Understanding Communication Networks

Ljiljana Trajković
ljilja@cs.sfu.ca

Communication Networks Laboratory
http://www.ensc.sfu.ca/cnl
School of Engineering Science
Simon Fraser University, Vancouver, British Columbia
Canada
Roadmap

- Introduction
- Traffic measurements and analysis tools
- Case study:
  - public safety wireless network: E-Comm
- Collection of BCNET traffic
- Internet topology and spectral analysis of Internet graphs
- Conclusions
ihr (535,102 nodes and 601,678 links)
Roadmap

- Introduction
- Traffic measurements and analysis tools
- Case study:
  - public safety wireless network: E-Comm
- Collection of BCNET traffic
- Internet topology and spectral analysis of Internet graphs
- Conclusions
Measurements of network traffic

- Traffic measurements:
  - help understand characteristics of network traffic
  - are basis for developing traffic models
  - are used to evaluate performance of protocols and applications

- Traffic analysis:
  - provides information about the network usage
  - helps understand the behavior of network users

- Traffic prediction:
  - important to assess future network capacity requirements
  - used to plan future network developments
Traffic modeling: self-similarity

- Self-similarity implies a “fractal-like” behavior
- Data on various time scales have similar patterns
- Implications:
  - no natural length of bursts
  - bursts exist across many time scales
  - traffic does not become “smoother” when aggregated
  - it is unlike Poisson traffic used to model traffic in telephone networks
  - as the traffic volume increases, the traffic becomes more bursty and more self-similar
Self-similarity: influence of time-scales

- Genuine MPEG traffic trace

![Graphs showing self-similarity](image)
Self-similarity: influence of time-scales

- Synthetically generated Poisson model
Traffic analysis: clustering analysis

- Clustering generates groups (clusters) of similar objects
- An object is described by a set of measurements
- Clustering algorithms can be used to analyze behavior of network users
- Users are grouped into clusters based on the similarity of their behavior
- Traffic prediction based on clusters is simplified to predicting users' traffic from few clusters
- Clustering tools:
  - k-means algorithm
  - AutoClass tool
Traffic prediction: SARIMA model

- Auto-Regressive Integrated Moving Average (ARIMA) model:
  - general model for forecasting time series
  - past values: AutoRegressive (AR) structure
  - past random fluctuant effect: Moving Average (MA) process
- Seasonal ARIMA (SARIMA) is a variation of the ARIMA model:
  - it captures seasonal patterns
Roadmap

- Introduction
- Traffic measurements and analysis tools
- Collection of BCNET traffic
- Case study:
  - public safety wireless network: E-Comm
- Internet topology and spectral analysis of Internet graphs
- Conclusions
Case study: E-Comm network

- An operational trunked radio system serving as a regional emergency communication system
- The E-Comm network is capable of both voice and data transmissions
- Voice traffic accounts for over 99% of network traffic
- A group call is a standard call made in a trunked radio system
- More than 85% of calls are group calls
- A distributed event log database records every event occurring in the network: call establishment, channel assignment, call drop, and emergency call
E-Comm network
E-Comm network

E-Comm’s Wide-Area Radio System: Police Customers

- West Vancouver
- North Vancouver
- North Vancouver District
- Anmore
- Coquitlam
- Port Coquitlam
- Pitt Meadows
- Maple Ridge
- Richmond
- New Westminster
- Surrey
- Delta
- Langley City
- Langley Township
- White Rock

Police departments using E-Comm’s Wide-Area Radio System

GVTAPS not illustrated
E-Comm network
E-Comm network

E-Comm’s Wide-Area Radio System: Ambulance Service

The BC Ambulance Service uses the E-Comm Radio System throughout the GVRD.
E-Comm network architecture

- Users
- Transmitters/Repeaters
- PSTN
- PBX
- Dispatch console
- Network switch
- Database server
- Data gateway
- Management console
- Vancouver
- Burnaby
- Other EDACS systems
E-Comm traffic data

- 2001 data set:
  - 2 days of traffic data
    - 2001-11-01 to 2001-11-02 (110,348 calls)
- 2002 data set:
  - 28 days of continuous traffic data
    - 2002-02-10 to 2002-03-09 (1,916,943 calls)
- 2003 data set:
  - 92 days of continuous traffic data
    - 2003-03-01 to 2003-05-31 (8,756,930 calls)
E-Comm traffic data

