JOURNAL OF COMPUTER AND SYSTEM SCIENCES: 5, 1–16 (1971)

Dot-Depth of Star-Free Events*

RINA S. COHEN AND J. A. BRZOZOWSKI

Department of Applied Analysis and Computer Science, University of Waterloo, Waterloo, Ontario, Canada

Received June 15, 1970

A regular event is star-free if it can be denoted by a regular expression involving only Boolean operations and concatenation (dot). The family of star-free events can be constructed by alternately applying Boolean operations and concatenation. This approach leads to a hierarchy of star-free events, and to the definition of "dot-depth" of a star-free event which appears to be useful as a measure of the complexity of the event.

Properties of dot-depth are examined; for example, it is shown that the dot-depth of a star-free event cannot be increased by the quotient operation with respect to any language, nor can it be increased by multiplying the event by a finite language. The use of two-sided quotients adds insight to the theory of star-free events and permits the derivation of some new properties of these events; in particular, every star-free event has at least one quotient which is either empty or full.

The family of star-free events has been shown to be equivalent to the family of regular noncounting events, and also corresponds to the family of finite automata whose semigroups have no nontrivial subgroups (group-free automata). In the final section an algorithm is developed for constructing a star-free expression for the event accepted by a group-free finite automaton. An upper bound for the dot-depth of the event is found.

We conjecture that for each $n \geqslant 0$, there exist star-free events of dot depth n. We have been able to show this only for $n \leqslant 2$; the general problem remains open.

1. NOTATION

It is assumed that the reader is familiar with the basic results in the theory of finite automata, regular languages, and derivative techniques [3, 7]. This section provides the terminology and notation used.

A — alphabet.

$I = A^*$ — free monoid generated by A.

* Most of the results in this paper were presented at the Hawaii International Conference on System Sciences, University of Hawaii, Honolulu, Hawaii, January 1968. The research reported here was supported by the National Research Council of Canada, Grants No. A-5255 and No. A-1617.

$\ell(w)$ — lenght of w.

$R \subseteq I$ — event.

$|E|$ — event denoted by expression E.

λ — the empty word.

ϕ — the empty event.

Set operations

$P \cup Q$(union), $P \cap Q$(intersection)

$\bar{P}$ (complement), $P \cdot Q$ — concatenation (product), $P^* = \bigcup_{n \geqslant 0} P^n$.

$x \setminus R = \{w \mid xw \in R\}$, left quotient of event R by word x.

$x \setminus R / y = \{w \mid xwy \in R\}$, two-sided quotient of event R by words x and y.

$U \setminus R = \{w \mid uw \in R$ for some $u \in U\}$, left quotient of event R by event U.

$U \setminus R / V = \{w \mid uwv \in R$ for some $u \in U, v \in V\}$ — two sided quotient of event R by events U and V.

$\mathscr{A} = \langle Q, M, q_1, F\rangle$ — finite automaton over alphabet A.

Q — set of states.

$q_1 \in Q$ — initial state.

$F \subseteq Q$ — set of final states.

$M : Q \times A \rightarrow Q$ — transition function.

$\mathscr{S} = \langle A, Q, M\rangle$ — semiautomaton over alphabet A, with Q and M as above.

2. Generation of Star-Free Events and the Notion of Dot-Depth

Star-free events have been studied by several authors [2–6]. For an extensive treatment of the subject the reader is referred to the forthcoming monograph by Papert and McNaughton [1].

Let $A = \{a_1, a_2, ..., a_k\}$ be a finite, nonempty alphabet, and let $I = A^*$ be the free monoid with unit λ generated by A.

The family E_0 of the $k + 2$ events: $\{a_1\}, \{a_2\}, ..., \{a_k\}, \{\lambda\}$ and ϕ will be called the family of *basic* events. An event is *star-free* iff it belongs to the smallest family X of events containing E_0 and closed under concatenation ("dot"), and the Boolean operations. This family is the set of all events that can be denoted by extended regular expressions without stars. To simplify notation, we shall use regular expressions to denote events. Thus $\{a_i\}$ is represented by a_i, $\{\lambda\}$ by λ, etc. Note that the event $\{A\}^*$ denoted by $A^* = I$ is star-free, for $I = \bar{\phi}$.

In constructing star-free events from E_0 it is convenient to consider Boolean operations together, because they are essentially combinatorial. On the other hand, concatenation has a sequential nature, in that the words in a product must appear in proper sequence.

We can begin constructing the family of star-free events by starting with the basic family E_0 and closing it under Boolean operations. We then close the resulting family under concatenation, then again under Boolean operations, etc. In general, if Y is a family of events, let $B(Y)$ be the smallest Boolean algebra containing Y. Let $M(Y)$ be the smallest family containing Y and closed under concatenation. Using this notation, define the following sequence of families of events:

$$E_0 \subset B_1 \subset M_1 \subset B_2 \subset M_2 \subset \ldots,$$

where $B_1 = B(E_0)$, $B_n = B(M_{n-1})$, for $n > 1$, and $M_n = M(B_n)$, for $n \geqslant 1$.
The following proposition is easily verified:

PROPOSITION 2.1. *B_1 is a finite Boolean algebra of 2^{k+2} elements and is characterized by the $k + 2$ atoms: λ, $a_1, a_2, \ldots, a_k$ and $\overline{\lambda \cup A} = A^2I$. Alternatively, $R \in B_1$, iff either $R \subseteq (A \cup \lambda)$ or $\bar{R} \subseteq (A \cup \lambda)$.*

Since B_1 is not closed under concatenation, we examine M_1.

PROPOSITION 2.2. *M_1 is the set of all events which can be expressed as products of events of type P, where $P \subseteq (A \cup \lambda)$ and of type Q, where $\bar{Q} \subseteq (A \cup \lambda)$.*

M_1 contains all the events consisting of single words, but does not contain all finite events. For example, if $A = \{0, 1\}$, then $0 \cup 11 \notin M_1$. In turn, we examine B_2.

