Traffic Pattern-based Adaptive Routing in Dragonfly Networks

Peyman Faizian , Shafayat Rahman
Atiqul Mollah , Xin Yuan
Florida State University

Scott Pakin
Mike Lang
Los Alamos National Laboratory

Motivation

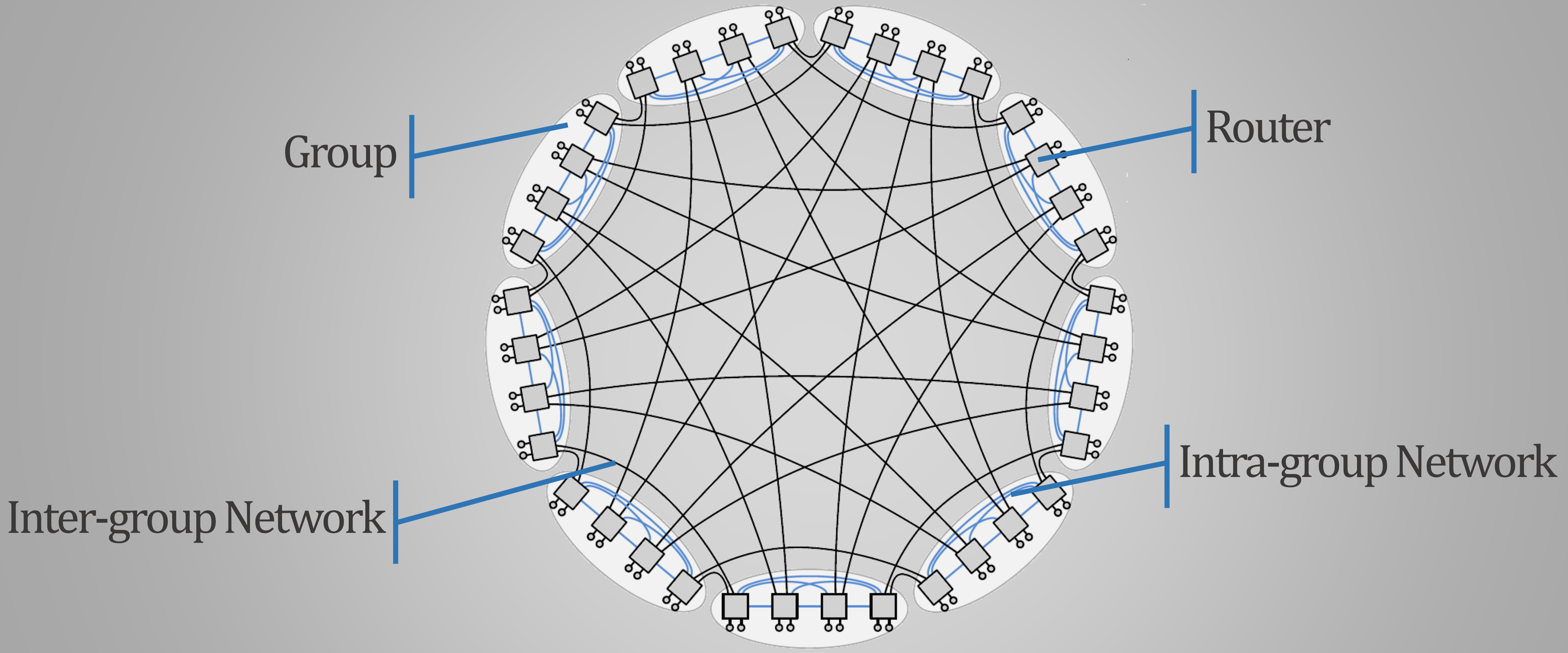
Dragonfly has been known as a potential topology for next generation of HPC systems

Effective routing in Dragonfly depends on the traffic pattern: minimal routing for uniform traffic and non-minimal routing for adversarial traffic

Adaptive routing is required to achieve good performance under various traffic patterns which chooses between minimal and non-minimal paths based on respective queue lengths

We will show that the available methods have some limitations and propose a traffic pattern-based adaptive routing to address these issues

Dragonfly

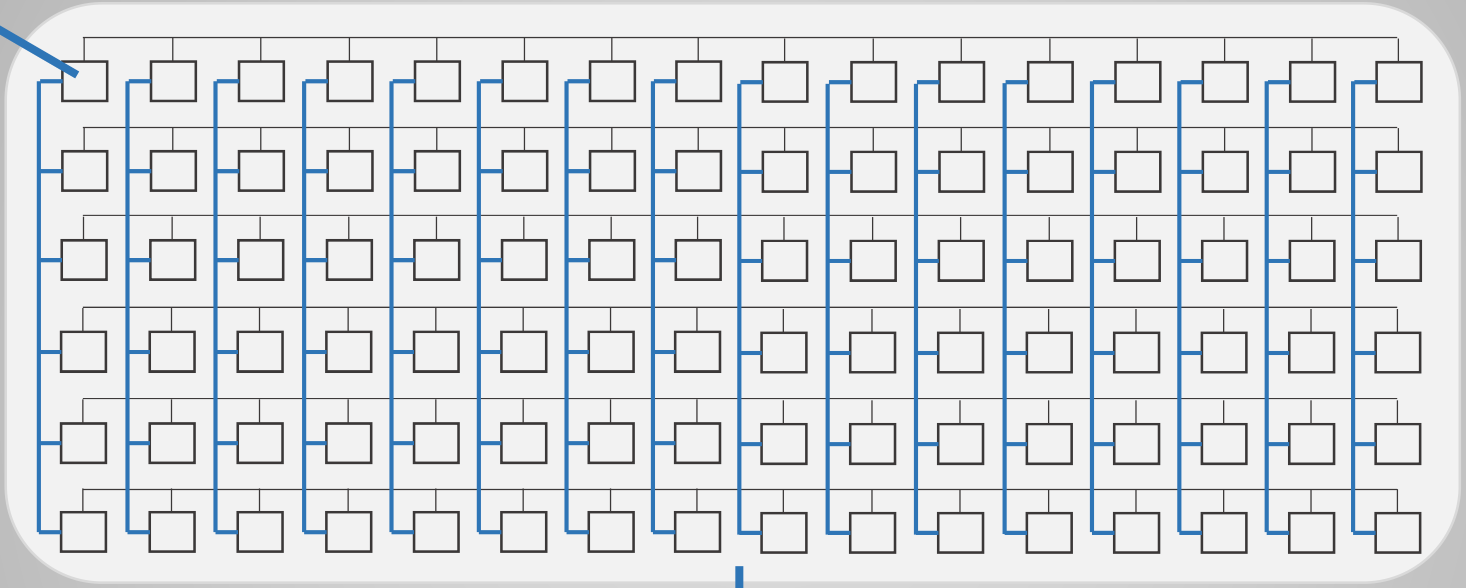


[Garcia et al, INA-OCMC '13]

