Scalable Ubiquitous Data Access in Clustered WSNs

Yuehhua Lee, Alex Thomo, Kui Wu, Valerie King
University of Victoria, Canada

Outline

- Sensor Network
- Motivation and Challenges
- System Architecture
- Coding Scheme
- Performance

Sensor Networks Overview

- Sensor node equipment: micro-controller, radio transceiver, memory module, sensing devices
- Sensor networks: a set of wirelessly interconnected sensors
- Limited resources

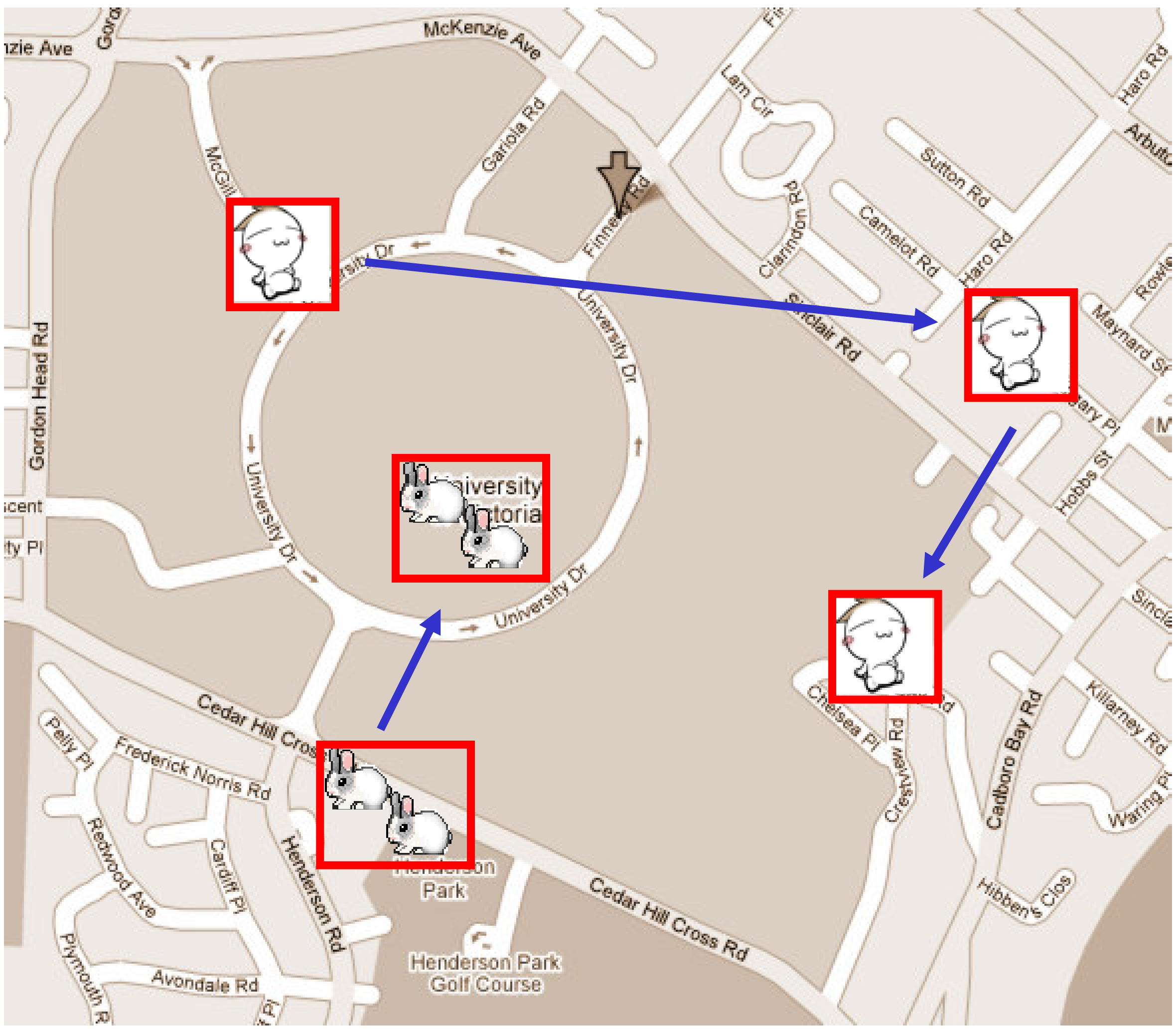
Motivation

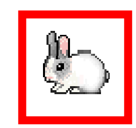
- Monitoring mobile objects
- Network may disconnect;
  sensors may die.
- Limited resources
- Goal 1: Cost-efficient
- Goal 2: Ubiquitous (obtaining any k data items by querying any k nodes with high probability)

