Supporting Bottleneck Structure Graphs in ALTO: Use Cases and Requirements

Jordi Ros-Giralt, Sruthi Yellamraju, Qin Wu, Richard Yang, Luis Contreras, Kai Gao, Jensen Zhang

I-Draft: draft-giraltyellamraju-alto-bsg-requirements-00.txt

https://datatracker.ietf.org/doc/draft-giraltyellamraju-alto-bsg-requirements/

IETF Plenary 113

ALTO WG Session

3/23/2022

Table of Contents

- Brief Introduction to Bottleneck Structures
- Bottleneck Structure Graph (BSG) Service: ALTO Use Cases
- Introducing Bottleneck Structures in the IETF ALTO WG: Requirements

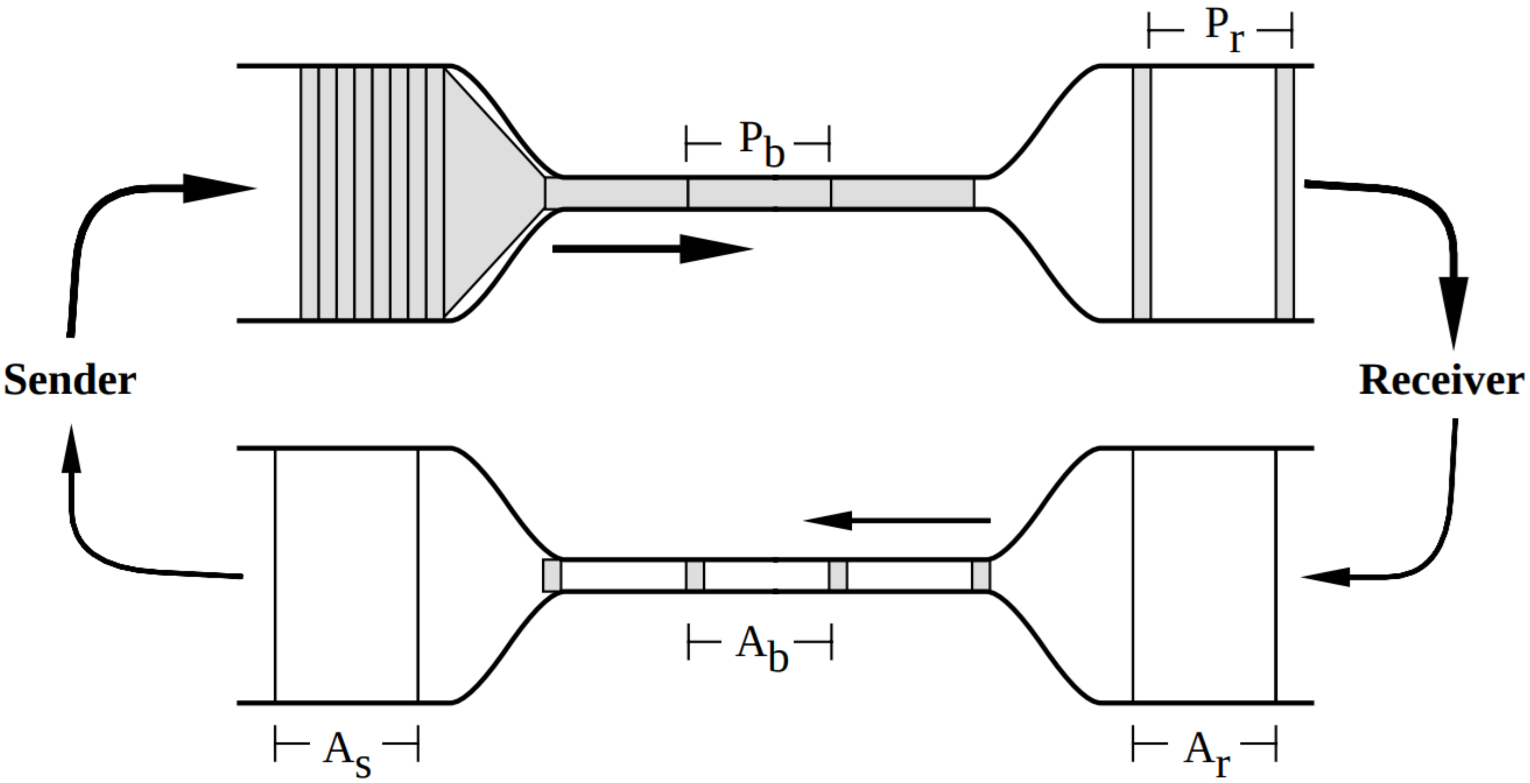
Brief Introduction to Bottleneck Structures