PROPOSITION 2.3. *B_2 consists of all the events which are Boolean functions of products of finite and cofinite events. (R is cofinite iff $\bar{R}$ is finite.)*

Proof. Since M_1 contains all events of the form $R = w$, $w \in I$, and B_2 must be closed under union, it follows that B_2 contains all finite events. Since it must be closed under complementation, all cofinite events must be included. Next it is clear that M_1 contains all products of words and I, i.e., all $R = w_1 I w_2 I \cdots I w_n I w_{n+1}$, $w_i \in I$, $i = 1, \ldots, n + 1$. Hence B_2 must contain all finite unions of products of words and I. However, note that every cofinite event can be written in the form $R = F \cup A^n I$ where F is finite (i.e., for each cofinite R there exists an integer $n \geqslant 0$ such that R contains all words of length $\geqslant n$). Any product of finite and cofinite events can be transformed into a finite union of words and I, by applying the distributive law of concatenation over union. Thus B_2 must contain all products of finite and cofinite events. Therefore, it is necessary for B_2 to contain all Boolean functions of products of finite and cofinite events. This is also sufficient, since all products in M_1 are products of finite or cofinite events.

Unfortunately, it becomes increasingly difficult to make meaningful statements regarding M_2, B_3, etc. However, we can return to E_0 and consider the possibility of applying concatenation first. In this way, we define the sequence.

$$E_0 \subset \hat{M}_1 \subset \hat{B}_1 \subset \hat{M}_2 \subset \hat{B}_2 \subset \ldots,$$

where $\hat{M}_1 = M(E_0)$, $\hat{M}_n = M(\hat{B}_{n-1})$ for $n > 1$, and $\hat{B}_n = B(\hat{M}_n)$ for $n \geqslant 1$.

PROPOSITION 2.4. *$\hat{M}_1$ consists of events of the form $R = w$, $w \in I$, or $R = \phi$.*

PROPOSITION 2.5. *$\hat{B}_1$ consists of all events which are either finite or cofinite.*

Proof. We need to verify that $\hat{B}_1$ is closed under union. This is immediate, since the union of two finite events is finite and the union of two cofinite events or of a finite and cofinite event is cofinite.

PROPOSITION 2.6. *$\hat{M}_2$ consists of all products of finite and cofinite events.*

PROPOSITION 2.7. $\hat{B}_2 = B_2$.

Proof. This follows from Propositions 2.6 and 2.3.

Thus it is seen that the two sequences of families of events are identical after $B_2 = \hat{B}_2$ as shown in Fig. 1. The family X of star free events

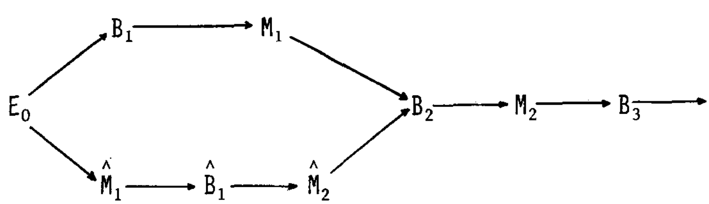

FIGURE 1

is obviously

$$X = \bigcup_{i=1}^{\infty} M_i = \bigcup_{i=1}^{\infty} B_i = \bigcup_{i=1}^{\infty} \hat{M}_i = \bigcup_{i=1}^{\infty} \hat{B}_i$$

Since the sequence B_1, M_1, B_2 $\cdots$ is shorter we will not refer further to the second sequence. Using the definition of star-height of regular events as an analogy we can define a measure of complexity of star-free events as follows:

DEFINITION 2.1. Let R be a star-free event. The *dot-depth* $d(R)$ is defined as $d(R) = n$ iff $R \in B_{n+1}$ but $R \notin B_m$ for $m < n + 1$.

The dot-depth indicates the minimum number of levels of concatenation, and can be thought of as a measure of the sequential nature of the event. Events of dot-depth 0 are subsets of $A \cup \lambda$ and their complements. Note, in particular, that $d(I) = 0$ since $I = \bar{\phi}$. The symbol I is used simply for convenience. $B_2 - B_1$ represents all events of dot-depth 1. For example, let $A = \{0, 1\}$ and let

$$R = 0I1 \cap \overline{I00I} = (0 \cdot \bar{\phi} \cdot 1) \cap \overline{\bar{\phi} \cdot 0 \cdot 0 \cdot \bar{\phi}}.$$

Then $d(R)$ is apparently 1, since R is a Boolean function of two events $0I1$ and $I00I$, both of which are in M_1. The event $P = \overline{(0 \cdot I \cdot 1)} \cdot 0$ appears to be of depth 2. However, one can verify that an equivalent expression is $P' = 0 \cup 00 \cup 1I0 \cup 0I00$, which shows that $d(P) = d(P') = 1$.

The following argument shows that there are events of dot-depth 2. We conjecture that for each n there exists a star-free event of dot-depth n, i.e., that the containment $B_{n+1} \supset B_n$ is in fact proper. However, we were unable to prove this, and the problem remains open.

LEMMA 2.8. *Every event in B_2 can be denoted by a regular expression of the form*

$$R = \bigcup_{k=1}^{l} \left(\left[\bigcap_{i=1}^{m(k)} \overline{w_0^{k,i} I w_1^{k,i} I \cdots I w_{s(k,i)}^{k,i}} \right] \cap \left[\bigcap_{j=1}^{n(k)} u_0^{k,j} I u_1^{k,j} I \cdots I u_{t(k,j)}^{k,j} \right] \right), \qquad (*)$$

where the $w_p^{k,i}$ and the $u_q^{k,j}$ are words, $m(k)$, $n(k)$, l, $s(k, i)$ and $t(k, j)$ are nonnegative integers and for $0 < p < s(k, i)$ and $0 < q < t(k, j)$, all words $w_p^{k,i}$, $u_q^{k,j}$ differ from λ.

Proof. In the proof of Proposition 2.3 we have shown that every event in B_2 is a Boolean function of a finite number of finite products of words and I. Every such function can be expressed in the disjunctive normal form ("sum-of-products form"), which is the form (*) above.

LEMMA 2.9. *The event $R = I0 \cdot \overline{I1I} = I02^*$ over alphabet $A = \{0, 1, 2\}$ is of dot-depth* 2.

