Improving Routing Scalability in Networks with Dynamic Substrates

Boris Drazic
Department of Electrical and Computer Engineering
University of Toronto
boris@comm.utoronto.ca

Jörg Liebeherr
Department of Electrical and Computer Engineering
University of Toronto
jorg@comm.utoronto.ca

***Abstract*—We consider routing between large collections of interconnected networks, referred to as substrate networks, which do not assume permanent connectivity to the Internet, and which support dynamic changes of connectivity due to mobility. Whereas scalable routing schemes, such as compact routing or greedy forwarding, are suitable for very large networks, they generally ignore the routing methods already available in the substrate networks. In this paper, we present a routing scheme, referred to as *Landmark domains routing (LDR)*, which maximally exploits available routing in the substrate networks, and establishes paths between connected regions of substrate networks. We analyze the scheme by numerical analysis and simulation, and compare its performance with compact and greedy routing methods. We demonstrate that leveraging existing routing can lead to a significant reduction in the required routing state information, while providing paths that are, on average, close to the lengths of shortest paths.**

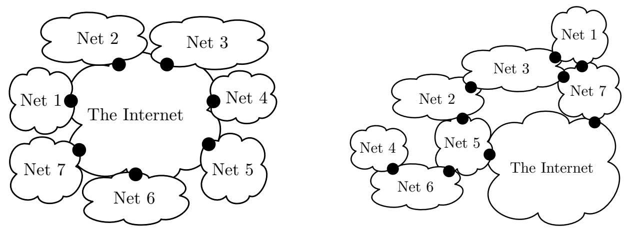
I. Introduction

The ever increasing number of networked devices and the emerging ability to connect to multiple communication networks using a range of modalities lays the foundation for a new internetworking architecture. Specialized networks, such as sensor networks, multi-hop vehicular networks, as well as countless private networks, create a collection of heterogeneous networks. Traditionally, these networks are assumed to be connected to the Internet, leading to a view of connectivity where the Internet is a central network, as illustrated in Fig. 1(a). All networks connect to the Internet, but not to each other. Here, routing between networks is built on top of that for the Internet, and can be reduced to the problem of tunneling messages through the Internet, e.g., [1]–[5]. However, in an environment where many networks are mobile and have only intermittent connectivity, the assumption of permanent connectivity to the Internet can be limiting. This suggests an alternative design, where the Internet is connected to some but not all networks, as illustrated in Fig. 1(b). As examples of such a design, FARA [6] and Plutarch [7] present internetworking architectures that operate on a collection of networks without assuming an inherent hierarchy or a central entity. More recently, concepts of a non-centralized architecture have been studied, e.g., in [8], [9], however, the question of routing, i.e., the computation of paths across multiple networks, is rarely addressed, with notable exceptions. For example, Pathlet routing [10] devises a packet forwarding method for arbitrarily connected networks with small forwarding tables. However, it does not address how routes are computed. Another example is the SpoVNet project [11], [12], which addresses the construction of overlay networks across heterogeneous collections of networks.

In this paper we are concerned with routing schemes for very large collections of networks, as in Fig. 1(b), where we refer to each individual network in the collection as a *substrate network*. We seek to support a virtually unlimited number of substrate networks of arbitrary sizes, with dynamic changes due to node mobility. If one views the internetwork as a single (flat) network of nodes, a routing scheme could be drawn from any of the available scalable routing methods, e.g., compact routing. However, this would ignore the routing methods already available in the substrate networks. The main finding of our research is that exploiting routing available in the substrate networks can significantly improve the scalability of a global routing architecture. In fact, a relatively simple routing scheme that provides paths between substrates is sufficient to generally achieve better performance than significantly more complex routing methods ignoring the available routing protocols.

In a network with n nodes, shortest path routing requires state information that scales linearly in the number of nodes, i.e., the state stored at a node is proportional to $O(n)$. Routing schemes are said to be scalable if the amount of required state information grows slower than linearly with n. Approaches to reduce the amount of state information by increasing the length of routing paths were first presented by Kleinrock and Kamoun [13] as *hierarchical routing*. The authors show that in a network with n nodes the minimum number of routing entries at a node is $e \ln(n)$, with path lengths approaching that of shortest paths as $n \to \infty$. The trade-off between the required state information and path lengths is studied under the umbrella of *compact routing*. The increase in the path length produced by a routing algorithm compared to the shortest path is referred to as *path stretch*. Gavoille and Gengler [14] show that all routing schemes with path stretch strictly below three need at least $\Omega(n)$ state at each node, while Thorup and Zwick [15] show that at least $\Omega(\sqrt{n})$ state at each node is necessary for a path stretch below five. They present a compact routing scheme [16] with path stretch at most three and state information that is bounded by $O(\sqrt{n \log n})$. Disco [17] is a compact routing scheme for general mobile networks, which bounds the state at a node by $O(r_S \sqrt{n \log(n)})$, where r_S is the maximal size of a source route. Compact routing performed on a collection of substrate networks (as in Fig. 1(b)) does not account for the paths that are already available within the substrate networks. We will show that leveraging such existing paths can significantly reduce the required state information at nodes.

An alternative approach to routing is *greedy routing*, where



(a) Internet as the central network. (b) No central network.

Fig. 1: Interconnected collection of (substrate) networks.

nodes are assigned coordinates, and messages are forwarded to the neighbor node whose coordinates are closest to that of the destination. Coordinates can consist of geographical locations (*geographical routing*) or can be derived from a virtual coordinate system, e.g., [18]–[21]. With greedy routing, the state information at nodes is bounded by the maximum node degree. However, greedy routing may run into *local minima*, which occur when a node does not have a neighbor that is closer to the destination than itself. Structured overlay networks, e.g., [22]–[24], can be viewed as performing greedy routing, however, they generally assume that all nodes are located in the same substrate network (the Internet). By using virtual multi-hop links, structured overlays can be adapted to provide routing in general interconnected networks, e.g., UIP [25] and VRR [26]. These schemes require state information bounded by $O(\log(n))$ and bound the path stretch in the number of overlay links, usually also $O(\log(n))$. Since overlay links may be realized by virtual paths of nodes, the actual path stretch can be significantly larger. UIP [25] adds heuristics to reduce the path stretch by searching for nodes located in the same substrate.

