Wireless Social Networks: It takes two to Tango

Anil Vullikanti and Madhav V. Marathe
Network Dynamics and Simulation Science Laboratory
Virginia Bio-Informatics Institute & Dept. of Computer Science
Virginia Tech
{vsakumar, marathe}@vt.edu
Web Site: http://ndssl.vbi.vt.edu
These slides are a version of the lecture given as a part of the Summer School organized by Wireless@VT.

Organizers: Wireless@VT, Virginia Polytechnic Institute and State University

Venue & Date: June 2010, Blacksburg VA
Acknowledgements

Members, Network Dynamics & Simulation Science Laboratory, VBI

External Collaborators:

S. S. Ravi, (SUNY Albany), Hari Balakrishnan (MIT), Ravi Sundaram (Northeastern), Lukas Kroc (Cornell), Riko Jacob (ETH), Kai Nagel (Berlin), Goran Konjevod (ASU), Aravind Srinivasan, Sri Parthasarathy (U. Maryland), Nan Wang (Goldman Sachs) Stephan Eidenbenz, Sunil Thulasidasan, Gabriel Istrate, Anders Hansson, Jim Smith (LANL), Mayur Thakur (Google)
Integrated representation of Social, vehicular and telecommunication networks
- Understanding wireless networks requires more than just packet simulations
- Open Systems: pot pouri of protocols, providers and standards
- Activity Based models for Synthetic Sessions
- Wireless networks “cannot” be defined without the underlying social network
What we’d like to have

For individuals in a population (representation of individuals):

- Their demographics (Who)
- The sequences of activities they do (What)
- The times they do them (When)
- The places they do them (Where)
- The reasons they do them (Why)

And their interactions with devices, environment and other individuals (and their context)

- The devices they carry
- How and where they use them (why)
- Whom do they interact with (whom)

Combined with dynamic models of processes (messages, services and packets) and their co-evolution to obtain

A causal modeling framework of multi-theory multi-layered social and communication dynamic networks
Challenges: these networks co-evolve

• Dynamic ad-hoc radio networks
  – Social Networks, mobility of devices, the specific calling patterns and network protocols (e.g. power and frequency assignment) all decide the time varying adhoc radio networks
  – Conversely, the underlying network decides the performance of network protocols, and potentially calling patterns

• Epidemics
  – Social Network, public policy and individual behavior affect the disease outcome
  – Conversely, as disease spreads, behavior and thus social networks changes.
Challenges: Multi-layered, multi-theory networks

- Emerging applications in ubiquitous computing and communications
  - Dynamic spectrum access and trading
  - Location aided services
- Integrated representation of Social and Wireless Networks needed for developing new applications
  - more than just packet simulations
  - Activity Based models for Synthetic Sessions

Wireless networks cannot be **effectively** designed, analyzed and controlled in isolation without taking into account the social context – *the social and communication networks co-evolve*
Our approach: Integrated Modeling of Co-evolving Social and Communication Networks

- A unique end-to-end modeling environment to represent *integrated coupled social communication networks (SoCom)*
  - Designed to scale to $10^7$-$10^9$ mobile entities
  - Inter-operable with existing simulations of specific modules
  - Highly Detailed on spatial and temporal scale
- Used in practical case studies
  - E.g. multi-sector crisis management for DHS
Illustrative Application Areas

- Worm propagation
- Denial of service attacks at various scales
- Network Design
- Identify critical assets (DHS study)
- Framework for trading Spectrum
- Demand modeling

**Integrated coupled social and wireless Network environment**

**Cyber-vulnerability**

**Dynamic Spectrum Markets**

**Network Planning and Vulnerability assessment**
Scenario: Spectrum Management

• Wireless companies want to bid for restricted bandwidth
  - Time varying demands needed for making good bids
  - Intelligent bids can be made based on geographic call patterns
  - FCC needs to ensure no collusion and bidding is fair

• Tools needed
  - Models for mobility and call patterns
  - Efficient methods to study detailed agent based market mechanisms
  - Behavioral models of market player: e.g. speculation and collusive behavior
Scenario: cybervulnerability to worm attacks

- Growth of Smart phones
  - 1.2 billion smart devices to be sold by 2010
  - Applications: m-commerce, banking, social-networking

- Increase in incidences of mobile malware
  - Increasingly vulnerable to attacks similar to PCs
  - Affected by cross-over (infect smart devices through PCs) worms
Scenario: cybervulnerability to worm attacks

- Designing strategies to protect networks
  - Understand the dynamics: outbreak size, duration, etc.
  - How to detect quickly
  - Interventions: choose subset of nodes to force patches

- Tools needed
  - Realistic mobility model
  - Tools for large scale simulation and analysis of epidemics

"Human mobility and wireless networking could combine to abet the spread of computer viruses”
- Jon Kleinberg  [Nature 2007]
Scenario: survivability analysis and network planning

- What is the maximum loss in capacity if some nodes fail randomly?
- If k nodes could be reinforced (e.g. high capacity mobile base stations), which should be the ones so that the capacity is least affected?
- Tools needed
  - Mobility matters: models for node mobility
  - Time varying demands
  - Methods for estimating capacity and critical nodes
Scenario: cellular network offloading in DTNs

- Communication at two levels
  - *Opportunistic communication*: each node gets information from friends and other contacts in social network who are in the vicinity (assume probabilistic model for diffusion)
  - *Direct transmission* from cellular network
  - Combined transmission to ensure bounded delays
- *Goal*: choose initial *target set* to seed the opportunistic communication, so that amount of cellular offloading is minimized
This talk

• Social, vehicular, communication networks are coupled

• Integrated communication networks are complex systems
  – Open, pot pouri of protocols, legacy systems
  – Understanding requires more than just packet simulations

• Today’s Tutorial: Outline an end-to-end approach
  – High resolution modeling of coupled social and communication networks spanning large urban areas
  – Illustrate the approach with realistic case studies
Outline of this talk

- Overall architecture
- Varied uses
- Dynamic Urban Agent Synthesis
- Dynamic Social Network Construction
- Tele-traffic analysis on Integrated Communication Network
- Case studies
Overall architecture

Dynamic Urban Agent Synthesis

Dynamic Social Network Construction

Tele-traffic Analysis on Integrated Communication Network
Focus: Coupled social & 3G+ communication networks spanning large urban areas

