An Energy-Aware Dynamic Clustering Algorithm for Load Balancing in Wireless Sensor Networks

Mudasser Iqbal
GSIT, Faculty of IT, MONASH University, Churchill, 3842 Australia
Email: Mudasser.Iqbal@infotech.monash.edu.au

Iqbal Gondal, Laurence Dooley
GSIT, Faculty of IT, MONASH University, Churchill, 3842 Australia
Email: {Iqbal.Gondal, Laurence.Dooley}@infotech.monash.edu.au

Abstract **– Energy efficient dynamic clustering offers a flexible paradigm to reconfigure the network in order to maximise network's life-time in resource constrained ad hoc sensor networks. The load profiles of parent nodes (PNs) can be used to define its current state as well as to predict potential failures caused by energy loss due to high loads on particular PNs. This paper proposes a novel dynamic clustering algorithm for load balanced routing based upon route efficiency. The algorithm exploits the pattern and load of traffic and energy dissipation rate of each node on the route to calculate the node and route efficiencies. The proposed algorithm maintains PNs in a state whereby the network life requirement is met by the PNs comprising the backbone of the network. Results prove that the proposed methodology balances the load effectively to meet the network life requirement while concomitantly preserving the network capacity.**

Index Terms – Dynamic Clustering, Load Balanced Routing, Wireless Sensor Networks

I. Introduction

Wireless Sensor Networks (WSN) typically comprise many cheap sensor nodes that have low computation, communication, storage and energy resources. Due to lack of resources, an independent network backbone of *Parent nodes* (PN) which have relatively high resources has to be established. These nodes are responsible for such tasks as in-network data processing, communication delay minimization and routing of data. A WSN must also be self-monitoring and able to proactively reconfigure to mitigate certain malfunctions before they actually occur.

The life of the backbone is crucial to the network since the sensor nodes connect to the backbone while collecting information from a site, such as a disaster area. The load profile of each PN is a critical factor impacting upon this life span, since as a PN starts to become overloaded, the likelihood of congestion, energy loss and the PN becoming inactive becomes higher. If the particular PN that fails may be a key element of the backbone, the sensor network reliability becomes problematic. So it is vital to the steady state operation of the network that a self configuration strategy is in place so the network can continually monitor the state of individual PNs. The rationale behind such a monitoring system is to act proactively as soon as a PN enters into the so-called *disaster state* where it stops functioning in a short period of time. The objective of a proactive action solicited in such situations is to redistribute the load in such a manner that PNs are kept functioning in an *ideal state,* where a PN guarantees to meet the time-to-live requirements, without major intermission in the availability of communication services to sensor nodes.

As nodes consume their limited initial energy at a rate proportional to their activities in their respective cluster, it transpires that to maximize the lifetime of the network, traffic should be evenly routed for balanced energy consumption amongst the nodes. Moreover, the balance of energy drainage should be in proportion to the energy reserves of the PNs in the backbone, instead of routing to minimize the absolute consumed power. For this reason, the routing and PN duty cycle are inexorably linked in any time-scheduled ad hoc network as the sleep/wake cycles for PNs will be adjusted depending on the absolute scheduled load of the PNs.Traditionally, the lifetime of a device in the network at any instant is measured by the energy resources left at that time, whereas the decision to grant connections to the devices in the network is solely based upon the availability of connections at that instant. This approach has a major flaw, as it may not provide a true picture of the network when there are momentary abnormal outcomes in the load profile. Also, estimating the life of a PN based *only* upon current energy resources has the potential for erroneous estimates, as the difference

Based on "A Novel Load Balancing Technique for Proactive Energy Loss Mitigation in Ubiquitous Networks", by M. Iqbal, I. Gondal, and L. Dooley which appeared in the Proceedings of the IEEE International Conference on Consumer Communications and Networks Conference 2006, Las Vegas, USA, 8-10 January 2006. © 2006 IEEE.

in load profile of two PNs having the same energy resources at an instant, causes a proportional difference in the time-to-live for both PNs.

This fundamental limitation provides the main motivation behind this paper, which introduces a new dynamic clustering methodology for proactive load balancing by taking into account previous load profile instances in both the formation and reconfiguration of clusters. The short term chronicle of load and energy depreciation rate information is then applied to estimate the life of the PN, so assisting to more accurately identify the present *state* of the network and reconfiguration requirements. The proposed methodology defines various states of PNs in the backbone and undertakes reconfiguration activates with the objective to maximize the time that a PN spends in the *ideal state*.

The remainder of the paper is organized as follows: Section II describes the related research while Section III details the underlying network. Various states of a PN are defined in Section IV with Section V presenting the core load-balancing model. Simulation results highlighting the performance comparison of the proposed model with traditional approaches are presented in Section VI, with some conclusions presented in Section VII.

II. Related Work

The choice of routing protocol significantly impacts on balance of load on various PNs and thus, the network life time, with certain power-aware routing protocols having been proposed based on a variety of power cost functions. For instance, when the battery level of a mobile host falls below a prescribed threshold, it no longer forwards packets for other hosts [1]. In [2], five metrics based on battery power consumption have been proposed and theoretically analysed. These include; i) minimize energy consumed per packet, ii) maximize time to network partition, iii) minimize variance in node power levels, iv) minimize cost per packet and v) minimize maximum node cost. A hybrid environment consisting of battery-powered and outlet-plugged hosts is considered in [3], while two distributed heuristic clustering approaches for multicasting are proposed in [4] to minimise transmission power. Other minimal energy routing strategies include [5, 6] which presents an algorithm guaranteeing strong connectivity and assumes limited node range with shortest path algorithm. The route however, may not be the minimum energy solution due to the possible omission of optimal links when the backbone is formed. Similarly, [7] developed a dynamic routing algorithm for establishing and maintaining connection-oriented sessions using the notion of predictive re-routing to manage unpredictable topology changes. Routing algorithms in mobile wireless networks [8, 9, 10, 11] use shortest-path routing where the number of hops equals the path length. This minimises the total energy consumed to reach a destination by minimising the energy consumed per packet (or unit flow). However, if traffic is always routed via the minimum energy path, nodes along that path become energy-drained more quickly, while others remain relatively inactive.

Recently, the focus in power aware routing has been to exploit the availability of redundant servers in a WSN to make *anycast* routing a viable option, since this ensures the most efficient delivery of datagrams to one of the *best* available servers [12, 13, 14]. Anycast routing algorithms are combined with server selection policy mechanisms to achieve improved load balancing and *Quality of Service* (QoS) [15, 16] provision. The packet forwarding protocol [15] for instance, employs a *weighted random selection* approach for multi-path selection to balance the network traffic, while an alternative load balancing approach involves equalising cluster membership [17], thereby relating network load to the number of nodes in a cluster, rather than to the actual cluster traffic pattern.

