EFFICIENT PROVISIONING FOR SERVICES IN LARGE-SCALE WIRELESS-WIRED NETWORKS

AHMED HELMY

DEPARTMENT OF ELECTRICAL ENGINEERING, UNIVERSITY OF SOUTHER CALIFORNIA

1. OBJECTIVE

This document outlines an architectural design for future internets, that are envisioned to include integrated wired and wireless (ad hoc) networks. Although such an environment presents several research challenges, we believe that the promise in such networks substantially outweighs the risk, and we shall show that the research directions proposed herein hold promise to overcome the challenges. One major challenge in ad hoc networks – that differentiates such networks from the Internet – is the lack of infrastructure (such as autonomous systems (ASs) or DNS). Another major challenge is *mobility*, which leads to a highly dynamic network that cannot be managed using conventional networking techniques. Moreover, many services in ad hoc networks may be *data-centric*. Meaning that the goal of the network may be to gather, store, and retrieve data with node*id*-independent semantics (e.g., spatial data). For such applications, conventional communication paradigms based on node IDs (such as those used in today's Internet, i.e., IP addresses) may not be suitable. Therefore, we introduce a novel architecture for ad hoc networks that provides efficient resource discovery (whether these resources are defined by physical node IDs, location, data, or in general *logical* IDs). In doing so, this proposal provides the building blocks for providing services in such emerging networks. We design our architecture to be self-configuring, scalable, efficient and robust. In addition, we provide an architecture for the necessary 'glue' that ties ad hoc and wired networks. Using a single mechanism (called multicast-based mobility) we provide efficient handoff solutions for unicast and multicast scenarios.

Although the solutions set forth in this proposal may be generalized for classes of ad hoc networks, we shall target specific example applications that illustrate the utility of such technology. In particular, we choose two applications: (a) person tracking and location, and (b) wildfire fighting. We believe that without our proposed architectures, these applications cannot be enabled efficiently in large-scale mobile wireless networks using existing technologies. This will be explained further in more detailed discussion of these applications, in addition to the related work section.

- Motivation

With the recent advances in microelectronics and communications, PDAs, wearable computers, and embedded systems are predicted to proliferate and provide the future platform for ubiquitous computing and communications. This will potentially enable several new classes of applications, such as location specific services, person and object localization, and search and rescue missions.
These networks may scale up to tens of thousands of wireless mobile nodes, some of which may be power constraint. Developing efficient protocols for such environments provides several research challenges. First, the protocols need to be designed with *power* consumption in mind. One main

factor of power consumption is communication. Hence, power awareness necessitates designing protocols with reduced communication overhead. Second, in such networks all nodes are potentially *mobile*. Such mobility produces a highly dynamic network, in which network connectivity is constantly changing. Unlike wired networks, these changes cannot be dealt with as failures, since they are part of the normal operation of the network. Ad hoc network protocols must adapt efficiently to mobility and network dynamics. Third, these networks may contain thousands of nodes. Hence, the protocols designed must scale with network size (i.e., performance of these protocols must be acceptable with the increase in number of nodes in the network). Fourth, the lack of infrastructure in ad hoc networks creates a barrier for service provisioning (such as multicast, location-based services or storage location). Without fixed servers with well-known addresses it is not possible to bootstrap these services with existing technologies. The protocols designed must be self-configuring and must enable efficient resource discovery without any reliance on existence of well-known servers or infrastructure. Fifth, we do not expect services in ad hoc networks to be necessarily linked to physical IDs (such as IP addresses). We envision various services (such as location-based services) to be linked to *logical IDs* instead. The designed protocols must have support for mapping logical IDs efficiently onto physical resources. This is what we call *data-centric* networking. Finally, these networks will be integrated with the wired Internet. The architecture proposed must allow efficient and seamless integration of ad hoc networks with the wired networks.

- Research hypothesis

The main architectural components of our design target large-scale ad hoc networks. As was mentioned above, the infrastructure-less nature of such networks brings about the need for efficient resource discovery. We provide two main mechanisms to achieve efficient resource discovery. The first is based on the small world phenomenon and establishes a contact-based architecture in wireless networks. We provide a unique approach for *mobility-assisted* contact selection that utilizes mobility to improve performance. The second mechanism provides a robust rendezvous mechanism to bootstrap services in such networks. To the best of our knowledge, this is the first work to address service bootstrap in large-scale ad hoc networks. In addition, we provide a component to tie wireless networks into wired networks efficiently using a multicast-based mobility scheme. Hence, the three main components of our architecture are (I) contact based mechanisms, (II) rendezvous mechanisms, and (III) multicast-based mobility.

I – Contact-based Mechanisms: Building a Small World in Mobile Wireless Networks

Self-configurability and infrastructure-less-ness of ad hoc networks provides resilience in the face of network dynamics and failures. In the design of infrastructure-less networks, efficient resource discovery is an essential component. Current approaches for resource discovery employ two main techniques. The first is flooding (or broadcast) where a request is sent to every node in the network to locate the resource. Many approaches for ad hoc networks depend on flooding for discovery (e.g., DSR [26], AODV[27]). This approach consumes a lot of communication resources and energy and does not scale well with the size of the network. The second is cluster-based hierarchies [47], where nodes would form clusters, electing a cluster-head within each of these clusters to be responsible for inter-cluster communication. Although in general they perform better than flooding, cluster-based hierarchies suffer from susceptibility to major re-configuration with failure or movement of nodes (especially cluster-heads). This is due to the complex coordination needed to setup the clusters. Frequent re-configuration wastes network resources and introduces undesirable network transients.

Other approaches use landmark hierarchy [44] (e.g., LANMAR[46] and SCOUT[45]). These approaches do not perform well in highly dynamic, high mobility environments. The zone routing protocol (ZRP)[4][29][30] uses the concept of zones with a table-driven protocol for intra-zone routing and on-demand protocol for inter-zone routing. The on-demand protocol uses flooding between borders of the zones (called 'bordercasting'). Although ZRP does not include complex hierarchical schemes, and performs better than flat flooding, it still uses global flooding between borders. In our approach we avoid bordercasting and instead use contacts, which proved to perform much more efficiently as we will show. We shall discuss these approaches further in the related work section.

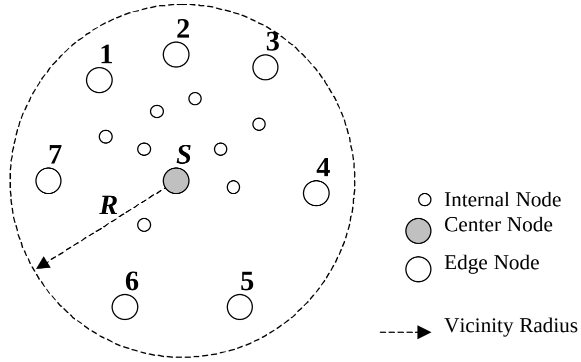

Figure 1 Vicinity from the perspective of node ***S*** as a center node. ***S*** knows about resources (internal and edge nodes) up to ***R*** hops away using table-driven protocol. Nodes at exactly ***R*** hops are called edge nodes.

