Capacity Deficit in Mobile Wireless Ad Hoc Networks Due to Geographic Routing Overheads

Alhussein A. Abouzeid
ECSE Department
Rensselaer Polytechnic Institute

Joint work with Nabhendra Bisnik

Maintaining State in VTNs

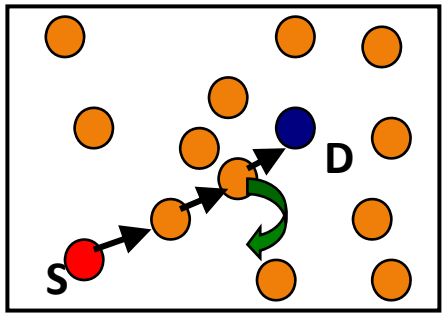
- Variable Topology Networks exhibit dynamically changing network topology e.g. due to mobility, fading, etc.
- *We view that routing typically involves a problem of maintaining state*
- Due to changing topology, it is very difficult to keep consistent state information
- Aggressive updates vs. use of stale information
  - Aggressive updates => large overheads
  - Stale information => low packet delivery rates
- Previous studies have shown that existing routing protocols overhead does not scale well with network size and mobility

Impact on Transport Capacity

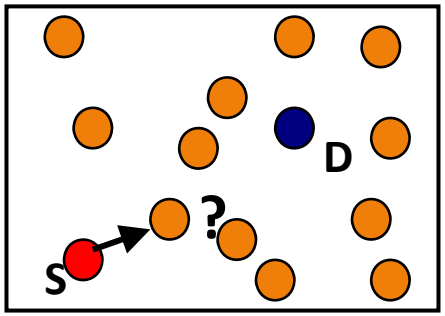
- The transport capacity scales poorly in MANETs (ignoring the capacity gains due to mobility in DTNs)
- The capacity analysis assumes that
  - ZERO overhead is required to *'figure out'* the network eg [GK00]
  - Or, that you don't need to figure out the network eg [GT01]
- However for many practical situations protocol information has to be continuously exchanged between nodes
- This protocol information reduces the *actual capacity* that is available for exchanging useful information

What is the deficit in capacity available to the end users caused by routing protocols?

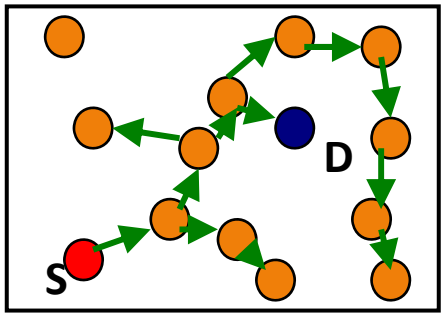
e.g. Mobility and Routing



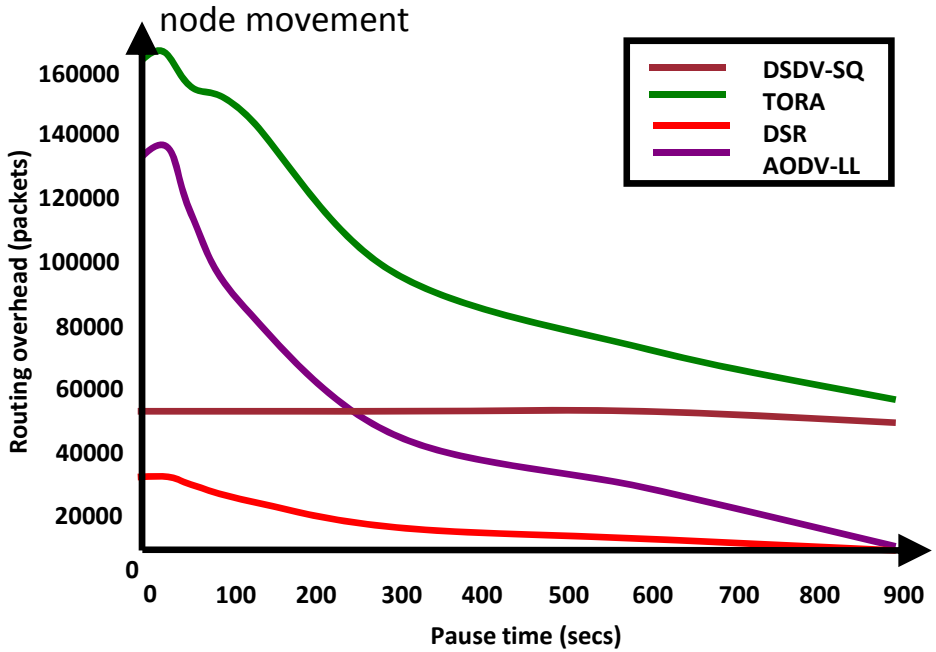
Paths frequently break due to node movement

