A Note on Star-Free Events

ALBERT R. MEYER

Carnegie-Mellon University, Pittsburgh, Pennsylvania*

ABSTRACT. It is shown that a short proof of the equivalence of star-free and group-free regular events is possible if one is willing to appeal to the Krohn-Rhodes machine decomposition theorem.

KEY WORDS AND PHRASES: regular event, regular set, automata, finite automata, decomposition, star-free, group-free, noncounting event, cascade product, semigroup, homomorphism

CR CATEGORIES: 5.22

1. *Introduction*

The star-free events are the family of regular events expressible in the extended language of regular expressions (using intersection and complementation as well as union and concatenation of events) without the use of the Kleene star (closure) operator. The equivalence of the star-free and group-free events was first proved by Schützenberger, 1966 [5]. Papert and McNaughton, 1966, [3] show that the star-free events are precisely those events definable in McNaughton's L-language, and they are thereby able to establish the above equivalence without extensive use of the properties of finite semigroups. However, if one is willing to appeal to the machine decomposition theorem of Krohn and Rhodes, the equivalence of star-free, group-free, and also noncounting regular events can be proved more simply.[1] We present such a proof in this paper.

2. *Preliminaries*

We assume the reader is already familiar with regular events and finite automata. Our notation follows Yoeli, 1965, [6] and Ginzburg, 1968, [2]. In particular, if f and g are functions from a set S into itself, arguments are written on the left (so that $sf = f(s)$), and the composition $f \circ g$ means that f is applied first (so that $s(f \circ g) = (sf)g$).

A *semiautomaton* is a triple $A = \langle Q^A, \Sigma^A, M^A \rangle$ with Q^A a finite set (of *states*), Σ^A a finite set (of *inputs*), and M^A a set of functions $M_\sigma{}^A$: $Q^A \rightarrow Q^A$ indexed by $\sigma \in \Sigma^A$. The mapping $M_\sigma{}^A$ is abbreviated σ^A. The element $q\sigma^A \in Q^A$ is the *next state* of $q \in Q^A$ under input $\sigma \in \Sigma^A$. For $x \in (\Sigma^A)^*$ the mapping x^A: $Q^A \rightarrow Q^A$

* Department of Computer Science. This work was supported by the Advanced Research Projects Agency of the Office of the Secretary of Defense (SD-146) and is monitored by the Air Force Office of Scientific Research.

[1] Several researchers in algebraic automata theory have noted that the above equivalence could be proved using the theorem of Krohn and Rhodes, although to the author's knowledge none have actually presented such a proof. Shortly after this paper was written, a version of Lemma 4 was stated in an abstract by Cohen and Brzozowski, 1968 [1].

is defined inductively: Λ^A is the identity map on Q^A where Λ is the null word in $(\Sigma^A)^*$, and if $x = y\sigma$ for $y \in (\Sigma^A)^*$ and $\sigma \in \Sigma^A$, then x^A is $y^A \circ \sigma^A$. Hence $(xy)^A = x^A \circ y^A$ for all $x,y \in (\Sigma^A)^*$. For $x \in (\Sigma^A)^*$ and integers $k \geq 0$, x^k is the concatenation of x with itself k times; $x^0 = \Lambda$ by convention. Clearly $(x^k)^A = (x^A)^k$ is the composition of x^A with itself k times. The (necessarily finite) set of distinct mappings x^A for $x \in (\Sigma^A)^*$ forms a semigroup G^A under composition. G^A is called the *semigroup of A*.

Let A and B be semiautomata. B is a *subsemiautomaton* of A providing $\Sigma^B \subset \Sigma^A$, $Q^B \subset Q^A$, and the mapping σ^B is the restriction of σ^A to Q^B for each $\sigma \in \Sigma^B$. B is a *homomorphic image* of A providing $\Sigma^A = \Sigma^B$ and there is an onto mapping η: $Q^A \rightarrow Q^B$ such that $\eta \circ \sigma^B = \sigma^A \circ \eta$ for each $\sigma \in \Sigma^A$. The mapping η is called a *homomorphism* of A onto B. A covers B, i.e. "$A \geq B$", if and only if B is a homomorphic image of a subsemiautomaton of A.

