A Pheromone-Aided Multipath QoS Routing Protocol and its Applications in MANETs

Paul Barom Jeon and George Kesidis

Abstract—In this paper, we present an ant-based multipath QoS routing protocol that utilizes a single link metric combining multiple weighted criteria. The metric is applied to the proposed energy efficient multipath algorithm that considers both energy and latency. Energy efficiency is an important issue in mobile ad hoc networks (MANETs) since node energy supplies are stored in batteries. In order to increase the network lifetime it is important to maximize the minimum node energy along a path. As the network topology changes, failures may occur on active routes, resulting in the need for new route discoveries if only single routes per flow are maintained. Frequent new route discovery would, however, increase routing overhead and increase mean and peak packet latency. Using multiple routes simultaneously per flow can be a solution to these problems. Also, a special case of the multipath QoS routing protocol that considers throughput is applied to a security context. A compromised node can obstruct network communication by simply dropping packets that are supposed to be forwarded. In our approach, messages are distributed over multiple paths between source and destination using ant-based QoS routing. In proportion to the throughput of each path, a pheromone-aided routing table is updated and, subsequently, paths that contain malicious nodes are naturally avoided.


I. INTRODUCTION

Rapid changes in the topology of mobile ad hoc networks (MANETs) lead to staleness of route information that, in turn, may result in poor routing decisions and a squandering of the overhead associated with routing. In the absence of a viable route, route discovery latency would have a direct impact on packet and flow-level quality-of-service (QoS). Such QoS degradation may not result if multiple routes from the source to the destination are simultaneously maintained, i.e., the network could switch immediately to a “backup” route when an active one fails. This will augment the reliability of data transmission. Multipath routing is also used when load balancing is required. Load balancing is an important field of MANETs because of limited bandwidth and energy restrictions. The sender can utilize multiple paths to satisfy the required bandwidth collectively. To increase the network lifetime, data packets are sent through different multiple paths.

This paper focuses on multipath routing protocols that are both timely and energy-efficient for nodes of MANETs accommodating multiple traffic classes. MANETs are often required to be able to operate under “volatile” conditions. Examples of sources of network volatility include finite energy supply (batteries), communication traffic, node mobility, changing environmental conditions affecting the channel and terrain, and enemy activity targeting the network itself. Consequently, there are large existing literatures on single path routing for MANETs, e.g., [26], [42], [43]. In addition, several multipath routing protocols [32], [35], [64] have been proposed to overcome the routing overhead problem of single path routing protocols in a volatile networking context.

For packets that are not delay-sensitive, a routing protocol may choose paths that are rich in energy. The “bottleneck” energy of a path is roughly defined to be the minimum energy supply among its associated nodes. So, the routing protocol could seek to determine the maximum bottleneck (“max-min”) energy path to a given destination. Again, this could be done using either a link-state or distance-vector approach. “Energy-aware” routing protocols, taking into account node energy in their link metrics, were proposed previously in, e.g., [9], [50], [63]. In [15], [46], [54], routes with the highest amount of residual energy are chosen to maximize the network’s “lifetime” under the basic assumption, of course, that node energy supplies are stored in batteries that cannot be easily, or can never be, replenished. Clearly, energy issues should be considered for latency-critical packets as well as it would be unwise to send them on paths with any component nodes that do not have sufficient energy to transmit them.

In this paper, we propose multipath QoS routing algorithms that jointly manage both delay and energy concerns. The principle challenges of such protocols is to be able to route in a highly volatile topology with minimum overhead. We will review the relatively small existing literatures on dual-purpose routing protocols [16], [29], [38].

Multipath routing [6] gives “path diversity” so that communication is less vulnerable to a localized attack, e.g., a relaying node that is either passively eavesdropping or maliciously dropping/modifying data or control packets. Also, judicious replication of important (high-demand) data together with multipath anycasting can make a content distribution network robust against DoS attacks targeting individual peers. Finally, security issues have motivated the use of routing proxies or indirect addressing [55] that may naturally lead to robust
multipath routing.

Since most of the conventional routing protocols in ad hoc wireless networks assume nodes are trustworthy and cooperative, even a small number of compromised nodes can obstruct network communication by discreetly dropping packets [27]. Several schemes [37], [44] have been proposed to detect malicious packet dropping under connection-oriented transport layer protocols, such as TCP. However, TCP is well known not to perform well in ad hoc wireless networks for various reasons such as misinterpretation of packet loss, frequent packet breaks, effects of path length, misinterpretation of congestion window, etc. [39]. Just et al. [27] proposed a probing technique that can detect malicious packet dropping under both connection-oriented and connectionless transport layer protocols.

In the security application of the proposed multipath QoS routing, we focus on avoiding malicious packet dropping, a class of DoS attack. If a node is already compromised at the routing phase, it is assumed to cooperate with other nodes to build legitimate routes between flow sources and destinations. Therefore, a source will not know if the path is insecure at first. At the data transmitting phase, the malicious node drops/modifies packets discreetly to evade detection. If the multiple paths are assumed to be node-disjoint [34], then a system of packet broadcast (replication) on all paths will be robust against such nodes, but this will, of course, generate a significantly large additional network load. So, we propose an ant-based multipath routing protocol that can adaptively avoid malicious packet dropping without additional packet duplication. Once a certain number, N, of packets are delivered through a given path, the destination generates a backward ant (BANT) [18] packet and sends it back along that path to source. After receiving the BANT packet, the source node updates its related pheromone-aided routing table. With the aid of pheromone mechanisms, the path that contains a malicious node is naturally avoided in a computationally and bandwidth efficient fashion.

This paper is organized as follows. In Section II, we overview a number of previous protocols that are related to our work. In section III, routing overhead under various multiple node-disjoint paths algorithms is discussed. We propose in section IV a more adaptive routing protocol to accommodate both energy and delay metrics under volatile conditions. In section V, we provide a simulation study of our proposed multipath QoS routing protocol. In section VI, the proposed multipath QoS routing protocol is applied to a security context in order to adaptively avoid malicious packet dropping. Finally, conclusions are drawn in Section VII.

