Probabilistic Analysis of Network Availability

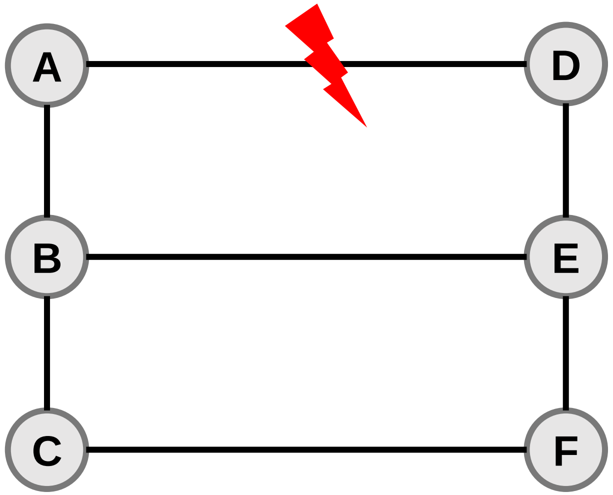
Yunmo Zhang[1], Hong Xu[2], Chun Jason Xue[1], Tei-Wei Kuo[1]

[1]City University of Hong Kong [2]Chinese University of Hong Kong

IEEE ICNP 2022

Prior Network Verification

HSA[1], Batfish[2], Minesweeper[3], …



Reachability

Is E reachable from A?

[1]Kazemian et al., “Header Space Analysis: Static Checking for Networks,” in Proc. USENIX NSDI, 2012.
[2]Fogel et al, “A General Approach to Network Configuration Analysis,” in Proc. USENIX NSDI, 2015
[3]Beckett et al,, “A General Approach to Network Configuration Verification,” in Proc. ACM SIGCOMM, 2017

Prior Network Verification

HSA[1], Batfish[2], Minesweeper[3], …

Are there other properties?

Reachability

√ For path-related (qualitative) properties

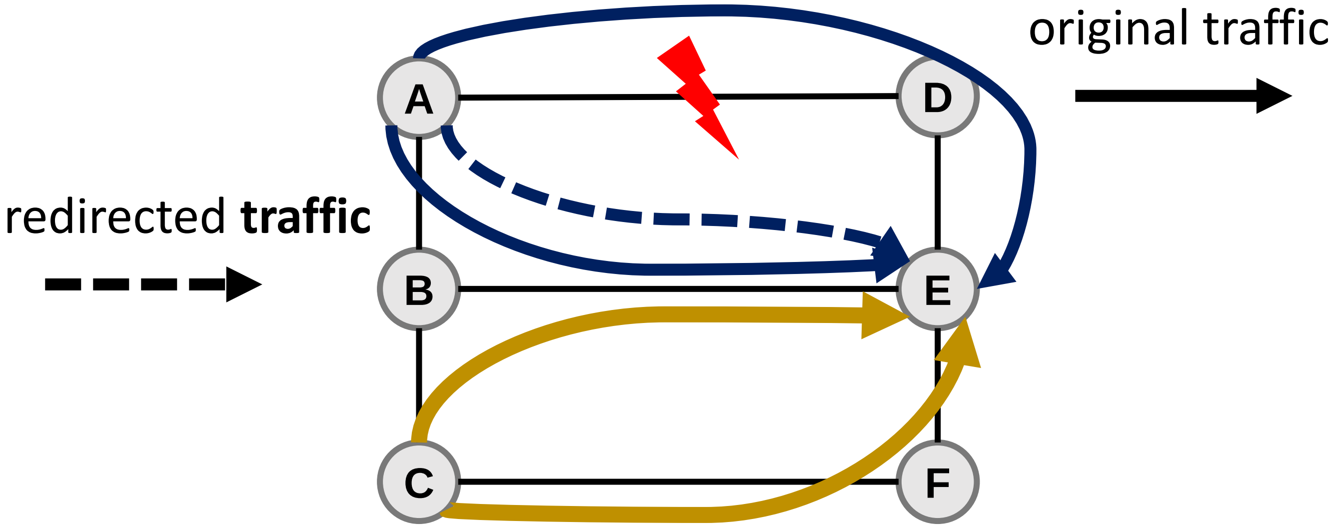
√ Answering a deterministic question

Is E reachable from A?

[1]Kazemian et al., “Header Space Analysis: Static Checking for Networks,” in Proc. USENIX NSDI, 2012.
[2]Fogel et al, “A General Approach to Network Configuration Analysis,” in Proc. USENIX NSDI, 2015
[3]Beckett et al,, “A General Approach to Network Configuration Verification,” in Proc. ACM SIGCOMM, 2017

Prior Network Verification

QARC[1], Chang *et al*[2].

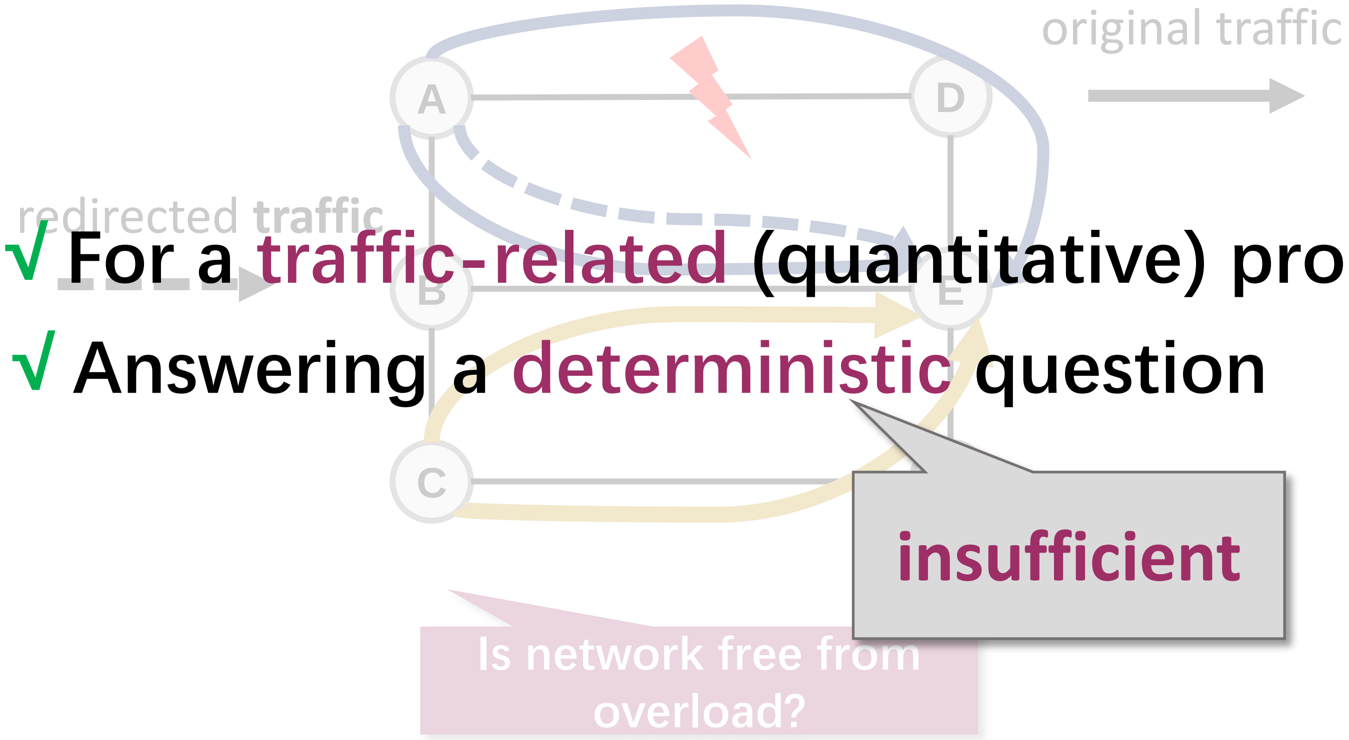


Is network free from overload?