Gupta [18], Chiasserini [19]and Subramanian [20] have focused on energy-efficient, hierarchical modelling of the sensor network through dynamic configuration of the tree nodes. The success of their dynamic tree models is based however, on the rather inappropriate assumption for sensor networks, that a sensor is capable of connecting to many parent nodes simultaneously. Cerpa [21] emphasized the need for a higher degree of synchronization between network components in order to reconfigure correctly. Hongwei [22] illustrated a wireless network design based upon a strict hexagonal topography, which ultimately renders the approach unsuitable for ad hoc sensor networks. Policy-based, self-managing systems were described by Joakim [23], but these impose a high computational and storage requirement on individual sensing units, while extending an already existing network was discussed by Bulusu [24, 25],though no suitable strategy for self configuration was proposed. Congzhou [26] and Kung [27] both focused on application-based parent selection techniques. While this work highlighted design issues, it lacked a quantitative description of methodologies to control the functional response of the PSN and self configuration.

In all these power-saving strategies, the routing decision is dynamically based upon instantaneous information about the minimum distance [28], residual energy, load profile of a PN [29] or a combination of them [2]. They all have one major flaw however, namely that the true network picture may be overlooked during momentary abnormal conditions in the load profile. Also, estimating the life of a PN using just the *current* energy resources has the potential for erroneous estimates, as a different PN load pattern using identical energy resources, causes a proportional difference in their life.

III. THE UNDERLYING NETWORK

The proposed model targets a far more hostile working environment, where node failures are the norm, while memory and computing resource constraints upon the nodes afford relatively complex protocols. The network topology is assumed either to be static or changing sufficiently slowly that there is time for optimally balancing the traffic in the intervals between consecutive topology changes. The performance objective is to maximise the network lifetime by balanced energy depreciation in the backbone.

The network is organized into clusters as a 3-level hierarchical design as shown in Figs. 1 and 2, with nodes at each level distributed either randomly or in a controlled topological pattern. The first level comprises the various sensor node clusters which are connected to a PN residing at level 2. The set of PNs may or may not form so-called *super clusters* [30], depending upon the connectivity in the network, an aspect that will be explored further in Section 3. The top level in the hierarchy is the *Central Commanding Infrastructure* (CCI), which is the final destination of the traffic and is not compulsory because its primary function is only to establish centralized control of the network, and thus is irrelevant in any decentralised WSN. The model exploits the PN distribution methodology proposed in [31] to ensure that any packet originating from a sensor node will take at most three hops to reach a PN. The validity of assuming that a network has at most q-hop clusters, where q is kept to a minimum, is well founded for resource constrained ad hoc networks. A large q leads to the formation of fewer clusters causing extra hops for a packet when routing between source and destination. In terms of performance, this means higher latency, greater energy consumption and more information processing per PN [17].

All sensors are assumed to transmit at the same power level and so have the same radio range r. The distance d between any sensor and its PN is equivalent to r/d hops, with maximum r/d=3. To ensure the model is sufficiently generic to be valid in a myriad of different network infrastructures including wired, wireless, mobile and ad hoc domains, no assumptions are made on: the homogeneity of node dispersion in the field, the communication media, network density or diameter and

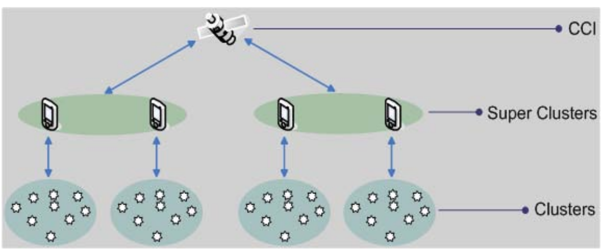

Fig.1. A hierarchy of clusters in ad hoc network architecture

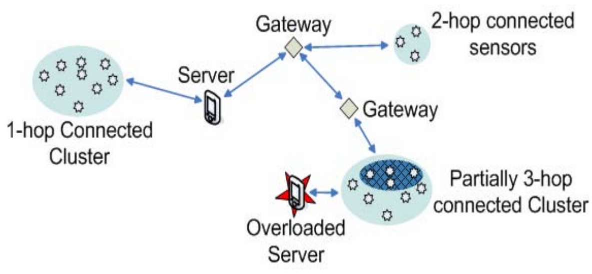

Fig.2. A clustered WSN with one, two and three hop connections, and a PN overloading scenario

distribution of energy consumption among backbone nodes. The model is therefore resource-based clustering, rather than zone-based [32].

The network thus has PNs well-distributed over the field, which can be represented by an undirected graph $G = (N,L)$, where N and L respectively represent the set of nodes n_i and links l_i. Note, while the cardinality of N remains the same, the cardinality of L may change with the creation and removal of links. An arbitrary node n_j is the *neighbour* of a node n_i iff:

$$dist(n_j, n_i) < r$$

where r is the transmission range of n. The complete neighbourhood $H(n_i)$ of n_i is then given by:

$$H(n_i) = \bigcup_{n_j \in N, n_j \neq n_i} \left\{ n_j \,\middle|\, dist(n_j, n_i) < r \right\}$$

which means that the neighbourhood of a node is the set of nodes that lie within its transmission range. There may also be a possibility that a node is physically nearer to a PN, but belongs to another PN. This is due to the fact that the clustering algorithm establishes associations between sensors and PNs based on the efficiency of the link, i.e., the closest node with the highest energy profile. So a sensor node may connect to either a closest PN or a 3-hops away PN via multi-hop links. It may also be able to connect to multiple PNs simultaneously and so, become a candidate for being a gateway node. Section 5 discusses this clustering algorithm in detail.

In order to respond to impairments and malfunctions in abnormal load situations, the algorithm reconfigures the network by directing sensor nodes to either connect to alternate PNs within their range or dynamically define gateway nodes between neighbouring clusters. Such gateways work as connectors between clusters and assist in rerouting excessive traffic from an overloaded node [33], as illustrated in the load balancing mechanism in Fig. 2. The algorithm is simulated in a custom built discrete event simulator in Microsoft .NET environment. As discussed earlier that the reconfiguration undertaken by the algorithm is based on defining various states of a PN, the next section defines these states in detail.

