A Realistic Weighted Clustering Algorithm for Data Gathering in Single Hop Cell Phone Based Sensor Network

M. B. Shah, P. S. Tamhankar, S. N. Merchant, U. B. Desai
Electrical Engineering Department
Indian Institute of Technology Bombay, India
Email: mehulshah@ee.iitb.ac.in,tamhankarp@gmail.com,merchant@ee.iitb.ac.in,ubdesai@iith.ac.in

Abstract—We present Weight based Realistic Clustering algorithm (WRCA) for energy efficient data collection from human based wireless network such as cell phone based sensor network. The WRCA creates better load balanced and stable clusters for data gathering in a realistic human mobility scenario as compared to WCA algorithm of mobile adhoc networks. Traditional way of stable clustering used to select stable Cluster Heads (CHs) based on instantaneous mobility criteria. While our algorithm incorporate only a pause time weighting parameter to leverage pause time distribution of human mobility for selection of more stable CHs. Load balancing aspect of WRCA tackles realistic scenario of inhomogeneous node distribution. Improvement in load balancing with WRCA assure better aggregation of sensor data and MAC layer performance. This is accomplished by incorporating a density center weighting parameter. This information is used to select density centered CHs. WRCA requires fewer messages to find density centered nodes as compared to TASC. Simulation results demonstrate the overall superiority in performance of the WRCA when mobility according to a realistic mobility model called Self similar Least Action Walk (SLAW) is considered. Simulation result shows that our proposed algorithm consumes 20% less energy and 50% more network lifetime as compared to WCA algorithm.

Keywords: Weighted Clustering, Density Based Clustering, Human Mobility, Mobile Sensing

I. INTRODUCTION

The advances in cell phone technology and ubiquitous use of cell phones has motivated the idea of participatory sensing with cell phones[1]. Embedded sensors of mobile phones has provided an ideal platform to researchers for developing novel personal, group and community scale sensing applications[2]. Mobile sensing with cell phones has presented new challenges not faced by Wireless Sensor Network (WSN) communities. The fundamental challenge from network designer perspective is to gather sensor data efficiently in a realistic mobility scenario of cell phone users. Realistic scenario consists inhomogeneous node distribution, power law distribution of mobile node’s flight length and pause time[7]. This realistic scenario of human mobility is taken care in the SLAW mobility model proposed in [7].

In this paper we consider the fact of human mobility to design data collection protocol for human mobility based mobile wireless sensor network. In our work we consider a community scale environmental monitoring application where the end user makes a query, specifying area of interest and the frequency of data collection with amount of tolerable delay.

In our system model the cell phones transmit data to the BS in a single hop fashion. The Single hop Cellular Sensor Networks (SCSN) (Fig. 1) are advocated in situations where sensor networks are built upon the Cellular infrastructure[3][12][15]. SCSN scenario consists of mobile hosts having the capability to communicate directly with another mobile host in their vicinity and forwarding sensed data packets.

Fig. 1. Single hop cell-phone based sensor network

Traditional WSN protocols [8][9] group nodes in to clusters and each node has its own Cluster head(CH). The CH gathers the data and transmit them to the BS directly. This reduces the total number of packets to be transmitted and thus saves energy. As CH gathers the data from its members, they spend energy at higher rate. So the CH role need to be periodically rotated to achieve energy efficiency or to prolong the network life-time. However the issue with SCSN is the node mobility. To tackle node mobility and to achieve scalability, stable clustering methods[5][6] are used in mobile wireless network. This technique tries to predict stable CH so as to stabilize the network backbone under mobile scenario. The WCA[4] algorithm uses instantaneous mobility parameter to predict stable CHs. The authors of WCA has evaluated WCA on random type mobility models where each node moves with random angle, velocity and pause times. We propose modification in the conventional WCA[4] algorithm to select more stable CH for node movements as per SLAW mobility model[7]. WCA [4] is a weight based clustering algorithm which does take into account the topological factors such as degree differences, sum
of distances with neighbors, mobility measure and cumulative time for which a node is a cluster head. TASC [13] calculates weight based on distance contribution and then density reachability criteria is used for grouping similar density into clusters. In [15] we propose a HMRECA clustering algorithm for homogeneously distributed nodes of TLW mobility model.