Our objective is to devise a routing scheme for a collection of substrate networks that takes maximal advantage of the available routing in the substrate networks. When considering substrate networks with mobile nodes, substrate networks may become partitioned and routing paths in a substrate network may become unavailable. We propose a mechanism by which nodes belonging to the same substrate network self-organize into maximal groups, where routing paths exists, and refer to these groups as *reachability domains*. Then, the problem of routing is reduced to that of creating routing paths across reachability domains. Our approach of detecting reachability domains is also pursued in SpoVNet [11], [12].

We seek to explore the benefits of a routing scheme, referred to as *landmark domains routing*, which fully leverages the existing routing available in substrate networks. Landmark domains routing sets up paths from all reachability domains to a subset of globally known reachability domains, called landmark domains. The established routing paths first forward messages to a landmark domain, and from the landmark domain to the destination, where the paths from landmark domains to nodes are realized by source routes, which are carried in the locator addresses of nodes. Through numerical analysis we show that this method generally requires less state information at nodes than scalable routing schemes, which ignore the routing paths available within substrate networks. We also show that paths set up by landmark domains routing are, on average, close to shortest paths. We present simulation experiments that evaluate the overhead of maintaining reachability domains and routing paths, and compare the overhead to that of a compact routing scheme (Disco [17]) and two greedy routing schemes (UIP [25] and VRR [26]).

Remark: If one views the Internet as a multi-substrate network, where autonomous systems are substrate networks, then intra-domain routing corresponds to routing within a substrate, and inter-domain routing corresponds to routing between substrates. In this interpretation, the proposed landmark domains routing scheme can be viewed as an inter-domain routing protocol that is suitable when subnets have frequent partitions. However, in this paper, we do not pursue this interpretation. Moreover, we do not address inter-domain routing issues, such as routing policies or security considerations.

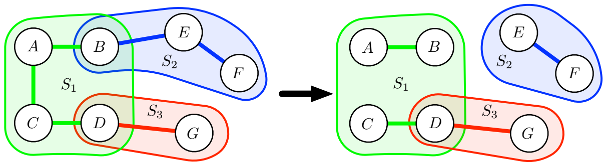
The rest of the paper is organized as follows: In Sec. II, we introduce some concepts of multi-substrate routing. In Sec. III, we provide the details of landmark domains routing. Sec. IV discusses numerical analysis of state information and path stretch. In Sec. V, we present results of simulations. We provide conclusions in Sec. VI.

II. Concepts of Multi-substrate Routing

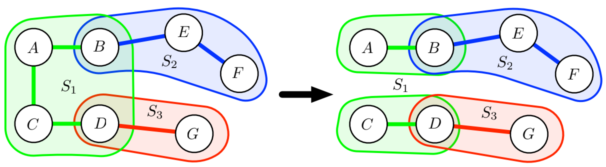
We have referred to constituting networks of the internetwork in Fig. 1(b) as substrate networks. In particular, the Internet is viewed merely as a single substrate network. Substrate networks may operate at the data link, network, transport, or even application layer. The global network, with all substrate networks, is referred to as a multi-substrate network. In the following we provide a more precise definition of substrate networks, partitions of substrate networks, and other relevant terminology and concepts.

We refer to nodes as network elements that perform forwarding of messages. Each node has one or more network attachment points (network interfaces). Network interfaces can be physical or logical and include, WiFi, Bluetooth, or IP interfaces, but also TCP server ports, or application-defined interfaces. We refer to attachment points of the same type, which implement the same protocol and use the same configuration, as *compatible attachment points*. For example, WiFi attachment points implementing IEEE 802.11a are compatible if they use the same service set identifier (SSID), and IP interfaces are compatible if they use addresses from the same IP address space. Communication is enabled when compatible attachment points are linked by a bidirectional communication channel. The communication channel may be a point-to-point link, a shared broadcast link, a switched network, or an internetwork. We refer, in general, to communication channels as *links*, and assume that all links are bidirectional. When nodes are connected by a link, we say that the nodes are *one-hop connected*. Nodes that are one-hop connected are referred to as *neighbors*.

We define a *substrate network* as a collection of nodes with compatible attachment points and links between one-hop connected nodes. A node with two or more attachment points connected to different substrate networks is called a *multi-homed* node. Multi-homed nodes connect substrate networks. We refer to a *multi-substrate network* as a collection of connected substrate networks.



(a) External change of connectivity.



(b) Internal change of connectivity.

Fig. 2: Dynamic substrates.

We refer to *routing* as the ability to set up multi-hop paths between nodes, where adjacent nodes on the path are one-hop connected, and we use the terms routing scheme or routing protocol to refer to the ability to set up multi-hop paths. We refer to *intra-substrate routing* as the routing within a substrate network. Examples of intra-substrate routing are the Spanning Tree Protocol (STP) in a switched Ethernet network, or Open Shortest Path First (OSPF) in a collection of IP subnets. Some substrate networks may not provide intra-substrate routing, for example, a collection of WiFi nodes in ad-hoc mode. In such substrate networks, a node can communicate only with its neighbors.

Substrate networks that include mobile wireless nodes may become partitioned as a result of node mobility, while, in wireline substrate networks, a partitioned network is generally a result of a failure. To account for node mobility, we consider substrate networks that are dynamic. *Dynamic substrate networks* are characterized by exhibiting two types of changes of connectivity between nodes: (1) *external* changes to the connectivity between substrate networks, and (2) *internal* changes to the availability of paths between nodes in the same substrate network. In Fig. 2(a), we show an external change of connectivity for a multi-substrate network composed of three substrate networks (S_1, S_2, and S_3). Initially, node B is attached to substrate networks S_1 and S_2, creating a connection between S_1 and S_2. After node B loses its attachment to substrate network S_2, substrate networks S_1 and S_2 are no longer connected. Internal changes of connectivity occur due to changes of connectivity between nodes within the same substrate network, and may cause partitions. A substrate network is *partitioned* if there are nodes that cannot be connected by a path of one-hop connected nodes from the substrate network. Fig. 2(b) depicts an internal change of connectivity in substrate network S_1, when one-hop connectivity between nodes A and C is lost. Substrate S_1 is partitioned since there is no end-to-end path from nodes A and B to nodes C and D.

