Harvard University
Computer Science 121

Problem Set 2

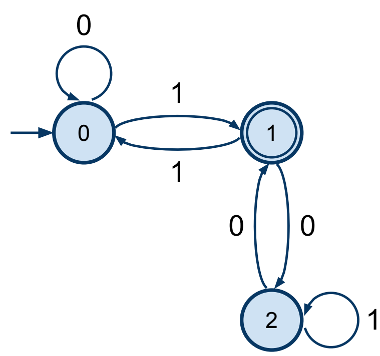
Due Tuesday, September 24, 2013 at 11:59 PM.
Submit your solutions electronically to cs121+ps2@seas.harvard.edu with "ps2 submission" in the subject line. The solutions to each part should be attached as separate PDF files, called lastname+ps2a.pdf, lastname+ps2b.pdf, and lastname+ps2c.pdf.
Late problem sets may be turned in until Friday, September 27, 2013 at 11:59 PM with a 20% penalty.
See syllabus for collaboration policy.

PART B (Graded by Paul)

PROBLEM 1 (8 points)

Shown below is a DFA for:

$$L = \{w \in \{0,1\}^* : \text{the number represented in binary notation by } w \text{ is equal to 1 mod 3}\}$$

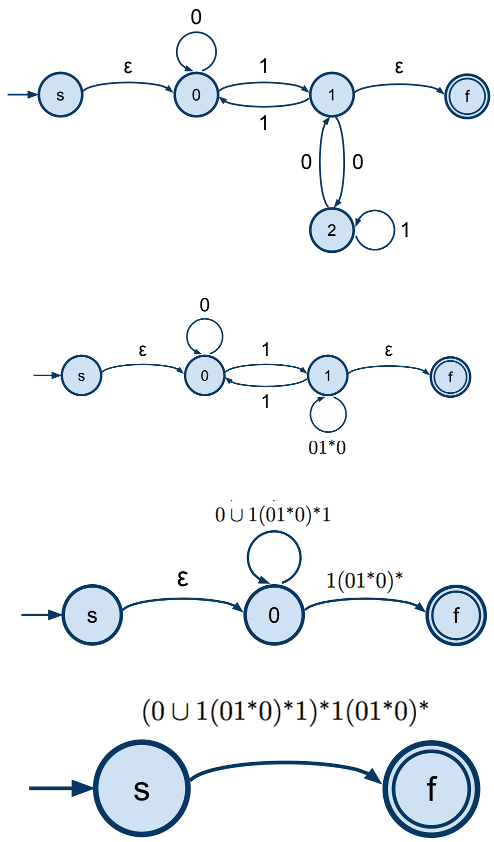


Convert this DFA to a regular expression for L, using the GNFA construction described in lecture. Show the full GNFA at each step of the construction.
Important: To receive credit for this problem, you *must* process the states in the order 2, 1, 0.

Solution:

See the figures below. We begin by adding new start/accept states with ε-transitions to the old ones. We then remove the original states one at a time, updating transitions as we go along. For example, in the first step, we delete state 2 and add a self-loop to state 0 with an appropriate regular expression. The final regular expression can be seen on the arrow in the last step of the construction.

GNFA construction for Problem 3.B:



PROBLEM 2 (1+1+1+1+1+1+1 points)

Classify the following sets as finite (in which case state the cardinality), countably infinite, or uncountably infinite. Give a sentence of justification.

(A) $\emptyset$
(B) $\{\emptyset\}$
(C) The set of all irrational numbers.
(D) The set of all syntactically valid C programs.
(E) The set of all languages over $\{a, b\}$ of strings of exactly 100 symbols.
(F) The set of all strings over $\{a, b\}$ longer than 100 symbols.
(G) The set of all languages over $\{a, b\}$ of strings longer than 100 symbols.

Solutions:
(A) $\emptyset$ is finite, and $|\emptyset| = 0$.
(B) $\{\emptyset\}$ is finite, and $|\{\emptyset\}| = 1$.
(C) The set of irrational numbers is uncountably infinite. We saw in section that $\mathbb{R}$ is uncountably infinite, and that $\mathbb{Q}$ is countably infinite; furthermore, $\mathbb{R}$ is the union of $\mathbb{Q}$ and the set of irrational numbers. So if the set of irrational numbers were countably infinite, then $\mathbb{R}$ would be as well, a contradiction.
(D) The set of syntactically valid C programs is countably infinite. There are infinitely many syntactically valid C programs (for instance, consider the set of programs that print all strings over $\{a, b\}$ before returning), but if we take Σ to be the set of all ASCII characters, the set of valid C programs is a subset of Σ^* and thus must be countable.
(E) This set is finite, and has cardinality $2^{2^{100}}$. There are 2^{100} strings of length exactly 100, as each symbol may be either an a or a b. A language is any subset of the set of all these strings, of which there are $2^{2^{100}}$.
(F) This set is countably infinite. We can order the strings by length then in lexicographic order (i.e. first list strings of length 101 in lexicographic order, then strings of length 102, and so on). Numbering the strings in this order, this gives a bijection between this set and $\mathbb{N}$.
(G) This set is uncountably infinite. This set is the power set of the set in (F) above, which we showed was countably infinite, and so this set is uncountably infinite.