Self-Adjusting Grid Networks to Minimize Expected Path Length

Chen Avin, Michael Borokhovich, Bernhard Haeupler, Zvi Lotker

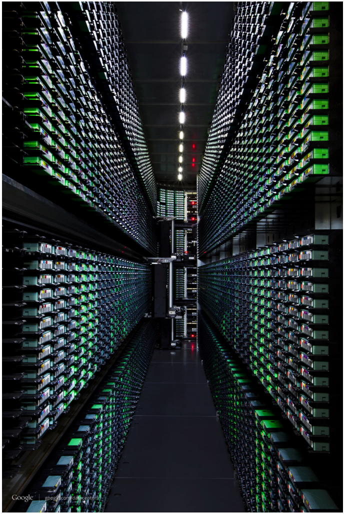
Communication Systems Engineering, BGU, Israel

Computer Science and Artificial Intelligence Laboratory, MIT, USA

SIROCCO 2013

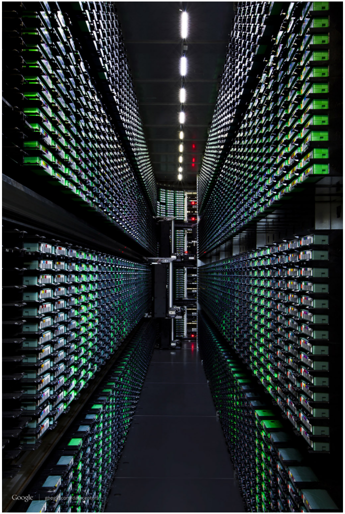
Motivation - Data Centers

- Energy cost ($50B in US alone 2008, doubles every 5 years!)
- Routing consumes about 20-30%



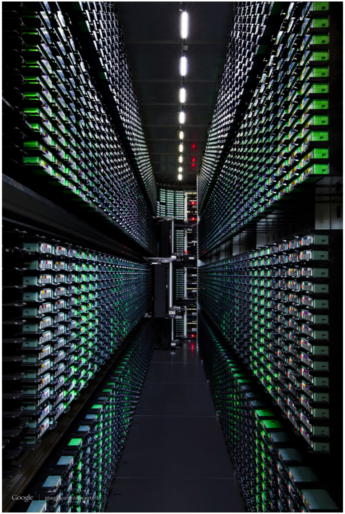
Motivation - Data Centers

- Energy cost ($50B in US alone 2008, doubles every 5 years!)
- Routing consumes about 20-30%
- Need to *adjust* the network, i.e., reduce the expected route length
- Fixed infrastructure...

