Computing and Communicating Functions over Sensor Networks

Solmaz Torabi

Dept. of Electrical and Computer Engineering
Drexel University
solmaz.t@drexel.edu

Advisor: Dr. John M. Walsh

Refrences

A. Giridhar and P. R. Kumar, "Computing and communicating functions over sensor networks," *Selected Areas in Communications, IEEE Journal on*, vol. 23, no. 4, pp. 755–764, 2005.

[1]

Outline

- Model of the problem
- Network topologies
  - Random planar network
  - Collocated network
- Classes of functions
  - Type threshold function
  - Type sensitive function
- Results

Summary

Order of difficulty of computations

- $\Theta(\frac{1}{n})$
  - Average, Mode, Type vector in Collocated network:
  - Data Downloading
- $\Theta(\frac{1}{\log n})$
  - Average, Mode, Type in Random Multi-hop network
  - Max, Min in Collocated network
- $\Theta(\frac{1}{\log \log n})$
  - Max, Min in Random Multi-hop network

Problem model

- n sensor nodes
- ρ_{ij} is the distance between two nodes i, j.
- Fusion node needs to calculate $f_n(x_1, ..., x_n)$ exactly.
- At time t, sensor i takes a measurement $x_i(t) \in \{1, .., |\mathcal{X}|\}$
- No probability distribution on $x_i(t)$
- Non-information theoretic formulation
- They adopt packet based collision model of wireless communication

Problem model

- All nodes share a common transmission range r.
- Receiver should be outside other transmitters' interference footprints
- Node i can succesuty transmit packet to j if
  - $\rho_{ij} \leq r$
  - and for every other simultaneously transmitting node k, $\rho_{kj} \geq (1+\Delta)r$
- Successful communication between two nodes takes place at a rate W bits/second.

Problem model

- Block coding allowed:
- M measurements of node 1: $x_1 \rightarrow x_1(1), x_1(2), ..., x_1(M)$
- M measurements of node 2: $x_2 \rightarrow x_2(1), x_2(2), ..., x_2(M)$
- ...
- M measurements of node n: $x_n \rightarrow x_n(1), x_n(2), ..., x_n(M)$
- M functions to compute: $f\left(x_1(1), .., x_n(1)\right), ..., f\left(x_1(M), ..., x_n(M)\right)$
- The matrix of measurements $\bar{x} \in \mathcal{X}^{Mn}$ is a $n \times M$ matrix.
- If all M functions are computed in time T, the computational rate is $R = \frac{M}{T}$

List of notations

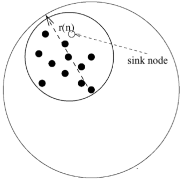
- $f_n : \mathcal{X}^n \to \mathcal{Y}_n$ is the function of interest.
- $\mathcal{R}(f, n)$ is the range of f_n
- $\mathcal{S}_{M,n}$ is a scheme or strategy
- $T(\mathcal{S}_{M,n})$ is the time taken by scheme $\mathcal{S}_{M,n}$ worst case over all $X \in \mathcal{X}^{nM}$
- $R(\mathcal{S}_{M,n}) = \frac{M}{T(\mathcal{S}_{M,n})}$ is the rate of the scheme.
- $R_{max}^{(n)}$ is the supremum of rates $R(\mathcal{S}_{M,n})$ over all schemes $\mathcal{S}_{M,n}$ and block length $M \Rightarrow R_{max} = \sup_{S,M} \frac{M}{T(\mathcal{S}_{M,n})}$
- $G^{(n)}$ is the spatial graph consists of the set of n nodes, with edges between nodes that are within a distance r of each other.

Outline

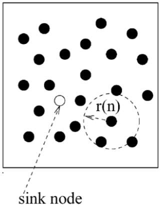
- Model of the problem
- Network topologies
  - Random planar network
  - Collocated network
- Classes of functions
  - Type threshold function
  - Type sensitive function
- Results

Network Topologies

- **Collocated Networks**: These are networks with $\rho_{ij} \leq r$ for all i, j so every transmission is heard by all nodes.



