Mehryar Mohri
Speech Recognition
Courant Institute of Mathematical Sciences
Homework assignment 1
Due: October 14, 2010

1. Generalities.

   (a) Give an example of an unambiguous automaton that is not deterministic.

   (b) Give an example of a non-deterministic automaton over the alphabet $\{a, b\}$ whose determinization results in an exponentially larger automaton.

2. String-matching automata.

   (a) Give a regular expression for the set of strings over the alphabet $\{a, b\}$ ending with aba.

   (b) Use the FSM library or OpenFst to create a binary representation of a non-deterministic automaton representing that expression.

   (c) Show a graphical representation of that automaton.

   (d) Indicate how this automaton can be used to find in a text the occurrences of string aba.

   (e) Determinize and minimize that automaton (use software library) and show the graphical representation of the result.

3. Division with finite-state transducers. The transducer of Figure 1 can take as input a binary sequence representing an integer n_1 and return a binary sequence representing the integer n_2.

   (a) Verify that the transducer is deterministic (or sequential) and complete. Is the inverse transducer sequential?

   (b) How is n_2 related to n_1? (hint: think division modulo some integer n that you should specify).

   (c) What does the number of the destination state of an accepting path indicate about the integer represented by the string labeling that path?

   (d) Give an automaton representing the set of binary sequences representing numbers equal to 1 modulo 4.

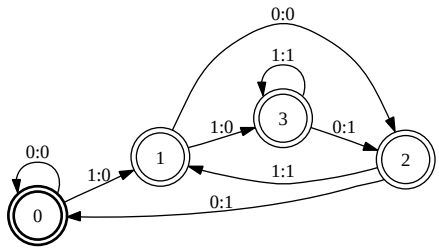

Figure 1: Finite-state transducer.

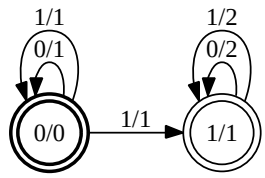

Figure 2: Weighted automaton computing integer values.

(e) Show how these results can be generalized: give a finite-state transducer that takes as input strings representing numbers in base x and that returns in the same base a string representing the result of the division by some integer p. Comment as before on the destination state of an accepting path. Note: you should describe the transducer in detail and prove its correctness.

4. Computing integer values with weighted automata. The previous exercise showed how transducers can be used for division, but the result was given as a sequence. Here, we wish to further *compute* the integer value of that sequence.

   (a) Give a *weighted regular expression* describing the weighted automaton of Figure 2 defined over the semiring $(\mathbb{R}, +, \times, 0, 1)$.

   (b) Show that it associates to any binary sequence its integer value (give a proof).

   (c) Give a weighted automaton with a similar property for sequences in an arbitrary base x.

   (d) Give a weighted automaton that takes as input a sequence in an arbitrary base x and returns the integer value of the division of

the number it represents by some integer p (hint: use previous exercise and composition).