To overcome the above limitations, we introduce a new loosely-coupled hierarchical architecture for resource discovery that is based upon the small world phenomenon. In our architecture, a wireless node knows about resources in its *vicinity* (up to *R* hops away). This is achieved through a localized table-driven routing protocol (e.g., DSDV[25]). Note that every node has its own vicinity, and hence has its own view of the network. This is shown in Figure 1. Unlike complex hierarchies, our architecture does not require major re-configuration due to dynamics, each node simply changes its own view of the network without coordination with other nodes. For out-of-vicinity resources, a node queries a small number of nodes (called *contacts*) outside its vicinity. Contacts increase the network view of a node and act as shortcuts to form a small world in the wireless network. We call our approach the 'contact-based architecture for resource discovery' (*CARD*). We have shown in [2][3] that *CARD* performs drastically better than the previous approaches. An interesting problem introduced by such architecture is the selection and maintenance of the contacts. We address such problem next.

Contact Selection and Maintenance Schemes

One approach for contact selection would be to send random recruiting messages outside of the vicinity to build those contacts. However, random contact selection may lead to incomplete network coverage and unpredictable performance. One improvement would be to choose contacts in a way that minimizes vicinity overlaps between the contacts and the original node. At the same time, we want to limit the distance at which the contacts may be selected (to reduce the maintenance overhead)[1]. To reduce overlap we include a TTL (hop count) field in the message. Unfortunately, once the selection message goes out of the vicinity, it has no sense of direction and it is not possible to depend only on hop count to infer overlap information (since a node knows about other nodes only within its vicinity). This leads to diminishing returns in terms of reachability with the increase in number of contacts; as more contacts are selected the number of nodes reachable tends to saturate. At the same time, looking for a good contact (with minimum overlap) becomes harder and the protocol incurs increased overhead (due to search and backtracking). To ameliorate this problem we include the list of edge nodes (those at the edge of the vicinity) in the contact selection message. This provides nodes along the selection path with information to infer overlap, and hence leads to increased reachability. This also leads to drastically reduced overhead due to reduction in backtracking [2].

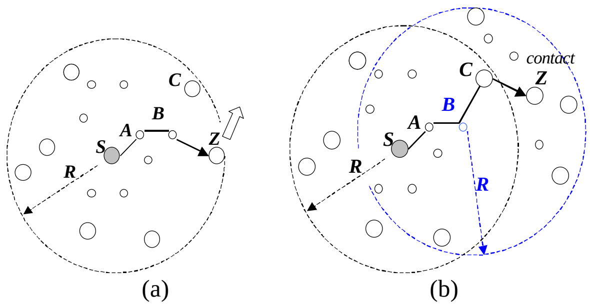

Figure 2. Lightweight contact maintenance protocol (a) *Z* is at the edge of *S*'s vicinity with a route '*S-A-B-Z*'. *Z* is moving out of vicinity. (b) *Z* is no longer within *S*'s vicinity, but it highly likely to exist in *B*'s vicinity. The route '*S-A-B-C-Z*' is identified and *Z* is selected as a *contact.*

For contact maintenance, we use a lightweight routing protocol that leverages already-existing vicinity information for each node. If a contact moves, then the previous hop node (in the original path) will have a valid route to the contact so long as it is within *R* hops away; a highly likely scenario in most cases. This is shown in Figure 2. In [2][3] we present more details on this approach and show that our protocol converges on stable contacts over time leading to reduced maintenance overhead and better reachability.

We have compared our *CARD* approach to flooding and bordercasting (employed by ZRP) using simulations over various topologies using random way point mobility and random queries. As shown in Figure 3, flooding performance degrades drastically with the scale of the network. Bordercasting (i.e., ZRP) performs better than flooding, but the best performance was achieved using *CARD*. Even after adding the overhead of contact selection and maintenance, *CARD*'s total overhead is less than 'third' that of ZRP query overhead for large networks. This is achieved by avoiding border flooding (or bordercasting) and instead using contact queries.

---

[1] In [1] we have shown that limiting the distance of contacts to around 20%-30% of the network diameter does indeed result in a small world graph and achieves the best reduction in average degrees of separation (ADS). ADS in our context denotes the average number of nodes queried before reaching the target. Reducing ADS translates into reduction in search overhead.

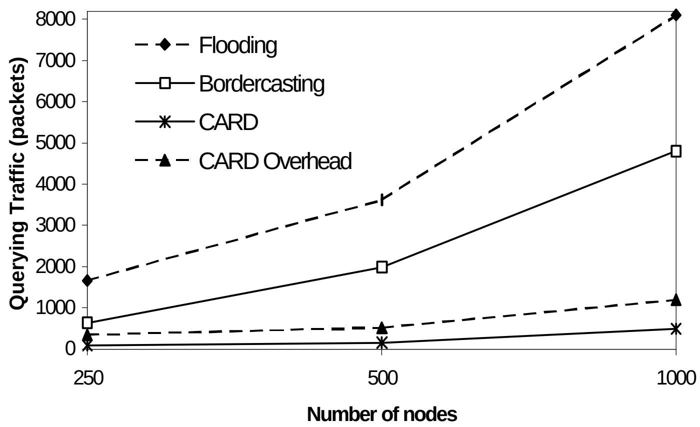

Figure 3. Simulation comparison of the contact-based approach to flooding and bordercasting (employed by ZRP)

In [1][6] we introduce another novel approach for contact selection that *takes advantage of mobility*. In that approach, a node first selects candidate contacts from its vicinity. While these nodes move and exchange messages with the original node (through the table-driven protocol) those nodes with desirable energy and mobility patterns are kept as candidate contacts. As the candidate contacts move out of the vicinity, such that they have minimum overlap with other contacts and the original node, they are promoted to become contacts and are used in resource discovery. This approach is illustrated in Figure 4. Unlike other approaches that consider mobility a liability, this novel protocol utilizes mobility to enhance its performance. In this sense, this protocol is quite unique as its performance *improves with mobility*, whereas other works on ad hoc networks clearly state that their performance degrades with mobility.

One drawback of this approach, is its dependence on mobility. Hence, in non-mobility or low-mobility scenarios, this approach may perform poorly.

As part of this research, we propose to design, develop and evaluate a hybrid, mobility-adaptive, approach that uses mobility-assisted protocols when suitable, but adapts by using efficient contact selection schemes during non-mobility scenarios. In addition, we shall re-visit mechanisms for minimum overlap detection. Currently, sending the list of edge nodes may incur high overhead. We plan to use encoding techniques (such as Bloom filters [22] and their variants [23][24]) to detect vicinity overlap statistics in a communication-efficient manner. Furthermore, we plan to evaluate the above protocols extensively using various topologies, mobility models [18] and communication patterns. This shall be done using network simulation (NS-2 [13]) as a first step, then using a laboratory test-bed for in-door and out-door environments in later stages.

