing $\overline{\mathbf{H}}(A^*) \subseteq \mathrm{SD}_{\mathbf{H}}(A^*)$

- Starting point: L recognized by $\varphi : A^* \to M \in \overline{\mathbf{H}}$.
- Wlog. M is not a group: Choose $c \in A$ s.t. $\varphi(c)$ is not a unit.
- We write $B = A \setminus \{c\}$, $[w] = \varphi^{-1}(w)$ and $T = \{[u] \mid u \in B^*\}$.
- Wlog. $w = u_0 c u_1 \ldots c u_{n+1}$.
- New homomorphism: $\psi : T^* \to M_c, \quad [u] \mapsto \varphi(cuc)$.

Then

$$\begin{aligned}\psi([u_1]\cdots[u_n]) &= \varphi(cu_1c) \circ \cdots \circ \varphi(cu_nc) \\ &= \varphi(cu_1c\cdots cu_nc) = \varphi(c) \cdot \varphi(u_1c\cdots u_nc).\end{aligned}$$

Main steps in showing $\overline{\mathbf{H}}(A^*) \subseteq \mathrm{SD}_{\mathbf{H}}(A^*)$

- Starting point: L recognized by $\varphi : A^* \to M \in \overline{\mathbf{H}}$.
- Wlog. M is not a group: Choose $c \in A$ s.t. $\varphi(c)$ is not a unit.
- We write $B = A \setminus \{c\}$, $[w] = \varphi^{-1}(w)$ and $T = \{[u] \mid u \in B^*\}$.
- Wlog. $w = u_0 c u_1 \ldots c u_{n+1}$.
- New homomorphism: $\psi : T^* \to M_c, \quad [u] \mapsto \varphi(cuc)$.

Then

$$\begin{aligned}\psi([u_1] \cdots [u_n]) &= \varphi(cu_1c) \circ \cdots \circ \varphi(cu_nc) \\ &= \varphi(cu_1c \cdots cu_nc) = \varphi(c) \cdot \varphi(u_1c \cdots u_nc).\end{aligned}$$

Define $\sigma : (B^*c)^* \to T^*$ with $\sigma(uc) = [u]$. Then

$$\forall w \in A^* : \varphi(cwc) = \psi\sigma(wc).$$

"Essentially" it remains to show

$$\sigma^{-1}(\mathrm{SD}_{\mathbf{H}}(T^*)) \subseteq \mathrm{SD}_{\mathbf{H}}(A^*).$$

Other results

1. $\mathrm{SD}_{\mathbf{H}} = \overline{\mathbf{H}}$ also holds over infinite words.

Other results

1. $\mathrm{SD}_{\mathbf{H}} = \overline{\mathbf{H}}$ also holds over infinite words.
2. Local divisor technique yields a Rees matrix decomposition result:

Other results

1. $\mathrm{SD}_{\mathbf{H}} = \overline{\mathbf{H}}$ also holds over infinite words.
2. Local divisor technique yields a Rees matrix decomposition result:
   - Every monoid is a Rees extension of a proper submonoid and a local divisor.
   - Thus, every monoid is an iterated Rees extension of subgroups.
   - $\overline{\mathbf{H}} = \mathrm{Rees}(\mathbf{H})$.
   - solves open problem of Almeida and Klíma: $\mathbf{V}$ is closed under Rees extensions if and only if $\mathbf{V} = \overline{\mathbf{H}}$.

Other results

1. $\mathrm{SD}_{\mathbf{H}} = \overline{\mathbf{H}}$ also holds over infinite words.
2. Local divisor technique yields a Rees matrix decomposition result:
   - Every monoid is a Rees extension of a proper submonoid and a local divisor.
   - Thus, every monoid is an iterated Rees extension of subgroups.
   - $\overline{\mathbf{H}} = \mathrm{Rees}(\mathbf{H})$.
   - solves open problem of Almeida and Klíma: $\mathbf{V}$ is closed under Rees extensions if and only if $\mathbf{V} = \overline{\mathbf{H}}$.

Thank you