On Decompositions of Regular Events

J. A. BRZOZOWSKI AND RINA COHEN

University of Waterloo, Waterloo, Ontario, Canada*

ABSTRACT. Decompositions of regular events into star events, i.e. events of the form $W = V^*$, are studied. Mathematically, the structure of a star event is that of a monoid. First it is shown that every regular event contains a finite number of maximal star events, which are shown to be regular and can be effectively computed. Necessary and sufficient conditions for a regular event to be the union of its maximal star events are found. Next, star events are factored out from arbitrary events, yielding the form $W = V^*T$. For each W there exists a unique largest V^* and a unique smallest T; an algorithm for finding suitable regular expressions for V and T is developed. Finally, an open problem of Paz and Peleg is answered: Every regular event is decomposable as a finite product of star events and prime events.

KEY WORDS AND PHRASES: events, regular events, decomposition, star events, prime events, finite automata

CR CATEGORIES: 5.22

1. *Introduction*

Let A be a finite alphabet and A^* the free monoid generated by A. An event W is a subset of A^*. If W happens to be a submonoid of A^*, then $W = W^*$, and W is called a star event. Star events and their properties have been studied previously [2, 4]. In this paper we study the structure of regular events by decomposing the events, using the star event as a basic building block. We deal mainly with regular events, although many results apply to arbitrary events.

In Section 2 the basic definitions and notation are given. The known results for star events are briefly summarized in Section 3. In Section 4 it is shown that every regular event contains a finite number of maximal star events, and an algorithm for finding these maximal star events is presented. Necessary and sufficient conditions for a regular event to be the union of all of its maximal star events are developed. Incidentally, an infinite maximal monoid contained in a regular event is related to a maximal submonoid of the well-known finite monoid [3] of the automaton accepting the regular event.

We next look for star events which can be factored out from a given event. Section 5 contains a discussion of (left) comet events of the form $C = ST$ where S is a star event and T is called the tail of C. Every comet event has a unique canonical form which uses the largest star S_M and the smallest tail T_m such that $C = S_M T_m$. An algorithm, using a nondeterministic state diagram, is developed for finding, in a convenient form, the smallest tail of a regular comet. Canonical forms for right comets and two-sided comets are also found.

This paper was presented at the Eighth Annual Symposium on Switching and Automata Theory, Austin, Texas, 1967.

* Department of Applied Analysis and Computer Science. The majority of this work was done while the authors were at the University of Ottawa in the Departments of Electrical Engineering and Mathematics, respectively. Part of this work was supported by the National Research Council of Canada, Grant No. A-1617.

Finally, in Section 6 we present an answer to an open problem suggested by Paz and Peleg [4]. It is proved that every regular event is decomposable as a product of a finite number of star events and prime events (an event is prime if it has no proper decomposition as a product). Unfortunately, such decompositions are not unique.

2. *Definitions and Notation*

Let $A = \{a_1, a_2, \cdots, a_k\}$ be a finite, nonempty *alphabet* of k distinct *letters* a_i. A string of $n \geq 0$ letters of A is a *word* w of length $l(w) = n$. The *empty word* is denoted by λ, $l(\lambda) = 0$. A^* denotes the free monoid generated by A. Any subset of A^* is called an *event*. The empty event is denoted by $\varnothing$, and $\Lambda = \{\lambda\}$. Let U and V be events; then $U \cup V$, $U \cap V$, and UV are the union, intersection, and product of U and V where $UV = \{w \mid w = uv, u \in U, v \in V\}$. Also $\bar{U} = A^* - U$ is the complement of U, and $U^* = \bigcup_{n=0}^{\infty} U^n$. For any word $w = a_{i_1}a_{i_2} \cdots a_{i_n}$, $l(w) > 0$, the *reverse* of w is the word $w^T = a_{i_n}a_{i_{n-1}} \cdots a_{i_1}$, and $\lambda^T = \lambda$. The reverse of an event U is defined by $U^T = \{u \mid u = w^T \text{ for some } w \in U\}$.

For $u \in A^*$, $W \subseteq A^*$, the *left quotient* of W by u is $u \backslash W = \{x \mid ux \in W\}$.

Regular expression is defined inductively:

(1) a_i for $1 \leq i \leq k$, λ, and $\varnothing$ are regular expressions.

