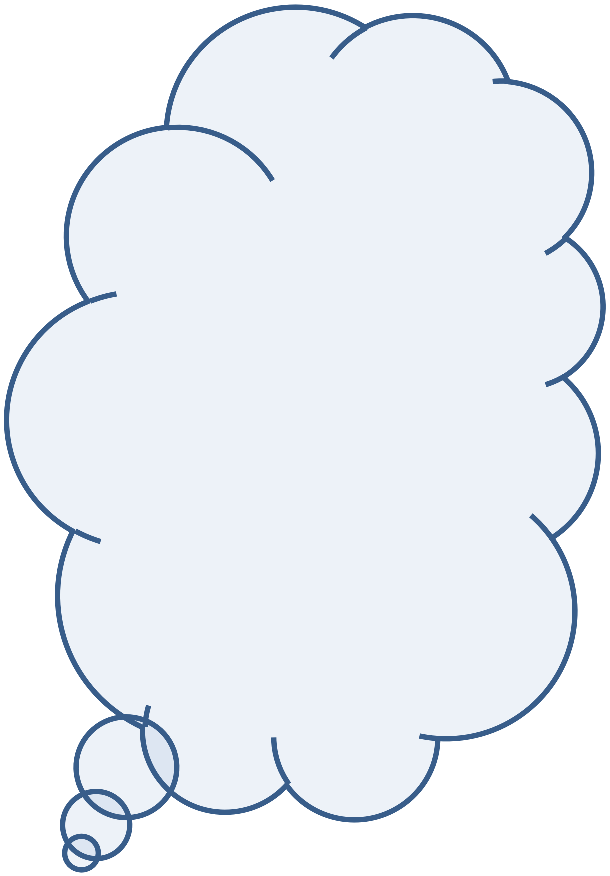
Shifting Network Tomography Toward A Practical Goal