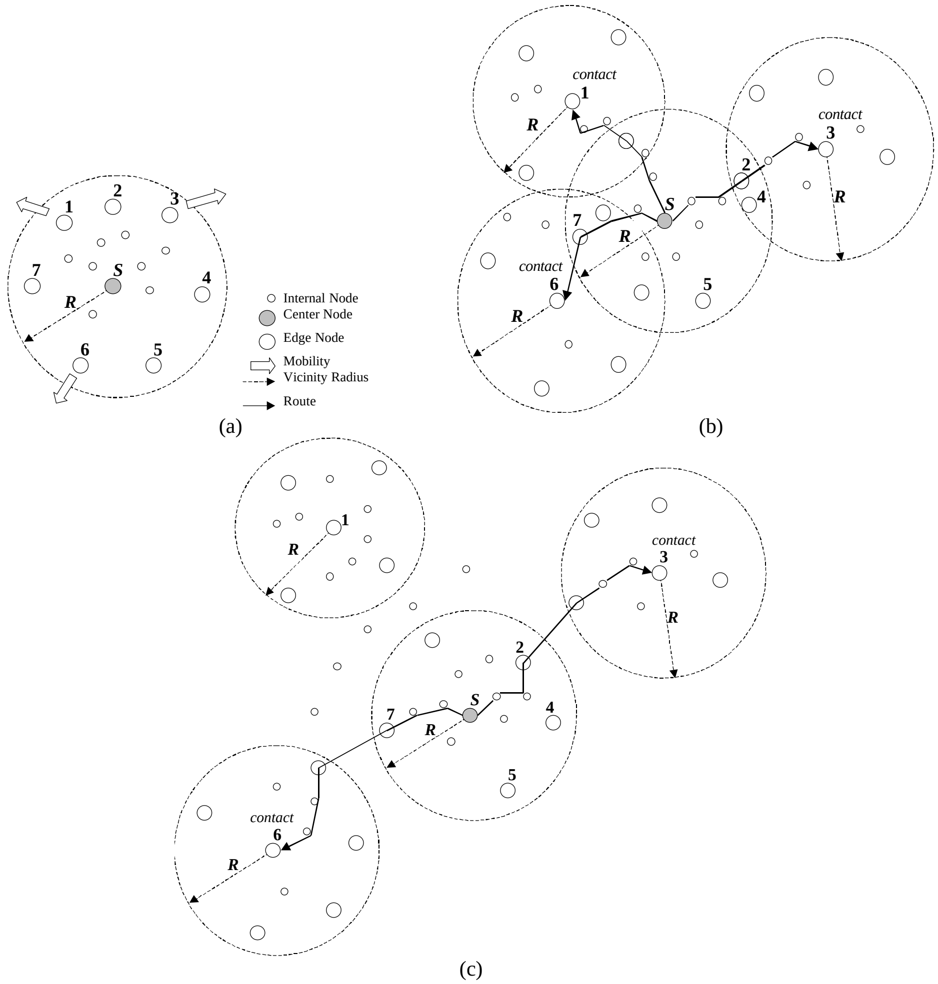

Figure 4 Example of vicinity, contacts and effect of mobility: (a) Vicinity for source node *S* is shown (with radius *R*). Edge nodes are numbered (1-7). Nodes 1,3 and 6 are moving/drifting out of vicinity. (b) Radii for vicinities of the drifting nodes are shown. *S* stays in *contact* with the drifting nodes, which enables it to obtain better network coverage with low overhead. (c) After moving away, contact nodes drift up to a point where their vicinities no longer intersect with *S*'s vicinity. In this example, *S* maintains contact with those nodes not more than (2*R*+1) hops away, i.e. nodes 3 and 6, and loses contact with node 1 as it drifts farther than the contact zone.

Summary I:

We propose a contact-based architecture for efficient resource discovery in large-scale ad hoc networks. Our architecture is based upon the small world phenomenon and leads to significant reduction in degrees of separation, and hence reduces the query overhead required for discovery. We have shown that our protocols improve upon existing approaches drastically. In addition, mobility-assisted techniques promise even more improved performance with mobility. This architecture provides the first component upon which we shall provide our service bootstrap schemes, discussed next.

II- Rendezvous Mechanisms: Bootstrapping Services using Rendezvous Regions (*RRs*)

The contact-based architecture presented above provides a base upon which we can build a model for bootstrapping services in large-scale ad hoc networks. Our bootstrap model uses a new concept that we call rendezvous regions (*RRs*). Our model extends to a large class of generic services. Examples of such services include (but are not limited to) multicast session establishment and directory, location-based services, and distributed storage and retrieval systems. The main idea is to self-configure the nodes into dynamically promoted servers that share responsibility of updating the service information. The only information the nodes need is a mapping scheme that maps service identifiers into physical regions. To illustrate, we consider a multicast service example. The multicast space is broken into group prefixes (*Gprefix*). Each *Gprefix* is assigned a *RR*. The mapping '$Gprefix_i \leftrightarrow RR_i$' is provided to all the nodes (either as a simple table or a hash function). Using that mapping, any sender and receiver for a group *G* can consistently determine the corresponding *RR* and send update/query messages to it. The messages are forwarded to the *RR* with the aid of contacts using approximate location information. For this rendezvous scheme, we assume that nodes know their approximate locations[2]. A subset of the nodes that reside in RR_i are elected using local promotion schemes to maintain $Gprefix_i$ information. We call those elected nodes service discovery servers (*SDSs*). Figure 5 shows the basic mechanisms of this scheme. If one of the servers moves out of the *RR* it would send a leave message that would trigger another node to promote itself as a server. Hence, this mechanism is self-replenishing and adapts gracefully to mobility and network dynamics.

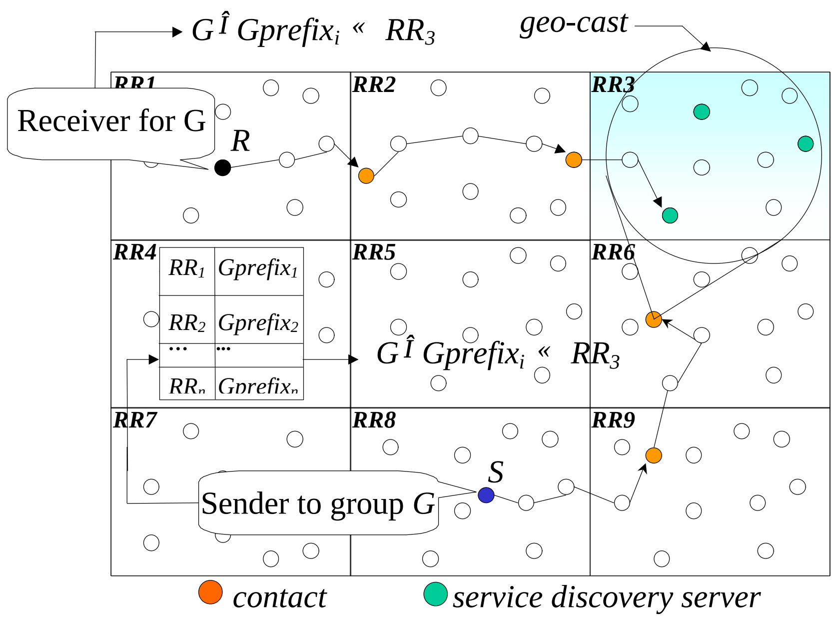

Figure 5. Solving the multicast rendezvous problem: In scalable multicast members of a group need only know the group ID (*G*) without prior knowledge of other members. Our approach: (1) Network is broken into *rendezvous regions (RR)* and the multicast address into group prefixes (*Gprefix)*. Nodes need only know mapping between *Gprefix* « *RR*. (2) In *RRi* servers are dynamically elected locally to serve $Gprefix_i$. (3) Senders use mapping to store their information at servers in *RRi*. (4) Receivers use mapping to retrieve information from servers in *RR*. (5) *Contacts* are used to discover servers efficiently.