In a partitioned substrate network, intra-substrate routing creates one routing domain for each partition. We refer to the partitions as *reachability domains*. (They are called *connectivity domains* in [11]). A reachability domain, or domain for short, is a maximal set of nodes, attached to the same substrate network, such that a path exists between every pair of nodes in the set. Two nodes are members of the same reachability domain if: (a) they are attached to the same substrate network; and (b) if they can exchange messages in this substrate network (possibly using intra-substrate routing). Fig. 3 shows a substrate network S_1 with nodes A through I, where links between nodes indicate one-hop connectivity. Assuming that the substrate network provides intra-substrate routing, the set of nodes is grouped into two reachability domains, labeled as D_1 and D_2. D_1 consists of nodes A through F, and D_2 contains nodes G, H, and I. Since there is no path in S_1 between nodes from D_1 and D_2, two separate reachability domains are formed. If a substrate network is not partitioned, all nodes in the substrate network belong to the same reachability domain.

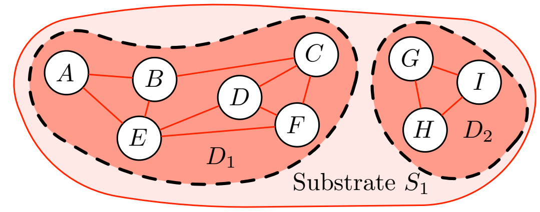


Fig. 3: Reachability domains.

III. LANDMARK DOMAINS ROUTING (LDR)

Given a collection of connected substrate networks as in Fig. 1(b), we are concerned with providing a routing scheme that establishes multi-hop paths between all nodes in the multi-substrate network, by maximally taking advantage of intra-substrate routing. We propose a routing scheme, referred to as *landmark domains routing (LDR)*, which provides end-to-end paths across the reachability domains of a multi-substrate network. A subset of reachability domains is designated as landmark domains. All nodes store information about next-hop nodes on paths to the landmark domains. *Locator addresses* of nodes specify a source route from a landmark domain. Using locators to denote destinations, LDR forwards messages in three phases: (1) forwarding to a landmark domain; (2) forwarding inside the landmark domain; and (3) forwarding along the source route in the locator address.

LDR adopts concepts from compact routing, but applies them at the level of reachability domains, as opposed to individual nodes. LDR benefits from the low amount of state information of a compact routing scheme. As we show in Sec. IV and Sec. V, LDR achieves a path stretch that is comparable to that of a pure compact routing scheme.

In the remainder of this section, we present details of the routing scheme. First, we discuss the management of reachability domains, then we discuss landmark domains and locators, and, finally, we present details of message forwarding.

A. Management of Reachability Domains

With dynamic substrates, membership of nodes in the same substrate network does not guarantee that intra-substrate routing is sufficient to route messages between these nodes. To ensure that intra-substrate routing can deliver messages between nodes they must belong to the same reachability domain. Thus, to fully utilize the available routing schemes within substrate networks, mechanisms are required for the discovery and maintenance of reachability domains. In LDR, nodes self-organize into reachability domains, and jointly determine identifiers for the domains.

The membership of nodes in reachability domains is detected through *domain discovery* messages, which are disseminated throughout a substrate network. Recipients of a domain discovery message have a path to the sender, and, therefore, belong to the same reachability domain as the sender. If available in a substrate network, domain discovery messages are transmitted by broadcast. If broadcast is not available, LDR uses structured overlay networks, e.g., Chord [22], for dissemination. Note that this requires that each node can rendezvous with an existing overlay, e.g., using configured address lists, servers, or out-of-band communication. When overlay networks exclusively use paths available through intra-substrate routing, an overlay network will contain all nodes in the same reachability domain of that substrate. We emphasize that the overlay networks are used only to disseminate control messages, and are not used for message forwarding.

For the discovery of a reachability domain for a particular substrate network, each node selects a random number. The node disseminates the number, in the overlay network available for the substrate, if it is larger than any number recently received on this overlay. The largest number becomes the identifier of the reachability domain, and its sender becomes the *leader node* for the domain. A leader node periodically sends a domain discovery message with the domain identifier to member nodes, using either a broadcast transmission or the described structured overlay network. Nodes record the last time they received the domain identifier. A node that has not recently received the domain identifier concludes it is no longer in the same domain as the leader node. A node may not receive the identifier from the leader for two reasons: (1) the leader node has failed or is no longer active, or (2) an intra-substrate path is no longer available to the leader node. In both cases, a new reachability domain is formed with a new identifier. Nodes without a leader assume the role of a leader node, and disseminate newly selected random numbers. The node with the largest number will emerge as the new leader, and its number will become the identifier of the reachability domain.

Reachability domains merge when a path between nodes from two different domains in the same substrate network becomes available. When using overlay networks, nodes must merge the overlay networks of the merged domains into a single overlay network, and the leader with the larger identifier becomes the leader of the merged domain. Details of the interactions of a node with the overlay networks of a reachability domain, in particular, the merging of overlay network using the Chord protocol, are described in [27].

B. Landmark Domains and Locators

We define *landmark domains* as a subset of reachability domains. A domain independently decides to become a landmark domain, based on the number of domains in the multi-substrate network, which is estimated using techniques from [28]. In a network with k reachability domains, the number of landmark domains can be selected as $\sqrt{k \log k}$, analogous to compact routing (which has $\sqrt{n \log n}$ landmark nodes in a network with n nodes). We will observe that even a much smaller number of landmark domains can result in good path stretch performance (see Sec. IV and Sec. V). Nodes participate in a routing protocol to discover paths from all domains to landmark domains, where the distance is measured in the number of intermediate domains. Every node has a *landmark domains routing table*, which stores next-hop nodes and distances to every landmark domain. Nodes in a domain, using the overlay network, exchange routing information to agree on shortest paths from the domain to landmark domains. Routing information is propagated to neighboring domains via multi-homed nodes.

As stated earlier, LDR uses source routes from landmark domains as locator addresses. Each node has a single locator, with a source route that originates at the landmark domain closest to the node. A node discovers its locator address by forwarding a locator discovery message to a landmark domain and recording the path of the message. If all links are bidirectional, the source route from a landmark domain to a node is the reverse path traversed by the discovery message. The locator of a node N, with respect to a landmark domain $\mathcal{L}$ is $\{id(\mathcal{L}) : sr(\mathcal{L}, N)\}$, where $id(\mathcal{L})$ is the domain identifier of $\mathcal{L}$ and $sr(\mathcal{L}, N)$ is a source route from some node in $\mathcal{L}$ to node N. The format of a source route $sr(\mathcal{L}, N)$ is

$$\left\langle id(D^{(1)}), sa(N^{(1)}) \right\rangle, \ \dots \ , \left\langle id(D^{(m)}), sa(N^{(m)}) \right\rangle,$$

where superscripts indicate the position in the source route, and $sa(N^{(i)})$ refers to the substrate address of node $N^{(i)}$ in the substrate network of $D^{(i)}$. Further, for $1 \leq i < m$, node $N^{(i)}$ is a multi-homed node, which is also a member of reachability domain $D^{(i+1)}$, i.e., node $N^{(i)}$ can send a message to node $N^{(i+1)}$ in reachability domain $D^{(i+1)}$. The first domain in the source route is the landmark domain $\mathcal{L}$ and the last node in the source route is node N, i.e., $D^{(1)} = \mathcal{L}$ and $N^{(m)} = N$. Note that, by construction of reachability domains, any node in $D^{(i)}$ can send a message to node $N^{(i)}$ using the substrate address $sa(N^{(i)})$.