II. RELATED WORK

A. Multipath Routing

Prominent routing protocols for wireless MANETs, such as AODV [43] and DSR [26], usually develop a single path between source and destination for a given flow. When the developed route fails, the sender would need to discover a new route to the receiver. This new route discovery will result in additional packet delay and network overhead. If multiple paths were developed in the first stage, when the first active path (possibly representing the shortest delay path) fails, a new path to the receiver can be immediately used.

Split Multipath Routing (SMR) [32] protocol is an extension of DSR, inheriting the on-demand, source routing characteristics of DSR. SMR builds maximally disjoint paths. SMR discovers two paths that share as few nodes and links as possible. One of the two routes is the shortest delay path. Maximally disjoint paths are preferred to prevent congestion on shared nodes/links, to accomplish efficient utilization of the network resources, and for fault tolerance. Initially, the source floods ROUTE REQUEST (RREQ) packets to the entire network to discover routes to the destination. Unlike DSR, the intermediate nodes do not maintain a route cache and, consequently, do not respond to the RREQs. Not every duplicate RREQs are discarded at intermediate nodes. Intermediate nodes forward the duplicate RREQ if its incoming link is different and its hopcount is not larger than that of previous received RREQs. As a result of flooding, several RREQs which traveled through different paths arrive at the destination. The shortest delay route is determined when the first RREQ arrives at the destination. After sending ROUTE REPLY (RREP) back to the source, the destination waits for a certain amount of time to receive more RREQs and learn more possible routes to the source. Among the possible paths, it selects two mostly disjoint paths. On discovering the first route, it is used to send data packets to the destination. When more than one route is discovered, data traffic is distributed into multiple paths using a per-packet allocation scheme.

AOMDV [35] is a multipath version of the well-known on-demand single path routing protocol AODV. It discovers loop-free and link-disjoint multiple paths between source and destination. By flooding the RREQs from the source to the destination multiple reverse paths are discovered both at the intermediate nodes and the destination node. Multiple RREP packets are sent from the destination to the source. The RREP packets travel back to the source through the previously established multiple reverse paths and form multiple forward paths to the destination at the source and intermediate nodes. The AOMDV protocol has two essential elements: establishing and maintaining multiple loop-free paths and finding link-disjoint paths. To establish loop-free paths, each node i maintains an “advertised hopcount” defined as the “maximum” hopcount of the multiple paths, for a destination d. Each node will accept an alternate route to the destination only if the hopcount of the alternate route is smaller than the advertised hopcount. Since the maximum hopcount is considered, the advertised hopcount will remain unchanged for the same sequence number. The advertised hopcount will be reset when node i receives a route advertisement for destination d with a greater sequence number. In order to discover multiple link-disjoint paths, each RREQ contains a “firsthop” field, defined as the neighbor of the source which RREQ has traveled at the first hop. In addition, “firsthop_list” for each RREQ is kept at each node. An intermediate node does not discard duplicate copies of RREQ immediately. On receiving a duplicate RREQ, the intermediate node checks whether it can form a new node-disjoint path to the source using “firsthop” and “firsthop_list”. 
Unlike intermediate nodes, the destination sends back several RREPs to the source, through unique neighbor of destination, regardless of the first-hop of the RREQ. Link-disjointness is assured at the first hop of the RREP by using different neighbors. After the first hop, the RREP follows the already established node-disjoint and thus link-disjoint reverse paths to the source.

AODVM [64] is also an extension of AODV. It discovers multiple node-disjoint paths. Instead of discarding all the duplicate RREPs, intermediate nodes keep information of each RREQ in the so called RREQ table. Also, intermediate nodes are not permitted to send RREPs directly to the source. Whenever the destination receives a RREQ it updates the sequence number and sends back a RREP to the source through the path which RREQ has traveled. When the RREP arrives at an intermediate node, a forward path from the intermediate node to the destination node is set up. Subsequently, the intermediate node selects a shortest path to the source from its RREQ table and sends the RREP via the path. In order to ensure the node-disjointness when nodes overhear any neighboring node sending a RREP packet, they delete the transmitting neighbor node from its RREQ tables.

Under multipath routing, a basic question is how to chose a path when multiple paths are available at a certain node to a destination. AOMDV simply selects the path that was discovered first, and SMR uses a per-packet allocation scheme. When there are more than one available path to a destination, SMR allocates these multiple paths in turn to a packet to be delivered to the destination. These simple schemes, however, will not fit in well with a multipriority paradigm, since the different priorities cannot be taken into consideration when selecting a path. In the following, we consider multipath and multipriority issues jointly when discovering routes to the destination. Instead of developing an optimal path for each priority, a suboptimal path is chosen among the multiple paths considering the priority. The proposed new method has low overhead since it does not execute multiple routing algorithms concurrently.

C. Ant Routing

In [12], the author proposed to maximize network lifetime by traffic flow augmentation and flow redirection to balance the energy consumption rates among the nodes in proportion to their energy reserves. In particular, they showed that delay-constrained routing was NP-complete. However in [13], the author proposed a heuristic algorithm to first reduce the NP-complete problem to a simpler one that could be solved in polynomial time, but the proposed solution was mainly for QoS routing in a multimedia network and energy issues could not be treated as a QoS parameter. Moreover, the time and computational complexity of the algorithms might not be ideal for ad hoc networks.

Taking inspiration from this, Feng and Doukleris [16] proposed a single weighted link metric combining both delay and cost considerations. However, this algorithm requires that the costs to be additive. For the problem at hand, delay is a cumulative parameter but the minimum energy resources should not be added along a route. Also, there is an assumption that both of the costs are of comparable magnitudes, which raises an issue of appropriate choice of dimensions.

A possible solution to route optimization while dealing with additive and non-additive metrics was proposed in [29]. However, this approach combines the two attributes to make routing decisions. In this way we lose control of the compromise with the energy for better delay performance. In other words, a route with lower cumulative delay may occasionally be selected even though the route is energy-poor. As a result of this, a node may at times only find paths with large delays as the network tries to preserve energy-poor paths.