Denisa Ghita, Can Karakus, Katerina Argyraki, Patrick Thiran

CoNext 2011, Tokyo, Japan

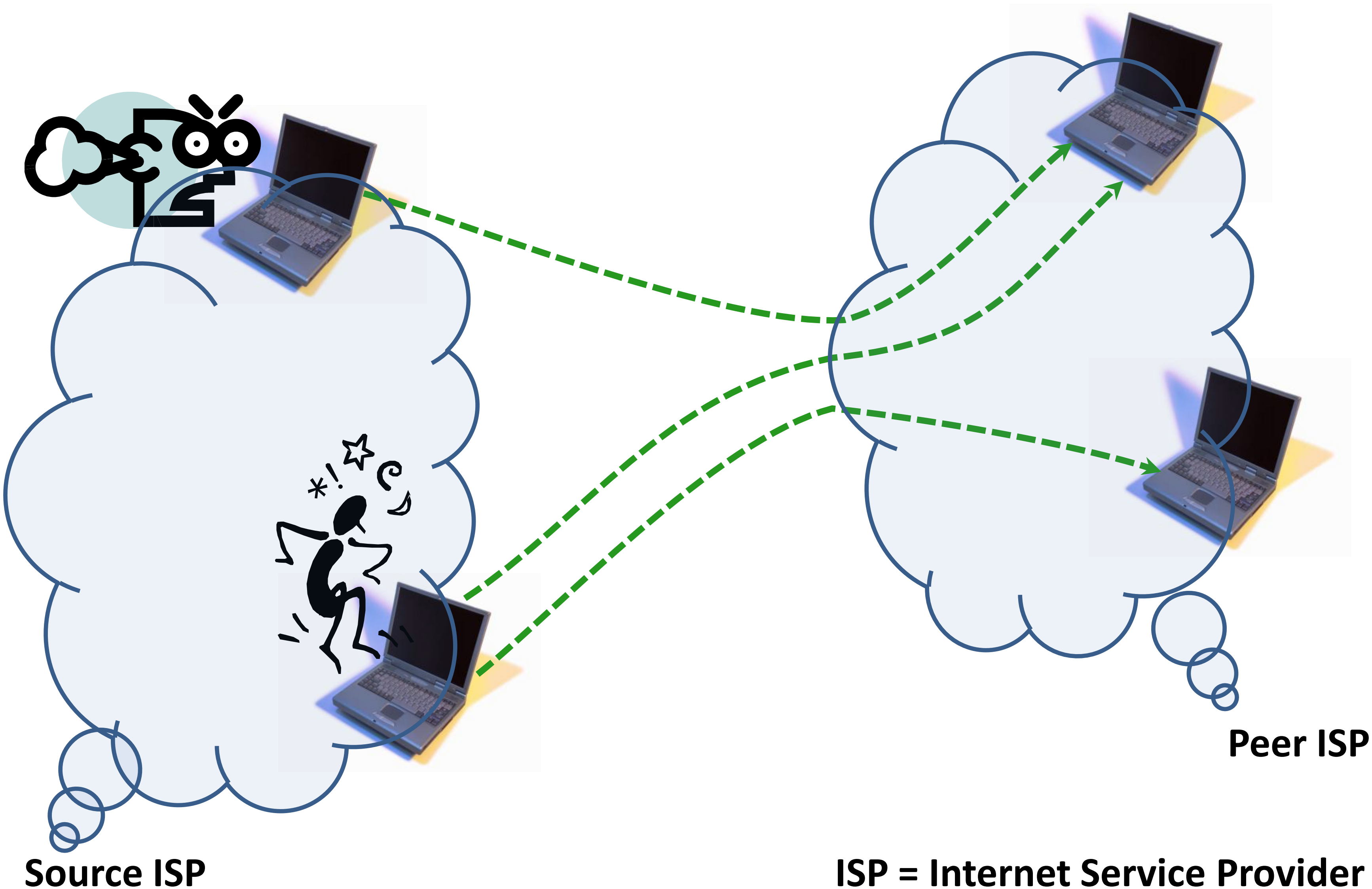
The ISP Curious About its Peers



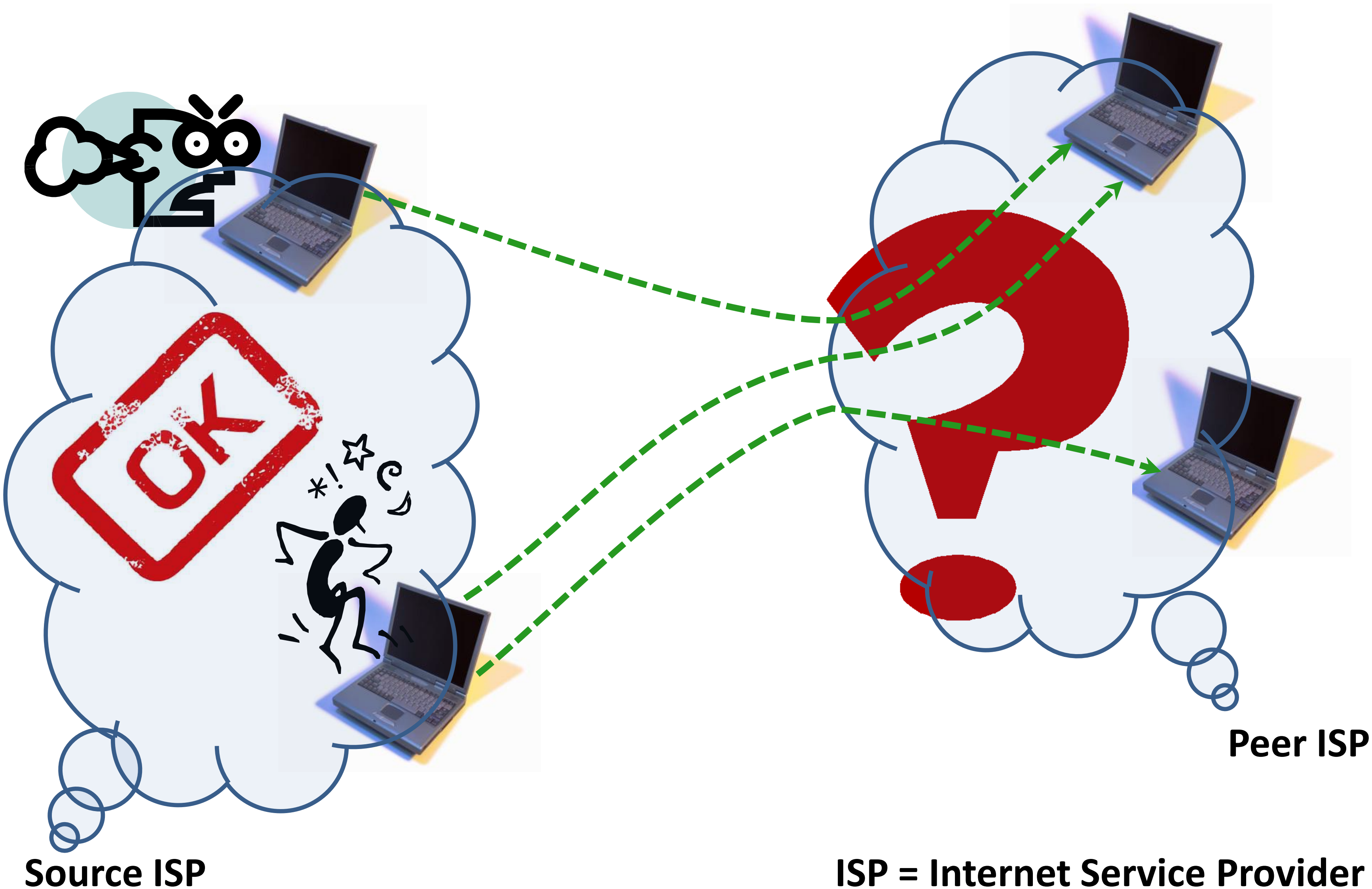
Source ISP

ISP = Internet Service Provider

The ISP Curious About its Peers

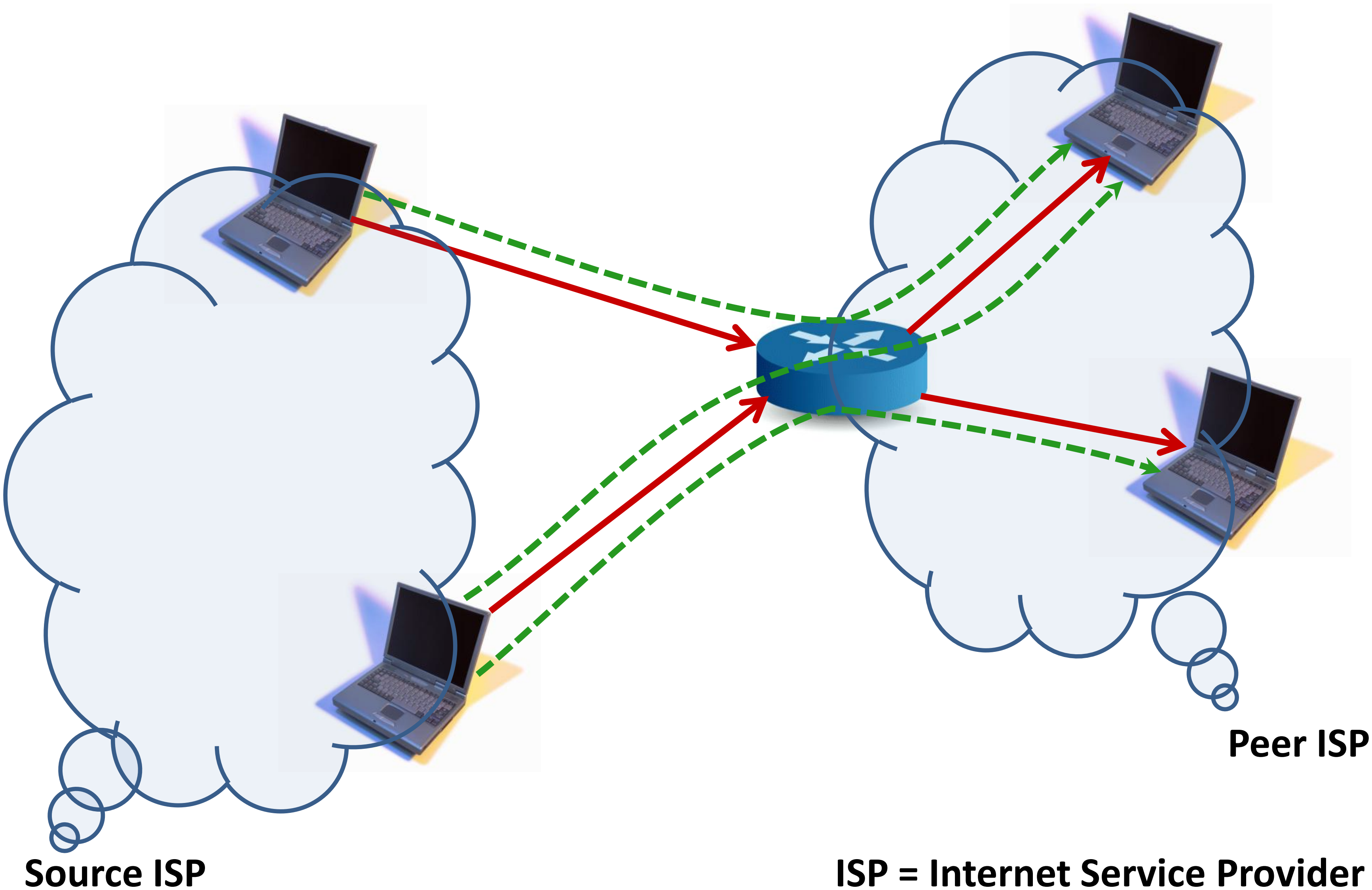


The ISP Curious About its Peers



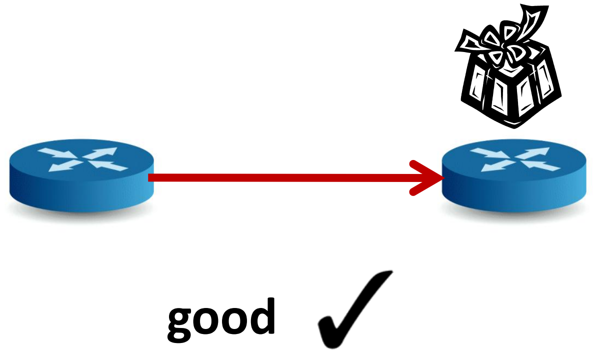
ISP = Internet Service Provider

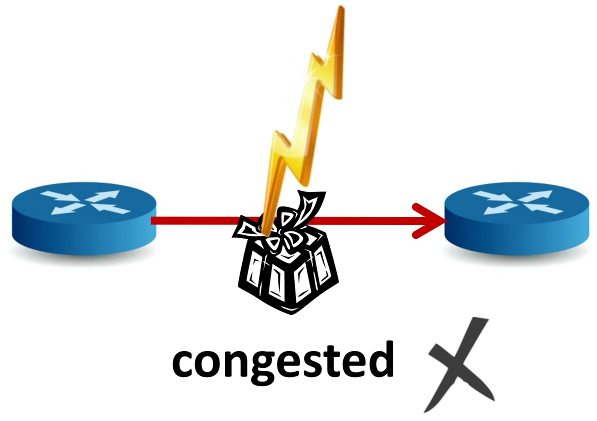
The ISP Curious About its Peers



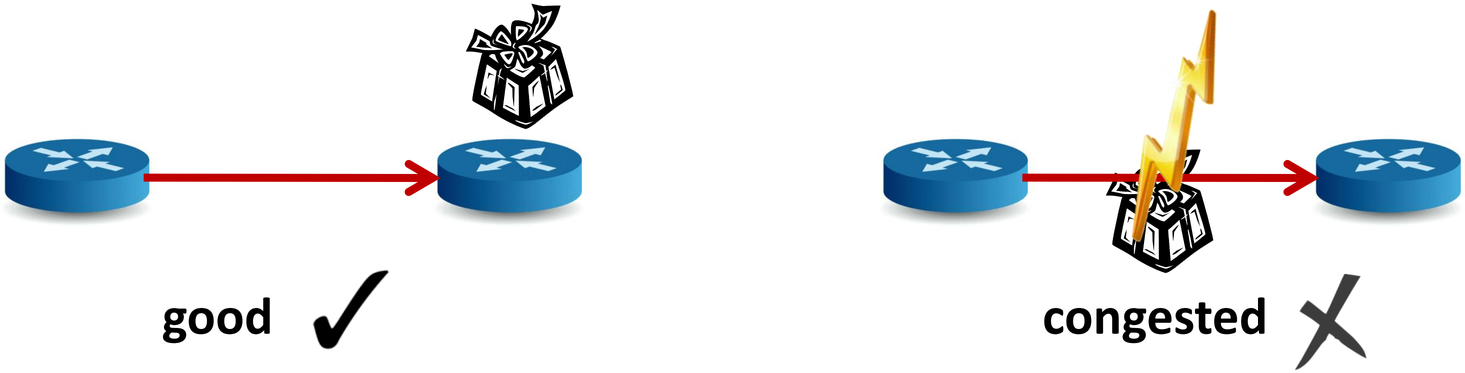
Boolean Inference

© N. Duffield '06



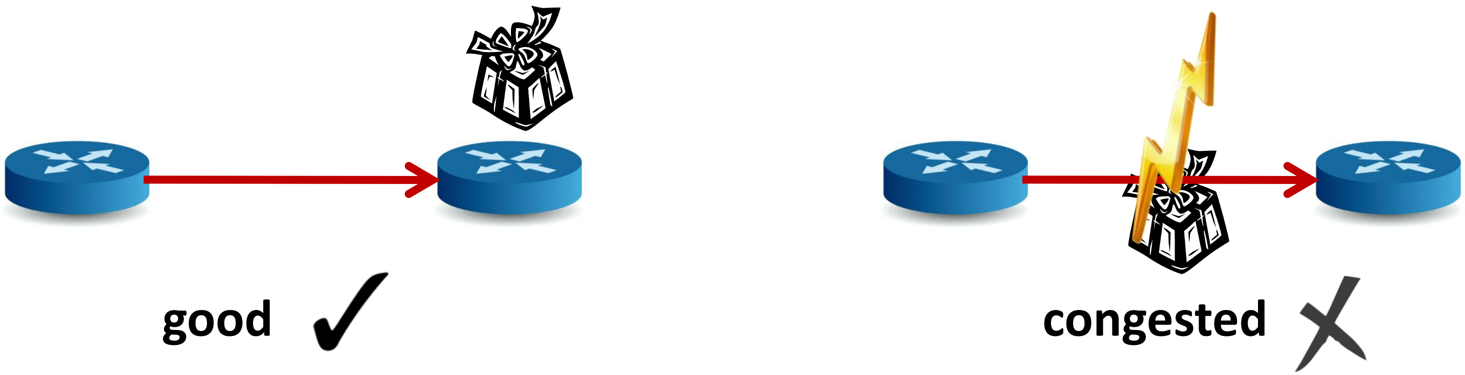


Boolean Inference

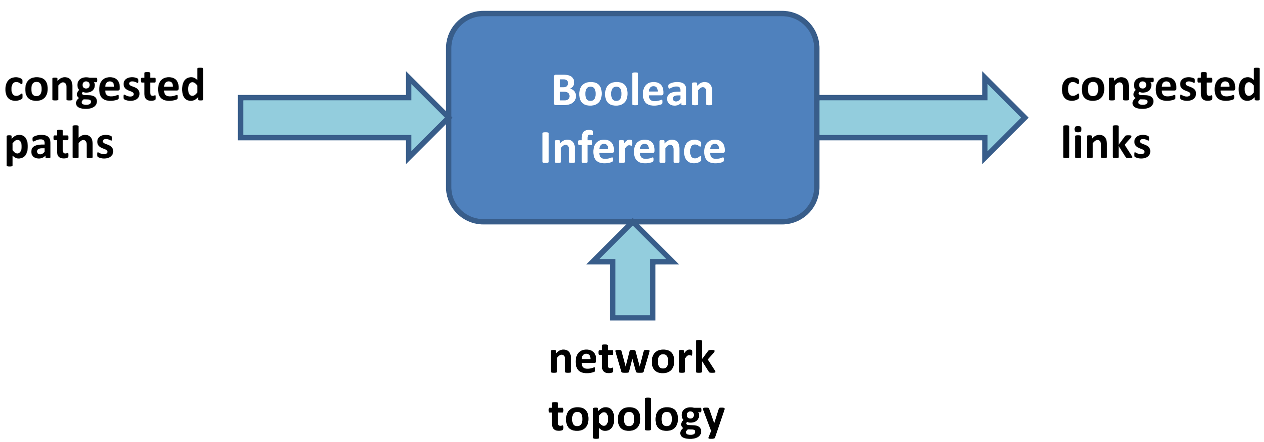


A path is good if and only if all its links are good.

Boolean Inference



A path is good if and only if all its links are good.

