Operations preserving regular languages*

J. Berstel[1], L. Boasson[2], O. Carton[2], B. Pettazoni[3], J.-E. Pin[2]

[1] IGM, Université de Marne-la-Vallée

[2] LIAFA, Université de Paris VII

[3] Lycée M. Berthelot, Saint-Maur

*Presented at FCT'2003, Malmö.

Filters

Filter: increasing sequence $(s_n)_{n\geq 0}$ of integers

Example: $s = 0, 1, 4, 9, 16, 25, \dots$

Filtering a word $w = a_0 \cdots a_n$ by s yields

$$w[s] = a_{s_0} a_{s_1} \cdots a_{s_k} \quad \text{where} \quad s_k \leq n < s_{k+1}$$

Example: $w = abracadabra$

w	a	b	r	a	c	a	d	a	b	r	a
s	0	1			4					9	
$w[s]$	a	b			c					r	

$$w[s] = abcr$$

Filtering a set $L \subset A^*$ of words : $L[s] = \{w[s] \mid w \in L\}$

Some examples

Let $L = (ab)^*$.

$$s_n = 3n + 1 \qquad abababababababababababa \cdots$$

$$s_n = n^2 \qquad abababababababababababab \cdots$$

$$s_n = n! \qquad abababababababababababababa \cdots$$

$$s_n = n(n+1) \qquad abababababababababa \cdots$$

Filtering problem

A filter preserves regular sets if, for any regular language L, the language $L[s]$ regular.

Problem: characterize filters preserving regular sets.

Regulator : A relation $R : A^* \to B^*$ such that $R(L)$ is regular for every regular L.

Examples: the following filters are regulators:

- $\{2n \mid n \geq 0\}$, (it is a rational transduction)
- $\{n^2 \mid n \geq 0\}$, (! it is not a rational transduction)
- $\{2^n \mid n \geq 0\}$, (!!)
- $\{n! \mid n \geq 0\}$. (!?!)

But $\{\binom{2n}{n} \mid n \geq 0\}$ **is not** a regulator.

A counter-example

Let $L = (ab)^*$. Let s be the filter with support

$\mathbb{N} \setminus \{n(n+1) \mid n \geq 0\} = \{1, 3, 4, 5, 7, 8, 9, 10, 11, 13, \ldots\}$

$$ababababababababababababababababa \cdots$$

$L[s]$ is the set of prefixes of the infinite word

$$b(ab)^0 b(ab)^1 b(ab)^2 b(ab)^3 \cdots$$

and $L[s]$ is not regular. Thus s is not a regulator.

Ultimately periodic sequences

- A sequence s is ultimately periodic modulo p if the sequence $s_n \mod p$ is ultimately periodic.

- A sequence s is ultimately periodic with threshold t if the sequence $\min(s_n, t)$ is ultimately periodic.

The sequence

$$010\mathbf{2}010\mathbf{3}010201\mathbf{0}4010201030102010\mathbf{5}\cdots$$

is ultimately periodic with threshold t, for each t.

The sequence s where s_n is the number of 1's in the binary expansion of n

$$011\mathbf{1}223\mathbf{1}2232334\mathbf{1}223\cdots$$

is not ultimately periodic with threshold 1.

Residually ultimately periodic sequences

A sequence s is residually ultimately periodic (r.u.p.) if it is both

- ultimately periodic modulo p for each $p > 0$,
- ultimately periodic with threshold t for each $t \geq 0$.

Proposition 1 *A sequence s is r.u.p. iff, for each morphism φ from $\mathbb{N}$ onto a finite semigroup, the sequence $\varphi(s_n)$ is ultimately periodic.*

Solution of the filtering problem

Theorem 2 *A filter* $(s_n)_{n\geq 0}$ *preserves regular sets iff the sequence* $\partial s_n = s_{n+1} - s_n$ *is residually ultimately periodic.*

The sequence $\partial s_n = s_{n+1} - s_n$ is the differential of s. A sequence s is differentially residually ultimately periodic (d.r.u.p.) if ∂s is r.u.p.