Nodes are responsible for creating their own locator addresses, which is initiated by sending a message to the closest landmark domain. Each time the source route to a node changes, so does its locator address. A modified locator address triggers an update in the mapping of node identifiers to node locators.

Note that source routes in LDR only require one node for each traversed reachability domain. Specifically, node $N^{(i)}$ is the last node (egress node) in domain $D^{(i)}$ that is traversed on the path from the landmark domain to node N. Such routes leverage existing intra-substrate routing: there is no need to record a path between nodes if a path is already available from intra-substrate routing. This mitigates the problem of long packet headers associated with long source routes.

C. Message Forwarding

Message forwarding in LDR proceeds similarly as in compact routing, with the key difference that forwarding exploits intra-substrate routing to traverse reachability domains. A message is first forwarded towards the landmark domain identified in the locator address of the destination. This is the

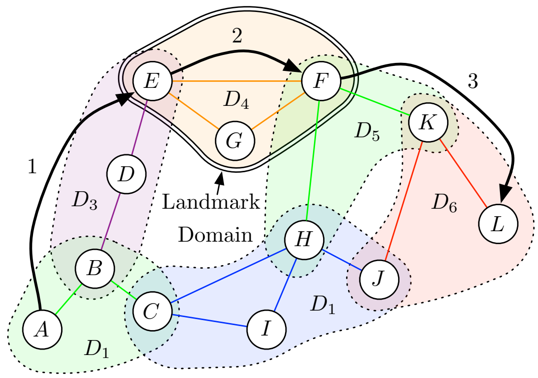


Fig. 4: Message forwarding in LDR.

landmark domain with the smallest distance to the destination. From the landmark domain, the message is forwarded on the shortest path to the destination, using the source route.

Additional heuristics, inspired by the landmark hierarchy [29], may be used to reduce the length of paths of LDR. For example, in LDR, any node may forward a message to the hindmost node in the source route, which is in the same reachability domain. Short cuts created in this fashion, may not require to forward a message to a landmark domain.

We illustrate the three phases of message forwarding of LDR in Fig. 4, for a message sent from node A to node L. The figure shows a multi-substrate network with six domains, where D_4 is a landmark domain. The locator of node L is:

$$\{id(D_4) : (\langle id(D_4), sa_4(F)\rangle, \langle id(D_5), sa_5(K)\rangle, \langle id(D_6), sa_6(L)\rangle)\} .$$

The forwarding of a message proceeds as follows.

1) **Next-hop forwarding to the landmark domain.** Node A makes a lookup in the landmark domains routing table, and obtains the address $sa_1(B)$ of the next-hop node B, on the path to landmark domain D_4. Using $sa_1(B)$, the message is delivered to node B. Node B forwards the message to node E, which is located in the landmark domain.

2) **Forwarding inside the landmark domain.** As ingress node to the landmark domain, node E forwards the message to the first node listed in the locator address of L, using intra-substrate routing available in D_4. This delivers the message to node F.

3) **Forwarding on the source route.** From the source route, node F obtains the address $sa_5(K)$ of the next node on the path towards the destination, and sends the message to node K, which forwards the message to the destination L. Messages are forwarded in D_5 and D_6 using intra-substrate routing.

IV. Numerical Analysis

In this section, we explore the inherent trade-offs and scalability properties of landmark domains routing. As a broader question, we wish to quantify the benefits of exploiting available routing schemes in substrates (intra-substrate routing) for the creation of a global routing scheme. We consider the following performance metrics:

- *State information* - the amount of information stored at each node by a routing protocol, i.e., the size of routing tables; and
- *Path stretch* - the ratio of the length of a path set up by a routing protocol and the shortest path between two nodes.

These two measures present a principal trade-off of a routing protocol, since state information can be reduced at the cost of increased path stretch. In this section, we assume a static network, where each substrate consists of one reachability domain. In Sec. V, we use simulation experiments to study networks with mobile nodes and dynamic substrates. To keep the parameter set to a moderate level, we assume that all substrate networks have equal size, and that all nodes have the same number of interfaces.

A. State Information

We consider a network with n nodes, each with n_I interfaces, which are distributed evenly over s substrate networks.

A shortest path routing protocol requires that each node has a routing table entry for each possible destination. Therefore, the state information of a node with shortest path routing (S_e^{SP}), measured in the number of entries, is given by

$$S_e^{SP}(n) = n .$$

Compact routing schemes, e.g., [16], [17], often bound the number of landmark nodes by $O(\sqrt{n \log(n)})$. Each node stores one routing table entry for each landmark node and one routing table entry for the at most $O(\sqrt{n \log(n)})$ closest nodes. Assuming a constant of $K > 0$, the state information at a node for a compact routing scheme (S_e^{CR}), measured in the number of entries, is

$$S_e^{CR}(n) = 2K\sqrt{n \log(n)} .$$

In our numerical examples, we set $K = 1$.