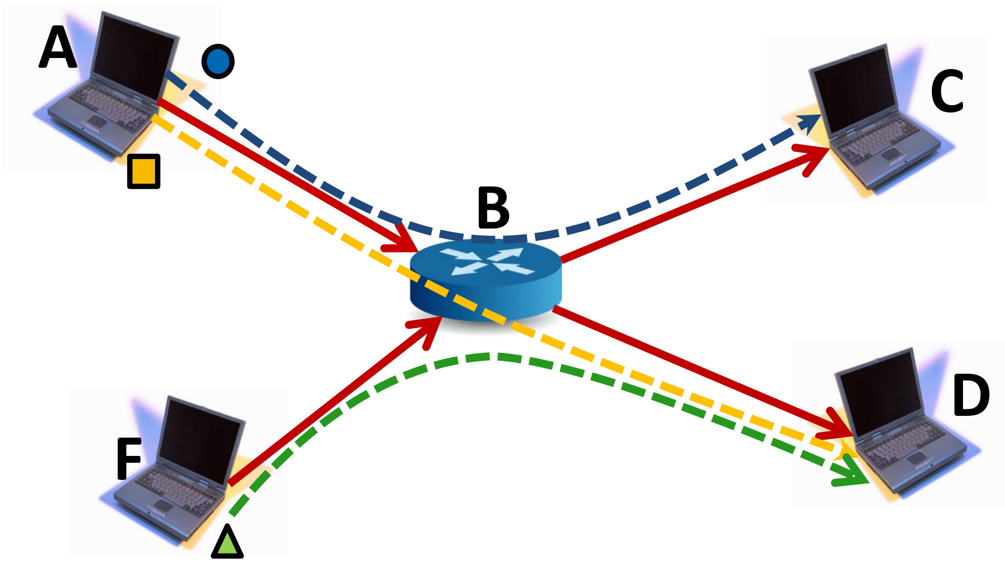


Contributions

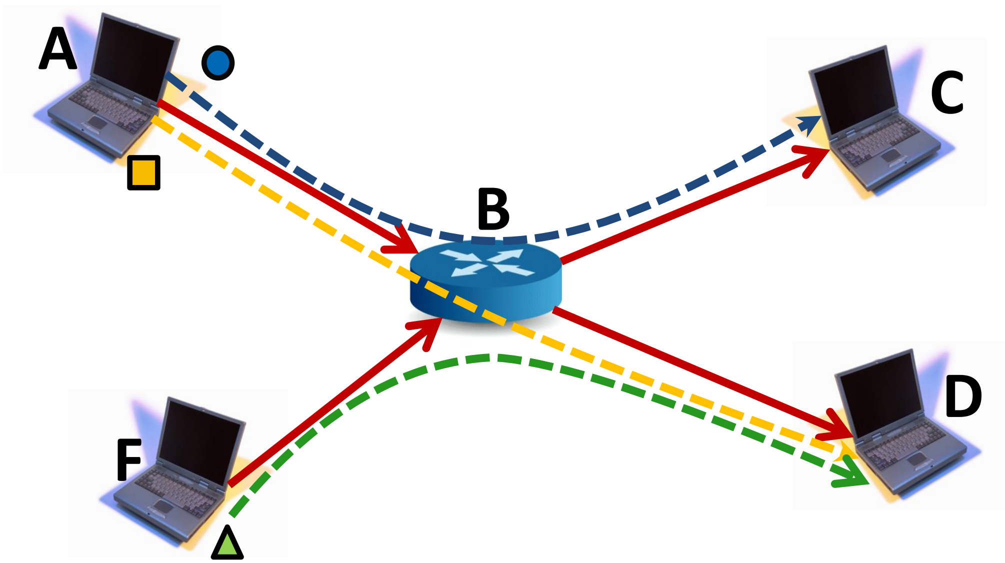
In the commercial ISP scenario, the information provided by Boolean Inference cannot be computed accurately.

We identify the "right" problem to solve in this scenario, which provides useful accurate information.

Challenge: Boolean Inference is Ill-Posed



Challenge: Boolean Inference is Ill-Posed



If all paths are congested,
then the congested links may be:

$$\{e_{AB}, e_{BD}\}$$

$$\{e_{AB}, e_{FB}\}$$

$$\{e_{BC}, e_{BD}\}$$

$$\{e_{AB}, e_{BC}, e_{BD}\}$$

$$\{e_{AB}, e_{BC}, e_{FB}\}$$

$$\{e_{AB}, e_{BD}, e_{FB}\}$$

$$\{e_{BC}, e_{BD}, e_{FB}\}$$

$$\{e_{AB}, e_{BC}, e_{BD}, e_{FB}\}$$

State-of-the-Art Boolean Inference Algorithms

Link Independence Assumption*:*
All links are independent.

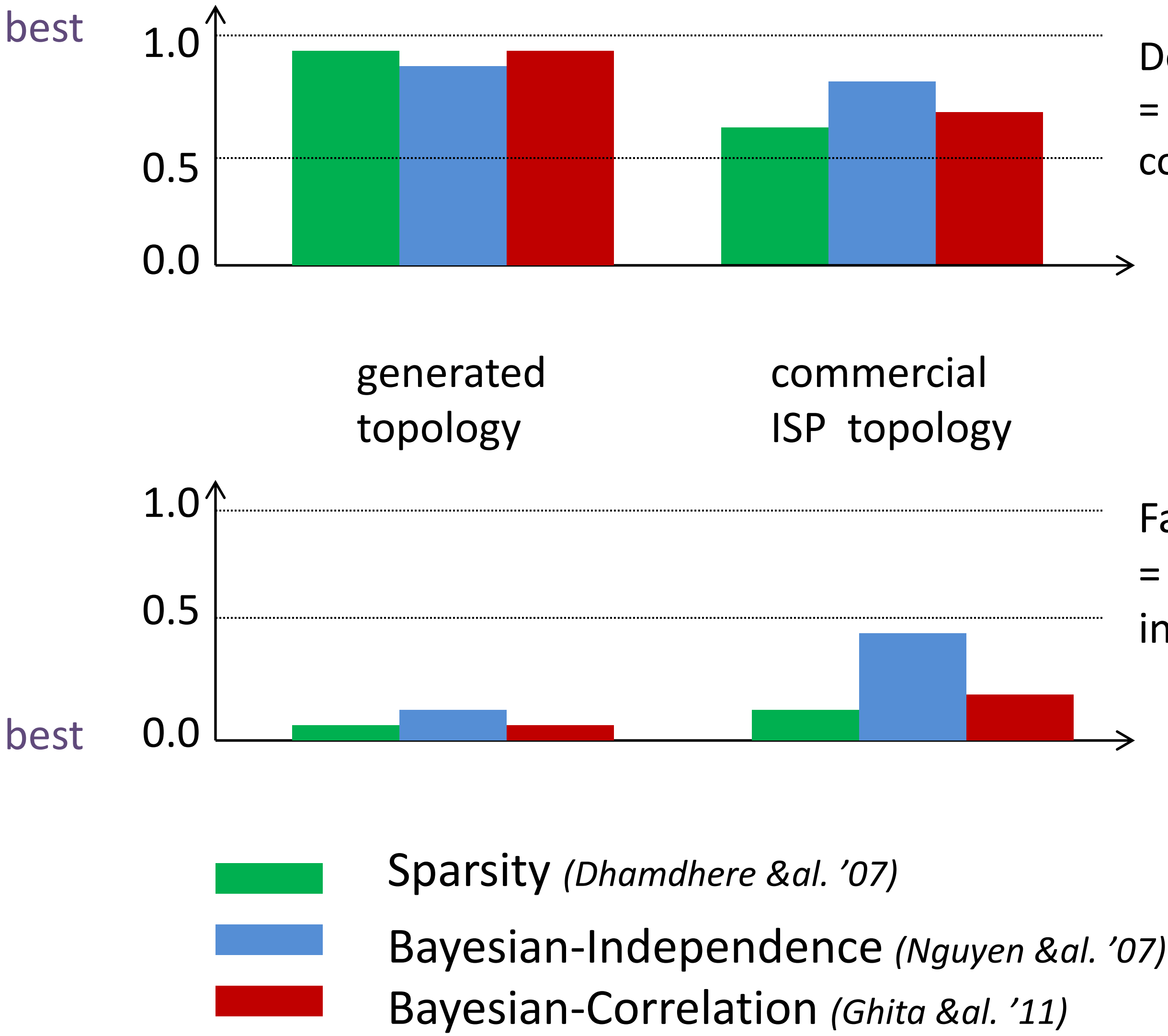
Link Homogeneity Assumption*:*
All links are equally likely to be congested.

Stationarity Assumption*:*
Stationary network dynamics.

State-of-the-Art Boolean Inference Algorithms

	Sparsity *Duffield ’06* *Dhamdhere &al. ’07*	Bayesian Independence *Nguyen &al. ‘07*	Bayesian Correlation *Ghita &al. ‘11*
Link Independence Assumption*:* *All links are independent.*	◆	◆	
Link Homogeneity Assumption*:* *All links are equally likely to be congested.*	◆		
Stationarity Assumption*:* *Stationary network dynamics.*		◆	◆

Boolean Inference
Not Accurate Enough in Our Scenario!



Detection Rate
= the fraction of congested links correctly identified as congested

False Positives Rate
= the fraction of links incorrectly identified as congested

Where's the Rub in Boolean Inference?

Our topology is sparser than generated topologies.

Paths that criss-cross yield more information.

The assumptions cannot be verified in practice.

Link Independence Assumption*: All links are independent.*

Link Homogeneity Assumption*: All links are equally likely to be congested.*

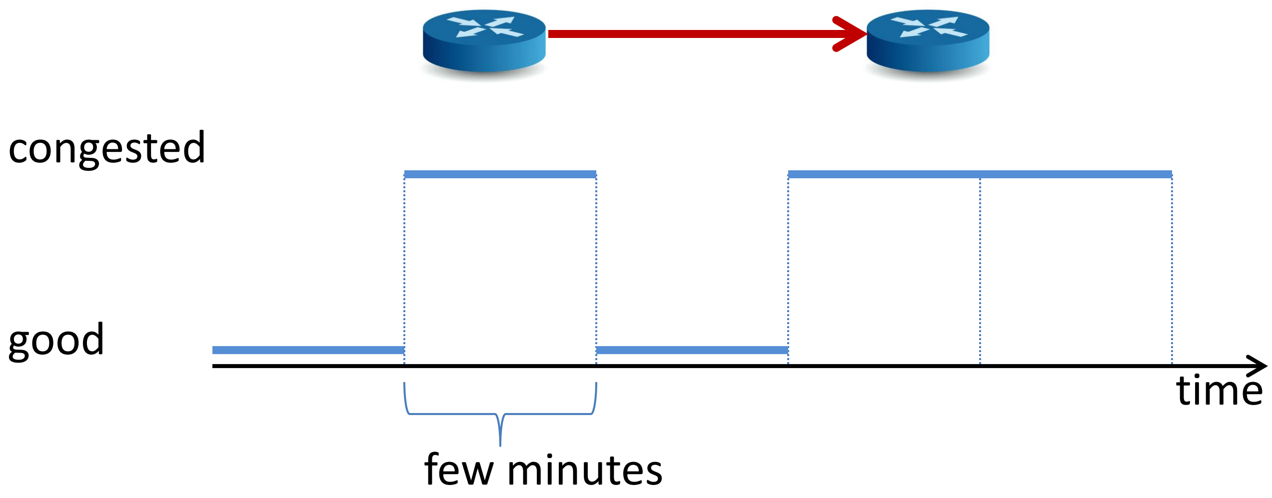
Stationarity Assumption*: Stationary network dynamics.*

Contributions

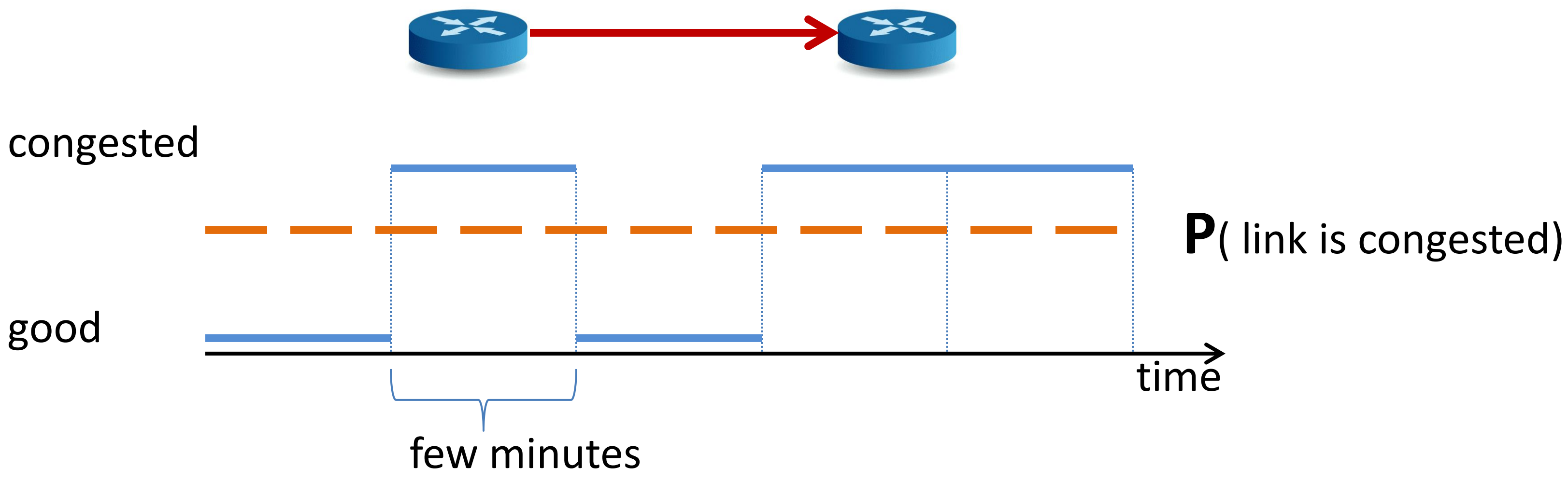
In the commercial ISP scenario, the information provided by Boolean Inference cannot be computed accurately.