System Architecture

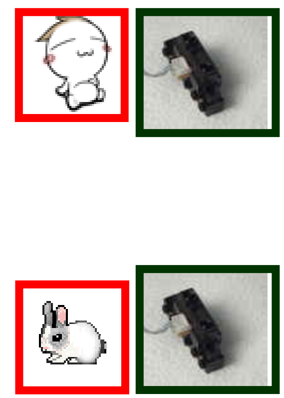
- Network: clustered sensor networks
- Data-generating nodes:
  - move with our objects
  - upload data to nearest storage nodes
- Data-storage nodes:
  - store, forward, encode data
- Data collector:
  - issue queries, collect and decode data

System Architecture

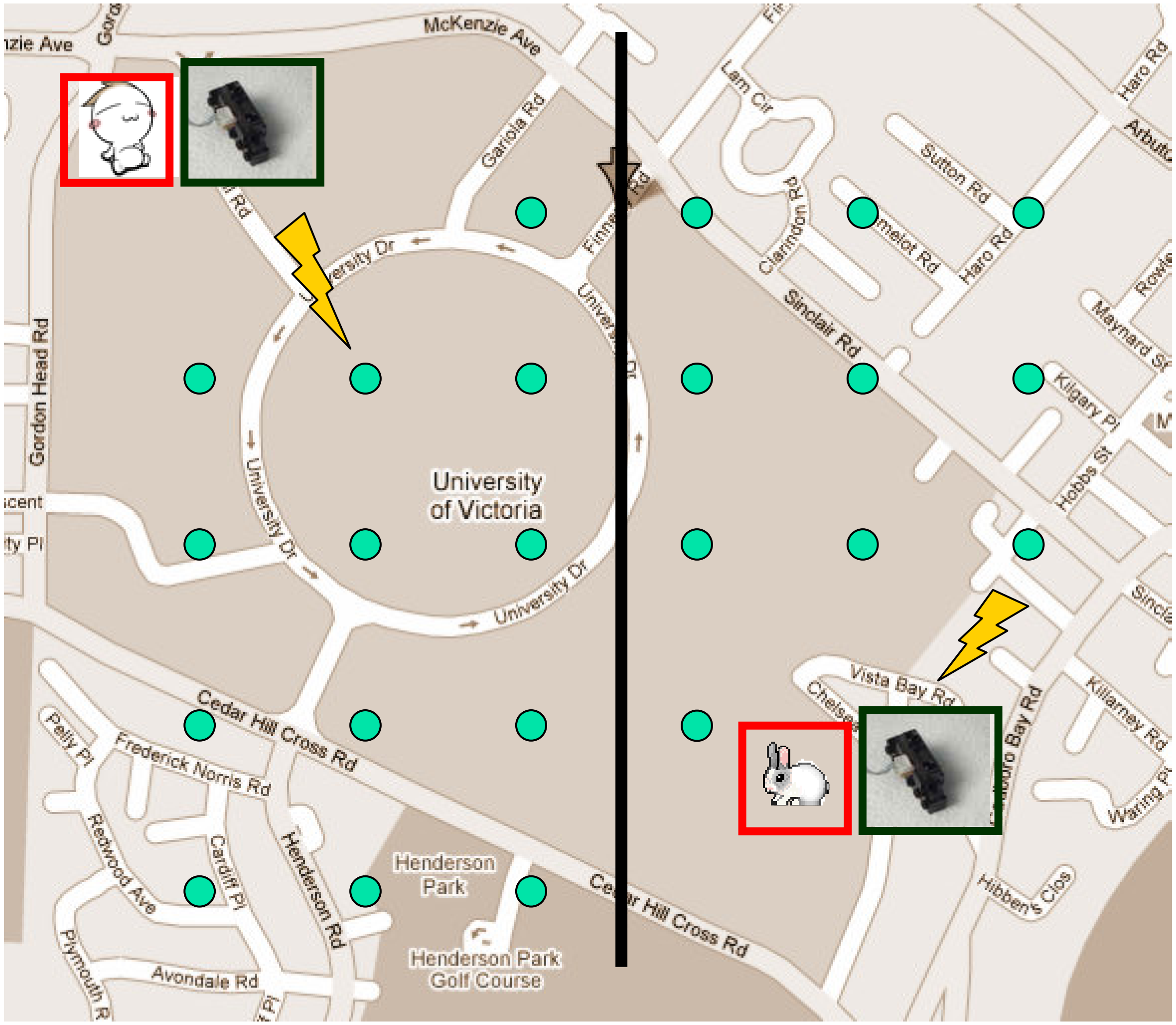




System Architecture



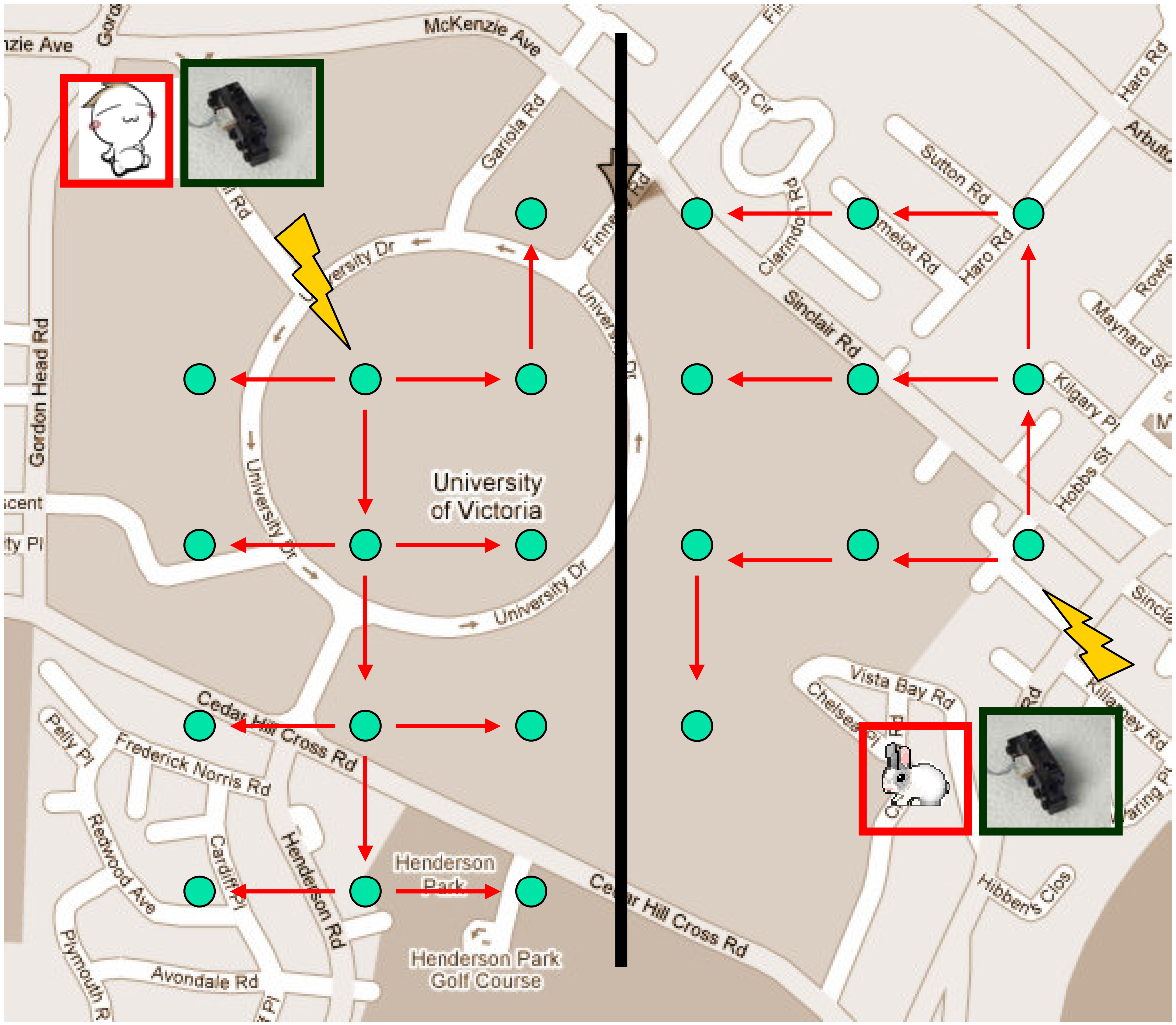
System Architecture



Encoding Scheme

- The storage node forwards the data to all storage nodes in the same cluster.
- According to the sensor ID, each storage node locates the coefficient in code matrix
- Update code checksum