- **Random Planar Networks**: The n nodes along with the sink node s are uniformly and independently distributed on a unit square



  - Note: The common range r of all the n nodes is so chosen that, by using multihop communication, the graph is connected

Review

Lemma:

For random planar networks, if range $r(n) = \sqrt{\frac{2 \log n}{n}}$ then $G^{(n)}$ is connected w.h.p. and $d(G(n)) \leq c \log n$ w.h.p.

Result follows from earlier paper on critical power for asymptotic connectivity.

A trivial upper bound

- The sink node can receive at most W bits/s
- Representing $f(.)$ requires $\log|\mathcal{R}(f,n)|$

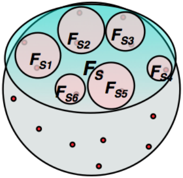
$$R_{max}^{(n)} \leq \frac{W}{\log|\mathcal{R}(f,n)|} \tag{1}$$

Outline

- Model of the problem
- Network topologies
  - Random planar network
  - Collocated network
- Classes of functions
  - Type threshold function
  - Type sensitive function
- Results

Types of functions

Divisible functions



Definition

- given a subset $S = \{i_1, ..., i_k\} \subset [n]$, denote by $x_S = [x_{i_1}, ..., x_{i_k}]$
- A function $f : \mathcal{X}^n \to \mathcal{Y}_n$ is divisible if
  - $|\mathcal{R}(f, n)|$ is nondecreasing in n
  - given any partition $\Pi(S) = \{S_1, .., S_j\}$ of $S \subset [n]$, $\exists$ a function $g^{\Pi(S)}$

$$f(\underline{x}_S) = g^{\Pi(S)}(f(\underline{x}_{S_1}), f(\underline{x}_{S_2}), ..., f(\underline{x}_{S_k})) \tag{2}$$

- **Example** $\max(1, 2, 3, 4, 5) = \max(\max(1, 2), \max(3, 4), \max(5))$

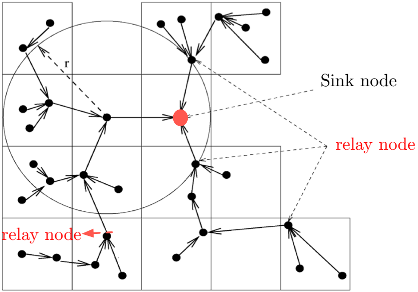
Computing divisible function over random planar network

Theorem

For a divisible function f, and $d(G^{(n)}) = \mathcal{O}(\log|\mathcal{R}(f,n)|)$,
$R_{max}(n) = \Theta(\frac{1}{\log|\mathcal{R}(f,n)|})$

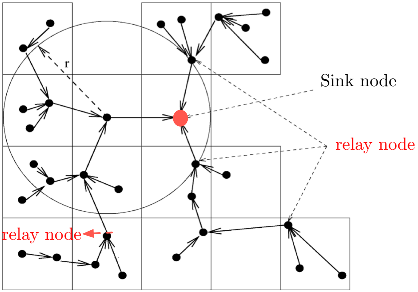
Proof

- Tessellation of plane into square cells of side $\frac{r}{\sqrt{2}}$
- Cell Graph: Define on a set of non-empty cells as vertices
- Two cells are adjacent if there are two nodes within each cell which are adjacent in G^n



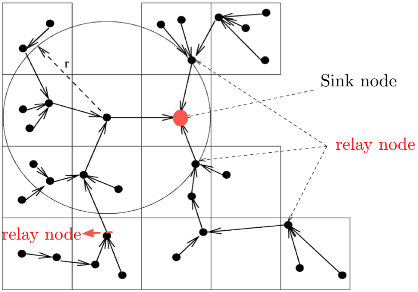
Proof

- Neighboring occupied cells can communicate with each other
- Consider a rooted spanning tree of the cell graph
- Designate the cell with the sink node as the root, s
- Choose a relay node (u) in each cell and a parent (v) in the next cell towards the root
- Each cell has one relay node (picked out of possibly multiple choices)