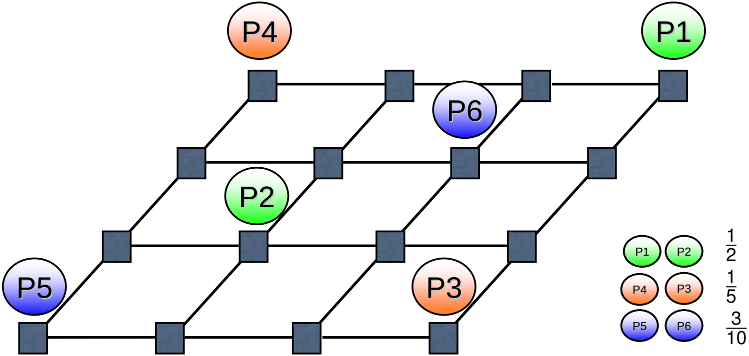


Motivation - Data Centers

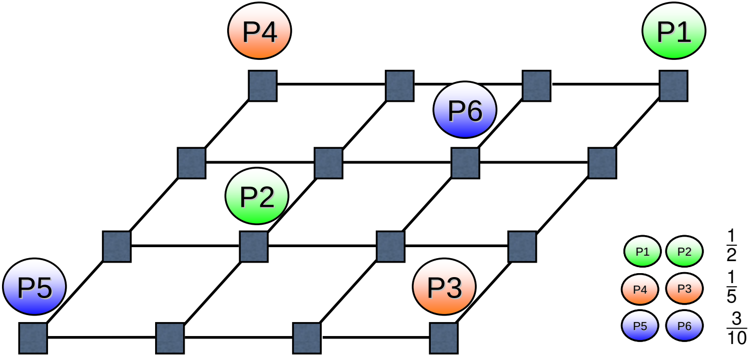
- Energy cost ($50B in US alone 2008, doubles every 5 years!)
- Routing consumes about 20-30%
- Need to *adjust* the network, i.e., reduce the expected route length
- Fixed infrastructure...
- Move processes (e.g., VM) between machines
- Virtualization and SDN (software defined networks), e.g., OpenFlow enable VM migration



Simple Example

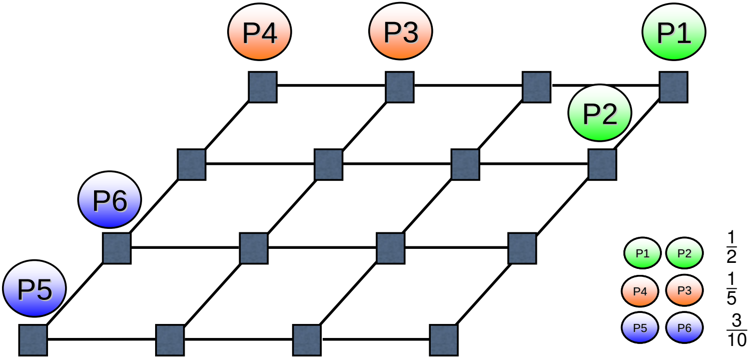


Simple Example



$$\mathbb{E}\left[\text{route length}\right] = \frac{1}{2}4 + \frac{1}{5}6 + \frac{3}{10}4 = 4.4$$

Simple Example

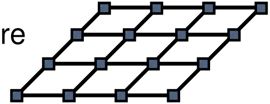


$$\mathbb{E}\left[\text{route length}\right] = \tfrac{1}{2}4 + \tfrac{1}{5}6 + \tfrac{3}{10}4 = 4.4$$

$$\mathbb{E}\left[\text{route length}\right] = \tfrac{1}{2}1 + \tfrac{1}{5}1 + \tfrac{3}{10}1 = 1$$

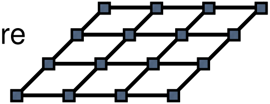
Model and Problem Definition

- ***Host*** Graph: $H(V, E)$: Physical Infrastructure



Model and Problem Definition

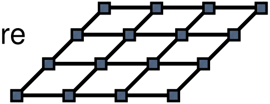
- ***Host*** Graph: $H(V, E)$: Physical Infrastructure



- Routing Requests: $\sigma = (\sigma_1, \sigma_2, \ldots, \sigma_m)$ $\quad \sigma_t = (u, v)$

Model and Problem Definition

- ***Host*** Graph: $H(V, E)$: Physical Infrastructure

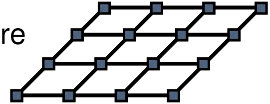


- Routing Requests: $\sigma = (\sigma_1, \sigma_2, \ldots, \sigma_m) \qquad \sigma_t = (u, v)$

- We assume the requests are i.i.d. from a given distribution

Model and Problem Definition

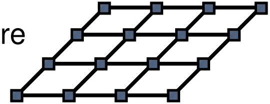
- ***Host*** Graph: $H(V, E)$: Physical Infrastructure



- Requests Distribution

	1	2	3	4	5
1		1/8		1/2	
2			0		
3					0
4		1/9			
5					

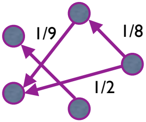
- ***Host*** Graph: $H(V, E)$: Physical Infrastructure