Properties of r.u.p. sequences

Theorem 3 (Zhang 98, Carton-Thomas 02) *Let* $(u_n)_{n\geq 0}$ *and* $(v_n)_{n\geq 0}$ *be r.u.p. sequences. The following seuquences are also r.u.p.:*

- u_{v_n} *(composition),* $u_n + v_n$, $u_n v_n$, $u_n^{v_n}$,
- $u_n - v_n$ *provided* $u_n \geq v_n$ *and* $\lim_{n\to\infty}(u_n - v_n) = +\infty$,
- *(generalized sum)* $\sum_{0\leq i\leq v_n} u_i$,
- *(generalized product)* $\prod_{0\leq i\leq v_n} u_i$.

Examples of r.u.p. sequences

- The sequences n^k and k^n (for fixed k).
- The exponential tower $k^{k^{k^{\cdot^{\cdot^{\cdot^{k}}}}}}$ of height n.
- The family of r.u.p. is not closed under quotient. Indeed, define
$$u_n = \begin{cases} 1 & \text{if } n \text{ is prime} \\ n!+1 & \text{otherwise} \end{cases}.$$
Then u_n is not r.u.p., but nu_n is r.u.p.
- For any (even non recursive) strictly increasing function $\varphi : \mathbb{N} \to \mathbb{N}$, the sequence $u_n = n!\varphi(n)$ is r.u.p. non recursive.
- If $\lim_{n\to\infty} u_n = +\infty$, then u is ultimately periodic with threshold t for each $t \geq 0$.

R.u.p. and d.r.u.p.

A sequence s is d.r.u.p. if its sequence of differences ∂s is r.u.p..

- D.r.u.p. sequences have closure properties very similar to r.u.p. sequences.
- Every d.r.u.p. sequence is r.u.p.
- There are r.u.p. sequences which are not d.r.u.p.

  Let b_n be a sequence of 0 and 1's which is not ultimately periodic. Then b_n is not r.u.p. because it is not ultimately periodic with threshold 1.

  The sequence $u_n = (\sum_{0 \leq i \leq n} b_i)!$ is r.u.p. but ∂u_n is not r.u.p. because $\min(\partial u)_n, 1) = b_n$.
- If s r.u.p. and $\lim \partial s_n = \infty \Rightarrow s$ then it is d.r.u.p.

Sequences which are not r.u.p.

- Spectra: $\{\lfloor \alpha n \rfloor \mid n \geq 1\}$ for irrational α.
- The sequence of positions of 1's in the Thue-Morse sequence.
- The sequence of Catalan numbers.

A filter preserving regular sets is drup

Proposition 4 *A filter s preserving regular sets is ultimately periodic for each $p > 0$.*

Let $A = \{0, 1, \ldots, p-1\}$. Set

$$x = (01 \cdots (p-1))^\omega$$

so $x(i) \equiv i \pmod p$, and set

$$y = x[s] = x(s_0)x(s_1) \cdots x(s_i) \cdots .$$

At position i, one gets

$$y(i) = x(s_i) \equiv s_i \pmod p.$$

Let L be the set of prefixes of x. Then L is regular. The set $L[s]$ is the set of prefixes of y. It is regular only if y is ultimately periodic. Thus s is ultimately periodic modulo p.

A filter preserving regular sets is drup (2)

Proposition 5 *If a filter s preserves regular sets, then ∂s is ultimately periodic with threshold t for each $t \geq 0$.*

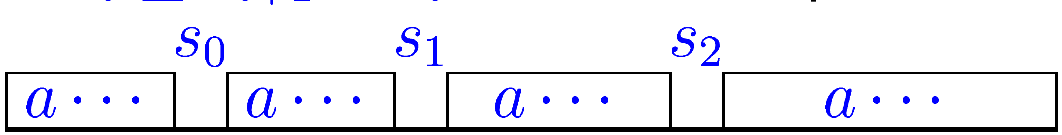
Set $d_i = \min(t, s_{i+1} - s_i - 1)$. We show that the infinite word $d = d_0 d_1 \cdots$ is ultimately periodic.
Define a prefix code over $B = \{0, 1, \ldots, t\} \cup \{a\}$ by

$$P = \{0, 1a, 2a^2, \ldots, ta^t, a\}$$

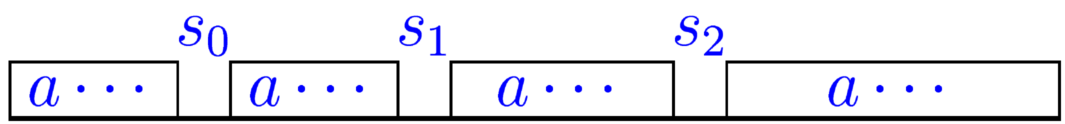
The language $P^*[s]$ is regular, and so is $R = P^*[s] \cap \{0, 1, \ldots, t\}^*$.

