Enhancing Energy Efficiency among Communication, Computation and Caching with QoI-Guarantee

Faheem Zafari[1], Jian Li[2], Kin K. Leung[1],
Don Towsley[2], Ananthram Swami[3]

[1]Imperial College London
[2]University of Massachusetts Amherst
[3]US Army Research Laboratory

DAIS-ITA Annual Fall Meeting 2017

Motivation

- Energy is one of the fundamental limitations of sensors

Motivation

- Energy is one of the fundamental limitations of sensors
- Sensors are responsible for
  - **Communication**
  - **Computation**
  - **Caching**

Motivation

- Energy is one of the fundamental limitations of sensors
- Sensors are responsible for
  - **Communication**
  - **Computation**
  - **Caching**
- All consume energy
  - Computation reduces communication cost
  - Caching reduces communication cost

The Big Question?

- Is *computation* the solution for energy consumption problem?
  - **No**. Computation reduces communication cost but also incurs energy cost[1]
  - Trade-off between communication and computation costs

[1] Barr, Kenneth C., and Krste Asanovi. "Energy-aware lossless data compression." ACM Transactions on Computer Systems (TOCS) 24.3 (2006): 250-291.

The Big Question?

- Is *computation* the solution for energy consumption problem?
  - **No**. Computation reduces communication cost but also incurs energy cost[1]
  - Trade-off between communication and computation costs
- Caching the data also incurs cost
  - Should the data be cached?
  - Where should the data be cached?

[1]Barr, Kenneth C., and Krste Asanovi. "Energy-aware lossless data compression." ACM Transactions on Computer Systems (TOCS) 24.3 (2006): 250-291.

The Big Question?

Decision to make

- How much to compress?

The Big Question?

Decision to make

- How much to compress?
- Where to cache the data?

System Model

- A directed graph $G = (V, E)$

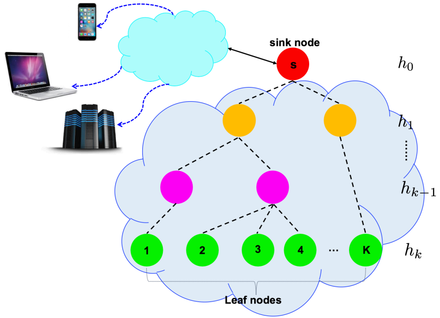

Figure: Tree-Structured Network

System Model

- A directed graph $G = (V, E)$
- $\mathcal{K} \subseteq V$: the set of leaf nodes with $|\mathcal{K}| = K$

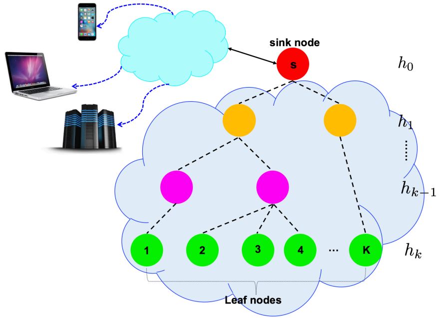

Figure: Tree-Structured Network

System Model

- A directed graph $G = (V, E)$
- $\mathcal{K} \subseteq V$: the set of leaf nodes with $|\mathcal{K}| = K$
- Only leaf nodes $k \in \mathcal{K}$ can generate data

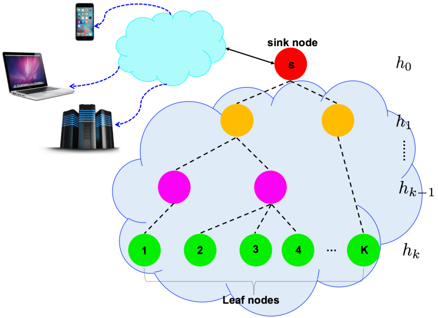

Figure: Tree-Structured Network

System Model

- A directed graph $G = (V, E)$
- $\mathcal{K} \subseteq V$: the set of leaf nodes with $|\mathcal{K}| = K$
- Only leaf nodes $k \in \mathcal{K}$ can generate data
- y_k : the amount of data generated by $k \in \mathcal{K}$