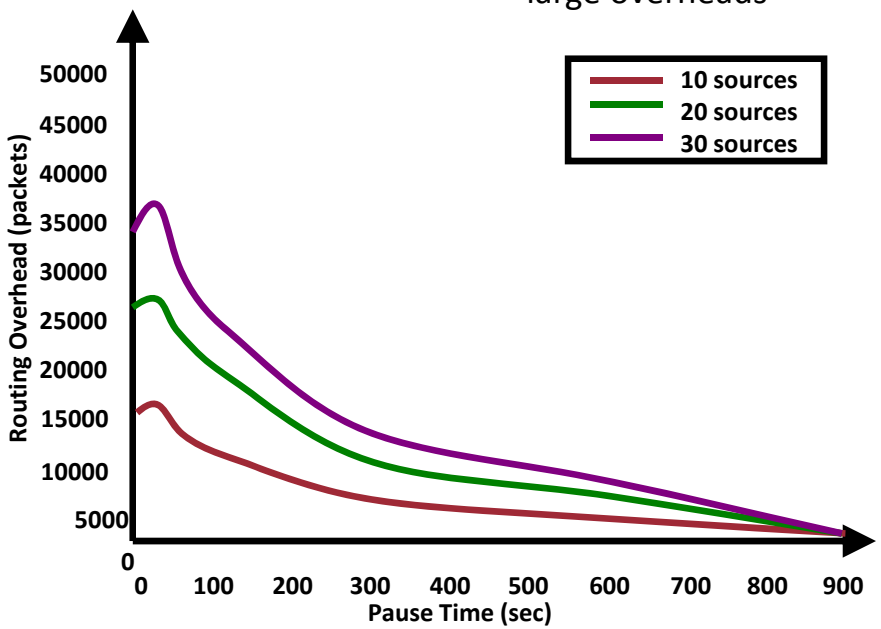


How to route packets now?



One option is to rediscover routes => large overheads

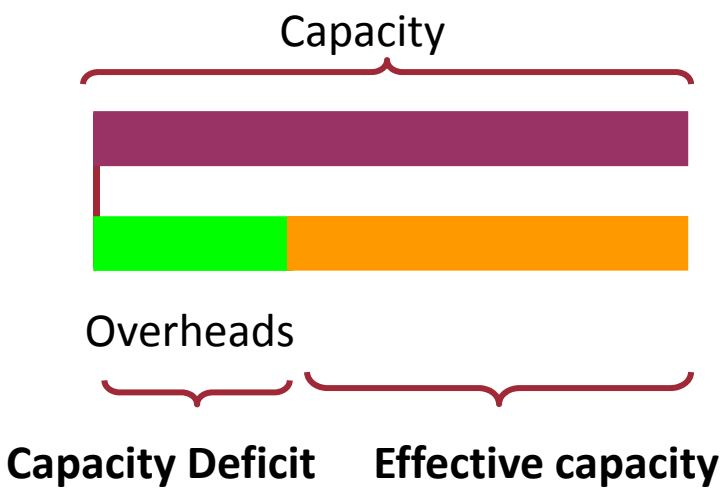




Existing studies [BM98] show that routing protocols are not able to handle high mobility or large number of sources

Objective: Limits on Protocol Information

- What is the lower bound on the routing overhead (state maintenance) that has to be incurred for reliable routing of packets in MANETs or VTNs in general
- Caveat: Function of the designer's definition of state
- Such bounds may
  - Provide yardstick for performance comparison
  - Inspire development of efficient routing protocols
  - Yield upper bounds on capacity deficit due to routing



In this paper…

- We consider geographical routing protocols i.e. state is geographic information
- Formulate the minimum routing overhead problem in geographic routing protocols as a rate-distortion problem
- Evaluate a lower bound on the rate at which a node has to transmit state information to ensure routing of packets with desired degree of reliability
  - What is minimum location update rate?
  - What is minimum beacon transmission rate?
- Evaluate an upper bound on the effective transport capacity available to the end users
- Characterize scenarios where complete transport capacity of a network may be consumed by routing overhead

Rate-distortion Motivation

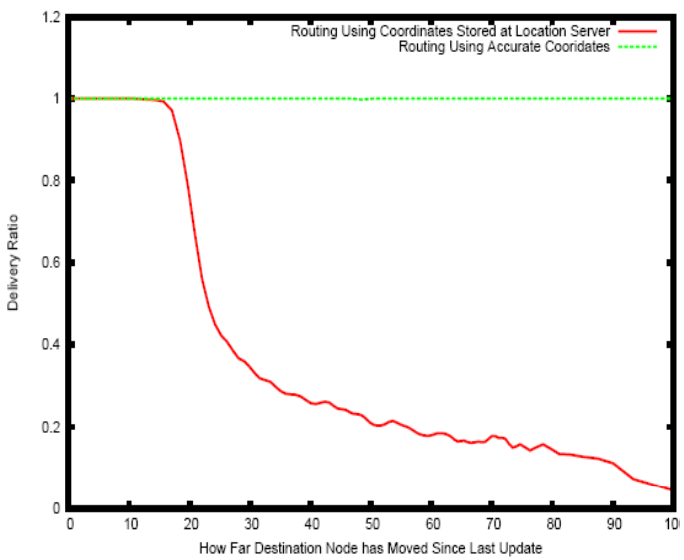
- State accuracy affects performance
- There is always a distortion between actual and perceived state
- In practice, can live with some distortion e.g. GPS driving directions are accurate within a few meters.
- So, we are interested in the overhead subject to a specified distortion bound
- Gives rise to a rate distortion problem
- Prior work
- [RG76]
  - R. Gallager. Basic limits on protocol information in data communication networks. IEEE Transactions on Information Theory, Vol.22, Iss.4, Jul 1976 Pages: 385- 398 .
  - "consider basic limitations on the amount of protocol information that must be transmitted in a data communication network to keep track of source and receiver addresses and of the starting and stopping of messages."
  - "a protocol is a source code for representing control information"
- [BA05]
  - Considered link state routing