- Requests Distribution

	1	2	3	4	5
1		1/8		1/2	
2			0		
3					0
4		1/9			
5					

- ***Guest*** Weighted Graph: $G(P, W)$
- $|V| = |P| = n$



$$p(u, v)$$

Expected Path Length

- A placement (arrangement):

$$\varphi : P \to V$$

Expected Path Length

- A placement (arrangement):

$$\varphi : P \to V$$

- For H, G, φ:

$$\mathrm{EPL}(\varphi) = \sum_{u,v \in P} \Pr(u,v)\mathrm{d}_H(\varphi(u), \varphi(v))$$

Expected Path Length

- A placement (arrangement):

$$\varphi : P \to V$$

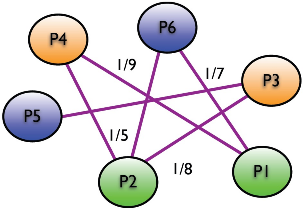
- For H, G, φ:

$$\mathrm{EPL}(\varphi) = \sum_{u,v \in P} \Pr(u,v)\, \mathrm{d}_H(\varphi(u), \varphi(v))$$

- **Minimum Expected Path Length Problem**

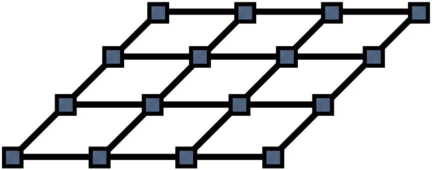
$$\mathrm{MEPL} = \min_{\varphi} \mathrm{EPL}(\varphi)$$

Example



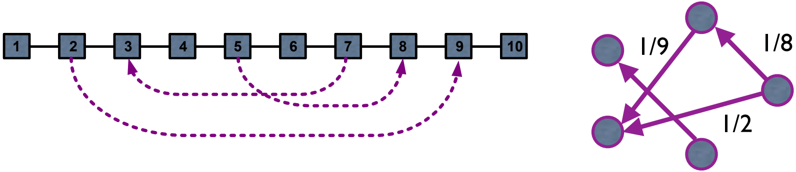
$G(P, W)$

$\varphi : P \to V$ **Find the best way to put processes on the graph to minimize expected path length**



$H(V, E)$

- VLSI layout
- Minimum Linear Arrangement (MLA)
- Known to be hard (NP-Complete)



$$MLA = \min_{\varphi} \sum_{u,v,\in P} w(u,v)|\varphi(u) - \varphi(v)|$$

Hardness of MEPL

- **Host Graph – Grid**
- **Guest Graph – Symmetric Product Distribution**
  - Activity level: $\mathrm{p}(u)$
  - Probability of request: $\mathrm{p}(u, v) = \mathrm{p}(u) \cdot \mathrm{p}(v)$

Hardness of MEPL

- **Host Graph – Grid**
- **Guest Graph – Symmetric Product Distribution**
  - Activity level: $\mathrm{p}(u)$
  - Probability of request: $\mathrm{p}(u, v) = \mathrm{p}(u) \cdot \mathrm{p}(v)$

Lemma

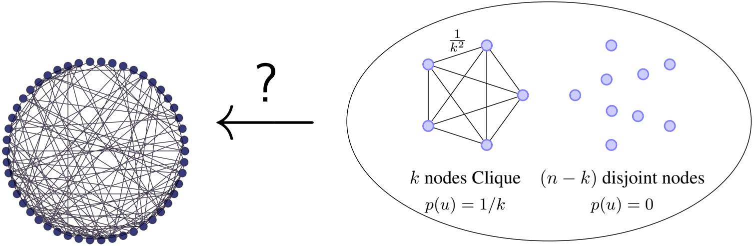
*If G is a **symmetric product distribution**, MEPL is still hard.*