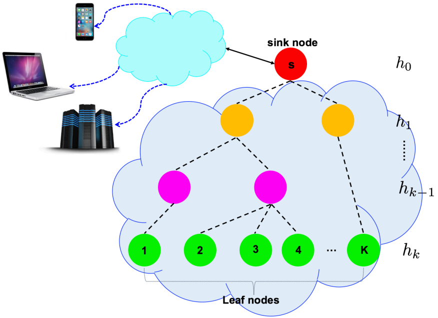

Figure: Tree-Structured Network

System Model

- Data are transmitted over a given path to sink node s

System Model

- Data are transmitted over a given path to sink node s
- Denote $\mathcal{H}^k$ as the path from node k to s, where $\mathcal{H}^k = \{h_0^k, h_1^k, \cdots, h_{h(k)}^k\}$, with $h_j^k \in V$, $(h_j^k, h_{j+1}^k) \in E$, $h_0^k \triangleq s$ and $h_{h(k)}^k \triangleq k$

System Model

- Data are transmitted over a given path to sink node s
- Denote $\mathcal{H}^k$ as the path from node k to s, where $\mathcal{H}^k = \{h_0^k, h_1^k, \cdots, h_{h(k)}^k\}$, with $h_j^k \in V$, $(h_j^k, h_{j+1}^k) \in E$, $h_0^k \triangleq s$ and $h_{h(k)}^k \triangleq k$
- Per-bit transmission cost ε_{vT}

System Model

- Data are transmitted over a given path to sink node s
- Denote $\mathcal{H}^k$ as the path from node k to s, where $\mathcal{H}^k = \{h_0^k, h_1^k, \cdots, h_{h(k)}^k\}$, with $h_j^k \in V$, $(h_j^k, h_{j+1}^k) \in E$, $h_0^k \triangleq s$ and $h_{h(k)}^k \triangleq k$
- Per-bit transmission cost ε_{vT}
- Per-bit reception cost ε_{vR}

System Model

- Data are transmitted over a given path to sink node s
- Denote $\mathcal{H}^k$ as the path from node k to s, where $\mathcal{H}^k = \{h_0^k, h_1^k, \cdots, h_{h(k)}^k\}$, with $h_j^k \in V$, $(h_j^k, h_{j+1}^k) \in E$, $h_0^k \triangleq s$ and $h_{h(k)}^k \triangleq k$
- Per-bit transmission cost ε_{vT}
- Per-bit reception cost ε_{vR}
- Per-bit computation cost ε_{vC}

System Model

- Data are transmitted over a given path to sink node s
- Denote $\mathcal{H}^k$ as the path from node k to s, where $\mathcal{H}^k = \{h_0^k, h_1^k, \cdots, h_{h(k)}^k\}$, with $h_j^k \in V$, $(h_j^k, h_{j+1}^k) \in E$, $h_0^k \triangleq s$ and $h_{h(k)}^k \triangleq k$
- Per-bit transmission cost ε_{vT}
- Per-bit reception cost ε_{vR}
- Per-bit computation cost ε_{vC}
- S_v : the storage capacity at node $v \in V$

System Model

- Data are transmitted over a given path to sink node s
- Denote $\mathcal{H}^k$ as the path from node k to s, where $\mathcal{H}^k = \{h_0^k, h_1^k, \cdots, h_{h(k)}^k\}$, with $h_j^k \in V$, $(h_j^k, h_{j+1}^k) \in E$, $h_0^k \triangleq s$ and $h_{h(k)}^k \triangleq k$
- Per-bit transmission cost ε_{vT}
- Per-bit reception cost ε_{vR}
- Per-bit computation cost ε_{vC}
- S_v : the storage capacity at node $v \in V$
- Node h_i^k along path $\mathcal{H}^k$ can compress the data generated by leaf node k with a *data reduction rate* $\delta_{k,i}$, where $0 < \delta_{k,i} \leq 1$, $\forall i, k$

System Model

- E_v : the total energy consumption at node v

$$E_v = E_{vR} + E_{vT} + E_{vC} + E_{vS}, \tag{1}$$

  - $E_{vR} = y_v \varepsilon_{vR}$ is the reception cost
  - $E_{vT} = y_v \varepsilon_{vT} \delta_v$ is the transmission cost
  - $E_{vC} = y_v \varepsilon_{vC} l_v(\delta_v)$ is the computation cost
  - $E_{vS} = w_{ca} y_v T$ is the storage cost
  - $l_v(\delta_v)$:a decreasing differentiable function of the reduction rate, e.g., $l_v(\delta_v) = \frac{1}{\delta_v} - 1$ [2]

[2] Eswaran, Sharanya, et al. "Adaptive in-network processing for bandwidth and energy constrained mission-oriented multihop wireless networks." IEEE Transactions on Mobile Computing 11.9 (2012): 1484-1498.