This rendezvous concept can also be applied to storage and retrieval networks where a file is mapped into *RR*, and so on. Also, location-based services can be naturally supported by this scheme where location-specific information would be stored at servers in *RR* containing that location. Parties interested in such information use the query scheme provided above to reach the corresponding *RR* efficiently and obtain the information.

In this proposal, we plan to further detail the design of our rendezvous scheme, provide protocol specifications for the mapping algorithm, approximate geographic routing using contacts, and the

---

[2] Since this scheme utilizes contacts, the location information used need only be approximate and rough, thus this scheme is more robust than other geographic based routing in the face of imprecise location information.

directory service bootstrap mechanism. We will further investigate enhanced robustness by using multiple mappings (or hash functions), and peer-to-peer network concepts (e.g., CAN[37], Chord[38], freenet[39]). Multiple mappings would lead to more replication of data in the network, which will also incur extra overhead. There is a spectrum of possibilities between replicating the data in all (or some) *RR*s, or only replicating forwarding pointers to the data and having the query follow the pointers. We shall investigate various replication/query strategies and study such design trade-off. Furthermore, we plan extensive evaluation of the proposed algorithms and comparison with other geographic routing schemes (e.g., LAR[35], GLS[28] and GPSR[36]). One aspect that we plan to investigate (that has not been investigated before) is the performance of these protocols under imprecise location information. We predict that our scheme (that operates using contacts on the order of *2R* hops apart) is more robust to errors and imprecision in location measurement and estimation than other geographic routing schemes. This is yet to be tested.

Summary II:

We propose a rendezvous mechanism to bootstrap services in large-scale ad hoc networks. Our mechanism uses simple mapping between service resources into physical regions. Those regions hold dynamically elected servers responsible for maintaining the service resource. Consistent mapping allows different nodes to rendezvous at a region. This scheme is robust to mobility and errors in location estimation. Example services supported include multicast, storage-retrieval systems (such as peer-to-peer networks), and location-based services.

III- Multicast-based Mobility (M&M) to Integrate Wired and Wireless Networks

Now that we have outlined several techniques that address efficient resource discovery and service bootstrap in large-scale ad hoc networks, we need to provide an efficient architecture by which mobile wireless nodes can communicate with wired nodes through the Internet infrastructure. A possible architecture for such heterogeneous network is to have multiple ad hoc networks, each of which may be connected to the Internet through one or more base stations (or access routers), and is considered as a LAN in that sense. One of the main problems associated with such architecture is the handoff problem; the problem of maintaining continuous communication while moving between access points. Several micro mobility protocols have been proposed to address this problem in the context of IP mobility (i.e., last hop wireless). All such solutions, however, target unicast data only. By contrast, we provide a solution that enables efficient handoff for both unicast and multicast data. We call our solution multicast-based mobility (M&M). Mobile nodes in ad hoc networks may be deployed to carry out specific tasks (e.g., search and rescue or disaster relief), and hence multicast communication plays an important role in such networks to enable their efficient collaboration. In [7] we introduced the main architecture. In [19][20][21] we introduced detailed mechanisms for using M&M for micro-mobility (intra-domain) and show that it outperforms most micro-mobility (e.g., hierarchical MIP[15], seamless handoff[50], Hawaii[17]), while achieving very similar performance (in terms of loss and delay during handoff) to the best micro-mobility approach (cellular IP 'CIP' [16]). Our architecture also allows M&M to co-exist with Mobile IP (MIPv4 [9]) and MIPv6 [11] as inter-domain protocols. As was pointed before, M&M is also suitable for multicast support to mobile users while CIP was designed for unicast only. Furthermore, M&M can support reactive and proactive handoff schemes that minimize handoff delays and losses, that no other unicast-based scheme can provide[21]. For a campus network to use M&M, the only requirement is to deploy multicast within that campus network or domain. Similarly, other micro-mobility approaches (e.g., CIP and Hawaii [17]) require deployment of protocol-aware routers within that domain. Also, dynamic configuration and address discovery (i.e., DHCP) may be used, as it would be for CIP or Hawaii, to obtain the regional mobile address from the visited sub-network. In addition, M&M (by virtue of using multicast routing) is more robust than other micro-mobility architectures in the face of failure of the root-router[3].

Following we shall illustrate how M&M is used to provide efficient handoff for unicast traffic, (when Mobile IP is the inter-domain mobility protocol). In basic multicast-based mobility, each mobile node (MN) is assigned a multicast address, instead of a unicast address. The MN, throughout its movement, joins this multicast address through locations it visits. Senders wishing to send to the MN send their packets to its *multicast* address, instead of unicast. Because the movement will be to a geographical vicinity, it is highly likely that the join from the new location, to which the mobile recently moved, will traverse a small number of hops to reach the already-established multicast distribution tree. Hence, performance during handoff improves considerably. An overview of this architecture is given in Figure 6. As the MN moves, it joins to the assigned multicast address through the new access router. Once the MN starts receiving packets through the new location, it sends a prune message to the old AR to stop the flow of the packets down that path. Thus completing the smooth handoff process.

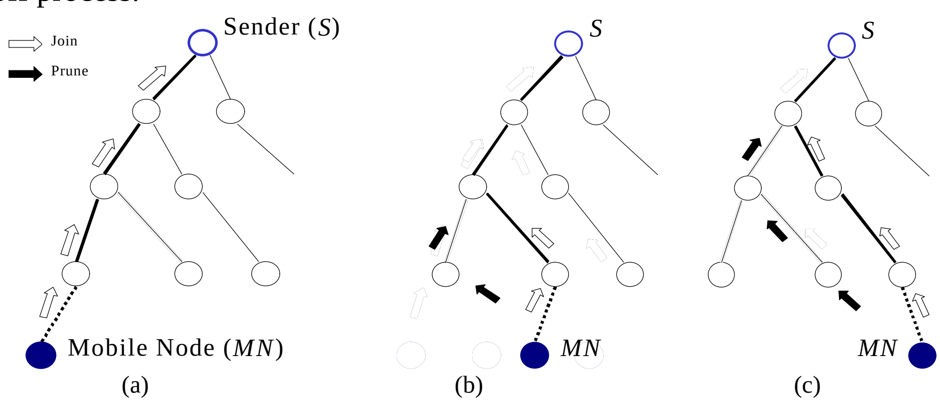

Figure 6. Multicast-based mobility. As the MN moves, as in (b) and (c), the MN joins the distribution tree through the new location and prunes through the old location.