- Focus on end-to-end packet level simulation of interdependent coupled social communication systems (including ad hoc, hybrid & mesh)
- Goal
  - $O(10^{7-9})$ mobile clients in an urban region, $O(10^{12-14})$ packets/hour
  - each demographically defined, each activity defined, each capable of creating or receiving realistic packet sessions
- Hooks for existing network simulators in end to end framework
Varied uses

- Design and architecture of next generation adhoc and mesh networks
  - Teletraffic modeling for wireless and mesh networks
  - Capacity of wireless networks
  - Design of cross layer protocols
- Assessing vulnerabilities associated with infrastructure inter-dependencies -- existing methods are not designed for this
  - Emergency management planning and restoration of communication systems in a built urban environment
  - Attack on network control operations of urban transport system
- Effects of regulations & policies on network level cyber-vulnerability
  - Distributed denial of service attack using fast moving wireless devices
  - Lack of available spectrum resources and prioritization schemes
Outline of this talk

• Overall architecture
• Varied uses
• Dynamic Urban Agent Synthesis
• Dynamic Social Network Construction
• Tele-traffic analysis on Integrated Communication Network
• Case studies
Dynamic urban agent synthesis

- Dynamic Urban Agent Synthesis
- Dynamic Social Network Construction
- Tele-traffic Analysis on Integrated Communication Network

- Population Synthesis
- Activity & Location Assignment
- Inter-modal Routing
- Vehicular Flow Simulation
Mobility models in literature

• Have significant impact on protocols [Barrett et al. MOBIHOC 2002]

• Number of different approaches
  – Random waypoint Model, e.g. [Johnson and Maltz, 1996]
  – Random Direction Mobility Model [Royer et al., ICC 2001]
  – Gauss-Markov Model [Liang and Haas, INFOCOM 1999]
  – Exponential Correlated Random Mobility [Gerla et al., MSWiM 1999]
  – City Section Model [Davies, 2000]
  – Column Mobility and other Group Mobility models [Sanchez and Manzoni, 2001]
  – Obstacle Mobility Model [Jardosh, Belding-Royer, et al., MOBICOM 2003]

Increasing amount of input data needed
Obstacle Mobility Model

- points move on voronoi graph of obstacles
- “random” movement pattern, with exponential waiting

[Jardosh, Belding-Royer, Almeroth, Suri, MOBICOM 2003]

Packet latency for various models
Mobility Models in literature

- **Advantages**
  - Simple to describe and implement: few parameters
  - Easy to analyze in many cases
  - Adequate to capture aggregate properties, e.g. density

- **Shortcomings**
  - Cannot represent realistic individual behavior
  - Unrealistic spatial and time variation
  - Do not take realistic urban features into account

- **Our approach**
  - Combines a wide variety of public and commercial data sets
  - Statistically matches traffic measurements in a city
  - Can be used to generate mobility in “unusual” settings
  - Significantly differs from other mobility models
Two Different mobility models

TRANSIMS mobility

Random WayPoint
Ad-hoc networks generated by TRANSIMS are structurally different from Random Waypoint and Erdos-Renyi Random Graphs.
Affect of Node/Edge Failures

Network Dynamics and Simulation Science Laboratory
• Instantaneous MAC layer capacity depends on topology and time
• Protocols need to be optimized for specific topologies
Route-lengths vs MAC Capacity

![Graph showing route-lengths vs MAC capacity with various data points and line segments.](image-url)
QoS Measures and Topology

Throughput

Average Latency

AODV/MIR

Radio Radius [m]

Average Packets Received Measure [%]

Average Latency Measure [s]
Mobility Matters!

• Realistic Urban environments (mobility, activities, sessions) yield structurally different communication networks

• The structure of the network affects the Quality of Service of digital traffic.

• Protocols optimized for random topologies and mobility models may perform poorly in practice

• Emergency response strategies, e.g. placing mobile base stations, need to consider mobility
Outline of this talk

• Overall architecture
• Varied uses
• **Dynamic Urban Agent Synthesis**
  – Population Synthesis
  – Activity Generation
  – Location Generation
  – Route Generation
  – Flow Simulation
• Dynamic Social Network Construction
• Tele-traffic analysis on Integrated Communication Network
• Case studies
Creating synthetic households: data flow

**Network Data**
- activity locations

**Forecast**
- marginals by block group

**STF-3A**
- summary tables of demographics
- available for block groups

**PUMS**
- 5% sample of census records
- PUMA consisting of census tracts, etc.
- approximately 5,000 people

**TIGER/Line**
- using MABLE/Geocorr
- geographic layout of census tracts and block groups

**Synthetic Households**
- location
- census tract / block group

**Synthetic Persons**
- gender
- age
- schooling
- employment (type, location, hours)
- transportation
- income

**Vehicles**
- vehicle id
- household
- initial network location
- type of vehicle
- emissions type

---

Network Dynamics and Simulation Science Laboratory
Step 1a Creating synthetic households: algorithm overview

- Use SF-3 marginal totals for each demographic variable to construct a multi-dimensional table with unknown values in each cell but known column and row sums

- Construct Multi-dimensional table using PUMS

- Use Iterative proportional fitting algorithm to construct a table that has right proportions based on SF3 data of households in each cell

- Randomly choose household from PUMS from each cell till the proportion is matched.
Step 1a Creating synthetic households: example

Proportion of Family Households, $n$, with Number of Workers in the Household

<table>
<thead>
<tr>
<th>Workers</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>&gt;2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prop.</td>
<td>0.000</td>
<td>0.336</td>
<td>0.594</td>
<td>0.069</td>
</tr>
</tbody>
</table>

Proportion of Family Households, $n$, with Householder Age in the Given Ranges

<table>
<thead>
<tr>
<th>Age</th>
<th>15-24</th>
<th>25-34</th>
<th>35-44</th>
<th>45-54</th>
<th>55-64</th>
<th>65-74</th>
<th>&gt;74</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prop.</td>
<td>0.011</td>
<td>0.372</td>
<td>0.261</td>
<td>0.128</td>
<td>0.128</td>
<td>0.100</td>
<td>0.000</td>
</tr>
</tbody>
</table>

Two SF-3 Tables giving marginal distributions for two demographics

Yields a multi-way table within unknown cell values
Step 1a Creating synthetic households: example