[1]Subramanian et al., Detecting Network Load Violations for Distributed Control Planes," in Proc. ACM PLDI, 2020
[2]Chang et al., "Robust Validation of Network Designs under Uncertain Demands and Failures," in Proc. USENIX NSDI, 2017

Prior Network Verification

QARC[1], Chang *et al*[2].



√ For a traffic-related (quantitative) property

√ Answering a deterministic question

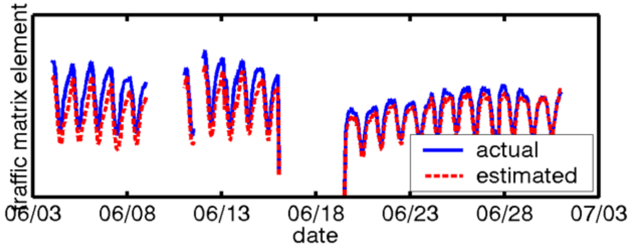
[1]Subramanian et al., Detecting Network Load Violations for Distributed Control Planes," in Proc. ACM PLDI, 2020

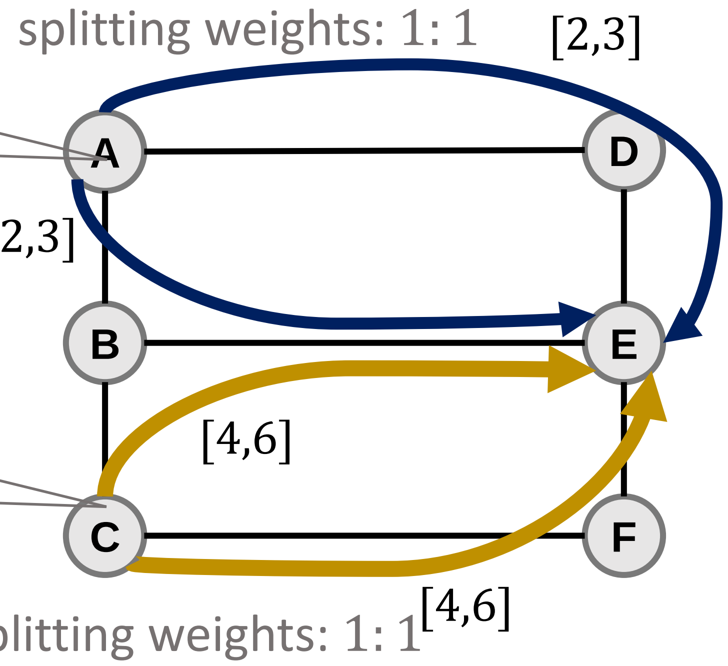
[2]Chang et al., "Robust Validation of Network Designs under Uncertain Demands and Failures," in Proc. USENIX NSDI, 2017

A Running Example

Link Capacity: 10

Demand (A,E) traffic volume fluctuation[1] range: [4,6]





[1]Matthew Roughan, Robust Network Planning (also the image source)

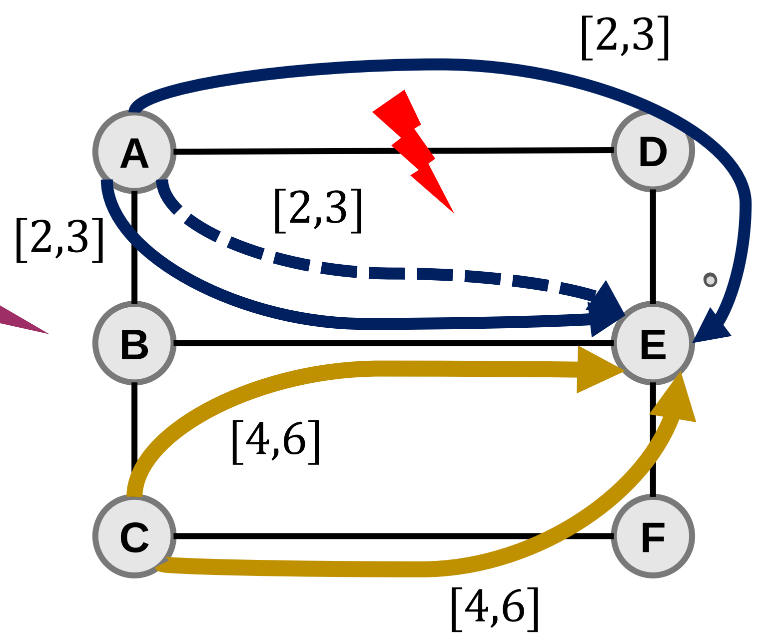
A Running Example

After failure…

Link Capacity: 10

How likely is BE to be overloaded?

Load on BE:
$(0.5 + 0.5)\ d_{AE} + 0.5\ d_{CE}$



Volume Range [4,6]

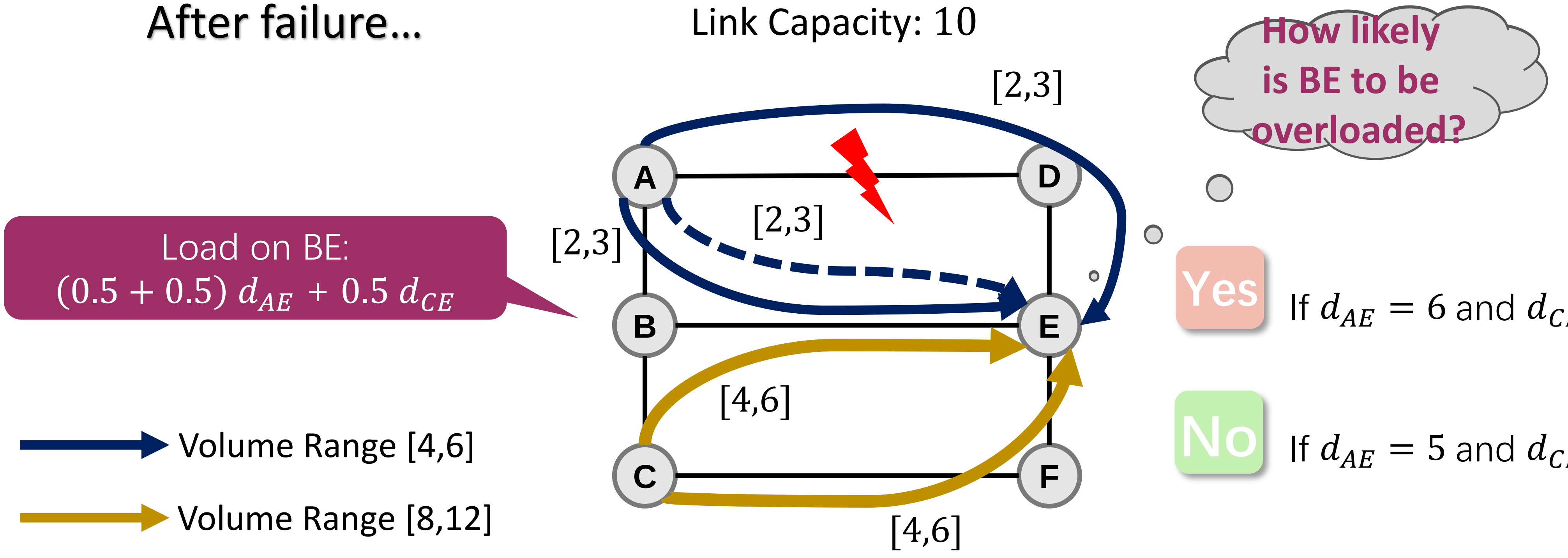
Volume Range [8,12]