$$d = x[s]$$

for the word x defined by $x(s_i) = d_i$ and $x(m) = a$ if $m \neq d_i$, for $i \geq 0$.
$x \in P^\omega$ because $d_i \leq s_{i+1} - s_i - 1$. So each prefix of d is in R.



A filter preserving regular sets is drup (3)



$$x = a^{s_0} d_0 a^{s_1 - s_0 - 1} d_1 a^{s_2 - s_1 - 1} \cdots d_i a^{s_{i+1} - s_i - 1} \cdots$$

The word $d_0 d_1 \cdots d_{n-1}$ is the maximal word of length n in R (for the order $0 < 1 < \cdots < t$). Indeed, if $d_i < d_i'$ then $d_i < t$, so $d_i = s_{i+1} - s_i - 1$ and d_i' is not followed by d_i' letters a.

The word d is read in a trim automaton recognizing R by taking at each state the edge with maximal label. Thus it is ultimately periodic.

Transductions

Transductions are relations from A^* into B^* and later into some monoid M.

Inverse filtering transduction: Let $s = (s_n)_{n\geq 0}$ be a sequence of integers. Define τ_s

$$\tau_s(a_0 \cdots a_n) = A^{s_0} a_0 A^{s_1 - s_0 - 1} \ldots A^{s_n - s_{n-1} - 1} a_n A^{\leq s_{n+1} - s_n - 1}$$

One has

$$L[s] = \tau_s^{-1}(L).$$

Transducers

Let A be an alphabet and M be a monoid.

states, initial states

$$\mathcal{T} = (Q, A \times \mathfrak{P}(M), E, I, F)$$

transitions, final states

Transitions: $q \xrightarrow{a|R} q'$ where $a \in A$ and $R \in \mathfrak{P}(M)$.

Initial and final labels: The entries of the vectors $I, F \in \mathfrak{P}(M)^Q$.

A transducer realizes a transduction τ from A^* to M defined as follows.