Framework and Implementation Details in the Following Papers

- [1] "On the Bottleneck Structure of Congestion-Controlled Networks," ACM SIGMETRICS, Boston, June 2020 [https://bit.ly/3eGOPrb].
- [2] "Designing Data Center Networks Using Bottleneck Structures," accepted for publication at ACM SIGCOMM 2021 [https://bit.ly/2UZCb1M].
- [3] "Computing Bottleneck Structures at Scale for High-Precision Network Performance Analysis," SC 2020 INDIS, November 2020 [https://bit.ly/3BriwaB].
- [4] "A Quantitative Theory of Bottleneck Structures for Data Networks", Reservoir Labs Technical Report, 2021 [https://bit.ly/38u8ARs].

Conventional View: Single Bottleneck Model

Figure 1: Window Flow Control ‘Self-clocking’

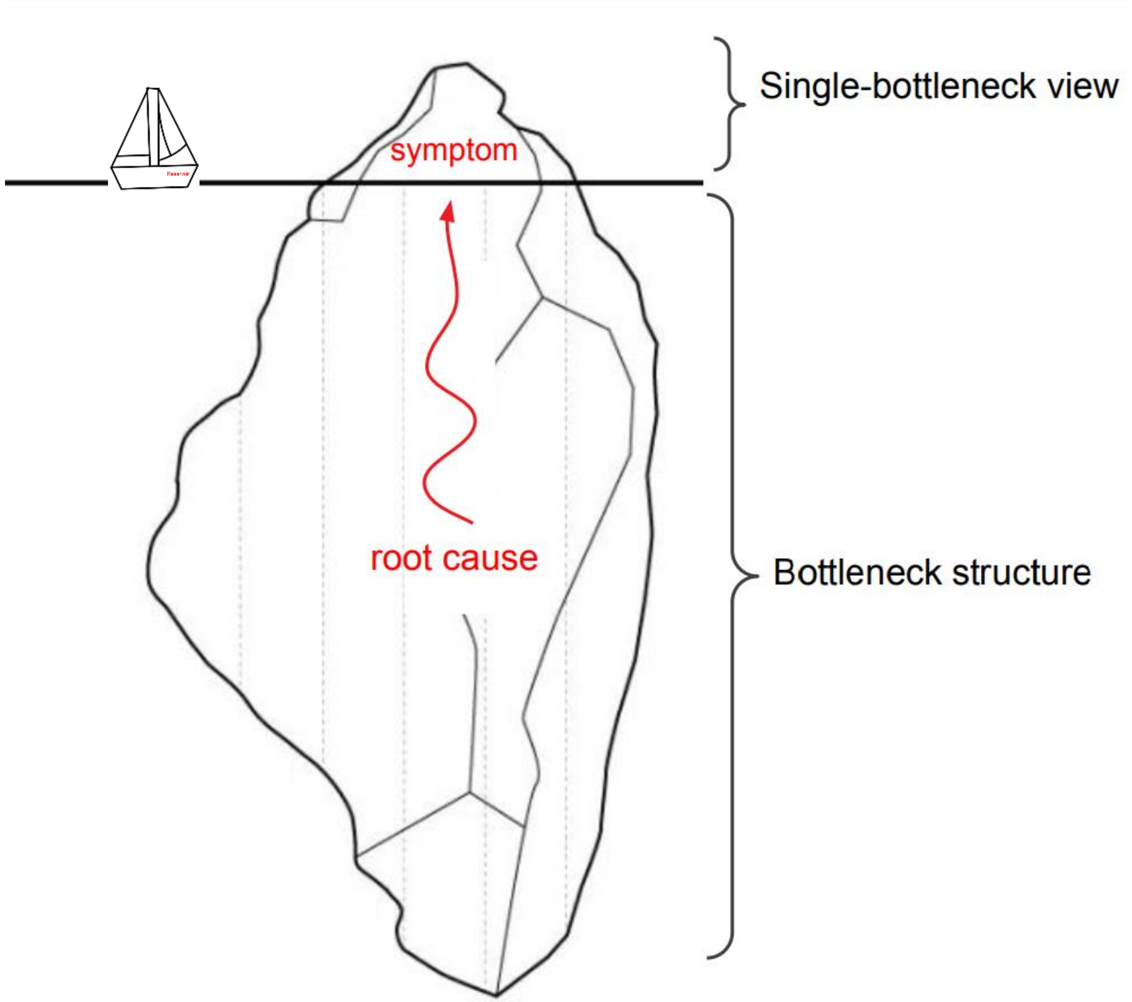


[*] Van Jacobson, "Congestion Avoidance and Control," SIGCOMM, 1988 [https://bit.ly/3FQouFf]

