This full text paper was peer reviewed at the direction of IEEE Communications Society subject matter experts for publication in the IEEE Globecom 2010 proceedings.

Energy-Aware Real-Time Opportunistic Routing for Wireless Ad Hoc Networks

Wei Yang*, Weifa Liang† and Wenhua Dou*
*School of Computer Science, National University of Defense Technology, ChangSha, HuNan, 410073, P.R.China
†School of Computer Science, Australian National University, Canberra, ACT 0200, Australia
Email: {weiyang, wliang}@cs.anu.edu.au, douwh@vip.sina.com

***Abstract*—Existing studies on the design of routing protocols for wireless ad hoc networks mainly focused on energy efficiency. However, in many real-time applications such as target tracking and bush fire surveillance, latency is an important concern, and little attention has been paid to it in the design of routing protocols for such applications to meet the specified Quality of Service (QoS) requirements like the end-to-end latency constraint. In this paper we propose an energy-aware, opportunistic routing protocol** `EARTOR` **for requests with QoS constraints, through striking the elegant balance between the energy consumption and the end-to-end latency. Our objective is to maximize the number of requests realized when dealing with a sequence of requests arrived one by one. The core techniques adopted include the cross-layer design that incorporates the duty cycle, a bidding mechanism for each relay candidate that takes its residual energy, location information, and relay priority into consideration. We finally conduct experiments by simulations to evaluate the performance of the proposed protocol against existing ones, in terms of the dynamic delivery ratio and the network capacity. The experimental results demonstrated that the proposed protocol outperforms the state-of-the-art protocols significantly.**

keywords: wireless ad hoc networks; localized routing; routing with QoS constraints; network capacity maximization; on-line routing; energy optimization; opportunistic routing.

I. INTRODUCTION

Energy efficiency has long been the major concern in the design of routing protocols for ad hoc networks to prolong the network operational time. However, in many real-time applications including intruder detection, target tracking, bush fire surveillance, and industrial process control monitoring [6], the end-to-end latency constraint is critical and of prominent concern. In general, most real-time applications can be categorized into two classes: the hard real-time and soft real-time applications. The former requires that the total time spent on routing a message from its source to its destination be no greater than the specified latency constraint, while the latter allows a certain degree of relaxations in the specified latency, for example, it is acceptable if the actual latency is $(1+\epsilon)\Gamma$, where Γ is the specified latency constraint and ϵ is a constant.

In practice, latencies of different requests can be various under different real-time applications, there is always a trade-off between energy and latency in terms of designing routing protocols to realize the requests in the network. With the further development of wireless technology recently, some commercial wireless sensors like XSM2 motes and WINS node have been equipped with multiple transmission power levels [1], [7]. To meet the latency constraint in real-time applications, it is urgent to have appropriate transmission power level scheduling at each relay node for message relay to meet its end-to-end latency constraint.

In the design of routing protocols for ad hoc networks, to save the energy of ad hoc nodes, the sleep and wake-up mechanism has been widely adopted, where nodes keep at sleep mode in most of time, and wake-up for a small fraction of a time period, consisting of sleep and wake up times. For convenience, we refer to the *duty cycle* of a node as the ratio of the wake-up time to the entire time period, and there is a trade-off between the duty cycle and the end-to-end communication latency in deterministic routing protocols [4], [8]. Although nodes can save more energy under lower duty cycles, the latency constraint may not be guaranteed, since the message holder has to wait for the next-hop designated node to wake up to relay the message.

Thus, by jointly taking the duty cycle and multiple transmission power levels into consideration, how to to design an on-line, distributed, and energy-aware routing protocol is quite challenging. The design aim is to maximize the *network capacity*, where the network capacity is defined as the total number of realized requests as the *delivery ratio* is no less than a specified threshold, while a delivery ratio is defined as the ratio of the number of realized requests to the total number of injected requests.

The main contributions of this paper are as follows. An Energy-Aware Real-Time Opportunistic Routing (`EARTOR`) protocol for routing requests with specified end-to-end latency constraints is proposed, which strikes the elegant balance between the energy consumption and the end-to-end latency and aims to maximize the number of the requests realized. The core techniques adopted include the cross-layer design that incorporates the duty cycle, a bidding mechanism for each relay candidate that takes its residual energy, location information, and relay priority into consideration. We finally conduct extensive experiments by simulations to evaluate the performance of the proposed protocol, in terms of the dynamic delivery ratio and the network capacity. The experimental results demonstrated that the proposed protocol outperforms the state-of-the-art protocols significantly.

