Traffic analysis of the Yukon mobile radio system

Savio Lau

saviol@cs.sfu.ca

December 01, 2004

communication
networks
laboratory

Outline

- Yukon mobile radio system data

- Traffic data relationships

- Clustering algorithms

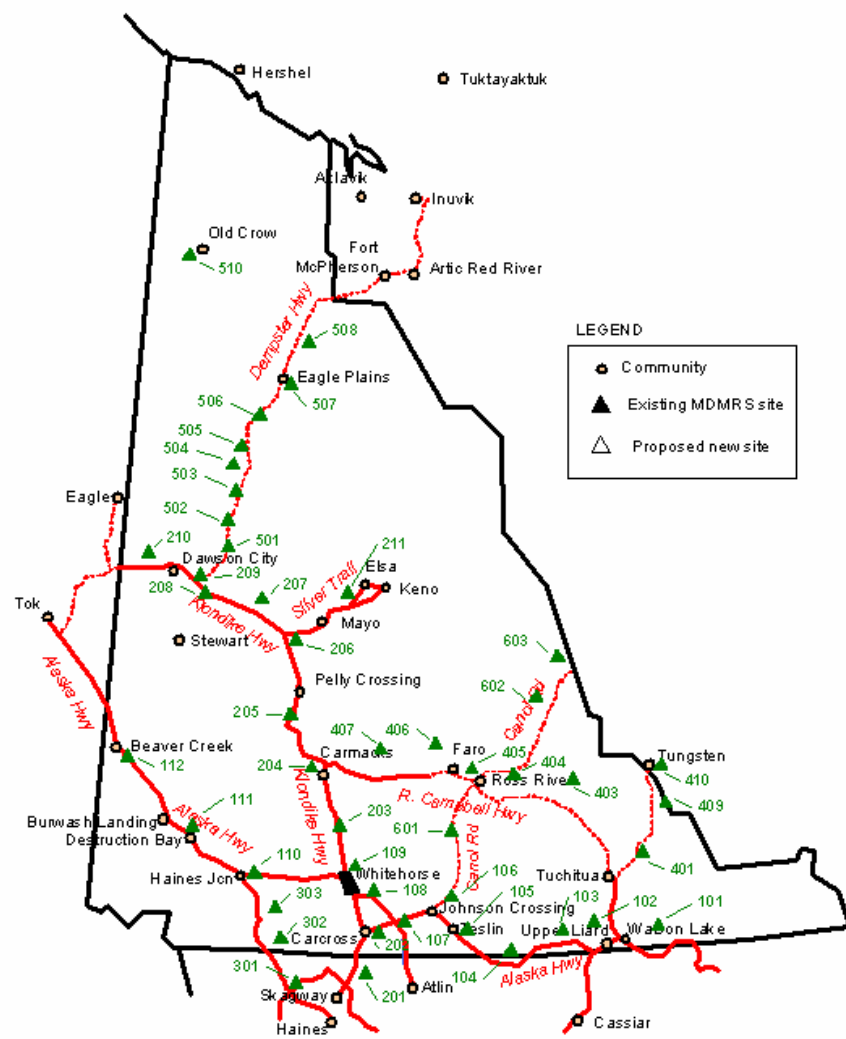- Analysis and results

- Conclusion and future work

Yukon mobile radio system data

- Multi-Departmental Mobile Radio System (M.D.M.R.S.)
- Utilizes trunking system, similar to E-COMM
- Two months of monthly summary data
- Consists of 47 base stations covering major roads
- 23 user groups

Yukon M.D.M.R.S map

Yukon data format

- Data in the first worksheet has the structure:

Month	SiteID	GroupID	Usage	Blocks	Pegs
1	0	B	0.0022	0	1
1	0	J	1.0351	0	153
1	0	S	33.3022	0	4969

- Second worksheet categorizes user groups into 4 categories: not used, public safety voice, non-public safety voice, and data

