Multitape NFA: Weak Synchronization of the Input Heads

Ömer Eğecioğlu[1], Oscar H. Ibarra[1] and Nicholas Tran[2]

[1]Department of Computer Science
University of California at Santa Barbara
{omer,ibarra}@cs.ucsb.edu

[2]Department of Mathematics & Computer Science
Santa Clara University
ntran@math.scu.edu

SOFSEM 2012

Outline

Motivation & FA Models

String synchronization:

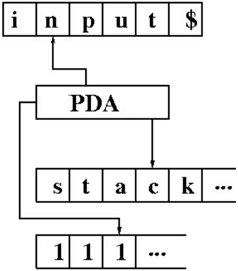
- Vulnerability analysis (SQL Injection Attacks)
- Verification
- Reachability (Reachability Analysis of String Systems)

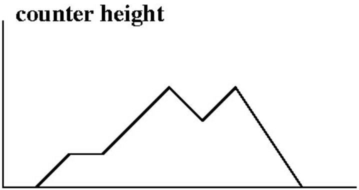
Finite Automata Models of String Systems:

- Single-track DFA to represent individual strings [Xu et al.], [Shannon et al.]
- Multi-track DFA to represent groups of strings [Yu et al.]
- Multi-tape and multi-head FA/PDA to represent groups of strings [Ibarra et al.]

NFA, Stack and Counter Machines

- *NFA*: *Nondeterministic*, *one-way* FA
- *PDA*: *Nondeterministic*, *one-way* FA with a stack
- *CM*: PDA where stack alphabet is unary
- *Reversal-bounded CM*: the stack height function has a bounded number of local maxima/minima





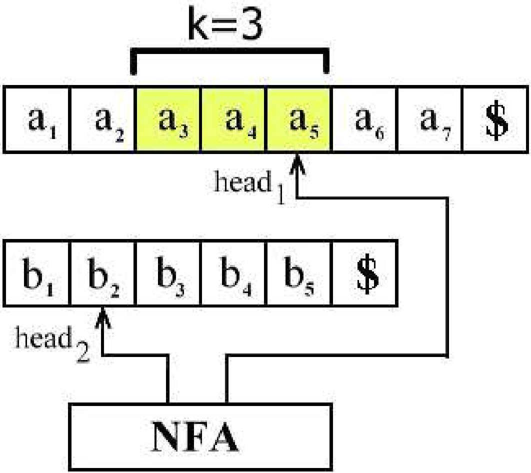
Multitrack vs. Multitape Automata

- Multitape automata are more expressive but have Turing-complete decision problems.
- Best of both worlds: expressiveness of multitape but decidability of multitrack automata.

  Naturally leads to *synchronization/synchronizability*:
- Does a *multitape* automaton "behave" like a *multitrack* automaton?
- Can a *multitape* automaton be converted to an equivalent *multitrack* automaton?

Basic Model

2-tape NFA, input $(a_1a_2a_3a_4a_5a_6a_7\$, b_1b_2b_3b_4b_5\$)$.



Weak/Strong Synchronization & Synchronizability

Given an n-tape NFA M,

- M is *strongly k-synchronized* if no two heads, neither of which is on \$, are more than k cells apart for *any* computation (accepting or not).
- M is *weakly k-synchronized* if for each *accepted* input tuple *there exists* an accepting computation in which no two heads, neither of which is on \$, are more than k cells apart.
- M is *strongly/weakly synchronized* if it is strongly/weakly k-synchronized for some k.
- M is *strongly/weakly k-synchronizable* if there is a strongly/weakly k-synchronized M' such that $L(M) = L(M')$.

Strong Synchronization $\neq$ Weak Synchronization

$L = \{(a^m\$, b^n\$) \mid m, n > 0\}$.

M: 2-tape NFA. On input $(a^m\$, b^n\$)$, executes one of:

1. M reads $a^m\$$ on tape 1 until head 1 reaches \$, and then reads $b^n\$$ on tape 2 until head 2 reaches \$, then accepts.
2. M reads the symbols on the two tapes simultaneously until one head reaches \$. Then the other head scans the remaining symbols on its tape and accepts.

M is not strongly synchronized, but it is weakly synchronized (weakly 0-synchronized).

Strongly synchronized $\rightarrow$ weakly synchronized, but not conversely.

A Strong Synchronization Result [DCFS11]

Theorem

It is decidable to determine, given an n-tape NFA M, whether it is strongly k-synchronized for some k. If this is the case, the smallest such k can be found.

Recall:
M is strongly k-synchronized if no two heads, neither of which is on \$, are more than k cells apart for *any* computation (accepting or not).

Sample Strong Synchronization Results [DCFS11]

- One-way 2-tape 2-counter DCM (for a fixed k): undecidable
- One-way n-tape PDA + reversal-bounded counters (for some k): decidable

- Two-way 2-tape DFA (for some k): undecidable
- Two-way n-tape NFA (for a fixed k): decidable

Definitions: Bounded and Unary Inputs

Definition

Let $a_1, a_2, \ldots, a_k$ be arbitrary symbols.