An *automaton* is a quintuple $\hat{A} = \langle Q^A, \Sigma^A, s^A, F^A, M^A \rangle$ where $A = \langle Q^A, \Sigma^A, M^A \rangle$ is called the *semiautomaton* of $\hat{A}$, s^A is an element of Q^A called the *start state*, and F^A is a subset of Q^A called the *final states*. The *event accepted by* $\hat{A}$ is $\{x \in (\Sigma^A)^* \mid s^A x^A \in F^A\}$. This definition of automaton is merely a notational variant of the usual finite state acceptor (cf. Rabin and Scott, 1959, [4]), and the events accepted by such automata are precisely the regular events.

3. *Star-Free and Noncounting Events*

The star-free events are defined inductively as follows:

Definition 1. Let Σ be a finite set (of inputs). The singleton $\{\sigma\}$ is a star-free event over Σ. If $U,V \subset \Sigma^*$ are star-free events over Σ, then $U \cup V$, $\bar{U}$ (the complement of U relative to Σ^*), and UV (the concatenation of U and V) are star-free events over Σ. An event is star-free over Σ only by implication from the preceding clauses.

By DeMorgan's law $U \cap V = \overline{\bar{U} \cup \bar{V}}$; so star-free events are also closed under intersection. Since the regular events over Σ include the singletons and are closed under union, relative complementation, and concatenation, it follows that every star-free event is regular.

Definition 2. (Papert-McNaughton) A regular event $U \subset \Sigma^*$ is a *noncounting* regular event over Σ if and only if there is an integer $k_U \geq 0$ such that for all $x,y,z \in \Sigma^*$

$$xy^{k_U}z \in U \Leftrightarrow xy^{k_U+1}z \in U.$$

Intuitively, an automaton accepting a noncounting event U need never count (even modulo any integer greater than one) the number of consecutive occurrences of any word y once k_U consecutive y's have occurred in an input word.

Lemma 1. (Papert-McNaughton) *Every star-free event is a noncounting regular event.*

Proof. The singleton $\{\sigma\}$ is trivially a noncounting regular event for every $\sigma \in \Sigma$ (choose $k_{\{\sigma\}} = 2$), so it is sufficient to show that if U and V are noncounting regular events over Σ, then so are $U \cup V$, $\bar{U}$, and UV.

Let $k = \max\{k_U, k_V\}$. Then for any $x, y, z \in \Sigma^*$,

$$xy^kz \in U \cup V \Leftrightarrow xy^{k_U}(y^{k-k_U}z) \in U$$

or

$$xy^{k_V}(y^{k-k_V}z) \in V \Leftrightarrow xy^{k_U+1}(y^{k-k_U}z) \in U$$

or

$$xy^{k_V+1}(y^{k-k_V}z) \in V \Leftrightarrow xy^{k+1}z \in U \cup V.$$

Thus $U \cup V$ is a noncounting regular event with $k_{U \cup V} = \max\{k_U, k_V\}$.

Similarly,

$$xy^{k_U}z \in \bar{U} \Leftrightarrow xy^{k_U}z \notin U \Leftrightarrow xy^{k_U+1}z \notin U \Leftrightarrow xy^{k_U+1}z \in \bar{U},$$

so that $\bar{U}$ is a noncounting regular event with $k_{\bar{U}} = k_U$.

Finally, let $k = k_U + k_V + 1$ and suppose $xy^kz \in UV$. Then $xy^kz = uv$ for some $u \in U$ and $v \in V$, and it must be the case that either $u = xy^{k_U}w$ for some $w \in \Sigma^*$, or that $v = w'y^{k_V}z$ for some $w' \in \Sigma^*$. In the first case it follows that $xy^{k_U+1}w \in U$, and in the second case $w'y^{k_V+1}z \in V$. Hence in either case $xy^{k+1}z \in UV$. Conversely, if $xy^{k+1}z \in UV$, the argument can clearly be reversed to show that $xy^kz \in UV$. Thus UV is a noncounting regular event with $k_{UV} = k_U + k_V + 1$. Q.E.D.