(2) If P and Q are regular expressions, then so are $P \cup Q$, $P \cap Q$, PQ, P^* and $\bar{P}$.

Given a regular expression R, $|R|$ is the event denoted by R. An event W is called *regular* iff there exists a regular expression R such that $W = |R|$. If R is a regular expression and $w \in A^*$, then $D_w R$ is a regular expression denoting $w \backslash |R|$, and it is called a *derivative* of R with respect to w [1].

If no formal distinction is necessary, for convenience we use regular expressions and derivatives to mean regular events and quotients.

A *finite automaton* (or simply *automaton*) $\mathfrak{A}$ over alphabet A is a quadruple $\mathfrak{A} = \langle Q, M, q_1, F\rangle$ where Q is a finite, nonempty set of *states*, $q_1 \in Q$ is the *initial state*, $F \subseteq Q$ is the set of *final states*, and M is the *transition function*: $Q \times A \rightarrow 2^Q$ (the range consisting of finite subsets of Q). If the range of M consists of subsets having a single state only, $\mathfrak{A}$ is said to be *deterministic*; otherwise, it is *nondeterministic*. The transition function is extended to A^*:

$$M(q_i, wa) = \{q' \mid \text{there exists } q \in M(q_i, w) \text{ such that } q' \in M(q, a)\}.$$

A word $w \in A^*$ is *accepted* by $\mathfrak{A}$ iff $(M(q_1, w) \cap F)$ is not empty. The set of all words accepted by $\mathfrak{A}$ will be denoted by $T(\mathfrak{A})$ and is regular [8]. Also, for each regular event R there exists a unique, reduced deterministic automaton [8] which will be denoted by $\mathfrak{A}_0(R)$. Since there exists a 1–1 correspondence between the set of states Q of $\mathfrak{A}_0(R)$ and the set of all distinct derivatives of R (see [1]), the derivative which corresponds to a state $q \in Q$ will be denoted by D_q.

We further extend the transition function M as follows: M: $2^Q \times 2^{A^*} \rightarrow 2^Q$ where for every $Q_1 \subseteq Q$ and $W \subseteq A^*$, $M(Q_1, W) = \{q \mid q \in Q \text{ and } q \in M(q_i, w)$ for some $q_i \in Q_1, w \in W\}$.

Let $\mathfrak{A} = \langle Q, M, q_1, F\rangle$ be an automaton. Then for any $q' \in Q$, $F' \subseteq Q$, $\mathfrak{A}(q', F')$ denotes $\langle Q, M, q', F'\rangle$.

For any event $W \subseteq A^*$, define a function $\delta(W)$ by:

$$\delta(W) = \begin{cases} \lambda & \text{iff } \lambda \in W, \\ \varnothing & \text{otherwise.} \end{cases}$$

3. *Star Events*

An event $W \subseteq A^*$ is a *star event* iff $W = V^*$ for some $V \subseteq A^*$. The basic results for star events are summarized below [2, 4].

(1) The following equivalent conditions are necessary and sufficient for $W \subseteq A^*$ to be a star event:

(a) $W = W^*$.

(b) $W = W^2$.

(c) $\lambda \in W$, and for each $u \in A^*$, $\lambda \in D_u W$ iff $D_u W \supseteq W$.

(2) If $W = V^*$ then V is called a *root* of W. Every star event W has a unique smallest root W_I which corresponds to the set of words u of W that have no proper decomposition $u = xy$, $x,y \in (W - \Lambda)$. Thus we have

$$W_I = (W - \Lambda) - (W - \Lambda)^2$$

for the minimum root of W.

(3) W is regular iff W_I is regular.

For a regular star event R, the minimum root can be constructed in a convenient form from the state diagram of a certain nondeterministic automaton recognizing R.

Definition 3.1. Let R be a regular star event. The nondeterministic finite automaton $\mathcal{B}(R)$ associated with R is defined as follows:

$$\mathcal{B}(R) = \langle P, N, p_1, G \rangle$$

(a) Each state of $\mathcal{B}(R)$ corresponds either to a derivative of R or to a difference $\gamma_j R$ of derivatives of R, where $\gamma_j R$ is any expression of the form $D_{w_1}R - D_{w_2}R - \cdots - D_{w_k}R$.