We identify the “right” problem to solve in this scenario, which provides useful accurate information.

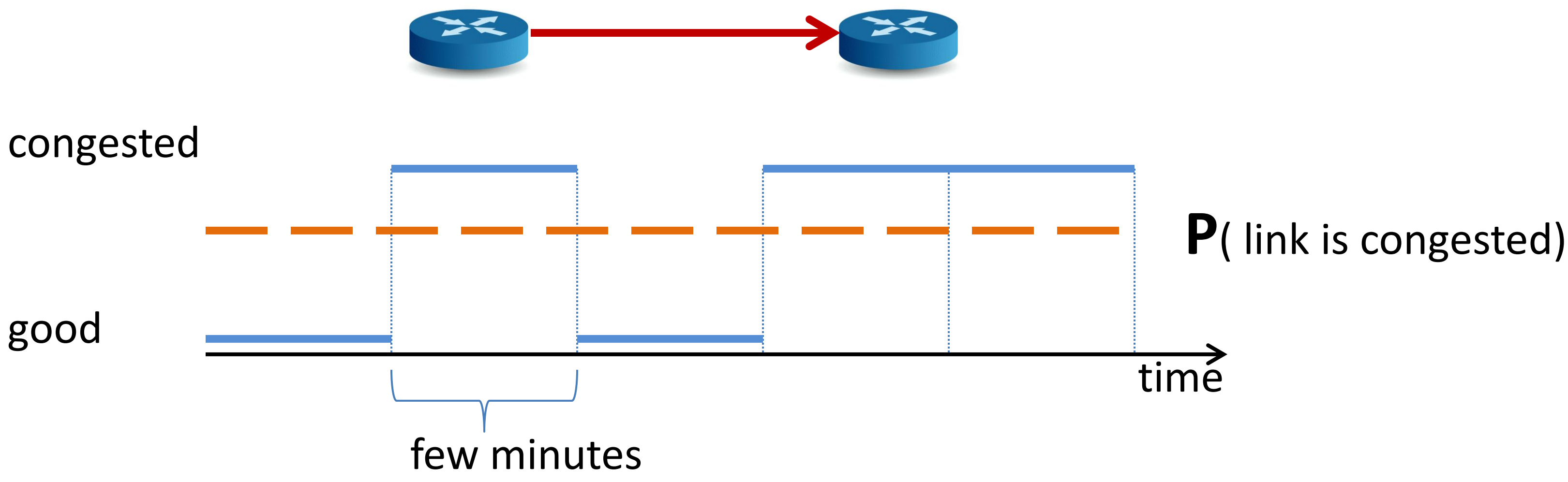
Inferring Congestion Frequency



Inferring Congestion Frequency



Inferring Congestion Frequency



congested paths → **Boolean Inference** → **congested links**

network topology

+1 congestion status of each link

+1 few minutes

not accurate enough (ill-posed, unverifiable assumptions)

congested paths → **Congestion Frequency** → **congestion probability of links**

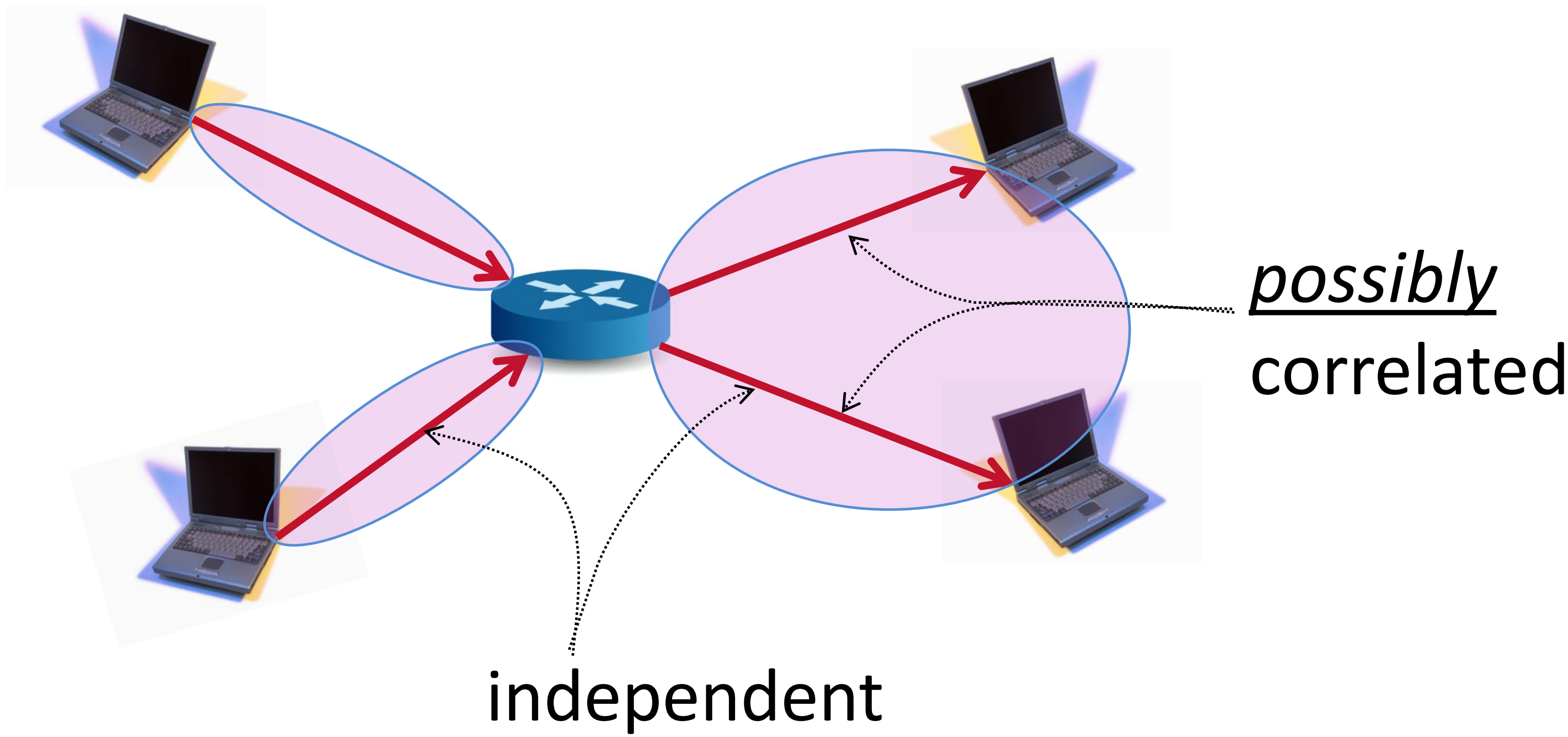
network topology

the probability that a set of links is congested

tens of minutes

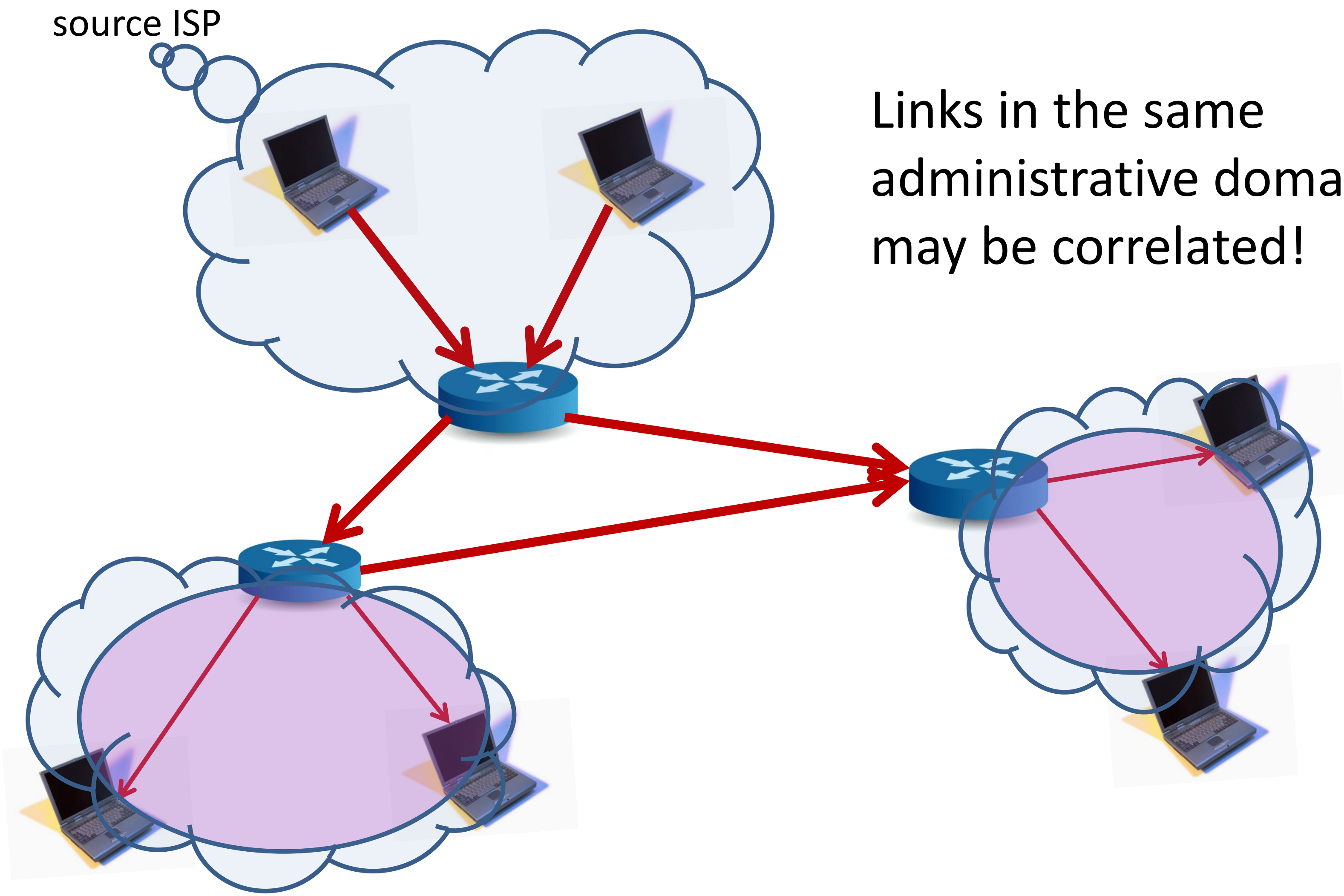
accurate (well-posed, realistic assumption) +∞