Problem Positioning: The Hiding Root Cause of System-Wide Performance

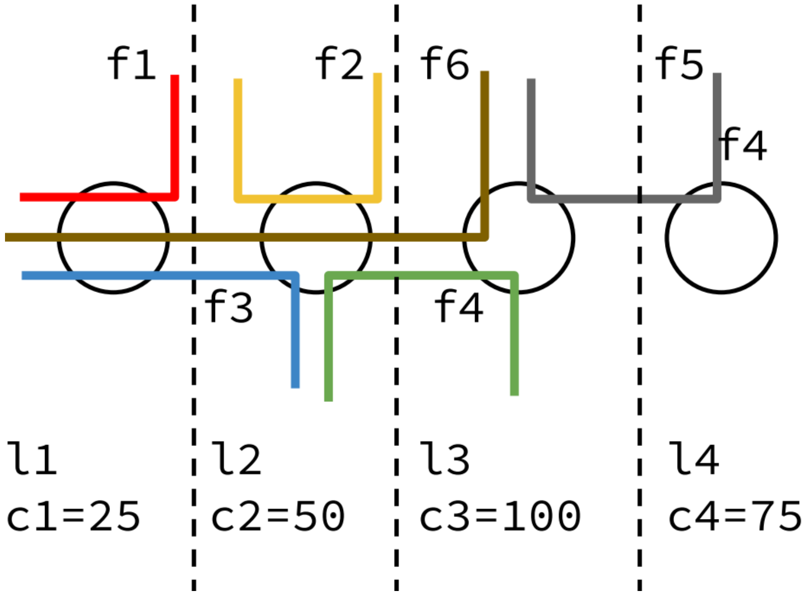
Analogy:

- Structure of the congestion problem in data networks:
  - The single-bottleneck problem is the tip of the iceberg (the symptom)
  - The bottleneck structure is the submerged portion (determines system-wide performance)



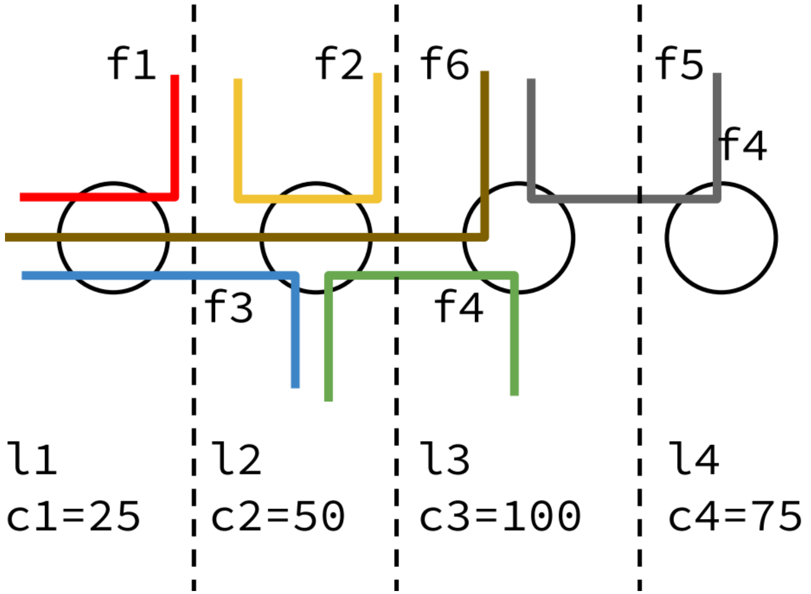
Simple Example of Bottleneck Structure

Communication Network:

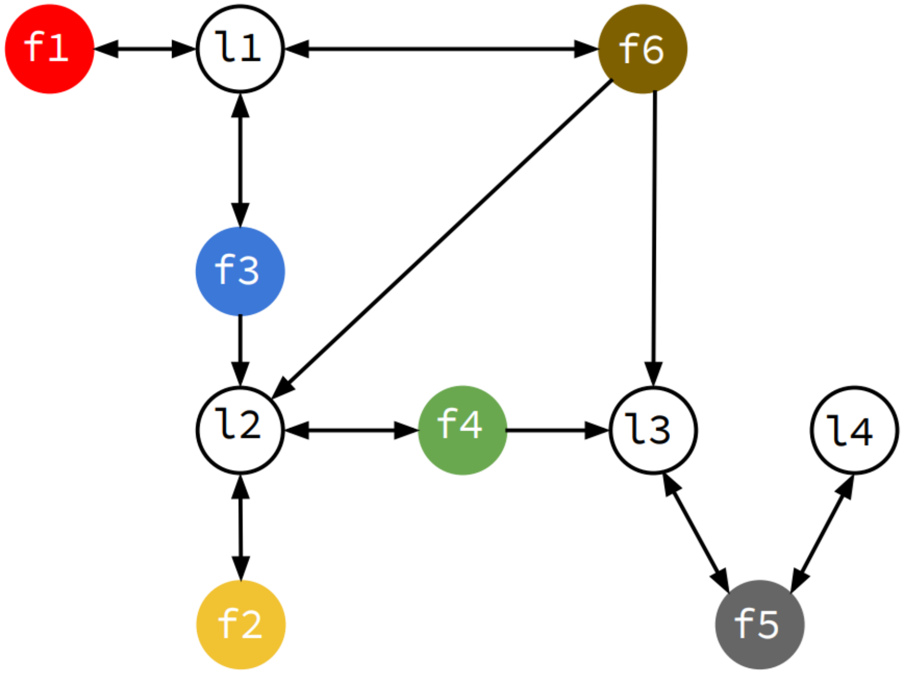


Simple Example of Bottleneck Structure

Communication Network:

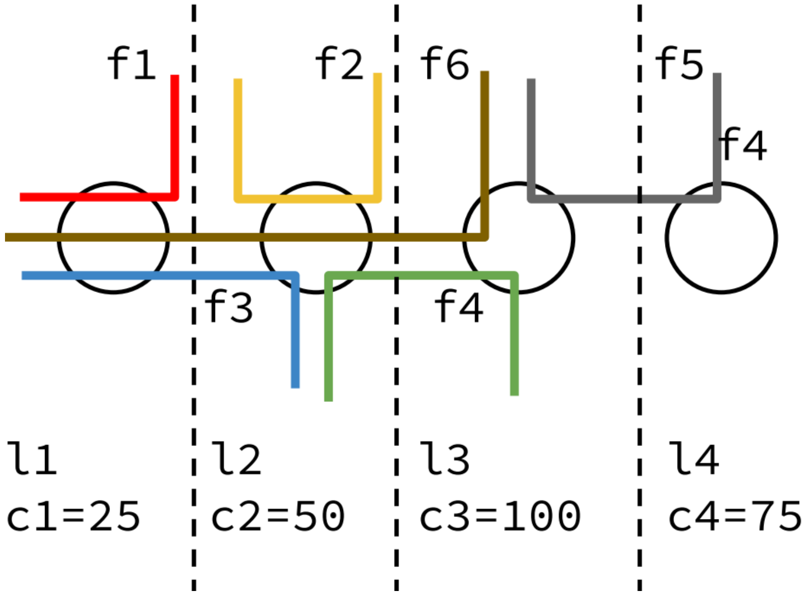


Bottleneck Structure:

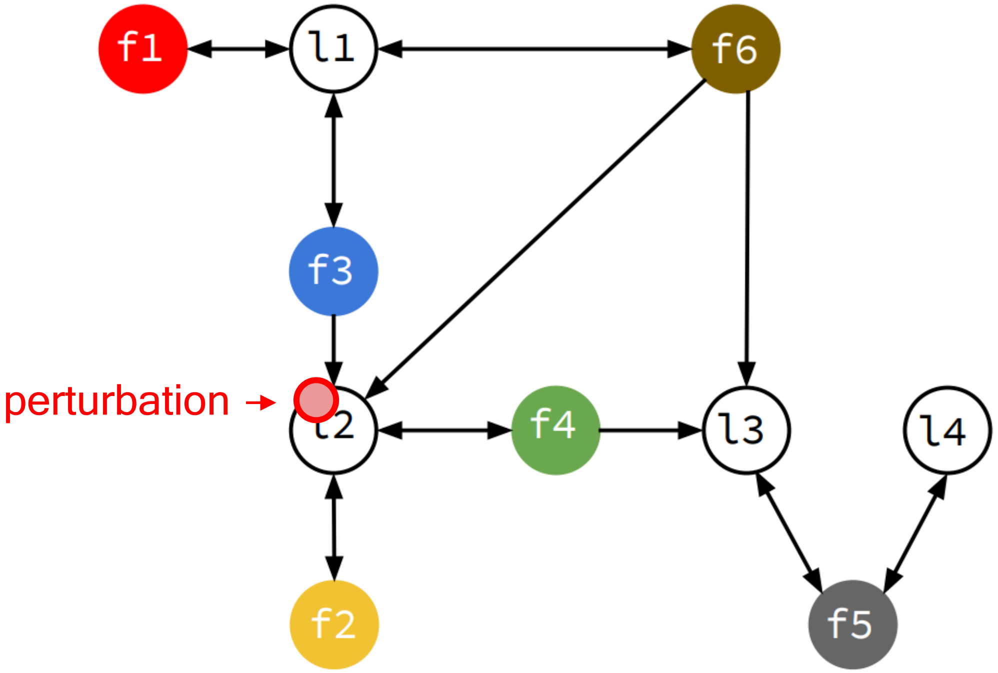


Simple Example of Bottleneck Structure

Communication Network:

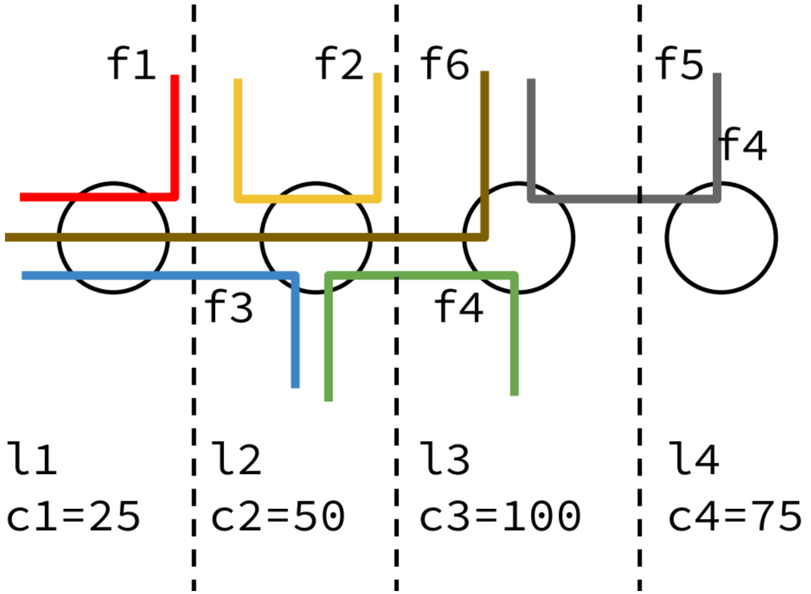


Bottleneck Structure:

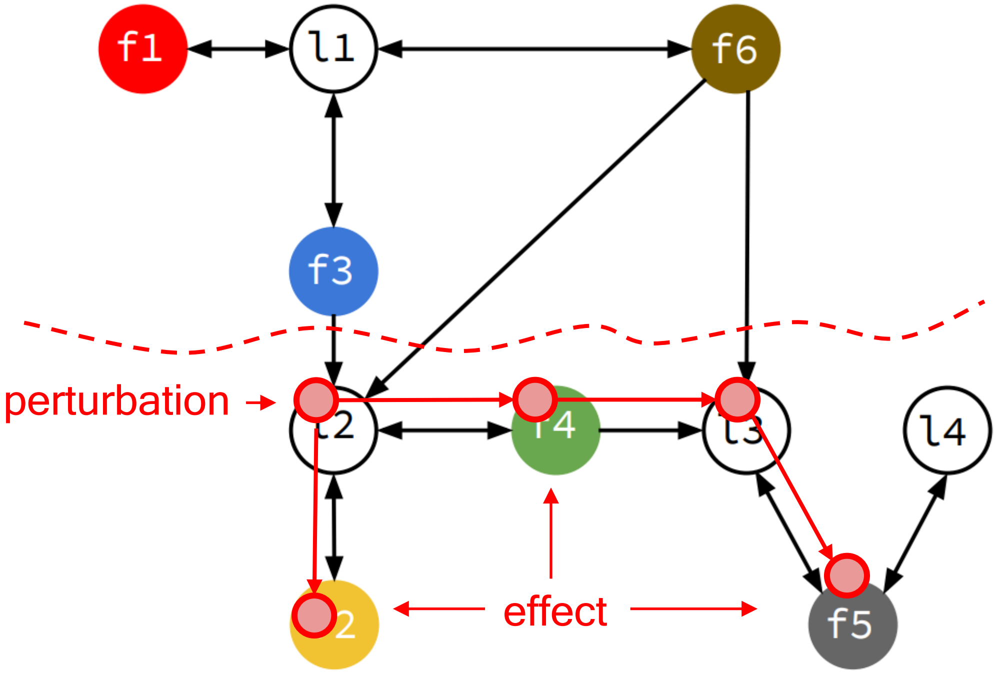


Simple Example of Bottleneck Structure

Communication Network:

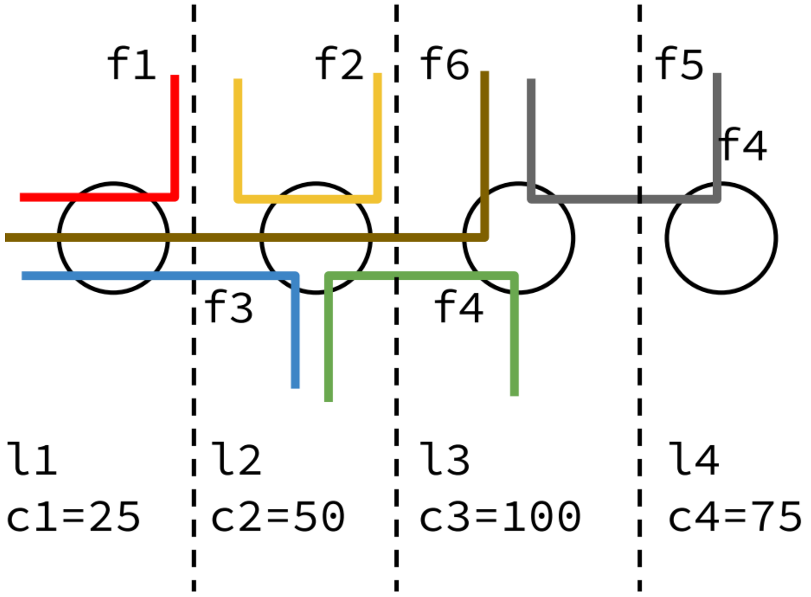


Bottleneck Structure:

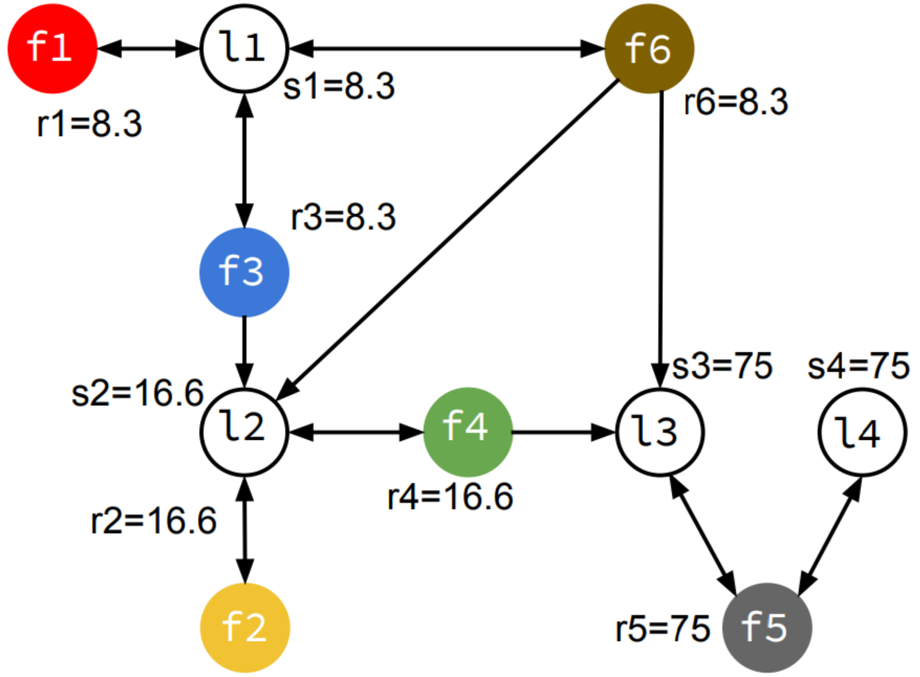


Simple Example of Bottleneck Structure

Communication Network:



Bottleneck Structure:

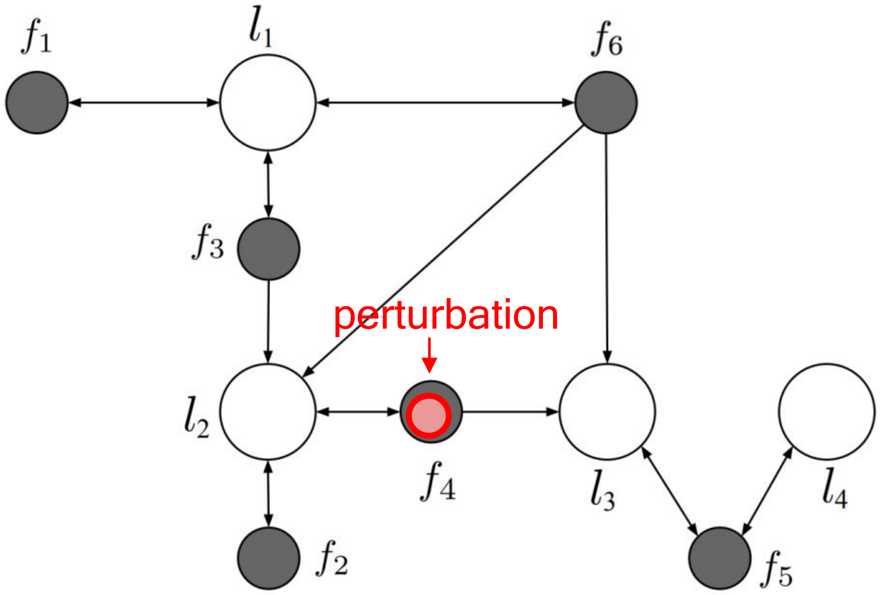


Flow bandwidth allocation: $\mathbf{r} = [\underset{f_1}{8.3}, \underset{f_2}{16.6}, \underset{f_3}{8.3}, \underset{f_4}{16.6}, \underset{f_5}{75}, \underset{f_6}{8.3}]$

Propagation Lemmas

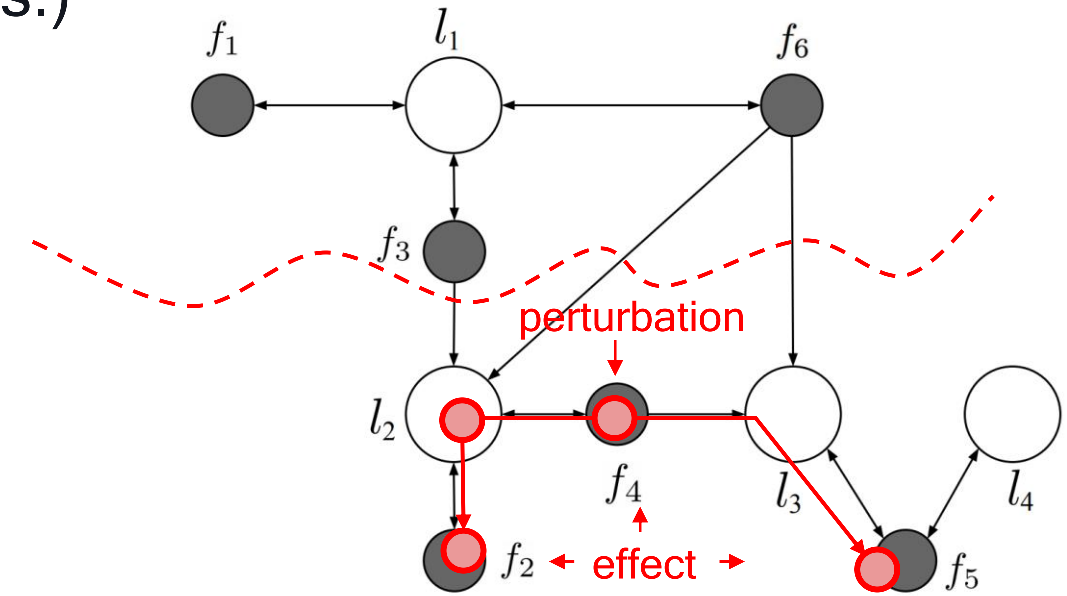
[*] SIGMETRICS 2020 and SIGCOMM 2021: https://bit.ly/3eGOPrb | [https://bit.ly/2UZCb1M]

Propagation Lemma: A flow f can influence the performance of another flow f' iff the bottleneck structure has a directed path from f to f'. (Same lemma exists for bottleneck links.)