Note that two independent routing algorithms running in parallel, one for each link metric, have the disadvantage of requiring twice the overhead as a single algorithm addressing both metrics. One could envision a synchronization of the two protocols so that a single packet would be required to update both metrics; however, both metrics would not necessarily require updates at the same times nor with the same frequency and, therefore, a single-packet update for both metrics may not result in substantial overhead savings. The routing information could also be piggybacked on new/additional data packet headers, but this has the disadvantage of reducing general transmission efficiency (header-to-packet ratio) of the network and may add latency to the routing algorithm as one would want to wait for a priority packet to piggyback. Typically, low-volume routing messaging would be treated as priority communication.

1 A set of neighboring nodes of node i is denoted by $S_i$. 
amount of pheromone at $k$ that is jointly associated with $d$ and $j$.

Initially, all pheromone levels could be equal (or, in the absence of pheromone to a destination, a purely random forwarding decision is made) but they will vary with time as described in the following. A FANT is converted to a backward ant (BANT) when it reaches $d$. The BANT follows the same path in the opposite direction as that of its corresponding forward ant. When the backward ant is transmitted to node $k$ by node $f$ ($k \in S_f$), the forwarding probability $P_{jd}$ at $k$ is increased and $P_{dn}$ is decreased for all $n \in S_k, n \neq f$. Once the backward ant returns to its source, $s$, the (one-way) trip time experienced by the FANT is measured and the pheromone levels associated with $d$ at $s$ are adjusted accordingly.

The Ant-Colony-Based Routing Algorithm (ARA) [18], suitable for MANETs, is based both on “swarm intelligence” and ant-colony meta-heuristics. ARA consists of three phases: route discovery, route maintenance, and route failure handling. In the route discovery phase, new routes between nodes are discovered with the use of a forward-and-backward ants, similar to AntNet. Routes are maintained by subsequent data packets, i.e., as the data traverse the network, node pheromone values are modified so that their paths are “reinforced.” Also, as in nature, pheromone values decay with time in the absence of such reinforcement. Routing (link) failures, usually caused by node mobility, are detected through missing acknowledgements. When a node detects a routing error, the pheromone value associated with the “missing link” is set to 0.

Termite [45] is another ant-based routing algorithm that is similar to ARA. However, unlike the ARA, pheromone is not considered in the route discovery phase. Instead of the forward and backward ants, route request (RREQ) and route reply (RREP) control packets are used to discover the routes. The RREQ packet randomly walks, not floods, through the network to discover a route to the destination. Pheromone levels are used for routing data packets and proactive “seed” packets are introduced for route maintenance.

Both energy and delay issues were considered in [38]. Only delay quantities, however, are considered when computing the pheromone values and forwarding probabilities. The dissipated energy of a node after each ant passes through it is calculated by

$$\Delta E_{ij} = \frac{K}{(D_{ij})^2}$$

(1)

where $K$ is the amount of energy to transmit the ant over a single unit distance, and $D_{ij}$ is the Euclidean distance between node $i$ and $j$ [36]. The residual node energy at time $t$ is computed by

$$E_i(t) = E_i(t-1) - \sum_j \Delta E_{ij}.$$  

(2)

When a node’s energy level simply drops below a pre-specified threshold value, the node is removed from the sensor network and alternative routes are found.

D. Secure Routing

In ad hoc wireless networks, intrusions can be classified into two categories: passive and active attacks. A passive intrusion does not disrupt the functioning of the network; instead, the attacker eavesdrops on the traffic flowing across the network to discover valuable information without modifying the data. Since a passive attack does not affect the functioning of the network, it is very difficult to detect. Encryption schemes are usually used to protect the data from an intruder/attacker. Unlike a passive attack, an active attack modifies or drops messages thereby obstructing the functioning of the network. Messages include both routing control packets and data packets. An adversary can attack routing packets resulting in an inefficient routing table at the source. On the other hand, an adversary can attack data packets resulting in incomplete transmission, even though it cooperates with other nodes to build legitimate routes between sources and destinations. Examples of active attacks and proposed solutions are Wormhole attacks (e.g., Packet Leashes [21]), Blackhole attacks (e.g., [14]), Byzantine attacks (e.g., [5]), and routing attacks (e.g., [22], SEAD [19], ARAN [48], ARIADNE [20]).

In a security context, multipath routing has several advantages including the avoidance of passive attacks, the avoidance of malicious packet dropping, and the robustness against focused DoS attacks. The avoidance of passive attacks (e.g., eavesdropping) can be achieved by simply dispersing the traffic over the existing multiple paths between source and destination so an attacker can only successfully intercept a portion of the transmitted message. In addition to this basic idea, various schemes have been added in order to increase confidentiality [7], [8], [41]. If there exist $k$ paths between source and destination and $n$ compromised intermediate nodes, although an attacker successfully drops part of the transmitted message, the original message can be safely delivered to the destination (unless $n$ is greater than $k$ and at least one node in each path are compromised). Note the tactic of multipath routing can be used in concert with intrusion detection systems [27], [37], [44] that are designed to detect intruders and then respond by, e.g., avoiding paths in which they reside. If an intruder attempts to substitute fraudulent packets, one employ a traceback mechanism to identify the intruders, e.g., trustworthy nodes could log packets at various points throughout the network and then use some extraction (“data mining”) techniques to find the path fraudulent packets traversed, see [47]. Snoeren et al. [51], [52] proposed a modification to this approach, called the Source Path Isolation Engine (SPIE), that hashes and stores only the first 28 bytes of a packet, thereby avoiding 99.9% of all collisions (false positives due to many-to-one mapping of the hash) while saving tremendous amounts of storage space. Alternatively, one could mark packets for traceback [53]. Finally, robustness against focused DoS attacks can also be achieved by transmitting replicated packets over multiple paths [30]. Although some portion of the intermediate nodes may be attacked and rendered unable to forward messages, with the help of redundant messages, the original message can be safely delivered to the destination.