To allow for gradual deployment of our scheme and co-existence with MIP, we limit our architecture to a domain. When a mobile node moves into a new domain, it is assigned a unicast address that is unique within the domain, called regional care of address (RCOA). The RCOA is chosen from a pre-defined subnet address used only for mobile nodes. This is needed to identify packets sent to mobile nodes. The RCOA maps to a multicast care of address (MCOA). The MCOA is used for routing packets within the domain, so there is no need to assign COA at every subnet. These multicast addresses are domain-scoped for micro-mobility (i.e., they have only local significance within the domain). In our scheme there is a one-to-one mapping between an RCOA and

[3] The root-router, as defined in Hawaii or CIP, is the router at which the micro mobility tree is rooted and through which traffic destined to mobile nodes funnels. This constitutes a single-point-of-failure. To alleviate such a problem, these micro-mobility protocols need to introduce replication or redundancy solutions that, in turn, introduce a new set of consistency problems during failure or healing of routers. M&M uses the underlying multicast routing protocol (e.g., PIM-SM) that already has built-in mechanism for dynamic election of rendezvous-points (RPs). M&M does not specify a single root-router, rather it uses the border routers to map the packets destined to mobile nodes into multicast packets. Failure of the RP or a border router will automatically trigger dynamic mechanims (in the underlying multicast routing) to adapt gracefully and recover from failure.

MCOA. When a mobile node moves into a new domain it is assigned RCOA by the access router (AR) and the mobile performs inter-domain handoff; i.e., it registers the RCOA with its home agent (HA) for MIPv4 or the corresponding node (CN) for MIPv6 with route optimization. The AR automatically infers the multicast address (MCOA) for the mobile node from the assigned unicast address (RCOA) through a straightforward *algorithmic mapping* (as described in [21]). The AR then triggers a Join message for MCOA to establish the multicast tree. Packets destined to the MN's home address are tunneled to its RCOA by the HA. When these packets arrive in the foreign domain, they are identified by the border router (BR) as being destined to a mobile node. As shown in Figure 7, the BR maps the destination unicast address to the *multicast address* and transmits the packets to the MN down the multicast tree. In this case the multicast tree consists of the branch leading to the mobile node (explicit join mechanisms are used, and so no broadcast is performed).

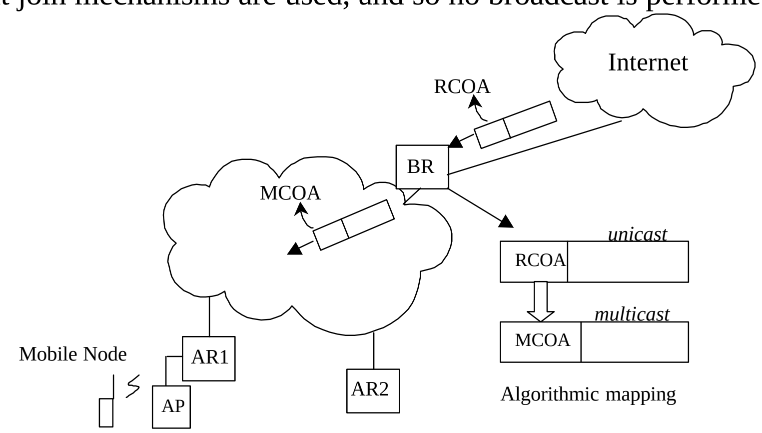

Figure 7. High level architectural view: Data packet is unicast over the Internet destined to the RCOA and arrives at the border router (BR) for the mobile node. The BR intercepts the packet and performs algorithmic mapping from the RCOA to MCOA. The packet is then multicast within the domain to the mobile node.

If a mobile node goes into 'sleep' or power-saving mode, when it wakes up it uses its RCOA to contact the nearest access router, the access router maps the RCOA into MCOA using algorithmic mapping, and joins to the multicast tree. We have evaluated our protocol using detailed NS-2 simulations over a rich set of topologies, mobility scenarios, and cell overlap conditions. We note that geographic vicinity (in mobility) may not necessary lead to network vicinity (i.e., handing off to a new access router that is directly connected to the previous access router). This was considered in our evaluation. Based on our previous studies, M&M holds a lot of promise for a very efficient handoff scheme, achieving low loss and delay during handoff, and traversing, on average, the least number of hops to the already established tree[20][21]. M&M's application and evaluation so far have been studied for single-hop (last-hop) wireless. As part of this proposal we shall investigate its application as a mechanism to tie large-scale ad hoc networks to the Internet efficiently. Note that multicast addresses are *logical* addresses (unlike unicast IP addresses), and thus are quite suitable for use with our overall scheme described above (based on contacts and *RR*s). Another aspect that will be studied is the interoperability between existing Internet protocols, M&M and ad hoc network protocols. Since our architecture allows operation with no change to the mobile node (the access router does the multicast functionality), then we expect smooth interoperation with existing protocols. The M&M can be used seamlessly to provide efficient handoff. A mobile connecting to an ad hoc network that, in turn, connects to the Internet through a new access router, would register its RCOA with the access router which maps it to the MCOA and triggers a join, so on. This, and the multicast scenarios, shall be investigated further and will be carried out through simulations and test bed experimentation. We may also investigate M&M between different heterogeneous LANs (e.g., 802.11 and GPRS). Another intriguing problem would be to explore the operation of M&M over a combination of IPv4 and IPv6.

Summary III:

We propose a multicast-based mobility scheme to integrate wired and wireless networks. Our approach leverages multicast technology to enable very efficient handoff performance between access routers. This was shown through extensive simulation and comparison studies. This architecture may be used for unicast and multicast traffic and we consider it a suitable choice for the glue required between wired and wireless ad hoc networks.

IV- Potential Target Applications

To illustrate the utility of our proposed protocols and architectures we discuss two of its potential applications; namely person/object tracking and wildfire fighting. In this work we propose to tailor the design, parameter setting and trade-offs in our architecture to each of the above applications. This will be performed as part of our efforts to provide a systematic procedure that can be followed for other applications that belong to the same (or similar) classes of applications.