A string x is

- *unary* if $x \in a_1^*$;
- *bounded* if $x \in a_1^* a_2^* \cdots a_k^*$.

A tuple of strings $(x_1, \ldots, x_n)$ is

- *unary* if each x_i is unary;
- *bounded* if each x_i is bounded;
- *all-but-one bounded (ABO)* if all x_i except one is bounded.

Definitions: Semilinear sets

Definition

Linear set in $\mathbb{N}^k$: $Q = \{v_0 + t_1 v_1 + \cdots + t_n v_n \mid t_1, \ldots, t_n \in \mathbb{N}\}$.
v_0 constant vector, $v_1, \ldots, v_n$ *period* vectors.

Semilinear set: Finite union of linear sets.

Properties:

- Every finite subset of $\mathbb{N}^k$ is semilinear
- Semilinear sets are closed under (finite) union, complementation and intersection
- Disjointness, containment, and equivalence problems for semilinear sets are decidable

Definitions: Parikh mapping

Definition

$\Sigma = \{a_1, \ldots, a_k\}$, $w \in \Sigma^*$.
$|w|_{a_i}$: the number of occurrences of a_i in w.

Parikh image of a word: $P(w) = (|w|_{a_1}, \ldots, |w|_{a_k}) \in \mathbb{N}^k$

Parikh image of a language L: $P(L) = \{P(w) \mid w \in L\} \subseteq \mathbb{N}^k$.

Properties:

- The Parikh image of a language L accepted by a PDA is an effectively computable semilinear set [Parikh].
- The Parikh image of a language L accepted by a PDA with 1-reversal counters is an effectively computable semilinear set [Ibarra].

Weak Synchronization & Synchronizability Results

2-tape 2-ambiguous NFA is k-synchronized for a given k?

2-tape 2-ambiguous NFA has an equivalent 0-synchronized version?

undecidable

decidable

n-tape unambiguous NFA + reversal-bounded counters is k-synchronized for some k?

n-tape ABO NFA is k-synchronized for a given k?

2-tape unary NFA is k-synchronized for some k?

Helpful Fact: 0-synchronization and Regularity

Definition

Let $(x_1, \ldots, x_n)$ be a tuple of strings. Define $\langle x_1, \ldots, x_n \rangle$ to be an n-track string where the symbols of x_i's are left-justified and the shorter strings are right-filled with blanks (λ) to make all tracks the same length.

For a language L of n-tuples, $\langle L \rangle = \{\langle x \rangle : x \in L\}$.

Lemma

Let L be a set of n-tuples. L is accepted by a weakly 0-synchronized n-tape NFA iff $\langle L \rangle$ is regular.

Proof Sketch of an Undecidability Result

Theorem

It is undecidable to determine whether a 2-tape NFA M is 0-synchronized.

Proof

By reduction from the Post Correspondence Problem: given a PCP instance $I = (u_1, \ldots, u_n); (v_1, \ldots, v_n)$ we construct a 2-tape NFA M to accept

$$\begin{aligned} L \;=\; & \{(xc^i, yd^j) : i, j > 0, x \neq y\} \cup \\ & \{(xc^i, xd^j) : i, j > 0, j = 2i, x \in solution(I)\} \end{aligned}$$

such that M is weakly 0-synchronized iff I has no solutions.

Proof Sketch (cont.): A 2-tape NFA for L

$$L = \{(xc^i, yd^j) : i, j > 0, x \neq y\} \cup \{(xc^i, xd^j) : i, j > 0, j = 2i, x \in solution(I)\}$$

Define a 2-tape NFA M as follows: on input $w = (xc^i, yd^j)$, M nondeterministically does one of the following:

- M verifies that $x \neq y$ by moving the two heads in 0-synchrony from left to right until both heads reach their endmarkers $ and noting that x and y differ at some point;
- M guesses a sequence of indices $i_1, \ldots, i_k$, and verifies (one index at a time) that $x = u_{i_1} \cdots u_{i_k}$ and $y = v_{i_1} \cdots v_{i_k}$ and that $j = 2i$.

Proof Sketch (cont.)

$$L = \{(xc^i, yd^j) : i, j > 0, x \neq y\} \cup \{(xc^i, xd^j) : i, j > 0, j = 2i, x \in solution(I)\}$$

If I has no solutions, the second part is empty and hence $\langle L \rangle$ is regular.

If I has a solution, then L contains a string $w = (xc^i, xd^{2i})$ for some x and i. If $\langle L \rangle$ is regular, pump w to obtain $\langle xc^{i+k}, xd^{2i+k} \rangle \notin \langle L \rangle$ for some k, a contradiction.