System Model

- E_v : the total energy consumption at node v

$$E_v = E_{vR} + E_{vT} + E_{vC} + E_{vS}, \tag{1}$$

  - $E_{vR} = y_v \varepsilon_{vR}$ is the reception cost
  - $E_{vT} = y_v \varepsilon_{vT} \delta_v$ is the transmission cost
  - $E_{vC} = y_v \varepsilon_{vC} l_v(\delta_v)$ is the computation cost
  - $E_{vS} = w_{ca} y_v T$ is the storage cost
  - $l_v(\delta_v)$:a decreasing differentiable function of the reduction rate, e.g., $l_v(\delta_v) = \frac{1}{\delta_v} - 1$[2]
- During a time period of T, R_k requests for the data y_k generated by leaf node k

[2] Eswaran, Sharanya, et al. "Adaptive in-network processing for bandwidth and energy constrained mission-oriented multihop wireless networks." IEEE Transactions on Mobile Computing 11.9 (2012): 1484-1498.

System Model

- $$b_{k,i} = \begin{cases} 1, \text{data from node } k \text{ is stored along path } \mathcal{H}^k \text{ at node } h_i^k, \\ 0, \text{otherwise.} \end{cases}$$

System Model

- $$b_{k,i} = \begin{cases} 1, \text{data from node } k \text{ is stored along path } \mathcal{H}^k \text{ at node } h_i^k, \\ 0, \text{otherwise.} \end{cases}$$
- The network is structured, i.e. sink node s knows which unique path to follow for a specific request

System Model

- $$b_{k,i} = \begin{cases} 1, \text{data from node } k \text{ is stored along path } \mathcal{H}^k \text{ at node } h_i^k, \\ 0, \text{otherwise.} \end{cases}$$
- The network is structured, i.e. sink node s knows which unique path to follow for a specific request
- C_v : the set of leaf nodes $k \in \mathcal{K}$ that are descendant of node v

System Model

- $$b_{k,i} = \begin{cases} 1, \text{data from node } k \text{ is stored along path } \mathcal{H}^k \text{ at node } h_i^k, \\ 0, \text{otherwise.} \end{cases}$$
- The network is structured, i.e. sink node s knows which unique path to follow for a specific request
- C_v : the set of leaf nodes $k \in \mathcal{K}$ that are descendant of node v
- $f(\delta_v) = \varepsilon_{vR} + \varepsilon_{vT}\delta_{k,i} + \varepsilon_{kC}(\frac{1}{\delta_{k,i}} - 1)$

System Model

- $$b_{k,i} = \begin{cases} 1, \text{data from node } k \text{ is stored along path } \mathcal{H}^k \text{ at node } h_i^k, \\ 0, \text{otherwise.} \end{cases}$$
- The network is structured, i.e. sink node s knows which unique path to follow for a specific request
- C_v : the set of leaf nodes $k \in \mathcal{K}$ that are descendant of node v
- $f(\delta_v) = \varepsilon_{vR} + \varepsilon_{vT}\delta_{k,i} + \varepsilon_{kC}(\frac{1}{\delta_{k,i}} - 1)$
- For convenience, let $f_{k,h(k)} \triangleq f_k$ and $\delta_{k,h(k)} \triangleq \delta_k$

Energy Efficiency Optimization

- E_k^C: energy for data received, transmitted, and possibly compressed by all nodes on the path from leaf node k to sink node s

$$E_k^C = \sum_{i=0}^{h(k)} y_k f(\delta_{k,i}) \prod_{m=i+1}^{h(k)} \delta_{k,m} \qquad (2)$$

Energy Efficiency Optimization

- E_k^{C}: energy for data received, transmitted, and possibly compressed by all nodes on the path from leaf node k to sink node s

$$E_k^{\mathrm{C}} = \sum_{i=0}^{h(k)} y_k f(\delta_{k,i}) \prod_{m=i+1}^{h(k)} \delta_{k,m} \tag{2}$$