(b) The initial state corresponds to the derivative $D_\lambda R = R$.

(c) $G = \{D_\lambda R\}$.

(d) For any $a \in A$, $N(R, a) = \{D_a R\}$ if $\lambda \notin D_a R$, and $N(R, a) = \{D_a R - R, R\}$ if $\lambda \in D_a R$.

(e) Let $\gamma_j R = D_{w_1}R - D_{w_2}R - \cdots - D_{w_k}R$, and let $\gamma_{ja}R = D_{w_1a}R - D_{w_2a}R - \cdots - D_{w_ka}R$. Then $N(\gamma_j R, a) = \{\gamma_{ja}R\}$ if $\lambda \notin \gamma_{ja}R$, and $N(\gamma_j R, a) = \{\gamma_{ja}R - R, R\}$ if $\lambda \in \gamma_{ja}R$.

THEOREM 3.1 [2] *Let R be a regular star event and $\mathcal{B}(R)$ its associated nondeterministic automaton. Then R_I is the set of all words of length greater than 0 which take the state diagram of $\mathcal{B}(R)$ from the initial state back to the initial state without going through the initial state.*

4. *Star Events Contained in a Regular Event*

We now consider the problem of finding star events contained in a regular event R.

Definition 4.1. Let S be a star event contained in a regular event R. S is a *maximal star* in R iff it is not properly contained in any other star event contained in R.

Definition 4.2. Let R be a regular event and let $\mathcal{A} = \mathcal{A}_0(R) = \langle Q, M, q_1, F \rangle$. For any subset $B \subseteq Q$ of states of $\mathcal{A}$, define

$$\text{str } B = \{x \in A^* \mid M(B, x) \subseteq B\}.$$

LEMMA 4.1. *Str B is a regular star event. Moreover, if $q_1 \in B \subseteq F$ then str $B \subseteq R$.*

PROOF. Clearly str $B = \bigcap_{q \in B} T(\mathcal{A}(q, B))$, and therefore str B is regular. Now obviously $\lambda \in$ str B, and for any $x,y \in$ str B, we get by definition of str B, for every $q \in B$,

$$M(q, xy) = M(M(q, x), y) \in B,$$

thus showing that $xy \in$ str B. Hence str $B = (\text{str } B)^2$, and str B is a star event. If $q_1 \in B$ and $B \subseteq F$, then by the definition of str B, $M(q_1, x) \in B \subseteq F$ for all $x \in$ str B, and therefore str $B \subseteq R$.

Notation. For a given automaton $\mathcal{A} = \langle Q, M, q_1, F \rangle$, denote by M_1 the function $M_1: A^* \to Q$ defined by $M_1(x) = M(q_1, x)$ for all $x \in A^*$. Extend this function to any subset $Y \subseteq A^*$: $M_1(Y) = \{M_1(y) \mid y \in Y\} \subseteq Q$. $M_1(Y)$ is the set of all states reachable by sequences of Y from the initial state q_1. Though it is well-defined, $M_1(Y)$ is not always computable. However it was proved in [6] that $M_1(Y)$ is computable whenever Y is a context-free language and, in particular, whenever Y is a regular set.

In the following theorems let R be a regular event and $\mathcal{A} = \mathcal{A}_0(R) = \langle Q, M, q_1, F \rangle$.

LEMMA 4.2. *Let S_1 be any star event contained in R. Then (a) $S_1 \subseteq$ str $(M_1(S_1))$, and (b) there exists a maximal star S of R containing S_1 such that $S =$ str B where $B \supseteq M_1(S_1)$.*

PROOF. (a) By definition of $M_1(S_1)$ there exists for any state $q \in M_1(S_1)$ a word $w_q \in S_1$ such that $M_1(w_q) = q$. Now let $w \in S_1$. Then for all $q \in M_1(S_1)$,

$$M(q, w) = M(M_1(w_q), w) = M(q_1, w_q w) = M_1(w_q w) \in M_1(S_1),$$

showing that $w \in$ str $(M_1(S_1))$, and hence $S_1 \subseteq$ str $(M_1(S_1))$.