A- Tracking people and objects using ad hoc networks

Tracking technologies are quite useful and sometimes are crucial in emergency situations. Unfortunately, usefulness of such technologies is limited to the coverage of the supporting systems (e.g., the cellular-network provider). The ability to configure the tracking nodes in an ad hoc network fashion could potentially increase coverage of the tracking system drastically. An interesting application of tracking systems is in monitoring and tracking of people (including children and elderly people). Recent advances in electronics and nano-technology have enabled new kinds of biomedical sensing. Utilizing such technologies, we envision near-future devices that monitor vital signs (e.g., pulse, blood pressure, temperature and blood glucose levels). This enables a device to detect alarming changes in ones health. These devices will have computing and wireless communication capabilities. Networking such devices in an ad hoc network and reporting health and location information during emergencies may prove quite valuable in providing much needed proper and medical assistance. This vision may constitute the heart of the future, fully automated, emergency 911 system. Without a scalable ad hoc architecture that adapts gracefully to mobility, such application cannot be realized. We believe that our mobility assisted contact-based architecture is the only existing scheme that possesses such desirable features. Furthermore, a rendezvous mechanism is necessary to provide scalable services in such a network. Our *RR* architecture is the only architecture, of which we are aware, that provides such a method without requiring accurate location information of all nodes. To enable such network to be connected to the Internet the M&M protocol may be used.

B- Efficient wildfire fighting using ad hoc networks

Wildfires often cover extended uninhabited geographical areas. The spread of such fires depends on several factors (e.g., wind, fuel and temperature) some of which are affected by the fire itself (e.g., fire-generated winds). These factors, among others (such as slopes, moisture distribution and canopy), are location-specific. Weather forecasts and wind measures in general provide helpful information in fighting such fires. However, currently such information is not real-time nor is it location specific. Ad hoc networks of sensing devices can potentially increase the efficiency of wildfire fighting by providing real-time information about temperature, moisture, and other factors mentioned above, along with location information. This information would help predict fire spread, increase efficiency in containing the fire and potentially save lives and property. The sensing devices may be deployed rapidly using airplanes, and they self-configure using our contact-based architectures. The measured information may need to be replicated rapidly for robustness (in case of device damage due to fire), and the network may be queried for specific data as needed. This data will help in predicting future fire spread, and hence will aid in effectively allocating fire fighting resources. To enable efficient operation of such a network we propose to use *RRs* to provide robust replication of monitored data in case of damage due to fire. Also, we propose the *CARD* (non-mobility) protocol to provide efficient query of the network for the stored/replicated data. In case the above monitoring network is to be linked to a mobile network of PDAs (e.g., carried by firemen), an adaptive (*CARD* + mobility-assisted) protocol may be used efficiently. The network may also be tied to a fire fighting management center (possibly through the Internet) using M&M.

2. RELATIONSHIP TO OTHER RESEARCH & PRACTICE

- Related Work

In wireless ad hoc networks perhaps the simplest form of resource discovery is global *flooding*. This scheme does not scale well. It also uses broadcast, which is usually unreliable at the data link layer (e.g., in 802.11). The synchronization of the broadcasts may lead to severe collisions and medium contention. Hence, it is our design goal to avoid global flooding. Expanding ring search uses repeated flooding with incremental TTL. This approach also does not scale well. Much of the work on routing protocols in ad hoc networks uses some form of flooding or ring search (e.g., DSR [26], AODV [27], ODMRP [34]).

Other approaches in ad hoc networks that address scalability employ hierarchical schemes based on clusters or landmarks (e.g., LANMAR [46], SCOUT [45] and [47]). These architectures, however, require complex coordination between nodes, and are susceptible to major re-configuration (e.g., adoption, re-election schemes) due to mobility or failure of the cluster-head or landmark. Furthermore, usually the cluster-head becomes a bottleneck. Hence, in general we avoid the use of complex coordination schemes for hierarchy formation, and we avoid using cluster-heads.

In the GLS architecture [28], all nodes know a *grid* map of the network. Nodes recruit location servers to maintain their location. Nodes update their location using an ID-based algorithm. Nodes looking for location of a specific ID use the same algorithm to reach a location server with updated information. This is a useful architecture when a node knows the network grid map, knows its own location, and knows the ID of the node it wishes to contact. Performance of such mechanism degrades with mobility as the location update messages are increased with every move. Furthermore, a source node may be looking for a target *resource* residing at a node with an ID *unknown* to the source node, in which case GLS fails.

The algorithm proposed in [48] and [49] uses global information about node locations to establish short cuts or friends, and uses geographic routing to reach the destination. Knowledge of the locations of all nodes over time under mobility conditions is infeasible. Also, the destination ID (and location) must be known in advance, which may not be the case in resource discovery.

In ZRP[4][29][30] the concept of hybrid routing is used, where table-driven routing is used intra-zone and on-demand routing is used inter-zone. Border-casting (flooding between borders) is used to discover inter-zone routes, which may not scale well. A good feature in ZRP is that a zone is node-specific. Hence, there is no complex coordination susceptible to mobility as in cluster-head approaches. We use the concept of zone in our architecture. However, we avoid border-casting by using contacts out-of-zone. The main concepts upon which contacts were designed (small world graphs) are fundamentally different than ZRP's bordercasting. We have compared the performance of ZRP and the contact-based approach through simulations. The contacts based approach incurs significantly lower overhead and has much more desirable scalability characteristics than ZRP.

In [5] an architecture based on intelligent agents is introduced for resource discovery in ad hoc networks. The concept of domains is used and global cluster-head election (using flooding) is needed to define a domain. This approach does not scale well in number of nodes in the network due to repeated global flooding. That architecture was not designed for a mobile network.

Our *mobility-assisted* contact selection scheme takes *advantage* of mobility, unlike any other previous work in this area. In this sense it is quite unique, and we expect that its performance *improve* with mobility, whereas all previous works clearly state that their performance degrades with mobility.

To the best of our knowledge, our rendezvous region scheme (RRs) described above, is the first to address explicitly the issues of service bootstrapping (e.g., multicast) in large-scale ad hoc networks. The idea of rendezvous points (RPs) was proposed (partially by the PI) for PIM-SM[8] for multicast in the Internet. However, discovery and management of the RP is fundamentally different in ad hoc networks. First, a single RP may fail and hence we introduce multiple RPs. Second, to avoid flooding RP information (as done in PIM-SM) we instead use a mapping from group prefixes into rendezvous regions. Third, mobility of the RPs may result in their migration from the respective RRs. This necessitates a dynamic election mechanism that we designed for our RR architecture.

Comparison of our M&M architecture to other micro-mobility techniques has already been presented in the main body of the document. To clarify further, M&M achieves handoff performance comparable to the best existing micro-mobility protocols. Furthermore, the fact that the multicast maybe sent to multiple (more than two) access routers, allows us to perform efficient reactive handoff (where a mobile node moves out of coverage and re-connects to a base station). Other micro-mobility approaches may allow bi-casting (sending to two base stations at a time), but this requires apriori knowledge of the movement, or connectivity to both base stations. Hence, reactive handoff cannot be handled well using bi-casting . We argue that allowing existing micro-mobility approaches to send to more than two base stations requires re-design of those protocols to re-invent what multicast has already established [8], in terms of loop duplication and black hole prevention mechanisms, among others. In addition, M&M naturally supports multicast traffic to the mobile nodes.