- E^{R} : the total energy consumed in responding to the subsequent $(R_k - 1)$ requests

$$\begin{aligned} E_k^{\mathrm{R}} = \sum_{i=0}^{h(k)} y_k(R_k - 1)\Bigg\{ f(\delta_{k,i}) \prod_{m=i+1}^{h(k)} \delta_{k,m}\Bigg(1 - \sum_{j=0}^{i-1} b_{k,j}\Bigg) \\ + \Bigg(\prod_{m=i}^{h(k)} \delta_{k,m}\Bigg) b_{k,i}\Big(\frac{w_{ca}T}{(R_k - 1)} + \varepsilon_k T\Big)\Bigg\}. \end{aligned} \tag{3}$$

Energy Efficiency Optimization

$$E^{\text{total}}(\boldsymbol{\delta}, \mathbf{b}) \triangleq \sum_{k \in \mathcal{K}} \left(E_k^{\text{C}} + E_k^{\text{R}} \right) \tag{4}$$

Non-convex Mixed Integer Nonlinear Programming (MINLP)

$$\begin{aligned}
\min_{\boldsymbol{\delta}, \mathbf{b}} \quad & E^{\textbf{total}}(\boldsymbol{\delta}, \mathbf{b}) \\
\textbf{s.t.} \quad & \sum_{k \in \mathcal{K}} y_k \prod_{i=0}^{h(k)} \delta_{k,i} \geq \gamma, \\
& b_{k,i} \in \{0, 1\}, \forall k \in \mathcal{K}, i = 0, \cdots, h(k), \\
& \sum_{k \in C_v} b_{k,h(v)} y_k \prod_{j=h(k)}^{h(v)} \delta_{k,j} \leq S_v, \forall\ v \in V, \\
& \sum_{i=0}^{h(k)} b_{k,i} \leq 1, \forall k \in \mathcal{K}.
\end{aligned} \tag{5}$$

Energy Efficiency Optimization

Theorem

The optimization problem defined in (5) *is NP-hard.*

Proof.

The optimization problem (5) can be reduced to a general non-convex MINLP problem. Since non-convex MINLP is NP-hard, the optimization problem described in (5) is NP-hard. □

Remark

The objective function E^{total} *defined in* (5) *is monotonically increasing in the number of requests* R_k *for all* $k \in \mathcal{K}$ *provided that* δ *and* $\mathbf{b}$ *are fixed.*

Notice that (2) is independent of R_k and (3) is linear in R_k, and its multipliers are positive. Hence, for any fixed $\mathbf{b}$ and δ, (4) increases monotonically with R_k.

Energy Efficiency Optimization

Remark

Given a fixed network scenario, if we increase the number of requests R_k for the data generated by leaf node k, then these data will be cached closer to the sink node or at the sink node, if there exists enough cache capacity, to reduce the overall energy consumption.

For fixed δ, observe from (3) that energy consumption decreases if the cache is moved closer to the root as the nodes deep in the tree do not need to retransmit.

Symbolic Reformulation

Non-Convex MINLP problem

$$
\begin{aligned}
\min \quad & \psi(X, Y) \\
\text{s.t.} \quad & G(X, Y) \leq 0 \\
& H(X, Y) = 0 \\
& X^L \leq X \leq X^U, X \in R \\
& Y \in [Y^L, \ldots, Y^U]
\end{aligned}
$$

Reformulated Problem

$$
\begin{aligned}
\min_{w} \quad & w_{obj} \\
\text{s.t.} \quad & Aw = b \\
& w^l \leq w \leq w^u \\
& Y \in [Y^L, \ldots, Y^U] \\
& w_k \equiv w_i w_j \quad \forall \quad (i, j, k) \in \tau_{bt} \\
& w_k \equiv \frac{w_i}{w_j} \quad \forall \quad (i, j, k) \in \tau_{lft} \\
& w_k \equiv w_i^n \quad \forall \quad (i, k, n) \in \tau_{et} \\
& w_k \equiv fn(w_i) \quad \forall \quad (i, k) \in \tau_{uft}
\end{aligned}
$$