Multi-way SF3 based table

| Householder Age | Workers 15-24 | 25-34 | 35-44 | 45-54 | 55-64 | 65-74 | >74 | %  
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>%</td>
<td>0.011</td>
<td>0.372</td>
<td>0.261</td>
<td>0.128</td>
<td>0.128</td>
<td>0.100</td>
<td>0</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Householder Age</th>
<th>Workers 15-24</th>
<th>25-34</th>
<th>35-44</th>
<th>45-54</th>
<th>55-64</th>
<th>65-74</th>
<th>&gt;74</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.001</td>
<td>0.007</td>
<td>0.006</td>
<td>0.002</td>
<td>0.017</td>
<td>0.042</td>
<td>0.028</td>
<td>0.104</td>
</tr>
<tr>
<td>1</td>
<td>0.077</td>
<td>0.072</td>
<td>0.081</td>
<td>0.032</td>
<td>0.053</td>
<td>0.040</td>
<td>0.012</td>
<td>0.297</td>
</tr>
<tr>
<td>2</td>
<td>0.019</td>
<td>0.090</td>
<td>0.182</td>
<td>0.103</td>
<td>0.056</td>
<td>0.015</td>
<td>0.004</td>
<td>0.468</td>
</tr>
<tr>
<td>&gt;2</td>
<td>0.000</td>
<td>0.002</td>
<td>0.043</td>
<td>0.050</td>
<td>0.027</td>
<td>0.007</td>
<td>0.002</td>
<td>0.131</td>
</tr>
<tr>
<td>Total</td>
<td>0.027</td>
<td>0.170</td>
<td>0.312</td>
<td>0.188</td>
<td>0.153</td>
<td>0.104</td>
<td>0.046</td>
<td></td>
</tr>
</tbody>
</table>

Two Tables are reconciled using Iterative proportional fitting
- Scale rows of PUMS table based on row sum of SF3
- Then scale each column based on column proportions

Proportions obtained from PUMS information
Yields synthetic households.....

..... that is statistically indistinguishable from census information.
Outline of this talk

• Overall architecture
• Varied uses
• **Dynamic Urban Agent Synthesis**
  - Population Synthesis
  - **Activity Generation**
  - Location Generation
  - Route Generation
  - Flow Simulation
• Dynamic Social Network Construction
• Tele-traffic analysis on Integrated Communication Network
• Case studies
Activity generator: data flow

Population Synthesis → Activity & Location Assignment → Inter-modal Routing → Vehicular Flow Simulation

Synthetic Population

Household Activity Survey
- representative sample of population
- including travel and activity participation of all household members
- recorded continuously for 24+ hours

Activity Generator

Activities
- participants
- activity type
- activity priority
- starting time, ending time, duration (preferences and bounds)
- mode preference
- vehicle preference
- possible locations

Network Data
- nodes
- links
- activity locations (includes land use and employment)
Step 1b Assigning activities patterns: algorithm overview

- Create skeletal patterns from the survey.
- Construct a classification and regression tree $T$ (CART) to partition the survey households.
- Each survey household is assigned to a leaf $l$ of $T$.
- Use household demographics as partitioning variables.
- Assign each synthetic household to a unique is leaf $l$ by applying decision rules in the tree $T$.
- Select a survey household at random from those assigned to $l$.
- Assign the skeletal patterns for the survey household members to the matching members in the synthesized household.
Step 1b Assigning activity patterns: using CART algorithm

Diagram showing decision tree based on variables such as household income, ages 5 to 17, and household size.
Times spent at these activities ....

<table>
<thead>
<tr>
<th>Activity</th>
<th>Time (Min)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Home</td>
<td>465</td>
</tr>
<tr>
<td>Work</td>
<td>225</td>
</tr>
<tr>
<td>Other</td>
<td>45</td>
</tr>
<tr>
<td>Work</td>
<td>245</td>
</tr>
<tr>
<td>Home</td>
<td>135</td>
</tr>
<tr>
<td>Other</td>
<td>60</td>
</tr>
<tr>
<td>Home</td>
<td>150</td>
</tr>
<tr>
<td>Home</td>
<td>465</td>
</tr>
<tr>
<td>Other</td>
<td>5</td>
</tr>
<tr>
<td>Other</td>
<td>30</td>
</tr>
<tr>
<td>Other</td>
<td>11</td>
</tr>
<tr>
<td>Home</td>
<td>141</td>
</tr>
<tr>
<td>Other</td>
<td>110</td>
</tr>
<tr>
<td>Home</td>
<td>240</td>
</tr>
<tr>
<td>Home</td>
<td>465</td>
</tr>
<tr>
<td>Other</td>
<td>5</td>
</tr>
<tr>
<td>Other</td>
<td>30</td>
</tr>
<tr>
<td>Other</td>
<td>11</td>
</tr>
<tr>
<td>Home</td>
<td>141</td>
</tr>
<tr>
<td>Other</td>
<td>110</td>
</tr>
<tr>
<td>Home</td>
<td>240</td>
</tr>
</tbody>
</table>

Person 1 Age = 40
Person 2 Age = 29
Person-3 Age = 28
Person-4 Age = 56

6 AM  noon  6 PM
Outline of this talk

• Overall architecture
• Varied uses
• **Dynamic Urban Agent Synthesis**
  – Population Synthesis
  – Activity Generation
  – Location Generation
  – Route Generation
  – Flow Simulation
• Dynamic Social Network Construction
• Tele-traffic analysis on Integrated Communication Network
• Case studies
Gravity models for location choice on tours

1. Choose anchor locations on tour:
   \[ P(i | j, a, m) \propto A(a,j) \exp(\beta_{am}D_{ij}) \]

2. Choose other locations on tour:
   \[ P(k | i, j, a, m) \propto A(a,k) \exp(\beta_{am}(D_{ik} + D_{kj})) \]
Yields activities locations

- Home
- Work
- Day Care
- Food
- Gym
- Lunch
- Shop

first person in household
second person in household
Outline of this talk

• Overall architecture
• Varied uses
  • Dynamic Urban Agent Synthesis
    – Population Synthesis
    – Activity Generation
    – Location Generation
    – Route Generation
    – Flow Simulation
• Dynamic Social Network Construction
• Tele-traffic analysis on Integrated Communication Network
• Case studies
Routing individuals: data flow

Population Synthesis → Activity & Location Assignment → Inter-modal Routing → Vehicular Flow Simulation

Link Travel Times
Vehicles
Activities
Transit Data
- route paths in network
- schedule of stops
- driver plans
- vehicle properties (e.g. bus capacity)

Network Data
- nodes
- links
- lane connectivity
- activity locations
- parking places & transit stops
- "process" links

Traveler Plans
- vehicle start and finish parking locations
- vehicle path through network
- expected arrival times along path
- travelers (driver and passengers) present in vehicle
- traveler mode changes