The rest of the paper is organized as follows. We first review the related work in Section II. We then introduce the system model and the problem definition in Section III. What followed is to present the details of the proposed protocol EARTOR in Section IV. We finally conduct the performance evaluation of the proposed protocol through experimental simulations in Section V, and we conclude in Section VI.

II. Related Work

There has been tremendous interest in the design of real-time geographic and opportunistic routing protocols. The pioneer work by Zorzi and Rao [10], [11] is a geographic routing protocol GeRaF, where each node follows a duty cycle, and different nodes can wake up and sleep asynchronously. Although GeRaF takes the duty cycle into consideration, it is a soft real-time routing protocol which neither has the specified latency constraint nor mentions the energy efficiency. He et al. [5] proposed a non-deterministic geographic routing protocol SPEED that can meet the soft end-to-end latency constraint. Felemban et al. [3] proposed an improved protocol MMSPEED, which further offers reliability guarantee in addition to latency guarantee. However, neither of the protocols takes the energy efficiency into consideration. Chipara et al. [1] devised a protocol RPAR, which considered the energy efficiency by adjusts the transmission power at nodes to tradeoff the energy consumption and the end-to-end latency. Note that [5], [3], [1] all require each node maintain an information table to store the its neighbors' information including their residual energy or the link quality, significant energy overhead is thus introduced by using beacon messages to update the table. The most related work with ours is a real-time opportunistic routing protocol ORTR proposed by Kim et al. [7], which can adjust the transmission power at nodes to meet the latency constraint. However, the forwarding area of ORTR may not be a shared carrier sensing area, consequently they cannot hear each other when one of them is broadcasting. In the worst scenario where the expected transmission radius at a relay node is its maximum transmission radius, the forwarding area may not contain any nodes when a low duty cycle is applied to each node. The protocol EARTOR is then proposed to bridge those gaps mentioned above.

III. Preliminaries

A. System model

We assume that an ad hoc network consists of homogeneous nodes densely deployed in a region of interest with density ρ nodes per unit area. Each node is equipped with omni-directional antenna and GPS. We further assume that each node has M fixed transmission power levels. For each power level i, $1 \leq i \leq M$, the corresponding transmission power is p_i, and $p_i < p_j$, $1 \leq i < j \leq M$. Each node can adjust its transmission power level dynamically from one to another. For a given transmission power p_i, there is a corresponding transmission radius d_i. Clearly, $d_i < d_j$ if and only if $p_i < p_j$, $1 \leq i < j \leq M$. To reduce the energy consumption of nodes, each node has an identical duty cycle ϕ with $0 < \phi \leq 1$, and different nodes can wake up and sleep asynchronously.

Let E_T be the amounts of transmission energy consumed at a node for transmitting an L-length message, using transmission power P_T and transmission rate R. Following the power loss model, E_T can be expressed as $E_T = L \cdot \epsilon_{fs} \cdot d^{\alpha}$, where α is the power loss component with $\alpha \in [2, 4]$, ϵ_{fs} $[J/(bit/m^{\alpha})]$ is the amount of energy consumed by an amplifier to transmit one bit data at an acceptable quality level. The relationship between E_T and P_T can be expressed $E_T = P_T \cdot \tau = P_T \cdot \frac{L}{R}$, where τ is the duration to transmit an L-length message with transmission rate R, thus, we have

$$P_T = R \cdot \epsilon_{fs} \cdot d^{\alpha}. \quad (1)$$

B. Problem definition

A routing request r is represented by a quadruple $r = (s, t, L, \Gamma)$, where s and t are the source and destination nodes, L is the message length of request r, $\Gamma \in \mathbb{R}^+$ is the end-to-end latency constraint of the request.

Given a wireless ad hoc network and a sequence of routing requests, arrived one by one, without the knowledge of future request arrivals, the *on-line, real-time, capacity maximization routing problem* is to realize as many requests as possible such that the total number of requests realized is maximized, provided that the *delivery ratio* is no less than a given threshold ω, where the delivery ratio is defined as the ratio of the number of realized requests to the total number of injected requests.

IV. Routing Protocol EARTOR

A. General framework

The basic strategy behind the proposed protocol EARTOR is as follows. We adopt cross-layer design by incorporating the sleep-wake up cycle of nodes. We design a bidding mechanism for all relay candidates such that the overhead on beacon messages and the coordination latency is significantly reduced. The core part of such mechanism is a novel response time function that takes the residual energy, the location coordinates, and the relay priority of the node into consideration.