(b) Denote $M_1(S_1)$ by B_1. Then $S_1 \subseteq$ str B_1, and if str B_1 is not maximal there exists a star event S_2 such that $S_2 \supsetneq$ str B_1. Thus $B_2 = M_1(S_2) \supsetneq B_1$, and also str $B_2 \supsetneq$ str $B_1 \supseteq S_1$. If str B_2 is again not maximal, there exists a star event $S_3 \supsetneq$ str B_2, and hence a set $B_3 = M_1(S_3) \supsetneq B_2$, such that str $B_3 \supseteq S_1$. By repetition of this procedure we get a sequence $B_1 \subsetneq B_2 \subsetneq \cdots \subsetneq B_i \subsetneq \cdots$ of subsets of the finite set F. Therefore the sequence is finite, and we get a set $B_n = B$ such that str B is maximal in R and contains S_1.

THEOREM 4.3. *Let S be a maximal star in R. Then $S =$ str $(M_1(S))$.*

PROOF. By Lemma 4.2, $S \subseteq$ str $(M_1(S))$. Since $\lambda \in S \subseteq R$ we have $q_1 \in M_1(S) \subseteq F$, and it follows from Lemma 4.1 that str $(M_1(S))$ is a star event contained in R. But S is maximal and therefore $S =$ str $(M_1(S))$. It follows that every maximal star contained in R is regular and the number of distinct maximal stars is finite.

COROLLARY 4.4. *If R is a star event, then $R =$ str F.*

THEOREM 4.5. *A regular event R can be represented as the union of its maximal stars iff for every $w \in R$, $w^* \subseteq R$.*

PROOF. *Necessity:* By assumption R is the union of its maximal stars. Thus for any $w \in R$, $w \in S$ for some maximal star event S in R; hence $w^* \subseteq S \subseteq R$.

Sufficiency: Let $w \in R$. Then w^* is in R. Since w^* is a star event, $w^* \subseteq S$ for

By construction, $\lambda \in \Delta_w C = D_{u_1 \cdots u_n v}C - \cdots - D_v C$. Thus $w \in C$ and for all $i = 1, \cdots, n$, $u_{i+1} \cdots u_n v \in C$. But this means that $w \in C - (S_M - \Lambda)C$, because if $u_1 \cdots u_i$ is a prefix of w that belongs to $S_M - \Lambda$, the corresponding suffix is not in C. Thus $T_m \supseteq W$. Now suppose $w \in T_m$, $l(w) > 0$. Expand w with respect to S_M. Then by Lemma 5.10, w takes the state graph of $\mathcal{B}(S_M)$ from the initial state to a state corresponding to $\Delta_w S_M$, and this path does not go through the initial state. Now $w \in T_m$ implies $w \in C$ and hence $\lambda \in D_w C$. If also $\lambda \in D_{u_{i+1} \cdots u_n v}C$, then there exists a suffix of w which is in C and the corresponding prefix is in S_m. This is impossible if $w \in T_m$. Hence $\lambda \in \Delta_w C$, $w \in W$, and $W \supseteq T_m - \Lambda$. Thus $T_m = W \cup \delta(C)$.

Example 5.1. Let $R = (110^*1)^*(11 \cup 0(0 \cup 1)^*)$. The derivative equations [1] for R and for its largest left star S are

$$
\begin{aligned}
R &= 0R_0 \cup 1R_1 & S &= 0S_0 \cup 1S_1 \cup \lambda \\
R_0 &= 0R_0 \cup 1R_0 \cup \lambda & S_0 &= 0S_0 \cup 1S_0 \cup \lambda \\
R_1 &= 1R_{11} & S_1 &= 1S_{11} \\
R_{11} &= 0R_{110} \cup 1R \cup \lambda & S_{11} &= 0S_{110} \cup 1S \\
R_{110} &= 0R_{110} \cup 1R & S_{110} &= 0S_{110} \cup 1S
\end{aligned}
$$

because $\{R, R_0\}$ is the set of all derivatives that contain R (see Theorem 5.7). The derivative equations for S can be reduced since $S_{11} = S_{110}$, and the minimum root of S can be found as in Theorem 3.1, giving $S = (0 \cup 01 \cup 011 \cup 110^*1)^*$. The minimum tail is found by using Theorem 5.8 as shown in Figure 3. The differences of derivatives of R that contain λ are $(R_0 - R)$, R_{11}, and $(R_0 - R_1 - R)$; thus the output states for tail computation are $(S_0 - S)$, S_{11}, and $(S_0 - S_1 - S)$, as shown. By inspection, $T_m = (0 \cup 01 \cup 11)$.