Route Planner

Network Dynamics and Simulation Science Laboratory
Chicago transportation network links

Multi-modal Transportation Network

Households with no vehicles
Route planning: algorithm

Network Dynamics and Simulation Science Laboratory
Examples of routes produced
Route density by Time of Day
Outline of this talk

- Overall architecture
- Varied uses
- **Dynamic Urban Agent Synthesis**
  - Population Synthesis
  - Activity Generation
  - Location Generation
  - Route Generation
  - Flow Simulation
- Dynamic Social Network Construction
- Tele-traffic analysis on Integrated Communication Network
- Case studies
Module 4: Cellular Automaton Microsimulation

- Population Synthesis
- Activity & Location Assignment
- Inter-modal Routing
- Vehicular Flow Simulation

Intersection with multiple turn buffers (not internally divided into grid cells)

- Single-cell vehicle
- Multiple-cell vehicle

7.5 meter × 1 lane cellular automaton grid cells
Synthetic dynamic urban population

Demographics:
- Age, Gender, Income, Job, Household size, vehicles, etc

Activities of every person

Statistics:
- 280 million synthetic people
- 129 million synthetic locations
- 1.5 billion activities
- 150 gigabytes
- 2800 compute-hours
Traffic model
Outline of this talk

- Overall architecture
- Varied uses
- Dynamic Urban Agent Synthesis
- Dynamic Social Network Construction
- Tele-traffic analysis on Integrated Communication Network
- Case studies
Constructing a social contact network

- The model knows where every person is at every second, this allows us to know:
  - who contacts who
  - how long the contact lasts
  - in what context this contact occurred (work, home)
Dynamic social contact networks based on co-location

People (8 million)

Vertex attributes:
- age
- household size
- gender
- income
- ...

Edge attributes:
- activity type: shop, work, school
- (start time 1, end time 1)
- (start time 2, end time 2)
- ...

Locations (1 million)

Vertex attributes:
- (x,y,z)
- land use
- ...

Network Dynamics and Simulation Science Laboratory
Social contact network of friends, family and business

Office Links: John, Ron
Family Links: Joe, Jill, Shawn, Mary, Jane
Friendship Links: Jill, Shawn, Joe, Mar, Tim
Outline of this talk

- Overall architecture
- Varied uses
- Dynamic Urban Agent Synthesis
- Dynamic Social Network Construction
- **Tele-traffic analysis on Integrated Communication Network**
  - Device assignment
  - Session generation
  - Network construction
  - Packet Simulation
  - Storage and regeneration of packet data
  - Mathematical Programming framework
- Case studies
A generic integrated communication network

System Mobility: UPMoST Technology

[Diagram showing various networks including Satellite Network, Wireline/Basestation Network, and Radio Packet Network]
Architecture for Tele-traffic analysis on Integrated Communication Network

Dynamic Social Network

Tele-traffic Generation

Device Assignment

Communication Network Construction

Packet/Data Flow Simulation

Storage, Analysis and Regeneration of Data

Mathematical Programming methods for Capacity
Outline of this talk

- Overall architecture
- Varied uses
- Dynamic Urban Agent Synthesis
- Dynamic Social Network Construction
- Tele-traffic analysis on Integrated Communication Network
  - Device assignment
  - Session generation
  - Network construction
  - Packet Simulation
  - Storage and regeneration of packet data
  - Mathematical Programming framework

- Case studies
Device assignment: data flow

- Dynamic Social Network
  - Locations
  - Demographics
  - Activities

- User Survey
  - Ownership statistics
  - Locations
  - Demographics

- Devices Characteristics
  - Power
  - Range
  - Wireless/Wireline

Device Assignment

- Tele-traffic Generation
- Packet/Data Flow Simulation
- Storage, Analysis and Regeneration of Data
- Mathematical Programming methods for Capacity

Device Information
- Time varying location
- Device time varying properties
- Device demographics
Wireless Device Assignment (ownership)

- People are assigned mobile devices to match CDC data
  - based on the demographic characteristics (household income, age and workers in the household, etc.).
  - Assignment based on classification and regression trees (CART) technique
Device assignment: example

- **Activity Locations**
- **Streets**
- **Block Group**

**John Doe**
- Age: 37
- Dest: Boeing
- Income: $37K

**Cell phone/PDA**
Outline of this talk

• Overall architecture
• Varied uses
• Dynamic Urban Agent Synthesis
• Dynamic Social Network Construction
• Tele-traffic analysis on Integrated Communication Network
  – Device assignment
  – Session generation
  – Network construction
  – Packet Simulation
  – Storage and regeneration of packet data
  – Mathematical Programming framework
• Case studies
Session generation: data flow

Dynamic Social Network
- Locations
- Demographics
- Activities

Device Information
- Time varying location
- Device time varying properties
- Device demographics

Social Network
- Friend, professional

Dynamic Spatial Calling Network
- Who is calling whom
- How long sessions last
- Kind of session

Tele-traffic Generation

Device Assignment

Packet/Data Flow Simulation

Communication Network Construction

Storage, Analysis and Regeneration of Data

Mathematical Programming methods for Capacity

Session Generation

Device Information

Network Dynamics and Simulation Science Laboratory
Session generation: architecture

- Session Generator (SG) is a discrete-event simulator to model who calls whom and when
  - Inputs
    - Statistics for call arrivals patterns
    - Social network of each individual
    - Activity of the individual
  - Method
    - Randomly select # calls in an interval
    - Select a random caller from the population
    - Select callee from caller’s social network
    - Determine call duration from input statistics
      - Depends on individual’s activities
  - Output
    - Session information for each mobile or landline calls
  - Validation
    - Session generation output matches the input statistics
Session generation: example

John Doe
- In car
- Age = 34
- Income > $26k

Jane Smith
- At Work
- Age = 57
- Income > $100k

Data
- 14.5 kbps
- 3.48 minutes
Spatio-Temporal Analysis: Experimental Design

- Location: Portland, OR
- Cell Size: 6.9 × 5.1mi$^2$ and 2.21 × 1.62mi$^2$
- Simulation During: 12:00am + 1 day
- Number of Seeds in Session Generation: 10
- Callers selected uniformly at random and callees from the caller’s social network.
- Metrics:
  - Call Duration Distribution
  - Hourly Call Arrivals
  - Peak Load Distribution
  - Cell Size Influence on Load CDF
  - Spatial View of Hourly Peak Load

Input distributions (Wilkomme et al., DySPAN, 2008)
Spatio-Temporal Analysis: Hourly Call arrival rate and intensity

- number of calls occurring within entire region during each hour of the day.
- Average matches the distribution from Sprint network.
- Call intensity: peak at downtown

Spatio-temporal variation in call intensity
Spatio-Temporal Analysis: Peak Load Distribution

- maximum number of simultaneous calls at a given cell tower during a given time interval (usually 1 hour).
- We study the difference between the hourly load CDF and the daily load CDF by Kolmogorov-Smirnov statistic.
- Load distribution does not simply follow one distribution for the duration of the day.
- Load distribution varies spatially.