We assume that the message of a request from its source has been transmitted to the current node v_i, as shown in Fig. 1. The forwarding area of node v_i is a *reuleaux triangle* [9], in which the nodes can hear each other when anyone broadcasts a message, using its maximum transmission power.

The idea is that the forwarding area of node v_i is partitioned into a number of annular regions, each corresponds a different transmission power or a transmission radius. Assume $d_0 = 0$, let A_k be the annular region between d_{k-1} and d_k, $1 \leq k \leq M$. For each annular region A_k, there is a corresponding Expected Progress of the Annular region, denoted as $EPA(k)$, which is the expected progress in terms of the distance of the message from its current node towards the destination, if one of the nodes[1] in annular region A_k becomes the next relay.

Node v_i then calculates an Expected Message Progress from node v_i, denoted as $EMP(v_i)$, based on the knowledge of the remaining latency of the message and the Euclidean distance between node v_i and the destination node t. To meet the latency constraint of the routing request message, intuitively, the message should be forwarded to those relay

[1]For the sake of simplicity, in the following, node implies that the node is in the wake-up mode.

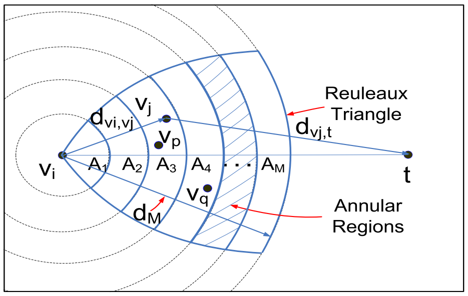

Fig. 1. The forwarding area and the partition of annular regions.

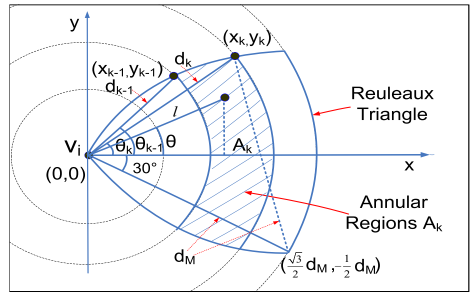

Fig. 2. The calculation of $EPA(k)$ for annular region A_k of node v_i.

candidates within the annular regions whose EPAs are no less than $EMP(v_i)$. Meanwhile, among all the relay candidates in different annular regions, the message is expected to be forwarded to the nodes in the region nearest to node v_i so as to minimize the energy consumption. Node v_i thus can adjust its power level appropriately so that the message is expected to reach its destination within the given end-to-end latency, if the transmission power of v_i is used by all nodes in the rest of the routing path. In other words, we sometimes trade the latency for energy gain, while in other cases, we use the energy to compensate the latency depending on the latency requirement of the request.

To realize this scheme, rather than performing a deterministic geographic routing, v_i uses the maximum transmission power p_M to broadcast a Beacon Request (BR) message, with endeavor to avoid the absence of nodes in the forwarding area. The BR message contains the location coordinates of v_i and the value of $EMP(v_i)$ that has been calculated by v_i in advance.

Once a node in the forwarding area receives the BR message, it starts its timer. Each node uses $EMP(v_i)$ to calculate its relay priority rank by comparing $EMP(v_i)$ to the EPA of the annular region to which the node belongs. Each node then calculates the response time using a novel local contention function. As shown in Fig. 1, nodes $v_j, v_p \in A_3$, node $v_q \in A_4$, if $EMP(v_i) > EPA(2)$ and $EMP(v_i) < EPA(3)$, the local contention function should ensure that nodes v_j and v_p in A_3 respond the faster to the relay request than that of node v_q in A_4.

An acknowledgment (ACK) message, which contains the index of the annular region to which the node belongs, is immediately broadcast by a node that has the fastest response time, using the maximum transmission power p_M. In our example, assume that node v_j has the fastest response time and broadcasts its ACK message, which contains the index of the annular region A_3. v_i hears this ACK message and knows that v_j is the next relay node, v_i then uses the corresponding transmission power p_3 to send the message to v_j. v_j will not go back to sleep until it receives and transmits the message to its next relay node or aborts the message if no relay node is found. On hearing the ACK message, all the other nodes that received the BR are suppressed.

To this end, a general framework of the proposed protocol has been revealed. What followed is how to calculate EPA of each annular region and EMP of each relay node, and how to design of the local contention function in the following.