The traditional criterions for CH selection as used in WCA would obviously be useful, but are inefficient in the realm of human mobility. Our proposed WRCA algorithm achieves improvement in achieving stability and load balancing in realistic human mobility scenario. The mobility criterion of WCA when applied for human mobility selects nodes with less mobility as CH, but is not optimal for predicting the most stable CHs. Instead of mobility if pause time characteristics of human mobility is exploited then more stable CH nodes can be predicted. Truncated power law distribution of pause time indicates bias on the past history of flight. The parameter \( P \) which is based on calculation of distances in 2-hop neighborhood.

In section II we discuss about the system model with brief introduction of realistic human mobility model called Self Similar Least Action Walk (SLAW) followed by a discussion of the WRCA in section III. The performance metrics and simulation set up is described in section IV and V followed by results and conclusion in VI and VII respectively.

II. SYSTEM MODEL

The Self-similar Least Action Walk (SLAW) model proposed in [7] captures the most important characteristics of human walk such as truncated power-law distributions of flights, pause-times and inter-contact times, fractal way-points, and heterogeneously defined areas of individual mobility in an environment like campus, state park, metropolitan area etc. SLAW first generates fractal waypoints. Then it utilizes the properties of fractal points to generate power law flights on top of them. The gaps among fractal points follow power law distribution. SLAW then uses Least Action Trip Planning points to generate points that resemble the real traces within 10\% error. The main assumptions made for the system are:

1) The monitoring application requests for samples at the BS specifying the area of interest and the frequency of data collection.
2) The nodes are mobile and move as per SLAW mobility model.
3) Each node is capable of transmitting beaconing messages so that it can know its neighbor in one hop and two hop.
4) Each node estimates its inter-node distances by measuring its received signal strength indicator (RSSI) values.

5) The BS can send the broadcast message simultaneously to all nodes via a separate control channel.

III. WEIGHT BASED REALISTIC CLUSTERING ALGORITHM

The WRCA clustering algorithm include cluster formation, maintenance and termination.

Cluster Formation:

This phase starts after receiving query from the end user. The Base Station broadcasts packets containing (id, \( R_{tx} \)) to each node through a dedicated control channel with in a specified area of interest. \( R_{tx} \) is cluster radius and related with the resolution of the sensor data to be monitored. Each node then starts periodically broadcasting \( Nbr_{onehop} \) message containing its own ID within transmission range \( R_{tx}/2 \). Each node will come to know its one hop neighbors along with its approximate distances within time \( T_1 \). After time \( T_1 \) each node will start transmitting \( Nbr_{twohop} \) message within its one hop neighborhood containing its one hop neighbor’s id. Within time \( T_2 \) each node will come to know its 2-hop neighbors along with their inter node distances.

(Fig.- 2) shows the important steps of the algorithm. After time \( T_2 \) each node can compute its density centered weight \( G_k \) based on its two hop neighbor information. To compute \( G_k \) each node first finds the all pair shortest paths i to j(APSP) in its 2-hop environment. Then every time a node k is used in the path, the weight of \( G_k \) is incremented as a function of the distance that node contributes to the path.

\[
G_k = \sum_{(i,j) \in APSP} g(i,j) \tag{1}
\]

where,

\[
g_{i,j} = \frac{l_{a,k} + l_{k,b}}{l_{i,j}} \tag{2}
\]

\( l_{a,k} \) is distance between node a and k and \( l_{k,b} \) is distance between node b and k. Each node also finds \( P_u \) where

\[
P_{(u)} = \begin{cases} 0 & \text{if } P_{scount(u)} > 0 \\ 1 & \text{if } P_{scount(u)} = 0 \end{cases} \tag{3}
\]

\( P_{scount(u)} \) in (3) is incremented every time if the node pauses for \( P_{stime(u)} \), where \( P_{stime(u)} \) is super pause time of node u. \( P_{stime(u)} \) is the threshold time of mobile node u above which it is likely to pause for a very long time. This is as per the power law distribution of pause time. \( P_u \) favors mobile node to become a CH if its pause time crosses certain threshold value. The determination of this threshold time is subjective and depends upon human mobility scenario. For e.g in a state fair scene smaller value of threshold pause time is expected compare to campus area scene.