Lemma

*If H is a **2-dimensional grid**, MEPL is still hard.*

Lemma

*If G is a **symmetric product distribution**, MEPL is still hard.*



Host H
arbitrary graph

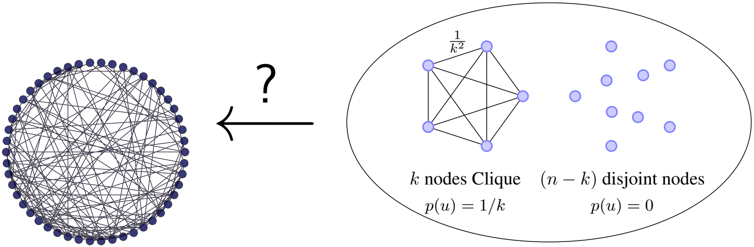
Guest G
symmetric product distribution

$$\mathrm{p}(u, v) = \mathrm{p}(u) \cdot \mathrm{p}(v)$$

Is there a CLIQUE of size k in H ?

Lemma

*If G is a **symmetric product distribution**, MEPL is still hard.*



Host H
arbitrary graph

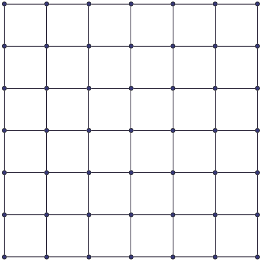
Guest G
symmetric product distribution
$\mathrm{p}(u, v) = \mathrm{p}(u) \cdot \mathrm{p}(v)$

H has a clique of size k if and only if $MEPL = \frac{k(k-1)}{k^2} = 1 - \frac{1}{k}$

Can we embed Tree into Grid?

Lemma

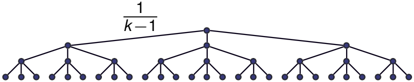
*If H is a **2-dimensional grid**, MEPL is still hard.*



Host H

grid (k^2 nodes)

$\xleftarrow{?}$



Guest G

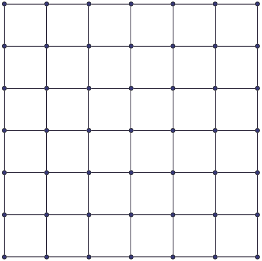
tree (k nodes)

- Can we embed G to H?
- This is hard [Bhatt et al., 1987]

Can we embed Tree into Grid?

Lemma

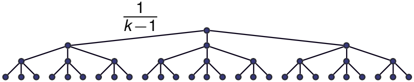
If H is a ***2-dimensional grid****, MEPL is still hard.*



Host H

grid (k^2 nodes)

$\overset{?}{\longleftarrow}$



Guest G

tree (k nodes)

- Can we embed G to H?
- This is hard [Bhatt et al., 1987]

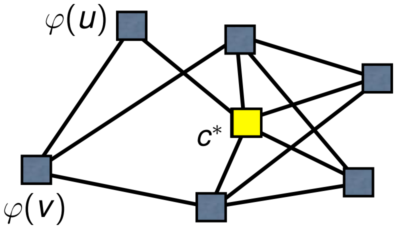
G **can be embedded into** ***H*** **if and only if** $MEPL = \frac{k-1}{k-1} = 1$

Main Result

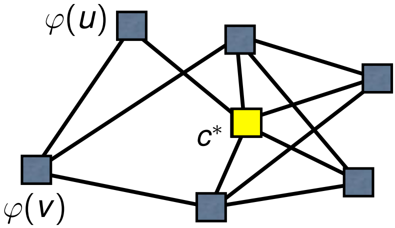
Theorem

*For **d-dimensional grid** (H) and a **symmetric product distribution** (G) there is a **simple distributed algorithm** with a local switching policy between processes and their neighbors that achieves a **constant** approximation to* MEPL

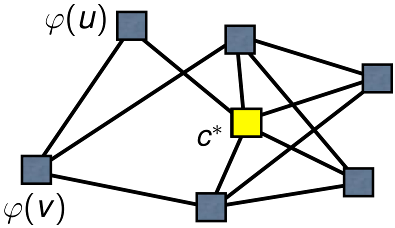
- Expected center: $c^*(\varphi) = \arg\min_x \sum_u \mathrm{p}(u)\mathrm{d}(\varphi(u), \varphi(x))$