B. Calculation of expected progress of annular region

Since the nodes in the network are homogeneous, for the annular regions with the same index of different nodes, the value of their EPAs are identical. Each node can assume it is at the origin of coordinates, a reuleaux triangle is then constructed on the coordinate plane, as shown in Fig. 2.

Thus, given a node density ρ, the duty cycle ϕ, a node v_i, and an annular region A_k, $1 \leq k \leq M$, using polar coordinates, the $EPA(k)$ can be calculated as follows.

$$EPA(k) = \frac{\iint_{S(A_k)} \rho\phi l \cos\theta \cdot ds}{\iint_{S(A_k)} \rho\phi ds} = \frac{\iint_{S(A_k)} l \cos\theta \cdot ds}{\iint_{S(A_k)} ds}, \quad (2)$$

where $S(A_k)$ is the area of annular region A_k, $ds = ldld\theta$. The numerator and the denominator of the right-hand side of Eq. (2) can be expanded as $2\Big(\int_0^{\theta_k}\int_{d_{k-1}}^{d_k} l\cos\theta ds + \int_{\theta_k}^{\theta_{k-1}}\int_{d_{k-1}}^{f(\theta)} l\cos\theta ds\Big)$, and $2\Big(\int_0^{\theta_k}\int_{d_{k-1}}^{d_k} ds + \int_{\theta_k}^{\theta_{k-1}}\int_{d_{k-1}}^{f(\theta)} ds\Big)$, respectively, where $f(\theta) = d_M(\sqrt{3}\cos\theta - \sin\theta)$, θ_k can be calculated using the following equations $x_k^2 + y_k^2 - d_k^2 = 0$, $(x_k - \frac{\sqrt{3}}{2}d_M)^2 + (y_k + \frac{d_M}{2})^2 = d_M^2$, and $\theta_k = \arctan(\frac{y_k}{x_k})$. θ_{k-1} can also be calculated, using the similar method.

C. Calculation of expected message progress

Assume that the message from its source has already been relayed to node v_i, v_i now tries to find the next relay for the message. v_i is responsible to calculate $EMP(v_i)$ and broadcast it with the BR message. We assume that $\Gamma(v_i)$ is the remaining latency of the message when it is received by v_i, which can be acquired by a piggyback on the message from the predecessor of v_i. Since $EMP(v_i)$ must meet the end-to-end latency constraint of the message in the rest of relays, we thus have

$$\Big(E(t_c(v_i)) + \frac{L + L(BR) + L(ACK)}{R}\Big) \cdot \frac{d_{v_i,t}}{EMP(v_i)} \leq \Gamma(v_i), \quad (3)$$

where $d_{v_i,t}$ is the Euclidean distance between node v_i and destination node t (can be obtained through GPS), the term

$(L + L(BR) + L(ACK))/R$ is the total transmission latency for an L-length message, a BR message, and a ACK message using transmission rate R in one hop. $E(t_c(v_i))$ is the expected coordination latency at node v_i, which is the expected duration between the moment when the BR is broadcast by v_i to the moment when the first ACK is broadcast by one of the nodes in the forwarding area, in other words, it is the expected response time to the relay request. The value of $E(t_c(v_i))$ is estimated based on the node density ρ, the duty cycle ϕ, and the expected message progress of the previous hop. In comparison with the total transmission latency of the messages at v_i, the value of $E(t_c(v_i))$ usually is very small, and we shall further analyze it later. In Inequality (3), the term $d_{v_i,t}/EMP(v_i)$ is the expected number of hops from node v_i to the destination node t if the transmission power level of the current node will be used by all nodes in the rest of the routing path. Thus, the left-hand side of the inequality is the expected end-to-end latency constraint, which should be no greater than $\Gamma(v_i)$. Thus,

$$EMP(v_i) > \frac{(E(t_c(v_i)) + (L + L(BR) + L(ACK))/R) \cdot d_{v_i,t}}{\Gamma(v_i)}. \tag{4}$$

Once $EMP(v_i)$ is calculated out, v_i broadcasts a BR message to all the nodes in the forwarding area. This BR message contains the location coordinates of v_i and the value of $EMP(v_i)$, which will be used by each node in the forwarding area to calculate its response time using the contention function.

D. The response time function

Having received a BR message from the sender, each node in the forwarding area calculates its response time to the BR message independently, based on its local knowledge including its residual energy, physical location, and relay priority rank.