Proof

- For each node u, define the descendant set D_u as
  - If u is a relay node of a c, D_u is the set of all nodes that either belong to c or to descendants of c
  - If u is the relay parent of $\{u_1, \dots, u_l\}$, $D_u = \{u\} \cup D_{u_1} \cup \dots \cup D_{u_l}$.
  - Else $D_u = u$
- Locally compute and pass on along tree to root
  - Collect data from $deg(G^n)$ nodes within cell



- Collect functional value of $\log|\mathcal{R}(f_n)|$ bits from child cells
- Pass on functional value of $\log|\mathcal{R}(f_n)|$ bits to parent cell

Computing a divisible function

Special cases

- Data Downloading:

$$\log|\mathcal{R}(f,n)| = \log|\mathcal{X}^n| = \mathcal{O}(n) \Rightarrow R_{max}(n) = \Theta(\frac{1}{n}) \quad (3)$$

if $d(G^n) = \mathcal{O}(n)$ (For any connected graph)

- Frequency histogram or the type-vector

$$\underline{\tau}(\underline{x}) = [\tau_1(\underline{x}), \tau_2(\underline{x}), ..., \tau_{|\mathcal{X}|}(\underline{x})] \quad (4)$$

where

$$\tau_i(\underline{x}) = |\{j : x_j = i\}| \quad (5)$$

is the number of occurrences of i in $\underline{x}$.

Histogram

The number of type vectors of a vector of size n is the number of nonnegative integer solutions to the equation

$$y_1 + y_2 + \ldots + y_{|\mathcal{X}|} = n \tag{6}$$

$$|\mathcal{R}(f, n)| = \binom{n + |\mathcal{X}| - 1}{|\mathcal{X}| - 1} \tag{7}$$

and

$$\left(\frac{n}{|\mathcal{X}|}\right)^{|\mathcal{X}|} \leq \binom{n + |\mathcal{X}| - 1}{|\mathcal{X}| - 1} \leq (n+1)^{|\mathcal{X}|} \tag{8}$$

$$R_{max}(n) = \Theta\left(\frac{1}{\log n}\right) \text{ if } d(G^n) = \mathcal{O}(\log n) \tag{9}$$

if $r(n)$ is chosen properly $\Rightarrow$ $d(G^n) = \mathcal{O}(\log n)$ w.h.p

Symmetric Function

Definition

Functions which are invariant with respect to permutations of their arguments.

$$f(x_1, ..., x_n) = f(\sigma(x_1, ..., x_n)) \quad \text{for all permutations} \quad \sigma \tag{10}$$

Note

Symmetric functions depend only on type vector $\tau = (\tau_1, \tau_2, ..., \tau_{|\mathcal{X}|})$

$$f_n(x_1, ..., x_n) = \bar{f}(\tau_1, \tau_2, ..., \tau_{|\mathcal{X}|}) \tag{11}$$

Note

The data generated by a sensor is of primary importance, rather than its identity.

Computing a Symmetric Function

- An obvious strategy to compute any symmetric function is to simply communicate the entire type or frequency-histogram

$$R_{max} = \Omega(\frac{1}{\log n}) \tag{12}$$

- Is it possible to do better?
- Two disjoint subclasses
  - Type-Sensitive Functions
  - Type-Threshold Functions

Type-Sensitive Functions

A symmetric function is type-sensitive if $\exists\ 0 < \gamma < 1$, and $k \in Z^+$, and any $j \leq n - [\gamma n]$, given any subset $\{x_1, ..., x_j\}$, there are two subsets of values $\{y_{j+1}, ..., y_n\}$ and $\{z_{j+1}, .., z_n\}$ such that

$$f(x_1, .., x_j, z_{j+1}, .., z_n) \neq f(x_1, ..., x_j, y_{j+1}, .., y_n) \tag{13}$$

- There is a γ such that a fraction γ of values is never enough to pin down the value of the function f_n
- The value of a type-sensitive function cannot be determined if a large enough fraction of the arguments are un-known.