Yes If $d_{AE} = 6$ and $d_{CE} = 12$

No If $d_{AE} = 5$ and $d_{CE} = 10$

A Running Example

After failure...



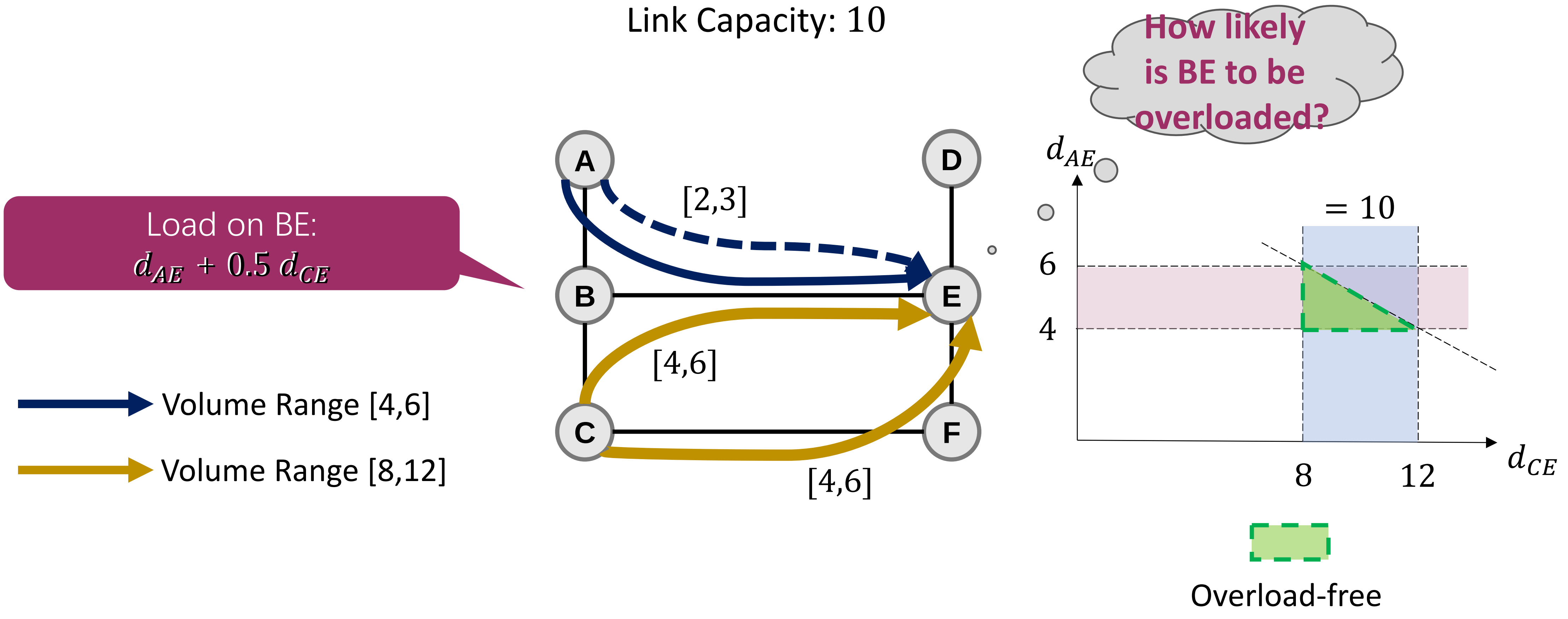
If $d_{AE} = 6$ and $d_{CE} = 12$

If $d_{AE} = 5$ and $d_{CE} = 10$

Simple yes or no answers could not profile the network availability comprehensively.

Probabilistic analysis naturally fits here.

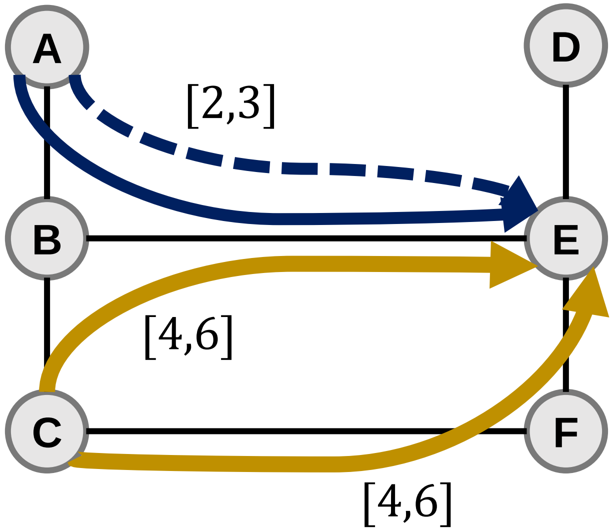
Probabilistic Analysis



Probabilistic Analysis

Link Capacity: 10

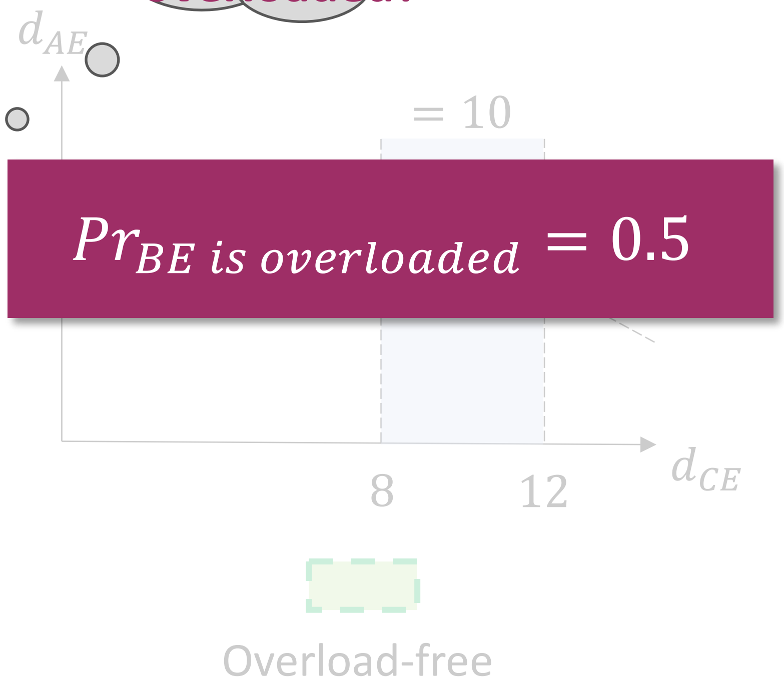
Load on BE:
$d_{AE} + 0.5\, d_{CE}$



Volume Range [4,6]

Volume Range [8,12]

How likely is BE to be overloaded?