Our Assumption: Correlation Sets



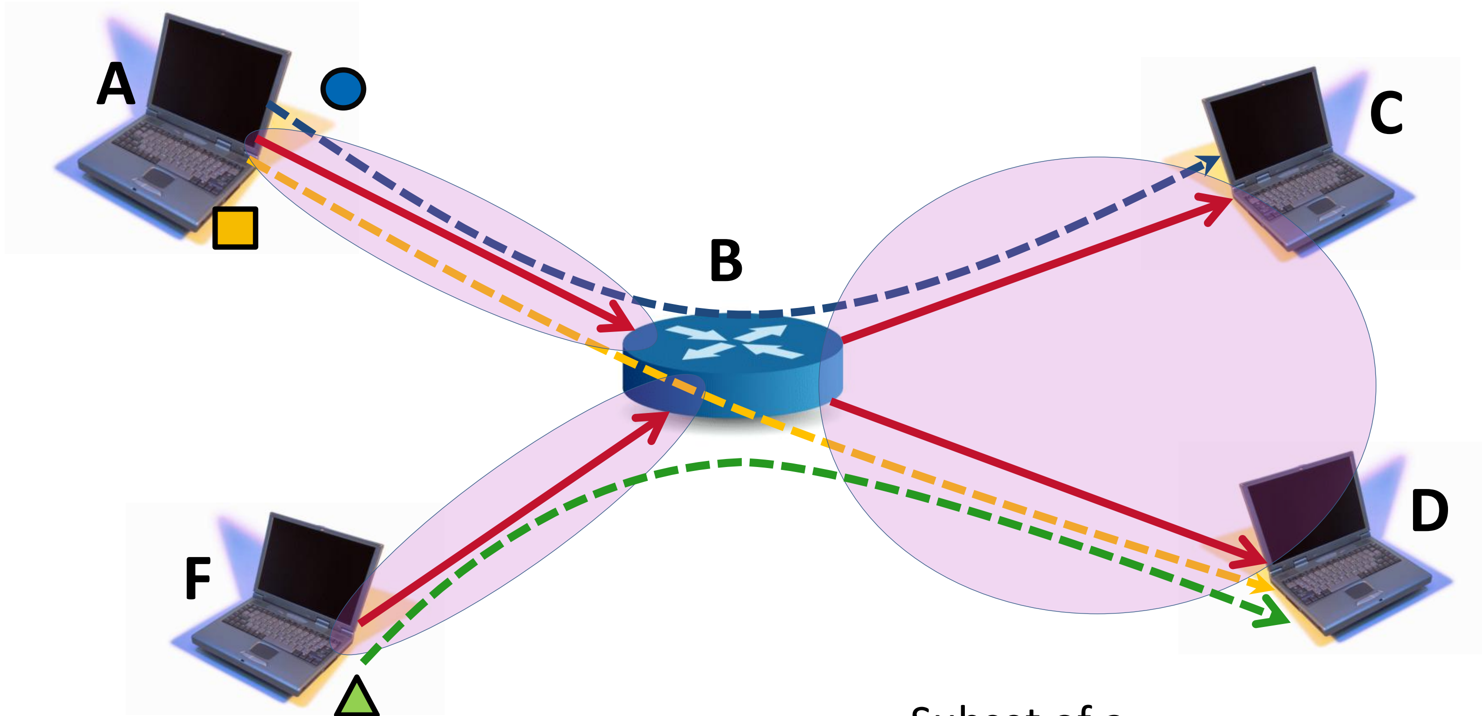
Independence among correlation sets.

How to Know the Correlation Sets?



Links in the same administrative domain may be correlated!

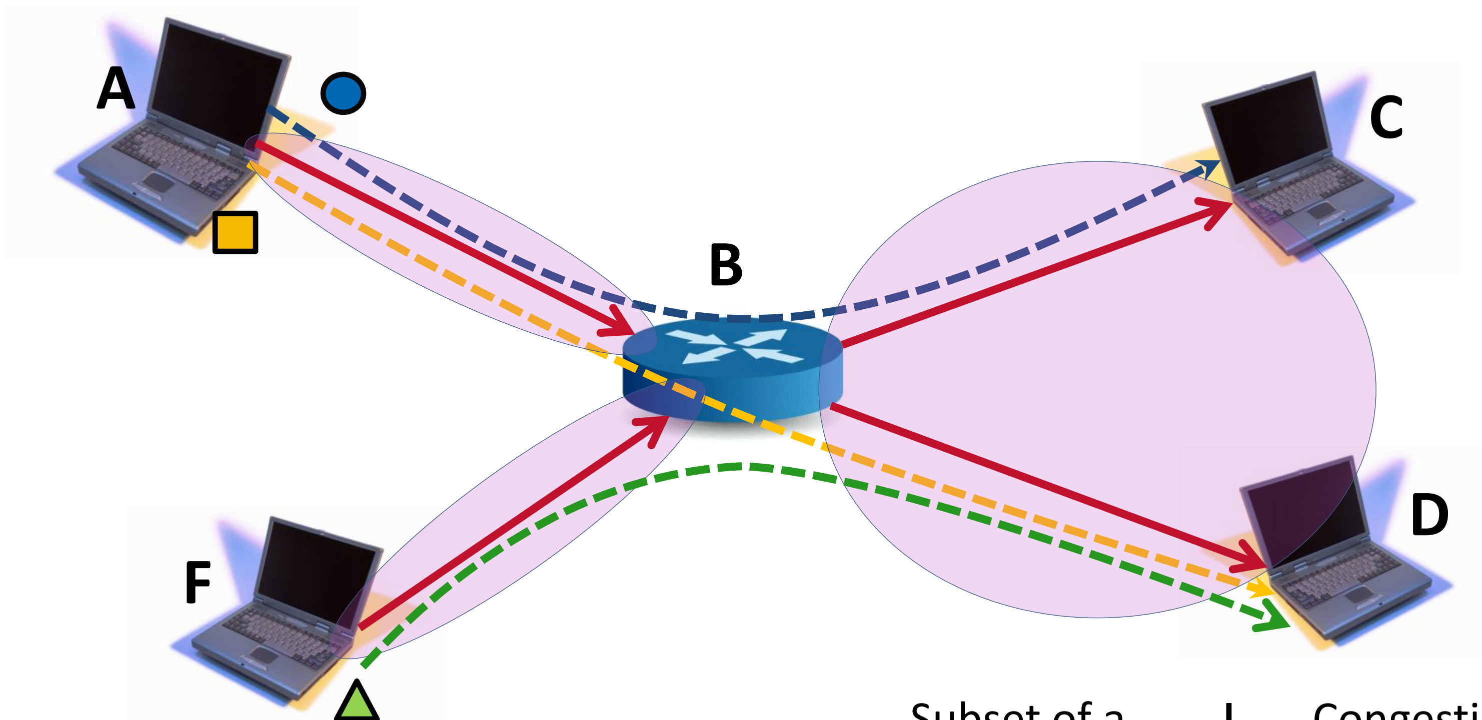
Is Congestion Frequency Well-Posed?



Each subset of a correlation set must be covered by a different set of paths!

Subset of a Correlation Set	Covered Paths
e_{AB}	● ■
e_{BC}	●
e_{BD}	■ ▲
e_{BC}, e_{BD}	● ■ ▲
e_{FB}	▲

Is Congestion Frequency Well-Posed?

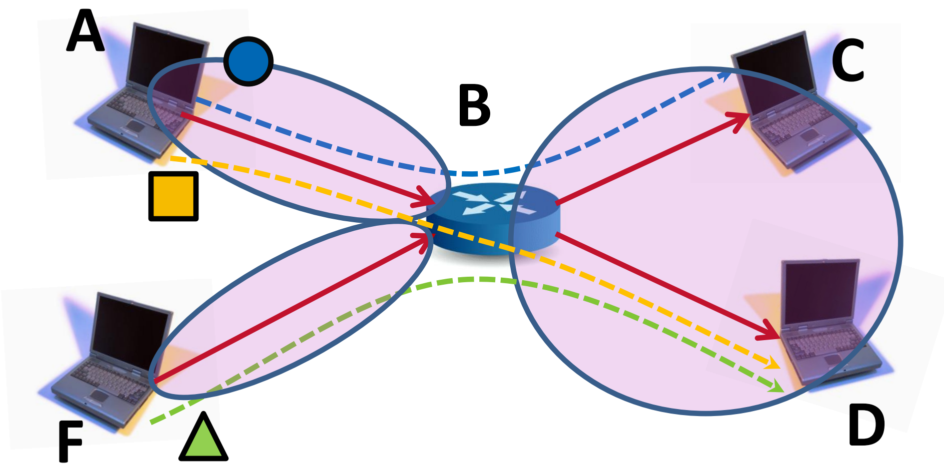


Each subset of a correlation set must be covered by a different set of paths!

Subset of a Correlation Set	Congestion Frequency
e_{AB}	$\mathbb{P}(e_{AB}\ congested)$
e_{BC}	$\mathbb{P}(e_{BC}\ congested)$
e_{BD}	$\mathbb{P}(e_{BD}\ congested)$
e_{BC}, e_{BD}	$\mathbb{P}(e_{BC}, e_{BD}\ congested)$
e_{FB}	$\mathbb{P}(e_{FB}\ congested)$

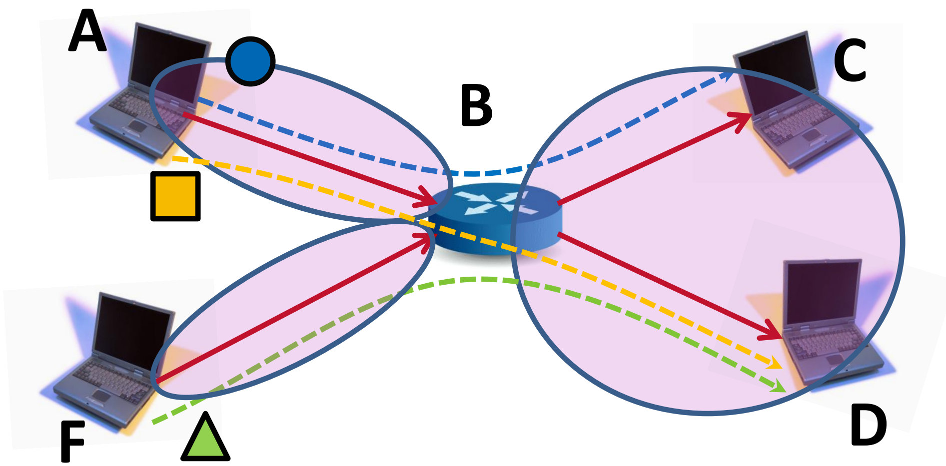
The Core Idea

Each set of paths generates an equation!



The Core Idea

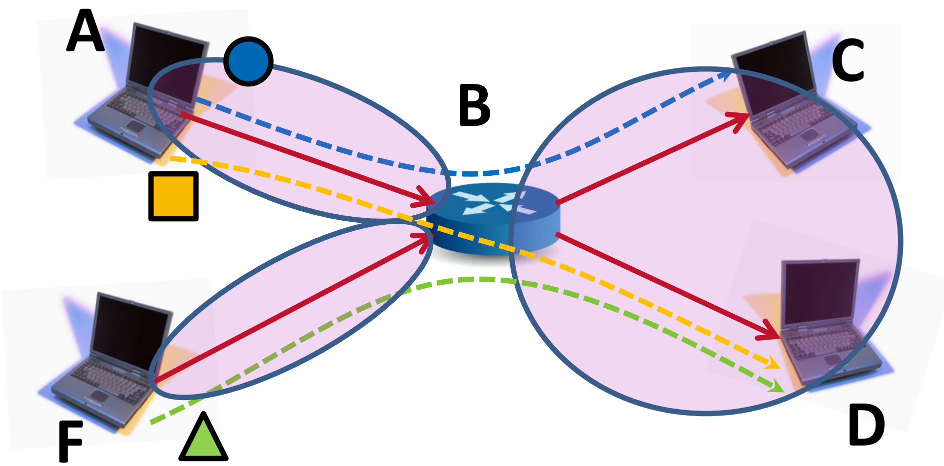
Each set of paths generates an equation!