For landmark domains routing, the state information of a node consists of: (1) the landmark domains routing table, storing a next-hop entry for each landmark domain; (2) entries for reachability domain identifiers and last times they were received from the leader node, with one entry for each interface; (3) state information needed to maintain the overlay network of a reachability domain; and (4) routing entries used for routing inside the substrate. The last component depends on the routing method used inside the substrate networks. For a comparison with compact routing, we set the number of landmark domains of LDR in a network with s substrates to $\sqrt{s \log(s)}$. In this way, we use the same criteria for the number of landmarks, with the difference that landmarks in LDR are reachability domains, while they are nodes in compact routing. In particular, when all substrate networks have only one node, the number of nodes and domains is equal. Hence, the number of landmark domains in LDR is equal to the number of landmark nodes of compact routing. For maintenance of each structured overlay network, a node must keep a small, constant, number of entries for finger nodes (n_m) and alternate successor nodes (r). We use the common choices $n_m = 7$, which is sufficient for 128-bit identifiers, and $r = 2$, which provides some robustness against failure of successor nodes. The number of entries for routing between the n/s nodes in a substrate network depends on the intra-substrate routing scheme. We denote the state information requirement by S_e^{Sub}, where we consider both shortest-path routing as well

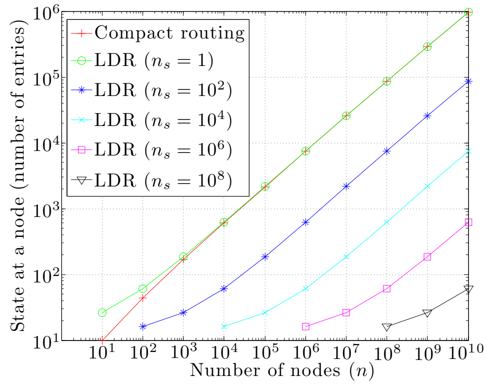


Fig. 5: State information (without intra-substrate routing).

as compact routing. Then, the stored state at a node in LDR, denoted by S_e^{LDR} and measured in the number of entries, is given by

$$S_e^{LDR}(n, s) = \sqrt{s \log(s)} + n_I \left(1 + n_m + r + S_e^{Sub}(n/s)\right) .$$

We first compare LDR with compact routing, without accounting for intra-substrate routing within the reachability domains (that is, $S_e^{Sub}(n/s) = 0$). This allows us to compare the scaling properties of LDR in isolation. Under the above assumption, in Fig. 5, we show the state information at each node as a function of the total number of nodes n. For LDR, the state information is shown for different substrate network sizes $n_s := n/s$. This is compared to the state information of compact routing (S_e^{CR}). We observe that the state information of LDR grows slower than linear with increasing number of nodes, regardless of the size of substrate networks. The larger the substrate networks, the smaller is the required state information. When each substrate network contains only a single node ($n_s = 1$), the state information with LDR and compact routing are similar. We conclude that the scaling of LDR with respect to stored state information is similar to that of compact routing.

Note that Fig. 5 does not account for the state information required to perform routing within a substrate. This is addressed now in a comparison of routing schemes that do and do not exploit substrate routing.

For methods without consideration of substrates, that is, without exploiting intra-substrate routing, we consider shortest path routing (SP) and compact routing (CR). We refer to these schemes as global schemes, since they compute routing paths between all nodes in the multi-substrate network. For routing schemes that exploit intra-substrate routing we consider several variations:

	Routing between substrates	Routing within substrates
SP+SP	shortest path	shortest path
LDR+SP	LDR	shortest path
LDR+CR	LDR	compact routing

SP+SP uses shortest path routing within substrate networks and between substrates networks. The required state information is therefore

$$S_e^{SP+SP}(n) = s + n_I(n/s) .$$

LDR+SP uses the LDR scheme between substrates and shortest path routing within each substrate. Here, the state in-

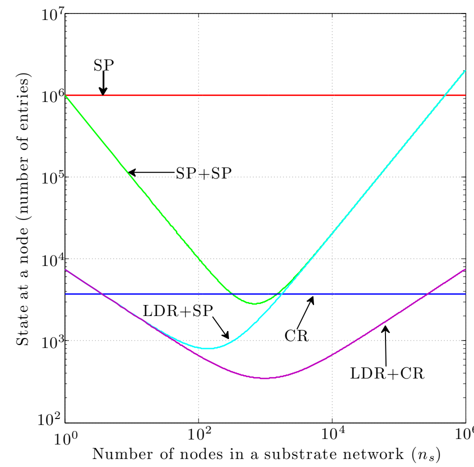


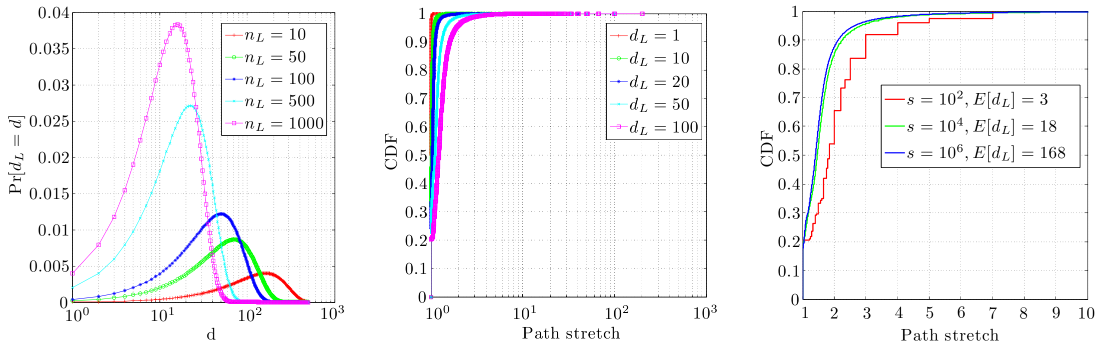
Fig. 6: State information (with intra-substrate routing) with $n = 10^6$ nodes.

formation is given by $S_e^{LDR}(n, s)$, where we use $S_e^{Sub}(n/s) = S_e^{SP}(n/s)$. Likewise, for LDR+CR we set $S_e^{Sub}(n/s) = S_e^{CR}(n/s)$. We analyze a multi-substrate network with $n = 10^6$ nodes, where we vary the number of substrate networks s. In Fig. 6 we plot the total stored state at a node as a function of the number of nodes in a substrate network ($n_s = n/s$), using the previously derived expressions. Obviously, the global routing schemes are not sensitive to n_s, and provide a constant line. The values between SP and CR reflect the range achievable with a global routing scheme, since it is appropriate to view SP as worst-case, and CR as the ideal global routing scheme.