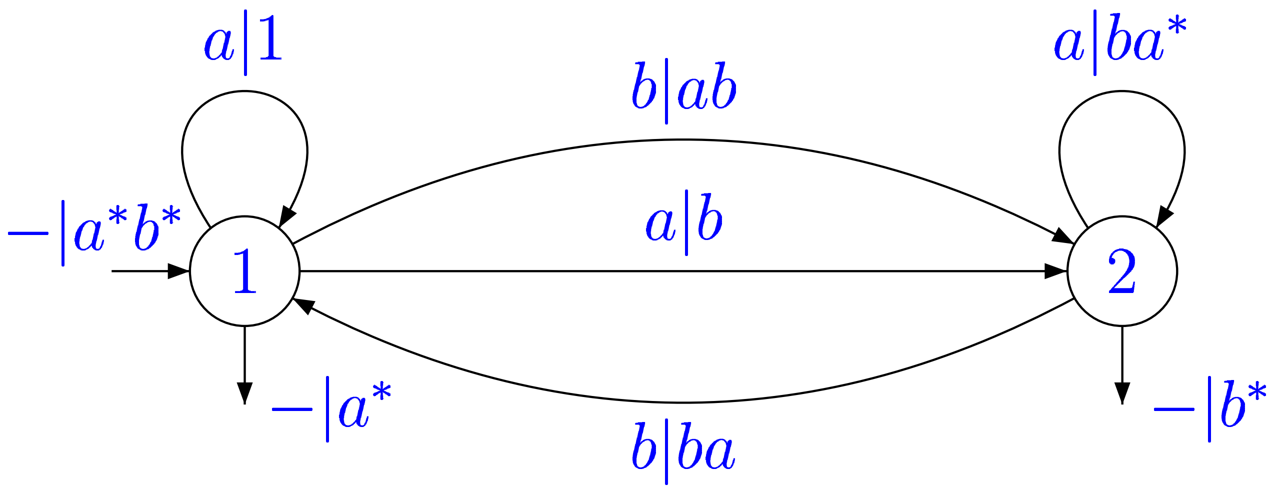
For $w = a_1 \cdots a_n$,

$\tau(w)$ is the union of all products $I_0 R_1 \cdots R_n F_n$ for all paths

$$\xrightarrow{I_0} q_0 \xrightarrow{a_1|R_1} q_1 \xrightarrow{a_2|R_2} q_2 \quad \cdots \quad q_{n-1} \xrightarrow{a_n|R_n} q_n \xrightarrow{F_n}$$

A transducer

$$\tau(ab) = a^*b^*(ab \cdot b^* \cup b \cdot ba \cdot a^*)$$



Rational and recognizable sets

Let M be a monid.

$\mathrm{Rat}(M)$ denotes the set of rational subsets of M obtained from the singletons using the operations union, product and star.

$\mathrm{Rec}(M)$ denotes the set of recognizable subsets of M, that is subsets P of M for which there exists a morphism φ of M onto a finite monoid F, and a subset Q of F such that $P = \varphi^{-1}(Q)$.

Rational transductions

A transduction is rational if it can be realized by a finite transducer with output labels that are rational subsets of M.

Theorem 6 *Let τ be a rational transduction from A^* to M. If K is a regular language over A, then $\tau(K)$ is rational subset of M. If L is a recognizable subset of M, then $\tau^{-1}(L)$ is a regular language over A.*

In order to show that d.r.u.p. filters preserve regular sets, it would be sufficient to show that the inverse filtering transduction is a rational transduction.

However, the inverse filtering transduction is **not** rational.

Residually rational transductions

A transduction τ from A^* to B^* is residually rational if for any morphism μ from B^* into a **finite** monoid M, $\mu \circ \tau$ is rational.

$$\begin{tikzcd} A^* \arrow[r, "\tau"] \arrow[dr, "\mu \circ \tau"'] & B^* \arrow[d, "\mu"] \\ & M \end{tikzcd}$$

Theorem 7 *If τ is residually rational and $L \subset B^*$ is regular, then $\tau^{-1}(L)$ is also regular, i.e. τ^{-1} is a regulator.*

Proof. Let $\mu : B^* \to M$ be the syntactic morphism of L. Then

$$\tau^{-1}(L) = (\mu \circ \tau)^{-1}(P).$$

where $P = \mu(L)$.

Residually rational transductions (2)

Theorem 8 *A transduction* $\tau : A^* \to B^*$ *is residually rational if and only if* τ^{-1} *is a regulator.*

Inverse of filtering transduction

Proposition 9 *Let* s *be a d.r.u.p. sequence. Then the inverse* τ_s *of the corresponding filtering transduction is residually rational (and consequently the filtering transduction of* s *is a regulator).*

$$\tau_s(a_0 \cdots a_n) = A^{s_0} a_0 A^{d_1} a_1 \cdots a_{n-1} A^{d_n} a_n (1 + A)^{d_{n+1}}$$

where $d_n = s_{n+1} - s_n - 1$.
Let $R = \mu(A)$ be the image of A in a finite monoid M. Since $\mathfrak{P}(M)$ is finite, there r and q such that