Finally, we remark that ultimate definite events [5] are simply comet events whose star is A^*.

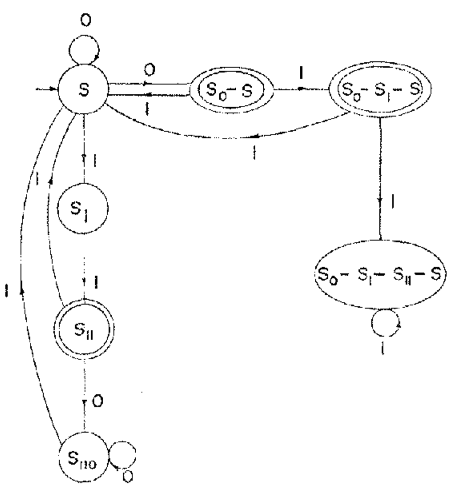

FIG. 3

6. *Products of Stars and Primes*

Definition 6.1. An event R is called a *prime event* iff $R = T_1T_2$ implies $T_1 = \Lambda$ or $T_2 = \Lambda$.

The following problem was stated in [4]: Given a regular event R, is R decomposable in the form $R = T_1T_2 \cdots T_k$ (k finite) with T_i prime or $T_i = T_i^*$, $i = 1, 2, \cdots, k$?

It is the purpose of this section to establish a positive answer to this question.

Definition 6.2. Let $\mathfrak{A} = \langle Q, M, q_1, F\rangle$ and $B \subseteq Q$. Then the *set* $\hat{B}$ *associated with* B is

$$\hat{B} = \{q' \mid q' \in Q \text{ and } D_{q'} \supseteq \bigcap_{q \in B} D_q\}.$$

Obviously $B \subseteq \hat{B}$, and also $\hat{\hat{B}} = \hat{B}$.

The following two lemmas are essentially due to Paz and Peleg [4].

LEMMA 6.1. *For any* $B \subseteq Q$, $\bigcap_{q\in B} T(\mathfrak{A}(q, \hat{B})) = \text{str } \hat{B}$.

PROOF. Denote $\bigcap_{q\in B} T(\mathfrak{A}(q, \hat{B}))$ by S. By the definition of str $\hat{B}$, str $\hat{B} = \bigcap_{q\in \hat{B}} T(\mathfrak{A}(q, \hat{B}))$, and since $B \subseteq \hat{B}$, str $\hat{B} \subseteq S$. To show $S \subseteq$ str $\hat{B}$, let $x \in S$. Then for any $y \in \bigcap_{q\in B} D_q$ and for any $q \in B$,

$$M(q, xy) = M(M(q, x), y) = M(q', y) \in F$$

since $q' = M(q, x) \in \hat{B}$ and $y \in D_{q'}$. Thus

$$x(\bigcap_{q\in B} D_q) \subseteq \bigcap_{q\in B} D_q.$$

In order to show that $x \in$ str $\hat{B}$, take any $\hat{q} \in \hat{B}$ and show that $M(\hat{q}, x) \in \hat{B}$. For any $y \in \bigcap_{q\in B} D_q$, $M(M(\hat{q}, x), y) = M(\hat{q}, xy) \in F$ since $xy \in \bigcap_{q\in B} D_q$ and $\bigcap_{q\in B} D_q \subseteq D_{\hat{q}}$. It follows from the definition of $\hat{B}$ that $M(\hat{q}, x) \in \hat{B}$, and hence $x \in$ str $\hat{B}$.

In the sequel, let $\mathfrak{A} = \mathfrak{A}_0(R) = \langle Q, M, q_1, F\rangle$.

LEMMA 6.2. *Let* $R = FT$ *and* $M_1(F) = B$. *Then* $R = F(\text{str } \hat{B})T$.

LEMMA 6.3. *Let* $R = FST$ *where* S *is a star event and* $B = M_1(FS)$. *Then* $S \subseteq \text{str } B \subseteq \text{str } \hat{B}$.