From Fig. 6, we observe that leveraging intra-substrate routing has significant benefits as long as substrate networks are not too small or too large. A comparison of SP+SP with LDR+SP and LDR+CR shows the benefits of using LDR for routing between substrate networks. We point out that the required state is reduced by up to an order of magnitude over a global compact routing approach. As the size of substrate networks increases, the state stored by shortest path routing in substrate networks dominates the total stored state, leading to SP+SP and LDR+SP having the same amount of stored state. If we accept that shortest path routing is an upper bound and compact routing is a lower bound with respect to the required state information, the two curves show the range of values achievable with different intra-substrate routing methods. When n_s is small ($n_s > 1$, otherwise the network is not connected), the substrate networks contain only few nodes and there is a small number of existing communication paths to leverage. Here, additional state needed to establish communication between the substrate networks outweighs the savings from using intra-substrate routing. When $n_s \to n$, the multi-substrate network contains only a few substrate networks and the number of nodes in a substrate network becomes comparable to the number of nodes in the multi-substrate network. Then nodes have to store almost the same amount of state for intra-substrate routing as a global routing scheme.

B. Path Stretch

We next address the path stretch achievable with LDR. A message forwarded by LDR is first sent to a landmark domain,



(a) $\Pr[d_L = d]$. (b) Sensitivity of path stretch to d_L. (c) Path stretch for $n_L = 1$.

Fig. 7: Numerical analysis of path stretch ($s = 10^6$ substrate networks).

and then from the landmark domain to the destination node. The landmark domain is selected by the destination node, since it is included in the locator of the destination. Recall that a node selects the closest landmark domain when creating its locator. In the worst case, the length of the shortest path from a node to the closest landmark domain is the diameter of the network. The largest path stretch occurs when source and destination are close to each other (but not able to route directly to each other). In this case, messages are forwarded to the landmark domain and back, thus incurring a path that can be twice as long as the diameter of the network.

For evaluation of path stretch, we consider a multi-substrate network as a grid of substrate networks of equal sizes, where each substrate network connects to four other substrate networks, i.e., directly above, below, left, and right. Since there is no substrate with higher connectivity offering more direct paths, the grid will provide larger diameter and longer paths than found in actual networks (for non-degenerated topologies). If we assume that shortest path routing is used within substrates, i.e., LDR+SP, then, as long as the number of substrate networks is large, and each substrate has the same size, the path stretch can be obtained from the number of traversed substrate networks. We refer to the traversal of one substrate network as a *hop*.

We pick one node as the destination, and refer to its reachability domain as the destination domain. We then compute the path stretch from all other domains to the destination domain. Landmark domains are randomly uniformly distributed across the substrate networks. If the network is sufficiently large, we may assume that the destination domain is located in the center of the grid. Let d_L be the random variable denoting the number of hops from the destination domain to the closest landmark domain. If there is only one landmark domain, the probability that it is more than d hops away from the destination domain is given by

$$\Pr[d_L > d] = 1 - \frac{1 + 4\sum_{i=1}^{d} i}{s} ,$$

where $\left(1+4\sum_{i=1}^{d} i\right)$ is the number of domains that are at most d hops away from the destination domain, and s is the total number of domains. For n_L landmark domains, $\Pr[d_L > d]$ is given by

$$\Pr[d_L > d] = \left(1 - \frac{1 + 4\sum_{i=1}^{d} i}{s}\right)^{n_L} .$$

In Fig. 7(a), we present the distribution of d_L for a multi-substrate network with 10^6 substrate networks, for different values of the number of landmark domains (n_L). The figure shows how, as the number of landmark domains is increased, it is more likely that the destination domain is close to a landmark domain. We next investigate the sensitivity of path stretch to the minimum distance to a landmark domain (d_L), by computing the path stretch from all domains to the destination domain, for different values of d_L. In Fig. 7(b), for a network with $s = 10^6$ substrate networks, we show the CDF of the path stretch for routes to the destination domain. For all values of d_L, the path stretch is small. The average path stretch is about 1.3 for $d_L = 100$, and decreases to 1.0028 when $d_L = 1$. We conclude that the path stretch is not very sensitive to d_L, and even a large distance between a destination domain and its landmark domain results in a short path stretch.

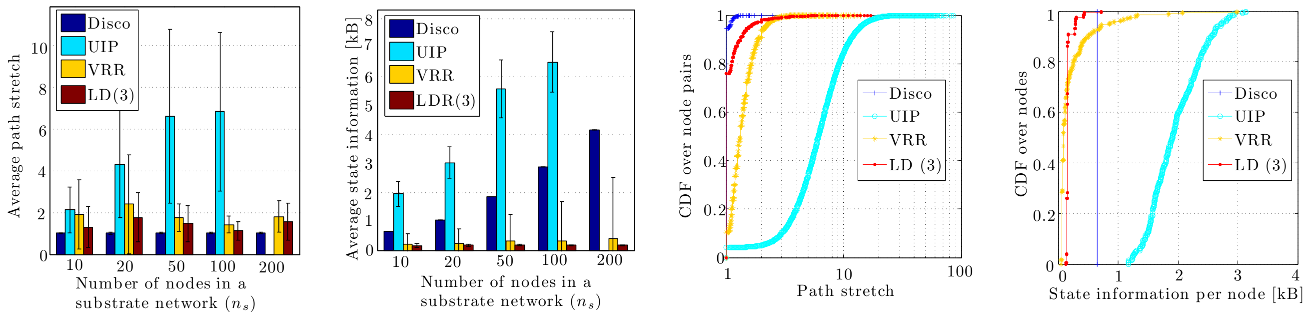
We next evaluate a scenario with only one landmark domain ($n_L = 1$). This is certainly a worst-case for the path stretch offered by LDR. Here, the number of substrate networks is varied with $s = 10^2$, 10^4, and 10^6. In Fig. 7(c), we show the CDF of the path stretch from all domains to the destination domain, via the single landmark domain. The legend of the shown plot includes the average minimum distance between the destination domain and the landmark domain, denoted by $E[d_L]$. Fig. 7(c) shows that even with only one landmark domain, the path stretch remains small. The average path stretch ranges from 1.58 (with $s = 10^6$) to 2.0439 (with $s = 10^2$).

V. Simulation Experiments

In this section, we present simulation experiments that compare LDR to protocols providing compact routing (Disco [17]) and greedy routing (UIP [25] and VRR [26]). These protocols are representative of the state-of-the-art of global routing schemes with sublinear scaling of state information. We developed a detailed packet-level simulation of these schemes for multi-substrate networks using OMNeT++ [30], which can support simulations with up to 5,000 nodes, including mobile nodes. Routing within substrate networks uses shortest path routing.