Proof. Suppose that R has dot-depth less than 2. Then R is in B_2 and thus can be denoted by a regular expression of the form (*) above, i.e., $R = \bigcup_{k=1}^{l} (S_k \cap T_k)$, where $S_k(T_k)$ is the expression within the left (right) pair of square brackets in (*).

Let r be the maximum length of all words $w_p^{k,i}$, $u_q^{k,j}$ appearing in (*), and let x be any word belonging to the event

$$R' = (0 \cup 1 \cup 2)^r \left(\bigcap_{\substack{k=1 \\ i=1}}^{\substack{m(k) \\ l}} [I w_1^{k,i} I \cdots I w_{s(k,i)-1}^{k,i} I] \right)$$

(clearly, R' is nonempty). Now consider the set $W = \{x02^n \mid n = 0, 1,...,\}$; W is a subset of R and hence for some k, $1 \leqslant k \leqslant l$, the event $(S_k \cap T_k)$ contains an infinite subset of W. Then S_k contains words of the form $x02^n$ for arbitrarily large n. From the manner in which x was constructed it follows that for each $i = 1,..., m(k)$, either $w_0^{k,i}$ is not a prefix of x,[1] or $w_{s(k,i)}^{k,i} \notin 2^*$ (otherwise $x02^n$ would belong to $w_0^{k,i} I w_1^{k,i} \cdots I w_{s(k,i)}^{k,i}$ for $n > r$). But this implies that the whole event $x12^r$ is contained in S_k. Furthermore, let $w = x02^s$, $s \geqslant r$, be any word of W contained in T_k. Then by definition of T_k and x, the whole set $R_1 = x02^r12^s$ is contained in T_k. Thus we get $R_1 \subseteq S_k \cap T_k \subseteq R$, which is false since $x02^r12^s$ is in R_1 but not in R. Q.E.D.

COROLLARY 2.10. *B_2 is properly contained in the family of star-free events.*

We now return to some general properties of dot depth. For our purposes it is more convenient not to consider concatenation as a binary operation but to express each star-free expression E in the *full product form* defined by induction on the number of operators as follows:

Basis. If $E \in E_0$ then E is in full product form.

Induction Step. If F_1 and F_2 are in full product form then so are

(i) $F_1 \cup F_2$

(ii) $F_1 \cap F_2$

(iii) $\bar{F}_1$

and

(iv) $F_1 F_2 \cdots F_m$ is in full product form only if F_i, $i = 1,..., m$,

are in full product form as a result of satisfying (i), (ii) or (iii) or are in the basis.

Clearly, such a form can be obtained by using the associative law for concatenation. If an expression is a product, then we ensure that the factors are not themselves proper products. Unless otherwise specified, we assume all expressions to be in this form.

DEFINITION 2.2. *The apparent dot-depth d_α of a star-free expression E is defined inductively*:

$$d_\alpha(E) = 0 \quad \text{if} \quad E \in E_0$$
$$d_\alpha(F_1 \cup F_2) = \max\{d_\alpha(F_1), d_\alpha(F_2)\}$$
$$d_\alpha(F_1 \cap F_2) = \max\{d_\alpha(F_1), d_\alpha(F_2)\}$$
$$d_\alpha(\bar{E}) = d_\alpha(E)$$
$$d_\alpha(F_1 F_2 \cdots F_m) = 1 + \max\{d_\alpha(F_i) \mid i = 1, 2,..., m\}.$$

[1] We consider also the empty word λ as a prefix (or suffix) of any word w.

Obviously, the *dot-depth of a star-free event* R is

$$d(R) = \min\{d_\alpha(E) \mid E \text{ is a star-free expression and } |E| = R\}.$$

We now define the *dot-depth of an expression* E to be the dot-depth of the corresponding event, i.e.,

$$d(E) = d(|E|) \leqslant d_\alpha(E).$$

A star-free expression E is said to be of *proper dot-depth* iff $d(E') = d_\alpha(E')$, for all sub-expressions E' of E, including E. One can show, using induction on dot-depth, that for each star-free event R there exists a star-free expression of proper dot-depth.

It is easy to verify that any quotient of a star-free event is again star-free. We now examine the properties of dot-depth under quotient operations.

LEMMA 2.11. *Let R be star-free and let $a \in A$. Then $d(a \diagdown R) \leqslant d(R)$.*

Proof. We first prove that for every star-free expression E of proper dot-depth, $d(a \diagdown E) \leqslant d(E)$. The proof is by induction on the number n of star-free operators $(\cup, \cap, {}^{-}, \cdot)$, where concatenation is considered as an m-ary operation since we are using the full product form.

Basis, $n = 0$. One easily verifies that $E \in E_0$ implies $(a \diagdown E) \in E_0$. Thus $d(E) = d(a \diagdown E) = 0$.

Induction step, $n > 0$. Assume that for all star-free expressions E of proper dot-depth and with at most n operators $d(a \diagdown E) \leqslant d(E)$. Now let E have $m + 1$ operators. We have the following cases.

(1) $E = F_1 \cup F_2$. We have $a \diagdown E = (a \diagdown F_1) \cup (a \diagdown F_2)$. Since E is of proper dot-depth, $d(E) = \max\{d(F_1), d(F_2)\}$. Now $d(a \diagdown E) = d((a \diagdown F_1) \cup (a \diagdown F_2)) \leqslant \max\{d(a \diagdown F_1), d(a \diagdown F_2)\} \leqslant \max\{d(F_1), d(F_2)\} = d(E)$, where the last inequality follows from the inductive assumption.

(2) $E = F_1 \cap F_2$. The argument is the same as in case 1 with $\cup$ replaced by $\cap$.

(3) $E = \bar{F}$. $d(a \diagdown E) = d(a \diagdown \bar{F}) = d(\overline{a \diagdown F}) = d(a \diagdown F) \leqslant d(F) = d(\bar{F}) = d(E)$

(4) $E = F_1 F_2 \cdots F_m$. $d(E) = 1 + \max\{d(F_i) \mid i = 1, 2, ..., m\}$. Recall that $a \diagdown E$ can be written in the form

(a) $a \diagdown E = (a \diagdown F_1) F_2 \cdots F_m$, if $\lambda \notin F_1$

(b) $a \diagdown E = (a \diagdown F_1) F_2 \cdots F_m \cup a \diagdown (F_2 \cdots F_m)$, if $\lambda \in F_1$.