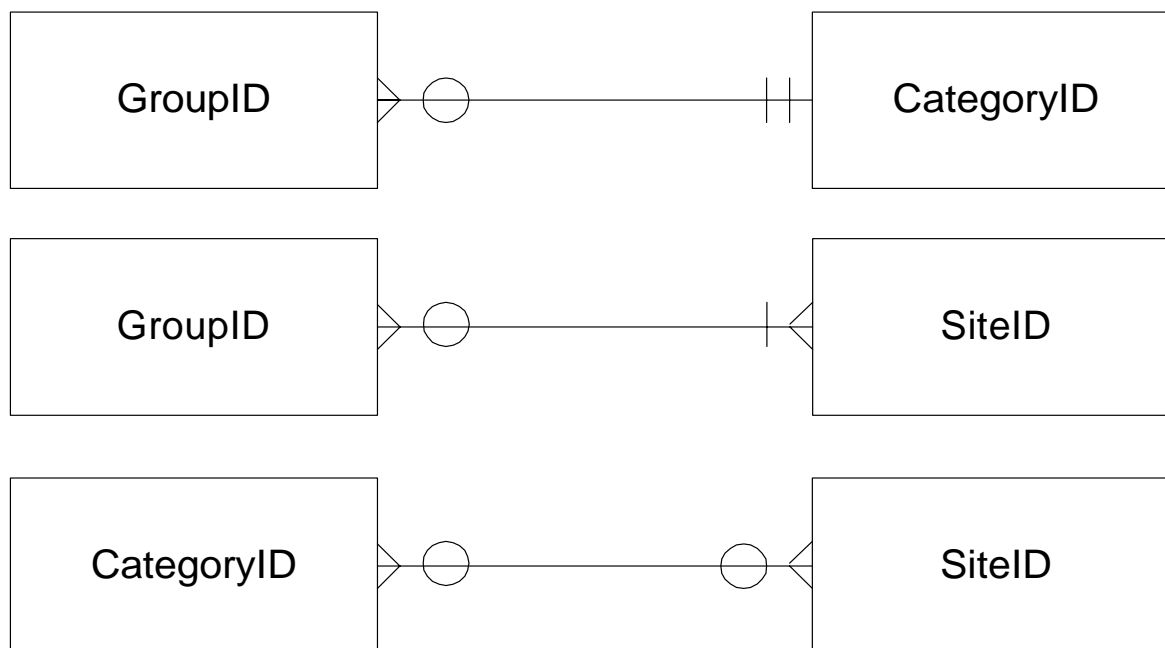Peg: number of calls
Blocks: number of blocked calls

Outline

- Yukon mobile radio system data
- **Traffic data relationships**
- Clustering algorithms
- Analysis and results
- Conclusion and future work

Traffic data relationships

GroupID	CategoryID

GroupID	SiteID

CategoryID	SiteID

Entities: GroupID, CategoryID, SiteID

Analysis of data relationships

- Relationships are used for analysis:
    - user groups' usage at all sites
        - cluster sites usage into groups
    - category's usage at all sites
    - clustering usage among users or categories within each site
- Clustering is also performed within each entity

Outline

- Yukon mobile radio system data
- Traffic data relationships
- **Clustering algorithms**
- Analysis and results
- Conclusion and future work

Clustering algorithms

- Geographical clustering:
  - transform M.D.M.R.S. map into adjacency matrix
- Algorithms:
  - ISODATA: single pass clustering based on likeness
  - mean-shift: adaptive kernel resistant to noise

Geographical clustering: problems

- Adjacency matrix is too sparse

- Adjacency matrix proved to be difficult to construct

- Simple k-means clustering is used:

  - divide data into k groups

  - minimize the sum of MSEs from each group

MSE: mean square error

Outline

- Yukon mobile radio system data
- Traffic data relationships
- Clustering algorithms
- **Analysis and results**
- Conclusion and future work

Analysis implementation

- Data is modified using Excel:
  - GroupIDs and CategoryIDs are changed from letters to numbers
- Modified data is exported to Matlab in a comma delimited form
- Manipulation in Matlab was used to group data according to entity relationship
- Matlab's k-means algorithm is used for clustering: k=3

Analysis implementation

- Average call duration is calculated by dividing usage by pegs:
  - used in Kolmogorov-Smirnov GoF test matching the log-normal distribution
- M.D.M.R.S. map is imported to Matlab for graphical representation:
  - site coordinates are determined manually

GoF: Goodness of Fit

Analysis results

- Two sites have no activity during the two months:
  - siteID 303 and 603
- Public safety and non-public safety user groups have the most traffic:
  - 135.61 hrs and 156.38 hrs in month 1
  - 117.37 hrs and 149.16 hrs in month 2
- Data traffic are 38.11 hrs for month 1 and 45.04 hrs for month 2

Analysis results

- Category called "not used" contains traffic data
- siteID clustering:
    - moderate usage group dominates: 29 sites
    - number of high usage sites: 12 sites
    - number of low usage sites: 11 sites
- siteID 0 (Whitehorse dispatch) consistently has the highest usage over both months: 34 hrs and 28 hrs