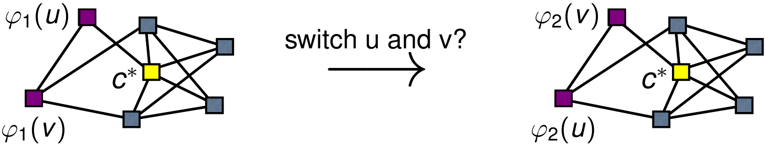
- Expected center: $c^*(\varphi) = \arg\min_x \sum_u \mathrm{p}(u)\mathrm{d}(\varphi(u), \varphi(x))$
- Expected distance to center: $C(\varphi) = \sum_u \mathrm{p}(u)\mathrm{d}(\varphi(u), c^*(\varphi))$



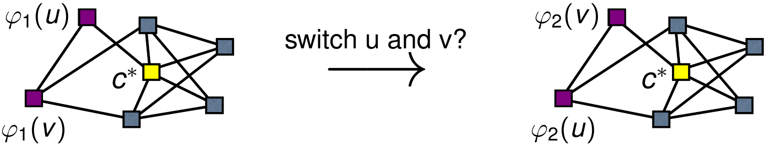
- Expected center: $c^*(\varphi) = \arg\min_x \sum_u \mathrm{p}(u)\mathrm{d}(\varphi(u), \varphi(x))$
- Expected distance to center: $C(\varphi) = \sum_u \mathrm{p}(u)\mathrm{d}(\varphi(u), c^*(\varphi))$
- Minimum expected distance: $C_{\min} = \min_\varphi C(\varphi)$



Switching Rule – Optimize Expected Distance to the Center



Switch only if: $C(\varphi_2) \leq C(\varphi_1)$

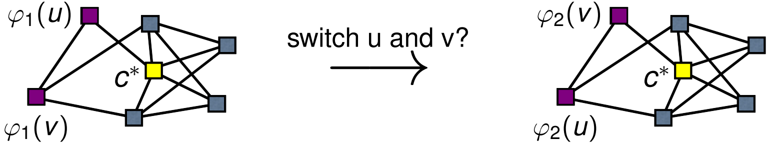


Switch only if: $C(\varphi_2) \leq C(\varphi_1)$

Assumptions:

- Recall that: $C(\varphi) = \sum_u \mathrm{p}(u)\mathrm{d}(\varphi(u), c^*(\varphi))$

Switching Rule – Optimize Expected Distance to the Center

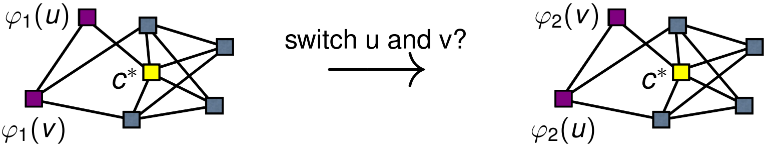


Switch only if: $C(\varphi_2) \leq C(\varphi_1)$

Assumptions:

- Recall that: $C(\varphi) = \sum_u \mathrm{p}(u)\mathrm{d}(\varphi(u), c^*(\varphi))$
- Every node knows current φ (locations of all nodes in H)
  - centralized directory

Switching Rule – Optimize Expected Distance to the Center



Switch only if: $C(\varphi_2) \leq C(\varphi_1)$

Assumptions:

- Recall that: $C(\varphi) = \sum_u \mathrm{p}(u)\mathrm{d}(\varphi(u), c^*(\varphi))$
- Every node knows current φ (locations of all nodes in H)
  - centralized directory
- Every node knows activity level $\mathrm{p}(u)$ of all nodes
  - observing requests over time

- Greedy approach

Switching Rule – Optimize Expected Distance to the Center

- Greedy approach
- Every switch decreases $C(\varphi)$

Switching Rule – Optimize Expected Distance to the Center

- Greedy approach
- Every switch decreases $C(\varphi)$
- Will stop at some local minimum placement $\widehat{\varphi}$

Switching Rule – Optimize Expected Distance to the Center

- Greedy approach
- Every switch decreases $C(\varphi)$
- Will stop at some local minimum placement $\widehat{\varphi}$
- How far local minimum $C(\widehat{\varphi})$ from global minimum $C_{\min}$?