Cray Cascade

All-to-all inter-group network

Radix-48 Aries
8 processing nodes/router



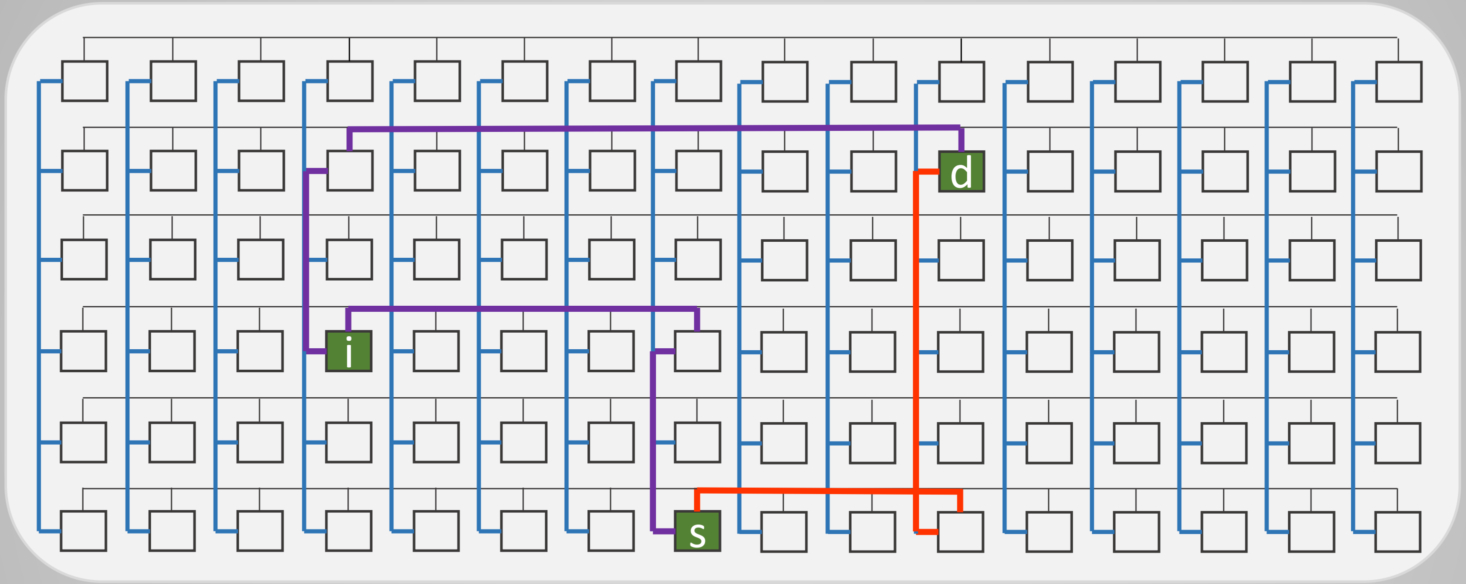
2D HyperX 16x6
Intra-group network

Dragonfly Routing

Basic intra-group routing

— Minimal

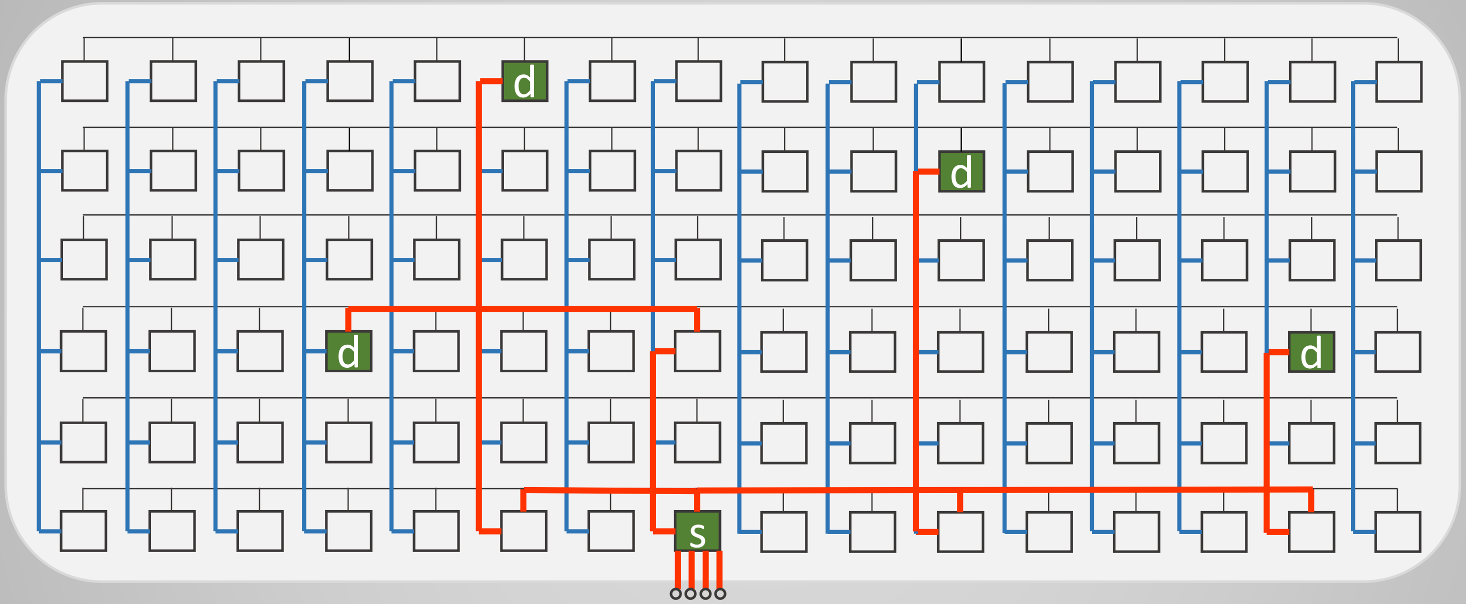
— VLB



Minimal path length ≤ 2 hops
VLB path length ≤ 4 hops

Dragonfly Routing

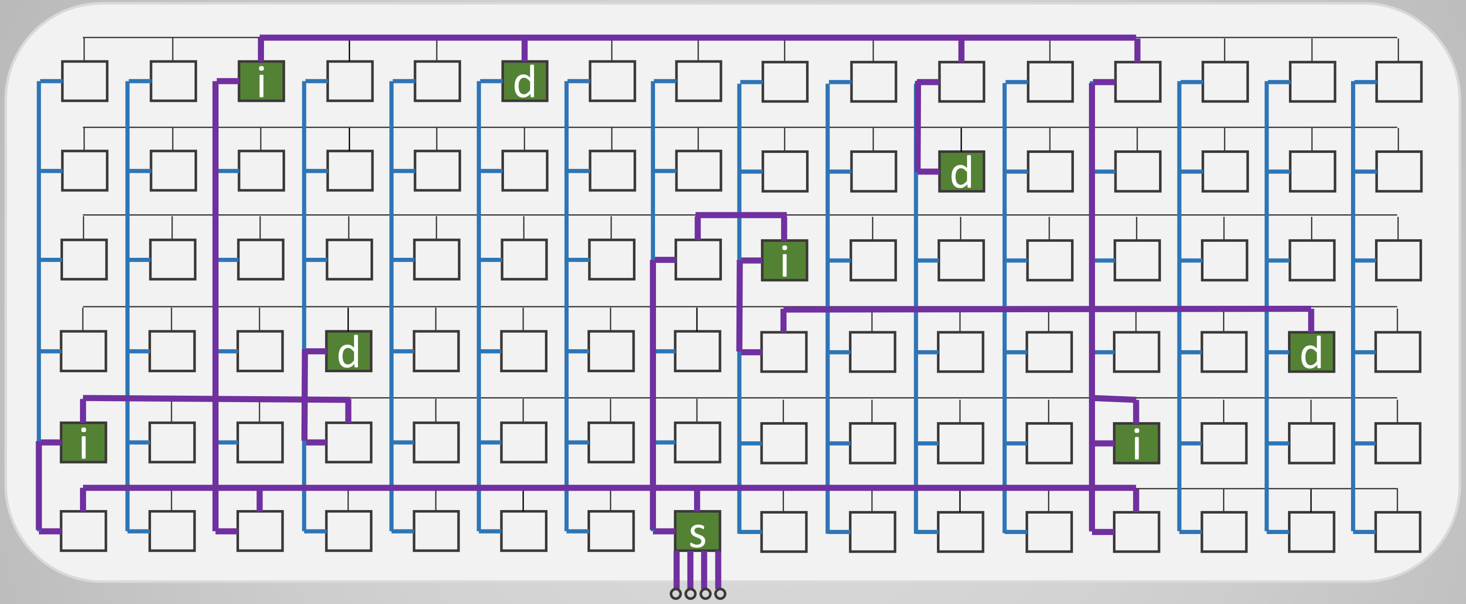
MIN with uniform random traffic



4 packets
8 links used

Dragonfly Routing

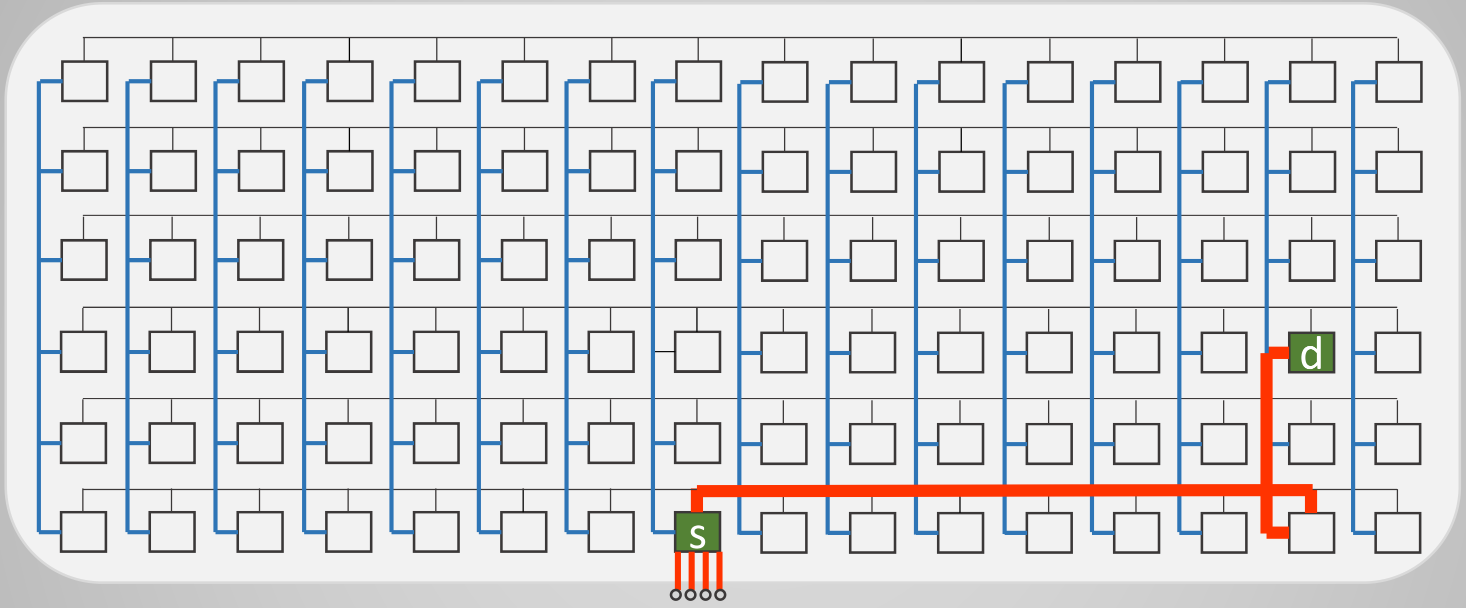
VLB with uniform random traffic



4 packets
16 links used

Dragonfly Routing

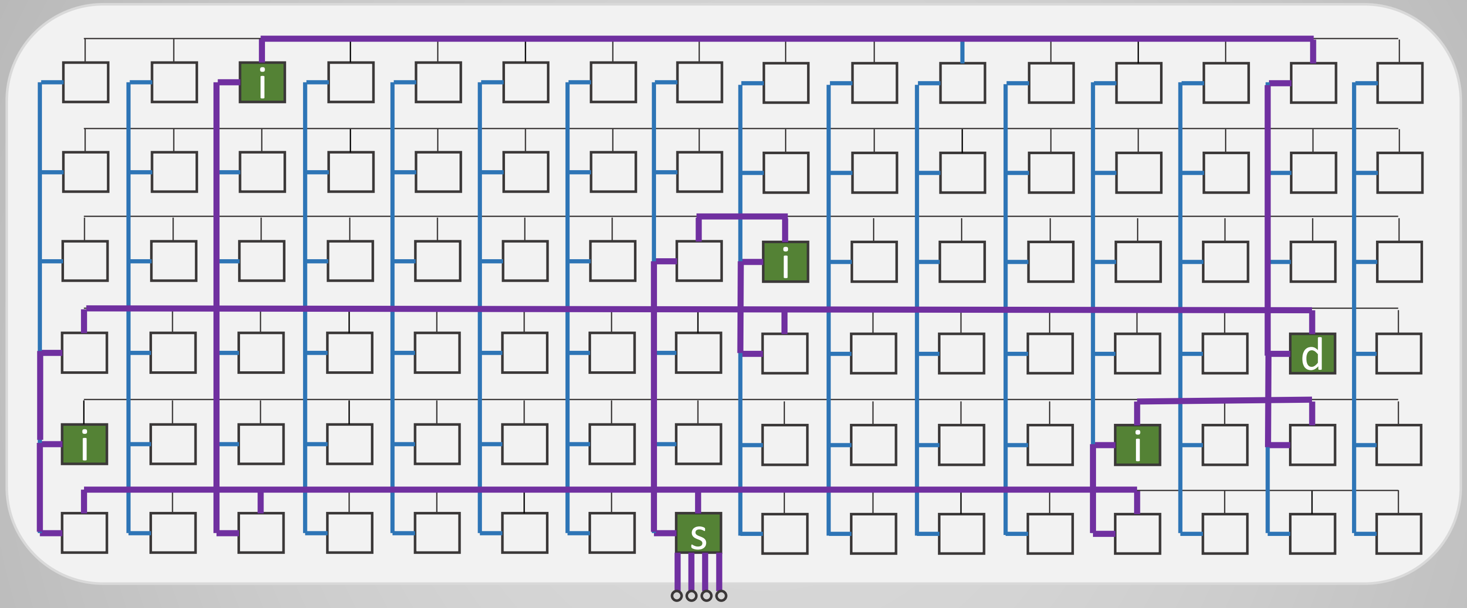