IV. The Busy State of a Parent Node

In order to identify the time and nature of proactive reconfiguration that suit a particular network situation, it is important to monitor a PN, especially while *Busy*. In case of steady state operation, a PN undergoes different sub states before it turns idle or its energy runs down, as shown in Fig. 3. These sub states reflect the *efficiency* of the PN which is a measure of its active participation and significance to the backbone. PN efficiency is defined in terms of two factors; the ability of a PN to grant connections to sensor nodes and its remaining life. The transition from one sub state to the other occurs when the ability of PN i to bear more load *(V)* and remaining life of the PN (∂) shift in either direction (low, high) across threshold levels u and w respectively. Analytically, efficiency measure η of PN at T given by:

$$\eta_T(i) = V_T(i) * \partial_T(i) \tag{1}$$

where
$$V_T(i) = 1 - \frac{\hat{x}_{T+1}(i)}{X(i)} \tag{2}$$

and
$$\partial_T(i) = E_T \Big/ \sum_{j=T-k}^{T} (E_j - E_{j-k}) \tag{3}$$

$V_T(i)$ is calculated in terms of average number of connections that PN granted to sensor nodes over the period T. x_j is the number of connected sensor nodes at time j while X is the total capacity of connections with a PN. The remaining life of a PN is not calculated in terms of a usual remaining-energy measure. Instead, equation (3) defines the remaining life of the PN at time T more realistically in terms of *Energy-Time* i.e. the length of time for which the remaining energy can sustain. This measure takes care of the fact that due to random deployment of sensing devices and, possibly, unbalanced load profiles, every PN can undergo different load situations. This measure is calculated using the rate of energy degradation and remaining energy level at time T. E_x in (3) is the amount of remaining energy at time x. The forecast of future load on the PN in terms of connection requests is calculated as the Moving Average (MA) of the number of connections granted from time T-k to T. It is given by:

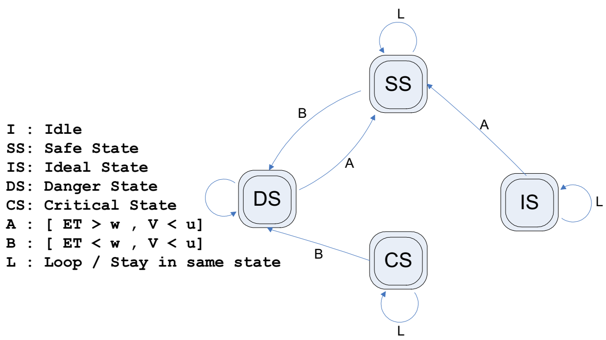

Fig. 3. PN Sub States while in Busy State

$$\hat{x}_{T+1}(i) = \frac{1}{k} \sum_{j=T-k}^{T} x_j(i) \tag{4}$$

In calculating PN efficiency, the historical connection and energy profile of a PN helps to undo the effects of spikes in connection requests, and reflects the actual trend of traffic at a PN.

The constant k used in both (2) and (3) is the width of the averaging window. It defines how quick PN states will change as a response to most recent changes in the load profile. Greater the k, lesser will be the response of the PN on current load situation and vice versa. The rationale behind is to prevent the network from responding too abruptly to momentary abnormal changes and too late to well sustaining load situations. The simulation results show the empirically established criteria for the selection of k.

While *busy*, a PN stays in *Ideal State* (IS) when the energy resources are sufficient enough to cope with prevalent load which is also low compared to the load bearing capacity of the PN. Given a load bearing capacity $V_T(i)$ at time T, a PN i should adhere to the following criteria to stay in IS:

$$\left.\begin{aligned} \partial_T(i) \geq w \mid w = L_{\mathrm{Re}q} - T \\ V_T(i) \geq u \mid u = b_T(i) * X \end{aligned}\right\} \tag{5}$$

where L_{Req} is the required life of the network and b is a tuneable dynamic variable, which is maintained by each PN to ensure equalized load in the network, so minimizing b helps also to minimize the maximum load at any instant on any PN. Selection and updation of b is detailed later in Section V-B. The thresholds u and w play key role in defining the instant of transition from one state to the other and, thus, define the nature and instant of proactive reconfiguration activities.

A PN enters the *Safe State* (SS) from IS when it experiences high load conditions while concomitantly meeting the time to live criteria in (5). If PN i transits to SS at time T through edge A (Fig. 3) then the following condition is held true:

$$\eta_T(i) < \eta_{T-1}(i) \mid V_T(i) < u \wedge \partial_T(i) > w \tag{6}$$

where w is selected so that (5) is upheld. It is now evident that the efficiency of a PN deteriorates as it transits from IS to SS, though this condition is safe as long as the time to live condition remains satisfied.

There are occasions when backbone devices start losing energy for other unexpected reasons like energy leakage and high control traffic. This situation is crucial since it leads to the transition of even highly efficient PNs from IS to *Critical State* (CS). The efficiency of a PN i in CS satisfies the condition:

$$\eta_T(i) < \eta_{T-1}(i) \mid V_T(i) < u \wedge \partial_T(i) > w \tag{7}$$

While the load on a PN in CS may be within an acceptable range, this state is undesirable since the time to live condition is violated. A PN in this state communicates its expected time of failure to neighbouring

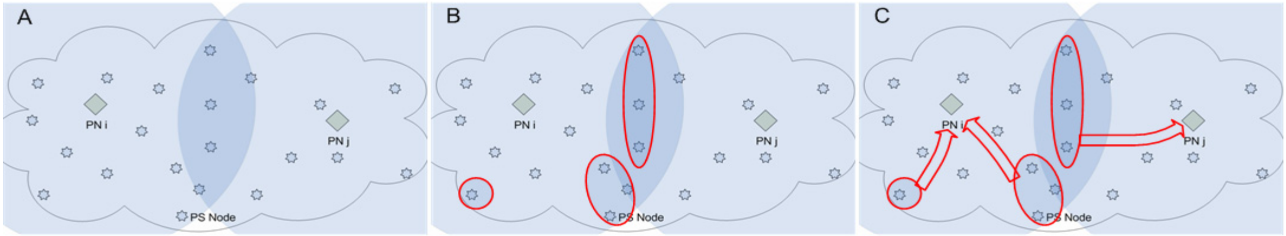

Fig. 4A: Load Imbalance Problem in clusters, 4B: GOR applied with objective (a) satisfied,
4C: GOR and LBR applied with both objectives (a), (b) satisfied

PNs so proactive action can be taken to avoid the prospective malfunction.

One of the predictable anomalies in a wireless network backbone is PN failure due to the loss of energy. Unless the energy resources are sufficient enough, the transition that is highly likely in such circumstances is to *Danger State* (DS) whereby the PN i is least efficient in such a way that:

$$\eta_T(i) < \eta_{T-1}(i) \mid V_T(i) < u \wedge \partial_T(i) < w \qquad (8)$$

A PN in DS fails to function very soon ahead of task completion. An estimate of time to fail for PN i in DS within which any reconfiguration methodology should affectively balance the load to save the network from malfunctions is given by:

$$T_{f(T)}(i) <= \partial_T(i) \qquad (9)$$

Since, a PN in IS ensures the time to live requirements of the network, this state is a high priority desirable state of our load balancing algorithm that reconfigures the network such that the PNs who have transited to less efficient states are brought back at least closer to IS. The following section presents the solution to this load balancing problem in detail.

