Implementation on USRP/GNU Radio CR Platform

1.0 Cluster Formation

Algorithm 1: Cluster formation at SU node i

Initialize

n_i; $H_{C,min} = 2$; $H_{H,min} = 1$; $j \in J = \{1,2,\dots,N(i)\}$;

Repeat

#Part I: Clusterhead election and member node joining#

if $(n_i > n_j | \forall j \in J)$ && $(i \leftarrow NN_i)$, **then**

$nodeState_i \leftarrow CH_i$

else if $(n_{c.i} \geq H_{C,min} \mid \forall i \in I)$ && $(i \leftarrow NN_i)$, **then**

$nodeState_i \leftarrow MN_{j,i}$

else *create a cluster* C_i

$nodeState_i \leftarrow CH_i$

end if

#Part II: Gateway node selection at $MN_{j,i}$ *of cluster* C_i*#*

if $(n_{h.i} = H_{H,min} \mid \forall i \in C_i)$ && $(i \leftarrow MN_{j,i})$ **then**

$nodeState_i \leftarrow GN_{j,i,k}$

else

$nodeState_i \leftarrow MN_{j,i}$

end if

Table 1: List of notations used in algorithms

Notation	Description
$N(i) = j \in J$	Neighbouring SU nodes of SU i
NN_i	Non-clustered SU node i
CH_i	Clusterhead of cluster i
$MN_{j,i}$	Member SU node i of cluster j
$GN_{j,i,k}$	Gateway SU node i of cluster j one-hop away from neighbouring cluster k
n_i	Number of available channels at SU node i
$n_{c.i}$	Set of available common channels among SU node i and its neighbouring SUs
$n_{h.i}$	Number of hops among SU node i and its neighbouring SUs
$H_{C,min}$	Minimum threshold for cluster formation among SU nodes
$H_{H,min}$	Minimum threshold for gateway SU node selection
p_{ij}	Reward estimate
$\gamma_{i,j}$	Dynamic discount factor
WF_j	OFF-state probability of channel j
PF_j	Packet delivery ratio at SU node j