 $\mathbb{P}(P_{AC}\ good) = \mathbb{P}(e_{AB}\ good)\mathbb{P}(e_{BC}\ good)$

The Core Idea

Each set of paths generates an equation!

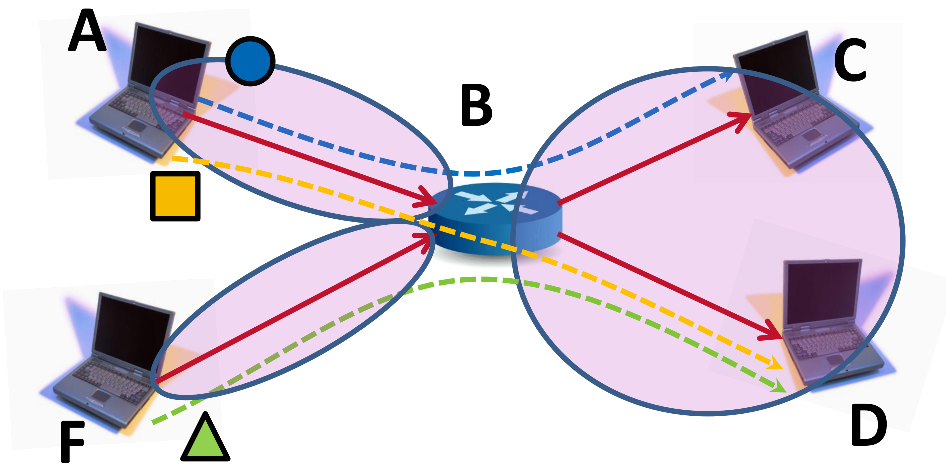


$$\mathbb{P}(P_{AC}\ good) = \mathbb{P}(e_{AB}\ good)\mathbb{P}(e_{BC}\ good)$$

$$\mathbb{P}(P_{AC}, P_{FD}\ good) = \mathbb{P}(e_{AB}\ good)\mathbb{P}(e_{FB}\ good)\mathbb{P}(e_{BD}, e_{BC}\ good)$$

The Core Idea

Each set of paths generates an equation!

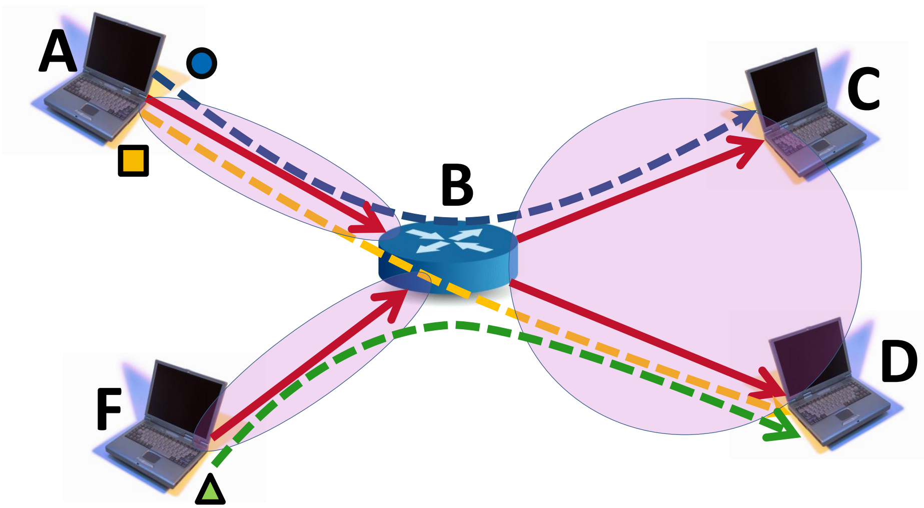


$$\mathbb{P}(P_{AC}\ good) = \mathbb{P}(e_{AB}\ good)\mathbb{P}(e_{BC}\ good)$$

$$\mathbb{P}(P_{AC}, P_{FD}\ good) = \mathbb{P}(e_{AB}\ good)\mathbb{P}(e_{FB}\ good)\mathbb{P}(e_{BD}, e_{BC}\ good)$$

$$= \mathbb{P}(P_{AC}, P_{AD}, P_{FD}\ good)$$

The Core Idea



● P_{AC}

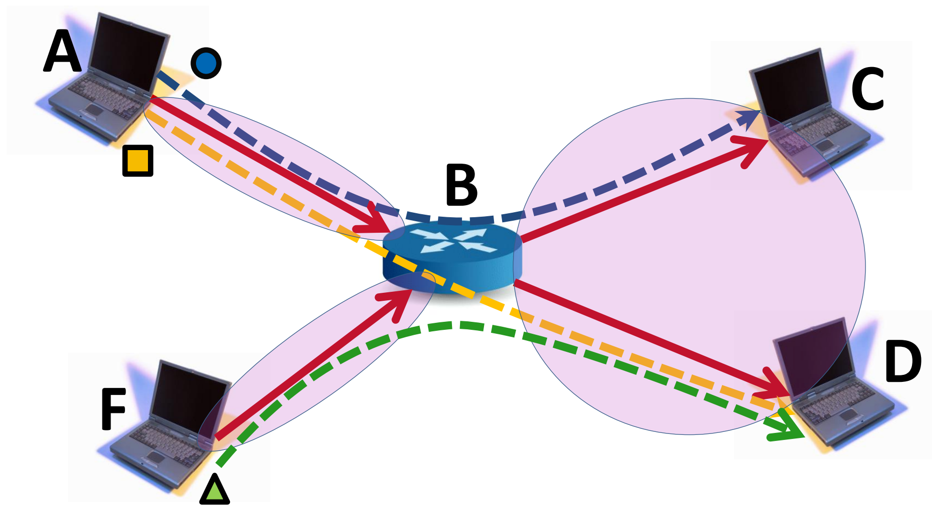
▲ P_{FD}

■ ● P_{AC}, P_{AD}

■ ▲ P_{AD}, P_{FD}

■ ● ▲ P_{AC}, P_{AD}, P_{FD}

The Core Idea



The system of equations has a unique solution!

$$\mathbb{P}(P_{AC}\ good) = \mathbb{P}(e_{AB}\ good)\mathbb{P}(e_{BC}\ good)$$

$$\mathbb{P}(P_{FD}\ good) = \mathbb{P}(e_{FB}\ good)\mathbb{P}(e_{BD}\ good)$$

$$\mathbb{P}(P_{AC}, P_{AD}\ good) = \mathbb{P}(e_{AB}\ good)\mathbb{P}(e_{BC}, e_{BD}\ good)$$

$$\mathbb{P}(P_{AD}, P_{FD}\ good) = \mathbb{P}(e_{AB}\ good)\mathbb{P}(e_{FB}\ good)\mathbb{P}(e_{BD}\ good)$$

$$\mathbb{P}(P_{AC}, P_{AD}, P_{FD}\ good) = \mathbb{P}(e_{AB}\ good)\mathbb{P}(e_{FB}\ good)\mathbb{P}(e_{BC}, e_{BD}\ good)$$

The Challenge

Combinatorial explosion: 1500 paths => 2^{1500} equations!

The Challenge

Combinatorial explosion: 1500 paths => 2^{1500} equations!

Our theorem identifies a large portion of the space that is redundant.

The Challenge

Combinatorial explosion: 1500 paths => 2^{1500} equations!

Our theorem identifies a large portion of the space that is redundant.

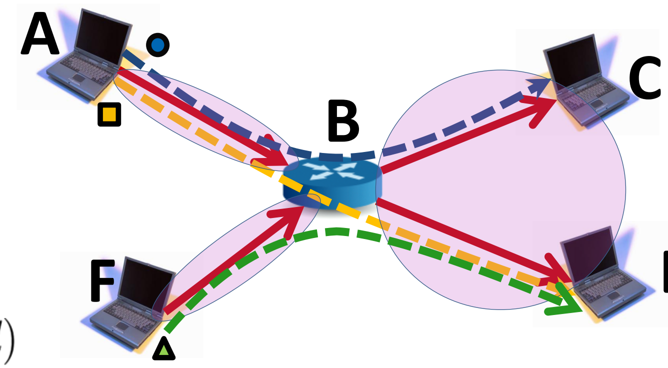
Our algorithm efficiently computes the probability that each set of links is congested.

Key Points of the Algorithm

Pick the sets of paths that are most likely to yield useful information (linearly independent equations).

Efficiently check if a set of paths yields useful information (a linearly independent equation).

How to Pick the Most Likely Sets of Paths



$$\mathbb{P}(P_{AC}\ good) = \mathbb{P}(e_{AB}\ good)\mathbb{P}(e_{BC}\ good)$$

$$\mathbb{P}(P_{FD}\ good) = \mathbb{P}(e_{FB}\ good)\mathbb{P}(e_{BD}\ good)$$

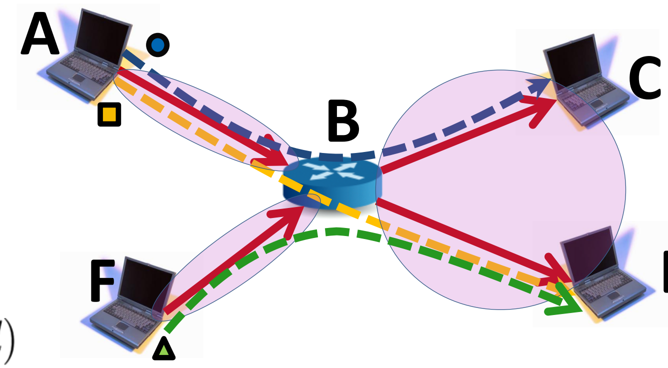
$$\mathbb{P}(P_{AC}, P_{AD}\ good) = \mathbb{P}(e_{AB}\ good)\mathbb{P}(e_{BC}, e_{BD}\ good)$$

$$\mathbb{P}(P_{AD}, P_{FD}\ good) = \mathbb{P}(e_{AB}\ good)\mathbb{P}(e_{FB}\ good)\mathbb{P}(e_{BD}\ good)$$

$$\mathbb{P}(P_{AC}, P_{AD}, P_{FD}\ good) = \mathbb{P}(e_{AB}\ good)\mathbb{P}(e_{FB}\ good)\mathbb{P}(e_{BC}, e_{BD}\ good)$$

The system of equations has an unique solution!