Symbolic Reformulation

Example

We consider $k = 1$ and $h(k) = 1$ in (5), i.e., one leaf node and one sink node. Then (2) and (3) reduce to

$$\begin{aligned} E_1^C &= y_1 f(\delta_{1,0})\delta_{1,1} + y_1 f(\delta_{1,1}), \\ E_1^R &= y_1(R_1 - 1)\left[f(\delta_{1,0})\delta_{1,1} + \delta_{1,0}\delta_{1,1} b_{1,0}(w_{ca}T + \delta_{1T})\right] \\ &+ y_1(R_1 - 1)\left[f(\delta_{1,1})(1 - b_{1,0}) + \delta_{1,0}\delta_{1,1} b_{1,1}(w_{ca}T + \delta_{1T})\right] \end{aligned} \tag{6}$$

$$\begin{aligned} \min_{\boldsymbol{\delta},\mathbf{b}} \quad & E^{\text{total}}(\boldsymbol{\delta}, \mathbf{b}) = E_1^C + E_1^R \\ \text{s.t.} \quad & y_1\delta_{1,0}\delta_{1,1} \geq \gamma, \\ & b_{1,0}, b_{1,1} \in \{0, 1\}, \\ & b_{1,0} y_1 \delta_{1,0}\delta_{1,1} \leq S_0, \\ & b_{1,1} y_1 \delta_{1,1} \leq S_1, \\ & b_{1,0} + b_{1,1} \leq 1. \end{aligned} \tag{7}$$

Symbolic Reformulation

$$
\begin{aligned}
\min_{\boldsymbol{\delta},\mathbf{b}} \quad & w_{\text{obj}} \\
\text{s.t.} \quad & y_1 w_{1,0}^{\text{bt}} \geq \gamma, \\
& b_{1,0}, b_{1,1} \in \{0,1\}, \\
& y_1 \overline{w}_{1,0}^{\text{bt}} \leq S_0, \\
& y_1 \tilde{w}_{1,1}^{\text{bt}} \leq S_1, \\
& b_{1,0} + b_{1,1} \leq 1, \\
& w_{1,0}^{\text{bt}} = \delta_{1,1} \times \delta_{1,0}, \\
& w_{1,0}^{\text{lft}} = \delta_{1,1} / \delta_{1,0}, \\
& \overline{w}_{1,0}^{\text{bt}} = b_{1,0} \times w_{1,0}^{b}, \\
& \tilde{w}_{1,1}^{\text{bt}} = b_{1,1} \times \delta_{1,1}, \\
& \tilde{w}_{1,0}^{\text{bt}} = \delta_{1,1} \times b_{1,0}, \\
& \tilde{w}_{1,0}^{\text{lft}} = b_{1,0} / \delta_{1,1},
\end{aligned}
$$

$$
\begin{aligned}
w_{\text{obj}} = {} & y_1 \varepsilon_{1R} \delta_{1,1} + \varepsilon_{1T} y_1 w_{1,0}^{\text{bt}} + y_1 \varepsilon_{1C} w_{1,0}^{\text{lft}} - y_1 \varepsilon_{1C} \delta_{1,1} \\
& + y_1 \varepsilon_{1R} + \varepsilon_{1T} y_1 \delta_{1,1} + y_1 \varepsilon_{1C} / \delta_{1,1} - y_1 \varepsilon_{1C} \\
& + y_1 (R_1 - 1) \Big[\varepsilon_{1R} \delta_{1,1} + \varepsilon_{1T} w_{1,0}^{\text{bt}} + \varepsilon_{1C} w_{1,0}^{\text{lft}} - \varepsilon_{1C} \delta_{1,1} \\
& + w_{ca} T \overline{w}_{1,0}^{bt} / (R_1 - 1) + \varepsilon_{1T} \overline{w}_{1,0}^{bt} \Big] + y_1 (R_1 - 1) \Big[\varepsilon_{1R} \\
& + \delta_{1,1} \varepsilon_{1T} + \varepsilon_{1C} / \delta_{1,1} - \varepsilon_{1C} - \varepsilon_{1R} b_{1,0} - \varepsilon_{1T} \tilde{w}_{1,0}^{bt} \\
& - \varepsilon_{1C} \tilde{w}_{1,0}^{lft} + \varepsilon_{1C} b_{1,0} + \tilde{w}_{1,1}^{bt} \Big(w_{ca} T / (R_1 - 1) + \varepsilon_{1T} \Big) \Big]
\end{aligned} \tag{8}
$$