PROOF. By a similar argument to that used in Lemma 4.2, $S \subseteq$ str B. Now str $B = \bigcap_{q\in B} T(\mathfrak{A}(q, B)) \subseteq \bigcap_{q\in B} T(\mathfrak{A}(q, \hat{B}))$, and applying Lemma 6.1 we obtain str $B \subseteq$ str $\hat{B}$.

LEMMA 6.4. *Let* $R = T_1T_2T_3$ *and* $M_1(T_1) \subseteq M_1(T_1T_2)$; *then* $\lambda \in T_2$.

PROOF. Suppose t_1, t_3 are words of minimum length in T_1, T_3, respectively. Since $M_1(t_1) \in M_1(T_1T_2)$ by assumption,

$$M_1(t_1t_3) = M(q_1, t_1t_3) = M(M(q_1, t_1), t_3) = M(M_1(t_1), t_3)$$
$$\subseteq M(M_1(T_1T_2), t_3) \subseteq M_1(T_1T_2T_3) \subseteq F.$$

Hence $t_1t_3 \in T_1T_2T_3$, and $\lambda \in T_2$.

LEMMA 6.5. *Let* $R = F(\text{str } B)\, T_m$ *where* $M_1(F \text{ str } B) = B$ *and* T_m *is the minimum tail of str* B. *Then for every nontrivial decomposition* $T_m = T_1T_2$,

$$M_1(F(\text{str } B)T_1) \neq B.$$

(*Note that str* B *may be equal to* Λ.)

PROOF. Assume $M_1(F(\text{str } B)T_1) = B$. Then for any $t_1 \in T_1$, $M(B, t_1) \subseteq B$, and hence $t_1 \in \text{str } B$ and $T_1 \subseteq \text{str } B$. Also, by the previous lemma, since $M_1(F \text{ str } B) = M_1(F(\text{str } B)T_1)$, $\lambda \in T_1$, and hence $T_2 \subseteq T_1T_2$. Concatenating by $(T_1 - \Lambda)$ (which is nonempty by assumption),

$$(T_1 - \Lambda)T_2 \subseteq (T_1 - \Lambda)T_1T_2 \subseteq (\text{str } B - \Lambda)T_1T_2 ,$$

or

$$(T_1 - \Lambda)T_2 \subseteq (\text{str } B - \Lambda)T_m .$$

But $(T_1 - \Lambda)T_2 \subseteq T_m$, and T_m is the minimum tail of str B; thus by Theorem 5.2, $T_m \cap (\text{str } B - \Lambda)T_m = \varnothing$, and therefore $(T_1 - \Lambda)T_2 = \varnothing$, which is a contradiction.

In a similar way the same result can be proved for the symmetric case.

LEMMA 6.6. *Let $R = F_m(\text{str } B)\ T$ where $M_1(F_m \text{ str } B) = B$, and F_m is the minimum front of str B. Then for every nontrivial decomposition $F_m = F_1F_2$, $M_1(F_1) \neq B$.*

LEMMA 6.7. *Let $R = FT$ where T is a comet event. Then R can be written in the form $R = F(\text{str } B_1)(\text{str } B_2) \cdots (\text{str } B_r)T_m$ where $r \geq 1$, $M_1(F) \subseteq B_1 \subsetneq B_2 \subsetneq \cdots \subsetneq B_r \subseteq Q$, and T_m is the minimum tail of str B_r and is not a comet event.*

PROOF. First factor a maximum star on the left of T by Theorem 5.1: $T = S_{1_M}T_1$. Define $B_1 = \hat{M}_1(FS_{1_M})$, and replace S_{1_M} by str B_1. This is allowed by Lemma 6.2 and the fact that $S_{1_M} \subseteq \text{str } B_1$ (Lemma 6.3). Thus $R = F(\text{str } B_1)T_{1_m}$ where T_{1_m} is the minimum tail of str B_1. If T_{1_m} is again a comet event, factor a maximal star on the left of it: $T_{1_m} = S_{2_M}T_2$. Since $\lambda \in S_{2_M}$, $M_1(F \text{ str } B_1) = B_1 \subseteq \hat{M}_1(F \text{ str } B_1S_{2_M}) = B_2$. But by Lemma 6.5, $B_2 \neq B_1$, and thus $B_1 \subsetneq B_2$. As before replace S_{2_M} by str B_2 and concatenate by the minimum tail T_{2_m}: $R = F(\text{str } B_1)\ (\text{str } B_2)T_{2_m}$. By repeating this process we obtain sets $B_1 \subsetneq B_2 \subsetneq \cdots \subsetneq B_k \subsetneq \cdots$, all contained in the finite set Q. Therefore after a finite number of steps we get a tail T_{r_m} which is not a comet event, and hence the lemma is proved.

