Study of a Bus-Based Disruption Tolerant Network: Mobility Modeling and Impact on Routing

Ellen (Xiaolan) Zhang, Jim Kurose, Brian Levine, Don Towsley, Honggang Zhang
Univ. of Massachusetts, Amherst,
Fordham University

Supported in part by NSF awards ANI 0325868 and CNS 0519881 and U.S. Army Research Laboratory and the U.K. Ministry of Defense under Agreement Number W911NF-06-3-0001.
UMass DieselNet

- 10 routes span 150 square miles
- 40 buses installed with brick computer; GPS receiver; 802.11b AP and adapter
- Bus-to-bus transfer when within range
- Network formed by buses is disconnected
  - often no contemporaneous end-to-end path between a pair of buses at a given time
  - Disruption Tolerant Network

Bus-to-bus transfers during 3/23-4/24, 2005
Other DTN Scenarios

- Networks for remote areas
- Scientific applications
- Military networks, disaster recovery networks, etc.
- Pocket-switched network

DakNet (posta)

Our focus: DTNs where disruptions are caused by random node mobility and sparse density
Routing despite disconnection, also called mobility-assisted routing

Epidemic routing [Vahdat and Becker,00]
- packet propagation => disease spreading
- recovery process on delivery to dest

Node status: infected, susceptible, recovered
Understanding DTN Mobility

- Mobility: important determinant of DTN routing performance
- Many works assume synthetic mobility models
- Real DTN mobility and implications to routing
  - trace collection: Haggle project, DieselNet, etc.
  - trace-driven simulation studies
  - trace characterization and modeling
    - power-law of aggregate inter-contact times [Chaintreau et al., 06, Chen et al., 06]
    - power law and exponential decay of aggregate inter-contact times [Karagiannis et al., 07]
    - pair-wise inter-contact times [Conan et al., 07]
Goal: a generative mobility model

- Our goal: a generative model based on real mobility traces that accurately predicts DTN routing performance

1. Original Traces
2. Trace-driven Simulation
3. Modeling Study

Modeling Study → Model
Model → Trace Generator
Trace Generator → Synthetic Traces

Compare

Performance metrics

Performance metrics
Outline

- Background and motivation
- Traces description, metrics of interest
- Modeling of inter-contact times
- Model comparison
- Summary and future directions
UMass DieselNet trace

- 10 routes span 150 square miles
- 40 instrumented buses
  - Linux computer; GPS receiver; 802.11b AP and adapter
- Bus-to-bus contact
  - when within range: 802.11 affiliation, TCP connection, transfer max. amount of data
  - <time, duration, amount of data transferred, GPS location>
UMass DieselNet: bus routes/schedules

- 3 most popular routes
  - linear/butterfly route
  - during 7am to 7pm for non holiday weekdays
  - multiple shifts: each shift starts at different bus stops at different time, runs on route continuously
Performance Metrics of Interests

- **Performance metrics:**
  - best case delay, copies made
  - hop count of epidemic path

- **Trace-driven simulation**
  - evaluate above metrics for packets generated at any time [7am, 7pm] between any (src, dest) pair

- **Aggregate distribution** of performance metrics, assuming:
  - src pkts arrive uniformly randomly to each unicast pair, at time uniformly randomly between [7 am, 7 pm]
Outline

- Background: Disruption Tolerant Network
- Traces description, metrics of interest
- Modeling of inter-contact times
- Model comparison
- Summary and future directions
How to model the trace?

- Focus on inter-contact time: duration of time between two subsequent contacts for a bus pair

- Choose a modeling granularity
  - shift-level: buses running on given shift-pair
  - route-level: buses running on given route-pair
  - aggregate: all-bus-pair

Goal: A simple model with good prediction accuracy
Preliminary: artifacts of finite length measurements

Different observations:
- fully observed, start-censored, end-censored, “no-meeting”
Preliminary: considering censored observations

- Ignoring censored samples => under-estimation
  - long inter-contact time likely to be censored
  - empirical aggregate CDF gives larger weights to pairs with small avg. inter-contact time

- Kaplan-Meier estimator
  \( \hat{S}(t) \) for \( S(t) := \Pr(X>t) \)
  - using all observations
  - Nonparametric maximum likelihood estimate

\[ \hat{S}(t) \]

\[ \Pr (X>t) \]

\[ \text{CDF with censored data} \quad \text{KM estimated} \]

\[ \text{empirical CCDF for all observations} \]

\[ \text{empirical CCDF for fully observed inter-contact time} \]
Evaluating Aggregate Model

- Trace generated based on aggregate statistics
  - similar total no. of contacts, matching statistics
  - delivers more packets
  - fewer copies made, similar path hop count
  - insight: contacts equally distributed to all pairs

Need finer-grained model!
Route-level: aggregate inter-contact times for buses running on route-pair
Route-level inter-contact time

Many short inter-contact time

Periodic behavior

Same no. of start, end-censored inter-contact time

Many instances of no-meeting
Understand/Model the structure

- Two buses on same linear routes meet every half round trip time
- When within range, two buses may fail to set up connection => inter-contact time made up of several physical inter-meeting times
- A mixture normal model

\[
f_{GEO_{1BM}}(x) = \sum_{i=1}^{\infty} p^{i-1}(1-p)f_N(x | i\mu, \sigma^2)
\]

- \(p\): prob. that two buses fail to set up connection in a meeting
- \(\mu\): physical inter-meeting time
- \(\sigma^2\): variance to account for random factors: traffic/road conditions.
Model Parameter Estimation

\[ f_{GEO_{-1BM}}(x) = \sum_{i=1}^{\infty} p^{i-1}(1 - p)f_N(x | i\mu, \sigma^2) \]