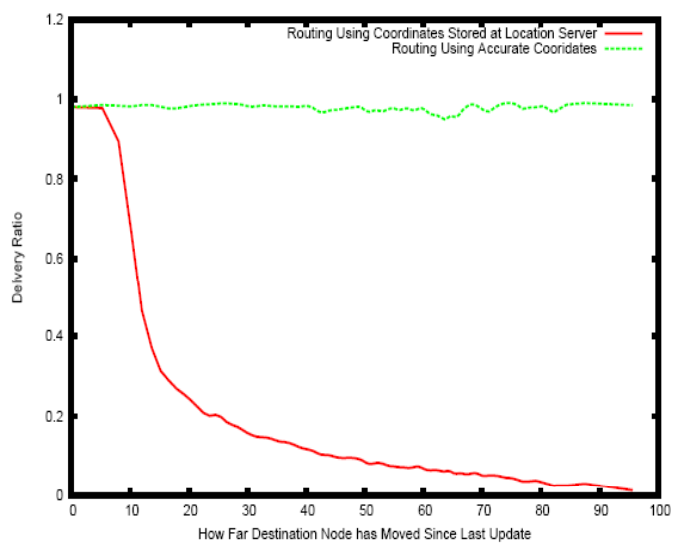
Location Error

- Should $\hat{X}_i(t)$ be equal to $X_i(t)$ at all times?
  - No!... Only when the location server is queried
- So to ensure high delivery ratio

$$\hat{X}_i(t) \approx X_i(t) \text{ when } t = T_i(j) \ \forall \ j \geq 1$$

- How does packet delivery ratio vary with error in location information?





Thus to ensure high packet delivery ratio, the error in location information must be greater than some threshold

Network Model

- The network consists of *n* nodes, each performs Brownian motion with variance σ^2
- We consider two network deployments
  - One dimensional network – Circle with perimeter is *L*
  - Two dimensional network – A torus, with area *A*
- The dimensions of network is large in comparison to σ^2
- The network is assumed to be connected
- Topology change negligible during packet traversal
- $X_i(t)$ denotes the position of node *i* at time t

$$X_i(t) = \{X_{i1}(t), X_{i2}(t)\}$$ For 2-D case

- $\hat{X}_i(t)$ denotes the position of node *i* that is stored at its location server at time t

Coordinate System

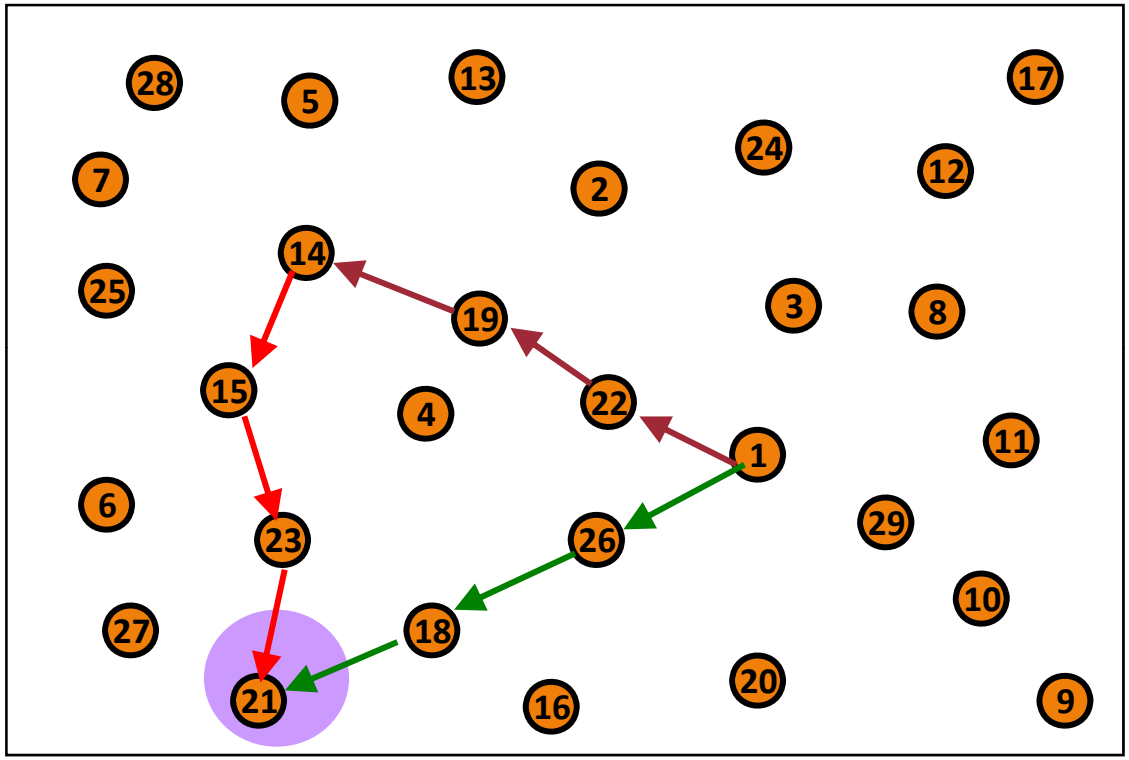
- For deployment along circle, the coordinates of a node are determined by opening the circle into a straight line about a fixed point
- Similarly for the 2-D case, we may open up the torus into a rectangle about a reference point
- Similar to considering projecting Brownian motion on infinite plane to a rectangle
- Let $X_i(t)$ denote the position of node *i* at time t