In case 4(a), $a \diagdown F_1$ will be either in E_0 or will be a Boolean function of other expressions. Thus $d((a \diagdown F_1) F_2 \cdots F_m) \leqslant 1 + \max\{d(a \diagdown F_1), d(F_2), ..., d(F_m)\} \leqslant 1 + \max\{d(F_1), d(F_2), ..., d(F_m)\} = d(E)$, where the last inequality follows from the inductive assumption. For case 4(b), we can expand $a \diagdown (F_2 \cdots F_m)$ to

$$(a \diagdown F_2) F_3 \cdots F_m \quad \text{if } \lambda \notin F_2$$
$$(a \diagdown F_2) F_3 \cdots F_m \cup a \diagdown (F_3 \cdots F_m) \quad \text{if } \lambda \in F_2.$$

Clearly, the same argument will apply to all such terms of the form

$$(a \setminus F_i) F_{i+1} \cdots F_m \text{ , i.e., } d((a \setminus F_i) F_{i+1} \cdots F_m) \leqslant d(E).$$

Since, in case 4(b), $a \setminus E$ is a union of such terms, it follows that $d(a \setminus E) \leqslant d(E)$.

We have shown that for each star-free E of proper dot-depth $d(a \setminus E) \leqslant d(E)$. But since every star-free event can be denoted by a star-free expression of proper dot-depth, it follows that $d(a \setminus R) \leqslant d(R)$ for all star-free R.

THEOREM 2.12. *Let U and V be any events and let R be star-free. Then $d(U \setminus R/V) \leqslant d(R)$, $d(u \setminus R) \leqslant d(R)$.*

Proof. We first show that $d(u \setminus R) \leqslant d(R)$ for all $u \in I$. Let $l(u) = n$. In Lemma 2.11 we have shown that $d(u \setminus R) \leqslant d(R)$ for $l(u) = 1$. This provides a basis for a proof by induction on $l(u)$:

$$d((xa) \setminus R) = d(a \setminus (x \setminus R) \leqslant d(x \setminus R).$$

Thus dot-depth cannot be increased by taking left quotients with respect to words.

The same argument shows that right quotients cannot increase dot-depth. Next $u \setminus R/v = (u \setminus R)/v$; hence two-sided quotients cannot increase dot-depth.

Next $U \setminus R = \bigcup_{u \in U} u \setminus R$. However, R has a finite number of distinct quotients $u_i \setminus R$, $i = 1, 2, \ldots, r$. Hence $U \setminus R$ is a finite union of some or all of the $u_i \setminus R$ and $d(U \setminus R) \leqslant \max\{d(u_i \setminus R),\ i = 1, 2, \ldots, r\} \leqslant d(R)$. The same argument applies to right quotients and two-sided quotients by events.

COROLLARY 2.13. *If the finite automaton $\mathscr{A}$ accepting star-free R is strongly connected, then*

$$d(w \setminus R) = d(R) \qquad \text{for all } w \in I.$$

Proof. Since $\mathscr{A}$ is strongly connected, for each $w \in I$ there exists $x \in I$ such that $x \setminus (w \setminus R) = R$. Now if $d(w \setminus R) < d(R)$ for some w, then also $d(R) = d(x \setminus (w \setminus R)) \leqslant d(w \setminus R) < d(R)$ which is a contradiction.

We can also show that dot-depth of R cannot be increased by multiplying by a finite event provided $d(R) \geqslant 1$.

THEOREM 2.14. *For any star-free event R with $d(R) \geqslant 1$ and for any finite event P, $d(PR) \leqslant d(R)$.*

Proof. Let $P = \{w_1, w_2, \ldots, w_r\}$, $w_i \in I$, $i = 1, 2, \ldots, r$. Then $PR = \bigcup_{i=1}^{r} w_i R$ and $d(PR) \leqslant \max\{d(w_i R) \mid 1 \leqslant i \leqslant r\}$. Thus it is sufficient to prove that for any $w \in I$, $d(wR) \leqslant d(R)$. Let E be a star-free expression of proper dot-depth for R. The proof is by induction on the number n of regular operators in E.

Basis, $n = 1$. The only operator resulting in $d \geqslant 1$ is concatenation. Thus E is a single word and so is wE and $d(wE) = d(E) = 1$.

Induction step $n > 1$. Now assume the theorem holds for all star-free E of proper dot-depth $\geqslant 1$.

(a) $E = F_1 \cup F_2$. If both $d(F_1)$ and $d(F_2)$ are greater than 0, then by the induction hypothesis,

$$d(wF_i) \leqslant d(F_i),$$

and

$$d(wE) = d(wF_1 \cup wF_2) \leqslant \max\{d(wF_1), d(wF_2)\} \leqslant \max\{d(F_1), d(F_2)\} = d(E).$$

If, however, $d(F_1) = 0$, then $d(E) = d(F_2)$ and $d(wF_1) \leqslant 1$. Thus $d(F_2) \geqslant 1$ and by the hypothesis $d(wF_2) \leqslant d(F_2)$. Thus $d(wE) \leqslant \max\{1, d(wF_2)\} = d(wF_2) \leqslant d(F_2) = d(E)$. The same applies if $d(F_2) = 0$.

(b) $E = F_1 \cap F_2$. One verifies that $wE = (wF_1) \cap (wF_2)$ and the argument is the same as for union.

(c) $E = \bar{F}$. One can verify that $wE = w\bar{F} = (wI) \cap \overline{wF}$. By the hypothesis, $d(wF) \leqslant d(F)$. Hence $d(wE) = d(wI \cap \overline{wF}) \leqslant \max\{d(w\bar{\phi}), d(\overline{wF})\} = \max\{1, d(wF)\} \leqslant \max\{1, d(F)\} = d(F) = d(E)$.

(d) $E = F_1 F_2 \cdots F_m$. Clearly, $d(wE) \leqslant d(E)$.

Thus the induction step holds.

