# UMC-205 Assignment-3

Aditya Gupta SR No: 22205

# Question 1

Consider the language L defined on the alphabet A as:

$$L = \{w \in A^* \mid |w| \text{ in base } 10 \text{ has odd number of digits}\}$$

Clearly, we can define  $L^c$  as:

$$L^c = \{ w \in A^* \mid |w| \text{ in base 10 has even number of digits} \}$$

Claim: Both L and  $L^c$  do not contain an infinite regular set.

**Proof:** We can see that both L and  $L^c$  are infinite sets since there are an infinite number of odd and even numbers, which allows us to construct an infinite number of strings in L and  $L^c$  respectively.

Now, we will show that L does not contain an infinite regular set. We proceed by contradiction. Let us assume that L does contain an infinite regular set, say R. Since R is regular, it must satisfy the ultimate periodicity criteria of regular languages, which says that the set of lengths of strings in R must be ultimately periodic. More precisely, the set:

$$len(R) = \{ |w| \mid w \in R \}$$

must be ultimately periodic. This implies that it must contain an infinite arithmetic progression. However, since  $R \subseteq L$ , the set  $len(R) \subseteq len(L)$ . But len(L) is a set containing natural numbers with odd number of digits. Thus, len(R) is also a set containing natural numbers with odd number of digits.

Now assume the set len(R) has an infinite AP. Let this AP be represented by the sequence  $\{a+nd\}, \forall n \in \mathbb{N}$ . Let the number of digits in d be m. Since we have an infinite increasing sequence, we can always find a number p with k digits where k=m+2 if m is odd and k=m+1 if m is even. Since m < k, we will eventually get a k+1 digit number in the sequence if we keep adding d to p. But k+1 is an even number, which contradicts the fact that len(R) only contains numbers with odd number of digits. Thus, len(R) cannot contain an infinite AP, which implies that R cannot be an infinite regular set.

Thus we have shown that the language L cannot contain an infinite regular set. By a similar argument, we can show that the language  $L^c$  also does not contain an infinite regular set.

### Question 2

For a language L over an alphabet A, we define:

$$first-halves(L) = \{ x \in A^* \mid \exists \ y : |x| = |y| \text{ and } xy \in L \}$$

Claim: If L is regular, then so is first-halves(L)

**Proof:** To construct a DFA for first-halves(L), we need to find states which can be reached from the start state by some strins of some length n, and then there exists a string of length n which can take us from this state to a final state. We construct the DFA for first-halves(L) as follows:

Let the DFA for L be  $A = (Q, s, \delta, F)$ . We define the DFA for first-halves(L) as  $A' = (Q', s', \delta', F')$  where:

- $Q' = \{(q,T) \mid q \in Q, T \subset Q, \text{ where there exists } w \in A^* \text{ such that } \hat{\delta}(s,w) = q \text{ and for all } t \in T, \text{ there exists } x \in A^* \text{ such that } |x| = |w| \text{ and } \hat{\delta}(t,x) \in F\}$
- s' = (s, F)
- $\delta'((q,T),a) = (r,U)$  where  $r = \delta(q,a), U = \{u \in Q \mid \text{there exists } a \in A, t \in T \text{ such that } \delta(u,a) = t\}$
- $F' = \{(q, S) \mid q \in S\}$

We can see that the transition function  $\delta'$  is well defined since  $\delta$  was well defined originally and the set U can be constructed uniquely for every T. Also since Q is finite, the set Q' is finite. Thus this is a DFA.

Now we need to show that A' accepts first-halves(L). First assume  $x \in L(A')$ . This means that there exists a state q such that  $\hat{\delta}(s,x) = q$  and there exists a string  $y \in A^*$  such that |x| = |y| and for every state  $s \in S$ ,  $\hat{\delta}(s,y) \in F$ . But the state q also belongs to S. Hence  $\hat{\delta}(q,y) \in F$ , which implies that  $xy \in L(A)$ . Thus  $x \in first$ -halves(L).

Now assume  $x \in first$ -halves(L). Then there exists  $y \in A^*$  such that |y| = |x| such that  $xy \in L$ . Thus there exists a state  $q \in Q$  such that  $\hat{\delta}(s,x) = q$  and  $\hat{\delta}(q,y) \in F$ . Consider  $(q, \{q\})$ . For this, we have x and y which satisfies the criteria for it to be in F'. Thus  $x \in L(A')$ .