$$X_i(t) = \{X_{i1}(t), X_{i2}(t)\}$$ For 2-D case

- Let $\widehat{X}_i(t)$ denote the position of node i that is stored at its location server at time t

Geographic Routing Model

- A location service scheme is used to assign location server to each host
- Hosts periodically update their location servers at some rate
- Each node maintains neighborhood information by transmitting beacons
- The source node contacts destination's location server before sending out the packet
- The source host includes the location information of the destination in each packet
- Each intermediate node forwards packet to a neighbor that is closer to the destination than itself



Sources of overhead

1. **Beacons**
2. **Location update packets**

Traffic Model

- The jth packet destined to node i is generated in the network at time $T_i(j)$
- $T_i(j) - T_i(j-1)$ is distributed according to p.d.f. $f_S(t)$ $\forall$ $j \geq 1$ and $1 \leq i \leq n$
- Node i forwards k^{th} packet at time $\tau_i(k)$
- $\tau_i(k) - \tau_i(k-1)$ is distributed according to p.d.f $f_\tau(t)$ $\forall$ $k \geq 1$ and $1 \leq i \leq n$
- $T_i(0) = \tau_i(0) \triangleq 0$

Distortion Measure

- The vector of positions of node i when first N packets are routed to it

$$X_i^N = \{X_i(T_i(1)), X_i(T_i(2)), \ldots, X_i(T_i(N))\}$$

- The vector of positions used to route the packets is

$$\hat{X}_i^N = \{\hat{X}_i(T_i(1)), \hat{X}_i(T_i(2)), \ldots, \hat{X}_i(T_i(N))\}$$

- We use squared-error distortion measure

$$d(T_i(j)) = |X_i(T_i(j) - \hat{X}_i(T_i(j)|^2$$

- Distortion measure

$$D_N = \frac{1}{N}\sum_{j=1}^{N} d(T_i(j))$$

$$\overline{D} = \lim_{N\to\infty} D_N$$

- To ensure high delivery ratio $\overline{D} \leq \epsilon^2$

Rate-Distortion Formulation for Minimum Update Rate

- At what rate should a node transmit its location information to the server such that the distortion bound is satisfied?
- Natural to formulate it as rate distortion problem

Let $\mathcal{P}_N(\epsilon^2)$ be the family of probability distributions for which

$$\frac{1}{N}\sum_{j=1}^{N}\sum_{\mathcal{X}_i^N,\hat{\mathcal{X}}_i^N} P_N[X_i^N,\hat{X}_i^N](X_i(T_i(j))-\hat{X}_i(T_i(j)))^2 \leq \epsilon^2$$

Then

$$R_N(\epsilon^2) = \min_{P_N\in\mathcal{P}_N(\epsilon^2)} \frac{1}{N} I_{P_N}(X_i^N;\hat{X}_i^N)$$

Minimum rate which the node needs to transmit for first N packets

$$R(\epsilon^2) = \lim_{N\to\infty} \inf R_N(\epsilon^2)$$

The minimum rate at which node should transmit location information

Rate Distortion Analysis

We show that mutual information between X_i^N and $\widehat{X}_i^N$ satisfies

$$\min_{P_N \in \mathcal{P}_N(\epsilon^2)} I_{P_N}(X_i^N; \widehat{X}_i^N) \geq N R_1(\epsilon^2)$$