942- Tower in central business area
Spatio-Temporal Analysis: Cell Size Influence on Load CDF

- Cell size: area covered by one cell tower
- Lower load on the smaller cells. (Large Size: 247 cells in the region; Small Size: 2109 cells in the same region)
- This result indicates SSRSM can help service provider to optimize the power and avenue for by devising optimal cell location and size.
Spatio-Temporal Analysis

- **Spatial View of Hourly Peak Load:** maximum number of simultaneous calls at a given cell tower during the hour.
- **Natural variations associated with urban mobility**
  - Load is concentrated in business areas during working hours (9:00am-5:00pm).
  - Load is dispersed to suburban area during offpeak hours.
  - Blank areas are regions with low or no inhabitants.
- **During working hours (9:00am-5:00pm), spatial load patterns (spatial profiles) are very similar.**
Application: Effect of Activity Change on Spectrum Usage

- Assume altered calling pattern during morning commute hours
  - Increased calls by people on the way to work and at work during morning calls
- Causal Behavioral modeling
  - Yields increase (spatially and temporally heterogeneous) in traffic as a result of behavioral change rather than statistically assuming that it will change by a fixed fraction

Increased call arrivals
Application: Impact of Cascading Hotspots

- Hotspots form due to traffic congestion or emergencies
- Hotspots can cascade
  - Simple model: if a tower becomes heavily loaded, it can spill over to other neighboring heavily loaded cells, with some probability $p$.
- Goal: quantify impact on total load affected
Outline of this talk

• Overall architecture
• Varied uses
• Dynamic Urban Agent Synthesis
• Dynamic Social Network Construction
• **Tele-traffic analysis on Integrated Communication Network**
  - Device assignment
  - Session generation
  - Network construction
  - Packet Simulation
  - Storage and regeneration of packet data
  - Mathematical Programming framework
• Case studies
Construction of dynamic wireless Network

Occlusion

No connection

radio range

radi

Tele-traffic Generation

Device Assignment

Pocket/Data Flow Simulation

Communication Network Construction

Storage, Analysis and Regeneration of Data

Mathematical Programming methods for Capacity
Dynamic vehicular ad-hoc network

Timestep: 200

Snapshot of ad hoc network

Dynamic network; Radio range = 75m
Building realistic Bluetooth networks

- **Step 1:** TRANSIMS [Beckman et al. 1996, Barrett et al. 2000] generates data for Activity-based mobility model (ABMM)
- **Step 2:** *Sub-location Modeling* – constructs a wireless network within each location
  - Assign an area to each location based on occupancy
  - Assign random positions to each individual
  - Construct a geometric random graph
- Can be used to model different networks

Degree distribution at different times at a single location

**Grid Approximation Model:**
To construct device contact network
Building realistic Bluetooth networks

- **Step 1:** TRANSIMS [Beckman et al. 1996, Barrett et al. 2000] generates data for Activity-based mobility model (ABMM)
- **Step 2:** Sub-location Modeling – constructs a wireless network within each location
  - Assign an area to each location based on occupancy
  - Assign random positions to each individual
  - Construct a geometric random graph
- Can be used to model different networks

![Degree distribution at different times at a single location](image)

**Grid Approximation Model:**
To construct device contact network
Outline of this talk

• Overall architecture
• Varied uses
• Dynamic Urban Agent Synthesis
• Dynamic Social Network Construction
• Tele-traffic analysis on Integrated Communication Network
  – Device assignment
  – Session generation
  – Network construction
  – Packet Simulation
  – Storage and regeneration of packet data
  – Mathematical Programming framework
• Case studies
Packet level simulators

- Detailed protocol level simulation on general ad-hoc wireless networks: ns-2, GloMoSim, Opnet, Qualnet
- Additional sensor network simulators: TOSSIM, Sensorsim
- Hybrid packet/fluid flow simulators: [Liu et al., 2001], [Kiddle et al, 2003]
- Testbeds and Emulation systems
  - Netbed from University of Utah
  - Winlab from Rutgers University
Outline of this talk

• Overall architecture
• Varied uses
• Dynamic Urban Agent Synthesis
• Dynamic Social Network Construction
• Tele-traffic analysis on Integrated Communication Network
• Case studies
Case Study 1: Dynamic spectrum analysis and management
Scenario: Spectrum Management

- Wireless companies want to bid for restricted bandwidth
  - Time varying demands needed for making good bids
  - Intelligent bids can be made based on geographic call patterns
  - FCC needs to ensure no collusion and bidding is fair

- Tools needed
  - Models for mobility and call patterns
  - Efficient methods to study detailed agent based market mechanisms
  - Behavioral models of market player: e.g. speculation and collusive behavior
Dynamic Spectrum Market Operation

Market clearing

FCC

Bids for spectrum

AT&T

Sprint

Verizon

WSP demands based on user demands

Allocation may not be adequate - affects QoS for consumers

Users switch provider if low quality - affects demands

VBI
Overall Architecture of SIGMA-SPECTRUM

- Synthetic demand model
- Market clearing models: e.g. efficient ascending bid auction
- Behavioral models of market player: e.g. speculation and collusive behavior
- Physical interference models for channel allocation
Market Model

Efficient market clearing mechanism:
- FCC: Ausubel’s ascending bid auction method to allocate the spectrum licenses

Advantage over current methods by FCC:
- Motivates bidders to bid truthfully
- Efficiently allocates licenses to bidders who value them the most
- Operates in open and transparent manner
- Preserves the privacy of the bidders
- Shares the virtues with Vickrey auction but prevents possible corruption and more efficient
Auction Mechanism

Example: 6 licenses and 4 SPs in auction

<table>
<thead>
<tr>
<th>Bid $</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>Total Demand</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 M</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>10</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Value</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>unit 1</td>
<td>5.5</td>
<td>7.5</td>
<td>7</td>
<td>4.5</td>
<td></td>
</tr>
<tr>
<td>unit 2</td>
<td>5</td>
<td>6</td>
<td>3</td>
<td>3.5</td>
<td></td>
</tr>
<tr>
<td>unit 3</td>
<td>4</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
</tbody>
</table>