$$R^r = R^{r+q}.$$

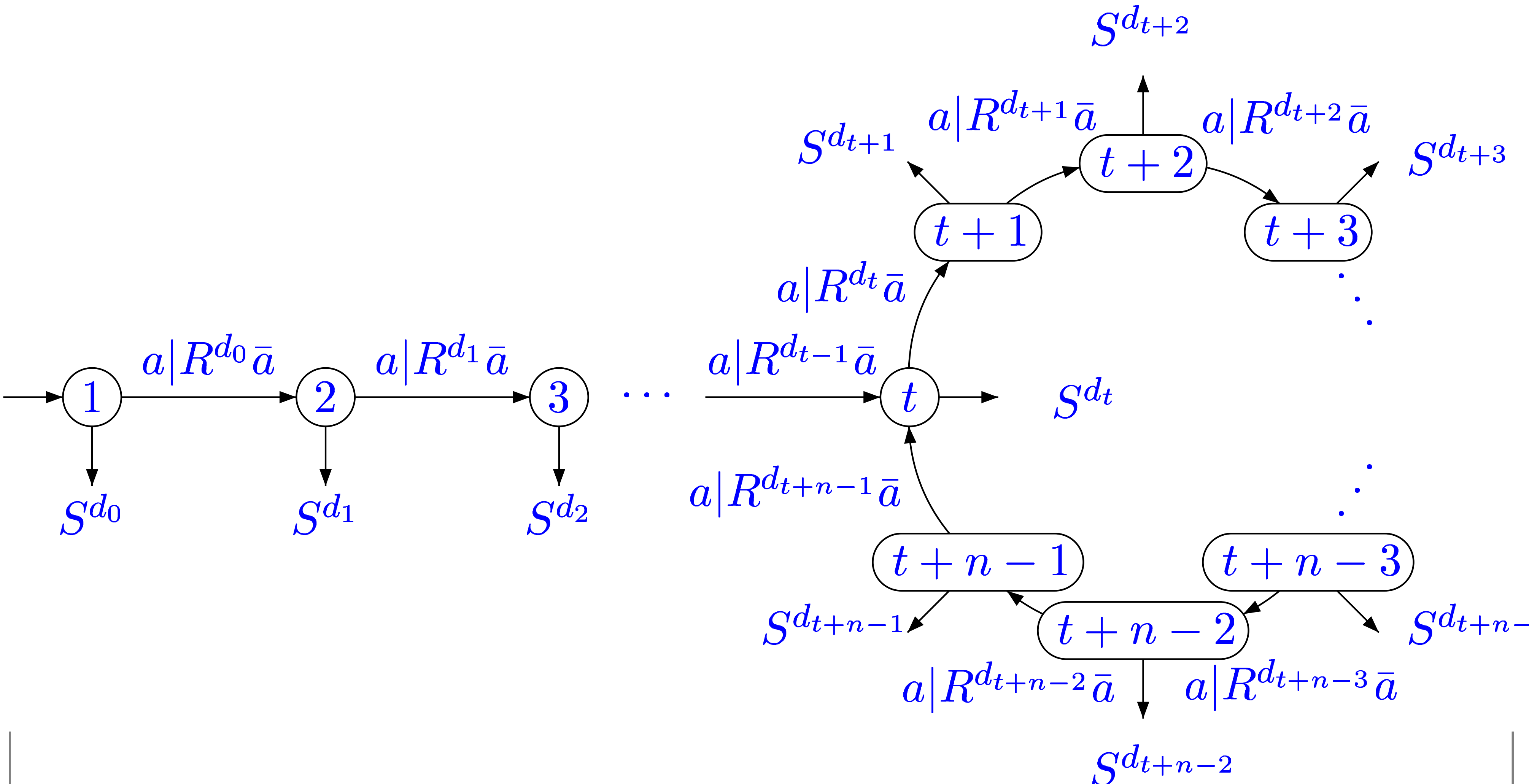
Since $(d_n)_{n \geq 0}$ is residually ultimately periodic, there are t and p such that