Switching Rule – Optimize Expected Distance to the Center

- Greedy approach
- Every switch decreases $C(\varphi)$
- Will stop at some local minimum placement $\widehat{\varphi}$
- How far local minimum $C(\widehat{\varphi})$ from global minimum C_{min}?
- What can we say about $\mathrm{EPL}(\widehat{\varphi})$?

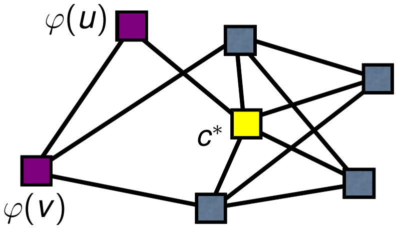
- Greedy approach
- Every switch decreases $C(\varphi)$
- Will stop at some local minimum placement $\widehat{\varphi}$
- How far local minimum $C(\widehat{\varphi})$ from global minimum C_{min}?
- What can we say about $\mathrm{EPL}(\widehat{\varphi})$?
- We show:

$$\frac{C(\widehat{\varphi})}{C_{\text{min}}} = O(1) \quad \text{and} \quad \frac{\mathrm{EPL}(\widehat{\varphi})}{\mathrm{MEPL}} = O(1)$$

MEPL and Minimum Expected Distance to Center

Lemma

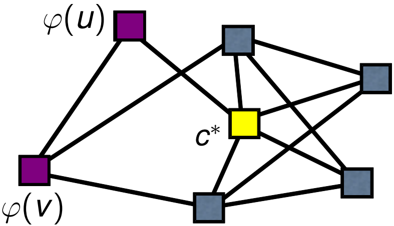
$$\forall \varphi : \quad C(\varphi) \leq \mathrm{EPL}(\varphi) \leq 2C(\varphi)$$



Lemma

$$\forall \varphi : \quad C(\varphi) \leq \mathrm{EPL}(\varphi) \leq 2C(\varphi)$$

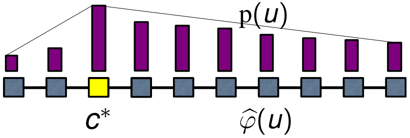
$$\overbrace{\mathrm{d}(\varphi(u), \varphi(v)) \leq \mathrm{d}(\varphi(u), c^*) + \mathrm{d}(c^*, \varphi(v))}$$



Expected Rank

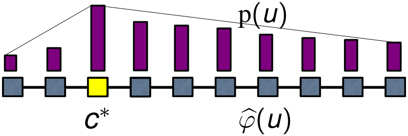
- Rank of a node $\mathrm{r}(u)$ is the position of the node in the ordered list of nodes' activity levels.
- Node with the highest activity level has rank 0.
- $\mathbb{E}[R] = \sum_{u} \mathrm{p}(u)\mathrm{r}(u)$

For any local optimum $\widehat{\varphi}$: $C(\widehat{\varphi}) \leq \mathbb{E}[R]$



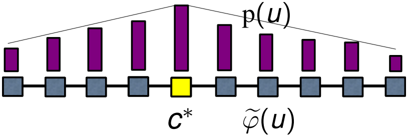
$$\mathrm{d}(\widehat{\varphi}(u), c^*) \leq \mathrm{r}(u)$$