This implies that

$$R(\epsilon^2) \geq R_1(\epsilon^2)$$

$$\boxed{R(\epsilon^2) \geq h(X_{i1}(T_1)) + h(X_{i2}(T_1)) - \log(\pi e \epsilon^2) \text{ bits/packet}}$$

where

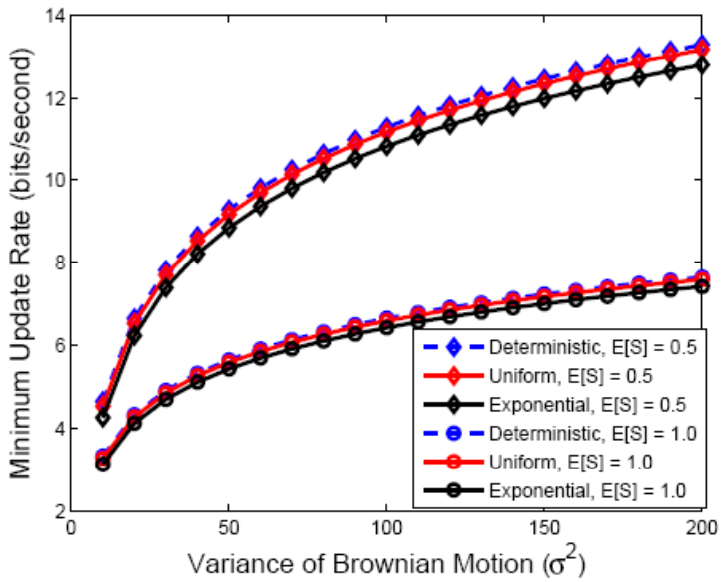
$$h(X_{i1}(T_1)) = -\int_{x=-\infty}^{\infty} f_X(x) \log\left(f_X(x)\right) dx$$

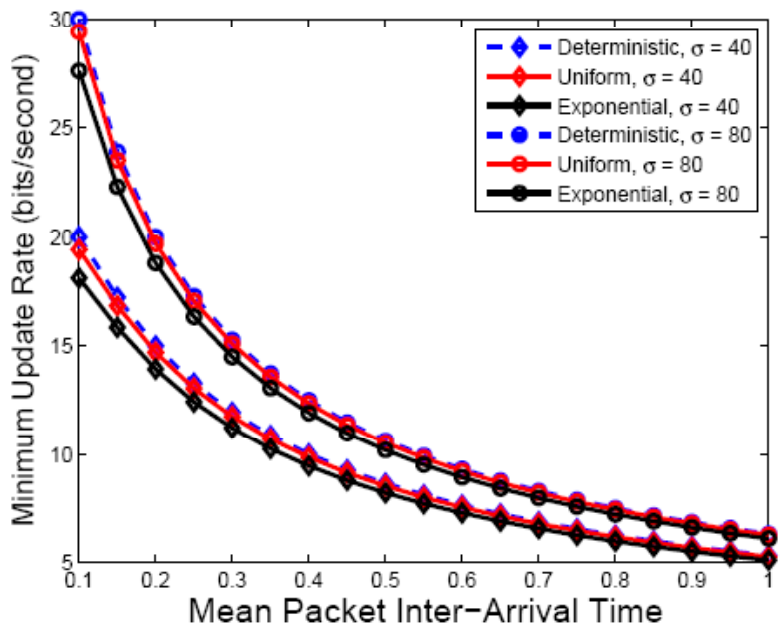
Differential entropy of $X_i(T_i(1))$ – depends on mobility pattern and packet inter-arrival process

$$f_X(x) = \int_{\tau=0}^{\infty} \frac{1}{\sqrt{\pi \sigma^2 \tau}} e^{-\frac{x^2}{\sigma^2 \tau}} f_S(\tau) d\tau$$

Distribution of $X_i(T_i(1))$ when $X_i(0) = 0$

Location Update Overhead for Different Packet Processes





- Interesting Insights from the paper for:
  - Deterministic arrival: has closed form expressions; can construct optimal update strategies for this case; has highest update rate among all distributions (Gaussian location change with highest variance)
  - Uniform/exponential inter-arrival: has highest entropy among cont time dist with finite/infinite base; yields max update rate

Location overhead increases with mobility and packet arrival rate

Neighborhood Information

- Nodes need to know who their neighbors are in order to make packet forwarding decisions
- Neighborhood information exchanged through exchange of beacons
- Again, consistent neighborhood information is required only at time instants when a node is forwarding packets
- At what rate must a node transmit its beacons such that its neighbors have consistent neighborhood information with high probability $(1 - \delta)$ when forwarding packets?
- We perform a similar rate-distortion analysis

Distortion Measure

Define indicator random variables

$$Z_{ij}(t) = \begin{cases} 1, & \text{if } j \in N_i(t) \\ 0, & \text{otherwise} \end{cases} \qquad \hat{Z}_{ij}(t) = \begin{cases} 1, & \text{if } j \in \hat{N}_i(t) \\ 0, & \text{otherwise} \end{cases}$$

Define vectors of indicator random variables:

$$Z_{ij}^N \triangleq \{Z_{ij}(\tau_i(1)), Z_{ij}(\tau_i(2)), \ldots, Z_{ij}(\tau_i(N))\}$$
$$\hat{Z}_{ij}^N \triangleq \{\hat{Z}_{ij}(\tau_i(1)), \hat{Z}_{ij}(\tau_i(2)), \ldots, \hat{Z}_{ij}(\tau_i(N))\}$$