$$R^{d_n} = R^{d_{n+p}} \quad \text{for every} \quad n \geq t.$$

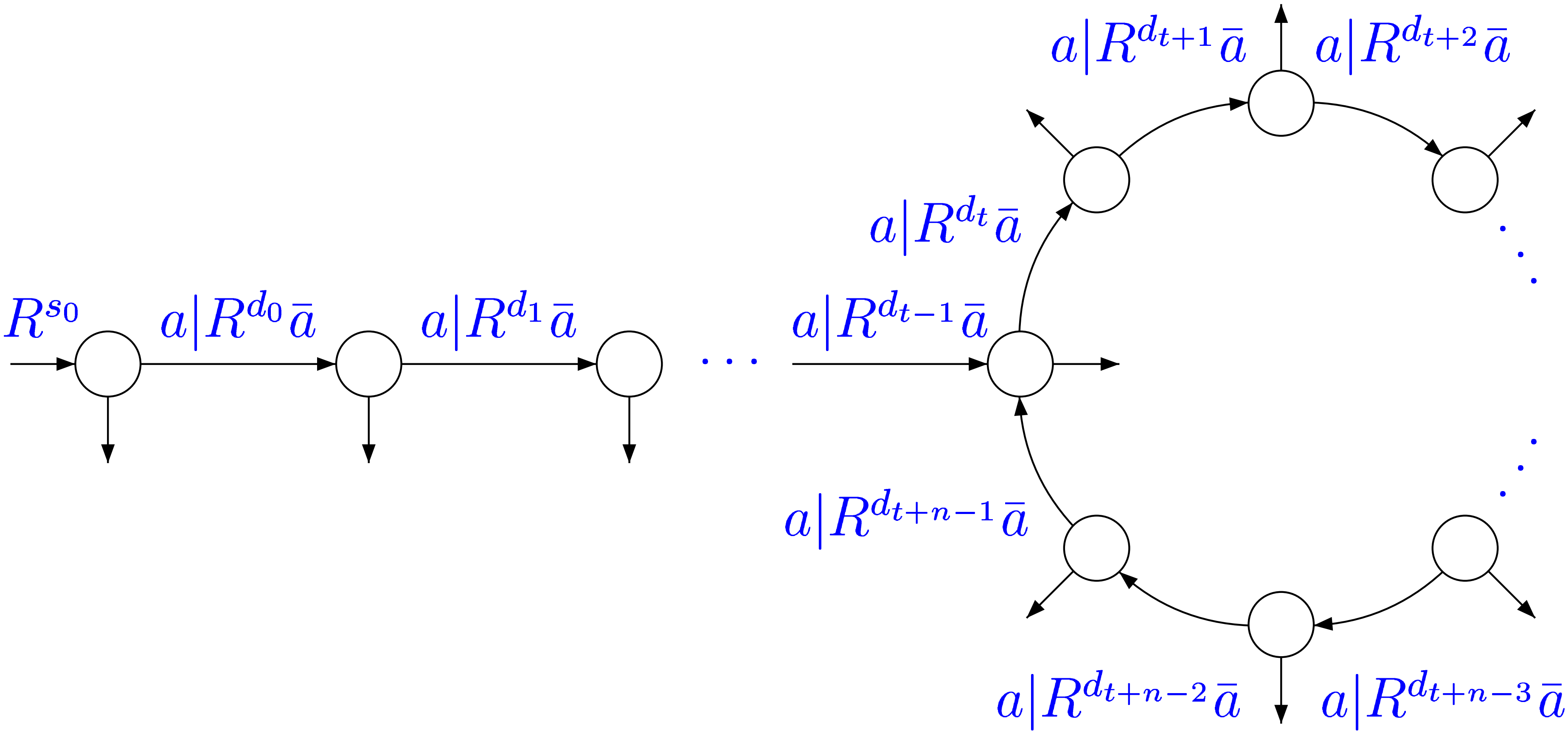
Thus, $\mu \circ \tau_s$ is realized by the following transducer:

Filtering transducer

A transducer realizing $\mu \circ \tau_s$. Here $S = 1 + R = \mu(1 + A)$.



Filtering transducer



Removal problem

Let S be a relation over $\mathbb{N}$ and $L \subset A^*$. Define

$$L/S = \{u \mid \exists v \;\; (|u|, |v|) \in S \text{ and } uv \in L\}$$

Example : Let $S = \{(n, n) \mid n \in \mathbb{N}\}$. Then L/S is the set of first halves of words in L.

A relation S of $\mathbb{N}^2$ is said to preserve recognizable sets over $\mathbb{N}$ if, for any recognizable $K \subseteq \mathbb{N}$, the set $S(K)$ is recognizable over $\mathbb{N}$ (i.e. a finite union of arithmetic progressions and of a finite set).

Theorem 10 (Seiferas, McNaughton)
L/S is recognizable for any recognizable set L iff S^{-1} preserves recognizable sets over $\mathbb{N}$.

Removal transduction

Proposition 11 *If* S *preserves recognizable sets over* $\mathbb{N}$*, then the inverse of the removal transduction is residually rational.*

The inverse of the removal transduction is defined by

$$\tau_S(u) = \bigcup_{(|u|,m)\in S} uA^m.$$