$Pr_{BE\ is\ overloaded} = 0.5$

Probabilistic Phenomenon

*For all RDS instances hosted in multiple Availability Zones (with the 'Multi AZ' parameter set to 'True'), Amazon guarantees **99.5%** uptime in any monthly billing cycle.*

aws

*The Covered Service will provide a Monthly Uptime Percentage to Customer of at least **99.9%** (the "Service Level Objective " or "SLO")*

Google Cloud

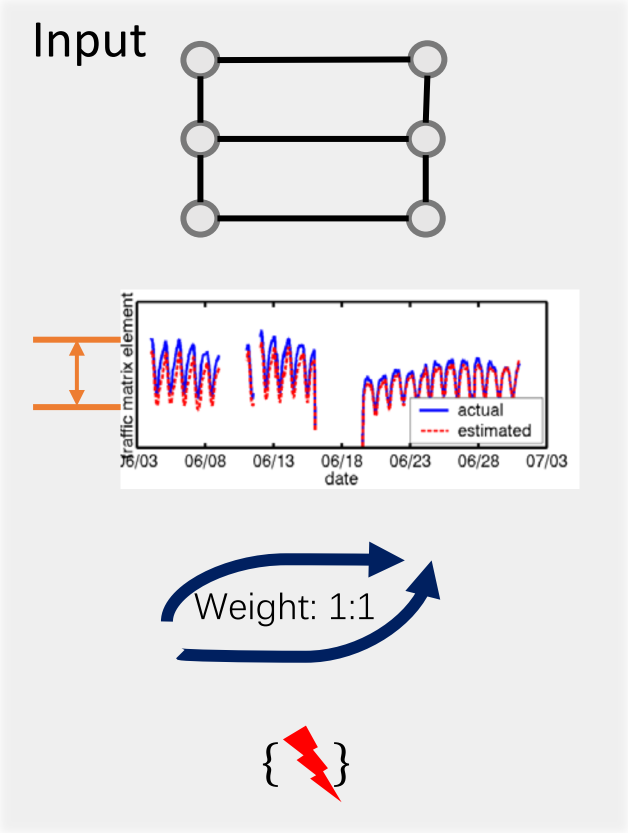
*We guarantee that at least **99.9%** of the time CDN will respond to client requests and deliver the requested content without error.*

Microsoft Azure

Probabilistic Analysis of Network Availability: Pita

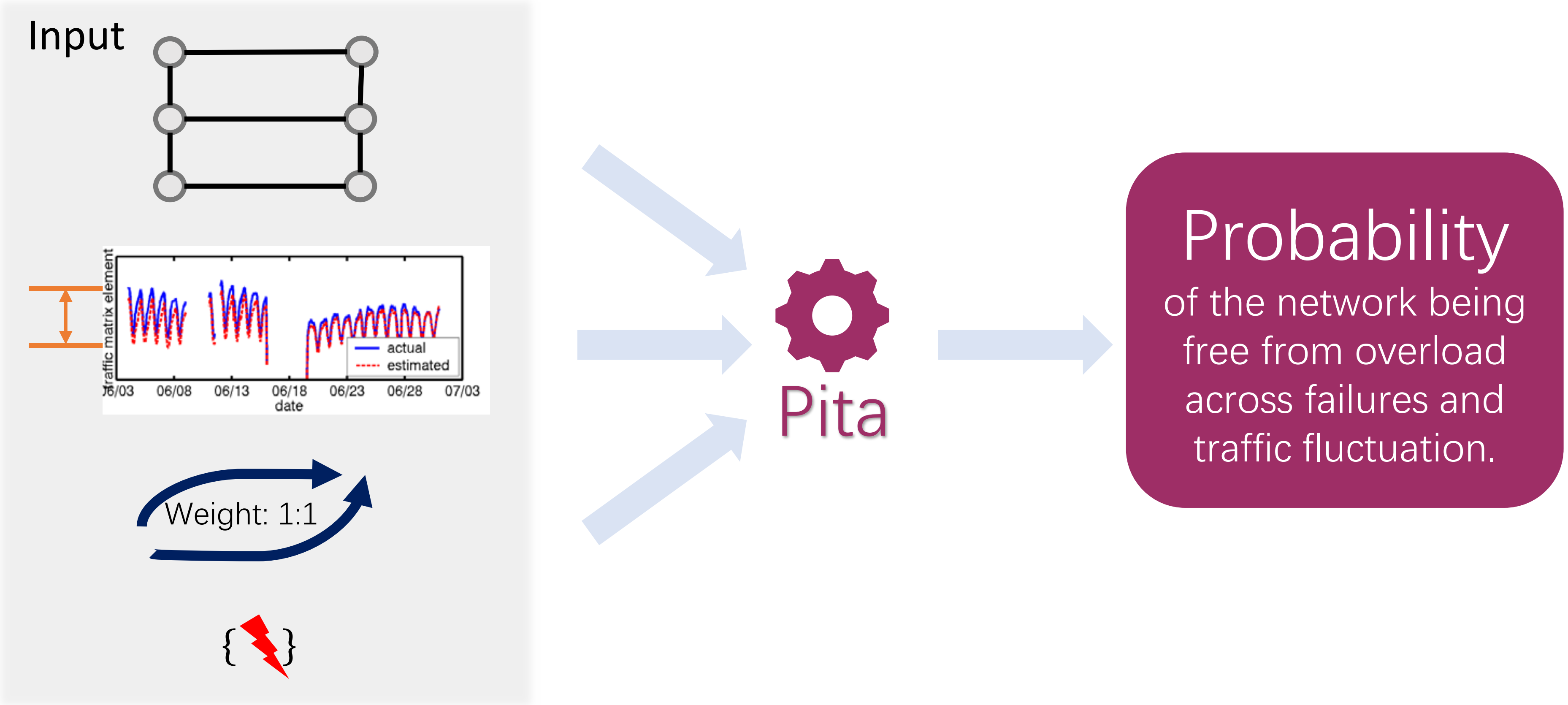
Pita Overview

Given the network topology, the range of traffic fluctuation, traffic tunnels with their splitting weights and a range of failure scenarios,



Pita Overview

Pita outputs the overall probability of the network being available.