How to Pick the Most Likely Sets of Paths



$$\mathbb{P}(P_{AC}\ good) = \mathbb{P}(e_{AB}\ good)\mathbb{P}(e_{BC}\ good)$$

$$\mathbb{P}(P_{FD}\ good) = \mathbb{P}(e_{FB}\ good)\mathbb{P}(e_{BD}\ good)$$

$$\mathbb{P}(P_{AC}, P_{AD}\ good) = \mathbb{P}(e_{AB}\ good)\mathbb{P}(e_{BC}, e_{BD}\ good)$$

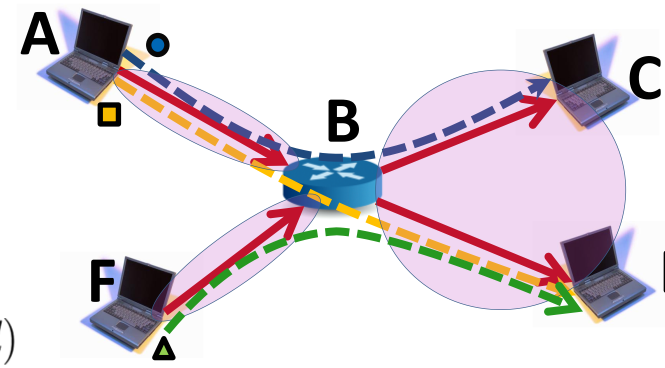
$$\mathbb{P}(P_{AD}, P_{FD}\ good) = \mathbb{P}(e_{AB}\ good)\mathbb{P}(e_{FB}\ good)\mathbb{P}(e_{BD}\ good)$$

$$\mathbb{P}(P_{AC}, P_{AD}, P_{FD}\ good) = \mathbb{P}(e_{AB}\ good)\mathbb{P}(e_{FB}\ good)\mathbb{P}(e_{BC}, e_{BD}\ good)$$

The system of equations has an unique solution!

	e_{AB}	e_{FB}	e_{BC}	e_{BD}	e_{BC}, e_{BD}
P_{AC}			1	0	
P_{FD}					
P_{AC}, P_{AD}					
P_{AD}, P_{FD}					
P_{AC}, P_{AD}, P_{FD}					

How to Pick the Most Likely Sets of Paths



● $$\mathbb{P}(P_{AC}\ good) = \mathbb{P}(e_{AB}\ good)\mathbb{P}(e_{BC}\ good)$$

▲ $$\mathbb{P}(P_{FD}\ good) = \mathbb{P}(e_{FB}\ good)\mathbb{P}(e_{BD}\ good)$$

■ ● $$\mathbb{P}(P_{AC}, P_{AD}\ good) = \mathbb{P}(e_{AB}\ good)\mathbb{P}(e_{BC}, e_{BD}\ good)$$

■ ▲ $$\mathbb{P}(P_{AD}, P_{FD}\ good) = \mathbb{P}(e_{AB}\ good)\mathbb{P}(e_{FB}\ good)\mathbb{P}(e_{BD}\ good)$$

■ ● ▲ $$\mathbb{P}(P_{AC}, P_{AD}, P_{FD}\ good) = \mathbb{P}(e_{AB}\ good)\mathbb{P}(e_{FB}\ good)\mathbb{P}(e_{BC}, e_{BD}\ good)$$

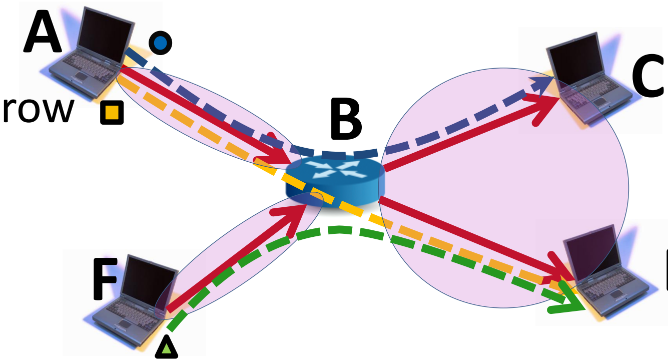
The system of equations has an unique solution!

		e_{AB}	e_{FB}	e_{BC}	e_{BD}	e_{BC}, e_{BD}
●	P_{AC}	1	0	1	0	0
▲	P_{FD}	0	1	0	1	0
■ ●	P_{AC}, P_{AD}	1	0	0	0	1
■ ▲	P_{AD}, P_{FD}	1	1	0	1	0
■ ● ▲	P_{AC}, P_{AD}, P_{FD}	1	1	0	0	1

The matrix has full column rank!

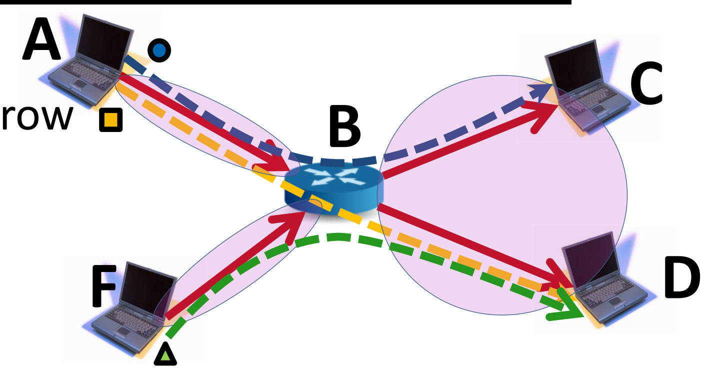
How to Pick the Most Likely Sets of Paths

1. make sure each unknown appears in at least one row



e_{AB}	e_{FB}	e_{BC}	e_{BD}	e_{BC}, e_{BD}
1				

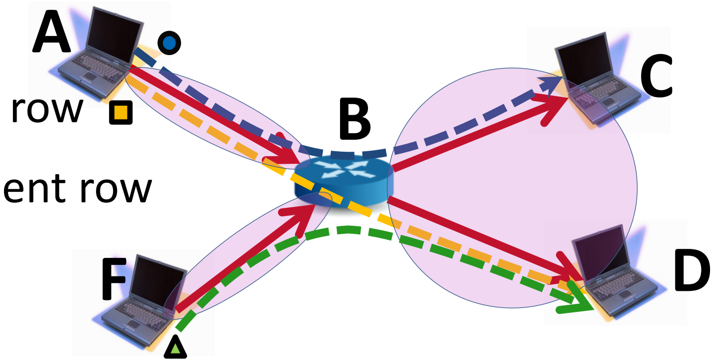
How to Pick the Most Likely Sets of Paths



1. make sure each unknown appears in at least one row

	e_{AB}	e_{FB}	e_{BC}	e_{BD}	e_{BC}, e_{BD}
P_{AC}	**1**	**0**	**1**	**0**	**0**

How to Pick the Most Likely Sets of Paths



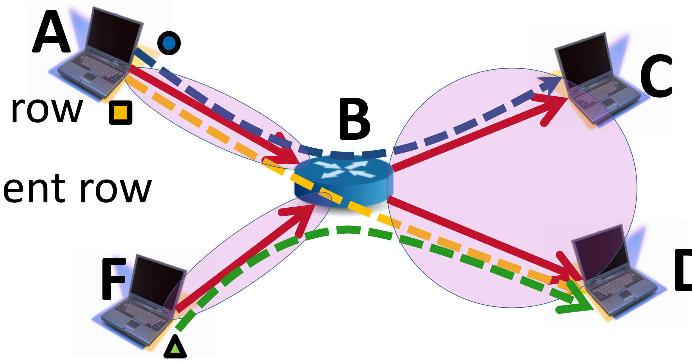
1. make sure each unknown appears in at least one row
2. augment the matrix with a new linearly independent row

		e_{AB}	e_{FB}	e_{BC}	e_{BD}	e_{BC}, e_{BD}
●	P_{AC}	1	0	1	0	0
▲	P_{FD}	0	1	0	1	0
■ ●	P_{AC}, P_{AD}	1	0	0	0	1

$= \mathcal{R}$

How to Pick the Most Likely Sets of Paths

1. make sure each unknown appears in at least one row
2. augment the matrix with a new linearly independent row

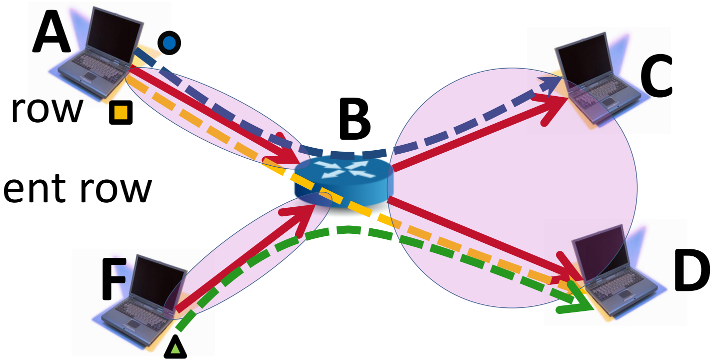


		e_{AB}	e_{FB}	e_{BC}	e_{BD}	e_{BC}, e_{BD}
●	P_{AC}	1	0	1	0	0
▲	P_{FD}	0	1	0	1	0
■ ●	P_{AC}, P_{AD}	1	0	0	0	1

$= \mathcal{R}$

path set $\mathcal{P}$ s.t. $Row(\mathcal{P})$ is linearly independent with $\mathcal{R}$

How to Pick the Most Likely Sets of Paths



1. make sure each unknown appears in at least one row

2. augment the matrix with a new linearly independent row

		e_{AB}	e_{FB}	e_{BC}	e_{BD}	e_{BC}, e_{BD}
●	P_{AC}	1	0	1	0	0
△	P_{FD}	0	1	0	1	0
■ ●	P_{AC}, P_{AD}	1	0	0	0	1

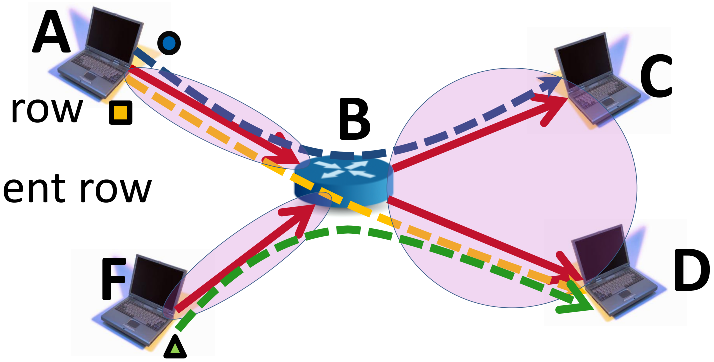
$= \mathcal{R}$

path set $\mathcal{P}$ s.t. $Row(\mathcal{P})$ is linearly independent with $\mathcal{R}$

$$Row(\mathcal{P}) \times Nullspace(\mathcal{R}) \neq 0$$

How to Pick the Most Likely Sets of Paths

1. make sure each unknown appears in at least one row
2. augment the matrix with a new linearly independent row



		e_{AB}	e_{FB}	e_{BC}	e_{BD}	e_{BC}, e_{BD}
●	P_{AC}	1	0	1	0	0
▲	P_{FD}	0	1	0	1	0
■ ●	P_{AC}, P_{AD}	1	0	0	0	1

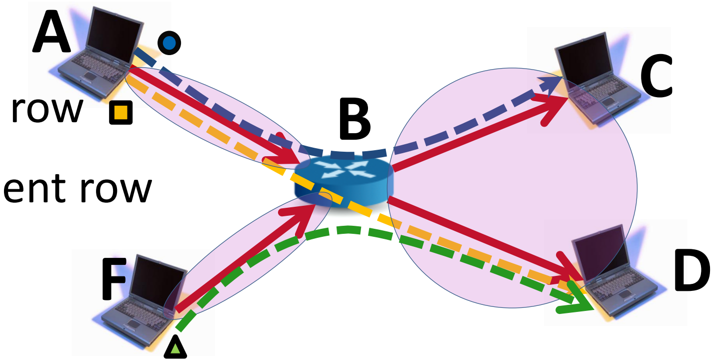
$= \mathcal{R}$

path set $\mathcal{P}$ s.t. $Row(\mathcal{P})$ is linearly independent with $\mathcal{R}$

$$Row(\mathcal{P}) \times Nullspace(\mathcal{R}) \neq 0$$

$$\begin{bmatrix} 0.0 & -0.4 \\ -0.7 & 0.0 \\ 0.0 & -0.7 \\ 4.0 & -4.3 \\ 0.2 & 2.5 \end{bmatrix} \begin{matrix} e_{AB} \\ e_{FB} \\ e_{BC} \\ e_{BD} \\ e_{BC}, e_{BD} \end{matrix}$$