MIN with adversarial traffic



4 packets
2 links used, max bandwidth = 1/4

Dragonfly Routing

VLB with adversarial traffic



4 packets
16 links used, max bandwidth = 1

Dragonfly Routing

Adaptive routing

MIN
VLB

$$Q_{min} \times H_{min} \leq Q_{vlb} \times H_{vlb} + T$$

MIN
or
VLB

CHOOSE
MIN path
VLB path

SELECT
least loaded path

FORWARD

Dragonfly Routing

Adaptive routing

$$\boxed{Q_{min} \times H_{min} \leq Q_{vlb} \times H_{vlb} + T}$$

T Bias towards selecting MIN path

T Bias towards selecting VLB path

"T is used to balance the performance under uniform random and worst case traffic patterns"

[Jiang et al, ISCA'09]

"Value of T needs to be determined empirically"

[Jiang et al, ISCA'09]

… so the performance of UGAL is influenced by T which is driven by the traffic pattern …

... so the performance of UGAL is influenced by T which is driven by the traffic pattern ...

Thus, Identifying the traffic pattern could help to improve the performance of UGAL

… so the performance of UGAL is influenced by T which is driven by the traffic pattern …

Thus, Identifying the traffic pattern could help to improve the performance of UGAL

But how do we identify the traffic pattern???