V. Load Balancing and Redistribution (LBR)

The objective of the LBR model is to shift the load from a PN i such that its efficiency at time T is increased, given by:

$$\eta_T(i) > \eta_{T-1}(i) \mid V_T(i) >= u \qquad (10)$$

i.e. the efficiency increases as the load on PN i increases. To achieve this goal, the following objectives for the LBR model are defined:

a) Maximize the balance of load on all PNs so the energy resources drain is evenly in each cluster
b) Retain the traffic flow without major intermission.

The first of these objectives involves equalizing the time to live for every PN in a cluster so that if $L_{Req}(i)$ is the required life time for cluster i, then the estimated time to live for every PN satisfies the following condition:

$$\partial_0 >= L_{Req}(i) \qquad (11)$$

where ∂_0 is the estimated time to live for a PN at T=0. Equation (11) illustrates that LBR should regularize the traffic in (10) in such a way that the use of energy resources is optimized in accordance with (11). However, the time to live is not the only bounding factor on the amount of regularization of traffic as the second of the above objective implies. The sensor connections with the network backbone must also be preserved in order to satisfy (10).

An important issue in employing a LBR mechanism is that it can be used only in situations where sensor nodes have more than one option for selecting PN, or where mutually agreed gateways exist between neighbouring PNs to route the traffic. The algorithm presented in this paper refuses a connection request, if required, to meet objective (a), while concomitantly arranging an alternate route for the sensor that requested that connection to meet objective (b). The next sections detail this method.

A. *Grounding of Request* (*GOR*)

Consider the unbalanced load scenario shown in Fig. 4A, where because of the random deployment of sensor nodes, the number of sensors that form a cluster with PN i are more than those with PN j. Under steady state network conditions, and without any load balancing mechanism, the efficiency of the two PNs at time T are observed as:

$$V_T(i) << V_T(j) \Rightarrow \left. \begin{array}{l} \partial_T(i) << L_{\mathrm{Req}}(i) - T \\ \partial_T(j) >> L_{\mathrm{Req}}(j) - T \end{array} \right\} \qquad (12)$$

These energy-time measures unravel the states for both PNs whereby PN i has already approached the critical DS while PN j remains comfortably within the IS. This imbalance in cluster lives is at first conjectured to be due to imbalance load profiles of the two PNs, however, the application of LBR to this scenario reveals the validity of this argument.

The core idea behind *Grounding of Request* (GOR) is to decline a request for connection raised by a sensor node, if:

$$x_T(i) + 1 > b_T(i) * (X) \quad \text{where } 0 \le b \le 1 \qquad (13)$$

where $x_T(i)$ is the number of connections at PN i at time T. b is a tuneable dynamic variable. The rationale is to increase the efficiency so that (10) is satisfied. If b_T is the same for each PN in the backbone and the density of the network is high, GOR guarantees that for a pair of PNs i and j in a cluster, following condition holds:

$$V_T(i) \simeq V_T(j) \qquad (14)$$

i.e. the average load on each PN is equalized. Employing this method of declining the connection requests to the load imbalance example in Fig. 1A, we get:

$$\partial_T(i) \simeq \partial_T(j) >> L_{\mathrm{Req}} - T$$

i.e. the load on the PNs is regulated in such a way that both PNs are moved into IS and the objective of optimality in (10) is achieved. The new configuration is shown in Fig. 4B whereby both PNs have the same number of connected sensor nodes. In order to form these clusters, the traditional methodology of selecting sensor nodes based on *Closest Node First* (CNF) approach is employed. While (11) is upheld, it is observed that the maximization objective (a) is satisfied, the second objective (b) is not. The example reveals that (a) is achieved at the cost of denying the connection requests raised by sensor nodes marked by circles, so they are disconnected from the network.

To also justify LBR objective (b), it is required that the disconnected sensor nodes highlighted by circles must also be given access to the backbone. For this purpose, a new dynamic cluster formation algorithm is introduced, the key feature of which is that it employs the concept of *Connection Efficiency* (CE) to evaluate the choice between multiple connectivity options. Unlike traditional connectivity protocols which make a PN as a decisive entity to grant the connection, the new algorithm gives this choice to a sensor node to decide whether it wants to connect to a particular PN based on a route bearing minimum cost. However, this does not incur a processing overhead on the sensor since the calculation of efficiency measures is done by PNs and are only communicated to sensor nodes as state beacons during connection establishment steps.

B. Dynamic Clustering Algorithm (DCA)

Consider the scenario shown in Fig. 5 in the form of a graph where a sensor node *p* (one of the disconnected sensor nodes from Fig. 4B) is able to connect to two PNs *i* and *j* respectively. The position of sensor node is such, that if the usual method of CNF is employed, it would connect to PN *i*. The load profile of the two PNs at time *T* is given by (14). In addition to the ability to directly connect to PN *j*, a multihop path through gateway *a*, designated by PN *i*, and gateway *b*, representing cluster headed by PN *j*, is also available. As shown in Fig. 5, each of these connectivity routes has an associated communication cost, labelled on the edges connecting the two entities as A, B, C, D and E. The shorter the distance, the lower the communication cost and vice versa.

To avoid the load imbalance caused by CNF (Fig. 4A), this cost measure is combined with the energy-time measure of the PN. If Ŝ is a set of nodes that form the route from a sensor *p* to a PN *i* and Â is a set of communication costs between every two nodes from Ŝ, then the connection efficiency between sensor *p* and PN *i* is given by:

$$\lambda_T(p,i) = \sum \frac{\eta_T(n)}{c} \forall (n \in \hat{S} \wedge c \in \hat{A}) \qquad (15)$$

TABLE I – The Dynamic Clustering Algorithm (DCA)

Preconditions: Sensor p needs to connect to PN i. R is a set of all possible available routes. C is a set of communication costs associated with each element from R. E is a set of energy-time measures for each element in R. R, C and E are known to sensor p.
1. LET Q is an empty set of connection efficiencies. 2. Each PN calculates CEs using E and C, and sends to its neighbourhood. Sensor p populates Q with REs received from each connected PN 3. Route(p) = max[Q] … equation (17) 4. Sensor p sends connection request to Route(p) PN 5. PN i receives connection request from p 6. PN i evaluates $x_T(i) + 1 > b_T(i) * (X)$… equation (13) 7. IF (13) is false then connection is granted, End 8. ELSE connection request is grounded 9. PN i raises the value of $b_T(i)$ as $b_T(i) = b_T(i) + .01$ 10. $b_T(i)$ is communicated to all PNs to ensure balance of load 11. PN i again evaluates (13) 12. IF (13) is false, connection granted, ELSE GOTO Step 9
Postconditions: Sensor p connected to PN i through most efficient route. The balance of efficiency of PNs around PN i, is improved.