LEMMA 6.8. *If R is not a comet event, then there exists a prime event P and an event T such that $R = PT$.*

PROOF. If R is a prime event, take $T = \Lambda$, and the lemma is proved. Otherwise, R is decomposable, i.e. $R = F_1T_1$ for some F_1, $T_1 \neq \Lambda$. By Lemma 6.2, $R = F_1(\text{str } \hat{B}_1)T_1$ where $B_1 = M_1(F_1)$. Replace F_1 by the minimum front F_{1_m} of str $\hat{B}_1$: $R = F_{1_m}(\text{str } \hat{B}_1)T_1$. Note that $F_{1_m} \neq \Lambda$ since otherwise R would be a comet event. Now if F_{1_m} is a prime event, the desired result follows. If not, however, continue decomposing it as follows: $F_{1_m} = F_2T_2$, F_2, $T_2 \neq \Lambda$. Again replace this by $F_{1_m} = F_{2_m}(\text{str } \hat{B}_2)T_2$ where $B_2 = M_1(F_2)$ and F_{2_m} is the minimum front of str $\hat{B}_2$. Thus $R = F_{2_m}(\text{str } \hat{B}_2)T_2(\text{str } \hat{B}_1)T_1$. If F_{2_m} is prime, the lemma is proved. If not, proceed in the same way decomposing it: $F_2 = F_3T_3 = F_{3_m}(\text{str } \hat{B}_3)T_3$ where $B_3 = M_1(F_3)$ and F_{3_m} is the minimum front with respect to str $\hat{B}_3$. Repeat this decomposition procedure until an F_{k_m} is obtained which is a prime. This must happen after at most $2^{\# Q}$ steps,[1] since in each step i, F_{i_m} was defined to be the minimum front with respect to str $\hat{B}_i$ in the decomposition

$$R = F_{i_m}(\text{str } \hat{B}_i)T_i(\text{str } \hat{B}_{i-1})T_{i-1} \cdots (\text{str } \hat{B}_1)T_1 ;$$

thus by further decomposition of F_{i_m} we get sets $\hat{B}_j$, $j > i$, none of which can be equal to $\hat{B}_i$ (Lemma 6.6). Thus for any i, j, $i \neq j$, we get $\hat{B}_i \neq \hat{B}_j$; i.e. the

[1] $\# Q$ denotes the number of elements in Q.

$\hat{B}_i$'s are all distinct, and their number can't exceed $2^{\#Q}$. Hence the decomposition procedure terminates and a prime event F_{k_m} is obtained which is factored on the left of R, as required.

THEOREM 6.9. *Every regular event R can be represented as a concatenation of a finite number n of star events and prime events where $n \leq 2^{\#Q}\ \#Q$.*

PROOF. If R is either a star event or a prime event, the theorem certainly holds. Otherwise, decompose R in the following way.

(a) If R is a comet event, apply the decomposition of Lemma 6.7 to R; i.e. $R = (\text{str } B_1)(\text{str } B_2) \cdots (\text{str } B_{r_1})T_{1_m}$ where for all $i = 1, \cdots, r_1 - 1$, $B_i \subsetneq B_{i+1}$ and T_{1_m} is the minimum tail with respect to str B_{r_1} and is not a comet event.

(b) If R is not a comet event, use Lemma 6.8 and factor a prime event P_1 on the left of R: $R = P_1T_1$, and by Lemma 6.2 replace this by $P_1(\text{str } B_1)T_{1_m}$ where $B_1 = \hat{M}_1(P_1)$ and T_{1_m} is the minimum tail with respect to str B_1. In this case define $r_1 = 1$.