- Records of network events:
  - established, queued, and dropped calls in the Vancouver cell
- Traffic data span periods during:

<table>
<thead>
<tr>
<th>Trace (dataset)</th>
<th>Time span</th>
<th>No. of established calls</th>
</tr>
</thead>
<tbody>
<tr>
<td>2001</td>
<td>November 1–2, 2001</td>
<td>110,348</td>
</tr>
<tr>
<td>2002</td>
<td>March 1–7, 2002</td>
<td>370,510</td>
</tr>
</tbody>
</table>
E-Comm traffic: observations

- Presence of daily cycles:
  - minimum utilization: ~2 PM
  - maximum utilization: 9 PM to 3 AM

- 2002 sample data:
  - cell 5 is the busiest
  - others seldom reach their capacities

- 2003 sample data:
  - several cells (2, 4, 7, and 9) have all channels occupied during busy hours

- The busiest hour: around midnight
- The busiest day: Thursday
- Useful for scheduling periodical maintenance tasks
E-Comm traffic: hourly traces

- Call holding and call inter-arrival times from the five busiest hours in each dataset (2001, 2002, and 2003)

<table>
<thead>
<tr>
<th></th>
<th>2001</th>
<th>2002</th>
<th>2003</th>
</tr>
</thead>
<tbody>
<tr>
<td>Day/hour</td>
<td>No.</td>
<td>Day/hour</td>
<td>No.</td>
</tr>
<tr>
<td>02.11.2001 15:00–16:00</td>
<td>3,718</td>
<td>01.03.2002 04:00–05:00</td>
<td>4,436</td>
</tr>
<tr>
<td>01.11.2001 00:00–01:00</td>
<td>3,707</td>
<td>01.03.2002 22:00–23:00</td>
<td>4,314</td>
</tr>
<tr>
<td>02.11.2001 16:00–17:00</td>
<td>3,492</td>
<td>01.03.2002 23:00–24:00</td>
<td>4,179</td>
</tr>
<tr>
<td>01.11.2001 19:00–20:00</td>
<td>3,312</td>
<td>01.03.2002 00:00–01:00</td>
<td>3,971</td>
</tr>
<tr>
<td>02.11.2001 20:00–21:00</td>
<td>3,227</td>
<td>02.03.2002 00:00–01:00</td>
<td>3,939</td>
</tr>
</tbody>
</table>
E-Comm traffic: statistical distributions

- Fourteen candidate distributions:
  - exponential, Weibull, gamma, normal, lognormal, logistic, log-logistic, Nakagami, Rayleigh, Rician, t-location scale, Birnbaum-Saunders, extreme value, inverse Gaussian

- Parameters of the distributions: calculated by performing maximum likelihood estimation

- Best fitting distributions are determined by:
  - visual inspection of the distribution of the trace and the candidate distributions
  - Kolmogorov-Smirnov test of potential candidates
### Call inter-arrival and call holding times: observations

<table>
<thead>
<tr>
<th></th>
<th>2001</th>
<th></th>
<th>2002</th>
<th></th>
<th>2003</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Day/hour</td>
<td>Avg. (s)</td>
<td>Day/hour</td>
<td>Avg. (s)</td>
<td>Day/hour</td>
<td>Avg. (s)</td>
</tr>
<tr>
<td>inter-arrival</td>
<td>02.11.2001 15:00–16:00</td>
<td>0.97</td>
<td>01.03.2002 04:00–05:00</td>
<td>0.81</td>
<td>26.03.2003 22:00–23:00</td>
<td>0.73</td>
</tr>
<tr>
<td>holding</td>
<td></td>
<td>3.78</td>
<td></td>
<td></td>
<td>4.07</td>
<td></td>
</tr>
<tr>
<td>inter-arrival</td>
<td>01.11.2001 00:00–01:00</td>
<td>0.97</td>
<td>25.03.2003 23:00–24:00</td>
<td>0.85</td>
<td></td>
<td>4.12</td>
</tr>
<tr>
<td>holding</td>
<td></td>
<td>3.95</td>
<td></td>
<td></td>
<td>3.84</td>
<td></td>
</tr>
<tr>
<td>inter-arrival</td>
<td>02.11.2001 16:00–17:00</td>
<td>1.03</td>
<td>01.03.2002 23:00–24:00</td>
<td>0.86</td>
<td>26.03.2003 23:00–24:00</td>
<td>0.85</td>
</tr>
<tr>
<td>holding</td>
<td></td>
<td>3.99</td>
<td></td>
<td></td>
<td>3.88</td>
<td></td>
</tr>
<tr>
<td>inter-arrival</td>
<td>01.11.2001 19:00–20:00</td>
<td>1.09</td>
<td>29.03.2003 02:00–03:00</td>
<td>0.87</td>
<td></td>
<td>4.14</td>
</tr>
<tr>
<td>holding</td>
<td></td>
<td>3.97</td>
<td></td>
<td></td>
<td>3.95</td>
<td></td>
</tr>
<tr>
<td>inter-arrival</td>
<td>02.11.2001 20:00–21:00</td>
<td>1.12</td>
<td>29.03.2003 01:00–02:00</td>
<td>0.88</td>
<td></td>
<td>4.25</td>
</tr>
<tr>
<td>holding</td>
<td></td>
<td>3.84</td>
<td></td>
<td></td>
<td>4.06</td>
<td></td>
</tr>
</tbody>
</table>