III. ROUTING OVERHEAD ANALYSIS

In an on-demand protocol, when a node needs to discover a route to a destination, it broadcasts route request (RREQ)
packets into the network. In order to reduce the routing overhead, each intermediate node typically forwards only one RREQ originating from the same source node. Most of the existing on-demand single path routing protocols, such as AODV [43], DSR [26], and LAR [28], only forward the first arrived RREQ. Consider the simple network shown in Figure 1. Suppose that a RREQ that followed the path $S \rightarrow a \rightarrow b \rightarrow e \rightarrow f$ arrived first at node $f$. When another RREQ that followed the path $S \rightarrow i \rightarrow j \rightarrow k \rightarrow f$ arrives at node $f$, it will be dropped since node $f$ already has seen a RREQ that originated from the same source $S$. Such a RREQ forwarding method limits the possibility of finding feasible multiple routes, e.g., only the route $S \rightarrow a \rightarrow b \rightarrow c \rightarrow d \rightarrow D$ is discovered although there exists another feasible route $S \rightarrow i \rightarrow j \rightarrow k \rightarrow f \rightarrow g \rightarrow h \rightarrow D$.

![Fig. 1. Example network topology](image)

Unlike single path routing protocols, multipath routing protocols do not always forward only one RREQ. In Split Multipath Routing (SMR) [32], not all duplicate RREQs are discarded at intermediate nodes. Intermediate nodes forward the duplicate RREQ if its incoming link is different and its hopcount is not larger than that of previous received RREQs. Therefore, node $f$ in Figure 1 forwards both RREQs from node $k$ and $e$. In SMR, the destination node $D$ discovers two node-disjoint paths from the three received RREQs.

In AOMDV [35], an intermediate node does not discard duplicate copies of a RREQ immediately. On receiving a duplicate RREQ, the intermediate node checks whether it can form a new node-disjoint path to the source using firsthop, firsthop_list, and advertised_hopcount, see [35] for detail definitions. On receiving a route reply (RREP) packet sent from the destination, an intermediate node selects one reverse path from the early stored reverse paths. Similar to AODV, an intermediate node only forwards the first arriving RREP in AOMDV. For the example network shown in Figure 1, node $f$ does not set up a reverse path to node $k$ although it receives a RREQ via node $k$ since advertised_hopcount of $f$ is not greater than advertised_hopcount of $k$. Therefore, only one path $S \rightarrow a \rightarrow b \rightarrow e \rightarrow c \rightarrow d \rightarrow D$ is found.

In AODVM [64], instead of discarding all the duplicate RREQs, intermediate nodes keep information of each RREQ in the so called RREQ table. When the RREP arrives at an intermediate node, a forward path from the intermediate node to the destination node is set up. Subsequently, the intermediate node selects a shortest path to the source from its RREQ table and sends the RREP via the path. In order to ensure the node-disjointness when nodes overhear any neighboring node sending a RREP packet, they delete the transmitting neighboring node from its RREQ tables. Similar to AOMDV, AODVM cannot discover two node-disjoint paths for the example network shown in Figure 1. At node $f$, the RREP is forwarded to node $e$ since two paths stored in RREQ table of node $f$ have same number of hops to the source and the RREQ via node $e$ arrived first at node $f$. Just as in AOMDV, only one path $S \rightarrow a \rightarrow b \rightarrow c \rightarrow d \rightarrow D$ is found.

In [60], a multipath routing algorithm based on selective broadcast (SB) method was proposed. When an intermediate node receives a RREQ, it caches the RREQ and rebroadcasts the RREQ only when the traveled path is node-disjoint with the paths included in previously cached same RREQs. Similar to SMR, node $f$ in Figure 1 rebroadcasts both RREQs from node $k$ and $e$. Just as in SMR, the destination node $D$ discovers two node-disjoint paths from the three received RREQs.

The route discovery result of different routing algorithms for the example network topology is summarized in Table I.

<table>
<thead>
<tr>
<th>Number of node-disjoint Paths</th>
<th>SMR</th>
<th>AOMDV</th>
<th>AODVM</th>
<th>SB</th>
</tr>
</thead>
<tbody>
<tr>
<td>Received RREQs at $D$</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>3</td>
</tr>
</tbody>
</table>

The result clearly shows that more routing overhead is required to discover more node-disjoint paths. However, the increased RREQs, which are broadcasted at each intermediate node, can degrade the performance of the routing protocol [57], [59]. When an active path gets broken and a source node has no cached alternative path, the source node broadcasts RREQ to discover a route to a destination. However, in a congested network, these broadcasted RREQs are more likely to collide with data packets of other connections.

In order to reduce the routing overhead of the selective broadcast method, a heuristic redirection method was proposed in [61]. Our proposed multipath routing algorithm adopts heuristic redirection in order to discover multiple paths with less overhead. Details of our route discovery method is described in Section IV-A.

Figure 2 shows the average ratio of the number of discovered node-disjoint paths with 95% confidence interval using two multipath algorithms to the number of hops of the shortest path. The sub-ideal number of node-disjoint paths are discovered using a method similar to that of [64]. The source first discovers a path to the destination using AODV [43]. The nodes on the first found path are then eliminated and AODV is executed again to find another path. These steps are repeated until no node-disjoint paths are found between the source and the destination. Details of our simulation environments are described in Section V.
IV. PPRA: PRIORITIZED PHEROMONE AIDED ROUTING ALGORITHM

Rapid changes in network topology leads to staleness of routes resulting in poor routing decisions and a squandering of the overhead associated with routing. This is especially costly when energy supplies are dwindling or successful communication requires heightened transmission energy due to increased channel noise. For such a volatile networking context, a highly distributed, quickly responsive (adaptive) and bandwidth-efficient multipath routing algorithm is required.

A. Route Discovery

Conventional ant routing algorithms require significant overhead for preliminary route set-up [17] possibly resulting in slow route convergence which is not satisfactory for dynamic ad hoc networks. A convergence result for pheromone load balancing, resolving the optimal path(s) among pairs of nodes, was given by us in [24]. Formal convergence results for ant routing were first announced in [31], in particular Theorem 2 of [31] for “regular” ants (in the terminology of [56]). In order to overcome the slow route convergence problem with less routing overhead, the route discovery characteristic of our scheme is inherited from the heuristic redirection method [62].