COROLLARY 2.15. *Let R be a star free event with $d(R) > 1$. Then for any two nonnegative integers m, n there exist words u, $v \in I$, $l(u) = m$, $l(v) = n$, such that $d(u \setminus R/v) = d(R)$.*

Proof. First let $n = 0$; then we are dealing with $u \setminus R/\lambda = u \setminus R$. We can always express any regular event as $R = \bigcup_{l(w)=m} w(w \setminus R) \cup P$ where P consists of words of length less than m. Now suppose for all w, $l(w) = m$, $d(w \setminus R) < d(R) > 1$. From the expression for R it is clear that $d(R) = \max\{d(w(w \setminus R)) \mid l(w) = m\}$. By Theorem 2.12 $d(w(w \setminus R)) \leqslant d(w \setminus R)$. Thus $d(R) \leqslant d(w \setminus R) < d(R)$, which is a contradiction. Thus for each m we can find u such that $d(u \setminus R) = d(R)$. Similarly, for the event $(u \setminus R)$ and for each n, we can find $v \in I$ such that $(u \setminus R)/v = (u \setminus R/v)$ has dot-depth $d(u \setminus R) = d(R)$.

3. TWO-SIDED QUOTIENTS OF STAR-FREE EVENTS

As was indicated in the previous section, in certain cases the dot depth of a star-free event is preserved under the quotient operation. However, it is shown in this section that under no circumstances can the dot depth of a star free event R be preserved under all two-sided quotients. In fact, every star free event has at least one quotient of dot depth 0, namely ϕ or $\bar{\phi} = I$, as indicated by the next theorem.

THEOREM 3.1. *For any star-free event R there exist words u, v such that either* $u \setminus R/v = \phi$ or $u \setminus R/v = I$.

Before proceeding with the proof of this theorem we prove two lemmas.

LEMMA 3.2. *Let R be a nonempty star-free event such that for any two words u, v there exist words x, y such that $ux \setminus R / yv = \phi$. Then there exist words w_1, w_2 such that $R \subseteq \overline{Iw_1Iw_2I}$.*

Proof. Let $P = \{p_1 = R, p_2, ..., p_n\}$ be the set of all distinct two-sided quotients of the form $w \setminus R/w'$, $w, w' \in I$, of R. By assumption, there exist words x_1, y_1 such that $x_1 \setminus R/y_1 = \phi$; we may assume without loss of generality that this empty quotient is p_n. Clearly $x_1 \setminus p_n / y_1 = x_1 \setminus p_1 / y_1 = p_n$. Thus the set

$$P_1 = \{p_j \mid p_j = x_1 \setminus p_i / y_1 \quad \text{for some } 1 \leqslant i \leqslant n\}$$

is a proper subset of P. If $P_1 = \{p_n\}$, then define $w_1 = x_1$, $w_2 = y_1$ and then we have $xw_1 \setminus R / w_2 y = \phi$ for any two words x, y, showing that $R \subseteq \overline{Iw_1Iw_2I}$ as required. Now assume $\#P_1 > 1$;[2] then there exists a quotient p_{i_1} in P_1 for some $1 \leqslant i_1 < n$. Then $p_{i_1} = u_1 \setminus R / v_1$ for some words u_1, v_1. By assumption of the lemma, there exist words x_2, y_2 such that $u_1 x_2 \setminus R / y_2 v_1 = \phi$. Thus $x_2 \setminus p_{i_1} / y_2 = x_2 \setminus p_n / y_2 = p_n$, indicating that the set $P_2 = \{x_2 \setminus p_i / y_2 \mid p_i \in P_1\}$ is properly contained in P_1. Continuing in this fashion we obtain a sequence $(x_1, y_1), (x_2, y_2), ..., (x_r, y_r)$, $1 \leqslant r \leqslant n - 1$, of pairs of words and a sequence $P = P_0 \supsetneqq P_1 \supsetneqq P_2 \supsetneqq \cdots \supsetneqq P_r = \{p_n\}$ such that $P_i = \{p_k \mid p_k = x_i \setminus p_j / y_i \text{ for some } p_j \in P_{i-1}\}$ for each $i = 1, 2, ..., r$. Define $w_1 = x_1 x_2 \cdots x_r$ and $w_2 = y_r \cdots y_2 y_1$. Now for any two words u, v, $u \setminus R/v = p_i \in P$ and thus by the construction of the sets P_i $uw_1 \setminus R / w_2 v = w_1 \setminus p_i / w_2 = p_n = \phi$. This implies that $R \subseteq \overline{Iw_1Iw_2I}$, which completes the proof.

LEMMA 3.3. *Let R be a star-free event such that for any two words u, v, there exist words x, y such that $ux \setminus R / yv = I$. Then there exist words w_1, w_2 such that $Iw_1Iw_2I \subseteq R$.*

Proof. Consider the event $R' = \bar{R}$. Since, clearly, $z_1 \setminus \bar{R} / z_2 = \overline{z_1 \setminus R / z_2}$ for any words z_1, z_2, R' satisfies the conditions of Lemma 3.2; hence there exist w_1, w_2 such that $R' \subseteq \overline{Iw_1Iw_2I}$, and the result follows.

Proof of Theorem 3.1. We shall prove the following equivalent assertion: (*) For any star-free event R and for any words u, v, there exist words x, y such that $ux \setminus R / yv = \phi$ or $ux \setminus R/yv = I$.

[2] $\#A$ denotes the number of elements in set A.

Since $ux \setminus R / yv = x \setminus (u \setminus R / v) / y$ and since a quotient $u \setminus R / v$ of a star-free event R is also star-free, the above assertion (*) is equivalent to the assertion of the theorem.