Proceed by induction:

(c) If $T_{i_m}(i = 1, 2, \cdots)$ is a comet event, apply step (a) for T_{i_m}; i.e. replace it by $(\text{str } B_{r_i+1})(\text{str } B_{r_i+2}) \cdots (\text{str } B_{r_{i+1}})T_{(i+1)_m}$ where $B_{r_i+1} \subsetneq B_{r_i+2} \subsetneq \cdots \subsetneq B_{r_{i+1}}$, and $T_{(i+1)_m}$ is the minimum tail with respect to str $B_{r_{i+1}}$.

(d) If T_{i_m} $(i = 1, 2, \cdots)$ is not a comet event, apply step (b) to it; i.e. replace it by $P_{i+1}(\text{str } B_{r_{i+1}})T_{(i+1)_m}$ where P_{i+1} is a prime event, $r_{i+1} = r_i + 1$, $B_{r_{i+1}} = \hat{M}_1(\cdots(\text{str } B_{r_i})P_{i+1})$, and $T_{(i+1)_m}$ is the minimum tail with respect to str $B_{r_{i+1}}$.

By reasoning similar to that used in proving Lemma 6.8, using Lemma 6.5, it can be seen that all sets B_{r_i}, $i = 1, 2, \cdots$, are distinct. Therefore this decomposition procedure must terminate after at most $2^{\#Q}$ steps. Since at each step we add at most $\#Q$ factors, the total number of factors in the decomposition cannot exceed $\#Q \cdot 2^{\#Q}$.

Remark. The decomposition procedure described above is not unique; i.e. a regular event R may be decomposed by this process into a concatenation of stars and primes in more than one way. This is due to the fact that the factoring of a prime event described in Lemma 6.8 is not necessarily unique. This will be illustrated by the following example.

Example 7.1. Let R be the event accepted by the automaton shown in Figure 4, where the empty derivative is not shown. The reader may easily verify that R can be denoted by the following two regular expressions,

$$R = (0 \cup \lambda)\ [10 \cup (3 \cup 21)(21)^*(10 \cup 0)]^*\ (\lambda \cup 3 \cup 21)(21)^*\ 1,$$
$$R = (3 \cup \lambda)\ [21 \cup (0 \cup 10)(10)^*(21 \cup 3)]^*\ (\lambda \cup 0 \cup 10)(10)^*\ 1.$$

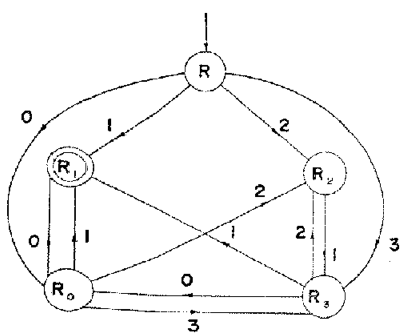

FIG. 4

All factors are either star events or primes, as may be easily verified, and both decompositions may be obtained by the decomposition procedure of Theorem 7.6.

REFERENCES

1. Brzozowski, J. A. Derivatives of regular expressions. *J. ACM 11*, 4 (Oct. 1964), 481–494.
2. Brzozowski, J. A. Roots of star events. *J. ACM 14*, 3 (July 1967), 466–477.
3. Hartmanis, J., and Stearns, R. E. *Algebraic Structure Theory of Sequential Machines*. Prentice-Hall, Englewood Cliffs, N. J., 1966.
4. Paz, A., and Peleg, B. On concatenative decompositions of regular events. *IEEE Trans. EC-17* (Mar. 1968), 229–237.
5. Paz, A., and Peleg, B. Ultimate-definite and symmetric-definite events and automata. *J. ACM 12*, 3 (July 1965), 399–410.
6. Stearns, R. E., and Hartmanis, J. Regularity preserving modifications of regular expressions. *Inform. Contr. 6* (1963), 55–69.
7. Ljapin, E. S. *Semigroups*. Translations of Math. Monographs (3). Amer. Math. Soc., Providence, R. I., 1963.
8. Rabin, M. O., and Scott, D. Finite automata and their decision problems. *IBM J. Res. Dev. 3*, 2 (Apr. 1959), 114–125.

RECEIVED MARCH, 1968