Encoding Scheme



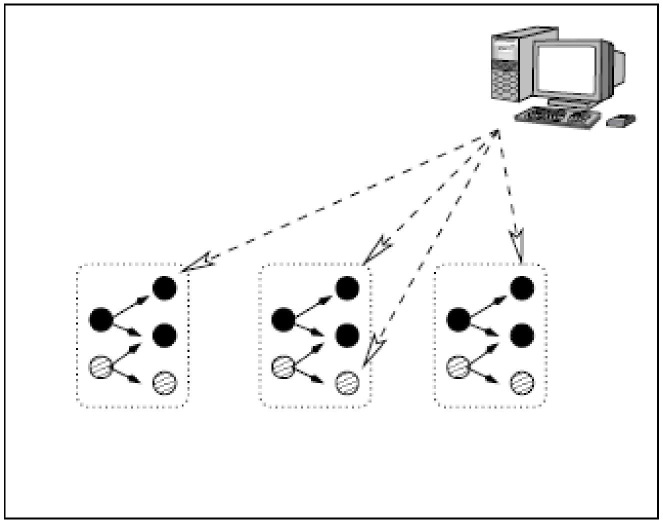
Vandermonde-based Reed-Solomon Code

- Apply a (N_m + K) by K code matrix
- Decode any K_m < K data items from any K_m equations

$$H \cdot \begin{bmatrix} d_1 \\ \vdots \\ d_K \end{bmatrix} = \begin{bmatrix} 1 & 0 & \dots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & 1 \\ 1 & b_1 & \dots & b_1^{K-1} \\ \vdots & \vdots & & \vdots \\ 1 & b_{N_m} & \dots & b_{N_m}^{K-1} \end{bmatrix} \begin{bmatrix} d_1 \\ \vdots \\ d_K \end{bmatrix} = \begin{bmatrix} d_1 \\ \vdots \\ d_K \\ s_1^m \\ \vdots \\ s_{N_m}^m \end{bmatrix}$$

Decoding scheme

- The data collector sample the network.
- Hopefully, in our random sample S, there are more than K_m sensors of cluster m.



Decoding Probability

- The sample space is C (N, |S|)
- We need more than K_m sensors from cluster m.
- K_m+X_m is the number of sensors in the intersection of our random sample and the cluster m.

$$P(S) = \frac{\sum_{X_1} \sum_{X_2} \cdots \sum_{X_M} \binom{N_1}{K_1+X_1}\binom{N_2}{K_2+X_2} \cdots \binom{N_M}{K_M+X_M}}{\binom{N}{|S|}}$$

Decoding Probability

- 3 clusters
- 21 storage nodes (7 each)
- 12 data nodes (4 each)
- Min sample size: 12
- When |S| = 13,

$$P(S) = \frac{\binom{7}{4+1}\binom{7}{4}\binom{7}{4}+\binom{7}{4}\binom{7}{4+1}\binom{7}{4}+\binom{7}{4}\binom{7}{4}\binom{7}{4+1}}{\binom{21}{13}}$$

Bounding the decoding probability

Theorem 1:
P(S) is no less than $1-\sum_{m=1}^{M}\exp[\frac{-(|S|p_m-K_m)^2}{2|S|}]$