Thus let E be a star-free regular expression denoting R. The proof will be given by induction on the number n of regular operators $\cup$, $\cdot$, $^{-}$ appearing in E.[3] If $n = 0$ then the assertion is trivial. Thus assume (*) holds for all regular events R' denoted by star-free expressions E' with n or less regular operators, and suppose the star-free expression E denoting R has $n + 1$ operators. There are three cases:

(a) $E = E_1 \cup E_2$. Consider any given pair of words u, v. Since E_1 and E_2 have no more than n regular operators each, we can apply the induction hypothesis to $R_1 = | E_1 |$ and $R_2 = | E_2 |$. If for $i = 1$ or 2, there exist words x, y such that $ux \setminus R_i / yv = I$, then clearly also $ux \setminus R / yv = I$ and (*) holds for R and the pair u, v. Thus suppose that there exist no words x, y such that $ux \setminus R_i / yv = I$ for either $i = 1$ or $i = 2$. Then by induction hypothesis there exist words x_1, y_1 such that $ux_1 \setminus R_1 / y_1 v = \phi$. Now, applying the induction hypothesis to R_2 with the pair of words $ux_1, y_1 v$, there exist words x_2, y_2 such that $ux_1 x_2 \setminus R_2 / y_2 y_1 v$ equals ϕ or I. But we have assumed above that $ux \setminus R_2 / yv \neq I$ for any pair of words x, y. Hence $ux_1 x_2 \setminus R_2 / y_2 y_1 v = \phi$ and clearly also $ux_1 x_2 \setminus R_1 / y_2 y_1 v = x_2 \setminus (ux_1 \setminus R_1 / y_1 v) / y_2 = \phi$, which implies $ux_1 x_2 \setminus R / y_2 y_1 v = \phi$ as required.

(b) $E = \bar{E}_1$, where E_1 has n regular operators. Then by induction hypothesis, for any two words u, v, there exist words x, y such that $ux \setminus | E_1 | / yv$ coincides with ϕ or I. But then also $ux \setminus R / yv = \overline{ux \setminus | E_1 | / yv}$ is ϕ or I.

(c) $E = E_1 E_2$. Let $| E_i | = R_i$, $i = 1, 2$, as before. Clearly(*) holds if the words u, v satsify $u \setminus R / v = \phi$. Thus we may assume $u \setminus R / v \neq \phi$. There are three possibilities:

(1) $R_2 / v \neq \phi$, and there exists a pair of words x', y' such that $ux' \setminus R_1 / y' = I$. In this case, for any $w \in R_2 / v$, we have $ux' \setminus R_1 R_2 / y'wv = I$; thus the pair of words $x', y'w$ yields the required result.

(2) $u \setminus R_1 \neq \phi$ and there exist words x', y' such that $x' \setminus R_2 / y'v = I$. This case is symmetric to case (1) and the result follows in a similar way.

(3) Neither case (1) nor (2) holds. The there are four cases:

(i) $u \setminus R_1 = R_2 / v = \phi$. Then clearly $u \setminus R / v = \phi$, contradicting our assumption.

(ii) For any pair of words x', y', $ux' \setminus R_1 / y' \neq I$ and $x' \setminus R_2 / y'v \neq I$. Applying the induction hypothesis to R_1 we get: For any pair of words u', v' there exists a pair of words x, y such that $uu'x \setminus R_1 / yv' = \phi$. But since $uu'x \setminus R_1 / yv' =$

[3] By De Morgan's Law the intersection operator need not be included. Also concatenation is considered here as a binary operator; thus E is not in full product form.

$u'x \backslash (u \backslash R_1) / yv'$, $u \backslash R_1$ satisfies the requirements of Lemma 3.2 and hence there exist words w_1, w_2 such that $u \backslash R_1 \subseteq \overline{Iw_1Iw_2I}$. Now, applying the induction hypothesis to R_2, we obtain in a similar way two words w_1', w_2' such that $R_2 / v \subseteq \overline{Iw_1'Iw_2'I}$. But then $u \backslash R_1R_2 / v \subseteq \overline{Iw_1Iw_2Iw_1'Iw_2'I}$. This implies $uw_1w_2 \backslash R / w_1'w_2'v = \phi$, which shows that (*) holds for R.

(iii) $u \backslash R_1 = \phi$ and $R_2 / v \neq \phi$. Now let $u_1, u_2, \ldots, u_k$ be the set of all distinct prefixes of u such that $u_i \in R_1$, and let $u = u_iv_i$, $i = 1, \ldots, k$. We now have two subcases:

(iii)$_A$ There exists i, $1 \leqslant i \leqslant k$, for which there exists a pair of words x, y such that $v_ix \backslash R_2 / yv = I$; then $u_iv_ix \backslash R_1R_2 / yv = ux \backslash R / yv = I$, showing that R with the pair u, v satisfy (*).

(iii)$_B$ For any i, $1 \leqslant i \leqslant k$, and for any two words $x, y, v_ix \backslash R_2 / yv \neq I$. The induction hypothesis (*) applied to R_2 and the words v_1, v yields the existence of words x_1, y_1 such that $v_1x_1 \backslash R_2 / y_1v = \phi$. Moreover, considering R_2 and the words v_2x_1, y_1v, (*) implies the existence of words x_2, y_2 such that $v_2x_1x_2 \backslash R_2 / y_2y_1v = \phi$. Proceeding in this fashion, one eventually ends up with sequences $x_1, x_2, \ldots, x_k$, $y_1, y_2, \ldots, y_k$ satisfying

$$v_ix_1x_2 \cdots x_i \backslash R_2 / y_iy_{i-1} \cdots y_1v = \phi \qquad \text{for all } i = 1, \ldots, k \qquad (**)$$

Now consider the event

$$W = ux_1x_2 \cdots x_k \backslash R_1R_2 / y_ky_{k-1} \cdots y_1v.$$

Suppose $W \neq \phi$, and let $w \in W$. Then $ux_1 \cdots x_kwy_k \cdots y_1v \in R_1R_2 = R$; but since, by our assumption, $u \backslash R_1 = \phi$, there exists a decomposition of u, $u = u_iv_i$, $1 \leqslant i \leqslant k$, such that $u_i \in R_1$ and $v_ix_1 \cdots x_kwy_k \cdots y_1v \in R_2$. But then

$$x_{i+1} \cdots x_kwy_k \cdots y_{i+1} \in v_ix_1 \cdots x_i \backslash R_2 / y_i \cdots y_1v,$$

which is impossible since the latter quotient is empty by (**). Hence $W = \phi$, i.e., $ux_1 \cdots x_k \backslash R / y_k \cdots y_1v = \phi$, showing that (*) holds for R.

(iv) $R_2 / v = \phi$ and $u \backslash R_1 \neq \phi$. The proof is similar to that of case (iii) and is, therefore, omitted.