Avg. call inter-arrival times: 1.08 s (2001), 0.86 s (2002), 0.84 s (2003)

## Busy hour: best fitting distributions

<table>
<thead>
<tr>
<th>Busy hour</th>
<th>Distribution</th>
<th>Call inter-arrival times</th>
<th>Call holding times</th>
<th>Weibull</th>
<th>Gamma</th>
<th>Lognormal</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>a</td>
<td>b</td>
<td>µ</td>
</tr>
<tr>
<td>02.11.2001 15:00–16:00</td>
<td></td>
<td></td>
<td></td>
<td>0.9785</td>
<td>1.1075</td>
<td>1.0326</td>
</tr>
<tr>
<td>01.11.2001 00:00–01:00</td>
<td></td>
<td></td>
<td></td>
<td>0.9907</td>
<td>1.0517</td>
<td>1.0818</td>
</tr>
<tr>
<td>02.11.2001 16:00–17:00</td>
<td></td>
<td></td>
<td></td>
<td>1.0651</td>
<td>1.0826</td>
<td>1.1189</td>
</tr>
<tr>
<td>01.03.2002 04:00–05:00</td>
<td></td>
<td></td>
<td></td>
<td>0.8313</td>
<td>1.0603</td>
<td>1.1096</td>
</tr>
<tr>
<td>01.03.2002 22:00–23:00</td>
<td></td>
<td></td>
<td></td>
<td>0.8532</td>
<td>1.0542</td>
<td>1.0931</td>
</tr>
<tr>
<td>01.03.2002 23:00–24:00</td>
<td></td>
<td></td>
<td></td>
<td>0.8877</td>
<td>1.0790</td>
<td>1.1308</td>
</tr>
<tr>
<td>26.03.2003 22:00–23:00</td>
<td></td>
<td></td>
<td></td>
<td>0.7475</td>
<td>1.0475</td>
<td>1.0910</td>
</tr>
<tr>
<td>25.03.2003 23:00–24:00</td>
<td></td>
<td></td>
<td></td>
<td>0.8622</td>
<td>1.0376</td>
<td>1.0762</td>
</tr>
<tr>
<td>26.03.2003 23:00–24:00</td>
<td></td>
<td></td>
<td></td>
<td>0.8579</td>
<td>1.0092</td>
<td>1.0299</td>
</tr>
</tbody>
</table>
E-Comm traffic: clustering

- E-Comm network and traffic data:
  - data preprocessing and extraction
- Data clustering
- Traffic prediction:
  - based on aggregate traffic
  - cluster based
E-Comm traffic: preprocessing

- Original database: ~6 GBytes, with 44,786,489 record rows
- Data pre-processing:
  - cleaning the database
  - filtering the outliers
  - removing redundant records
  - extracting accurate user calling activity
- After the data cleaning and extraction, number of records was reduced to only 19% of original records
## E-Comm traffic: data preparation