When a traffic source does not have an active route to a destination, the source initiates route discovery by broadcasting a forward ant (FANT) packet. The source address is included in the FANT, and as it propagates, the addresses of intermediate nodes that it visits are appended prior to forwarding. Unlike other single path routing algorithms, in order to discover multiple paths, intermediate nodes do not discard duplicate FANTs, instead the latter received FANTs are cached at the intermediate nodes. When a FANT reaches the destination node, the destination generates a backward ant (BANT) packet for the source node. Since the destination has a route to the source contained in the FANT it received, it does not flood the BANT but forwards it to one of the neighboring nodes if the path contained in the received FANT is node-disjoint with all cached paths of previously received FANTs. The intermediate nodes heuristically redirect the received BANTs originated from the destination node in order to increase the chance of finding a new node-disjoint path. More detail of the heuristic redirection method is well described in [62].

B. Energy Efficient Source Routing

One of the main problem of source routing is that the packet length is proportional to the path length. Therefore, the energy consumption also increases proportionally. In order to overcome this problem, an energy efficient method is added to [25] as following.

At the source and intermediate nodes, PPRA checks if the selected path has been used before. If the path has not been used before, full route information is added to the data packet similar to the conventional source routing protocols. Before forwarding the data packet, PPRA builds a routing table using the destination and next hop information extracted from the route information added to the data packet. When the first data packet reaches the destination node, proper routing tables are built at the corresponding source and all intermediate nodes. When the next data packet is to be sent, PPRA notices that the selected path is used before and does not add any additional route information to the data packet. Instead, PPRA utilizes the routing table built upon first use of the path.

By using such a method, the additional route information does not have to be always added to the data packet. This will reduce the total packet size resulting in a less energy consumption.

When a node is responsible for forwarding multiple traffic flows, energy consumption of the node increases proportionally. In order to overcome this problem, a path with no node responsible for multiple traffic flows is preferred. At the route discovery phase, the number of traffic flows that utilize the node is collected and sent to the source. The number of traffic flows of each node is examined and the maximum is cached as $\text{max}_{\text{node used}}$. The $\text{max}_{\text{node used}}$ is used when selecting a path to be cached. If two paths have same path length, the path with smaller $\text{max}_{\text{node used}}$ is preferred. Note that maintaining more than two alternate paths usually does not affect the performance considerably [40]. For this reason, we maintain only two multiple paths.

C. Route Maintenance

The source nodes maintain a routing table that contains entries of neighboring nodes to reach destination nodes. For example, in Figure 3, if node Z forwards a BANT from node W to node X, node X creates a pheromone state entry for reaching node W through node Z. An example routing table is shown in Table II. When the source receives the BANT, it has an entry for reaching the destination through node X. Since duplicate FANTs are not discarded, the destination node may send multiple BANTs back to the source. Note that we specifically consider node-disjoint paths for simulation simplicity [62].

FANTs are periodically sent from the source to the destination, after sending every $N$ data packets along the selected

![Fig. 2. Ratio of the number of discovered node-disjoint paths to the number of hops of the shortest path.](image-url)
D. Routing using Energy and Delay Metrics

The ant routing framework has two types of feedback: positive feedback increases the pheromone levels on routes actively carrying ant packets and negative feedback periodically decreases pheromone values to limit the effects of stale information. Routing decisions tend to favor paths with higher pheromone levels and, when allowed to converge, shortest end-to-end paths are empirically observed to be favored [18]. In the following, we propose a modified ant mechanism algorithm that uses energy and delay metrics to perform updates of pheromone levels. Assuming a control packet containing both energy and delay information, a separate pheromone level will be maintained for each traffic type.

The time-to-live (TTL) field for anti-like routing provides explicit distance (hopcount) information to a packet’s source. TTL (hopcount) information is widely used in existing routing mechanisms, see, e.g., [43]. We will exploit a “TTL pheromone” based on TTL data to improve the performance of an ant-like routing algorithm.

Two routing algorithms are developed. In the first algorithm, ant packet headers are assumed to have two fields used for routing: one to indicate bottleneck residual energy of a path and the other being TTL. This data will be used to maintain both bottleneck residual energy and TTL pheromone levels in the nodes. More specifically, if the residual energy of a forwarding node is lower than the bottleneck residual energy inscribed in the packet, the node overwrites the packet’s energy field with its own residual energy level. Latency-critical traffic will be routed based on both pheromones whereas latency-non-critical packets will use only energy pheromone levels. Note that we continue to assume that latency-critical and latency-non-critical packets are separately queued at each node.

In the second algorithm, packet headers have fields that:

- track the minimum residual energy of the nodes that relay them (as in the first algorithm) and
- track the cumulative delay based on backlog information of queued packets destined to the packet’s source.

So, when a packet reaches its destination, it contains the minimum residual energy and the cumulative queuing delay of its route back to its destination. Thus, energy and delay (instead of TTL) pheromone levels will be maintained at each node.

1) TTL-pheromone and Energy-pheromone: Consider Figure 3. Upon receipt of a new ant packet transmitted from node W to node X via Z, with TTL value $TTL(w, z)^2$, node X calculates the TTL pheromone value $\delta^T(w, z)$ to node W through node Z using:

$$\delta^T(w, z) = \beta_1TTL(w, z) + \delta^T(w, z)$$

where $\beta_1$ is a scaling constant.

Node X maintains a routing table, as shown in Table II, containing both a TTL-pheromone $\delta$ and energy-pheromone $\epsilon$ levels. Pheromone values will decrease in time in the absence of positive feedback. The periodic decay for the TTL-pheromone is governed by the iteration:

$$\delta^T(w, z) = \delta^T(w, z)\beta_2$$

where $0 < \beta_2 < 1$.

For a highly volatile network, we can increase the frequency of decay or decrease the $\beta_2$ value. This ensures that stale routes decay faster.

Similar to equation (3), the energy-pheromone is calculated as:

$$\epsilon^T(w, z) = \alpha_1E_{\text{min}}(w, z) + \epsilon^T(w, z)$$

where $\alpha_1$ is a scaling constant. Periodic decay for energy-pheromone levels is governed by:

$$\epsilon^T(w, z) = \epsilon^T(w, z)\alpha_2$$

where $\alpha_2$ is a constant satisfying $0 < \alpha_2 < 1$. Again for highly volatile networks, periodic decays are performed more frequently to ensure faster decay of stale route information.