where *c* is the distance between two nodes. The use of energy-time measure for evaluating a route prevents sensor nodes from connecting to a highly loaded PN, while other PNs with better energy resources are available. From Fig. 5, for all possible routes, the CE measures are given as:

$$\lambda_T(i,p) = \eta_T(i)/A \ , \ \lambda_T(j,p) = \eta_T(i)/B$$
$$\lambda_T(\bar{j},p) = \lambda_T(a,p) + \lambda_T(b,a) + \lambda_T(j,b)$$
$$\lambda_T(\bar{j},p) = \frac{\eta_T(a)}{C} + \frac{\eta_T(b)}{D} + \frac{\eta_T(j)}{E} \qquad (16)$$

where $\lambda_T(\bar{j},p)$ is the CE measure of an alternate route from sensor *p* to PN *j* that includes the gateways *a* and *b*. It should be noted that in the case of a multihop route, the route efficiency includes the energy-time measures of all the intermediate devices as well as the communication cost for each edge that is included in the route (16). Upon receiving the efficiency measures from an array of nodes, a sensor processes the efficiencies vector to select the best

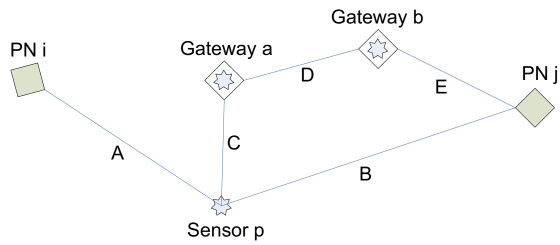

Fig. 5. Multiple connectivity routes and associated communication costs

route using the following relationship:

$$Route(p) = \max[\, \vec{\lambda}_T(s,d)\,] \qquad (17)$$

i.e. the route with maximum efficiency is selected. The vector $\vec{\lambda}_T(s,d)$ contains the efficiency measures of all possible routes from sensor s to each PN d within three hops. The route selected by (17) establishes the connection of a sensor with an efficient PN via the most efficient route. However, if the energy-time measures for the given PNs are equal, CE works same as CNF. Fig. 4C shows the connections established as a result of employing cluster formation algorithm. It must be noted that in spite of being closer to PN i, the sensor nodes which are able to connect to PN j as well, decide to connect to PN j, because of significantly lower $V_T(i)$, thereby balancing the load on the clusters. Due to this change in load, the levels of V_T and energy-time measure for both the PNs are regulated such that both PNs transit to IS while justifying both the LBR objectives as well. The complete cluster formation algorithm is detailed in Table I.

C. *Energy Restoration Action* (*ERA*)

Energy restoration becomes necessary when the entire backbone undergoes high load situations in such a way that applying GOR and DCA does not help. This situation arises when the majority of the nodes in the network switch to either CS or DS where the remaining energy of a PN is insufficient to meet the network life requirements. In these circumstances, re-distribution of network load to maintain connectivity across the backbone (objective b) pushes the PNs further into a high risk situation, where the energy drainage can be very uneven and the likelihood of collapse in different parts of the backbone is high.

In such circumstances, the deployment of more resources becomes inevitable. Depending upon the level of PN access and time criticalness, two methods can be adopted to *refuel* the backbone:

i. If PNs are not easily accessible, or it is difficult to replace/recharge power resources, then deploy redundant PNs at network design time, or at a time when the *in-situ* proactive monitoring system detects the majority of the network backbone nodes in either CS or DS.
ii. Energy resources should be recharged.

In this work, we assume that the backbone is always accessible, so the ERA component always recharges PNs when load redistribution fails to assist further in bringing a PN back into either IS or SS.

VI. Results and Analysis

Extensive simulation experiments were conducted to evaluate the performance of the LBR model. The model was implemented for different load profiles and node densities. In developing the simulation test bed, the guidelines proposed in [34, 35, 36] were considered to test both the validity of the conceptual model of the network, and all the various underlying assumptions and limitations concerning the network's internal mechanisms and calculating precision of results at 95% confidence level. Where various system configurations were tested with various densities of sensor nodes and parent nodes, simulating ad hoc networks, the results of one case study are presented in this paper for illustration purposes. For this purpose, a major disaster site in Australia, the Granville train collapse was taken as a test bed shown in Fig. 6. The ad hoc network comprised a backbone of four PNs and densely deployed mobile sensor nodes.

The results shown are the illustration of similar results observed for 80 contiguous experiment runs conducted to control the errors of final results and establish 95%

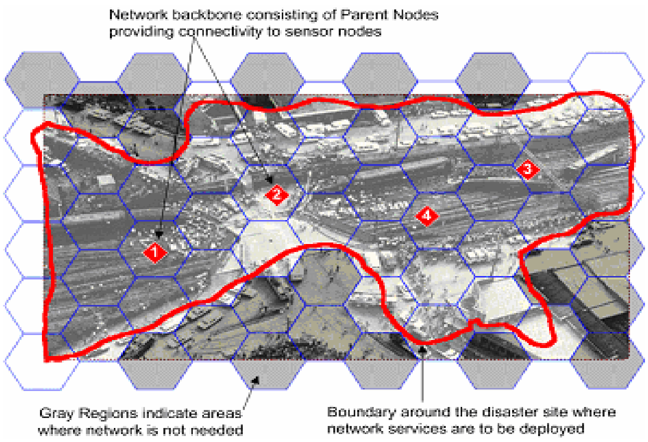

Fig. 6 The disaster recovery network simulated for Granville train collapse

TABLE II – Simulation Environment Parameters

Attribute	Value			
Area under Surveillance	Open irregular Terrain of appx. 10000m² dimensions			
Deployment Topology	Random			
PN Comm. Range	15m			
Sensor Comm. Range	3m-13m			
Density of Sensor nodes	125 Randomly Deployed			
Density of PNs	4			
Mobility	Stationary, Semi Mobile			
Performance Metrics	Network Efficiency, Load Balancing Network Life			
Control Packet Size	500bytes			
Network Activity Time	10 min			
Power Consumption (mW)	Tx	Rx	Idle	Sleep
	14.88	12.50	12.36	0.016

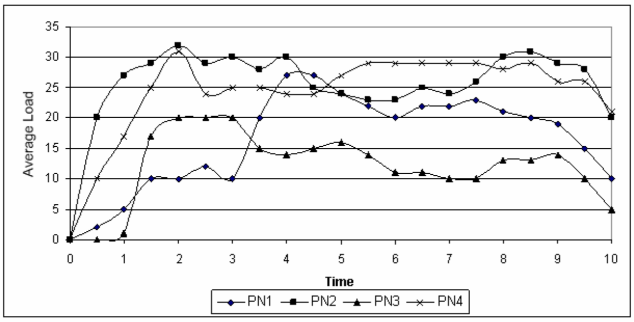

Fig. 7. Load Balancing and Convergence among PNs

confidence levels. Each experiment uses a different randomly generated topology for sensor nodes, while the parent nodes are deployed according to the design model in [31]. The gateway locations vary depending upon the combined topological connectivity of sensor nodes and PNs, and are determined by employing the LACON protocol [37]. Each sensor node is assigned identical randomly-generated residual energy level between 0 and 2 Joules and each PN is assigned an initial energy between 15 and 20 Joules. If the energy level of any node reaches zero, it is considered non-functional. Two nodes are said to have a wireless link between them if they are within communication range of each other.