For object tracking, in SCOUT [45] an architecture was presented that is based on hierarchy formation. Using concepts borrowed from landmark hierarchy [44], where wireless devices self-configure in a multi-level hierarchy of parent nodes and children nodes. Each level is associated with a radius to which the device advertises itself. To configure the hierarchy complex mechanisms for promotion, demotion, and adoption are used. These mechanisms are susceptible to major re-configurations with mobility. This is mentioned clearly in their work. The root nodes of the hierarchy use global flooding to send advertisements. These advertisements are sent periodically. If the root nodes fail or move, new root nodes may be elected, and all nodes in the network may need to re-map all tracked objects. This does not scale well under dynamic conditions. The work presents two schemes that may be supported by the hierarchy, called SCOUT-AGG and SCOUT-MAP. In SCOUT-AGG object information is aggregated up the hierarchy. Queries travel up the hierarchy tree until the object is found. This query scheme may degenerate to flooding if the object summaries in many devices in the hierarchy indicate that they may have the object. SCOUT-MAP uses indirection through a device locator. A hashing scheme is used to route to the device locator (which has information about the tracking device). The hash depends on the number and identity of children of each device in the hierarchy that will be involved in this routing. Hence, a change in the number or identity of children for any of the en-route devices will cause re-hashing. Children for any devices at any level often change with mobility. A re-hash of the objects and their device locators is needed with mobility, in addition to re-configuration of the hierarchy. The performance of the SCOUT architecture degrades drastically with node mobility and network dynamics.

- Complementary Research

Advances in technologies and architectures that enable low-power, low-cost computing and communications will significantly help the realization of large-scale ad hoc networks that we are addressing in this research. In addition, research on mobility modeling [18] will provide richer evaluation and test scenarios for our protocols. Scalable simulation tools (e.g., NS-2 [13]) provide a powerful virtual environment to study design trade-offs and protocol performance. Research on small wireless devices that enable monitoring human health and vital signs, or those that enable monitoring temperature, wind, humidity or slope, will in turn enable applications similar to those we mention in this document. For example, the digital angel project [33] provides small devices (e.g., watches) for medical monitoring. However, they provide limited service (using specific cellular service provider) to track persons and objects. Coverage, however, is limited to that ISP's coverage. Ad hoc networks will potentially increase such coverage drastically; coverage that is crucially needed in cases of emergency. We plan to leverage digital angel's sensing and GPS capabilities when adequate. Also, for fire fighting the smart dust project at UC-Berkeley [32] is building devices that will enable monitoring forests for wild fires [31]. That project concentrates on the device technology while our project deals with protocol and architectural design. Hence, the two efforts are complementary.

Research on security in ad hoc networks (e.g., TESLA[41], BEBA[42], and Ariadne[40]) may prove to be useful in providing security for our protocols and architectures. TESLA is based on efficient symmetric cryptography, and does not use one-way functions that are expensive to compute. It can be used to provide secure authentication in point-to-point communication and broadcast communication. Ariadne is a protocol based on TESLA that provides secure ad hoc routing. It is a variant of DSR, and hence has similar shortcomings in terms of scaling. We shall leverage these protocols and investigate ways in which they may be applied to our architecture. One possible direction to design secure scalable ad hoc networks is to create a *small world of trust*. The main idea is to establish security relationships with selected nodes (such as the contacts used in our architecture), then use those relationships to establish further security relationships. For example, node *A* establishes a secure relationship with node *B* using TESLA. If node *B* already has a secure relationship with node *C*, then *A* can trust *C* by transitivity. Then the problem of establishing a secure relationship between two nodes becomes that of finding a chain of secure nodes leading to a node that both nodes consider as secure. In order to solve this problem efficiently we need to reduce the degrees of security separation between the nodes. This is achieved by using the small world concept using secure contacts, similar to what we described above. One possible problem is that establishing a secure relationship (using TESLA) requires synchronization between nodes, which is hard to achieve with distant nodes, especially if not all nodes are GPS capable. To overcome this problem, nodes start by establishing secure relationships with their direct neighbors, with which they may establish synchronization easily [43]. As those secure neighbors move away, they in turn establish secure relationships with new neighbors, and the small world of trust gets constructed (similar to the way we establish a small world of contacts above).

References

[1] A. Helmy, "*Small* Large-Large Scale Wireless Networks: Mobility-Assisted Resource Discovery", USC-TR, July 2002. (Submitted for Review) [http://ceng.usc.edu/~helmy]

[2] N. Nahata, P. Pamu, S. Garg, A. Helmy, "Efficient Resource Discovery for Large Scale Ad hoc Networks using Contacts", ACM SIGCOMM *Conference* (Refereed poster), August 2002. [http://ceng.usc.edu/~helmy]

[3] S. Garg, P. Pamu, N. Nahata, A. Helmy, "Contact Based Architecture for Resource Discovery (CARD) in Large Scale MANets", USC-TR, July 2002. (Submitted for Review). [http://ceng.usc.edu/~helmy]

[4] M. Pearlman, Z. Haas, "Determining the optimal configuration for the zone routing protocol", IEEE JSAC, p. 1395-1414, vol. 17, 8, Aug 1999.

[5] J. Liu, Q. Zhang, W. Zhu, J. Zhang, B. Li, "A Novel Framework for QoS-Aware Resource Discovery in Mobile Ad Hoc Networks", *IEEE ICC '02*.

[6] A. Helmy, "Architectural Framework for Large-Scale Multicast in Mobile Ad Hoc Networks", *IEEE ICC '02*. [http://ceng.usc.edu/~helmy]

[7] A. Helmy, "A Multicast-based Protocol for IP Mobility Support", ACM Second International ACM SIGCOMM Workshop on Networked Group Communication (NGC 2000), Palo Alto, November 2000. [http://ceng.usc.edu/~helmy]

[8] D. Estrin, D. Farinacci, A. Helmy, D. Thaler, S. Deering, V. Jacobson, M. Handley, C. Liu, P. Sharma, "Protocol Independent Multicast – Sparse Mode (PIM-SM): Protocol Specification", *RFC 2362/2117 of IETF/IDMR*, March '97/'98. [http://ceng.usc.edu/~helmy]

[9] C. Perkins, "IP Mobility Support", *RFC 2002, Internet Engineering Task Force*, October 1996.

[10] C. Perkins, D. Johnson, "Route Optimization in Mobile IP", *Internet Draft, Internet Engineering Task Force*, February 2000.

[11] C. Perkins and D. Johnson, "Mobility Support in IPv6", *Proceedings of MobiCom'96*, November 1996.

[12] D. Johnson, C. Perkins, "Mobility Support in IPv6", *Internet Draft, Internet Engineering Task Force*, March 2000.

[13] L. Breslau, D. Estrin, K. Fall, S. Floyd, J. Heidemann, A. Helmy, P. Huang, S. McCanne, K. Varadhan, Y. Xu, H. Yu, "Advances in Network Simulation", *IEEE Computer*, vol. 33, No. 5, p. 59-67, May 2000. [http://ceng.usc.edu/~helmy]

[14] A. T. Campbell and J. Gomez IP Micro-Mobility Protocols, ACM SIGMOBILE Mobile Computer and Communication Review (MC2R), 2001

[15] E. Gustafsson, A. Jonsson, C. Perkins, Mobile IP Regional Registration, Internet-draft, draft-ietf-mobileip-reg-tunnel-02, March 2000.