The latency-non-critical packets are routed according to the probabilities that are simply proportional to energy-pheromone levels. Specifically, latency-non-critical packets at X destined to node W are routed through Z with probability:

$$p^E_{\text{z}}(w, z) = \frac{\epsilon^T(w, z)}{\epsilon^T(w, z) + \epsilon^T(w, y) + \epsilon^T(w, u)}$$

For latency-critical packets, first define a decision probability proportional to the TTL-pheromone:

$$p^T_{\text{z}}(w, z) = \frac{\delta^T(w, z)}{\delta^T(w, z) + \delta^T(w, y) + \delta^T(w, u)}$$

Note that the TTL value is decremented as packet propagates the network.

**TABLE II**

<table>
<thead>
<tr>
<th>Destination Node</th>
<th>Neighbor Node</th>
<th>TTL-Pheromone</th>
<th>Energy-Pheromone</th>
</tr>
</thead>
<tbody>
<tr>
<td>W</td>
<td>Y</td>
<td>$\delta^T(w, y)$</td>
<td>$\epsilon^T(w, y)$</td>
</tr>
<tr>
<td>W</td>
<td>Z</td>
<td>$\delta^T(w, z)$</td>
<td>$\epsilon^T(w, z)$</td>
</tr>
<tr>
<td>W</td>
<td>U</td>
<td>$\delta^T(w, u)$</td>
<td>$\epsilon^T(w, u)$</td>
</tr>
</tbody>
</table>

**Fig. 3.** Small example network topology

path. Once the destination receives the FANT, it sends a BANT back to the source using the same path the FANT has traveled. Therefore, both energy and delay information of the path are collected and delivered to the source. The pheromone levels of the source are updated using equations (3) and (5). Similar method to maintain the route periodically was introduced in [11].
Routing of latency-critical packets should consider both energy and delay pheromone levels. We will use the following decision probability to route through Z packets at X destined to W:

\[ p^p_{out}(w, z) = \frac{p^e_{out}(w, z) + \theta p^c_{out}(w, z)}{\sum_{i \in \{x, y, u\}}[p^e_i(w, i) + \theta p^c_i(w, i)]} \]  

(9)

where \( 0 < \theta < 1 \).

Note that \( p^p_{out}(w, z) \) combines both energy-pheromone and TTL-pheromone values into one quantity with a comparable magnitude, which is the main contribution of this paper. By normalizing the pheromones, we can make both pheromones have the same dimension. It is similar to the combination of normalizing the pheromones, we can make both pheromones magnitude, which is the main contribution of this paper. By

converting the TTL and cost metric in QoS routing [16], but unlike [16] they have the same dimension. It is similar to the combination of normalizing the pheromones, we can make both pheromones have the same dimension. It is similar to the combination of normalizing the pheromones, we can make both pheromones magnitude, which is the main contribution of this paper. By

where \( \gamma_1 \) is a scaling constant. The periodic decay equation is:

\[ d^e(w, z) = \gamma_1 D(w, z) + d^d(w, z) \]  

(10)

with \( \gamma_2 > 1 \). The component \( p^e_{out}(w, z) \), that replaces \( p^c_{out}(w, z) \), of the routing decision probabilities (9) for latency-critical packets is based on delay-pheromone is computed as shown in (12).

\[ p^e_{out}(w, z) = \frac{1/d^e(w, z)}{1/d^e(w, z) + 1/d^e(w, y) + 1/d^e(w, u)} \]  

(12)

V. PERFORMANCE OF PPRA

We have simulated PPRA using ns-2 [1] to validate its efficiency and ability under volatile MANETs environments. Packet delivery ratio, end-to-end delay, routing overhead, route discovery overhead, and residual node energy were used as metrics to compare the performance of PPRA with AOMDV and AODV routing protocols. Each simulation result (each reported point on each curve) represents an average of 10 independent trials.

Unlike the results of [24], the simulation results did not converge to a single optimal path. As the minimum residual energy of a node along a path varied with time, the amount of delivered pheromone changed. Therefore, instead of converging to a single path, the multiple paths were used alternatively. Most of the time, as most of the traffic was assumed to be low priority traffic, the source tried to use energy-rich path, which was determined by equation (9) and changed with time. Hence, load balancing was achieved and the network lifetime was maximized.

A. Simulation Environment

In our simulation study, we assumed a 1000m x 1000m area, where fifty nodes were disbursed randomly with the communication range of a node being 250 meters. The nodes employed the IEEE 802.11b medium access mechanism with 2 Mbps channel bit rate. To examine the effect of the routing algorithm only, the network was moderately loaded and ten source-destination connection pair was considered. A constant bit rate (CBR) traffic source model was used to generate data traffic of 1024 Byte packets. Using the CBR model, the source sent one data packet per second to the destination on average. Among all of the data, 10% was designated latency-critical. Latency-critical type had priority over latency-non-critical type and whether a packet was of the latency-critical or latency-non-critical type was determined at random and independently. In the simulation, we assumed that nodes had 20.0 Joules of battery energy initially. The transmitting and receiving power required by a node was set to 1.3272 W and 0.96696 W respectively [3].

Each simulation was run for 900 seconds using random waypoint mobility model with maximum speed of 20 m/sec. Seven different pause times, from 0 to 900, were used to observe the effect of the mobility. The pheromone values of each path were updated every 10 data packets. At this update phase, BANT packets, which carried the energy and delay information of the path, were used. The destination node sent back a BANT packet to the source along a path after receiving 10 data packets that traveled along the path.

B. Performance Metrics

Five metrics were taken into consideration: Packet delivery ratio is the ratio of successfully delivered data packets to the total data packets sent from the source to the destination. End-to-end delay is the amount of time needed to successfully deliver a packet from the source to the destination. End-to-end delays were observed separately for both priority (latency-critical) packets and nonpriority (latency-non-critical) packets. Routing overhead is the ratio of routing packets transmitted to the total data packets delivered. Routing packets include control packets used for route discovery, route maintenance, and pheromone updates. Route discovery overhead is the total number of route discovery phases and the total number of route request (FANT) packets distributed over the entire network. Finally, node energy is the average node energy and minimum node energy that were observed.