Thus L(A') = first-halves(L).

### Question 3

The McNaughton-Yamada construction is defined as:

$$L_{pq}^{X \cup \{r\}} = L_{pq}^X + L_{pr}^X \cdot (L_{rr}^X)^* \cdot L_{rq}^X$$

where

$$L_{pq}^{\{\}} = \begin{cases} \{a \in A \mid \delta(p,a) = q\} & \text{if } p \neq q \\ \{a \in A \mid \delta(p,a) = q\} \cup \{\epsilon\} & \text{if } p = q \end{cases}$$

Thus we can write:

$$L_{ss}^{\{s,p,q\}} = L_{ss}^{\{s,p\}} + L_{sq}^{\{s,p\}} \cdot (L_{qq}^{\{s,p\}})^* \cdot L_{qs}^{\{s,p\}}$$

Expanding the terms on the RHS, we get:

$$\begin{split} L_{ss}^{\{s,p\}} &= L_{ss}^{\{s\}} + L_{sp}^{\{s\}} \cdot (L_{pp}^{\{s\}})^* \cdot L_{ps}^{\{s\}} \\ L_{sq}^{\{s,p\}} &= L_{sq}^{\{s\}} + L_{sp}^{\{s\}} \cdot (L_{pp}^{\{s\}})^* \cdot L_{pq}^{\{s\}} \\ L_{qq}^{\{s,p\}} &= L_{qq}^{\{s\}} + L_{qp}^{\{s\}} \cdot (L_{pp}^{\{s\}})^* \cdot L_{pq}^{\{s\}} \\ L_{qs}^{\{s,p\}} &= L_{qs}^{\{s\}} + L_{qp}^{\{s\}} \cdot (L_{pp}^{\{s\}})^* \cdot L_{ps}^{\{s\}} \end{split}$$

Once again, we expand the terms on the RHS to get:

$$L_{ss}^{\{s\}} = L_{ss}^{\{\}} + L_{ss}^{\{\}} \cdot (L_{ss}^{\{\}})^* \cdot L_{ss}^{\{\}}$$

$$= (a + \epsilon) + (a + \epsilon) \cdot (a + \epsilon)^* \cdot (a + \epsilon)$$

$$= (a + \epsilon) + (a + \epsilon) \cdot (a^*) \cdot (a + \epsilon)$$

$$= (a + \epsilon) + (a + \epsilon) \cdot (a^* + a^*)$$

$$= (a + \epsilon) + (a + \epsilon) \cdot (a^*)$$

$$= (a + \epsilon) + (a^*)$$

$$= a^*$$

$$\begin{split} L_{sp}^{\{s\}} &= L_{sp}^{\{\}} + L_{ss}^{\{\}} \cdot (L_{ss}^{\{\}})^* \cdot L_{sp}^{\{\}} \\ &= b + (a + \epsilon) \cdot (a + \epsilon)^* \cdot b \\ &= b + (a^*) \cdot b \\ &= a^* \cdot b \end{split}$$

$$\begin{split} L_{pp}^{\{s\}} &= L_{pp}^{\{\}} + L_{ps}^{\{\}} \cdot (L_{ss}^{\{\}})^* \cdot L_{sp}^{\{\}} \\ &= (b + \epsilon) + \phi \cdot (a + \epsilon)^* \cdot b \\ &= b + \epsilon \end{split}$$

$$\begin{split} L_{ps}^{\{s\}} &= L_{ps}^{\{\}} + L_{ps}^{\{\}} \cdot (L_{ss}^{\{\}})^* \cdot L_{ss}^{\{\}} \\ &= \phi + \phi \cdot (a + \epsilon)^* \cdot (a + \epsilon) \\ &= \phi \end{split}$$

$$L_{sq}^{\{s\}} = L_{sq}^{\{\}} + L_{ss}^{\{\}} \cdot (L_{ss}^{\{\}})^* \cdot L_{sq}^{\{\}}$$
  