Branch-and-Bound

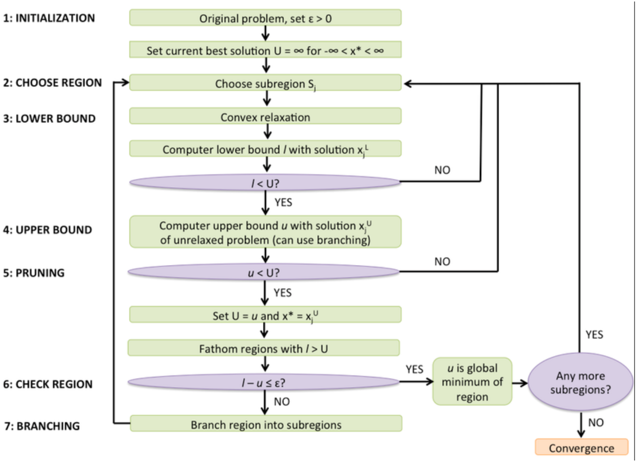

Figure: BBM example (taken from `https://optimization.mccormick.northwestern.edu/index.php/File:SBB_flowchart.png`)

Branch-and-Bound

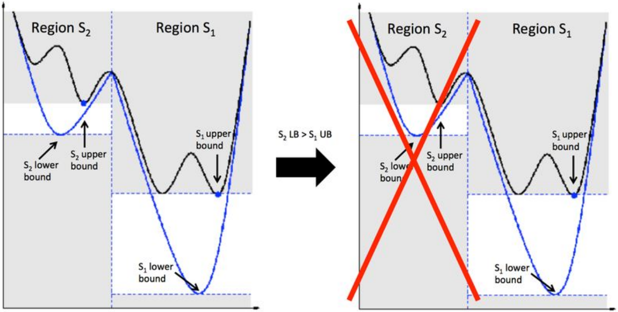

Figure: BBM example (taken from `https://optimization.mccormick.northwestern.edu/index.php/File:SBB.png`)

Spatial Branch and Bound

- Decomposes non-linear functions of the original problem symbolically and recursively with simple operators into simple functions

Table: Summary of notations

Notation	Description
ϕ^u	upper bound of the objective function
$\mathcal{L}$	list of regions
$\mathcal{R}$	any sub-region in $\mathcal{L}$
$\phi^{\mathcal{R},u}$	upper bound on the objective function in subregion $\mathcal{R}$
$\phi^{\mathcal{R},l}$	lower bound on the objective function in subregion $\mathcal{R}$
ϵ	difference between the upper and lower bound
$w_i^{\mathcal{R},l}$	lower bound on auxiliary variable w_i in subregion $\mathcal{R}$
$w_i^{\mathcal{R},u}$	upper bound on auxiliary variable w_i in subregion $\mathcal{R}$

V-SBB

Algorithm 1 Variant of Spatial Branch-and-Bound (V-SBB)

Step 1: Initialize $\phi^u := \infty$ and $\mathcal{L}$ to a single domain
Step 2: Choose a subregion $\mathcal{R} \in \mathcal{L}$ using *least lower bound rule*
if $\mathcal{L} = \emptyset$ **then** Go to Step 6
if for chosen region $\mathcal{R}$, $\phi^{\mathcal{R},l}$ is infeasible or $\phi^{\mathcal{R},l} \geq \phi^u - \epsilon$ **then** Go to Step 5
Step 3: Obtain the upper bound $\phi^{\mathcal{R},u}$
if upper bound cannot be obtained or if $\phi^{\mathcal{R},u} > \phi^u$ **then** Go to Step 4
else $\phi^u := \phi^{\mathcal{R},u}$ and, from the list $\mathcal{L}$, delete all subregions $\mathcal{S} \in \mathcal{L}$ such that $\phi^{\mathcal{S},l} \geq \phi^u - \epsilon$
if $\phi^{\mathcal{R},u} - \phi^{\mathcal{R},l} \leq \epsilon$ **then** Go to Step 5
Step 4: Partition $\mathcal{R}$ into new subregions $\mathcal{R}_{\text{right}}$ and $\mathcal{R}_{\text{left}}$
Step 5: Delete $\mathcal{R}$ from $\mathcal{L}$ and go to Step 2
Step 6: Terminate Search
if $\phi^u = \infty$ **then** Problem is infeasible
else ϕ^u is ϵ-global optimal

Evaluation

Table: Parameters used in simulations

Parameter	Value
y_k	1000
R_k	100
w_{ca}	1.88×10^{-6}
T	10s
ε_{vR}	50×10^{-9}
ε_{vT}	200×10^{-9}
ε_{cR}	80×10^{-9}
γ	$[1, \sum_{k \in \mathcal{K}} y_k]$