As shown in Fig. 1, assume that node v_j is one of the nodes receiving the BR message in the forwarding area, it first extracts the value of $EMP(v_i)$ from the BR message and then calculates the following three factors in order to compute its response time.

1. The energy factor F_r, which is expressed as follows.

$$F_r(v_j) = \frac{e^{(1-RE(v_j)/IE(v_j))} - 1}{e - 1} \in (0, 1), \tag{5}$$

where $RE(v_j)$ is the residual energy of node v_j and $IE(v_j)$ is the initial energy of the node v_j, which is identical to all nodes. The term $1 - RE(v_j)/IE(v_j)$ is the energy utilization ratio of node v_j. This factor indicates that a node responds earlier to the relay request if its energy utilization ratio is relatively smaller, since it has plenty of residual energy.

2. The location factor F_l, which is calculated as follows.

$$F_l(v_j) = 1 - \frac{d_{v_i,t} - d_{v_j,t}}{d_M} \in (0, 1], \tag{6}$$

where $d_{v_i,t}$ and $d_{v_j,t}$ are the Euclidean distances between node v_i and destination node t and between v_j and t, respectively. This measurement favors the nodes near to the destination node respond earlier to the relay request when more than one node is in the forwarding area.

3. The relay priority factor F_p. Denote by $IAR(v_j)$ the Index of the Annular Region to which node v_j belongs. Suppose $\exists k$, $1 \le k \le M$, such that $EPA(k-1) < EMP(v_i) < EPA(k)$. If $IAR(v_j) \ge k$, the relay priority rank of node v_j is defined as

$$F_p(v_j) = IAR(v_j) - k, \tag{7}$$

otherwise, v_j will not participate in the competition for relaying the message. The rationality behind is if v_j is allowed to relay the message, the end-to-end latency constraint may not be met. Note that the rank of a node with the highest relay priority is 0. All nodes in an annular region have identical ranks, and the nodes in $A(k)$ have the highest ranks in terms of relay priority, followed by nodes in $A(k+1)$, and so on.

The response time $t(v_j)$ of node v_j can be calculated then as follows.

$$t(v_j) = \frac{t_{cmax}}{M} \cdot \left[\frac{1}{2}(F_r(v_j) + F_l(v_j)) + F_p(v_j)\right], \tag{8}$$

where t_{cmax} is the maximum coordination latency which is given in advance, M is the number of transmission power levels of a node. For the sake of convenience, let $\Delta = t_{cmax}/M$ in the rest of analysis. We have the following lemma.

Lemma 1: Let $NFR(v_i)$ be the set of wake-up nodes in the forwarding area of v_i when v_i broadcasts a BR message. The contention function defined above meets:

(i) $\forall v_j \in NFR(v_i)$, $t(v_j) \in (\Delta \cdot F_p(v_j), \Delta \cdot (F_p(v_j) + 1))$;

(ii) $\forall v_j, v_j' \in NFR(v_i)$, if $F_p(v_j) < F_p(v_j')$, $t(v_j) < t(v_j')$.

If given two nodes are in different annular regions, the one with a higher relay priority rank (nearer to the sender) will respond to the request earlier than the other one. If they are in the same annular region, which one will respond first will be jointly determined by its energy factor and location factor.

Proof: We show Claim (i) first. $\forall v_j \in NFR(v_i)$, since $F_r(v_j) \in (0, 1)$ and $F_l(v_j) \in (0, 1]$, $(1/2)(F_r(v_j) + F_l(v_j)) \in (0, 1)$, thus $t(v_j) \in [\Delta \cdot F_p(v_j), \Delta \cdot (F_p(v_j) + 1)]$, by the definition of the contention function. We then prove Claim (ii). $\forall v_j, v_j' \in NFR(v_i)$, since $t(v_j) \in (\Delta \cdot F_p(v_j), \Delta \cdot (F_p(v_j) + 1))$ and $t(v_j') \in (\Delta \cdot F_p(v_j'), \Delta \cdot (F_p(v_j') + 1))$, by Claim (i). If $F_p(v_j) < F_p(v_j')$, then v_j and v_j' are in different annular regions, $IAR(v_j') \ge IAR(v_j) + 1$, and $F_p(v_j') = IAR(v_j') - k \ge IAR(v_j) - k + 1 \ge F_p(v_j) + 1$, which implies that $\Delta \cdot F_p(v_j') \ge \Delta \cdot (F_p(v_j) + 1)$, thus $t(v_j) < t(v_j')$, Claim (ii) then follows. ■