Actual and perceived neighborhood for N forwarded packets

Hamming distortion measure is used

$$d(Z_{ij}(t), \hat{Z}_{ij}(t)) = \begin{cases} 1, & \text{if } Z_{ij}(t) \neq \hat{Z}_{ij}(t) \\ 0, & \text{if } Z_{ij}(t) = \hat{Z}_{ij}(t) \end{cases}$$

$E_{ij}(t)$ is defined as $$E_{ij}(t) = Z_{ij}(t) - \hat{Z}_{ij}(t)$$

If $E_{ij}(t)$ is zero then neighborhood information is correct

Rate distortion formulation

Minimum beacon rate problem: What is the minimum rate at which node j must transmit beacons such that

$$P[E_{ij}(\tau_i(k)) = 0] \geq 1 - \delta \;\; \forall \; 1 \leq i \neq j \leq n, 1 \leq k < \infty$$

The rate distortion function is given by

$$R^{(b)}(\delta) = \lim_{N \to \infty} \min R_N^{(b)}(\delta)$$

$$R_N^{(b)}(\delta) = \min_{P_N \in \mathcal{P}_N^{(b)}(\delta)} \frac{1}{N} I_{P_N}(Z_{ij}^N; \hat{Z}_{ij}^N)$$

Where $\mathcal{P}_N^{(b)}(\delta)$ the family of joint distributions of Z_{ij}^N $\hat{Z}_{ij}^N$ for which the distortion constraint is satisfied

Lower Bound for Beacon Rate

Similar to the minimum update rate analysis, we show that

$$R^{(b)}(\delta) \geq R_1^b(\delta)$$

Lower bound on beacon transmission rate is thus given by

$$R^{(b)}(\delta) \geq \mathcal{H}(p(l^\star)) - \mathcal{H}\left(\frac{\delta}{2}, 1-\delta, \frac{\delta}{2}\right) \text{ beacons/pkt}$$