Outline

Background and Motivation

Pita Overview

Problem Formulation

Solution

Evaluation

Problem Formulation

The probability of network availability

$$\boxed{\text{Probability of the network being free from overload across failures and traffic fluctuation.}} = \sum_{f \in F} \boxed{Pr(\phi_f)} \cdot \boxed{Pr(f)}$$

Problem Formulation

F is a set of failure scenarios concerned (input from users).

f is a failure scenario (a set of links failed).

$$\underbrace{\text{Probability}}_{\text{of the network being free from overload across failures and traffic fluctuation.}} = \sum_{f \in F} Pr(\phi_f) \cdot Pr(f)$$

Problem Formulation

F is a set of failure scenarios concerned (input from users).

f is a failure scenario (a set of links failed).

Probability of the network being free from overload across failures and traffic fluctuation.

$$= \sum_{f \in F} Pr(\phi_f) \cdot Pr(f)$$

The probability that f happens, e.g., 0.001

Problem Formulation

Overload-free property ϕ_f: the network could accommodate the traffic fluctuation under a scenario f **without link overloaded** (other than the failed ones).

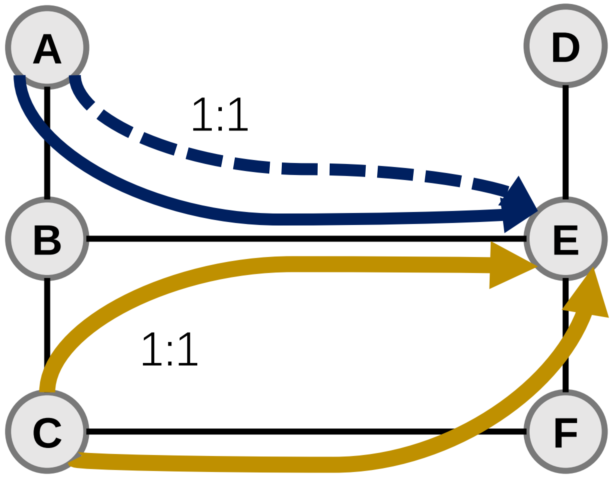
Probability of the network being free from overload across failures and traffic fluctuation.

$$= \sum_{f \in F} Pr(\phi_f) \cdot Pr(f)$$

Overload-free Probability $\Pr(\phi_f)$

Problem Formulation

Overload-free property ϕ_f: the network could accommodate the traffic fluctuation under a scenario f **without link overloaded** (other than the failed ones).



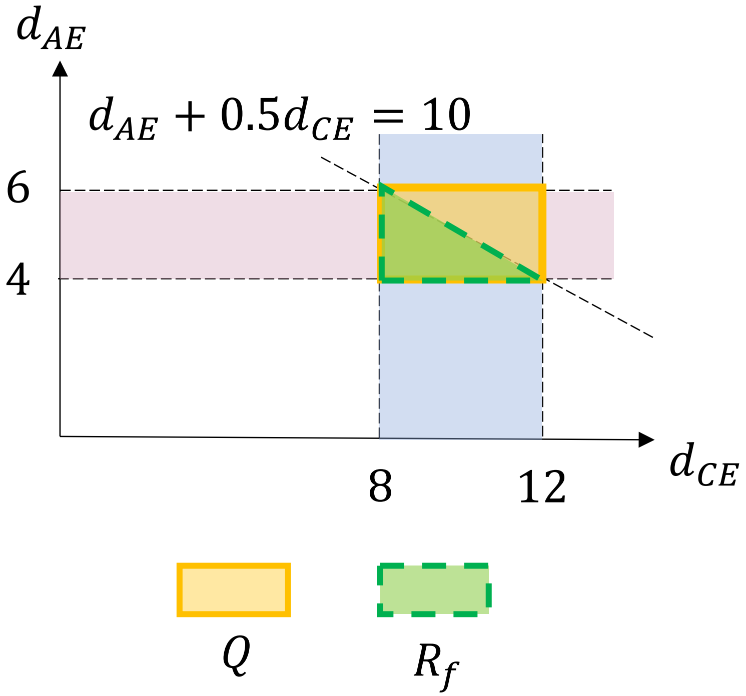
overload-free ϕ_f $=$

Link BE:	$d_{AE} + 0.5 d_{CE} \leq 10$
Link AB:	$\wedge\ d_{AE} \leq 10$
Link BC/CF/EF:	$\wedge\ 0.5 d_{CE} \leq 10$
	$\vdots$

Link capacity constraints

Problem Formulation

Overload-free probability $\Pr(\phi_f)$ is the Lebesgue measure of ϕ_f in the whole traffic set Q.



$$
\begin{aligned}
& d_{AE} && \leq 6 \\
\wedge\ & -d_{AE} && \leq -4 \\
\wedge\ & d_{CE} && \leq 12 \\
\wedge\ & -d_{CE} && \leq -8
\end{aligned}
$$

Demand (A,E) range constraints

Demand (C,E) range constraints

whole Set Q

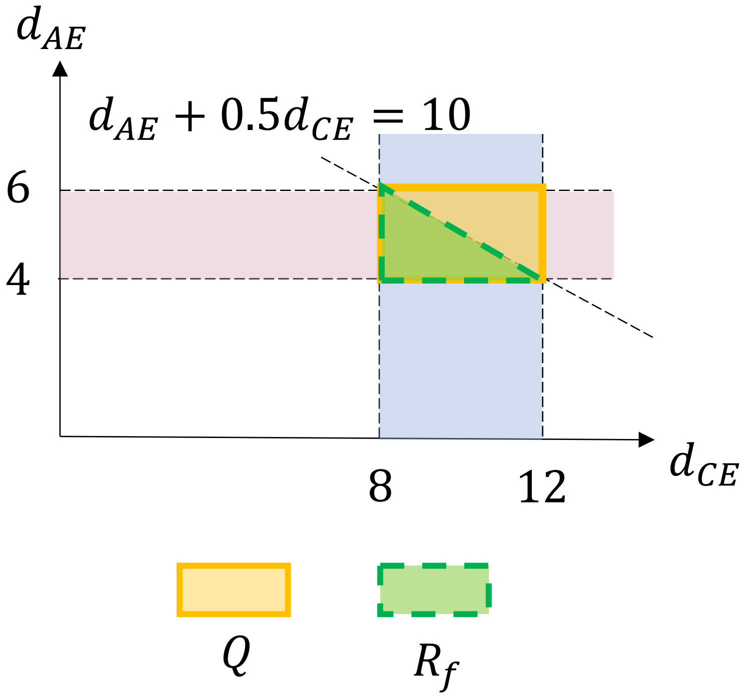
$$
d_{AE} + 0.5d_{CE} \leq 10
$$

$\vdots$

ϕ_f

Problem Formulation

Overload-free probability $\Pr(\phi_f)$ is the Lebesgue measure of ϕ_f in the whole traffic set Q.