Analysis results

- **Month 1**:
    - 20 of 23 groups participated in calls
    - no transmission from groups 7, 14, and 18
- **Month 2**:
    - 21 of 23 groups participated in calls
    - no transmission from groups 7 and 18
- Groups 19 and 21 dominate in usage (115–130 hours)
- Next cluster (group 25) has ~40 hours of activity
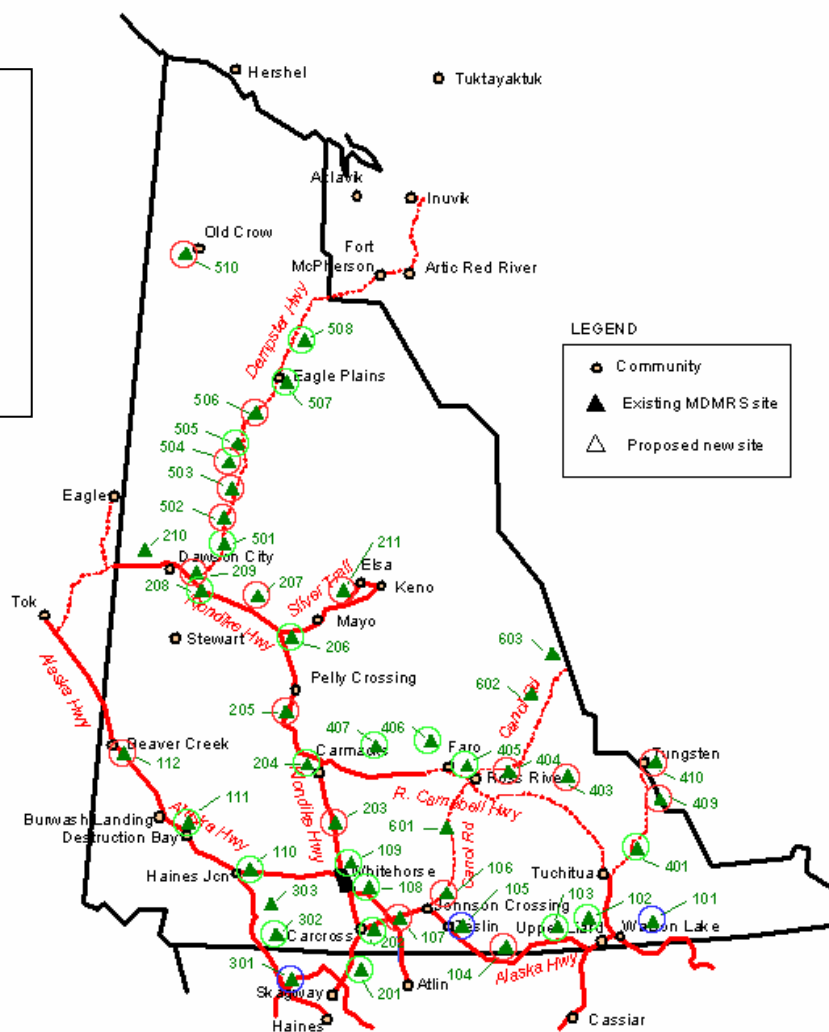- Most users have low system usage, with less than 10 hours of activity

Circles denote the level of activity for GroupID 21 at each site from the k-means algorithm

Red: high activity

Green: medium activity

Blue: low activity

Usage activity map

- User groups cannot be distinguished
- If activity map is used for E-COMM data, it may be possible to identify individual user groups by activity locations:

   - most emergency services operate in specific geographical areas

Kolmogorov-Smirnov GoF test

- Matlab kstest is used to compare average call duration with the normal distribution:
  - hypothesis failed
- Matlab kstest2 is used to compare average call duration with the lognormal distribution:
  - hypothesis fails for:
    - the combined data set
    - individual months: month 1 and month 2
    - p value is in the range $1e^{-47}$ to $1e^{-13}$