<table>
<thead>
<tr>
<th>Date</th>
<th>Original</th>
<th>Cleaned</th>
<th>Combined</th>
</tr>
</thead>
<tbody>
<tr>
<td>2003/03/01</td>
<td>466,862</td>
<td>204,357</td>
<td>91,143</td>
</tr>
<tr>
<td>2003/03/02</td>
<td>415,715</td>
<td>184,973</td>
<td>88,014</td>
</tr>
<tr>
<td>2003/03/03</td>
<td>406,072</td>
<td>182,311</td>
<td>76,310</td>
</tr>
<tr>
<td>2003/03/04</td>
<td>464,534</td>
<td>207,016</td>
<td>84,350</td>
</tr>
<tr>
<td>2003/03/05</td>
<td>585,561</td>
<td>264,226</td>
<td>97,714</td>
</tr>
<tr>
<td>2003/03/06</td>
<td>605,987</td>
<td>271,514</td>
<td>104,715</td>
</tr>
<tr>
<td>2003/03/07</td>
<td>546,230</td>
<td>247,902</td>
<td>94,511</td>
</tr>
<tr>
<td>2003/03/08</td>
<td>513,459</td>
<td>233,982</td>
<td>90,310</td>
</tr>
<tr>
<td>2003/03/09</td>
<td>442,662</td>
<td>201,146</td>
<td>79,815</td>
</tr>
<tr>
<td>2003/03/10</td>
<td>419,570</td>
<td>186,201</td>
<td>76,197</td>
</tr>
<tr>
<td>2003/03/11</td>
<td>504,981</td>
<td>225,604</td>
<td>88,857</td>
</tr>
<tr>
<td>2003/03/12</td>
<td>516,306</td>
<td>233,140</td>
<td>94,779</td>
</tr>
<tr>
<td>2003/03/13</td>
<td>561,253</td>
<td>255,840</td>
<td>95,662</td>
</tr>
<tr>
<td>2003/03/14</td>
<td>550,732</td>
<td>248,828</td>
<td>99,458</td>
</tr>
</tbody>
</table>

**Total 92 Days**

<table>
<thead>
<tr>
<th>Original</th>
<th>Cleaned</th>
<th>Combined</th>
</tr>
</thead>
<tbody>
<tr>
<td>44,786,489</td>
<td>20,130,718</td>
<td>8,663,586</td>
</tr>
</tbody>
</table>

- 44.95% Original
- 19.34% Cleaned

---

*2013 Consumer Electronic Forum, Taiwan*
User clusters with K-means: $k = 6$
Cluster sizes:
- 17, 31, and 569 for $K = 3$
- 17, 33, 4, and 563 for $K = 4$
- 13, 17, 22, 3, 34, and 528 for $K = 6$

$K = 3$ produces the best clustering results (based on overall clustering quality and silhouette coefficient)

Interpretations of three clusters have been confirmed by the E-Comm domain experts
E-Comm traffic: prediction

- Important to assess future network capacity requirements and to plan future network developments.
- A network traffic trace consists of a series of observations in a dynamical system environment.
- Traditional prediction: considers aggregate traffic and assumes a constant number of network users.
- Approach that focuses on individual users has high computational cost for networks with thousands of users.
- Employing clustering techniques for predicting aggregate network traffic bridges the gap between the two approaches.
Prediction: based on the aggregate traffic

- Two groups of models, with 24-hour and 168-hour seasonal periods:
  - SARIMA \( (2, 0, 9) \times (0, 1, 1)_{24} \) and \( (0, 1, 1)_{168} \)
  - SARIMA \( (2, 0, 1) \times (0, 1, 1)_{24} \) and \( (0, 1, 1)_{168} \)

- Models with a 168-hour seasonal period provided better prediction than the four 24-hour period based models, particularly when predicting long term traffic data.

- Prediction of traffic in networks with a variable number of users is possible, as long as the new users could be classified within the existing clusters.
Prediction of 168 hours of traffic based on 1,680 past hours: sample

Comparison of the 24-hour and the 168-hour models
- Solid line: observation
- o: prediction of 168-hour seasonal model
- *: prediction of 24-hour seasonal model
Prediction of 168 hours of traffic based on 1,680 past hours

Comparisons: model $(1,0,1) \times (0,1,1)_{168}$
- * observation
- * prediction without clustering
- o prediction with clustering
Roadmap

- Introduction
- Traffic measurements and analysis tools
- Case studies:
  - public safety wireless network: E-Comm
  - Collection of BCNET traffic
- Internet topology and spectral analysis of Internet graphs
- Conclusions
BCNET packet capture: physical overview