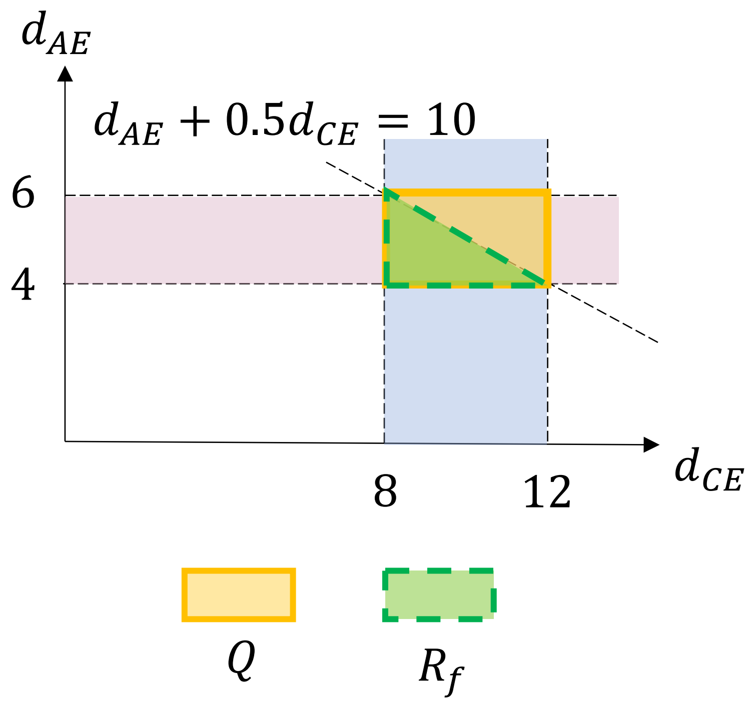


$$R_f = \{(d_{AE}, d_{CE}) \mid \phi_f \wedge Q\}$$

$$\Pr(\phi_f) = vol(R_f)/vol(Q)$$

Computing $\Pr(\phi_f)$ in the Running Example

Overload-free probability $\Pr(\phi_f)$ is the Lebesgue measure of ϕ_f in the whole traffic set Q.



How about R_f and Q in higher dimensions?

$$R_f = \{(d_{AE}, d_{CE}) \mid \phi_f \wedge Q\}$$

$$\Pr(\phi_f) = vol(R_f)/vol(Q)$$

The area of a convex polygon when R_f is **2-dimensional.**

The area of a rectangle when Q is **2-dimensional.**

Computing $\Pr(\phi_f) = vol(R_f)/vol(Q)$

- Geometrically, the whole set $Q = \{(d_1, \ldots, d_n) | \wedge_{i=1}^{n} \; L_i \leq d_i \leq U_i\}$ is an n-dimensional hyperrectangle defined by the ranges of demands.

Regular polytope

n = #demand

$$Q = \overbrace{\begin{array}{rr} d_{AE} & \\ \wedge -d_{AE} & \\ \wedge & d_{CE} \\ \wedge & -d_{CE} \end{array}}^{n = \#\text{demand}} \begin{array}{l} \leq 6 \\ \leq -4 \\ \leq 12 \\ \leq -8 \end{array}$$

$$vol(Q) = \prod_{i=1}^{n} (U_i - L_i)$$

Computing $\Pr(\phi_f) = vol(R_f)/vol(Q)$

Geometrically, $R_f = \{(d_1, \dots, d_n) | \phi_f \wedge Q\}$ is an n-dimensional polytope enclosed by m hyperplanes.

E.g., at most 100 for a network of ten nodes

n = #demand

$$R_f = \left\{ \begin{array}{l} \underbrace{\begin{array}{rl} \phi_f = & d_{AE} + 0.5 d_{CE} \le 10 \\ \wedge & d_{AE} \le 10 \\ \wedge & 0.5 d_{CE} \le 10 \\ & \vdots \end{array}}_{\text{#edge in the network } - \text{ #failed link in } f} \\ \underbrace{\begin{array}{rl} \wedge\ Q = & d_{AE} \le 6 \\ \wedge & -d_{AE} \le -4 \\ \wedge & d_{CE} \le 12 \\ \wedge & -d_{CE} \le -8 \end{array}}_{2 \cdot \text{#demand}} \end{array} \right\}$$

#edge in the network − #failed link in f

2 · #demand

$m = O$(#edge + 2 · #demand)

E.g., at most 220 for a network of ten nodes and twenty edges

Computing $\Pr(\phi_f) = vol(R_f)/vol(Q)$

Geometrically, $R_f = \{(d_1, \dots, d_n) | \phi_f \wedge Q\}$ is an n-dimensional polytope enclosed by m hyperplanes.

- **Irregular and high-dimensional:**

The volume could not be exactly computed when dimension is larger than 15[1].

Proof see paper

- **Convex:**

We could resort to Multiphase Markov Chain Monte Carlo (Multiphase MCMC) to approximate the volume.

[1]B. Bueler, A. Enge, and K. Fukuda, "Exact Volume Computation for Polytopes: A Practical Study," in Polytopes — combinatorics and computation, 2000, pp. 131–154

Outline

- Background and Motivation
- Pita Overview
- Problem Formulation
- Solution
- Evaluation

Solution Takeaways

$\Pr(\phi_f)$ boils down to the volume of a high-dimensional polytope R_f.

The volume of R_f is approximated by Multiphase MCMC:

- Constructing a series of convex bodies that volume ratios multiplication is R_f.
- Estimating a volume ratio by MCMC.

A domain-specific optimization on the random walk of MCMC:

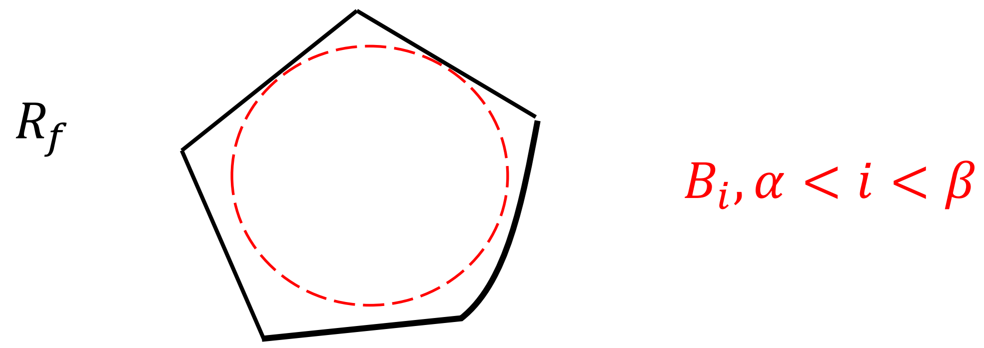
- Exploiting the structural property of our problem.

There are special cases where $\Pr(\phi_f)$ could be determined (See paper).

Approximating $Pr(\phi_f)$ by Multiphase MCMC

1. Constructing a series of convex bodies that volume ratios multiplication is R_f.

We first construct a sequence of convex bodies $K_\alpha \subseteq K_{\alpha+1} \dots \subseteq K_{\beta-1} \subseteq K_\beta$, where convex bodies $\{K_i\}$ are the intersections of R_f and a series of concentric balls $\{B_i\}$.