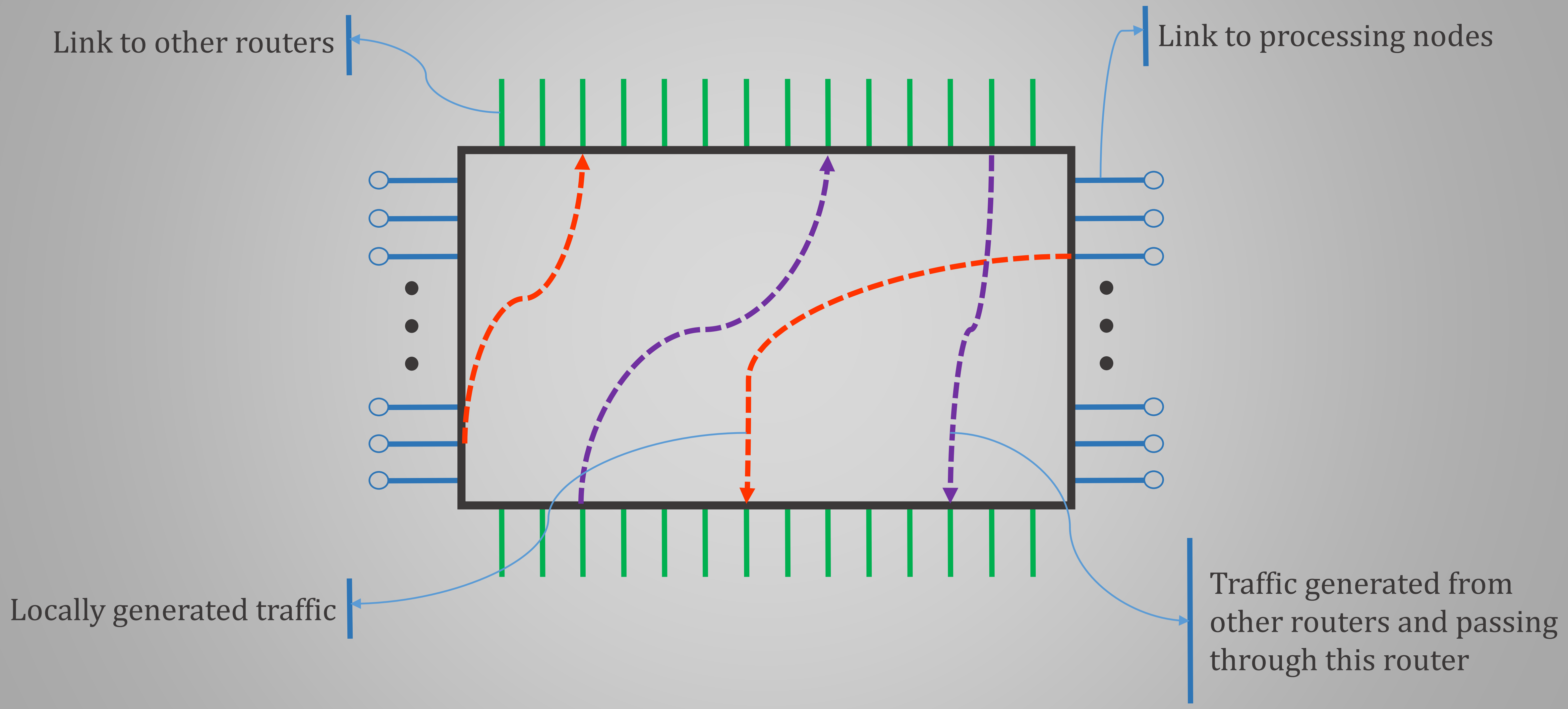
Why to Identify Traffic Pattern

Minimal routing works best under load balanced or uniform random traffic

Non-minimal routing is desirable when adversarial traffic is observed

By identifying these traffic patterns, we can decrease the number of false routing decisions made by the adaptive routing scheme

Observed Traffic at Each Router



Quantifying Traffic Pattern

Locally generated traffic

$DestC_i$

i =	0	1	2	3	4	5	…	*a*-3	*a*-2	*a*-1
	2	4	3	3	4	3	…	2	3	4

Uniform Random, h = 50, injection rate = 0.4

$DestC_i$

i =	0	1	2	3	4	5	…	*a*-3	*a*-2	*a*-1
	0	0	0	360	0	0	…	0	0	0

Adversarial, h = 50, injection rate = 0.4

Count the number of generated packets sent to each destination router over the past *h* cycles

Quantifying Traffic Pattern

Local traffic pattern can be quantified using *localimpact*

$$Localimpact = DestC_i/h$$

Injection Rate	*Pattern*	*$DestC_i$*	*Localimpact*
0.1	UR	1	0.02
	ADV	90	1.80
0.44	UR	4	0.08
	ADV	396	7.92
0.9	UR	8	0.16
	ADV	∞	∞

→

$Localimpact < low_l$	*Benign*
$Localimpact > high_l$	*Adversarial*
otherwise	*Mixed*

Quantifying Traffic Pattern

Globally generated traffic

$Port\_thr_j$

j =	0	1	2	3	4	5	…	k-3	k-2	k-1
	7	10	8	9	13	10	…	9	7	11

Uniform Random, h = 50, injection rate = 0.4

$Port\_thr_j$

j =	0	1	2	3	4	5	…	k-3	k-2	k-1
	30	35	36	33	36	28	…	29	37	32

Adversarial, h = 50, injection rate = 0.4

Count the number of packets generated from other routers and passing through each port over the past *h* cycles

Quantifying Traffic Pattern

Global traffic pattern can be quantified using *globalimpact*

$$Globalimpact = Port\_thr_j/h$$

Injection Rate	Pattern	$Port\_thr_j$	Globalimpact
0.1	UR	2.24	0.04
	ADV	5.45	0.11
0.44	UR	9.86	0.20
	ADV	33.7	0.67
0.9	UR	20.5	0.41
	ADV	∞	∞

$Globalimpact < low_g$	Benign
$Globalimpact > high_g$	Adversarial
otherwise	Mixed

Traffic Pattern Based Adaptive Routing

Based on ***localimpact*** and ***globalimpact*** , our routing scheme operates in nine operating regions

	globalimpact		
localimpact	*benign* / *benign*	*benign* / *mixed*	*benign* / *adversarial*
	mixed / *benign*	*mixed* / *mixed*	*mixed* / *adversarial*
	adversarial / *benign*	*adversarial* / *mixed*	*adversarial* / *adversarial*

We knew that by tuning T
under different traffic patterns
we can improve the
performance of UGAL

We knew that by tuning T under different traffic patterns we can improve the performance of UGAL

We introduced a mechanism to distinguish operating regions for our routing scheme based on local and global traffic info

We knew that by tuning T under different traffic patterns we can improve the performance of UGAL

We introduced a mechanism to distinguish operating regions for our routing scheme based on local and global traffic info

We can tailor T values to each operating region

Traffic Pattern Based Adaptive Routing

UGAL(T) | Larger T → prefer minimal path
Smaller T → prefer non-minimal path

		globalimpact		
		benign	**mixed**	**adversarial**
localimpact	**benign**	MIN/UGAL(64)	MIN/UGAL(64)	UGAL(48)
	mixed	UGAL(-4)	UGAL(-20)	UGAL(-40)
	adversarial	UGAL(-48)	UGAL(-64)	VLB

By observing higher local and global impact, routing moves toward using non-minimal paths to avoid congestion