For any local optimum $\widehat{\varphi}$: $C(\widehat{\varphi}) \leq \mathbb{E}[R]$



$$\mathrm{d}(\widehat{\varphi}(u), c^*) \leq \mathrm{r}(u)$$

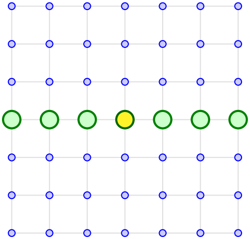
For the global optimum $\widetilde{\varphi}$: $C_{\min} \geq \frac{1}{2}\mathbb{E}[R]$



$$\mathrm{d}(\widetilde{\varphi}(u), c^*) \geq \mathrm{r}(u)/2$$

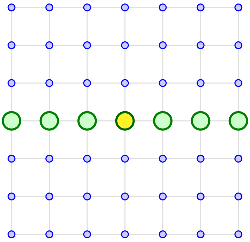
2-Dimensional Grid

$$C(\widehat{\varphi}) \approx \sqrt{n}$$

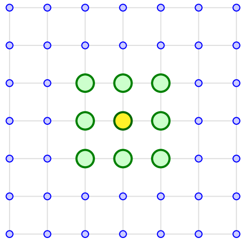


2-Dimensional Grid

$C(\widehat{\varphi}) \approx \sqrt{n}$

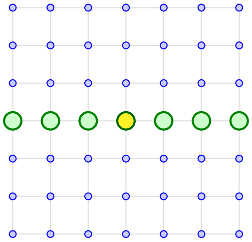


$C_{\min} \approx \sqrt[4]{n}$

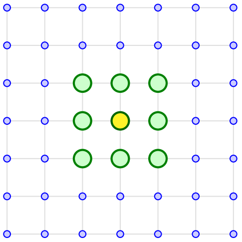


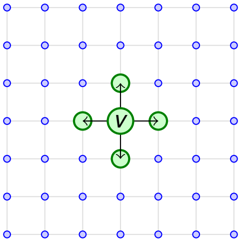
2-Dimensional Grid

$$C(\widehat{\varphi}) \approx \sqrt{n}$$



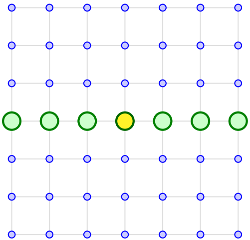
$$C_{\min} \approx \sqrt[4]{n}$$



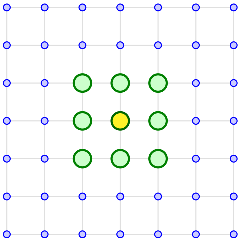


2-Dimensional Grid

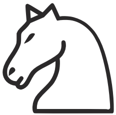
$C(\widehat{\varphi}) \approx \sqrt{n}$

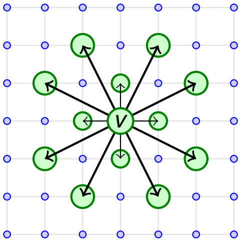


$C_{\min} \approx \sqrt[4]{n}$

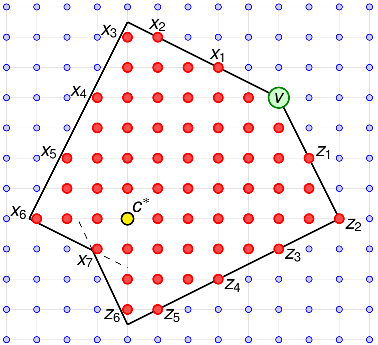


Allow *chess knight* moves





For any local optimum $\widehat{\varphi}$

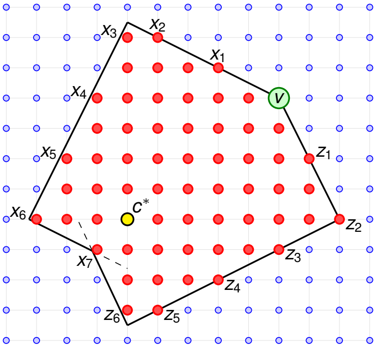


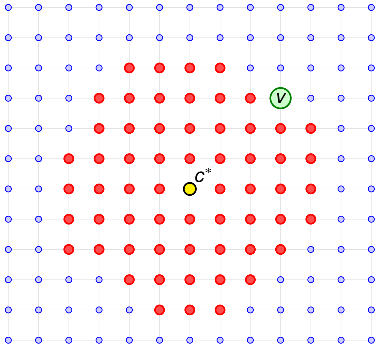
$$C(\widehat{\varphi}) \leq \frac{4}{\sqrt{6}}\mathbb{E}[\sqrt{R}]$$

$$\mathrm{d}(\widehat{\varphi}(v), c^*) \leq \frac{4}{\sqrt{6}}\sqrt{\mathrm{r}(v)}$$