The number of times the node has served as a CH is stored by the node as a parameter \( E_u \), which is indirectly related to the energy expenditure of the node. CH node consumes more energy as compared to non cluster head node. The BS periodically transmits information about the maximum value of the above three parameters attained in the network to all the nodes. All the nodes normalize their parameters with respect
to this value. After this all the nodes initialize their counter based on the count value $C_u$ given by:

$$C_u = w_1 E_u + w_2 G_u + w_3 P_u$$  \hspace{1cm} (4)$$

where $w_1$, $w_2$ and $w_3$ are arbitrarily chosen weighing factors for the clustering process satisfying $\sum_{i=1}^{3} w_i = 1$.

The counter with the lowest count value expires first and it declares itself as a Cluster Head by transmitting ‘CH$_{declare}$’ message within transmission range $R_{tx}$. This message consists of node id field, cluster id field and score field. In the case of CH, the first two fields are the same. Upon receiving ‘CH$_{declare}$’ packet the nodes in the neighborhood of CH reset their counters and join the cluster formed by the corresponding CH by transmitting back a ‘CH$_{accept}$’ message with transmission range $R_{tx}$. This also contains the node id and the cluster id field. If the neighboring node is already a member of some other cluster then it will ignore the CH$_{declare}$ message. This process will terminate after $T_{stop}$ (which is finite). The maximum value of $T_{stop}$ is proportional to the maximum value a counter can take (denoted by $C_{max}$).

After finite time $T_{stop}$ all the cluster member nodes transmit their sensed data to the CH as per 802.11 based RTS-CTS messaging scheme. CHs then aggregates the data packet and sends it to the BS as shown in figure 1. This communication is done with high transmitting power.

**Cluster Maintenance:** All the nodes of a particular cluster sends hello message periodically to its CH. Due to node mobility some of the nodes currently attached with a particular CH may move to the some other cluster. This is called reaffiliation. Reaffiliation is detected both at cluster head node and at member node. If CH does not get hello message periodically from its member node then it strikes that node off from its neighbor list. Similarly, if a node doesn’t hear a periodical or hello message from the CH, then it assumes that it has become a free node. If a free ordinary node hears a hello message from some other CH then it will join it.

**Cluster Termination:** Due to node mobility the CH might move from its own cluster and approach some other CH. Now these CHs no longer remain the dominant nodes. This will end the current round and again invoke cluster formation.

### IV. Performance Metrics

The performance of data aggregation schemes for monitoring applications can be evaluated based upon:

1. **Load Balance Factor (LBF):** It measures how well balanced the cluster heads are. It is described by the following equation.

$$LBF = \frac{nc}{\sum_{i=1}^{N} (x_i - \mu)^2}$$  \hspace{1cm} (5)$$

$nc$ is the number of cluster heads, $x_i$ is the cardinality of cluster $i$, and $\mu$ (N being the total number of nodes in the system) is the average number of neighbors of a CH. Clearly, a higher value of LBF signifies a better load distribution and it tends to infinity for a perfectly balanced system.

2. **Number of reaffiliations:** The reaffiliation count is incremented when a node gets dissociated from its cluster head and becomes a member of another cluster within the current dominant set.

3. **Number of DS updates:** The dominant set update takes place when two or more CHs comes with in the transmission range of one another due to mobility.