A General Undecidability Result

Theorem

The following problems are undecidable, given a 2-tape (and hence multitape) NFA M:

1. *Is M weakly k-synchronized for a given k?*
2. *Is M weakly k-synchronized for some k?*
3. *Is there a 2-tape (multitape) NFA M' that is weakly 0-synchronized (or weakly k-synchronized for a given k, or weakly k-synchronized for some k) such that $L(M') = L(M)$?*

An Undecidability Result for 2-ambiguous NFA

Theorem

The following problems are undecidable, given a 2-ambiguous 2-tape (and hence multitape) NFA M:

1. *Is M weakly k-synchronized for a given k?*
2. *Is M weakly k-synchronized for some k?*
3. *Is there a 2-tape (multitape) NFA M' that is weakly 0-synchronized (or weakly k-synchronized for a given k, or weakly k-synchronized for some k) such that $L(M') = L(M)$?*

Some Decidability Results: ABO and Unary Cases

Theorem

It is decidable to determine, given a 3-tape ABO NFA M and a nonnegative integer k, whether M is weakly k-synchronized.

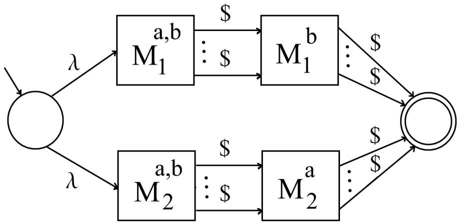
Unary case: input is of the form $(a^n\$, b^m\$)$.

Theorem

Suppose M is a unary 2-tape NFA. Then it is decidable whether or not M is weakly synchronized.

Unary Case

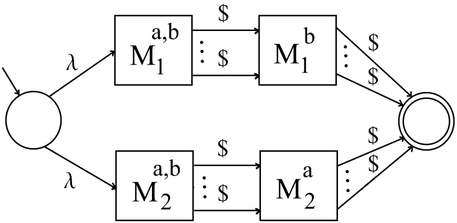
Transitions of $M \to$ letters from $\Sigma = \{a, b, \$\} \to$ NFA M'. Computation paths of $M \to$ words over Σ. Any accepting computation path w for input $(a^n\$, b^m\$)$ has $P(w) = (n, m)$.



There are two types of accepting computations:

1. $w = x\$y\$$ with $x \in \{a, b\}^*$ and $y \in \{b\}^*$,
2. $w = x\$y\$$ with $x \in \{a, b\}^*$ and $y \in \{a\}^*$.

Unary Case (cond.)



Whether M is weakly k-synchronized is equivalent to showing that for any given word $w \in L(M')$, there is a strongly k-synchronized word u such that either

- $ub^j \in L(M')$ and $P(w) = P(u) + (0, j)$, or
- $ua^i \in L(M')$ and $P(w) = P(u) + (i, 0)$.

Segueway to regular languages.

Synchronization and Regular Languages

Definition

$w \in \Sigma^*$ is *strongly k-synchronized* if for any factorization $x = uv$

$$-k \leq |u|_a - |u|_b \leq k.$$

A language L over Σ is:

Strongly k-synchronized if all of its words are strongly k-synchronized.

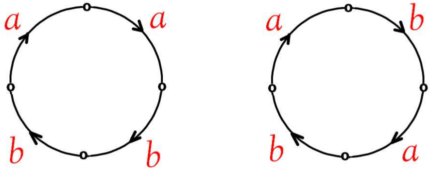
Strongly synchronized if it is strongly k-synchronized for some k.

Weakly k-synchronized if for every $w \in L$, there is a $w' \in L$ such that $P(w') = P(w)$ and w' is strongly k-synchronized.

Weakly synchronized if it is weakly k-synchronized for some k.

Balanced Cycles and Periods

- Balanced cycles of length 4:



- For a linear set

$$\{(a_0, b_0) + k_1(a_1, b_1) + \cdots + k_s(a_s, b_s) \mid k_1, k_2, \ldots, k_s \geq 0\},$$

that appears in $P(L)$, a period (a_i, b_i), $i > 0$ is balanced if $a_i = b_i$.

Synchronization and Regular Languages

Theorem

For a regular language L over $\Sigma = \{a, b\}$, the following are equivalent:

1. *L is strongly synchronized*
2. *L is weakly synchronized*
3. *The minimum state DFA for L has no* unbalanced simple cycles
4. *The Parikh image $P(L)$ has no* unbalanced periods

In Closing

An easy to state open problem on weak synchronization of multitape NFA.

An Open Decidability Problem

Problem

Is it decidable to determine, given a 2-tape NFA whose tapes are over bounded languages, whether it is weakly k-synchronized for some k?

Note:

- We only know this (in the positive) for the bounded unary case.

Sample Strong Synchronization Results [DCFS11]

- One-way 2-tape 2-counter DCM (for a fixed k): undecidable
- One-way n-tape PDA + reversal-bounded counters (for some k): decidable

- One-way 2-head DFA (for some k): undecidable
- One-way n-head PDA + reversal-bounded counters (for a fixed k): decidable

- Two-way 2-tape DFA (for some k): undecidable
- Two-way n-tape NFA (for a fixed k): decidable

- Two-way 2-head DCM (for a fixed k): undecidable
- Two-way n-head NFA (for a fixed k): decidable
- Two-way 2-head reversal-bounded DCM (for a fixed k): decidable