=  $\phi + (a + \epsilon) \cdot (a + \epsilon)^* \cdot \phi$   
=  $\phi$ 

$$\begin{split} L_{pq}^{\{s\}} &= L_{pq}^{\{\}} + L_{ps}^{\{\}} \cdot (L_{ss}^{\{\}})^* \cdot L_{sq}^{\{\}} \\ &= a + \phi \cdot (a + \epsilon)^* \cdot \phi \\ &= a \end{split}$$

$$L_{qq}^{\{s\}} = L_{qq}^{\{\}} + L_{qs}^{\{\}} \cdot (L_{ss}^{\{\}})^* \cdot L_{sq}^{\{\}}$$
$$= \epsilon + a \cdot (a + \epsilon)^* \cdot \phi$$
$$= \epsilon$$

$$L_{qp}^{\{s\}} = L_{qp}^{\{\}} + L_{qs}^{\{\}} \cdot (L_{ss}^{\{\}})^* \cdot L_{sp}^{\{\}}$$

$$= b + a \cdot (a + \epsilon)^* \cdot b$$

$$= b + a^+ \cdot b$$

$$= a^* \cdot b$$

$$L_{qs}^{\{s\}} = L_{qs}^{\{\}} + L_{qs}^{\{\}} \cdot (L_{ss}^{\{\}})^* \cdot L_{ss}^{\{\}}$$

$$= a + a \cdot (a + \epsilon)^* \cdot (a + \epsilon)$$

$$= a + a \cdot (a^*)$$

$$= a + a^+$$

$$= a^+$$

Substituting these values back into the original equation, we get:

$$\begin{split} L_{ss}^{\{s,p\}} &= L_{ss}^{\{s\}} + L_{sp}^{\{s\}} \cdot (L_{pp}^{\{s\}})^* \cdot L_{ps}^{\{s\}} \\ &= a^* + a^* \cdot b \cdot (b + \epsilon)^* \cdot \phi \\ &= a^* \end{split}$$

$$L_{sq}^{\{s,p\}} = L_{sq}^{\{s\}} + L_{sp}^{\{s\}} \cdot (L_{pp}^{\{s\}})^* \cdot L_{pq}^{\{s\}}$$

$$= \phi + a^* \cdot b \cdot (b + \epsilon)^* \cdot a$$

$$= a^* \cdot b \cdot (b + \epsilon)^* \cdot a$$

$$= a^* \cdot b^+ \cdot a$$

$$\begin{split} L_{qq}^{\{s,p\}} &= L_{qq}^{\{s\}} + L_{qp}^{\{s\}} \cdot (L_{pp}^{\{s\}})^* \cdot L_{pq}^{\{s\}} \\ &= \epsilon + a^* \cdot b \cdot (b + \epsilon)^* \cdot a \\ &= \epsilon + a^* \cdot b^+ \cdot a \end{split}$$

$$L_{qs}^{\{s,p\}} = L_{qs}^{\{s\}} + L_{qp}^{\{s\}} \cdot (L_{pp}^{\{s\}})^* \cdot L_{ps}^{\{s\}}$$

$$= a^+ + a^* \cdot b \cdot (b + \epsilon)^* \cdot \phi$$

$$= a^+$$

Finally, using these values we get:

$$\begin{split} L_{ss}^{\{s,p,q\}} &= L_{ss}^{\{s,p\}} + L_{sq}^{\{s,p\}} \cdot (L_{qq}^{\{s,p\}})^* \cdot L_{qs}^{\{s,p\}} \\ &= a^* + a^* \cdot b^+ \cdot a \cdot (\epsilon + a^* \cdot b^+ \cdot a)^* \cdot a^+ \\ &= a^* + a^* \cdot b^+ \cdot a \cdot (a^* \cdot b^+ \cdot a)^* \cdot a^+ \\ &= a^* + (a^* \cdot b^+ \cdot a)^+ \cdot a^+ \\ &= \epsilon + a^+ + (a^* \cdot b^+ \cdot a)^+ \cdot a^+ \\ &= \epsilon + (\epsilon + (a^* \cdot b^+ \cdot a)^+) \cdot a^+ \\ &= \epsilon + (a^* \cdot b^+ \cdot a)^* \cdot a^+ \end{split}$$

Thus, the regular expression corresponding to the language accepted by the DFA is:

$$\epsilon + (a^* \cdot b^+ \cdot a)^* \cdot a^+$$

## Question 4

To avoid using intermediate states, we start by considering all states and then inductively remove states one by one till we have none left. Thus we proceed with backward induction.