It is assumed the channel is collision free and a node sensing a target produces data packets at a rate tuned to generate a specific network-wide load profile. We assume that each PN can handle at most 15 nodes in its cluster in terms of resource allocation.

The purpose of the illustrative network was to relay information to help the rescue squad to communicate across the disaster site. Because of the criticality of domain, the traffic so generated is evaluated to be time critical. The performance of the LBR model was evaluated in terms of the following key QoS metrics: *load balancing among servers*, *conformance with network life requirement* and *network efficiency*. The results were compared with the traditional *Closest Node First* (CNF) and *Maximum Energy First* (MEF) approaches. The following sections discuss the results and analyse the new load balancing methodology proposed in this paper.

A. Load Balancing

Fig. 7 shows the way load was balanced among the PNs using the LBR methodology. It is observed that the load profiles of PNs 2 and PN 4 remain very close to each other. This effect was because of the close proximity of both PNs, with one PN being used by LBR to balance the high load on the other and vice versa. Closely observing the graphs, it is revealed that the loads profiles of PN1 and PN3 were also tailored by those of PN2 and PN4. This effect illustrates a very promising feature of the self configuration methodology since it revealed the effectiveness of the algorithm on outliers also. The network of gateways was utilized in this case and multihop routes were established in order to keep the objectives (a) and (b) intact.

B. Network Efficiency

The basis of the new methodology is the way network devices are evaluated to form the clusters as the network operates. The metric for such evaluation, called *efficiency*, is illustrated in Fig. 8. Figure 8A shows the efficiency of PNs when clusters were formed using the new LBR methodology, while the plots in Fig. 8B show the same for the PNs when the load balancing was done using instantaneous information regarding energy and status of connection availability of the PNs.

These plots clearly reveal the benefit derived by employing the new strategy for cluster formation. As the vertical axis shows the efficiency of PNs, it can be seen that for LBR, where every PN stays alive throughout the stipulated network lifetime, all stay in or close to IS. Conversely, Fig. 8B confirms that because of the instantaneous information input, the traditional approach suffers from a high number of efficiency fluctuations and two of the PNs are observed to collapse early and fail to meet the time-to-live criteria.

Another important aspect unraveled by simulations is the fast convergence of efficiencies of PNs while configured using LBR. Fig. 8A shows that after an initial

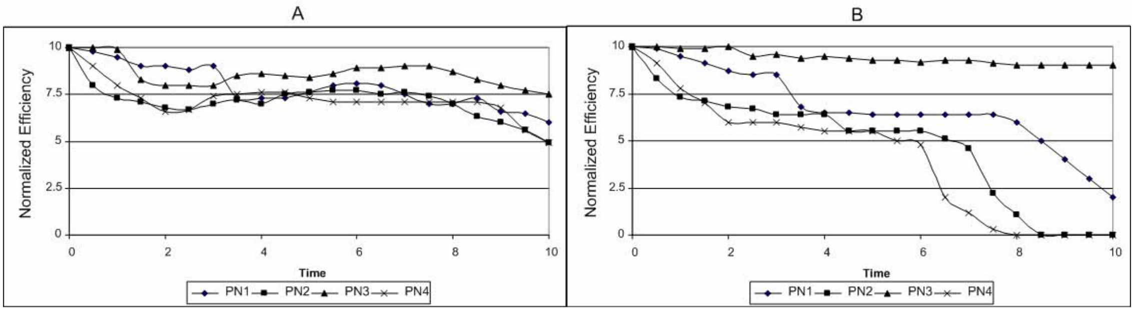

Fig. 8A: Efficiency of PNs when load balancing done using Energy-Time method,
8B: Efficiency as a result of load balancing with traditional method

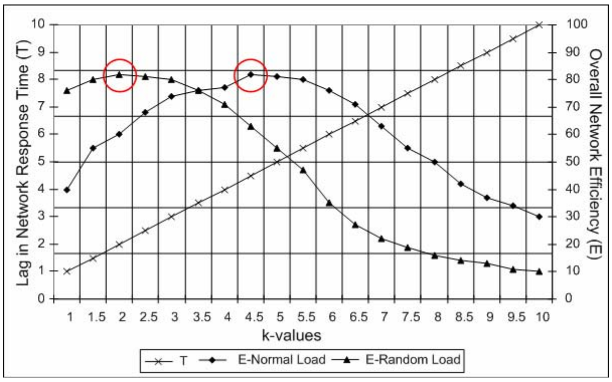

Fig. 9. Effect of different values of k on efficiency and network response time

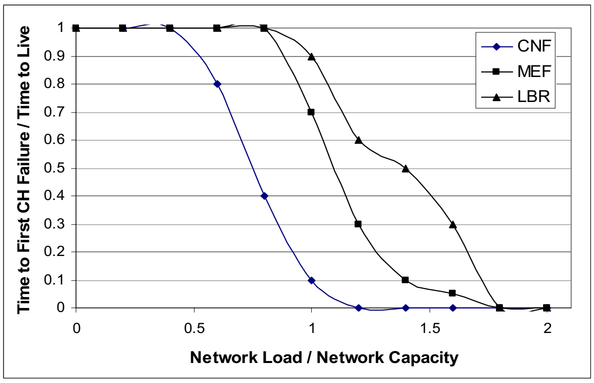

Fig. 10. Conformance analysis of CNF, MEF and LBR with required Network Life

setup, the LBR converges the various efficiencies of the backbone PNs to the same level. For instance, the efficiency of PN4 decreases as it takes more load from the cluster of PN2, thereby helping PN2 to improve its efficiency. The efficiencies of PNs in traditional configuration approach do not converge, resulting in the collapse of PN2 and PN4. The reason for this when compared with LBR is that since the sensor nodes rely only upon the current picture of the network to select the route, even a momentary drop in load keeps the sensor nodes inclined towards the closest PN, in spite of the overall high load on the PN. The superior efficiency of PN3 in Fig. 8B as compared with Fig. 8A is directly due to it being an outlier, that is, its distance from the other clusters prevented it from actively participating in sharing the load of other PNs, thereby rendering it idle for most of the time.

C. Analysis of k Values

The value of *k* impacts on cluster formation, with the higher the value of *k*, the longer time it will take for the self configuration to converge towards current load conditions. Fig. 9 illustrates this aspect where the left vertical axis is the time (*T*) that the network takes to adapt to current load, and the straight line shows a direct, proportion in the *k* and *T*. The higher the value of *k*, the load profile of the network will carry more historic details that may show a wrong picture of current state. Since, the state of the PN is crucial for sensor nodes to decide on whether to connect to a PN or not, this *k* value affects cluster formations and so does to the efficiency of the PNs.