Total demand: 10 > 6 => No winner in this round
Example: 6 licenses and 4 SPs in auction

<table>
<thead>
<tr>
<th>Bid $</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>Total Demand</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 M</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>10</td>
</tr>
<tr>
<td>3 M</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>8</td>
</tr>
</tbody>
</table>

A’s rivals’ demand: 5 < 6 => A is guaranteed to win 1 unit
Market Model

Example: 6 licenses and 4 SPs in auction

<table>
<thead>
<tr>
<th>Bid $</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>Total Demand</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 M</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>10</td>
</tr>
<tr>
<td>3 M</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>8</td>
</tr>
<tr>
<td>3.5 M</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>7</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Value</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>Total Demand</th>
</tr>
</thead>
<tbody>
<tr>
<td>unit 1</td>
<td>5.5</td>
<td>7.5</td>
<td>7</td>
<td>4.5</td>
<td></td>
</tr>
<tr>
<td>unit 2</td>
<td>5</td>
<td>6</td>
<td>3</td>
<td>3.5</td>
<td></td>
</tr>
<tr>
<td>unit 3</td>
<td>4</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
</tbody>
</table>

A’s rivals’ demand: 4 < 6 => A is guaranteed to win another unit
**Market Model**

Example: 6 licenses and 4 SPs in auction

<table>
<thead>
<tr>
<th>Bid $</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>Total Demand</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 M</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>10</td>
</tr>
<tr>
<td>3 M</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>8</td>
</tr>
<tr>
<td>3.5 M</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>7</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Value</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>unit 1</td>
<td>5.5</td>
<td>7.5</td>
<td>7</td>
<td>4.5</td>
<td></td>
</tr>
<tr>
<td>unit 2</td>
<td>5</td>
<td>6</td>
<td>3</td>
<td>3.5</td>
<td></td>
</tr>
<tr>
<td>unit 3</td>
<td>4</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
</tbody>
</table>

B’s rivals’ demand: 5 < 6 => B is guaranteed to win 1 unit

**Winning Board**

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>3 M</td>
</tr>
<tr>
<td>A</td>
<td>3.5 M</td>
</tr>
<tr>
<td>B</td>
<td>3.5 M</td>
</tr>
</tbody>
</table>
Example: 6 licenses and 4 SPs in auction

<table>
<thead>
<tr>
<th>Bid $</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>Total Demand</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 M</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>10</td>
</tr>
<tr>
<td>3 M</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>8</td>
</tr>
<tr>
<td>3.5 M</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>7</td>
</tr>
<tr>
<td>4 M</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>6</td>
</tr>
</tbody>
</table>

Total demand: 6 = Supplies => Auction ends
Market Model

Greedy Channel Allocation:

• A graph-coloring-based heuristic
• Allocate the licenses with the smallest number of channels to satisfy the load

Supply Assumptions:

• FCC is the only supplier in the primary market
• Fixed number of licenses in FCC auction
• No cost to FCC in obtaining and auctioning the spectrum (Auction revenue = FCC profit)
• Supply curve of FCC is a vertical line (FCC is willing to sell licenses at the highest possible price.)
Experiments: Setup

- Location: Portland, OR
- 10 Licenses in auction
- 6 Service Providers (A, B..., F) and 2 Speculators (G, H)
- Market Share of Providers: 29%, 30%, 18%, 12%, 6%, 5%
- Minimum Bid: $1 million
- Reservation Price: $350K
Experiments

- Demand is generated from the greedy channel allocation to satisfy the peak load in the Portland area.

Spatial View of Hourly Peak Load
Experimental Design: 4 Cases

- Base Case
- Variation in Demand: Service providers alter their true demands
- Collusive Behaviors
- Reduced License Capacity

| Base Case      | 1 | Original valuations. |
|               | 2 | Base0, allocations when only service providers bid. |
|               | 3 | Base1, allocations with service providers plus one speculator. |
|               | 4 | Base2, allocations with service providers plus two speculators. |
| Variation in Demand | 5 | Valuations under reduced demand. |
|               | 6 | Allocations under reduced demand. |
|               | 7 | Valuations under increased demand. |
|               | 8 | Allocations under increased demand. |
|               | 9 | Valuations under 50% increase, 50% decrease but no net change in demand. |
|               | 10 | Allocations, 50% increase, 50% decrease, no net change in demand. |
|               | 11 | Valuations under random change in demand. |
|               | 12 | Allocations, random change in demand. |

| Collusive Behavior | 13 | Allocations, collusion among service providers, valuations as in Table II. |
|                   | 14 | Allocations, collusion among speculators, valuations as in Table II. |

| Reduced License Capacity | 15 | Valuations under reduced license capacity of 5%. |
|                         | 16 | Allocations, reduced license capacity, only service providers bid. |
|                         | 17 | Allocations, reduced license capacity, service providers and one speculator bid. |
|                         | 18 | Allocations, reduced license capacity, service providers and both speculators bid. |
Analysis

- **Base Case**
  - The addition of speculators to the market raises the prices on all licenses.

- **Demand Manipulation**
  - Even when a truthful mechanism is in place, the uncertainties about the secondary market can greatly influence the bidding strategies of the market players.

- **Collusive Behavior**
  - Collusive partners can lead to substantially decreased cost for SPs

- **Reduced Capacity of Licenses**
  - Bidders can reduce their demand to better match their needs.
  - Finer split of the license capacity leads to higher market efficiency.
Analysis: Efficiency of License Allocation

- The excess ratio for all service providers drops significantly during peak hours.
- Service providers (SP) who get more than needed bandwidth have higher excess ratios.
Analysis: Effect of bidding strategies

- Entry of speculator in the market reduces the excess ratio because the speculator wins the license that was allocated to SP in base case.
- Reduced license capacity shows a lower positive excess ratio, i.e., higher efficiency, than base case 0.
- Spatial and temporal view of excess channels (base case +2 speculators).
  - Daytime hours have less surplus channels than night time hours
  - Downtown has significant variation in channel usage between daytime & nighttime
  - Other areas than downtown have sufficient channels to meet the load
Summary of Results