If U is a noncounting regular event over Σ and $\sigma \in \Sigma$, then $\sigma^{k_U} \in U$ implies that U contains all words in σ^* of length at least k_U. Therefore either $\bar{U} \cap \sigma^*$ or $U \cap \sigma^*$ is a finite event. The regular event $(\sigma\sigma)^*$ is neither finite nor has finite complement, which proves:

Corollary 1. *The noncounting (and hence the star-free) regular events are a proper subfamily of the regular events.*

4. *Group-Free Events*

Associated with any event $U \subset \Sigma^*$ is a congruence relation, $\equiv \pmod U$, on Σ^* defined for $w,y \in \Sigma^*$ as:

$$w \equiv y \pmod U \Leftrightarrow (\forall x,z \in \Sigma^*)(xwz \in U \Leftrightarrow xyz \in U).$$

Noncounting regular events are thus those regular events U such that $y^{k_U} \equiv y^{k_U+1} \pmod U$ for all $y \in \Sigma^*$.

The relation between this congruence and automata is an immediate consequence of the familiar theorems of Nerode and Myhill (cf. Rabin and Scott, 1959 [4]): if U is a regular event, then there is an automaton $\hat{A}$ accepting U (viz., the reduced automaton accepting U) such that $x \equiv y \pmod U \Leftrightarrow x^A = y^A$.

Definition 3. A *subgroup of a semigroup* S is a subsemigroup of S whose elements form an abstract group under multiplication in S. A semigroup is *group-free* if and only if all its subgroups are isomorphic to the trivial group with one element. A semiautomaton is *group-free* if and only if the semigroup of the semiautomaton is group-free. A regular set U is *group-free* if and only if there is an automaton $\hat{A}$ accepting U such that the semiautomaton A of $\hat{A}$ is group-free.

Lemma 2. *Let S be a semigroup. If there is an integer $k \geq 0$ such that $s^k = s^{k+1}$ for all $s \in S$, then S is group-free.*

Proof. Let G be a subgroup of S, and let g be an element of G. Then $g^k = g^{k+1}$ implies $e = g^k(g^{-1})^k = g^{k+1}(g^{-1})^k = g$ where g^{-1} is the inverse of g in G and e is the identity in G. Hence $G = \{e\}$ is the trivial group. Q.E.D.

Corollary 2. *Every noncounting regular event (and thus every star-free event) is a group-free regular event.*

PROOF. If U is a noncounting regular event, then $y^{k_U} \equiv y^{k_U+1} \pmod{U}$ implies that $(y^{k_U})^A = (y^{k_U+1})^A$ in the reduced automaton $\hat{A}$ accepting U. Hence $(y^A)^{k_U} = (y^A)^{k_U+1}$ for every element $y^A \in G^A$, and G^A is group-free by Lemma 2. Q.E.D.

5. *Decomposition into Resets*

The machine decomposition theorem of Krohn and Rhodes supplies the key step in the proof that group-free events are star-free.

Definition 4. Let A and B be semiautomata and ω: $Q^A \times \Sigma^A \to \Sigma^B$. The *cascade product* of A and B with mapping ω, $A \mathring{\omega} B$, is the semiautomaton C with $Q^C = Q^A \times Q^B$, $\Sigma^C = \Sigma^A$, and σ^C for $\sigma \in \Sigma^C$, defined for all $s^A \in Q^A$, $s^B \in Q^B$ by:

$$\langle s^A, s^B \rangle \sigma^C = \langle s^A \sigma^A, s^B ((\langle s^A, \sigma \rangle \omega)^B) \rangle.$$

A cascade product of three or more automata is defined by association to the left; e.g. a cascade product of semiautomata A, B, and C is any semiautomaton $(A \mathring{\omega}_1 B) \mathring{\omega}_2 C$ for any mappings ω_1 and ω_2 with appropriate domain and range.

Definition 5. A semiautomaton R is a *reset* providing $Q^R = \{1, 2\}$ and Σ^R is the union of three mutually exclusive sets Σ_1^R, Σ_2^R, Σ_I^R such that: $\sigma \in \Sigma_1^R \Leftrightarrow$ *range* $(\sigma^R) = \{1\}$; $\sigma \in \Sigma_2^R \Leftrightarrow$ range $(\sigma^R) = \{2\}$; and $\sigma \in \Sigma_I^R \Leftrightarrow \sigma^R$ is the identity on Q^R.