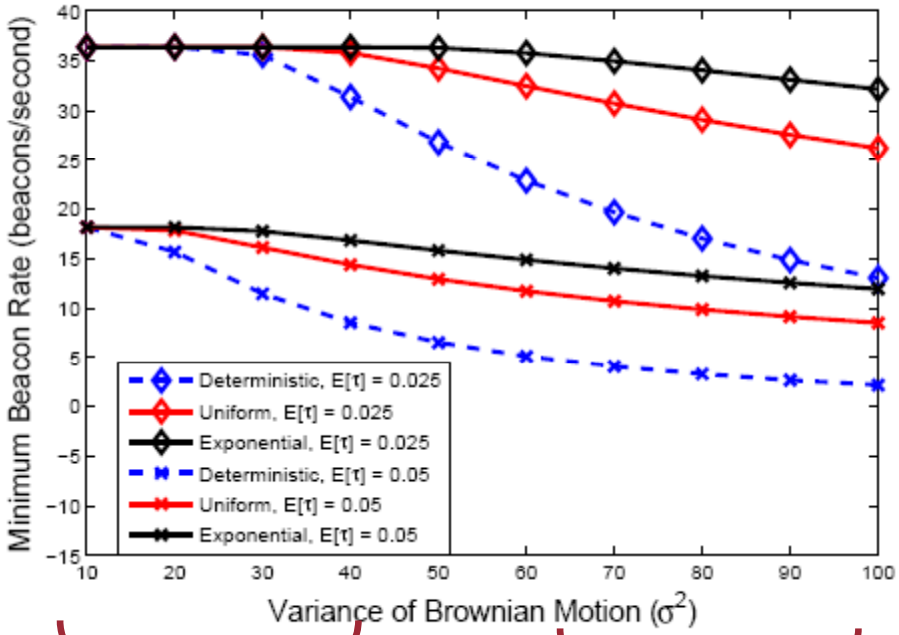
Where

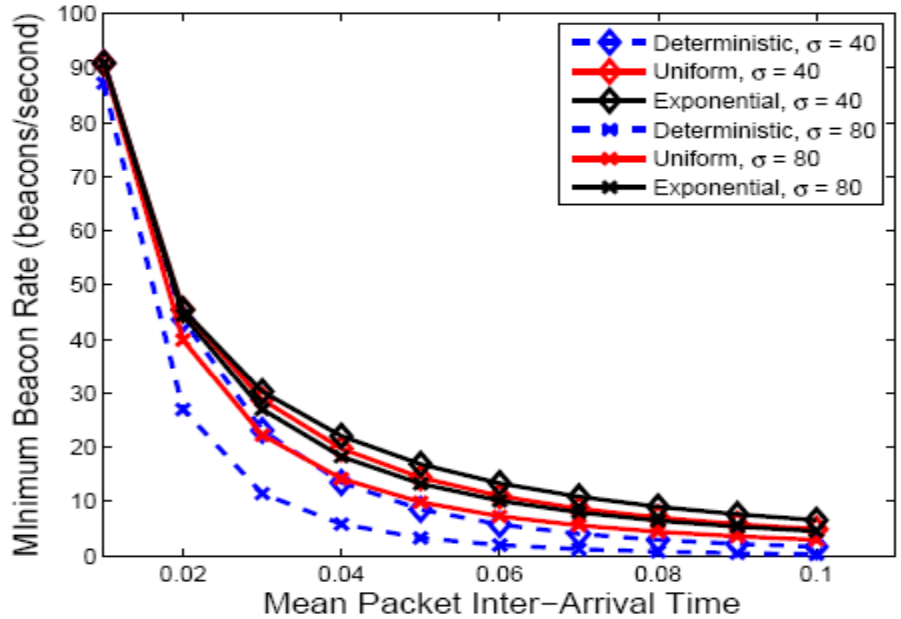
$$p(l^\star) = P[Z_{ij}(\tau_i(1)) = 1 | |X_j(0) - X_i(0)| = l^\star]$$

$$l^\star \triangleq \arg \min_{-r \leq l \leq r} \left| P[Z_{ij}(\tau_i(1)) = 1 | X_j(0) = l] - 0.5 \right|$$

l^* is the initial distance between nodes i and j that maximizes $H(Z_{ij}(\tau_i(1)))$

Numerical Results





Higher unpredictability

Change of neighborhood more predictable

When nodes are highly mobile they need to transmit beacons LESS frequently

Capacity Deficit Due to Routing Overheads

- If $\mathcal{C}(n)$s the transport capacity of a network, then the residual capacity, $\mathcal{C}'(n)$, for network users is given by

Deficit due to location updates

Deficit due to beacons

$$\mathcal{C}'(n) \leq \mathcal{C}(n) - n\left(\frac{\eta}{E[S]}R(\epsilon^2) + \frac{r}{E[\tau]}R^{(b)}(\delta)\right)$$

where

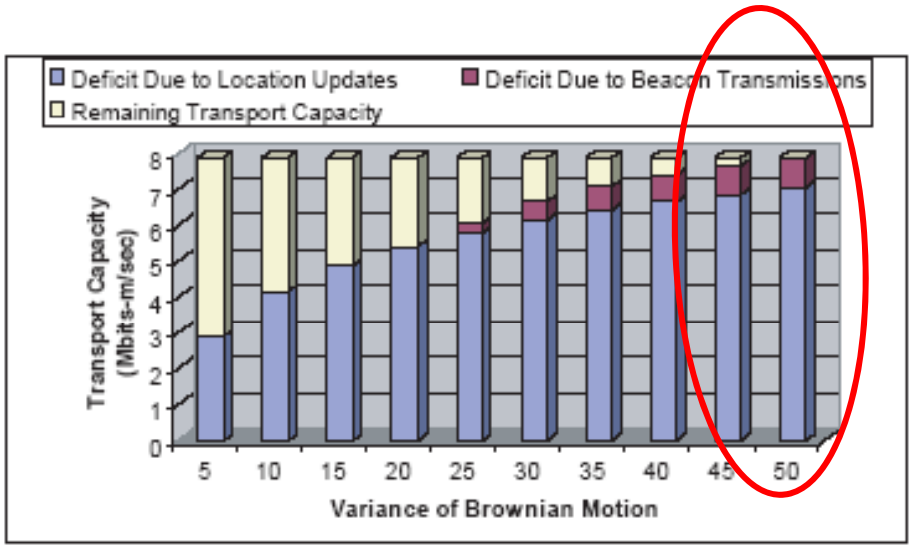
η: Average distance between a node and its location server.

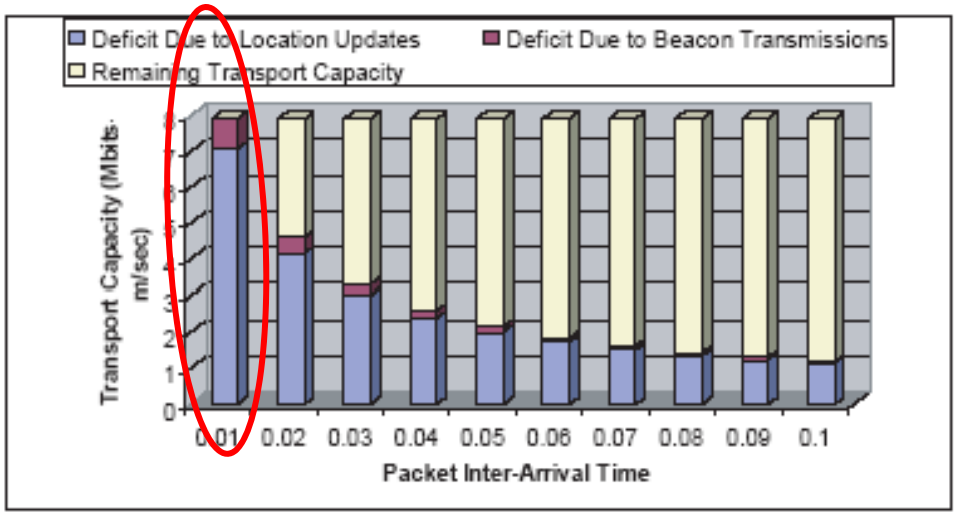
$E[S]$: Average inter-arrival time for packet destined to a node.