4. **Energy Consumption:** We are calculating the total energy consumption of all nodes for various cluster radius. Our energy model for the sensors is based on the first order radio model described in [8]. A sensor consumes $E_{elec} = 50nJ/bit$ to run the transmitter or receiver circuitry and $E_{amp} = 100pJ/bit/m^2$ for the transmitter amplifier. Thus, the energy consumed by a sensor $i$ in receiving a k-bit data packet is given by

$$E_{Rk} = (E_{elec} \cdot k)$$  \hspace{1cm} (6)$$

While the energy consumed in transmitting a data packet to sensor $j$ is given by

$$E_{Tx} = E_{elec} \cdot k + E_{amp} \cdot R_{tx}^2 \cdot k$$  \hspace{1cm} (7)$$

$R_{tx}$ is the transmission range. Each CH uses energy in receiving data signals from its members, aggregating the data and transmitting it to the BS represented using the multi path model. Thus, the energy spent by a CH node during a single round is

$$E_{CH} = (E_{elec} \cdot k \cdot CH_{degree} + E_{DA} \cdot k)$$

$$+ (E_{elec} \cdot k + E_{amp} \cdot d_{toBS}^2 \cdot k)$$  \hspace{1cm} (8)$$

where $k$ is the number of bits in each data message, $CH_{degree}$ is the number of ordinary nodes in the cluster, $d_{toBS}$ is the average distance from a CH to the base station, and we assume lossy data aggregation with the energy for aggregation to be $E_{DA}$. As for each non-CH node, it only needs to transmit its data to the CH once during a round. Since the distance to the CH is small, the energy dissipation follows the Friiss free space model. Thus, the energy used in each non-CH node is

$$E_{nonCH} = (E_{elec} \cdot k) + E_{amp} \cdot R_{tx}^2 \cdot k$$  \hspace{1cm} (9)$$
Table I

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sensor Field</td>
<td>500 × 500 rectangle area</td>
</tr>
<tr>
<td>Number of Nodes (N)</td>
<td>75</td>
</tr>
<tr>
<td>BS location</td>
<td>(250, 250)</td>
</tr>
<tr>
<td>Packet header size</td>
<td>25bytes</td>
</tr>
<tr>
<td>Initial Energy</td>
<td>1J/battery</td>
</tr>
<tr>
<td>Data packet size</td>
<td>500bytes</td>
</tr>
<tr>
<td>$E_{elec}$</td>
<td>50nJ/bit</td>
</tr>
<tr>
<td>$E_{fs}$</td>
<td>0.0013pJ/bit/m2</td>
</tr>
<tr>
<td>$E_{mp}$</td>
<td>0.00013pJ/bit/m4</td>
</tr>
<tr>
<td>$E_{fusion}$</td>
<td>5nJ/bit/signal</td>
</tr>
<tr>
<td>Sampling Time</td>
<td>1 minute</td>
</tr>
<tr>
<td>Total Simulation Time</td>
<td>120 minutes</td>
</tr>
<tr>
<td>Min. pause time for SLAW</td>
<td>30 seconds</td>
</tr>
<tr>
<td>Max. pause time for SLAW</td>
<td>3600 seconds</td>
</tr>
<tr>
<td>Hurst parameter for self similarity</td>
<td>0.75</td>
</tr>
<tr>
<td>Levy scale factor for pause</td>
<td>2</td>
</tr>
<tr>
<td>Levy exponent for pause time</td>
<td>1</td>
</tr>
<tr>
<td>Levy exponent for distance</td>
<td>3</td>
</tr>
<tr>
<td>Pause time threshold</td>
<td>20 minutes</td>
</tr>
</tbody>
</table>

Reaffiliation will consume $E_r$ units of energy, where

$$E_r = (E_{elec} \cdot k) + (E_{amp} \cdot d^2)$$ at CH \hspace{1cm} (10)

$$E_r = (E_{elec} \cdot k)$$ at ordinary node. \hspace{1cm} (11)

DS update will consume $E_d$ units of energy, where

$$E_d = \sum_{i=1}^{N} (E_{elec} \cdot k') + (E_{amp} \cdot R_{tx}^2) \cdot k'$$ \hspace{1cm} (12)

$$E_{tot} = \sum_{i=1}^{K} E(k) + (E_r \cdot R) + (E_d \cdot D)$$ \hspace{1cm} (13)

where $K$ is the total number of CH, $R$ is the total number of reaffiliations, $D$ is the total number of DS updates, $k'$ is size of ‘Hello’ packet.