[16] A. Campbell, J. Gomez, S. Kim, A. Valko, C. Wan, Z. Turanyi, "Design, implementation, and evaluation of cellular IP" IEEE Personal Communications , Volume: 7 Issue: 4 , Page(s): 42 –49, Aug. 2000.

[17] R. Ramjee, T. La Porta, L. Salgarelli, S. Thuel, K. Varadhan, L. Li, "IP-based access network infrastructure for next-generation wireless data networks", IEEE Personal Communications, Volume: 7 Issue: 4, Page(s): 34 –41, Aug. 2000.

[18] F. Bai, N. Sadagopan, A. Helmy, "IMPORTANT: A framework to systematically analyze the Impact of Mobility on Performance of RouTing protocols for Adhoc NeTworks", *USC-CS-TR-02-765*, July 2002. (Submitted for Review). [http://ceng.usc.edu/~helmy]

[19] A. Helmy, "State Analysis and Aggregation Study for Multicast-based Micro Mobility", IEEE International Conference on Communications (ICC), May 2002. [http://ceng.usc.edu/~helmy]

[20] A. Helmy, and M. Jaseemuddin, "Efficient Micro-Mobility using Intra-domain Multicast-based Mechanisms (M&M)", *USC-CS-TR-01-747*, Aug 2001. [http://ceng.usc.edu/~helmy]

[21] A. Helmy, M. Jaseemuddin, Ganesha Bhaskara, "Efficient Micro-Mobility using Intra-domain Multicast-based Mechanisms (M&M)", July 2002. (Submitted for Review). [http://ceng.usc.edu/~helmy]

[22] Burton Bloom**,** "Space/time trade-offs in hash coding with allowable errors", *CACM*, 13(7):422-426, July 1970.

[23] M. Mitzenmacher, "Compressed Bloom Filters**",** *PODC*, 2001.

[24] J. Byers, J. Considine, M. Mitzenmacher, S. Rost, "Informed Content Delivery Across Adaptive Overlay Networks", *ACM SIGCOMM*, Aug. 2002.

[25] C.E. Perkins and P. Bhagwat, "Highly Dynamic Destination-Sequenced Distance-Vector Routing (DSDV) for Mobile Computers", Comp. Comm. Rev., Oct. 1994, pp.234-244.

[26] David B. Johnson, Davis A. Maltz, "The Dynamic Source Routing Protocol for Mobile Ad Hoc Networks" October 1999 IETF Draft.

[27] Charles E. Perkins, Elizabeth M. Royer, Samir R. Das, "Ad Hoc On-demand Distance Vector Routing", October 99 IETF Draft.

[28] J. Li, J. Jannotti, D. Couto, D. Karger, R. Morris, "A Scalable Location Service for Geographic Ad Hoc Routing (GLS/Grid)", ACM Mobicom 2000.

[29] Z. Haas, M. Pearlman, "The Zone Routing Protocol (ZRP) for Ad Hoc Networks", IETF Internet draft for the Manet group, June '99.

[30] Z. Haas, M. Pearlman, "The Performance of Query Control Schemes for the Zone Routing Protocol", ACM SIGCOMM '98.

[31] P. Eng, "Tiny Fire Marshals", abcnews, Sept. 13, 02. http://abcnews.go.com/sections/scitech/CuttingEdge/cuttingedge020913.html.

[32] Smart dust project, UC-Berkeley. http://robotics.eecs.berkeley.edu/~pister/SmartDust/

[33] Digital Angel Corporation, http://www.digitalangel.net.

[34] S. Lee, M. Gerla, C. Chiang, "On-demand multicast routing protocol", IEEE WCNC, p. 1298-1302, vol. 3, 1999.
[35] Y. Ko, N. Vaidya, "Location-aided routing (LAR) in mobile ad hoc networks", Wireless Networks 6, 4, p. 307-321, July 2000.
[36] B. Karp, H. Kung, "GPSR: Greedy Perimeter Stateless Routing for Wireless Networks", MobiCom 2000.
[37] S. Ratnaswamy, P. Francis, M. Handley, R. Karp, and S. Shenker. A scalable content- addressable network. ACM SIGCOMM, 2001.
[38] I. Stoica, R. Morris, D. Karger, F. Kaashoek, and H. Balakrishnan. Chord: A peer-to-peer lookup service for internet applications. ACM SIGCOMM, 2001.
[39] I. Clarke, O. Sandberg, B. Wiley, and T.W. Hong. Freenet: A distributed anonymous information storage and retrieval system in designing privacy enhancing technologies. Int'l Workshop on Design Issues in Anonymity and Unobservability, LNCS 2009, 2001.
[40] Y. Hu, A. Perrig, D. Johnson, "Ariadne: A Secure On-Demand Routing Protocol for Ad Hoc Networks", Mobicom 2001.
[41] Adrian Perrig, Ran Canetti, Dawn Song, and J. D. Tygar. Efficient and Secure Source Authentication for Multicast. In *Network and Distributed System Security Symposium, NDSS '0*1, pages 35–46, February 2001.
[42] Adrian Perrig, Ran Canetti, J.D. Tygar, and Dawn Song. Efficient Authentication and Signing of Multicast Streams over Lossy Channels. In *IEEE Symposium onSecurity and Privac*y, pages 56–73, May 2000.
[43] J. Elson, L. Girod, D. Estrin, "Fine-Grained Network Time Synchronization using Reference Broadcasts", Fifth Symposium on Operating Systems Design and Implementation (OSDI) 2002.
[44] P. F. Tsuchiya, "The Landmark Hierarchy: A new hierarchy for routing in very large networks", CCR, Vol. 18, no. 4, pp. 35-42, Aug. 1988.
[45] S. Kumar, C. Alaettinoglu, D. Estrin, "SCOUT: Scalable object tracking through unattended techniques", ICNP 2000.
[46] P. Guangyu, M. Gerla, X. Hong, "LANMAR: landmark routing for large scale wireless ad hoc networks with group mobility", MobiHOC '00, p. 11-18, 2000.
[47] B. Das, V. Bharaghavan, "Routing in ad-hoc networks using minimum connected dominating sets", *in Proc. IEEE ICC '97.*
[48] L. Blazevic, S. Giordano, J.-Y. Le Boudec "Anchored Path Discovery in Terminode Routing". *Proceedings of the Second IFIP-TC6 Networking Conference (Networking 2002), Pisa, May 2002.*
[49] J. Kleinberg, "Navigating in a small world", *Nature*, 406, Aug. 2000.
[50] A. O'Neill, G. Tsirtsis, S. Corson, Edge Mobility Architecture, Internet-draft, draft-oneill-ema-02.txt, July 2000.