Let Q be the set of all states of the NFA. We define the base case as:

$$LA(p,Q,q) = \begin{cases} \{a \in A \mid q \in \Delta(p,a)\} & \text{if } p \neq q \\ \{a \in A \mid q \in \Delta(p,a)\} \cup \{\epsilon\} & \text{if } p = q \end{cases}$$

Now, for  $r \in X$ , we define the inductive step as:

$$LA(p, X \setminus \{r\}, q) = LA(p, X, q) + LA(p, X, r) \cdot (LA(r, X, r))^* \cdot LA(r, X, q)$$

Finally, to get the regular expression corresponding to the language accepted by the NFA, we set  $p = s, q = f, X = \phi$  and take its union over all start and final states, ie:

$$L = \bigcup_{s \in S, f \in F} LA(s, \phi, f)$$

where S is the set of all start states and F is the set of all final states.

# Question 5

The canonical MN relation  $\equiv_L$  for a language L is defined as:

$$x \equiv_L y \iff (\forall z \in A^*, xz \in L \iff yz \in L)$$

By the Myhill-Nerode theorem, a language L is regular if and only if it has a finite number of equivalence classes under  $\equiv_L$ .

#### MN relation for L

The language L is defined as:

$$L = \{ w \in \{a, b\}^* \mid |\#_a(w) - \#_b(w)| \le 2 \}$$

Claim: The equivalence classes under  $\equiv_L$  are:

$$C_j = \{w \in \{a, b\}^* \mid \#_a(w) - \#_b(w) = j\} \ \forall j \in \mathbb{Z}$$

**Proof:** Clearly, the above classes are disjoint since no string can have two different values for  $\#_a(w) - \#_b(w)$ . Also, the union of all these classes is the set of all strings over  $\{a,b\}^*$ , since for any string w,  $\#_a(w) - \#_b(w)$  can only take values in  $\mathbb{Z}$ . Thus, the classes  $C_j$  form a partition of the set of all strings over  $\{a,b\}^*$ .

The classes accepted by the DFA are  $C_{-2}, C_{-1}, C_0, C_1, C_2$  since the modulus of the difference should be less than or equal to 2.

Now we need to show that the classes are indeed equivalence classes under  $\equiv_L$ . Let  $x, y \in C_j$ . Then  $\#_a(x) - \#_b(x) = j$  and  $\#_a(y) - \#_b(y) = j$ . Let  $z \in \{a, b\}^*$ . Then:

$$\#_a(xz) - \#_b(xz) = \#_a(x) + \#_a(z) - \#_b(x) - \#_b(z)$$

$$= \#_a(x) - \#_b(x) + \#_a(z) - \#_b(z)$$

$$= j + \#_a(z) - \#_b(z)$$

$$\#_a(yz) - \#_b(yz) = \#_a(y) + \#_a(z) - \#_b(y) - \#_b(z)$$

$$= \#_a(y) - \#_b(y) + \#_a(z) - \#_b(z)$$

$$= j + \#_a(z) - \#_b(z)$$

Thus 
$$\#_a(xz) - \#_b(xz) = \#_a(yz) - \#_b(yz) \implies (xz \in L \iff yz \in L)$$
  
Thus,  $x \equiv_L y \implies (\forall z \in A^*, xz \in L \iff yz \in L)$ 

Now assume  $x \not\equiv_L y$ . Then  $\#_a(x) - \#_b(x) = z_1$  and  $\#_a(y) - \#_b(y) = z_2$  for some  $z_1 \neq z_2$ . WLOG, let  $z_1 > z_2$ . We consider three cases:

- 1. One of  $z_1$  or  $z_2$  is in the range [-2,2]: In this case,  $z=\epsilon$  differentiates between x and y.
- 2.  $z_1, z_2 > 2$ : In this case,  $z = b^{z_2-2}$  differentiates between x and y.
- 3.  $z_1, z_2 < -2$ : In this case,  $z = a^{-2-z_1}$  differentiates between x and y.

Thus for every  $z_1$  and  $z_2$  such that  $z_1 \neq z_2$ , there exists a z such that  $xz \in L \Leftrightarrow yz \in L$ . Thus,  $x \not\equiv_L y \implies (\exists z \in A^*, \ xz \in L \Leftrightarrow yz \in L)$ 

Taking the contrapositive of the above statement, we get:

$$(\forall z \in A^*, xz \in L \iff yz \in L) \implies x \equiv_L y$$

Hence we get that:

$$x \equiv_L y \iff (\forall z \in A^*, xz \in L \iff yz \in L)$$

Thus, the classes  $C_j$  are indeed equivalence classes under  $\equiv_L$ . Since the number of classes is infinite, the language L is not regular.