5) Network Lifetime (Energy Efficiency): It is a primary metric for evaluating the performance of a sensor network. The common definitions include the time until the first or the last node in the network depletes its energy. For cell phone based WSN it is defined as the time until a $\alpha$ percent of sensors consume $\beta$ percent of initial energy, where $\alpha$ is taken as 50 percent and beta is taken as 20 percent in our case. These parameters are studied for varying cluster radii for SLAW mobility model.

V. Simulation Scenario

The simulation is implemented in MATLAB. The WCA algorithm is popular for Mobile Adhoc Network. For SCSN environment it is called as WCA with some minor modifications[12]. We have applied WRCA and WCDA[12] on the SLAW mobility model. The various parameter settings for SLAW are given in table-I. Monte Carlo simulations have been carried out for 10 random seeds and for varying cluster radii (50m to 200m). The simulations are carried out for weight values as shown in table-II.

To show the relative effect of newly introduced parameters on energy we have kept weight values of $w_3$ same in all four cases and have analyzed the effect. WRCA-g++ is the WRCA algorithm but instead of parameter $G_k$ the sum of distance parameter $S_k$ is used. WRCA+m is the WRCA algorithm with fourth additional mobility parameter $M_K$ and WCDA1 is the WCDA algorithm with but instead of sum of distance parameter $S_k$ the parameter $G_k$ is used.

VI. Simulation Results

Figure 3 shows comparisons of load balancing factor of WCDA and WRCA algorithm for cluster radius of 150m. Result shows better performance of WRCA algorithm. This indicates effectiveness of newly introduced parameter $G_k$. However improvement is less but that is due to many single node clusters. Single node clusters are created due to inhomogeneous node distribution.

Figure 4 shows reaffiliation of mobile nodes due to mobility for WRCA and WCDA algorithms. Number of reaffiliations are less for smaller cluster radius due to frequent DS updates at these radii.
Figure 5 shows DS updates for WRCA and WCDA algorithms. Results shows 8.57% reduction in DS updates for WRCA algorithm compared to WCDA algorithm for cluster radius of 100m and 15.4% for cluster radius of 150m. When compared with WRCA+M, it is reduced by 12.8% for radius of 100m and 19.9% for radius of 150m.

Figure 6 shows energy consumption for WRCA and WCDA algorithms. The energy consumption of WRCA and WRCA+M decreases by 20.25% for cluster radius of 100m and for cluster radius of 150m the energy consumption of WRCA and WRCA+M is reduced by 35.2% and 35.7% respectively. From figure 6 it is evident that the WCDA,WCDA1 and WRCA+M is reduced by 35.2% and 35.7% respectively.

Figure 7 shows the 55% improvement obtained with the use of WRCA algorithm over WCDA and 37% improvement over WRCA-g+d and WCDA1 algorithm in terms of 50% node death. This indicate the energy efficiency of the WRCA algorithm.

VII. Conclusion
In this paper, we propose a new weighted clustering scheme for mobile sensor networks called WRCA. Present approach combines the essence of WCA and TASC. WRCA is effective in two ways. Its generates load balanced and stable clusters to reduced energy consumption. Simulation results indicate that in realistic human mobility scenario, it outperforms the existing WCA protocol.

VIII. Acknowledgement
This work has been supported by the Department of Science and Technology (DST)-Government of India, and the India-UK Advanced Technology Center of Excellence in Next Generation Networks (IUATC) Project.

References