How to Pick the Most Likely Sets of Paths



1. make sure each unknown appears in at least one row
2. augment the matrix with a new linearly independent row

		e_{AB}	e_{FB}	e_{BC}	e_{BD}	e_{BC}, e_{BD}
●	P_{AC}	1	0	1	0	0
▲	P_{FD}	0	1	0	1	0
■●	P_{AC}, P_{AD}	1	0	0	0	1

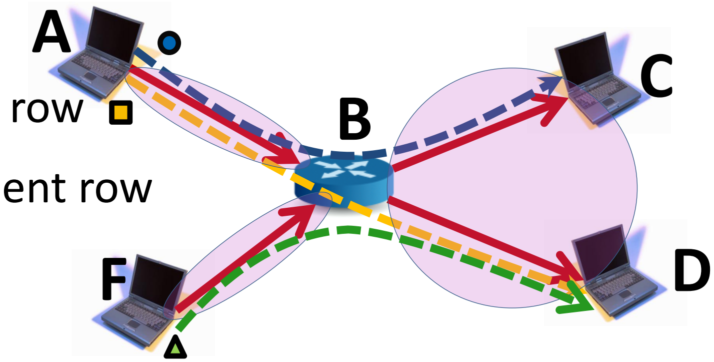
$= \mathcal{R}$

path set $\mathcal{P}$ s.t. $Row(\mathcal{P})$ is linearly independent with $\mathcal{R}$

$$Row(\mathcal{P}) \times Nullspace(\mathcal{R}) \neq 0$$

0.0	−0.4	e_{AB}
−0.7	0.0	e_{FB}
0.0	−0.7	e_{BC}
4.0	−4.3	e_{BD}
0.2	2.5	e_{BC}, e_{BD}

How to Pick the Most Likely Sets of Paths



1. make sure each unknown appears in at least one row
2. augment the matrix with a new linearly independent row

		e_{AB}	e_{FB}	e_{BC}	e_{BD}	e_{BC}, e_{BD}
●	P_{AC}	1	0	1	0	0
△	P_{FD}	0	1	0	1	0
■●	P_{AC}, P_{AD}	1	0	0	0	1

$= \mathcal{R}$

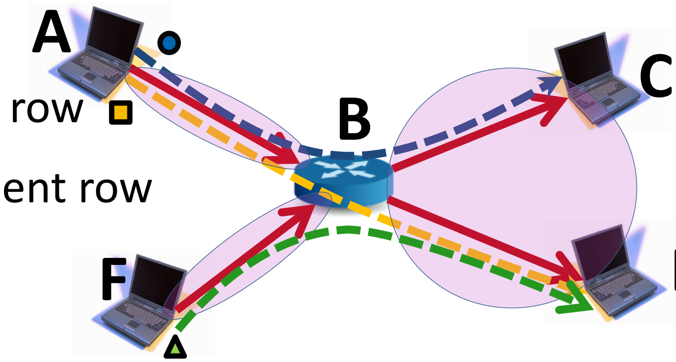
path set $\mathcal{P}$ s.t. $Row(\mathcal{P})$ is linearly independent with $\mathcal{R}$

$$Row(\mathcal{P}) \times Nullspace(\mathcal{R}) \neq 0$$

$$\begin{bmatrix} ? & ? & ? & 1 & ? \end{bmatrix} \times \begin{bmatrix} 0.0 & -0.4 \\ -0.7 & 0.0 \\ 0.0 & -0.7 \\ 4.0 & -4.3 \\ 0.2 & 2.5 \end{bmatrix} \begin{matrix} e_{AB} \\ e_{FB} \\ e_{BC} \\ e_{BD} \\ e_{BC}, e_{BD} \end{matrix}$$

How to Pick the Most Likely Sets of Paths

1. make sure each unknown appears in at least one row
2. augment the matrix with a new linearly independent row



	e_{AB}	e_{FB}	e_{BC}	e_{BD}	e_{BC}, e_{BD}
● P_{AC}	1	0	1	0	0
▲ P_{FD}	0	1	0	1	0
■● P_{AC}, P_{AD}	1	0	0	0	1

$= \mathcal{R}$

path set $\mathcal{P}$ s.t. $Row(\mathcal{P})$ is linearly independent with $\mathcal{R}$

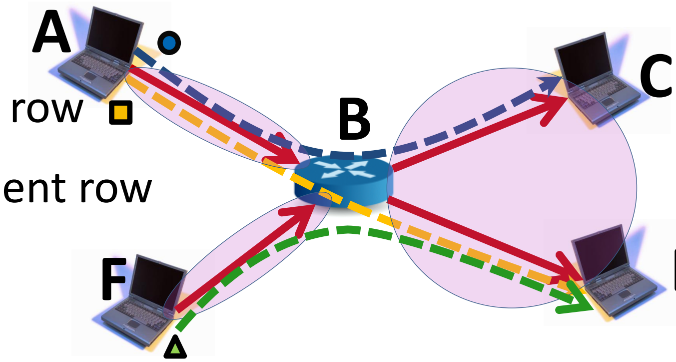
$$Row(\mathcal{P}) \times Nullspace(\mathcal{R}) \neq 0$$

$$\begin{bmatrix} ? & ? & ? & 1 & ? \end{bmatrix} \times \begin{bmatrix} 0.0 & -0.4 \\ -0.7 & 0.0 \\ 0.0 & -0.7 \\ 4.0 & -4.3 \\ 0.2 & 2.5 \end{bmatrix} \begin{matrix} e_{AB} \\ e_{FB} \\ e_{BC} \\ e_{BD} \\ e_{BC}, e_{BD} \end{matrix}$$

Path set $\mathcal{P}$ must cover link e_{BD}

How to Pick the Most Likely Sets of Paths

1. make sure each unknown appears in at least one row
2. augment the matrix with a new linearly independent row



	e_{AB}	e_{FB}	e_{BC}	e_{BD}	e_{BC}, e_{BD}
● P_{AC}	1	0	1	0	0
△ P_{FD}	0	1	0	1	0
□● P_{AC}, P_{AD}	1	0	0	0	1
□△ $\mathcal{P} = P_{AD}, P_{FD}$	1	1	0	1	0

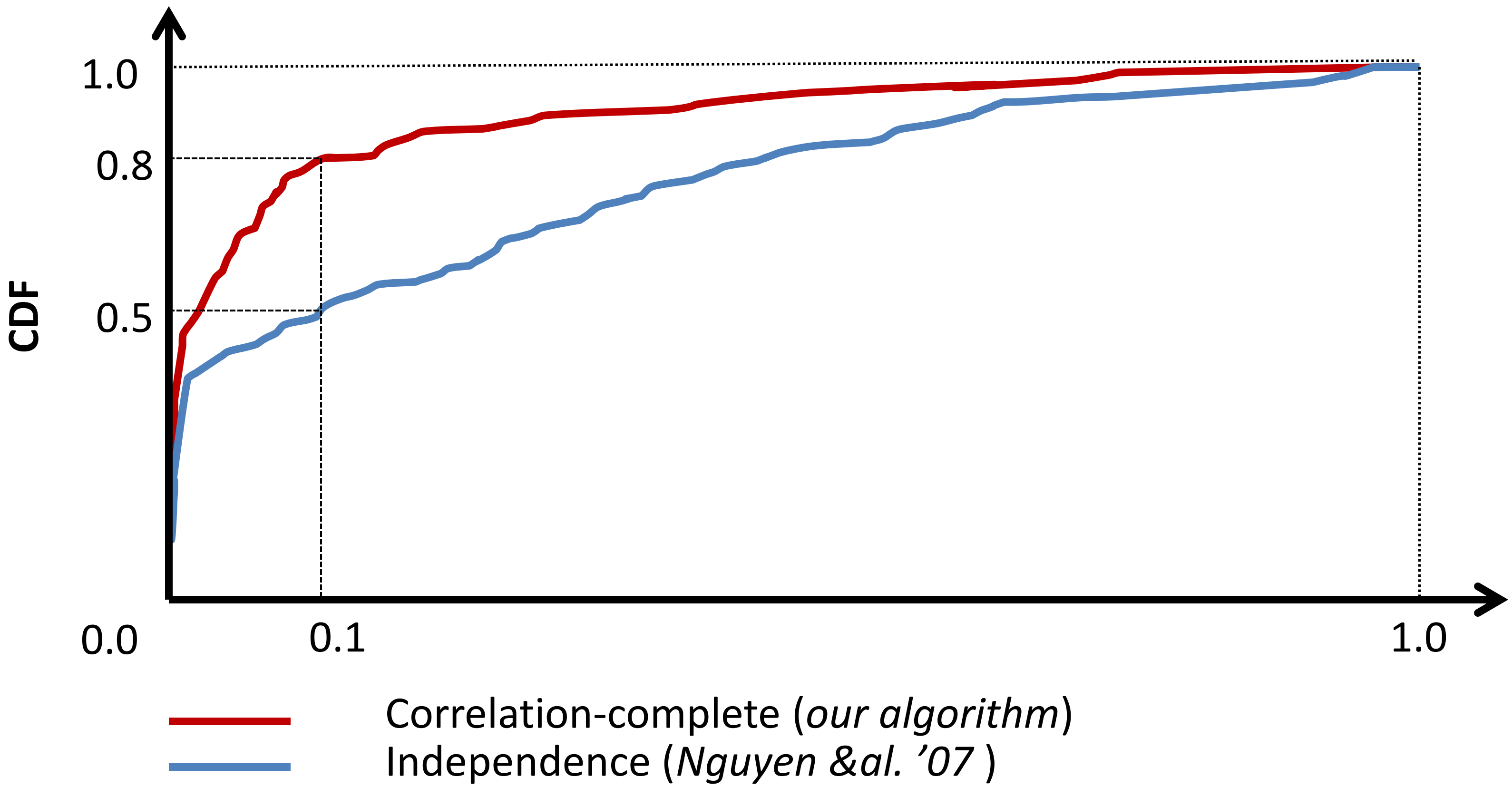
$= \mathcal{R}$ (first three rows)

$$Row(\mathcal{P}) \times Nullspace(\mathcal{R}) \neq 0$$

$$\begin{bmatrix} ? & ? & ? & 1 & ? \end{bmatrix} \times \begin{bmatrix} 0.0 & -0.4 \\ -0.7 & 0.0 \\ 0.0 & -0.7 \\ 4.0 & -4.3 \\ 0.2 & 2.5 \end{bmatrix} \begin{matrix} e_{AB} \\ e_{FB} \\ e_{BC} \\ e_{BD} \\ e_{BC}, e_{BD} \end{matrix}$$

Path set $\mathcal{P}$ must cover link e_{BD}

Simulations



absolute error between the actual probability that a link is congested, and the probability inferred by the algorithm.

Conclusions

In the commercial ISP scenario, the information provided by Boolean Inference cannot be computed accurately.

Boolean Inference is ill-posed.

The assumptions made by Boolean Inference algorithms are not verifiable in practice.

The “right” problem to solve which provides useful accurate information is Congestion Frequency.

Congestion Frequency is well-posed under certain well-defined conditions, and requires only one realistic assumption.

We propose a complete and efficient algorithm which computes the congestion frequencies of links.

Future Work

Our method gives insights on how to deal with noisy measurements.