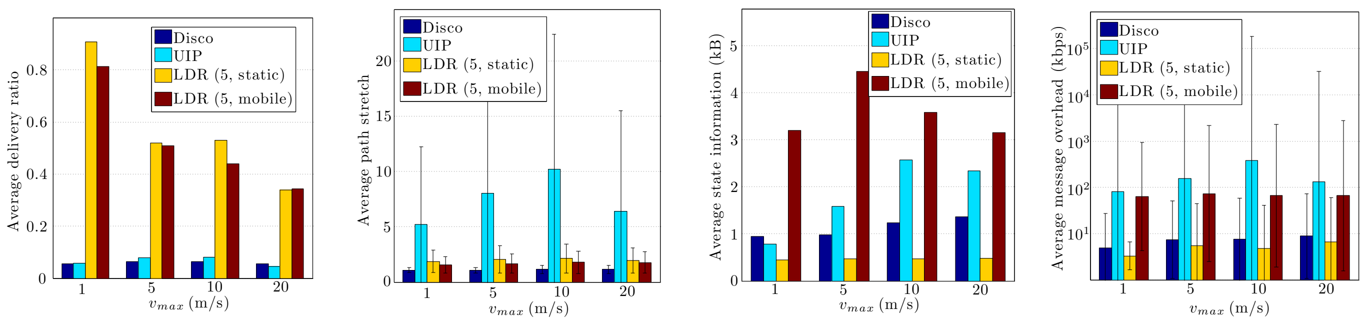
A. Fixed Network Topology

We first consider networks with a fixed topology. We randomly generate topologies for multi-substrate networks with node and substrate network degree governed by a power-law



(a) Average path stretch. (b) Average state information. (c) CDF of path stretch ($n_s = 100$). (d) CDF of state info. ($n_s = 100$).

Fig. 8: Static multi-substrate network ($\alpha = 2.2$, $\beta = 2.2$, $s = 25$).



(a) Average delivery ratio. (b) Average path stretch. (c) Average state information. (d) Average message overhead.

Fig. 9: Dynamic multi-substrate network.

probability distribution, with exponents α and β, respectively. The degree of a substrate network is given by the number of other substrate networks that are connected to the substrate network. We refer to [27] for details of the algorithms.

We generate multi-substrate networks, with parameters $\alpha = 2.2$, $\beta = 2.2$, and $s = 25$, which results in topologies with a moderate number of well connected substrate networks, and vary the number of nodes per substrate network (n_s). The total number of nodes is $n = s \times n_s$. All experiments are run on three network topologies, and performance metrics are presented as average results.

For LDR, we select three domains as landmark domains, which is well below $\sqrt{s \log(s)}$. The path stretch is based on measurements of $62,500$ randomly selected node pairs. We present the average values for path stretch in Fig. 8(a) and stored state in Fig. 8(b), where error bars show one standard deviation. Due to the large number of messages exchanged between nodes in UIP, we were unable to simulate UIP for n_s exceeding 100.

We observe in Fig. 8(a) that the path stretch in LDR is larger than in Disco, but smaller than that of the other schemes, and does not increase with n_s. In Fig. 8(b) we see that LDR requires the least state information for all simulated multi-substrate networks. The state information of VRR is similar to that of LDR, with VRR showing more variability. Note that the amount of stored state for Disco and UIP increases with n_s, but remains roughly constant in LDR.

For the same experiment setup, we show empirical cumulative distribution functions of the path stretch, in Fig. 8(c), and the state information, in Fig. 8(d), collected from three topologies generated with $n_s = 100$ nodes per substrate network and $n = 2,500$ total number of nodes. For LDR, 99.24% of paths have path stretch below 3, which is the bound on path stretch for Disco. LDR has on average the smallest amount of state information of all compared routing schemes, without major differences in the state at individual nodes.

B. Networks with Mobile Nodes

We present simulations of a multi-substrate network with dynamic substrates, by adding substrate networks with mobile wireless nodes to a static topology. We start by generating a network with static (wired) substrate networks as described earlier, with $s = 10$, $n_s = 50$, $\alpha = 2.2$, and $\beta = 2.2$. Then we add 10 wireless substrate networks and equip each static node with a wireless interface, connecting to randomly selected wireless substrate network. We create 200 mobile nodes, each with two wireless interfaces with a radio range set to 20 meters. We organize the mobile nodes into 20 groups of 10 nodes. All mobile nodes in the same group connect to the same wireless substrate network with one of their wireless interfaces, and to a randomly selected wireless substrate network with the other interface.

Node mobility is modeled using the Reference Point Group Mobility (RPGM) model [31], where nodes move in groups in a 400×400 m^2 area. Nodes in a group move independently within a group boundary, given by a circle with a radius of 75 m. The movement of nodes is a superposition of the group movement and the movement within a group. Waypoint mobility governs the movement of individual nodes as well as that of groups, where nodes (and groups) move from one randomly selected waypoint to another, randomly selecting a new speed (between 0 and v_{max} m/s) at each waypoint. Also, nodes (and groups) pause for a random time between 0 and 5 seconds, when they reach a waypoint.

For LDR, we select five landmark domains, which are all placed either in static substrate networks, designated as LDR (5, static), or in mobile substrate networks, designated as LDR (5, mobile). This allows us to observe the impact of placing landmark domains in mobile substrate networks. We measure the overhead of the routing protocols in terms of the average bit rate of received routing messages at nodes. Fig. 9 shows the average delivery ratio, path stretch, state information, and overhead as a function of the maximum node speed v_{max}. Due to its slow convergence, we were not able to simulate VRR for dynamic networks. The average delivery ratio of LDR, although decreasing as v_{max} increases, is considerably higher than the average delivery ratio of Disco and UIP. This should be expected since Disco and UIP do not have mechanisms that deal with dynamic networks. For the delivered messages, the average path stretch of LDR is low and comparable to that of Disco. The average state information and the message overhead show a significant difference between the selection of static and mobile substrate networks for landmark domains. Specifically, placing landmark domains in dynamic substrate networks results in significantly increased message overhead and state information. The root cause is that landmark domains in mobile substrate networks experience splits and merges, which require frequent changes of the node locators.

VI. Conclusion

Multi-substrate routing addresses the problem of determining paths in interconnected independent networks, without relying on a single central substrate network (e.g., the Internet). We explored routing for arbitrarily connected large-scale multi-substrate networks. Compact routing and greedy forwarding are well known to have excellent scalability properties, however, they do not exploit available routing schemes within the individual substrate networks. We presented a simple routing scheme, which leverages the paths set up by intra-substrate routing, and showed that it can achieve a significant reduction in the amount of stored state information, compared to schemes that do not account for intra-substrate routing. We found that the path stretch of such a routing scheme is low, approaching an ideal path stretch in very large networks. Many aspects of multi-substrate routing, such as routing policies between substrate networks that are operated by different entities, are unexplored and remain to be investigated.