#### MN relation for M

The language M is defined as:

$$M = \{w \in \{a, b\}^* \mid \text{ for all prefixes u of w}, |\#_a(u) - \#_b(u)| \le 2\}$$

Claim: The equivalence classes under  $\equiv_M$  are:

$$C_{-2} = \{w \in \{a,b\}^* \mid \text{ for all prefixes u of } w, |\#_a(u) - \#_b(u)| \le 2 \& \#_a(w) - \#_b(w) = -2\}$$

$$C_{-1} = \{w \in \{a,b\}^* \mid \text{ for all prefixes u of } w, |\#_a(u) - \#_b(u)| \le 2 \& \#_a(w) - \#_b(w) = -1\}$$

$$C_0 = \{w \in \{a,b\}^* \mid \text{ for all prefixes u of } w, |\#_a(u) - \#_b(u)| \le 2 \& \#_a(w) - \#_b(w) = 0\}$$

$$C_1 = \{w \in \{a,b\}^* \mid \text{ for all prefixes u of } w, |\#_a(u) - \#_b(u)| \le 2 \& \#_a(w) - \#_b(w) = 1\}$$

$$C_2 = \{w \in \{a,b\}^* \mid \text{ for all prefixes u of } w, |\#_a(u) - \#_b(u)| \le 2 \& \#_a(w) - \#_b(w) = 2\}$$

$$C_x = \{w \in \{a,b\}^* \mid \text{ for some prefix u of } w, |\#_a(u) - \#_b(u)| > 2\}$$

**Proof:** The classes  $C_{-2}$ ,  $C_{-1}$ ,  $C_0$ ,  $C_1$ ,  $C_2$  are clearly disjoint since no string can have two different values for  $\#_a(w) - \#_b(w)$ . The class  $C_x$  is also disjoint from the other classes since it requires some prefix u to have  $|\#_a(u) - \#_b(u)| > 2$ . Also, the union of all these classes is the set of all strings over  $\{a, b\}^*$ , since for any string w, the string itself can be a prefix of itself. Thus these classes form a partition of the set of all strings over  $\{a, b\}^*$ .

The classes accepted by the DFA are  $C_{-2}, C_{-1}, C_0, C_1, C_2$  since the modulus of the difference for each prefix should be less than or equal to 2.

Now we need to show that the classes are indeed equivalence classes under  $\equiv_M$ . We consider two cases here:

- 1.  $x, y \in C_j$  for some  $j \in \{-2, -1, 0, 1, 2\}$ : Then  $\#_a(x) \#_b(x) = j$ ,  $\#_a(y) \#_b(y) = j$  and for every prefix u of x and y,  $|\#_a(u) \#_b(u)| \le 2$ . Let  $z \in \{a, b\}^*$ . If for some prefix p of z,  $|\#_a(xp) \#_b(xp)| > 2$ , then  $|\#_a(yp) \#_b(yp)| > 2$  and vice versa. Hence both xz and yz will not be in M.
  - However, if for every prefix p of z,  $|\#_a(xp) \#_b(xp)| \le 2$ , then  $|\#_a(yp) \#_b(yp)| \le 2$  and vice versa. Hence both xz and yz will be in M.
- 2.  $x, y \in C_x$ : Since x and y have some prefix u such that  $|\#_a(u) \#_b(u)| > 2$ , adding any z will not change that property. Hence both xz and yz will not be in M.

Thus, 
$$x \equiv_M y \implies (\forall z \in A^*, xz \in M \iff yz \in M)$$

Now assume  $x \not\equiv_M y$ . We consider two cases for the classes of x and y:

- 1.  $x \in C_{j_1}$  and  $y \in C_{j_2}$  for some  $j_1, j_2 \in \{-2, -1, 0, 1, 2\}$ : Then  $\#_a(x) \#_b(x) = j_1$ ,  $\#_a(y) \#_b(y) = j_2$ . WLOG assume  $j_1 > j_2$ . Then  $z = a^{3-j_1}$  differentiates between x and y since xz will be in  $C_x$  (and hence not in M) while yz will be in  $C_{j_2+3-j_1}$  (and hence in M).
- 2. One of x or y is in  $C_x$ : Then  $z = \epsilon$  differentiates between x and y.