The following weak form of the decomposition theorem is sufficient for our purposes (for a constructive proof of the general theorem see Ginzburg, 1968 [2].

THEOREM. (*Krohn-Rhodes*) *Every semiautomaton A is covered by a cascade product of semiautomata A_1, A_2, $\cdots$, A_n such that for $1 \leq i \leq n$, A_i is a reset, or else G^{A_i} is a nontrivial homomorphic image of a subgroup of G^A.*

Since the trivial group has only itself as a homomorphic image, the following lemma is immediate:

LEMMA 3. *Every group-free semiautomaton is covered by a cascade product of resets.*

COROLLARY 3. *Every group-free regular event is accepted by an automaton whose semiautomaton is a cascade product of resets.*

PROOF. Let $\hat{A}$, with group-free semiautomaton A, be an automaton accepting a group-free regular event U. By Lemma 3 and the definition of covering, A is the image under a homomorphism η of a subsemiautomaton of a cascade product C of resets. There is no loss of generality in assuming that $\Sigma^A = \Sigma^C$ since the subsemiautomaton of C obtained by restricting Σ^C to Σ^A is also a cascade product of resets which covers A. Choose any $s^C \in Q^C$ such that $s^C \eta = s^A$ (the start state of $\hat{A}$), and define $F^C = \{q \in Q^C \mid q\eta \in F^A\}$. Then for any $x \in (\Sigma^A)^*$,

$$x \in U \Leftrightarrow s^A x^A \in F^A \Leftrightarrow s^C \eta \, x^A \in F^A \Leftrightarrow s^C x^C \eta \in F^A \Leftrightarrow s^C x^C \in F^C.$$

Hence the automaton $\hat{C}$ with semiautomaton C, start state s^C, and final states F^C is the required automaton accepting U. Q.E.D.

6. *The Main Theorem*

The behavior of cascades of resets can be described in terms of star-free events using

Definition 6. For a semiautomaton A and states $p,q \in Q^A$, the set $A_{p,q}$ is $\{x \in (\Sigma^A)^* \mid px^A = q\}$.

LEMMA 4. *Let $C = B\&R$ with B a semiautomaton, R a reset, and ω: $Q^B \times \Sigma^B \to \Sigma^R$. If $B_{p,q}$ is a star-free event (over Σ^B) for all $p,q \in Q^B$, then $C_{a,b}$ is a star-free event (over $\Sigma^C = \Sigma^B$) for all $a,b \in Q^C$.*

PROOF. Write "Σ" for the (equal) sets Σ^B and Σ^C. It will be clear from the context which machine an input acts upon, and so superscripts on input words will be omitted. By the definition of cascade product, the first component of $\langle p, 1\rangle y$ is simply py for any $p \in Q^B$, $y \in \Sigma^*$. Since R is a reset, in order for the second component of $\langle p, 1\rangle y$ to be 2, R must receive an input $\langle r, \sigma\rangle\omega \in \Sigma_2{}^R$ for some $r \in Q^B$, $\sigma \in \Sigma$.

Suppose $x \in C_{\langle p,1\rangle,\langle q,2\rangle}$. Then $px = q$ and so $x \in B_{p,q}$, but also x must equal $y\sigma z$ for some $y,z \in \Sigma^*$, $\sigma \in \Sigma$ such that $py = r$ for some $r \in Q_B$ and $\langle r, \sigma\rangle\omega \in \Sigma_2{}^R$. Choose the shortest z such that $x = y\sigma z$ for y and σ satisfying the preceding conditions. Then no prefix of z causes R to receive an input $\langle s, \delta\rangle\omega \in \Sigma_1{}^R$ (where $s \in Q^B$, $\delta \in \Sigma$); i.e. $z \notin B_{r\sigma,s}\delta\Sigma^*$.