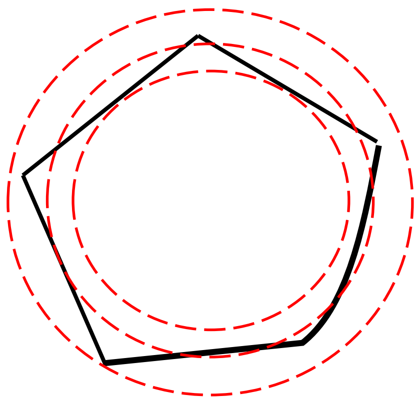


Approximating $Pr(\phi_f)$ by Multiphase MCMC

1. Constructing a series of convex bodies tha known volume s multiplication is R_f

We first construct a sequence of convex bodies $K_\alpha \subseteq K_{\alpha+1} \ldots \subseteq K_{\beta-1} \subseteq K_\beta$, where convex bodies $\{K_i\}$ are the intersections of R_f and a series of concentric balls $\{B_i\}$.

R_f



$B_i, \alpha < i < \beta$

B_α: the largest ball enclosed by R_f

B_β: a ball enclosing R_f

Approximating $Pr(\phi_f)$ by Multiphase MCMC

1. Constructing a series of convex bodies that volume ratios multiplication is R_f.

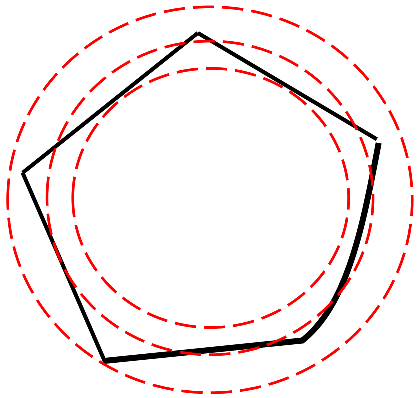
Then, $\mathrm{vol}(\mathrm{R_f})$ is computed by $vol(K_\alpha)\boxed{\frac{vol(K_{\alpha+1})}{vol(K_\alpha)}}\frac{vol(K_{\alpha+2})}{vol(K_{\alpha+1})}\cdots\frac{vol(K_\beta)}{vol(K_{\beta-1})}$

R_f

known volume

Each such ratio is estimated by MCMC (Markov Chain Monte Carlo)

R_f



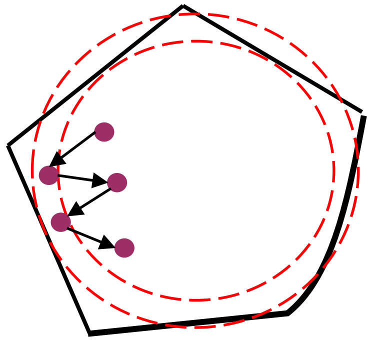
Approximating $Pr(\phi_f)$ by Multiphase MCMC

2. Estimating a volume ratio by MCMC.

$$\mathrm{vol}(\mathrm{R_f}) = vol(K_\alpha)\boxed{\frac{vol(K_{\alpha+1})}{vol(K_\alpha)}}\frac{vol(K_{\alpha+2})}{vol(K_{\alpha+1})}\cdots\frac{vol(K_\beta)}{vol(K_{\beta-1})}$$

We use CDHR as the random walk algorithm in Pita.

R_f



Using (Markov Chain) random walk to generate many (almost) uniformly distributed sample points in K_{i+1}

Counting the number sample points also residing in K_i

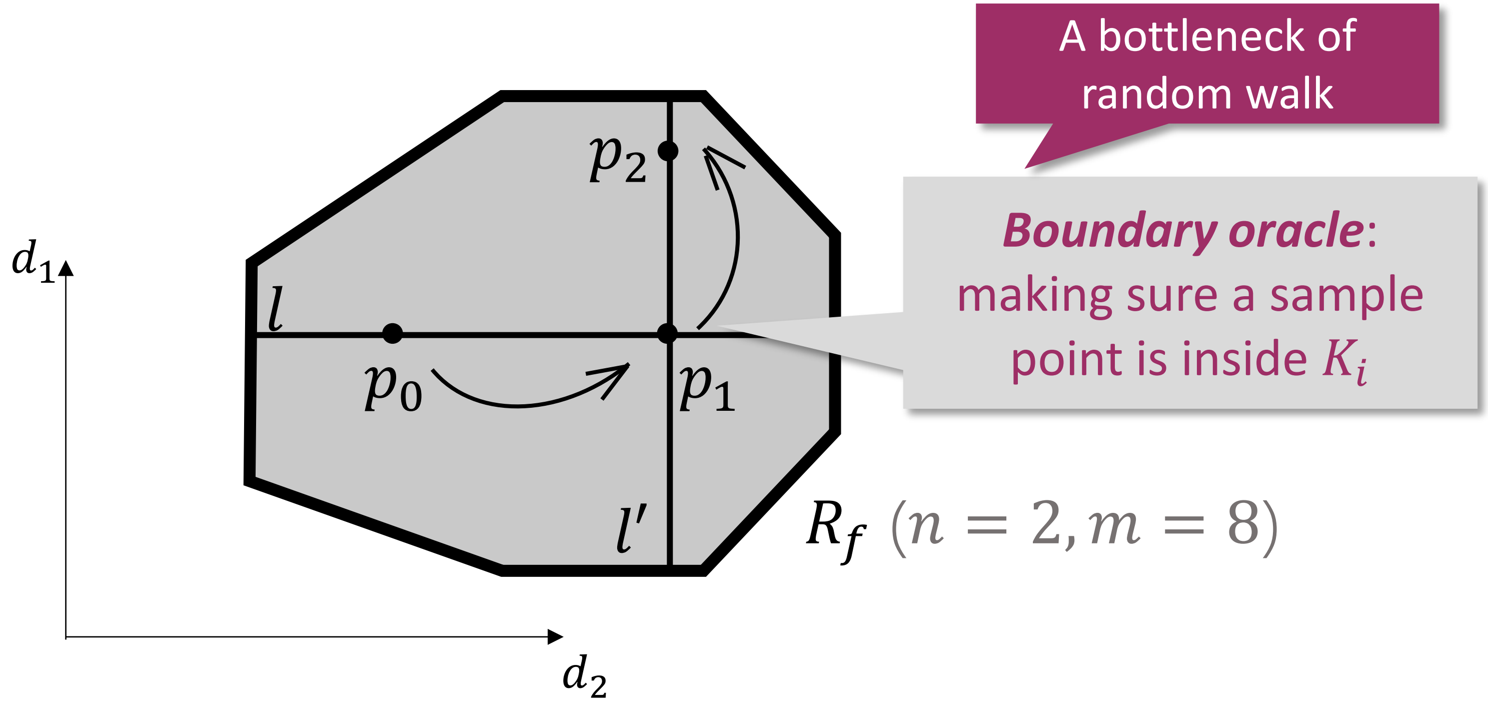
The ratio $\frac{vol(K_{i+1})}{vol(K_i)}$ is estimated by $\frac{\#sample\ points\ in\ K_{i+1}}{\#sample\ points\ in\ K_i}$

Random walk: CDHR

Coordinate Direction Hit-and-Run (CDHR): at each step, it samples (next) point by

(1) randomly picking a line l through current point p_0 who parallel to the axes and

(2) moving current point p_0 to a random point p_1 uniformly distributed on the chord $\mathrm{K_i} \cap l$



A domain-specific optimization on CDHR

The original boundary oracle in CDHR computes the intersection points of line l with m hyperplanes in R_f.