Thus for x and y in different classes, there always exists a z such that  $xz \in M \Leftrightarrow yz \in M$ . Thus,  $x \not\equiv_M y \implies (\exists z \in A^*, xz \in M \Leftrightarrow yz \in M)$ 

Taking the contrapositive of the above statement, we get:

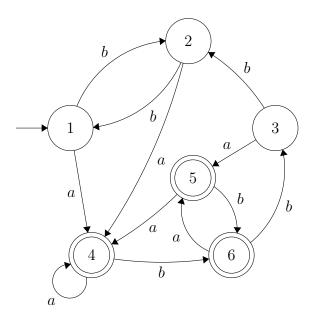
$$(\forall z \in A^*, xz \in M \iff yz \in M) \implies x \equiv_M y$$

Hence we get that:

$$x \equiv_M y \iff (\forall z \in A^*, xz \in M \iff yz \in M)$$

Thus, the classes  $C_{-2}, C_{-1}, C_0, C_1, C_2, C_x$  are indeed equivalence classes under  $\equiv_M$ . Since the number of classes is 6, which is finite, the language M is regular.

#### Question 6



Let the DFA be  $A = (Q, s, \delta, F)$  where:

$$Q = \{1, 2, 3, 4, 5, 6\}$$
  
$$s = 1$$
  
$$F = \{4, 5, 6\}$$

The  $\delta$  function is given by:

$$\delta(1, a) = 4$$

$$\delta(1, b) = 2$$

$$\delta(2, a) = 4$$

$$\delta(2, b) = 1$$

$$\delta(3, a) = 5$$

$$\delta(3, b) = 2$$

$$\delta(4, a) = 4$$

$$\delta(4, b) = 6$$

$$\delta(5, a) = 4$$

$$\delta(5, b) = 6$$

$$\delta(6, a) = 5$$

$$\delta(6, b) = 3$$

To minimise the DFA, we create a 6 x 6 table with the rows and columns labelled 1 to 6. We then iteratively mark elements in this table using the following procedure:

- 1. Ignore all the elements on the diagonals of the table by setting them to  $\cdot$
- 2. Mark all elements of the form (p,q) where one of them is a final state and the other is not.
- 3. Now, iteratively mark the states (p,q) such that (p,q) is unmarked and  $\exists a \in A$  such that  $(\delta(p,a), \delta(q,a))$  is marked.

The states that are not marked are equivalent and can be merged. Also, since the table is symmetric, we can work with only the lower triangular region.

The table after the first step is:

	1	2	3	4	5	6
1	•					
2						
3						
4						
$ \begin{vmatrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \end{vmatrix} $						
6						•

The table after the second step is:

	1	2	3	4	5	6
1						
$\begin{vmatrix} 2 \\ 3 \end{vmatrix}$						
3						
4	X	X	X			
5	X	X	X			
6	X	X	X			

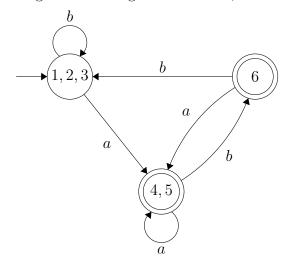
For the first iteration of the third step, we can mark the following:

- Mark (6,4) since  $(\delta(6,b),\delta(4,b)) = (3,6) \equiv (6,3)$  is marked

The table now looks like:

	1	2	3	4	5	6
1						
2						
3						
4	X	X	X			
5	X	X	X			
6	X	X	X	X	X	

For the second iteration of the third step, we can not mark any more elements. Thus the algorithm ends here. From the table, we can see that (2,1), (3,1), (3,2) and (5,4) are unmarked. Thus, the states 1, 2 and 3 can be merged into a single state and the states 4 and 5 can be merged into a single state. Thus, the minimised DFA is:



Now the DFA is  $A' = (Q', s', \delta', F')$  where:

$$Q' = \{\{1, 2, 3\}, \{4, 5\}, 6\}$$
$$s' = \{1, 2, 3\}$$
$$F' = \{\{4, 5\}, 6\}$$

The  $\delta'$  function is given by:

$$\delta'(\{1,2,3\},a) = \{4,5\}$$

$$\delta'(\{1,2,3\},b) = \{1,2,3\}$$

$$\delta'(\{4,5\},a) = \{4,5\}$$

$$\delta'(\{4,5\},b) = 6$$

$$\delta'(6,a) = \{4,5\}$$

$$\delta'(6,b) = \{1,2,3\}$$