Examples

- Mode : If more than half the x_i's are unknown, the mode is undetermined
- Mean, Median, Majority

Type-Threshold Functions

A symmetric function f is said to be type-threshold if $\exists$ θ-vector called the threshold vector, such that

$$f(\underline{x}) = \bar{f}(\underline{\tau}(\underline{x})) = \bar{f}(\min(\underline{\tau}(\underline{x}), \underline{\theta})) \text{ for all } x \in \mathcal{X}^n \tag{14}$$

- Only want to know whether each τ_i exceeds a threshold θ_i

Examples

- Max, Min: $\theta = [1, 1, ..., 1]$
- k^{th} largest value: $\theta = [1, 1, ..., 1]$
- Mean of the k largest values:
- Indicator function $I\{x_i = k \text{ for some } i\}$, with $\theta = [0, 0, ..., 1, ..., 0, 0]$

Collision-free strategies (CFS) in collocated Networks

- Without loss of generality, suppose that time is slotted, and one bit is transmitted in each slot. $W = 1$

Collision-free strategy

- is a strategy which is required to explicitly avoid collisions. consists of

$$\phi_m : \{0,1\}^{m-1} \to \{1, ..., n\} \tag{15}$$

$$\psi_m : \mathcal{X}^N \times \{0,1\}^{m-1} \to \{0,1\} \tag{16}$$

- Node ϕ_1 transmit packet $\psi_1(x_{\phi_1})$ at time 1.
- Node $\phi_2(\psi_1(x_{\phi_1}))$ transmit packet $\psi_2(\psi_1(x_{\phi_1}), x_{\phi_2})$ at time 2.
- ...

Collision-free strategies (CFS) in collocated Networks

- The node designated to transmit at time m is fixed by the value $\phi_m(\psi_{m-1}, \psi_{m-2}, ..., \psi_1)$, which can be computed by all the nodes.
- The identity of the transmitting node is automatically known to all.
- The medium access problem is resolved in a distributed but collision-free fashion.
- The strategies described above are required to explicitly avoid collisions

Type-Sensitive Functions in Collocated Networks

Theorem

The maximum rate for computing a type-sensitive function in a collocated network, using any CFS is $\Theta(\frac{1}{n})$, which is of the same order as communicating the entire data.

Type-Sensitive Functions in Collocated Networks

Proof

- Wlog suppose $|\mathcal{X}| = 2$
- initially x_{g1} is in the set S_{g1}^0 with the cardinality $|S_{g1}^0| = 2^M$
- After first transmission,x_{g1} can be in one of two sets depending on whether it transmits 0 or 1
  - Let the transmission correspond to the larger set, call it be S_{g1}^1
  - $|S_{g1}^1| \geq 1/2|S_{g1}^0|$
- After t-th transmission of node k, let x_k lie in S_k^t with $|S_k^t| \geq 1/2|S_k^{t-1}|$

Type-Sensitive Functions in Collocated Networks

- So at the end, uncertainty set is: $|S_1 \times S_2 \times ... \times S_n| \geq 2^{nM-T}$
- Thus at least $nM - T$ places in the nM values $(x_1, x_2, ..., x_n)$ are undetermined
- However to compute $f_n(x(1), x(2), ..., x(M))$, at least cnM values are needed
- So $nM - T \leq (1-c)nM$
- So $T \geq cnM$
- Hence $R = M/T \leq \mathcal{O}(\frac{1}{cn})$
- Thus $R_{max}(n) = \mathcal{O}(\frac{1}{n})$ for collocated case