$E[\tau]$: Average time interval between two packets forwarded by a node.

Residual Capacity

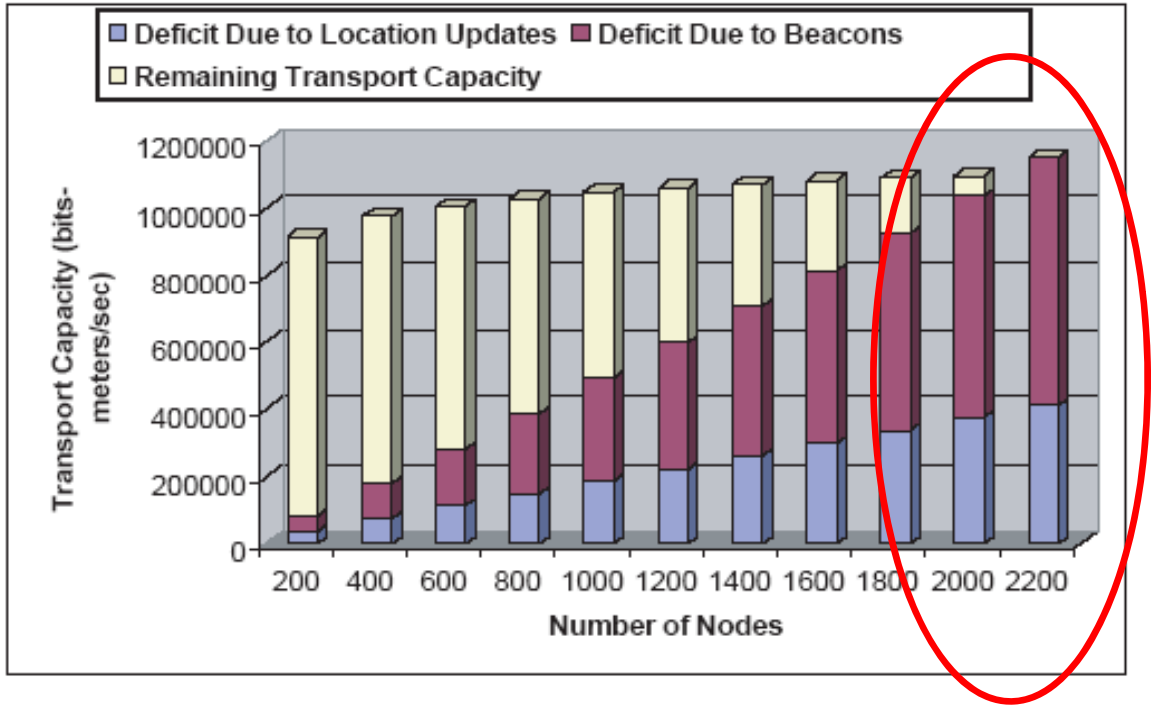
$$\mathcal{C}'(n) \leq \left(\frac{\sqrt{8}}{\pi} \frac{1}{\Delta} W \sqrt{nA} - n \frac{\eta}{E[S]} R(\epsilon^2) - n \frac{r}{E[\tau]} R^{(b)}(\delta) \right) \text{ bit} - \text{m/sec}$$





For high mobility and packet arrival rate, the entire transport capacity may be occupied by routing overheads

Residual Capacity



Not only does capacity not scale well with the number of nodes, routing overheads scale at a higher rate.

Critical Network Size

- For geographic routing, the upper bound on the maximum number of nodes that may be deployed in a network while ensuring that it has non-zero residual transport capacity is given by

$$n^{\star} \leq \left(\frac{\frac{\sqrt{8}}{\pi} \frac{1}{\Delta} W \sqrt{A}}{\eta U(\epsilon^2) + r U^{(b)}(\delta)} \right)^2$$

Critical size increases with
- **Transmission rate**
- **Deployment area**

Critical size decreases with
- **Mobility**
- **Packet arrival rate**
- **Number of hops to reach a location server**
- **Communication radius**

Conclusions

- We presented a rate-distortion framework for evaluating lower bounds on minimum routing overheads in geographic routing
- For high mobility and packet generation rate, entire capacity may be occupied by routing overheads
- Is it possible to design scalable routing protocols?
  - Exploit trade-offs to make routing protocols scalable
- For example, traffic shaping at source to reduce the effective packet generation rate

Future Work

- Design protocols whose update rates are close to the lower bounds
  - How to encode the change in location?
  - How to efficiently manage location servers?
  - What could be a more efficient way of maintaining neighborhoods
- Extend the model to other routing paradigms
- Unified cost measures for different types of state

Thanks.

Questions & Comments?