The other aspect as shown in Fig. 9 is the selection of *k* value. In order to justify the selection, different load profiles were applied which can be broadly classified in two categories, *Normal Load* and *Random Load*. In the case of normal, the sensor nodes were caused to generate regular traffic with lesser fluctuations in the load. On the other side, in a random load profile, sensor nodes were made to generate traffic unevenly.

The empirical results in Fig. 9 show that the *best* value of *k* for which the average efficiency of the network is a maximum remains under 5 for both categories of load. However, for the normal load profile, the value of *k* suggested to take the condition of 4.5 previous load instances into consideration. Because of high dynamicity of state in case of random load, *k* value appeared to be lower than for normal to reflect the states of the PNs accurately and was found to be 2.5.

It was also found that increasing the value of *k* further from the best, decreased the network efficiency and it appeared to deteriorate more quickly for random load because of the rapid loss of current network state.

D. Network Life

In the simulation, the performance of LBR, CNF and MEF was analyzed in terms of its conformance to network life. This measured the robustness of a model by applying high traffic compared to the capacity of the backbone, so their capability to balance the load is assessed such that the network life criterion is upheld. In this particular experiment, the network is considered alive until the first PN fails. Fig. 10 shows the conformance graphs for CNF, MEF and LBR, from which it is clear that in CNF, at least one PN is overloaded and fails to meet network life criteria as the network approaches it capacity, while the performance of MEF drops dramatically beyond the maximum network capacity. LBR however, provided superior network life conformance even beyond this point, by virtue of considering the ratio of the distance between the nodes to the estimate of remaining PN life, and giving both parameters equal weight in making the target server selection.

VII. Conclusions

This paper has presented a new *load balancing and redistribution* (LBR) technique for adhoc networks in general and wireless sensor networks in particular. Both analytic and simulation results have proven that using energy-time measure to estimate the life of the PNs and controlling the formation of clusters using *a priori* information about the network load profiles, provided a significant performance improvement over traditional way of relying solely on the instantaneous state. Dynamic numerical bounds on the maximum load on each PN have been developed and a load distribution technique devised to maximize the time a PN stays in ideal state and minimise the intermission of traffic. Simulations have shown promising results in the ay the proposed LBR methodology out performed the traditional CNF and MEF techniques by keeping the PNs in the ideal state for most of the time. The results also confirmed the model convergence behaviour in redistributing the load among the backbone PNs and stabilizing the network in its steady state. Using the averaged historic information about the network also meant the new model was very robust so limiting the network from acting upon instantaneous false loads and comfortably meeting network life requirements.

REFERENCES

[1] J. Gomez, A.T. Campbell, M. Naghshineh, C. Bisdikian. *"A distributed contention control mechanism for power saving in random-access ad-hoc wireless local area networks,"*. in *Proc. of IEEE International Workshop on Mobile Multimedia Communications*. 1999: p. 114-123

[2] Suresh Singh , Mike Woo , C. S. Raghavendra. *"Power-aware routing in mobile ad hoc networks,"*. in *Proc. of ACM/IEEE international conference on Mobile computing and networking*. Oct 25-30. 1998. Dallas, Texas, United States: p. 181-190

[3] J.H. Ryu, D.H. Cho. *"A new routing scheme concerning power-saving in mobile ad-hoc networks,"*. in *Proc. of IEEE International Conference on Communications*. 2000: p. 1719-1722

[4] J.H. Ryu, S. Song, D.H. Cho. *"A power-saving multicast routing scheme in 2-tir hierarchical mobile ad-hoc networks,"*. in *Proc. of IEEE Vehicular Technology Conference*. 2000: p. 1974-1978

[5] Dennis J. Baker, Anthony Ephremides, *"The Architectural Organization of a Mobile Radio Network via a Distributed Algorithm"*. IEEE Transactions on Communications, 1981. 29(11): p. 1694-1701.

[6] Baker, A. Ephremides J.E. Wieselthier D.J., *"A Design Concept for Reliable Mobile Radio Networks with Frequency Hopping Signaling,"*. Proceedings of IEEE, 1987. 75(1): p. 56-73.

[7] A. Michail, A. Ephremides. *"A Distributed Routing Algorithm for Supporting Connection-Oriented Service in Wireless Networks with Time-Varying Connectivity,"*. in *Proc. of the Third IEEE Symposium on Computers & Communications*. June. 1998. Athens, Greece: p. 587-591

[8] Maltz, David B Johnson and David A, *"Dynamic source routing in ad hoc wireless networks,"*. Mobile Computing (ed. T. Imielinski and H. Korth), Kluwer Academic Publishers, Dordrecht, The Netherlands., 1996. 353.

[9] Garcia-Luna-Aceves, Shree Murthy and J. J., *"An Efficient Routing Protocol for Wireless Networks,"*. ACM Mobile Networks and Applications Journal, Special issue on Routing in Mobile Communication Networks, 1996. 1(2): p. 183-197.

[10] Corson, Vincent D. Park and M. Scott. *"A Highly Adaptive Distributed Routing Algorithm for Mobile Wireless Networks,"*. in *Proc. of IEEE Infocom*. 1997. Kobe, Japan: p. 1405-1413

[11] Bhagwat, Charles Perkins and Pravin. *"Highly Dynamic Destination-Sequenced Distance-Vector Routing (DSDV) for Mobile Computers,"*. in *Proc. of the ACM SIGCOMM*. October. 1994: p. 234--244

[12] C. Partridge, T. Mendez, W. Milliken, *"Host Anycasting Service"*. IETF RFC 1546, 1993.

[13] E. Zegura, M. Ammar, Z. Fei, S. Bhattacharjee, *"Application-Layer Anycasting: A Server Selection Architecture and Use in a Replicated Web Service"*. ACM/IEEE Trans. on Networking, 2000. 8(4): p. 455-466.

[14] H. Miura, M. Yamamoto, *"Server Selection Policy in Active Anycast"*. IEICE Trans. on Comm, 2001. E84-B.

[15] D. Xuan, W. Jia, W. Zhao, H. Zhu, *"A Routing Protocol for Anycast Messages"*. IEEE Trans. on Parallel and Distributed Systems, 2000. 11(6).

[16] W.T. Zaumen, S. Vutukury, and J.J. Garcia-Luna-Aceve. *"Load-Balanced Anycast Routing in Computer Networks"*. in *Proc. Of IEEE Symposium on Computers and Comm.* 2000

[17] M. Chatterjee, S. K. Das, D. Turgut, *"WCA: A Weighted Clustering Algorithm for Mobile Ad Hoc Networks"*. Kluwer Cluster Computing, 2002. 5: p. 193-204.