Evaluation Methodology

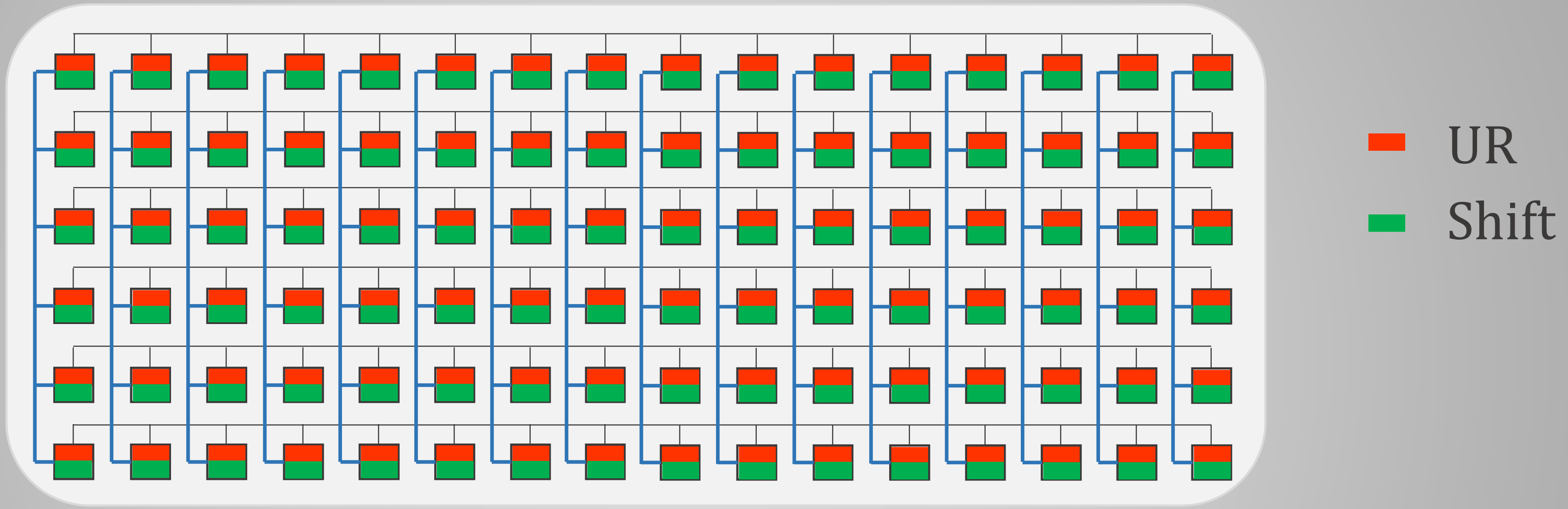
NETWORK
1 Group of a Cray Cascade machine
16x6 2D HyperX, a=96, p=18

SIMULATION
Booksim, 4 VCs, VC buffer size = 32
Single flit packets

ROUTING
MIN, VLB, UGAL-L, UGAL-G, TPR

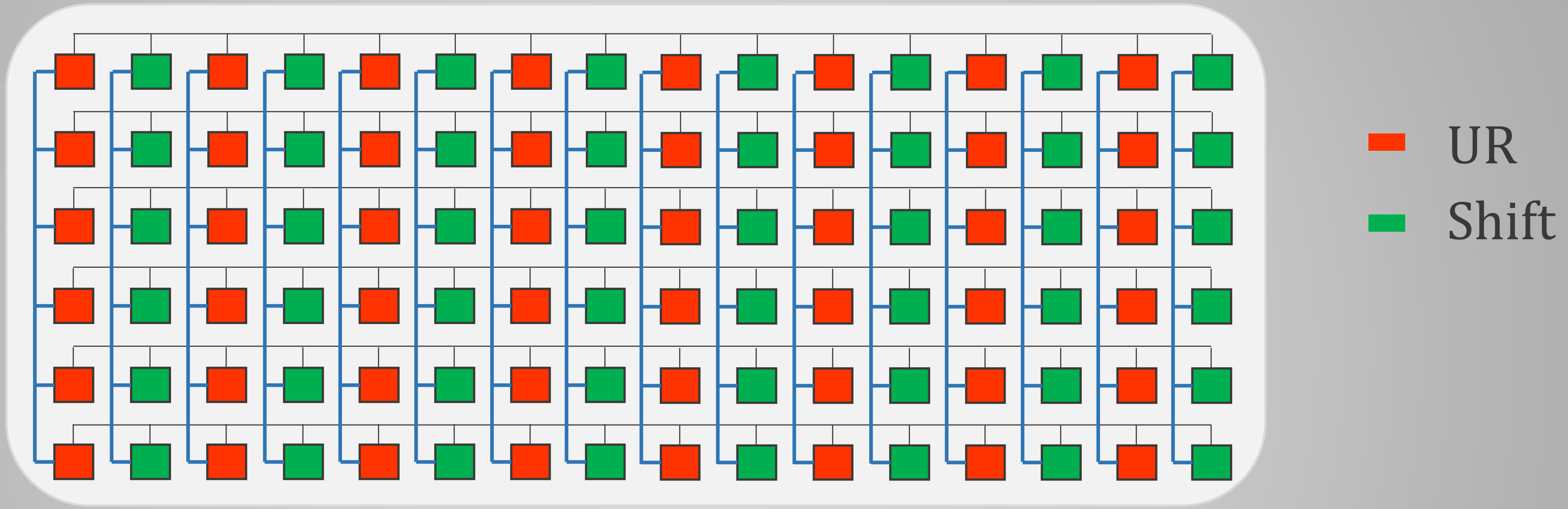
TRAFFIC
Uniform Random, $Shift_1$, NLC_URADV, RLC_URADV
Only intra-group communication