Acknowledgments

The research in this paper is supported in part by the Natural Sciences and Engineering Research Council of Canada.

References

[1] D. Farinacci *et al.*, "Locator/ID separation protocol (LISP)," *Internet Draft (draft-ietf-lisp-22)*, Feb. 2012.

[2] R. Atkinson and S. Bhatti, "ILNP architectural description," *Internet Draft (draft-irtf-rrg-ilnp-arch-06)*, Jul. 2012.

[3] X. Xu, "Routing architecture for the next generation Internet (RANGI)," *Internet Draft (draft-xu-rangi-04)*, Aug. 2010.

[4] B. Ahlgren *et al.*, "A node identity internetworking architecture," in *Proc. IEEE INFOCOM*, Apr. 2006, pp. 1–6.

[5] A. Feldmann *et al.*, "HAIR: hierarchical architecture for Internet routing," in *Proc. CoNext Workshop on Re-architecting the Internet (ReArch)*, Dec. 2009, pp. 43–48.

[6] D. Clark *et al.*, "FARA: reorganizing the addressing architecture," in *Proc. ACM SIGCOMM Workshop on Future Directions in Network Architecture (FDNA)*, Aug. 2003, pp. 313–321.

[7] J. Crowcroft *et al.*, "Plutarch: an argument for network pluralism," in *Proc. ACM SIGCOMM workshop on Future Directions in Network Architecture (FDNA)*, Aug. 2003, pp. 258–266.

[8] G. Bouabene *et al.*, "The autonomic network architecture (ANA)," *IEEE J. on Sel. Areas in Commu.*, vol. 28, no. 1, pp. 4–14, Jan. 2010.

[9] T. Koponen *et al.*, "Architecting for innovation," *SIGCOMM Comput. Commun. Rev.*, vol. 41, no. 3, pp. 24–36, Jul. 2011.

[10] P. B. Godfrey *et al.*, "Pathlet routing," in *Proc. ACM SIGCOMM*, Aug. 2009, pp. 111–122.

[11] S. Mies, O. P. Waldhorst, and H. Wippel, "Towards end-to-end connectivity for overlays across heterogeneous networks," in *Proc. IEEE ICC Workshops 2009*, Jun. 2009, pp. 1–6.

[12] S. Mies and O. Waldhorst, "Autonomous detection of connectivity," in *Peer-to-Peer Computing (P2P), 2011 IEEE International Conference on*, Aug 2011, pp. 44–53.

[13] L. Kleinrock and F. Kamoun, "Hierarchical routing for large networks; performance evaluation and optimization," *Computer Networks*, vol. 1, no. 3, pp. 155 – 174, Jan. 1977.

[14] C. Gavoille and M. Gengler, "Space-efficiency for routing schemes of stretch factor three," *J. on Parallel Distrib. Comput.*, vol. 61, no. 5, pp. 679–687, May 2001.

[15] M. Thorup and U. Zwick, "Approximate distance oracles," in *Proc. ACM STOC*, Jul. 2001, pp. 183–192.

[16] ——, "Compact routing schemes," in *Proc. ACM Symposium on Parallel Algorithms and Architectures (SPAA)*, Jul. 2001, pp. 1–10.

[17] A. Singla *et al.*, "Scalable routing on flat names," in *Proc. ACM CoNEXT*, Dec. 2010, pp. 20:1–20:12.

[18] A. Rao *et al.*, "Geographic routing without location information," in *Proc. ACM MobiCom*, Sep. 2003, pp. 96–108.

[19] J. Newsome and D. Song, "GEM: Graph EMbedding for routing and data-centric storage in sensor networks without geographic information," in *Proc. ACM SenSys*, Nov. 2003, pp. 76–88.

[20] R. Fonseca *et al.*, "Beacon vector routing: scalable point-to-point routing in wireless sensornets," in *Proc. NSDI*, May 2005, pp. 329–342.

[21] J. Herzen, C. Westphal, and P. Thiran, "Scalable routing easy as PIE: a practical isometric embedding protocol," in *Proc. IEEE ICNP*, Oct. 2011, pp. 49–58.

[22] I. Stoica *et al.*, "Chord: A scalable peer-to-peer lookup service for Internet applications," in *Proc. ACM SIGCOMM*, Aug. 2001, pp. 149–160.

[23] B. Y. Zhao, J. Kubiatowicz, and A. D. Joseph, "Tapestry: An infrastructure for fault-tolerant wide-area location and routing," Computer Science Division, University of California, Berkley, Tech. Rep. UCB/CSD-01-1141, Apr. 2001.

[24] A. Rowstron and P. Druschel, "Pastry: Scalable, decentralized object location, and routing for large-scale peer-to-peer systems," in *Proc. 18th IFIP/ACM International Conference on Distributed Systems Platforms (Middleware)*, Nov. 2001, pp. 329–350.

[25] B. Ford, "Unmanaged Internet protocol: taming the edge network management crisis," *ACM SIGCOMM Comput. Commun. Rev.*, vol. 34, no. 1, pp. 93–98, Jan. 2004.

[26] M. Caesar *et al.*, "Virtual ring routing: network routing inspired by DHTs," in *Proc. ACM SIGCOMM*, Aug. 2006, pp. 351–362.

[27] B. Drazic, "Scalable routing for networks of dynamic substrates," Master's thesis, Univeristy of Toronto, ECE Department, Jan. 2014.

[28] M. Bawa *et al.*, "Estimating Aggregates on a Peer-to-Peer Network," Stanford University, Tech. Rep. TR-2003-24, Apr. 2003.

[29] P. F. Tsuchiya, "The landmark hierarchy: a new hierarchy for routing in very large networks," in *Proc. ACM SIGCOMM*, Aug. 1988, pp. 35–42.

[30] A. Varga, "The OMNET++ discrete event simulation system," in *Proc. European Simulation Multiconference (ESM)*, Jun. 2001, pp. 319–324.

[31] X. Hong *et al.*, "A group mobility model for ad hoc wireless networks," in *Proc. ACM International Workshop on Modeling, Analysis and Simulation of Wireless and Mobile Systems*, Aug. 1999, pp. 53–60.