C. Simulation Results

Figures 4 to 10 shows the simulation results under different pause times. One data packet was generated every second (\( \lambda = 1 \)) and seven different pause times were considered. The maximum speed was set to 20 m/sec. Each simulation result for PPRA was compared to that of AOMDV and AODV.

Figure 4 shows the packet delivery ratio (PDR) of PPRA, AOMDV, and AODV. The PDR tends to increase as the pause time increases. This is manifest since the active path is less likely to break as the network becomes static. However,
the PDR first decreases as the pause time increases. Due to mobility, the active path may break. When all paths, including the backup paths, to the destination break, a new path can be discovered only after the change of topology of the network, i.e., a node that can form a path to the destination should come into the transmission range. Note that the change of topology is proportional to mobility. Hence, as mobility decreases it becomes more difficult to recover from the broken path. This reason explains the valley that appears when the pause time is 30 seconds in Figure 4. Note that AODV shows better PDR, especially under high mobility. Under high mobility, both the active path and backup path are more likely to break. Therefore, the packet sent out over the backup path just after the break of the active path is more likely to be dropped under high mobility.

In Figure 5, we can see that the end-to-end delay of latency-critical packets (P) is reduced by applying PPRA. Since multiple paths were discovered, when a path to the destination breaks, packets could immediately continue to be forwarded using a backup path without a new route discovery. Obviously, this reduced the end-to-end delay. We can observe that, in general, the end-to-end delay of latency-non-critical packets (NP) is greater than that of P packets. Note that AOMDV shows better performance from the end-to-end delay perspective under moderate and low mobility. Since PPRA considers both delay and energy jointly, not all priority packets were sent along the shortest path. For this reason PPRA generally requires more end-to-end delay than AOMDV. Unlike PPRA, AOMDV used the shortest path first. It switched to the alternate path when the shortest path was not available.

The routing overhead is shown in Figure 6. Since more control packets are required at the route discovery phase and extra control packets are required periodically to monitor the condition of the paths, the routing overhead of PPRA is higher than other protocols. The overhead for path monitoring can be reduced by piggybacking the pheromone information on data packets if appropriate traffic exists in opposite direction. Because of the periodic updates, PPRA requires certain amount of routing overhead constantly.

Figure 7 shows the result of number route discovery under various pause times. The number of route discoveries is less than that of the other two protocols, especially under high mobility, since PPRA discovers and maintains multiple routes to the destination. Figure 8 shows the result of total route request (FANT) packets distributed over the entire network. As explained in section III, PPRA requires more number of FANTs to be distributed per route discovery. However, since the number of route discovery is less than other protocols, the total number of distributed FANTs is comparable to that of others.

Figures 9 and 10 depict the average of residual node energy and minimum of residual node energy respectively.
Minimum node energy was defined as the residual energy of a node that had the smallest residual energy. Due to periodic pheromone updates, fundamentally, PPRA consumed more energy, in total, than other routing schemes under same circumstances. However, we can see that the average node energies under PPRA are similar to those of other protocols. This is due to the small packet size of the control packet. Since other routing schemes did not consider the priority of the data packets, all latency-critical and latency-non-critical packets were sent using the same path. Once the path was selected, the path was used until it became unavailable. PPRA, however, selected a path depending on the priority of data packets and pheromone values of multiple paths. Different type of data packets used different paths in PPRA. Hence, the discovered multiple routes were alternately used resulting in higher minimum node energy. From the results we can observe that PPRA increases the minimum residual node energy with comparable total energy consumption.

More simulations using different transmission rate were conducted by us, e.g., taking $\lambda = 2$, all leading to similar results as those reported above.

VI. AVOIDING MALICIOUS PACKET DROPPING

The ant routing framework has two types of feedback: positive feedback increases the pheromone levels on routes actively carrying ant packets and negative feedback periodically decreases pheromone values to limit the effects of stale information. Routing decisions tend to favor paths with higher pheromone levels and, when allowed to converge, shortest end-to-end paths are empirically observed to be favored [18].

In the following, we propose a modified ant mechanism algorithm that uses throughput metrics to perform updates of pheromone levels. Control packets (BANTs) are periodically sent from the destination to the source, after receiving every $N$ data packets, along a path that data packets have traveled. Once the source receives a BANT, it increases the pheromone level of the path that the BANT has traveled. Therefore, the pheromone level of a path is proportional to the throughput of the path. If all paths are secure, i.e., there are no malicious packet dropping along the paths, the data packets will eventually use the highest throughput as shown in [24].

Considering Figure 3, upon receipt of a new BANT packet transmitted from node W to node X via Z, node X calculates the throughput pheromone value $\delta^x(w, z)$ to node W through node Z using:

$$\delta^x(w, z) = C + \delta^x(w, z)$$

where $C$ is a constant. Node X maintains a routing table, as shown in Table II, containing a throughput-pheromone $\delta$ levels. Pheromone values will decrease in time in the absence of positive feedback. The periodic decay for the throughput-pheromone is governed by the iteration:

$$\delta^x(w, z) = \beta \delta^x(w, z)$$

where $0 < \beta < 1$. For a highly volatile network, we can increase the frequency of decay or decrease the $\beta$ value. This ensures that stale routes decay faster.

The data packets are routed according to the probabilities that are simply proportional to throughput-pheromone levels. Specifically, data packets at X destined to node W are routed through Z with probability:

$$p^z(w, z) = \frac{\delta^x(w, z)}{\delta^x(w, z) + \delta^x(w, y) + \delta^x(w, u)}. \tag{15}$$

If data packets are dropped partially (e.g., Jellyfish Attacks [2]) by a malicious node, the destination will receive fewer data packets and therefore send BANT packets back to the source less frequently along the path that contains...
the malicious node. Therefore, at the source, the correspond-
ing pheromone levels will slowly increase compared to the pheromone levels of other paths. Similarly, if data packets are dropped completely (e.g., Blackhole Attacks [2]) by a malicious node. The destination has no cause to send BANT packets to the source along that path. Therefore, the corre-
sponding pheromone level will stop increasing, and the path will be avoided. So, under ant-based multipath routing, mal-
licious packet dropping attacks (either Jellyfish or Blackhole attacks) are naturally avoided. Note that our scheme can avoid any attack that degrades the throughput of a path. Conventional single path routing algorithm cannot achieve this property. Even if an application layer sends back an acknowledgement to the source, the source cannot avoid the malicious node at the rerouting phase. Further more, the acknowledgement could be maliciously dropped.