This completes the proof of Theorem 3.1.

COROLLARY 3.4. *Every star-free event R satisfies one of the following*:

(a) There exist words u, v, u', v' such that $u \backslash R / v = \phi$ and $u' \backslash R / v' = I$.

(b) There exist words w_1, w_2 such that $R \subseteq \overline{Iw_1Iw_2I}$.

(c) There exist words w_1, w_2 such that $Iw_1Iw_2I \subseteq R$.

The anonymous referee has pointed out that an alternate proof of Theorem 3.1 can be obtained by applying Lemma 3.3 (p.456) of Ref. [9]. One can prove, using this lemma, that if M is a finite monoid with only trivial subgroups, then there exist elements m_1, m_2, m_3 of M such that $m_1 M m_2 = m_3$. Now take M to be the syntactic monoid of R and let γ be the homomorphism from I to M induced by Myhill's congruence relation mod R [1, 3]. Then from the above equation one obtains $\gamma^{-1}(m_1) I \gamma^{-1}(m_2) \subseteq \gamma^{-1}(m_3)$ and for any $u \in \gamma^{-1}(m_1)$, $v \in \gamma^{-1}(m_2)$, $uIv \subseteq \gamma^{-1}(m_3)$, hence $I = u \diagdown \gamma^{-1}(m_3) / v$. Since $\gamma^{-1}(m_3)$ must be contained either in R or in $\bar{R}$, it follows that $u \diagdown R / v$ equals either I or ϕ, which proves Theorem 3.1.

4. Construction of Star-Free Expressions and an Upper Bound for Dot-Depth

The family of star-free events has several interesting characterizations. First, it has been shown [2, 4] that an event is star-free if an only if it is group-free, i.e., its syntactic monoid [3] has no nontrivial subgroups. Furthermore, the group-free events have been shown to be equivalent to the regular noncounting languages [2] and, by the Krohn–Rhodes theory, also correspond to the events recognizable by a cascade product of reset automata [3, 8]. Finally, the star-free events are precisely the events definable in McNaughton's L-language [2].

In this section, we utilize the Krohn–Rhodes result with Zeiger's decomposition methods [8] to derive an algorithm for constructing star-free expressions for group-free events and to obtain an upper bound for the dot-depth of such events.

Definition 4.1. A semiautomaton $\mathscr{S}$ is a triple, $\mathscr{S} = \langle A, Q, M \rangle$, where A is the input alphabet, Q the set of states, and M the transition function, as in the definition of a finite automaton. ($\mathscr{S}$ can be considered as a finite automaton in which the initial state and the set of final states are not specified.)

Definition 4.2. Let $\mathscr{S}_1 = \langle A, Q_1, M_1 \rangle$ and $\mathscr{S}_2 = \langle A \times Q_1, Q_2, M_2 \rangle$ be two semiautomata. The *cascade product* $\mathscr{S}_1 \to \mathscr{S}_2$ of $\mathscr{S}_1$ and $\mathscr{S}_2$ is the semiautomaton $\mathscr{S}_1 \to \mathscr{S}_2 = \langle A, Q_1 \times Q_2, M \rangle$, where $M((q, q'), a) = (M_1(q, a), M_2(q', (a, q)))$ for all $q \in Q_1$, $q' \in Q_2$ and $a \in A$. Let $\mathscr{S}_1 = \langle A, Q_1, M_1 \rangle$, $\mathscr{S}_2 = \langle A \times Q_1, Q_2, M_2 \rangle, \ldots$ $\mathscr{S}_n = \langle A \times Q_1 \times Q_2 \times \cdots \times Q_{n-1}, Q_n, M_n \rangle$ be n semiautomata. The cascade product of $\mathscr{S}_1, \mathscr{S}_2, \ldots, \mathscr{S}_n$ is the semiautomaton $\mathscr{S}$ defined inductively by

$$\mathscr{S} = (\mathscr{S}_1 \to \mathscr{S}_2 \to \cdots \to \mathscr{S}_{n-1}) \to \mathscr{S}_n .$$

Definition 4.3. $\mathscr{S} = \langle A, Q, M \rangle$ is a *reset* semiautomaton iff for any $a \in A$, either $M(Q, a) = \{q_a\}$ for some $q_a \in Q$ or $M(q, a) = q$ for all $q \in Q$.

Notation. Let $A_q = \{a \in A \mid M(Q, a) = \{q\}\}$ be the set of input letters *resetting* $\mathscr{S}$ *to state* q. Let $A_R = \bigcup_{q \in Q} A_q$ be the set of all *reset* inputs. Let

$$A_I = \{a \in A \mid M(q, a) = q \qquad \text{for all } q \in Q\}$$

be the set of all *identity* inputs. Clearly, in a reset semiautomaton, $A = A_R \cup A_I$.

For any two states p, q of $\mathscr{S}$ let $\mathscr{S}_{pq} = \{x \in A^* \mid M(p, x) = q\}$.

Notation. For any $n = 1, 2, \ldots$, let $U(M_n)$ denote the closure under union of the monoid M_n (defined in Section 2). Due to the distributivity of concatenation over union, $U(M_n)$ is also a monoid.

THEOREM 4.1. *If* $\mathscr{S} = \langle A, Q, M \rangle = \mathscr{S}_1 \to \mathscr{S}_2 \to \cdots \to \mathscr{S}_n$, *where each* $\mathscr{S}_i$, $1 \leqslant i \leqslant n$, *is a reset semiautomaton, then for any* $q, q' \in Q$, $\mathscr{S}_{qq'} \in U(M_{n+1})$, *and a star-free expression of dot-depth at most* $n + 1$ *can be effectively constructed for* $\mathscr{S}_{qq'}$.

Proof. (by induction on n).

Basis. $n = 1$. One can verify that

$$\mathscr{S}_{qq'} = \delta_{qq'} \cup \bar{\phi} A_{q'} \cdot \overline{\bar{\phi} A_R \bar{\phi}}$$

where $\delta_{qq'} = \lambda$ iff $q = q'$, and ϕ otherwise. Hence $\mathscr{S}_{qq'} \in U(M_2)$.