Node-Level Combined Traffic



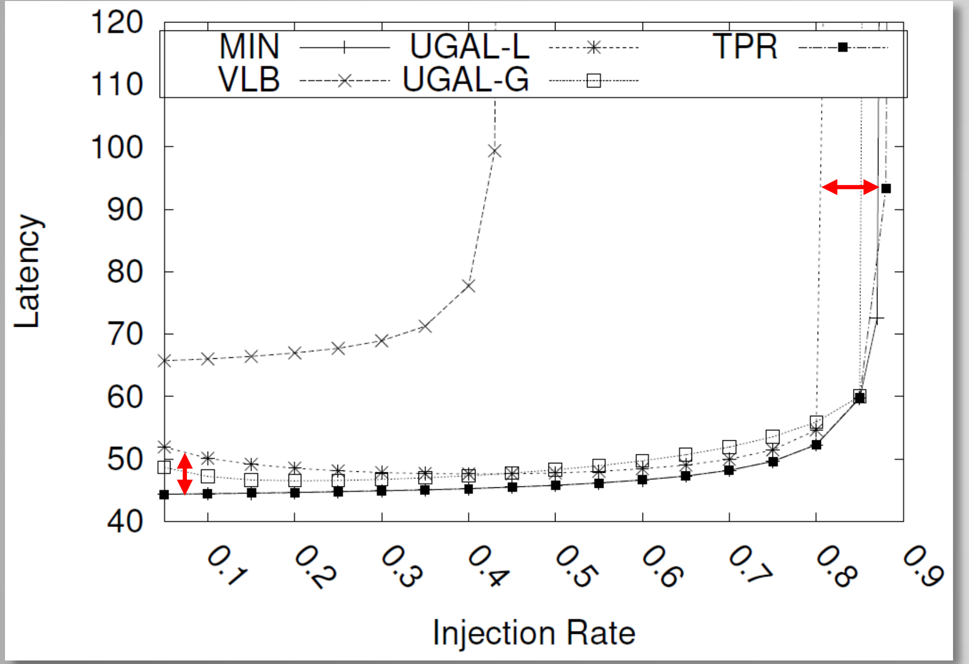
NLC_URADV(50,50)

Router-Level Combined Traffic

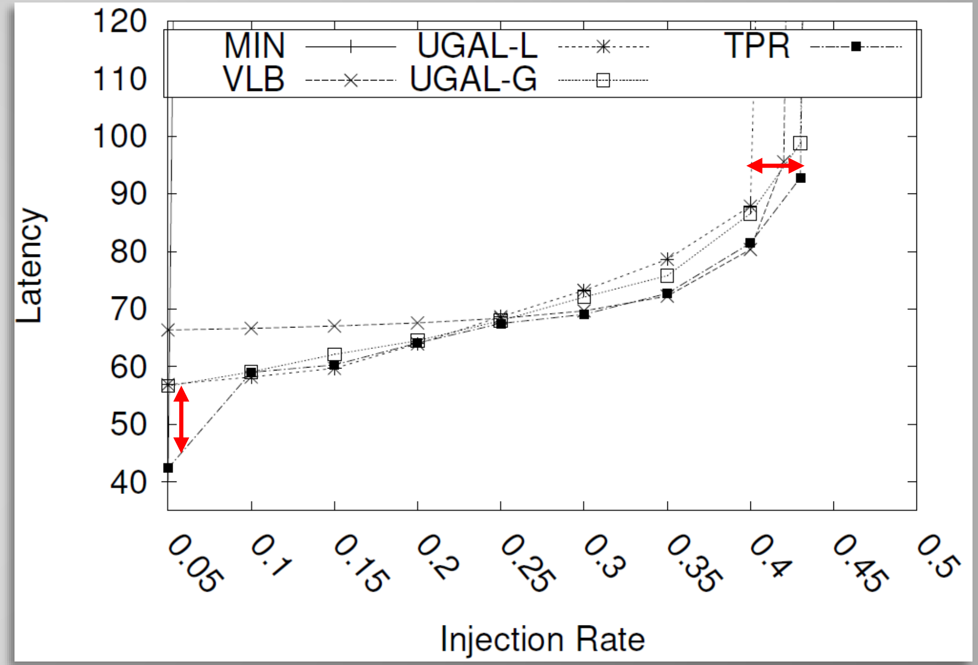


RLC_URADV(50,50)