M : number of clusters

$|S|$: size of sample

K_m: number of data nodes in cluster m

N_m: number of storage nodes in cluster m

P_m : K_m/N_m

Bounding the decoding probability

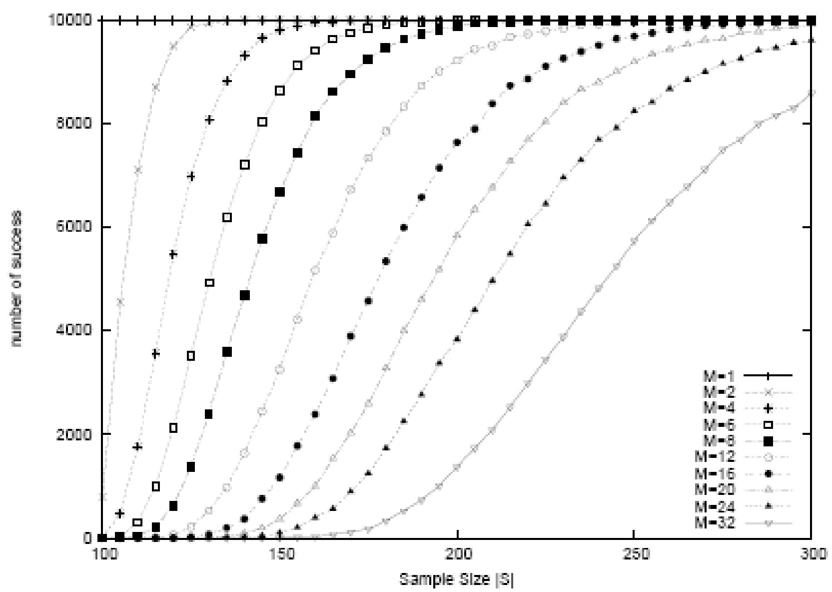
- *Theorem2*:
  We can choose |S| to be

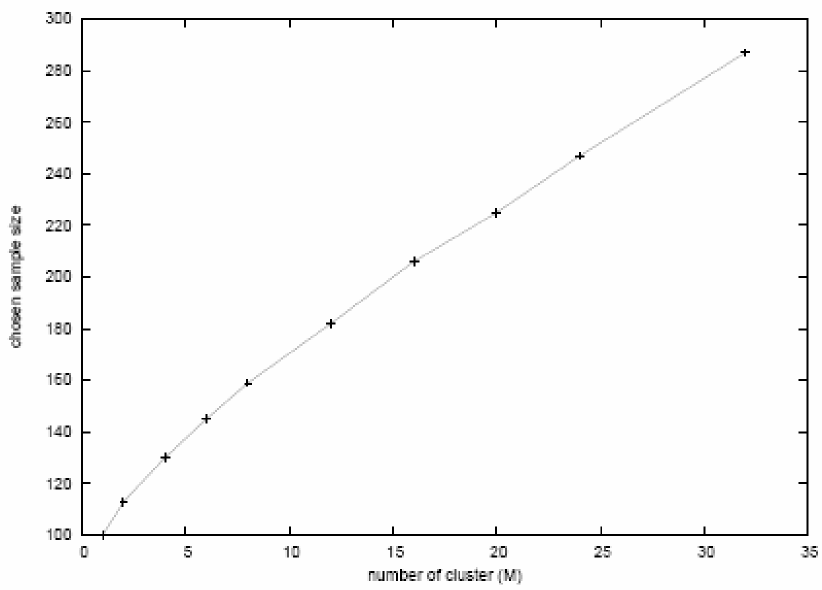
$$\frac{2(\hat{p}\hat{K} + \ln\frac{M}{\varepsilon}) + \sqrt{4(\hat{p}\hat{K} + \ln\frac{M}{\varepsilon})^2 - 4\hat{p}^2\hat{K}^2}}{2\hat{p}^2}$$

  in order to achieve at least 1- ε decoding probability

System performance

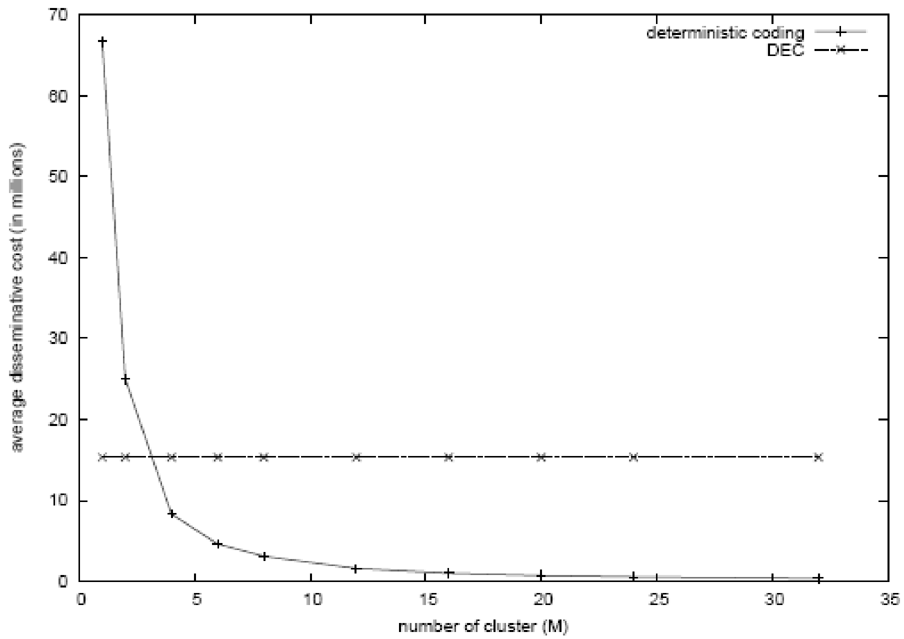
- M decreases the decoding probability





System performance

- M decreases the communication cost



- M>$(K/5\ln K)^{2/3}$

Thank you

Questions ?