The complexity is $O(m) = O(\#edge + 2 \cdot \#demand)$

$$R_f = \left.\begin{aligned} \phi_f = \quad & d_{AE} + 0.5 d_{CE} \le 10 \\ \wedge \quad & d_{AE} \le 10 \\ \wedge \quad & 0.5 d_{CE} \le 10 \\ & \vdots \\ \wedge\, Q = \quad & d_{AE} \le 6 \\ \wedge & -d_{AE} \le -4 \\ \wedge \quad & d_{CE} \le 12 \\ \wedge \quad & -d_{CE} \le -8 \end{aligned}\right\}$$

m hyperplanes

$2 \cdot \#demand$ hyper-planes **parallel to the axes** (the direction of walking)

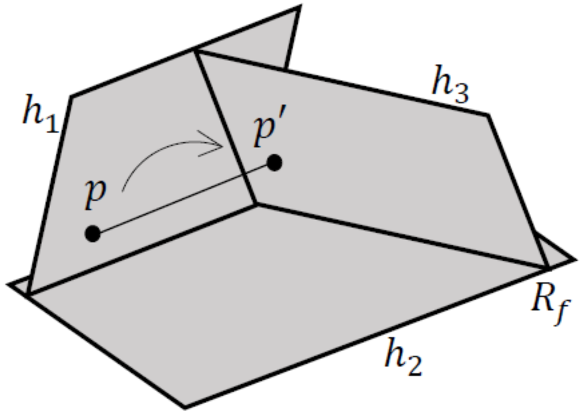
A domain-specific optimization on CDHR

The original boundary oracle in CDHR computes the intersection points of line l with m hyperplanes in R_f.

The complexity is O(m) $= O(\#edge + 2 \cdot \#demand)$

OptHR safely bypasses hyperplanes parallel to the axes.

Proof see paper



The complexity is $O(\#edge)$

OptHR bypasses checking whether p' steps outside the boundary defined by h_1 and h_2

Outline

- Background and Motivation
- Pita Overview
- Problem Formulation
- Solution
- Evaluation

Evaluation Setting

Real Topologies:

- GridNet, Abilene and ANS (from The Internet Topology Zoo)
- B4 [Jain et al.]

Number of demands: 81~300+

Synthetic Traffic Matrices:

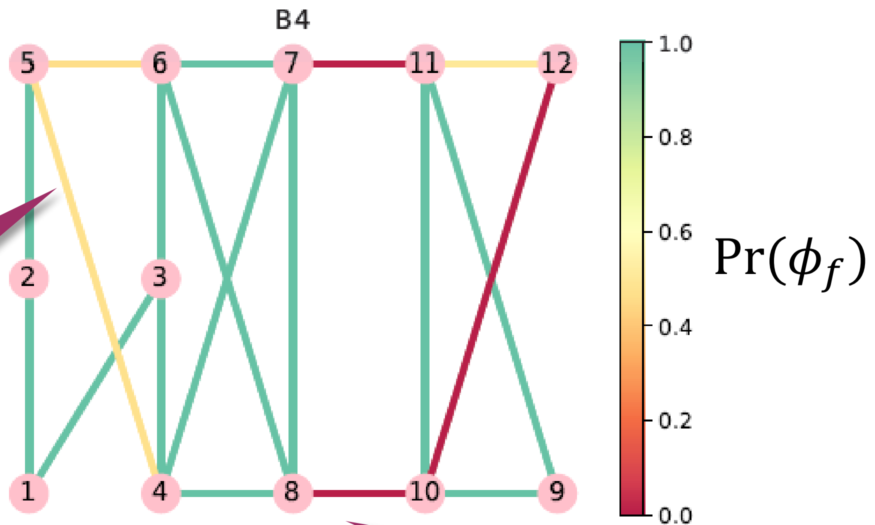
- Gravity Model

Failure model:

- Each link in the network fails independently at the probability of 0.001

Evaluation

A network's $\Pr(\phi_f)$ upon each single link's failure ($|f| = 1$).



A failure scenario that is not imperative

A definitely severe failure scenario

Pita quantifies the risk degrees of failure scenarios instead of only determining whether there could be a risk.

Evaluation

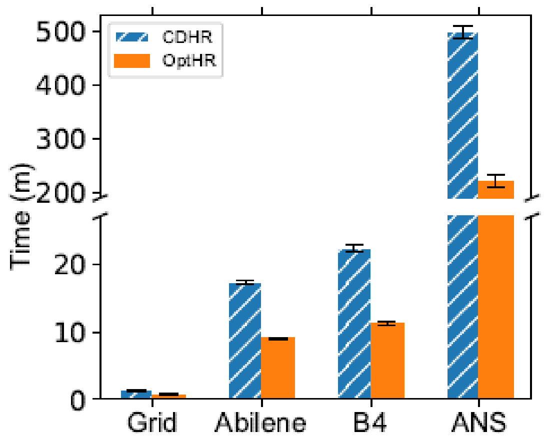
Network overall availability under a set of failure scenarios.

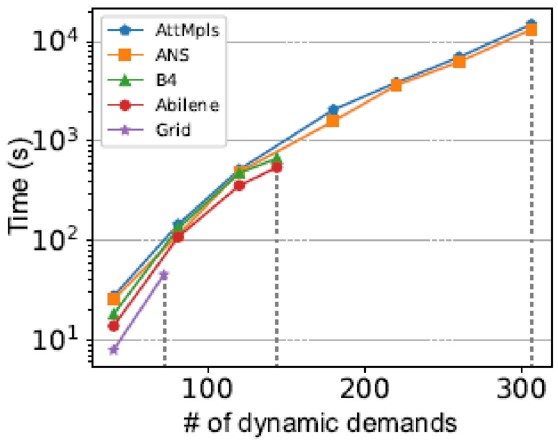
- $k \leq 1$: F includes all scenarios of at most 1 link failed.
- $k \leq 2$: F includes all scenarios of at most 2 link failed.

Network	KSP		MaxFlow		MinLatency	
	$k \leq 1$	$k \leq 2$	$k \leq 1$	$k \leq 2$	$k \leq 1$	$k \leq 2$
GridNet	99.833%	99.830%	99.786%	99.782%'	100%	100%
Abilene	99.900%	99.900%	100%	99.999%	100%	99.997%
B4	99.567%	99.561%	99.552%	99.548%	100%	99.996%
ANS	99.805%	99.800%	99.609%	99.601%	99.805%	99.802%

Evaluation

Pita's Running Time





OptHR reduces up to **55%** running time compared with SOTA random walk.

A few minutes for networks with ~100 demands

Summary

Pita: a probabilistic analysis framework for network availability.

Proof see paper

The problem of computing the availability is #P-hard. But the convexity of the problem makes it possible to be approximated by Multiphase MCMC.

A domain-specific optimization could make the SOTA random walk algorithm faster, which is theoretically proved and empirically validated.

Pita could probabilistically profile a network's availability with quantifying the overload-free probability for each failure scenario.

Thanks

yunmo.zhang@my.cityu.edu.hk