E. Estimation of the expected coordination latency

Assume that node v_i now holds a message for relay, to calculate $EMP(v_i)$, the expected coordination latency $E(t_c(v_i))$ at v_i is to be given. While it is impossible to be known before the actual coordination happens, an approximate estimation of $E(t_c(v_i))$ based on v_i's predecessor v_{i-1} instead is proposed. Assume that through the calculation of $EMP(v_{i-1})$, the nodes in A_k within the forwarding area of node v_{i-1} have the highest relay priorities. If $\rho\phi S(A_k) \ge 1$, then at least a node in A_k responds to the request within less than Δ time units, the value of $E(t_c(v_i))$ can be approximately estimated by Δ conservatively; otherwise, identify an annular region $A_{k_{min}}$ with the smallest index $k_{min} > k$ that satisfies the following inequalities, $\rho\phi \sum_{i=k}^{k_{min}} S(A_i) \geq 1$, $\rho\phi \sum_{i=k}^{k_{min}-1} S(A_i) < 1$, and $k < k_{min} \leq M$. Thus,$E(t_c(v_i)) = \sum_{i=k}^{k_{min}} \left(Pr\{Node\ in\ A_i\} \cdot (i-k+1) \cdot \Delta\right) = \Delta \cdot \sum_{i=k}^{k_{min}} \left(\frac{\rho\phi S(A_i)(i-k+1)}{\rho\phi \sum_{j=k}^{k_{min}} S(A_j)}\right)$, where $Pr\{Node\ in\ A_i\}$ is the existence probability of a wake up node in A_i, and $\Delta \cdot (i-k+1)$ is the amount of the coordination latency if none of wake up nodes has been found in A_i, for all i with $k \leq i < k_{min}$. For a special case where node v_i is the source node s itself, there is no predecessor for s, we just use $k = 1$ to estimate the value of $E(t_c(v_i))$.

F. Collision avoidance and recovery mechanism

Contention collision occurs when two or more nodes in the same annular region respond to the request at the same time. Once the collision occurs, the sender v_i as well as the other nodes hearing them cannot receive any ACK messages, therefore, the sender and all the receivers will wait for the amount of time t_{cmax} for the next round reply. If there is still no feedback ACK message, the nodes that responded simultaneously are aware of the collision, and adjust their response times randomly to avoid any further collision in the next round. Occasionally, if the sender v_i cannot receive any ACK message after t_{cmax} time units, it recalculates a new $EMP(v_i)$ due to the further delay. We set a failure threshold for a counter, which records the number of times of failure. We then set the threshold of $EMP(v_i)$ as the maximum transmission radius d_M. If both $EMP(v_i)$ and the counter reach their thresholds, the request is aborted. If the counter reaches its threshold while $EMP(v_i)$ does not, then the recovery mechanism starts up. This probability happens when there are no nodes in the forwarding area of v_i, we thus have to incorporate face routing mechanisms [2] to route the packets or change the forwarding angle [9] to another direction to find wake-up nodes.

G. Routing protocol

Having described each component of the proposed protocol EARTOR, we now provide the basic operations of the message sender v_i and its next relay node v_j in Figs. 3, respectively, assuming that v_j is at wake up mode when v_i broadcasts its BR message. Thus, given a wireless ad hoc network and a sequence of routing requests, arrived one by one, without the knowledge of future request arrivals. There exists a routing protocol EARTOR for all the requests such that the number of requests realized is maximized, while the latency of each realized request is met.

Sender(v_i)
1. Set a Failure Counter (FC) with a threshold Max_F;
2. Estimate the expected coordination latency $E(t_c(v_i))$;
3. Calculate the expected message progress $EMP(v_i)$;
4. Broadcast a BR message to the forwarding area of v_i, and set a Timer with expired time t_{cmax};
5. **if** Timer expires **then**
  FC=FC+1;
  if FC $< Max\_F$ **then**
    Update the remaining latency $\Gamma(v_i)$ and GoTo Step 2;
  else
    Abort the request;
else
  Receive an ACK message;
6. Extract the index k of the annular region to which the node that broadcast the ACK message belongs, use the corresponding transmission power p_k to transmit the message to the node that broadcast the ACK message;

Receiver(v_j)
1. Receive a BR message from node v_i;
2. Extract $EMP(v_i)$ from BR, figure out the annular region to which v_j belongs and calculate the EPA of the region, calculate $F_p(v_j)$ by comparing $EMP(v_i)$ to EPA of this region;
3. Calculate the response time $t(v_j)$ using the contention function, where $t(v_j) = \Delta \cdot [(1/2)(F_r(v_j) + F_l(v_j)) + F_p(v_j)]$, a Response Timer then starts;
4. **if** an ACK message overheard **then**
  Cancel Response Timer;
else
  Broadcast an ACK message to claim itself to be the next relay node;