Conversely, if $py = r$ for $\langle r, \sigma\rangle\omega \in \Sigma_2{}^R$ and $z \notin B_{r\sigma,s}\delta\Sigma^*$ for any $\langle s, \delta\rangle\omega \in \Sigma_1{}^R$, then $y\sigma z \in C_{\langle p,1\rangle,\langle q,2\rangle}$ providing $y\sigma z \in B_{p,q}$. Altogether, one has:

$$C_{\langle p,1\rangle,\langle q,2\rangle} = B_{p,q} \cap [\bigcup B_{p,r}\sigma\overline{(\bigcup B_{r\sigma,s}\delta\Sigma^*)}],$$

where the lefthand union is over all $r \in Q^B$, $\sigma \in \Sigma$ such that $\langle r, \sigma\rangle\omega \in \Sigma_2{}^R$, and the righthand union is over all $s \in Q^B$, $\delta \in \Sigma$ such that $\langle s, \delta\rangle\omega \in \Sigma_1{}^R$.

The unions in the expression for $C_{\langle p,1\rangle,\langle q,2\rangle}$ are infinite, and Σ^* is a star-free event ($\Sigma^* = \bar{\emptyset}$ and $\emptyset = \{\sigma\} \cap \overline{\{\sigma\}}$), so that $C_{\langle p,1\rangle,\langle q,2\rangle}$ is a star-free event. The set of $x \in C_{\langle p,1\rangle,\langle q,1\rangle}$ is precisely the set of $x \in \Sigma^*$ such that $px^B = q$ and $x \notin C_{\langle p,1\rangle,\langle q,2\rangle}$, i.e. $C_{\langle p,1\rangle,\langle q,1\rangle} = B_{p,q} \cap \overline{C_{\langle p,1\rangle,\langle q,2\rangle}}$, and so $C_{\langle p,1\rangle,\langle q,1\rangle}$ is also a star-free event.

Since the argument is symmetric in states 1 and 2 of Q^R, $C_{a,b}$ is a star-free event for all $a,b \in Q^C$. Q.E.D.

LEMMA 5. *If C is a cascade product of resets, then $C_{a,b}$ is a star-free event for all $a,b \in Q^C$.*

PROOF. Let R be a reset and B a semiautomaton such that $Q^B = \{p\}$ and $\Sigma^B = \Sigma^R$. For $\sigma \in \Sigma^B$ define ω: $Q^B \times \Sigma^B \to \Sigma^R$ by the condition $\langle p, \sigma\rangle\omega = \sigma$. In this trivial case of a cascade product, $R_{i,j} = (B\&R)_{\langle p,i\rangle,\langle p,j\rangle}$ for all $i,j \in Q^R$. Since $B_{p,p} = (\Sigma^B)^*$ is star-free, Lemma 4 implies that $R_{i,j}$ is star-free.

The rest of the proof follows immediately by Lemma 4 and induction on the number of resets in C. Q.E.D.

COROLLARY 4. *Every event accepted by an automaton $\hat{A}$ whose semiautomaton A is a cascade product of resets is a star-free event.*

PROOF. Let $a \in Q^A$ be the start state of $\hat{A}$, and F^A the final states. The event accepted by $\hat{A}$ is $\bigcup_{b \in F^A} A_{a,b}$, which is a star-free event since the union is finite and $A_{a,b}$ is star-free by Lemma 5. Q.E.D.

This completes the proof of the following.

THEOREM. (Schützenberger, Papert-McNaughton) *The following are equivalent for events $U \subset \Sigma^*$:*

(1) *U is a star-free event.*

(2) *U is a noncounting regular event.*

(3) *U is a group-free event.*
(4) *U is accepted by a cascade product of resets.*

REFERENCES

1. Cohen, R., and Brzozowski, J. A. On star-free events. Proc. of Hawaii Interna Conference on System Science, U. of Hawaii Press, Honolulu, 1968, pp. 1–4.
2. Ginzburg, A. *Algebraic Theory of Automata.* ACM Monograph Series, Academic New York, 1968.
3. Papert, S., and McNaughton, R. On topological events. In *Theory of Automata*, Michigan Engineering Summer Conferences, 1966.
4. Rabin, M. O., and Scott, D. Finite automata and their decision problems. *IBM J Develop. 3*, 2 (Apr. 1959), 114–125.
5. Schützenberger, M. P. On a family of sets related to McNaughton's *L*-langua *Automata Theory*, Caianiello, E. R. (Ed.), Academic Press, New York, 1966, pp. 320-
6. Yoeli, M. Generalized cascade decompositions of automata. *J. ACM 12*, 3 (July 411–423.

received march, 1968