- **Expectation-Maximization Algorithm**: find maximum likelihood estimates for \( p, \mu, \sigma \) from empirical data
  - hidden variables: \# of physical meetings within inter-contact time, i.e., which component observation is drawn from
  - account for censored observations

- Estimated model generates similar fully-observed inter-contact time, censored observations as the original trace
Route-level Model

- **Linear routes:** some shift pairs have higher failure probability:

\[
f_{GEO\_MP\_1BM}(x) = \sum_{i=1}^{2} w_i \sum_{l=1}^{\infty} p_i^{l-1} (1-p_i) f_N(x \mid i\mu, \sigma^2)
\]

- **Butterfly shape route** (campus shuttle):
  - shifts pair on same direction: very rarely meet
  - shifts pair on opposite direction: meet either every half round trip time or every round trip time

\[
f_{GEO\_2BM}(x) = \sum_{i=1}^{2} w_i \sum_{l=1}^{\infty} p_i^{l-1} (1-p_i) f_N(x \mid li\mu, \sigma^2)
\]
Outline

- Background: Disruption Tolerant Network
- Traces description, metrics of interest
- Modeling of inter-contact times
- Model comparison
- Summary and future directions
Model Comparison

- Compare different models in terms of performance prediction
  - **aggregate model**: sample from aggregate statistics, considering censorship
  - **route-level statistics**: sample from the route-level statistics, considering censorship
  - **route-level model**: derived based on route-level statistics, with additional consideration for Campus shuttle
Model Comparison: delivery delay

- **Route-level statistics**
  - predicts delay, copies more accurately than aggregate model
- **Route-level model**
  - better prediction: incorporate shift info
  - matches tail
  - predicts larger avg. delivery delay: small inter-contact times not considered; correlation between different bus pairs
Conclusion

- Generative model based on real mobility trace
  - importance of considering censorship
  - aggregate model cannot capture aggregate performance statistics
  - finer-grained model predicts performance more accurately, and reveals structures within mobility

- Potentially applicable to other transport-based networks
Future Directions

- Understand and model short inter-contact times
- Model contact duration
- Model the correlation between different shift-pairs
- Model validation: using a separate trace
- Impact of infrastructures: APs in garage, café
- Technique for identifying the structure without domain knowledge
Acknowledgement:

- DieselNet project: John Burgess, Mark Cornor, Brian Lynn, Adam Sherson, Glen Barrington (PVTA), Yuri Pyuro, et al.
- MobiCom reviewers and shepherd

Thanks!
Questions/Comments?
Backup slides
802.11b Trans. Range
[Anastasi et al. 03]

Table 3. Estimates of the transmission ranges at different data rates.

<table>
<thead>
<tr>
<th></th>
<th>11 Mbps</th>
<th>5.5 Mbps</th>
<th>2 Mbps</th>
<th>1 Mbps</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data</td>
<td>30 meters</td>
<td>70 meters</td>
<td>90-100 meters</td>
<td>110-130 meters</td>
</tr>
<tr>
<td>Control</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TX_range</td>
<td>90 meters</td>
<td>120 meters</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Is Aggregate Model Sufficient?

All-bus-pair-all-day aggregated inter-contact time
Kaplan-Meier Estimator
(Product Limit Estimator)

- Suppose in \{x_i\}, there are n distinct inter-contact times, sorted as
  - \( T_1 < T_2, \ldots, < T_n \)
  - \( n_i \): the no. of inter-contact times that are greater than \( T_i \) (including censored observations)
  - \( d_i \): the no. of inter-contact times of length \( T_i \)

- Kaplan-Meier estimator for \( S(t) := \Pr(X > t) \)

\[
\hat{S}(t) = \prod_{t_i < t} \frac{n_i - d_i}{n_i}
\]
Bus Pairs on SHUTTLE Route

- $T_1$: time to traverse A-B-C-D-A
- $T_2$: time to traverse C-D-E-F-C
- Meeting sequence for buses on opposite directions: $T_1/2, T_1/2, T_2/2, T_2/2, T_1/2, T_1/2,$ ...; or $T/2, T/2,$ ..., where $T = T_1 + T_2$ ($T_1 \approx T_2$ for SHUTTLE)
- Buses on same directions rarely meet

$$f_{GEO_{\_2BM}}(x) = \sum_{i=1}^{2} w_i \sum_{l=1}^{\infty} p_i^{l-1} (1 - p_i) f_n(x, il\mu, \sigma^2)$$
Model Validation: Copies Made

Relative accuracy of predictions by different models similar to delay performance
Some shift pairs have higher failure prob.

\[ f_{\text{GEO\_MP\_1BM}}(x) = \sum_{i=1}^{2} w_i \sum_{l=1}^{\infty} \left( p_i^{l-1} (1 - p_i) f_N(x \mid i \mu, \sigma^2) \right) \]

Model generates similar fully observed inter-contact time. Discrepancy in censored observations due to other failure conditions.
Routing in DTNs

- **Challenges**
  - opportunistic contacts: need to search for paths
  - resource constraints: bandwidth, power, buffer space
  - local knowledge with delayed feedback

- **Objective**
  - resource efficient scheme to achieve good performance, e.g., high packet delivery ratio, small delay

- **Routing schemes**
  - single-copy or multiple-copy routing
  - stateless or stateful routing
  - resource constraints assumptions: packet scheduling, buffer management strategies