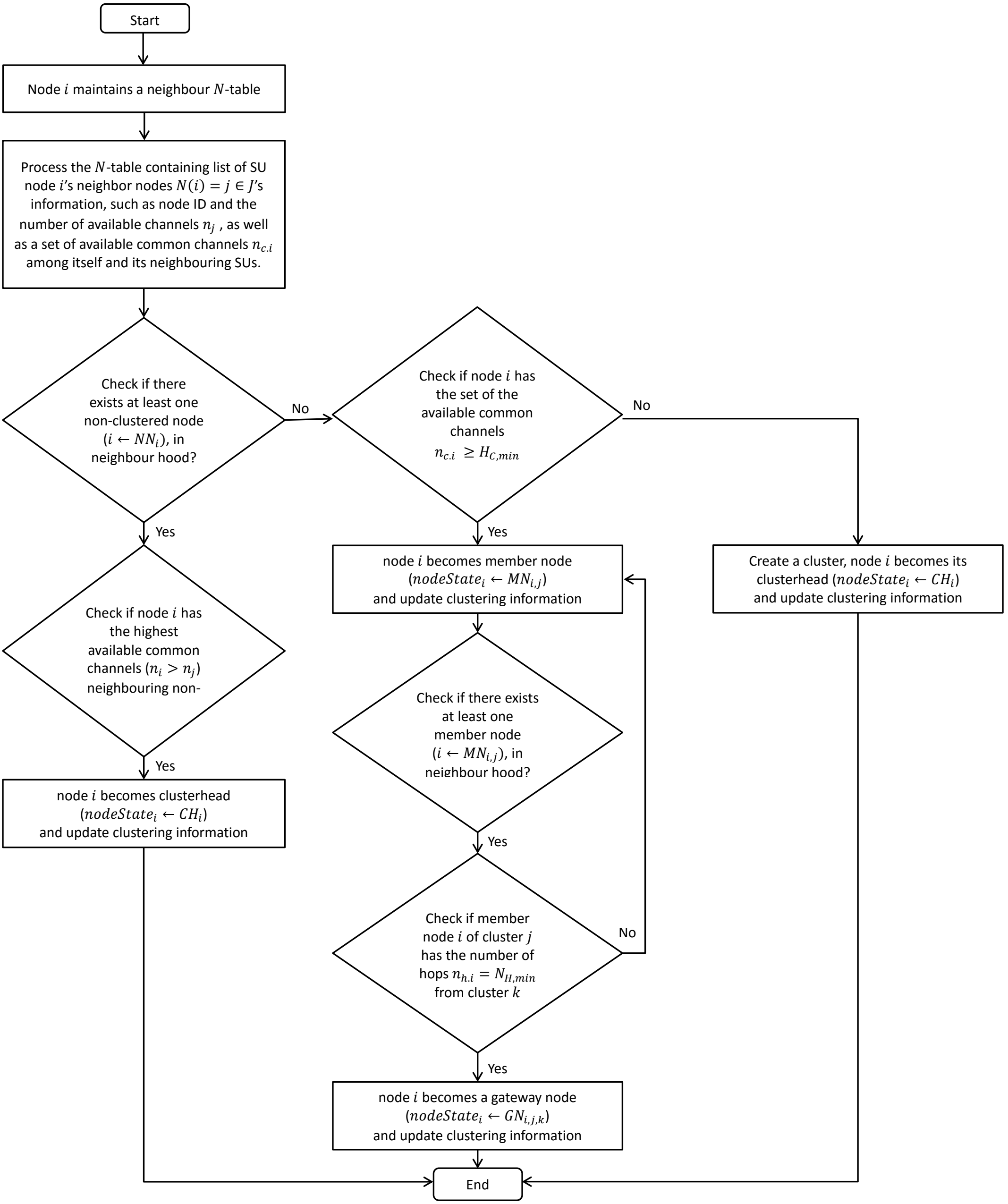

Figure 1: Flow chart for Algorithm 2.

2.0 RL-based Route Selection

Algorithm 2: Traditional Q-routing at SU node i

Initialize

$\gamma = [0,1]$; $\alpha = [0,1]$; $Q_t^i(s_t^i, a_t^i) = 0$; $a_t^i \in A$, $s_t^i \in S$

Repeat

1: Observe state s_t^i

2: Select exploitation or exploration action

a. If exploitation, choose best-known action $a_t^i = \max_{k \in A} Q_t^i(s_t^i, k)$

b. If exploration, choose random action $a_t^i \in A$

3: Perform action a_t^i

$t \rightarrow t+1$

4: Observe reward $r_{t+1}^i(s_t^i, a_t^i)$ and receive $\max_{a_t^j \in A} Q_t^j(s_t^j, a_t^j)$ from the next-hop neighbor node a_t^i

5:

6:Update Q-value $Q_t^i(s_t^i, a_t^i)$

$$Q_{t+1}^i(s_t^i, a_t^i) \leftarrow (1-\alpha) Q_t^i(s_t^i, a_t^i) + \alpha \left[r_{t+1}^i(s_t^i, a_t^i) + \gamma \max_{k \in N(i) = a_t^j} Q_t^j(s_t^j, a_t^j) \right] 1$$

Algorithm 3: Enhanced Q-routing with dynamic discount factor $\boldsymbol{\gamma_{i,j}}$ model at SU node i

Initialize

$\omega = [0,1]$; $\alpha = [0,1]$; $Q_t^i(s_t^i, a_t^i) = 0$; $a_t^i \in A$, $s_t^i \in S$

Repeat

$t \rightarrow t+1$

1: Observe reward $r_{t+1}^i(s_t^i, a_t^i)$ and receive $\max_{a_t^j \in A} Q_t^j(s_t^j, a_t^j)$ from the next-hop neighbor node a_t^i

2: For $\forall j \in J$ calculate dynamic discount factor $\gamma_{i,j}$ using Equation 2a.

3: Update Q-value $Q_t^i(s_t^i, a_t^i)$

$$Q_{t+1}^i(s_t^i, j) \leftarrow (1-\alpha) Q_t^i(s_t^i, j) + \alpha \left[r_t^i(s_{t+1}^i, j) + \gamma_{i,j} \max_{k \in N(i) = a_t^j} Q_t^j(s_t^j, k) \right] 3$$

$$\gamma_{i,j} = \omega \sqrt{WF_j . PF_j}, \qquad \forall j \in J \qquad 2a$$

$$WF_j = \cup_{ON}^k + \left(\cup_{OFF}^k \times e^{-(\lambda_{ON} + \lambda_{OFF})} \right) \qquad 2b$$

$$PF_j = \sum_j number\ of\ packets\ (receive/sent) \quad 2c$$

Table 2: Q-routing model for route selection embedded in SU node i

State	$s_t^i \in S = \{1,2,\dots,N-1\}$ represents a SU destination node , where N is the total number of SUs in the network.
Action	$a_t^i \in A = \{1,2,\dots,J\}$ represents the selection of a next-hop SU neighbor node j, where J the number of SU i's neighboring SU nodes.
Reward	$r_t^i(s_t^i, a_t^i) = \begin{cases} 1, & success \\ 0, & otherwise \end{cases}$ Node i receives a reward value of 1 if the packet it forwards has reached the SU destination node s_t^i, such as a SU BS, via its next-hop SU neighbour node $a_t^i = j$; otherwise, it receives a reward value of 0.