Evaluation Results

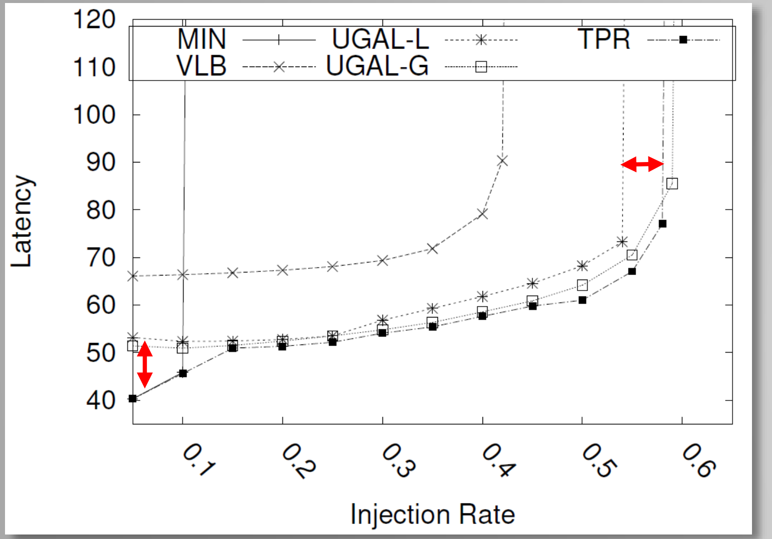


Uniform Random

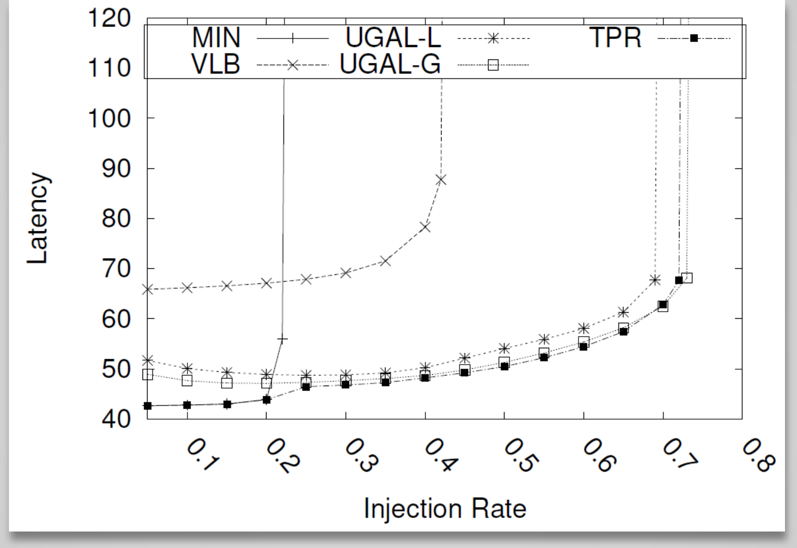


Shift_1

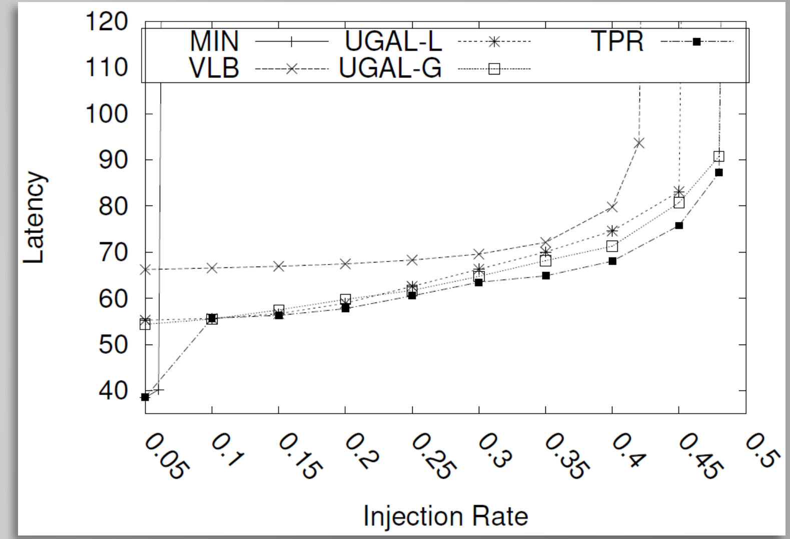
Evaluation Results



NLC_URADV
(50,50)

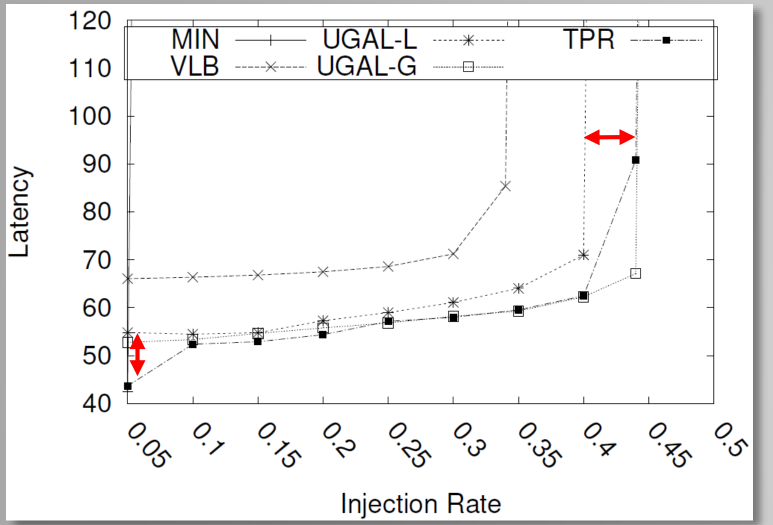


NLC_URADV
(80,20)

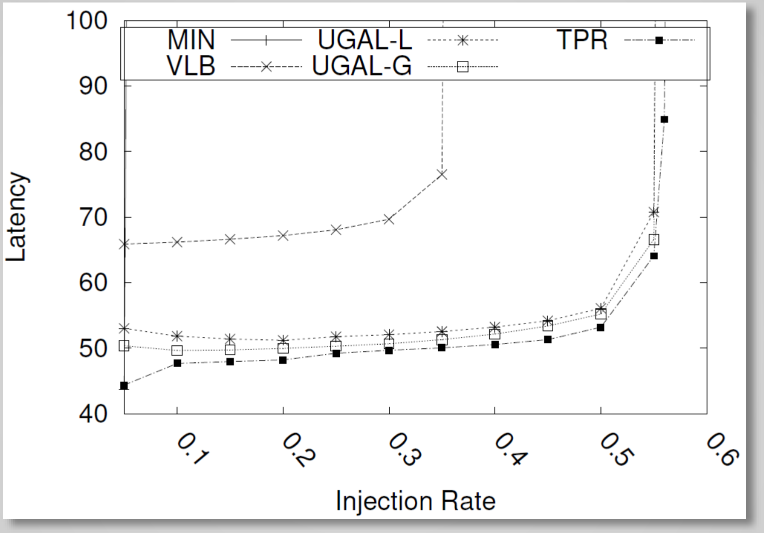


NLC_URADV
(20,80)

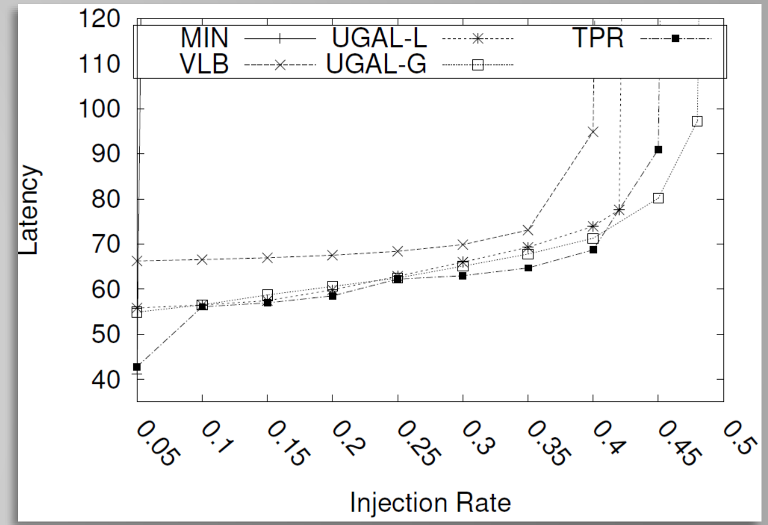
Evaluation Results



RLC_URADV (50,50)



RLC_URADV (80,20)



RLC_URADV (20,80)

Conclusion

By identifying local and global traffic conditions, TPR achieves the best latency results among all evaluated routing schemes

TPR improves the throughput performance of UGAL-L for almost every traffic pattern considered in this study

The same proposed method, can improve the performance of other similar routing schemes including Piggyback, Reservation and Progressive adaptive routing


THE
FLORIDA STATE
UNIVERSITY