For any local optimum $\widehat{\varphi}$

For the global optimum $\widetilde{\varphi}$





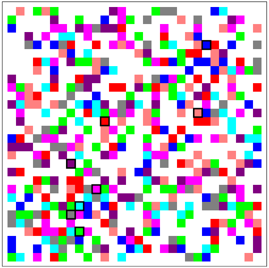
$$C(\widehat{\varphi}) \leq \frac{4}{\sqrt{6}}\mathbb{E}[\sqrt{R}]$$

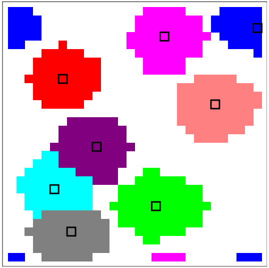
$$C_{\min} \geq \frac{1}{\sqrt{2}}\mathbb{E}[\sqrt{R}]$$

$$\mathrm{d}(\widehat{\varphi}(v), c^*) \leq \frac{4}{\sqrt{6}}\sqrt{\mathrm{r}(v)}$$

$$\mathrm{d}(\widetilde{\varphi}(v), c^*) \geq \sqrt{\frac{\mathrm{r}(v)}{2}}$$

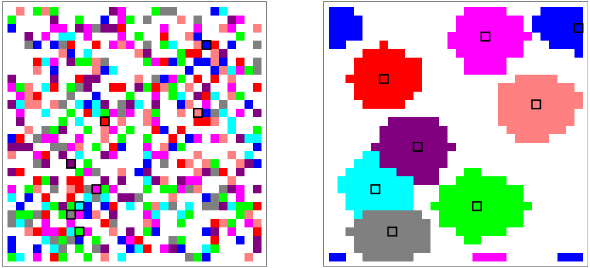
Simulations – Clustered Requests



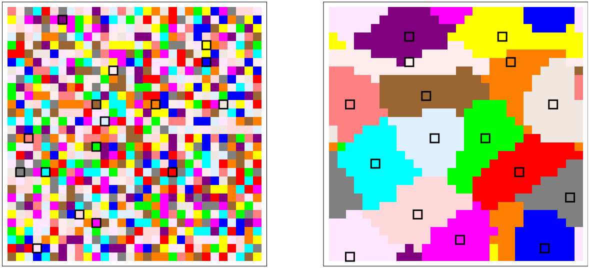


900 nodes
50% inactive
8 clusters

Simulations – Clustered Requests



900 nodes
50% inactive
8 clusters



900 nodes
16 clusters

Animation

- MEPL is hard for general graphs and requests patterns
- For **grids** and **symm. product distr.** we showed greedy approach that is a constant approximation

Summary

- MEPL is hard for general graphs and requests patterns
- For **grids** and **symm. product distr.** we showed greedy approach that is a constant approximation
- Future work:
  - Real datacenters infrastructure
  - More requests patterns

Summary

- MEPL is hard for general graphs and requests patterns
- For **grids** and **symm. product distr.** we showed greedy approach that is a constant approximation
- Future work:
  - Real datacenters infrastructure
  - More requests patterns

THANK YOU!