[18] H. Gupta, S. Das, Q. Gu. *"Connected Sensor Cover: Self-Organization of Sensor Networks for Efficient Query Execution"*. in *ACM,MobiHoc*. 2003

[19] C.F. Chiasserini, I.Chlamtac, P. Monti, A. Nucci, *"An Energy efficient method for nodes assignment in cluster-based ad hoc networks"*. ACM/Kluwer Wireless Networks Journal, 2003.

[20] L. Subramanian , R. H. Katz. *"An architecture for building self-configurable systems"*. in *ACM international symposium on Mobile ad hoc Networking & Computing*. November 20,. 2000. Boston, Massachusetts

[21] A. Cerpa, D. Estrin. *"ASCENT: adaptive self-configuring sensor networks topologies"*. in *IEEE INFOCOM*. 2002

[22] Arora, H. Zhang and A., *"GS^3: scalable self configuration and self-healing in wireless sensor networks"*. Computer Networks, 2003. 43.

[23] J. Eriksson, N. Finne, S. Janson. *"Towards Self-Managing Systems"*.http://www.ercim.org/publication/Ercim_News/enw58/eriksson.html

[24] N. Bulusu, J. Heidemann, and D. Estrin. *"Adaptive Beacon Placement"*. in *21st International Conference on Distributed Computing Systems*. April. 2001. Phoenix, AZ: p. 489-498

[25] N. Bulusu, J. Heidemann, and D. Estrin. *"Scalable Coordination for Wireless Sensor Networks: Self Configuring Localization Systems"*. in *ISCTA*. 2001. UK

[26] C. Zhou, B. Krishnamachari. *"Localized topology generation mechanisms for wireless sensor networks"*. in *GLOBECOM*. 2003

[27] Vlah, H. T. Kung and Dario. *"Efficient Location Tracking Using Sensor Networks"*. in *IEEE WCNC*. March. 2003. New Orleans

[28] A. D. Amis, R. Prakash, T. H. P. Vuong, D. T. Huynh. *"Max-Min D-Cluster Formation in Wireless Ad Hoc Networks"*. in *Proc. of IEEE INFOCOM*. March. 2000

[29] C. Y. Lin. J. H. Lo, S. Y. Kuo. *"Load-Balanced Anycast Routing"*. in *proc. of the 10th Intl. Conf. on Parallel and Distributed Systems*. 2004

[30] Wendi Heinzelman, Anantha Chandrakasan, and Hari Balakrishnan. *"Energy-Efficient Communication Protocols for Wireless Microsensor Networks"*. in *Proc. of Hawaaian Int'l Conf. on Systems Science*. 4-7 Jan. 2000: p. 10

[31] M. Iqbal, I. Gondal, L. Dooley. *"Investigating the Dynamics of Pervasive Sensor Networks through Parent Localization"*. in *Proc of Australian Telecommunication Networks and Applications Conference*. Dec. 2004: p. 268-275

[32] M. Iqbal, I. Gondal, L. Dooley. *"A Novel Load Balancing Technique for Proactive Energy Loss Mitigation in Ubiquitous Networks"*. in *Proc of IEEE Consumer Communications and Networking Conference (CCNC'2006)*. 2006. LAS VEGAS, US

[33] Y. Wang, W. Wang, X. Y. Li. *Distributed Low-Cost Backbone Formation for Wireless Ad Hoc Networks*. in *Proc of The Sixth ACM International Symposium on Mobile Ad Hoc Networking and Computing (MobiHoc'05)*. 25-28 May. 2005. Urbana-Champaign, Illinois, USA.: p. 2-12

[34] K. Pawlikowski, H. -D. J. Jeong and J. -S. R. Lee, *"On Credibility of simulation Studiesof Telecommunication Networks"*, in *IEEE Communications Magazine*. 2002. p. 132-139

[35] P. Bratley, B. L. Fox, L. E. Schrage, *"A Guide to Simulation"*. 1983: Springer Verlag.

[36] Kelton, Am Law and W., *"Simulation Modeling and Analysis, 3/E"*. 2000: McGraw Hill.

[37] M. Iqbal, I. Gondal, L. Dooley. *"LACON: Localized Autonomic Configuration in Pervasive Sensor Networks"*. in *Proc of IEEE Intelligent Sensors, Sensor Networks and Intelligent Processing Conference*. Dec. 2004. Melbourne Australia: p. 31-36

Mudasser IQBAL received a B.S. Honours in Computer Science in 2003 from University of Engineering and Technology, Pakistan. He has been a systems analyst at National Engineering Services Pakistan (NESPAK) playing lead roles in various e-Government projects. He is currently perusing a Ph.D. degree in the field of self-configuring wireless sensor networks, at Monash University, Australia. He is an IEEE and Australian Research Council Communications Research Network member. His research interests include pervasive health care, monitoring and self-configuration of ad hoc wireless networks, high performance cross layer protocol design and simulation.

Dr Iqbal GONDAL is working as a senior lecturer with Monash University Australia and teaches project management and Data communication and Networks. Prior joining Monash, he worked in the capacity of research fellow and senior systems engineer for seven years in Singapore and Australia with General Motors, Singapore Government and other industries. He has experience of ten years in network design and development, project management, System design and integration, SCADA, intelligent techniques, adaptive systems and wireless switches for financial services. Currently, he is supervising a team of researcher in ad hoc network, bioinformatics and tailored monitoring systems. He is a member of IEEE for fifteen years and is also member of IEE and Institute of Engineer Australia. He has authored forty six research papers in reputable International conferences and Journals. His research interests are: Bioinformatics, Pervasive networks, Technology transfer, Information fusion, On-line education, and Networking and Communication.

Prof. Laurence S. DOOLEY is Chair of Multimedia Technology in the Faculty of IT at Monash University, where his principal research interests are Multimedia Signal Processing, Wireless Sensor Networks (WSN), Mobile Communications, Bioinformatics, Technology Transfer and R&D Commercialisation strategies for small business. He has published over 150 peer-reviewed scientific scholarly papers and successfully supervised 13 higher degree research students to completion and is currently mentoring ten PhD students. He is the co-inventor of two patents upon missing value estimation in cDNA microarray data and WSN source localisation respectively, and was jointly awarded the Outstanding Paper Prize sponsored by the IEEE Communications Society at the International Conference on Next Generation Wireless Systems (ICNEWS'06) earlier this year. He has held a number of international visiting professorial positions including, in 2006 at the Deutsches Zentrum fur Luft und Raumfahrt (DLR)-Germany's National Institute for Aeronautics and Space Research in Berlin. He is a Senior Member of the Institute of Electrical and Electronic Engineers (IEEE) as well as being a full member of the British Computer Society and Engineering Council (Chartered Engineer).