Induction step. Assume the theorem holds for some $n \geqslant 1$ and let $\mathscr{S} = \mathscr{S}' \to \mathscr{S}_{n+1}$, where $\mathscr{S}' = \langle A, P, N \rangle = \mathscr{S}_1 \to \mathscr{S}_2 \to \cdots \to \mathscr{S}_n$. By the induction hypothesis $\mathscr{S}'_{pp'} \in U(M_{n+1})$ for any $p, p' \in P$, and the corresponding star-free expressions can be effectively constructed. Next, for any two states s, s' of $\mathscr{S}_{n+1}$,

$$(\mathscr{S}_{n+1})_{ss'} = \delta_{ss'} \cup \bar{\phi} C_{s'} \cdot \overline{\bar{\phi} C_R \bar{\phi}}$$

where $C = A \times P$, the above expression is over the alphabet C and complements are relative to C^*.

Let $p, p' \in P$, and let q, q' be the two states of $\mathscr{S}$ which correspond to (p, s) and (p', s') respectively in $\mathscr{S} = \mathscr{S}' \to \mathscr{S}_{n+1}$. Then from the expression $(\mathscr{S}_{n+1})_{ss'}$ over alphabet C we derive the following star-free expression for $\mathscr{S}_{qq'}$ over alphabet A.

$$\mathscr{S}_{qq'} = \delta_{qq'} \cup \Big(\bigcup_{(a, p_i) \in C_{s'}} \mathscr{S}'_{pp_i} a Y_i \Big),$$

where

$$Y_i = \overline{\Big(\bigcup_{(a', p_j) \in C_R} \mathscr{S}'_{N(p_i, a) p_j} a' \bar{\phi} \Big)} \cap \mathscr{S}'_{N(p_i, a) p'} .$$

Now since $\mathscr{S}'_{p_1p_2}$ for each $p_1, p_2 \in P$ is in $U(M_{n+1})$ by the inductive assumption and since $U(M_{n+1})$ is a monoid, Y_i is in B_{n+2} and so is $\mathscr{S}'_{pp_i}a$. Thus $\mathscr{S}_{qq'}$ is in $U(M_{n+2})$ and is therefore of dot-depth at most $n + 2$.

By the well-known decomposition theorem of Krohn and Rhodes [3, 8], any group-free event R is recognizable by an automaton whose semiautomaton is a cascade product of reset semiautomata. This cascade product of reset semiautomata can be effectively found for any given group-free R, using Zeiger's decomposition methods with set systems [8]. Combining this with Theorem 4.1, one obtains an algorithm for constructing a star-free expression for any given group-free event.[4] We shall employ here Zeiger's decomposition methods to obtain an upper bound for the dot-depth of a star-free R in terms of the number of states in the reduced automaton recognizing R.

DEFINITION 4.4. Let R be a star-free event and let $\mathscr{A}_0(R) = (Q, M, q_1, F)$ be the reduced deterministic automaton recognizing R. Define

$$i(R) = \max\{\#Q' \mid Q' \subsetneqq Q \quad \text{and} \quad Q' = M(Q, a) \quad \text{for some } a \in A\}$$

THEOREM 4.2. *Any group-free event R is recognizable by an automaton whose semiautomaton is a cascade product of $i(R)$ reset semiautomata.*

Outline of proof. Let $\mathscr{A}_0(R) = (Q, M, q_1, F)$, let $\mathscr{S}$ be the semiautomaton of $\mathscr{A}_0(R)$, and let $\sigma' = \{Q' \subsetneqq Q \mid Q' = M(Q, a)$ for some $a \in A\}$. Let σ be the set system obtained from σ' by removing all sets properly contained in others. Using the set system σ, decompose $\mathscr{S}$ (via Zeiger's method) into a cascade product $\mathscr{S}_1 \to \mathscr{S}_2$, so that $\mathscr{S}_1$ is a reset semiautomaton and $\mathscr{S}_2$ is a group-free automaton with $i(R)$ states. $\mathscr{S}_2$ can then be decomposed, using Zeiger's standard decomposition method, into a cascade product of at most $i(R) - 1$ reset semiautomata.

Combining Theorems 4.1 and 4.2 we obtain

COROLLARY 4.3. *The dot-depth of any group-free event R is bounded by $i(R) + 1$ (cf. Definition 4.4). In particular, $d(R)$ cannot exceed the number of states in the reduced deterministic automaton recognizing R.*

Remark. The dot-depth $d(R)$ may, in fact, equal $i(R) + 1$, as is the case for the event $R = I02^*$ (cf. Lemma 2.9).

Finally, we would like to point out that the problem of determining dot-depth of star-free events is open; in fact, it is not known whether there exist star-free events of arbitrary dot-depth.

[4] This algorithm has also been independently found by Meyer [6].

References

1. S. Papert and R. McNaughton, "Non-Counting Automata," (Monograph), to appear.
2. S. Papert and R. McNaughton, On topological events, *in* "Theory of Automata," University of Michigan, Engineering Summer Conferences, 1966.
3. A. Ginzburg, "Algebraic Theory of Automata," Academic Press, New York/London, 1968.
4. M. P. Schützenberger, "On a Family of Sets Related to McNaughton's L-Language," Automata Theory, (E. R. Caianiello, Ed.), pp. 320–324, Academic Press, New York/London, 1966.
5. R. S. Cohen and J. A. Brzozowski, On star-free events, Proc. Hawaii Internat. Conf. System Sci., Honolulu, Hawaii, 1968.
6. A. R. Meyer, A note on star-free events, *J. Assoc. Comput. Mach.* **16** (1969), 220–225.
7. J. A. Brzozowski, Derivatives of regular expressions, *J. Assoc. Comput. Mach.* **11** (1964), 481–494.
8. P. Zeiger, Cascade decomposition of automata using covers, *in* "The Algebraic Theory of Machines, Languages and Semigroups" (M. A. Arbib, Ed.), pp. 55–800, Academic Press, New York/London, 1968.
9. K. Krohn and J. Rhodes, "Algebraic Theory of Machines, I. Prime Decomposition Theorem for Finite Semigroups and Machines," *Trans. Am. Math. Soc.* **116** (1965), 450–464.