Fig. 3. Protocol EARTOR

V. Performance Evaluation

A. Simulation environment

We use the MATLAB to generate a random node distribution in a 10×10 unit square region. We assume that nodes in the region are densely deployed with node density ρ varies from 8 to 13 nodes per unit area. We further assume that each node has a maximum of M=10 different power levels that can be adjusted dynamically. Without loss of generality, we set the transmission radius of each power level from 0.2 to 2 units, with the increment of 0.2 units. We then unify the power consumption, following Eq. (1), where $\alpha = 2$, the power consumption of a node corresponding to the minimum transmission radius is 1 power unit, and the power consumption under different transmission radii is i^2 power units, where i is the index of the power level, $1 \leq i \leq M$. Furthermore we assume that the reception power and the idling power consumptions at each node are identical, which are both 0.5 power units. The initial energy capacity of each node is set to 5×10^6 energy units, where an energy unit is defined as a unit power consumption per second. We assume that the duty-cycle mechanism is adopted to save energy, which means that each node is available as a relay node with probability ϕ, $\phi \in (0, 1]$. The data transmission rate is $R = 20kbps$. The t_{cmax} for the coordination latency is set to $50ms$. The message length of the beacon request (BR) message and acknowledgment (ACK) message are 100 bits. For each routing request

$r = (s, t, L, \Gamma)$, unless otherwise specified, we assume that the message length L is drawn randomly from $1K$ bits to $10K$ bits. To simulate a reasonable latency constraint Γ for each request, we use $\Gamma = \frac{\lceil d(s,t)/d_M \rceil (\tau_c + ((L + L(BR) + L(ACK))/R))}{\eta}$, to generate the latency constraint for each request, where the latency constraint coefficient $\eta \in (0, 1]$ is used to adjust the degree of tightness of the latency constraint Γ. η is randomly drawn from 0.1 to 1.0 with increment of 0.1. The higher the value of η, the tighter the latency constraint will be. $d(s, t)$ is the Euclidean distance between s and t, which can be obtained through GPS. d_M is the maximum transmission radius when a node works at its maximum power level, which is 2 units. τ_c is the expected coordination latency of one hop, which is set $t_{cmax}/2$. L, $L(BR)$, and $L(ACK)$ are lengths of message of the routing request r, BR, and ACK, accordingly. The term $(L + L(BR) + L(ACK))/R$ is the one-hop transmission latency of the messages themselves.

Given a node density ρ, 20 network instances are generated and for each network instance, 10 different request sequences with each sequence containing 10^5 requests are randomly injected into the network. The value shown in all figures is the mean of 200 values.

B. Performance evaluation of various protocols

To evaluate the performance of the proposed protocol EARTOR, we compare it against the state-of-the-art protocols ORTR [7] and GeRaF [10]. We set $\rho = 10$ and $\phi = 0.5$, and monitor the dynamic delivery ratio as routing requests arrive one by one. We set the delivery ratio threshold to $\omega = 70\%$ and do not terminate the subsequent requests until the delivery ratio is below the threshold ω. As shown in Fig. 4, before the total number of injected routing requests reaches 2×10^4, the delivery ratio of each protocol is roughly identical, which achieves 100%, then the delivery ratio of protocol GeRaF deteriorates sharply, followed by protocols ORTR and EARTOR. Protocol EARTOR shows a much better performance in terms of network capacity as time goes by. This implies that protocol EARTOR is the best among the three protocols in terms of energy conservation. Fig. 5 further illustrates that by plotting the node distribution by their remaining energy, when the number of injected messages reaches 2×10^4. Protocol EARTOR saves energy through taking advantage of its energy-latency trade-off mechanism, while protocol ORTR focuses on the balance of the energy consumption where the energy consumption histogram is nearly uniformly distributed. Protocol GeRaF is a greedy routing protocol that does not consider the energy efficiency, as some nodes have already depleted their energy.

VI. Conclusions

In this paper we have studied an on-line, real-time capacity maximization problem in routing by proposing a novel energy-aware, real-time, opportunistic routing protocol EARTOR. We also conducted extensive experiments by simulations to evaluate the performance of the proposed protocol against existing ones, in terms of the dynamic delivery ratio and the network capacity. The experimental results showed that the proposed protocol EARTOR outperforms existing the state-of-the-art protocols including ORTR and GeRaF significantly.