Type-Threshold Functions in Collocated Networks

Theorem

The maximum rate for computing a nonconstant type-threshold function in a collocated network, using any CFS is $\Theta(\frac{1}{\log n})$

Proof

- First prove the result for the case $|\mathcal{X}| = 2$, and the max function $f(x_1, x_2, ..., x_n) = \max\{x_i : 1 \leq i \leq n\}$
- $\theta = [1, 1, ..., 1]$

Proof

Achievability

Goal : providing a sequence of CFS's $\mathcal{S}_{M,n}$, asymptotically achieving the rate $\Omega(\frac{1}{\log n})$

- Take block length $M = \ell n > n$
- Let the number of 1's in the vector $\underline{X}_i$ be M_i
- Define $S_i = \{1 \leq j \leq M : X_i(j) = 1, X_k(j) = 0, \text{for all } k < i\}$
- Define $\bar{M}_i = |S_i|$, for each $1 \leq i \leq n$
- Note : S_i's are disjoint.
- Note : $\sum_i \bar{M}_i \leq M$
- Note : Communicating the sets S_1, S_2,..,S_n to the sink suffices to reconstruct the function.

Proof

A collision free strategy $\mathcal{S}_{M,n}$

- Node i compute $\bar{M}_i$, and communicate its value in $\log M$ slots.
- Now S_i is one of the $\binom{M-\sum_{j<i}\bar{M}_j}{\bar{M}_i}$
- Node i communicate the identity of the set S_i in $\log \binom{M-\sum_{j<i}\bar{M}_j}{\bar{M}_i}$

$$T(\mathcal{S}_{M,n}) = n \log M + \sum_i \log \binom{M - \sum_{j<i} \bar{M}_j}{\bar{M}_i} \tag{17}$$

By bounding RHS

$$T(\mathcal{S}_{M,n}) < n \log \ell(n+1) + \ell(n+1) \log(n+1)e \tag{18}$$

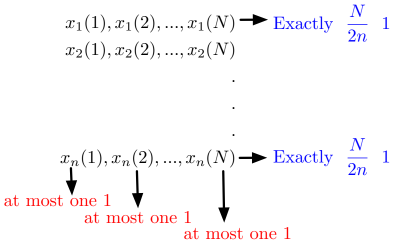
$$R = \frac{M}{T} = \Omega(\frac{1}{\log n}) \tag{19}$$

Proof

Upper bound

Goal: $R_{max}(n) = \mathcal{O}(\frac{1}{\log n})$

- Take $M > 2n$. Consider a set of measurement matrices:



- Claim: For each set of transmissions $P_1, P_2, ..., P_T$, there is unique such x in the above set that produces it.

- Suppose not. Then there are two: x and y which produce same transmissions. They differ in some $x_k \neq y_k$
- Then also produces same transmissions since node k hears the same under x_k or y_k and so reacts the same.
- But this has different "Max" values from x
- Thus "Max" functions are not determined from transmissions
- Number of such vectors $x = \prod_i \binom{M-(i-1)\frac{M}{2n}}{\frac{M}{2n}} > (n-1)^M$
- So $2^T > (n-1)^M$
- So $T > M \log(n-1)$
- So $R = \frac{M}{T} \leq \frac{1}{\log(n-1)} \Rightarrow R^n_{max} = \mathcal{O}(\frac{1}{\log n}) \Rightarrow R^n_{max} = \Theta(\frac{1}{\log n})$

Summary

Order of difficulty of computations

- $\Theta(\frac{1}{n})$
  - Average, Mode, Type vector in Collocated network:
  - Data Downloading
- $\Theta(\frac{1}{\log n})$
  - Average, Mode, Type in Random Multi-hop network
  - Max, Min in Collocated network
- $\Theta(\frac{1}{\log \log n})$
  - Max, Min in Random Multi-hop network

Thank You!

Question?