- Developed a microscopic agent based tool for analyzing wireless spectrum market
- Case Study:
  - The possibility of trading in the secondary market has significant repercussions on the bidding behavior of the service providers in the primary market.
  - With DSA, speculators have incentive to join the market and make profits through arbitrage.
  - Bidders can collude to save the auction cost and split the capacity later.
  - The finer split of the license capacity makes it a more efficient market.
Case Study II: Cybervulnerability of wireless networks: dynamics of worm propagation

Image from “Malware goes Mobile,” Scientific American, 2006
Epidemics in wireless cognitive networks

New issues in cyber-security

- Ubiquity of smart digital devices (20.7 million devices +): increased risk of malware attacks
- Multiple scales ranging from Bluetooth networks to Internet
- Self-forming and dynamic networks resistant to common regulation
- Need to guard against sophisticated worms that can attack and spread on multiple networks
- Goal: efficient tools for understanding and control of the spread of worms

Our approach

- EpiNet: scalable simulation tool for study of Bluetooth worms motivated by epidemics on human contact networks
- Synthetic urban mobility using activity based model

"Human mobility and wireless networking could combine to abet the spread of computer viruses”
- Jon Kleinberg  [Nature 2007]
The EpiNet modeling framework

- **Step 1:** Construct realistic human mobility patterns using TRANSIMS [Barrett et al., '00]

- **Step 2:** Construct Bluetooth proximity network by combining mobility pattern with location model

- **Step 3:** Build a within-host abstract model of the malware's behavior (how malware moves from one behavioral state to another)

- **Step 4:** Model for representing how malware spreads when devices interact (e.g. independent cascade or threshold models, deterministic versus probabilistic, dose model, etc)

- **Step 5:** Model for detection and response strategies for answering epidemiological science questions (Passive self detection, and signature dissemination)
**Step 1:** Synthetic population, activities and assigning devices

- **Step 1:** TRANSIMS [Beckman et al. 1996, Barrett et al. 2000] generates data for Activity-based mobility model (ABMM)
  - Census data to construct synthetic population
  - Activity surveys to construct activities
  - Device assignment based on National Health Institute Surveys
Step 2: Building realistic Bluetooth networks

- **Step 2: Sub-location Modeling** – constructs a wireless network within each location
  - Assign an area to each location based on occupancy
  - Assign random positions to each individual
  - Construct a geometric random graph

**Degree distribution**

At different times at a single location

**Grid Approximation Model**

To construct device contact network
Step 3: Building within-host model for the malware

- Using the protocol description of the malware
  - We construct a probabilistic timed transition system (PTTS) for the Bluetooth malware
  - Model is parameterized by the malware and Bluetooth protocol

Diagram:

- Worm Model
- Worm Behavior + Wireless Protocol
- Construction of Abstract Worm Model
Step 3: Calibration and validation of the model

- Calibrated by detailed simulations
  - UCBT model for Bluetooth
  - Calibration for small instances
- The model is validated with detailed simulations
  - The infection growth with EpiNet tracks the detailed simulation very closely
## Comparison with prior approaches

<table>
<thead>
<tr>
<th>Factors</th>
<th>Mathematical Models [Yan, ICDCS ‘06]</th>
<th>Simulation based computational models</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scope</td>
<td>1 location/city area</td>
<td>1 location</td>
</tr>
<tr>
<td>Temporal Scale</td>
<td>1 second</td>
<td>ms. / µs.</td>
</tr>
<tr>
<td>Spatial Scale</td>
<td>meters</td>
<td>meters</td>
</tr>
<tr>
<td>Mobility model</td>
<td>Random waypoint model</td>
<td>Random Waypoint, Random Walk, Random Landmark</td>
</tr>
<tr>
<td>Device interaction network</td>
<td>Dependent on mobility model parameters</td>
<td>Based on mobility models</td>
</tr>
<tr>
<td>Within-host Malware Model</td>
<td>Analytical expression</td>
<td>Detailed implementation</td>
</tr>
<tr>
<td>Detection</td>
<td>Can be implemented</td>
<td>Not studied, difficult to implement</td>
</tr>
<tr>
<td>Control mechanisms</td>
<td>Can be implemented, but limited by network size</td>
<td>Can be implemented</td>
</tr>
<tr>
<td>Network co-evolution</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Summary of results

1. Computational scaling
   - Sequential EpiNet 100x faster than NS-2
   - Parallel EpiNet can simulate networks with millions of devices (1.6 Million node system in about an hour)
   - Speedups are obtained with very little loss in accuracy (no more than 5%)

2. Mobility matters:
   - Dynamics of the malware spread are significantly affected by human mobility
   - Bluetooth malware propagates slowly providing opportunity for control

3. Network parameters have significant impact on spread

4. Targeted intervention schemes based adaptive detection more effective
   - Interventions based on static graph metrics have limited efficacy
   - Device-based detection and automatic signature generation approaches work better to control the spread
# Wireless epidemiology study: Simulation setup

<table>
<thead>
<tr>
<th>Factorial experiment design</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Network</strong></td>
</tr>
<tr>
<td>Area</td>
</tr>
<tr>
<td>Demographics</td>
</tr>
<tr>
<td>People (devices); locations</td>
</tr>
<tr>
<td>Smart device ownership</td>
</tr>
<tr>
<td><strong>Simulation</strong></td>
</tr>
<tr>
<td>Replicates</td>
</tr>
<tr>
<td>Duration of Simulation</td>
</tr>
<tr>
<td>Initially infected</td>
</tr>
<tr>
<td>Wallclock</td>
</tr>
<tr>
<td>Infection seed</td>
</tr>
<tr>
<td>Sensitivity analysis</td>
</tr>
<tr>
<td>Malware parameters</td>
</tr>
<tr>
<td>Network parameters:</td>
</tr>
<tr>
<td><strong>Response mechanisms</strong></td>
</tr>
<tr>
<td>Static</td>
</tr>
<tr>
<td>Device-based detection</td>
</tr>
<tr>
<td><strong>Results</strong></td>
</tr>
<tr>
<td>Cumulative infection size</td>
</tr>
<tr>
<td>$T(q,x)$: time taken to infect $q$ percent of devices when $x$ is varied</td>
</tr>
</tbody>
</table>
Scaling timeline of the EpiNet simulator

**Ns-2**
- 500 devices
- 30-40 hours

**EpiNet V1**
- 500 devices
- 25 minutes

**EpiNet V1**
- 30000 devices
- 40-45 hours

**EpiNet V2**
- 30000 devices
- 2 hours

**EpiNet V3**
- 1.6 million devices
- 50 minutes

**Scaling Studies**

**Direct conversion from EpiSimdemics**
1. A day → A second
2. Abstract malware model
3. Sub-location modeling

**Problem**
Communication overhead

**Major Modifications**
1. Simulation in seconds
2. Event structure optimized
3. Optimize communication
4. Mobility in 5 minute intervals