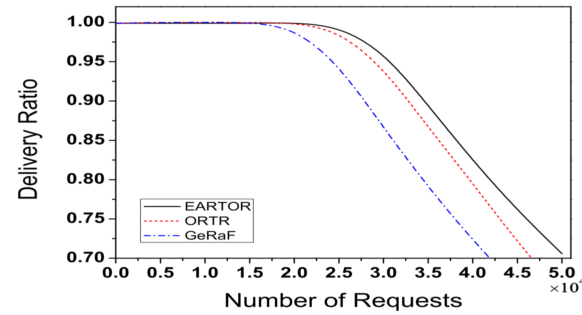

Fig. 4. The delivery ratio delivered by routing protocols EARTOR, ORTR, and GeRaF respectively, where $\rho = 10$, $\phi = 0.5$, $L \in [1K, 10K]$, $\eta \in (0, 1]$.

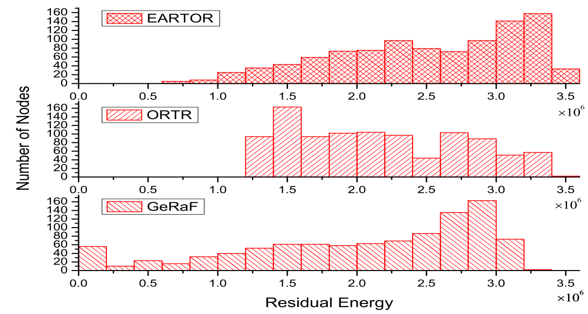

Fig. 5. The node distribution by the residual energy for routing protocols EARTOR, ORTR, and GeRaF respectively, where the total number of injected messages is 2×10^4, $\rho = 10$, $\phi = 0.5$, $L \in [1K, 10K]$, $\eta \in (0, 1]$, and $\omega = 70\%$.

References

[1] O. Chipara, Z. He, G. Xing, Q. Chen, X. Wang, C. Lu, J. Stankovic, and T. Abdelzaher. Real-time power aware routing in sensor networks. *Proc. IWQoS'06*, IEEE, pp.83-92, 2006.

[2] Q. Fang, J. Gao, and L. Guibas. Locating and bypassing routing holes in sensor networks. *Proc. INFOCOM'04*, IEEE, pp.2458-2468, 2004.

[3] E. Felemban, C. G. Lee, and E. Ekici. MMSPEED: multipath multi-speed protocol for QoS guarantee of reliability and timeless in wireless sensor networks. *IEEE Trans. Mobile Computing*, Vol.5, No.6, pp.738-754, 2006.

[4] Y. Gu and T. He. Data forwarding in extremely low duty cycle sensor networks with unreliable communication links. *Proc. SenSys'07*, ACM, pp.321-334, 2007.

[5] T. He, J. A. Stankovic, C. Lu, and T. Abdelzaher. SPEED: A stateless protocol for real-time communication in sensor networks. *Proc. ICDCS'03*, IEEE, pp.46-55, 2003.

[6] J. Heo, J. Hong, and Y. Cho. EARQ: energy aware routing for real-time and reliable communication in wireless industrial sensor networks. *IEEE Trans. Industrial Informatics*, Vol.5, No.1, pp.3-11, 2009.

[7] J. Kim and B. Ravindran. Opportunistic real-time routing in multi-hop wireless sensor networks. *Proc. Symposium on Applied Computing (SAC'09)*, ACM, pp.2197-2201, 2009.

[8] F. Wang and J. Liu. Duty-cycle-aware broadcast in wireless sensor networks. *Proc. INFOCOM'09*, IEEE, 2009.

[9] M. Witt and V. Turau. BGR: blind geographic routing for sensor networks. *Proc. 3rd International Workshop on Intelligent Solutions in Embedded Systems*, IEEE, pp.51-61, 2005.

[10] M. Zorzi and R. R. Rao. Geographic random forwarding (GeRaF) for ad hoc and sensor networks: energy and latency performance. *IEEE Trans. Mobile Computing*, Vol.2, No.4, pp.349-365, 2003.

[11] M. Zorzi and R. R. Rao. Geographic random forwarding (GeRaF) for ad hoc and sensor networks: multihop performance. *IEEE Trans. Mobile Computing*, Vol.2, No.4, pp.337-348, 2003.