- BCNET is the hub of advanced telecommunication network in British Columbia, Canada that offers services to research and higher education institutions.
BCNET packet capture

- BCNET transits have two service providers with 10 Gbps network links and one service provider with 1 Gbps network link.
- Optical Test Access Point (TAP) splits the signal into two distinct paths.
- The signal splitting ratio from TAP may be modified.
- The Data Capture Device (NinjaBox 5000) collects the real-time data (packets) from the traffic filtering device.
Net Optics Director 7400: application diagram

- Net Optics Director 7400 is used for BCNET traffic filtering
- It directs traffic to monitoring tools such as NinjaBox 5000 and FlowMon
Network monitoring and analyzing: Endace card

- Endace Data Acquisition and Generation (DAG) 5.2X card resides inside the NinjaBox 5000
- It captures and transmits traffic and has time-stamping capability
- DAG 5.2X is a single port Peripheral Component Interconnect Extended (PCIx) card and is capable of capturing on average Ethernet traffic of 6.9 Gbps
Real time network usage by BCNET members

- The BCNET network is high-speed fiber optic research network
- British Columbia's network extends to 1,400 km and connects Kamloops, Kelowna, Prince George, Vancouver, and Victoria
Roadmap

- Introduction
- Traffic measurements and analysis tools
- Case study:
  - public safety wireless network: E-Comm
- Collection of BCNET traffic
- Internet topology and spectral analysis of Internet graphs
- Conclusions
Internet topology

- Internet is a network of Autonomous Systems:
  - groups of networks sharing the same routing policy
  - identified with Autonomous System Numbers (ASN)
- Autonomous System Numbers: [http://www.iana.org/assignments/as-numbers](http://www.iana.org/assignments/as-numbers)
- Internet topology on AS-level:
  - the arrangement of ASes and their interconnections
- Analyzing the Internet topology and finding properties of associated graphs rely on mining data and capturing information about Autonomous Systems (ASes)
Variety of graphs

- **Random graphs:**
  - nodes and edges are generated by a random process
  - Erdős and Rényi model

- **Small world graphs:**
  - nodes and edges are generated so that most of the nodes are connected by a small number of nodes in between
  - Watts and Strogatz model (1998)
Scale-free graphs

- **Scale-free graphs:**
  - graphs whose node degree distribution follow power-law
  - rich get richer
  - Barabási and Albert model (1999)

- **Analysis of complex networks:**
  - discovery of spectral properties of graphs
  - constructing matrices describing the network connectivity
Analyzed datasets

- Sample datasets:
  - Route Views:
    TABLE_DUMP| 1050122432| B| 204.42.253.253| 267| 3.0.0.0/8| 267 2914 174 701| IGP|
    204.42.253.253| 0| 0| 267:2914 2914:420 2914:2000 2914:3000| NAG| |
  - RIPE:
    TABLE_DUMP| 1041811200| B| 212.20.151.234| 13129| 3.0.0.0/8| 13129 6461 7018| IGP|
    212.20.151.234| 0| 0| 6461:5997 13129:3010| NAG| |
Datasets collected from Border Gateway Protocols (BGP) routing tables are used to infer the Internet topology at AS-level.
The Internet topology is characterized by the presence of various power-laws:
- node degree vs. node rank
- eigenvalues of the matrices describing Internet graphs (adjacency matrix and normalized Laplacian matrix)
- Power-laws exponents have not significantly changed over the years
- Spectral analysis reveals new historical trends and notable changes in the connectivity and clustering of AS nodes over the years
Roadmap

- Introduction
- Traffic measurements and analysis tools
- Case study:
  - public safety wireless network: E-Comm
- Collection of BCNET traffic
- Internet topology and spectral analysis of Internet graphs
- Conclusions
**Conclusions**

- Traffic data from deployed networks can be used to:
  - evaluate network performance
  - characterize and model traffic (inter-arrival and call holding times)
  - classify network users using clustering algorithms
  - predict network traffic by employing models based on aggregate user traffic and user clusters
- Internet datasets reveal trends in the evolution of the Internet topology
- Spectral analysis indicate that clusters of connected Internet nodes have changed over time
Ihr (535,102 nodes and 601,678 links)
Ihr (535,102 nodes and 601,678 links)