**Problem**
Model Detail

**Model Reduction**
1. Bluetooth model abstracted through model reduction
2. Simulation in TUs

**Problem**
Model Detail
Model reduction to improve scalability

- Problems with detailed model
  - Simulation time resolution = 1 second
  - Detailed model state space is large
    - Requires more memory
    - Slows down simulation
- Solution: Perform simulation in TUs (discrete interval)
  - Gillespie’s algorithm to next event
  - Offline State traversal
    - Probability of infection \((p)\)
    - Time the device remains infectious \((T_{inf})\)
Model reduction to improve scalability

- Problems with detailed model
  - Simulation time resolution = 1 second
  - Detailed model state space is large
    - Requires more memory
    - Slows down simulation
- Solution: Perform simulation in TUs (discrete interval)
  - Offline State traversal
    - Probability of infection \( (p) \)
    - Time the device remains infectious \( (T_{inf}) \)
- Preliminary results
  - Simulation of 1.6 million devices in less than an hour
  - Error is about 5%
Summary of results

1. Computational scaling
   - Sequential EpiNet 100x faster than NS-2
   - Parallel EpiNet can simulate networks with millions of devices (1.6 Million node system in about an hour)
   - Speedups are obtained with very little loss in accuracy (no more than 5%)

2. Mobility matters
   - Dynamics of the malware spread are significantly affected by human mobility
   - Bluetooth malware propagates slowly providing opportunity for control

3. Network parameters have significant impact on spread

4. Targeted intervention schemes based adaptive detection more effective
   - Interventions based on static graph metrics have limited efficacy
   - Device-based detection and automatic signature generation approaches work better to control the spread
Mobility matters!

- **RWP** [Nodes: 109, Area: 100 m$^2$, Pause: 300s (600 s)]
  - 1 infected device
- **ABMM** [Nodes: 91-141, activity-based mobility]
  - 1%, 5%, 10% devices infected
- **Network structure**
  - Degree distribution
  - Density of the location
- **Conclusions**
  - Malware spreads to more devices for RWP
  - ABMM requires a larger number of initial infections to cause a noticeable spread
  - ABMM has a faster initial spread, but fairly quickly saturates
  - **Realistic mobility alters the conclusions completely:** we see completely different dynamics
Mobility matters!

- **RWP** [Nodes: 109, Area: 100 m², Pause: 300s (600 s)]
  - 1 infected device
- **ABMM** [Nodes: 91-141, activity-based mobility]
  - 1%, 5%, 10% devices infected
- Network structure
  - Degree distribution
  - Density of the location

**Conclusions**
- Malware spreads to more devices for RWP
- ABMM requires a larger number of initial infections to cause a noticeable spread
- ABMM has a faster initial spread, but fairly quickly saturates
- Realistic mobility alters the conclusions completely: we see completely different dynamics
Summary of results

1. Computational scaling
   - Sequential EpiNet 100x faster than NS-2
   - Parallel EpiNet can simulate networks with millions of devices (1.6 Million node system in about an hour)
   - Speedups are obtained with very little loss in accuracy (no more than 5%)

2. Mobility matters:
   - Dynamics of the malware spread are significantly affected by human mobility
   - Bluetooth malware propagates slowly providing opportunity for control

3. Network parameters have significant impact on spread

4. Targeted intervention schemes based adaptive detection more effective
   - Interventions based on static graph metrics have limited efficacy
   - Device-based detection and automatic signature generation approaches work better to control the spread
Network structure

- Union graph: combines contact graphs at all times
- Network consists of a large number of disconnected components
  - Largest component size is approx. 8000
  - Significant number of small components
- Degree distribution is for a particular hour is exponential

Component sizes in the Chicago network (Union graph)

Degree distribution of the Chicago network at different time of the day
Effect of network parameters: Market share ($m$)

- Market share of mobile device operating system
  - Defines set of devices with a particular vulnerability
  - $m$ defines the percentage of devices susceptible

- Conclusions
  - Network has several disconnected components
  - Market share fragments the network further
  - We observe a distinct threshold effect
    - Significant change in $T(q,m)$ to infect greater than 20% of devices
  - The speed of the spread allows for response mechanisms
  - Faster spread than reported by [Wang et. al., Science ‘09]
    - Fidelity of the model
    - Pair-interaction model being accurate
Summary of results

1. Computational scaling
   - Sequential EpiNet 100x faster than NS-2
   - Parallel EpiNet can simulate networks with millions of devices (1.6 Million node system in about an hour)
   - Speedups are obtained with very little loss in accuracy (no more than 5%)

2. Mobility matters:
   - Dynamics of the malware spread are significantly affected by human mobility
   - Bluetooth malware propagates slowly providing opportunity for control

3. Network parameters have significant impact on spread

4. Targeted intervention schemes based adaptive detection more effective
   - Interventions based on static graph metrics have limited efficacy
   - Device-based detection and automatic signature generation approaches work better to control the spread
Interventions to control spread of malware

- Control strategy: selecting devices to apply software patches
  - Static graph metrics
    - Degree based selection criteria not much better than random strategy
    - Requires a larger set of devices to achieve better control
- Centralized control based on ‘infection reports’
  - Accuracy depends on the detection strategy
  - Early and accurate detection achieves limited improvement
  - Large scale patching required for effective control
Summary

- Described a disaggregated simulation based methodology to
  - Represent synthetic coupled social and communication networks
  - Analyze practical applications that require such coupled representations

- Technology is scalable
  - 10 million individuals & devices, spatial resolution ~ few meters, temporal resolution ~ few seconds.
  - Simulation used for integrating diverse data sets and creating new dynamic information using interaction based models.

- Applications include
  - Design and analysis of cellular networks
  - Spectrum markets for cognitive radio networks
  - Design and analysis of vehicular ad-hoc networks
  - Sensing and monitoring applications involving sensor networks
Take Home Message

• Urban Communication networks do not operate in isolation
  – Detailed representation of coupled social and communication networks is not optional
• Coupled Social and Communication networks are complex systems
  – Open, varied in their ownership, dynamic
  – Validation/verification of these models represents hard and open questions
• An Interaction-based high resolution approach is possible to comprehend these systems
  – Necessarily uses high performance computing resources
  – Has been demonstrated to be useful in analyzing important scientific and practical questions
Thank You