Two attack scenarios were considered in our simulation study. First, we simulated Jellyfish [14] attacks. Specifically, we assumed, among three paths, one contained one malicious node and another contained two malicious nodes. The mal-
licious nodes were assumed to have benignly cooperated in the routing phase to build multiple routes between source and destination. However, subsequent data packets and BANT packets were dropped discreetly without an error message sent back to the source. Therefore, the source had no knowledge of malicious packet drops. Generally, however, a source receives an error message when normal packet loss occurs\(^3\), and it considers the path to be broken.

We simulated two different dropping rates \((\rho = 0.05 \text{ and } \rho = 0.10)\). Discrete intruders with small dropping rates were selected in order to validate our scheme. A malicious node with large dropping rate can easily be detected and avoided. Secondly, we simulated Jellyfish and Blackhole [14] attacks together. A Blackhole attack is a special case \((\rho = 1.00)\) of a Jellyfish attack, i.e., all packets are simply dropped. Specifically, we assumed that one path contained a malicious node performing Jellyfish attacks and another contained a malicious node performing Blackhole attacks. Unlike the first scenario, all nodes were assumed to be secure initially. The node that performed Jellyfish attacks was compromised at 100 seconds, and the node that performed Blackhole attacks was compromised at 200 seconds.

Path throughput, under the first scenario, is depicted in Figure 11. The secure path in this experiment was path 1. One of the nodes of path 2 became malicious at 100 seconds and performed Jellyfish attacks. Again, two different dropping rates \((\rho_1 = 0.05 \text{ and } \rho_1 = 0.10)\) were considered. At 200 seconds, one of the nodes of path 3 started to drop all data/BANT packets (Blackhole attack, \(\rho_2 = 1.00\)). We see the throughput of path 3 dropped to zero at 200 seconds. As expected, for both Figure 12(a) and 12(b), path 1 became dominant. Other performance metrics, under Jellyfish and Blackhole attacks, are summarized in Table IV.

\[\begin{array}{cccc}
\text{Drop Rate} & \text{Packet Delivery Ratio} & \text{End-to-End Delay} & \text{Routing Overhead} \\
\text{Rate} & \text{(\%)} & \text{(msec)} & \text{(\%)} \\
0.00 & 99.98 & 32.6 & 10.04 \\
0.05 & 96.83 & 32.8 & 10.02 \\
0.10 & 99.40 & 36.8 & 10.01 \\
\end{array}\]

Similarly, path throughput, under the second scenario, is depicted in Figure 12. The secure path in this experiment was path 1. One of the nodes of path 2 became malicious at 100 seconds and performed Jellyfish attacks. Again, two different dropping rates \((\rho_1 = 0.05 \text{ and } \rho_1 = 0.10)\) were considered. At 200 seconds, one of the nodes of path 3 started to drop all data/BANT packets (Blackhole attack, \(\rho_2 = 1.00\)). We see the throughput of path 3 dropped to zero at 200 seconds. As expected, for both Figure 12(a) and 12(b), path 1 became dominant. Other performance metrics, under Jellyfish and Blackhole attacks, are summarized in Table IV.

\[\begin{array}{cccc}
\text{Drop Rate} & \text{Packet Delivery Ratio} & \text{End-to-End Delay} & \text{Routing Overhead} \\
\text{Rate} & \text{(\%)} & \text{(msec)} & \text{(\%)} \\
0.05 & 97.02 & 28.6 & 10.05 \\
0.10 & 96.55 & 27.8 & 10.03 \\
\end{array}\]

VII. CONCLUSIONS

In this paper, we proposed a multipath QoS routing protocol that can accommodate both latency-critical and latency-non-
critical traffic under volatile network conditions. The mech-
anism was based on information obtained from periodically transmitted ants resulting in reinforced path-pheromone levels. For latency-critical traffic, routing decisions were based on combinations of normalized energy and delay pheromone lev-
els (9). For latency-non-critical traffic, only energy pheromone levels were used. With the help of PPRA, latency-non-critical traffic was transmitted over the most energy rich path, while latency-critical traffic was transmitted over low delay paths that are also energy rich for reliability (9), and the network lifetime was maximized. In simulation studies our algorithm rapidly responded to network volatility maintaining good delay and throughput performance with a comparable amount of (ant) overhead, that itself required little in the way of computa-
tion and no specialized hardware. In particular, for additional overhead, the algorithm showed good minimum energy performance which is an important factor of maximizing net-
work lifetime. Also, the algorithm was applied to a security context in order to avoid malicious packet dropping. The mechanism was based on throughput information obtained from periodically transmitted BANTs resulting in reinforced path-pheromone levels.

\[\text{3 Normal packet loss is caused by various reasons such as congestion (buffer overflow), excessive collision (limited retransmission), topology change, etc.}\]
Fig. 11. Path Throughput under Jellyfish Attacks

Fig. 12. Path Throughput under Jellyfish and Blackhole Attacks

REFERENCES


Paul Barom Jeon  
Paul Barom Jeon received his M.S. in Electronic Engineering from Sogang University, Korea in 1996. He was a research engineer in the Mechatronics Center of the Samsung Heavy Industries, Korea, from 1996 to 2001. He received his Ph.D. in Electrical Engineering from Pennsylvania State University in 2006. His research areas span QoS multipath routing, network protocol design, and security. Currently, he is a student member of the IEEE.

George Kesidis  
George Kesidis received his M.S. and Ph.D. in EECS from U.C. Berkeley in 1990 and 1992 respectively. He was a professor in the E&CE Dept. of the University of Waterloo, Canada, from 1992 to 2000. Since April 2000, he has taught in both the CS&S and EE Departments of the Pennsylvania State University. His research experience spans several areas of computer/communication networking including security, incentive engineering, efficient simulation, and traffic engineering. Currently, he is a senior member of the IEEE and the TPC co-chair of IEEE INFOCOM 2007.