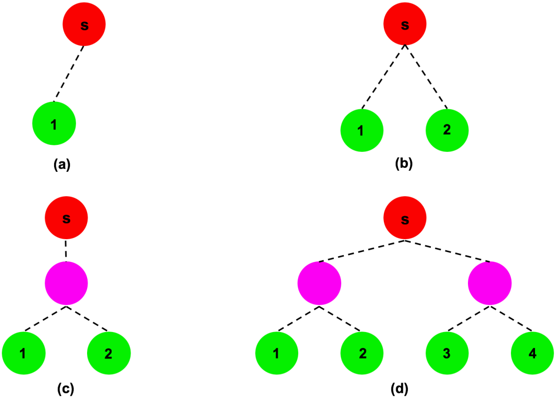

Figure: Candidate network topologies used in the experiments

Table: The Best Solution to the Objective Function (Obj.) and Convergence time for seven nodes network

Solver	$\gamma = 1$		$\gamma = 1000$		$\gamma = 2000$		$\gamma = 3000$		$\gamma = 4000$	
	Obj.	**Time** (s)	**Obj.**	**Time** (s)	**Obj.**	**Time** (s)	**Obj.**	**Time** (s)	**Obj.**	**Time** (s)
Bonmin	0.0002	0.214	0.039	0.164	0.078	0.593	0.117	0.167	0.156	0.212
NOMAD	0.004	433.988	0.121	381.293	0.108	203.696	0.158	61.093	0.181	26.031
GA	0.043	44.538	0.096	30.605	0.164	44.970	0.226	17.307	0.303	28.820
V-SBB	0.0001	1871.403	0.039	25.101	0.078	30.425	0.117	23.706	0.156	19.125
Relaxed	0.0002	0.201	0.039	0.111	0.078	0.095	0.117	0.102	0.156	0.105

Evaluations

Table: Infeasibility of Bonmin for different networks

Networks	(a)	(b)	(c)	(d)
# of test values	1000	2000	2000	4000
# of infeasible solutions	0	0	1	216
Infeasibility (%)	0	0	0.05	5.4

Table: Comparison between V-SBB and Bonmin for smaller values of γ in seven nodes network

Solver	γ =1		γ =5		γ =50	
	Obj.	Time (s)	Obj.	Time	Obj.	Time
Bonmin	0.0002	0.214	0.0003	0.224	0.0021	0.364
V-SBB	0.00011	1871	0.00019	1243	0.0020	3325
Imp. (%)	52.45		50.30		4.62	

Evaluations

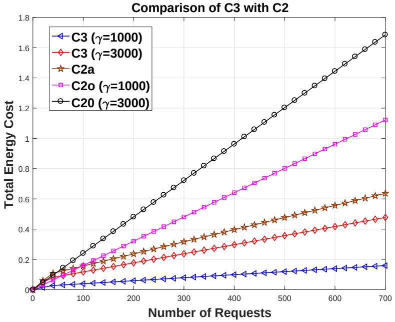

Figure: Comparison of C3 and C2 optimization for the seven nodes network in Figure 4.

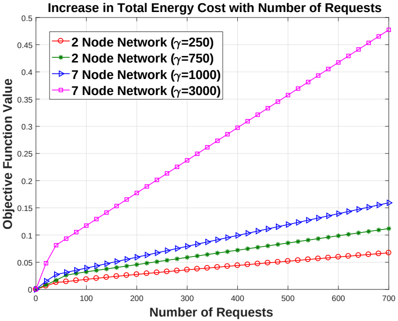

Figure: Total Energy Costs vs. Number of Requests.

Conclusions

- Investigated energy efficiency tradeoff among communication, computation and caching with QoI guarantee in communication networks

Conclusions

- Investigated energy efficiency tradeoff among communication, computation and caching with QoI guarantee in communication networks
- Proposed a variant of the spatial branch-and-bound (V-SBB) algorithm, which can solve the MINLP with ϵ-optimality guarantee

Conclusions

- Investigated energy efficiency tradeoff among communication, computation and caching with QoI guarantee in communication networks
- Proposed a variant of the spatial branch-and-bound (V-SBB) algorithm, which can solve the MINLP with ϵ-optimality guarantee
- Observed that C3 optimization framework improves energy efficiency by as much as 88% compared with either of the C2 optimizations

Thank you!