Kolmogorov-Smirnov GoF test
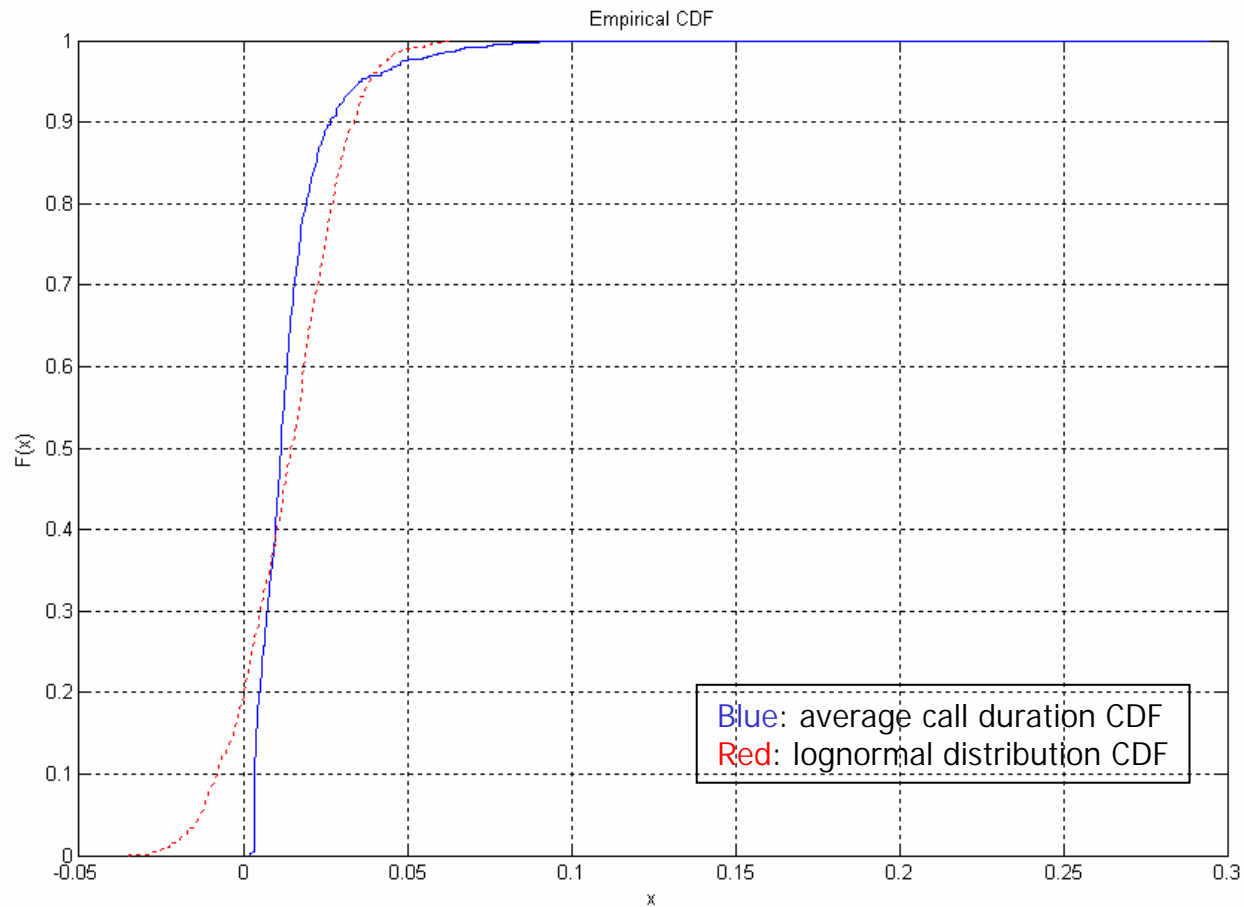
Outline

- Introduction and motivations
- Project goals
- Implementation and design
- Analysis and results
- **Conclusion and future work**

Conclusion

- Analyzed Yukon M.D.M.R.S. data:
  - used k-means clustering method to group usages based on entity relationships
  - applied Kolmogorov-Smirnov goodness-of-fit test on average call duration
  - represented geographical usage
- Future work:
  - analyze additional data from M.D.M.R.S to locate monthly trends
  - apply similar analysis to E-COMM data

References

- Y. Cheng, "Mean shift, mode seeking and clustering," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 17, pp. 790–799, 1995.

- J. R. Jensen, *Introductory digital image processing: A remote sensing perspective*, 2nd. Ed., Upper Saddler River, NJ: Pentice Hall, 1996.

- R. Ramakrishnan, J. Gehrke, *Database Management Systems*, 2nd Ed., Boston, Mass: McGraw-Hill, 2000.

- T. Kanungo, D. M. Mount, N. Netanyahu, C. Piatko, R. Silverman, and A. Y. Wu, "An efficient k-means clustering algorithm: Analysis and implementation," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 24, pp.881–892, 2002.

- N. Cackov, B. Vujicic, S. Vujicic, and Lj. Trajkovic, "Using network activity data to model the utilization of a trunked radio system," (special session) *Proc. SPECTS 2004*, San Jose, CA, July 2004, pp. 517–524.

Questions?