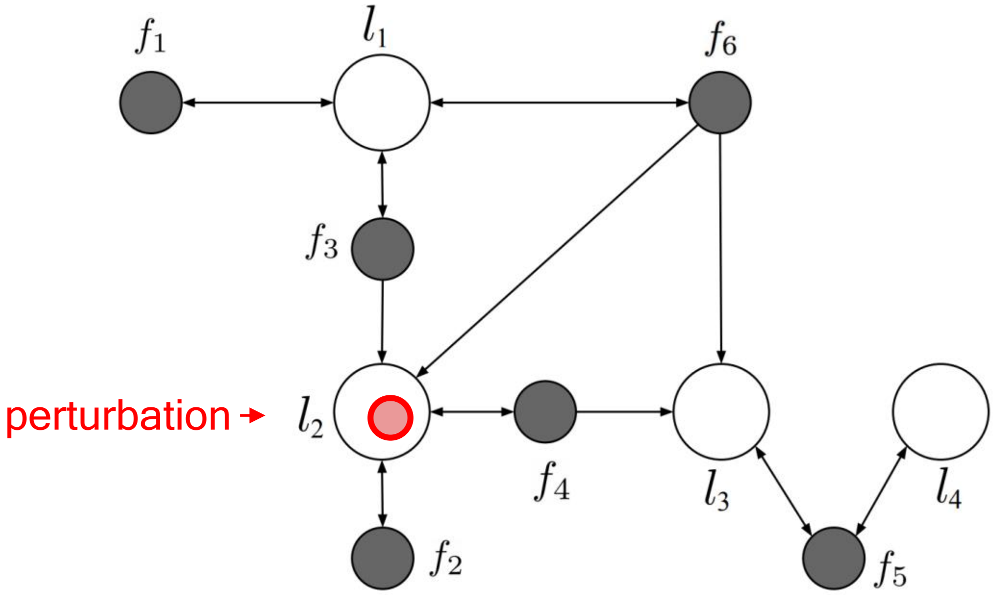
[*] SIGMETRICS 2020 and SIGCOMM 2021: https://bit.ly/3eGOPrb | [https://bit.ly/2UZCb1M]

Propagation Lemma: A flow f can influence the performance of another flow f' iff the bottleneck structure has a directed path from f to f'. (Same lemma exists for bottleneck links.)



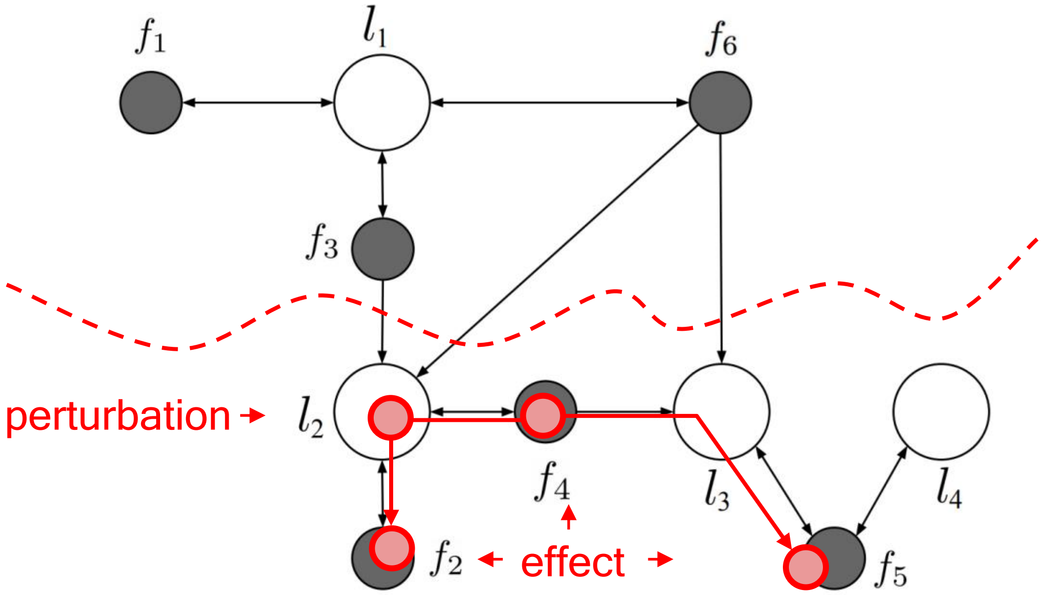
[*] SIGMETRICS 2020 and SIGCOMM 2021: https://bit.ly/3eGOPrb | [https://bit.ly/2UZCb1M]

Propagation Lemma: A flow f can influence the performance of another flow f' iff the bottleneck structure has a directed path from f to f'. (Same lemma exists for bottleneck links.)



[*] SIGMETRICS 2020 and SIGCOMM 2021: https://bit.ly/3eGOPrb | [https://bit.ly/2UZCb1M]

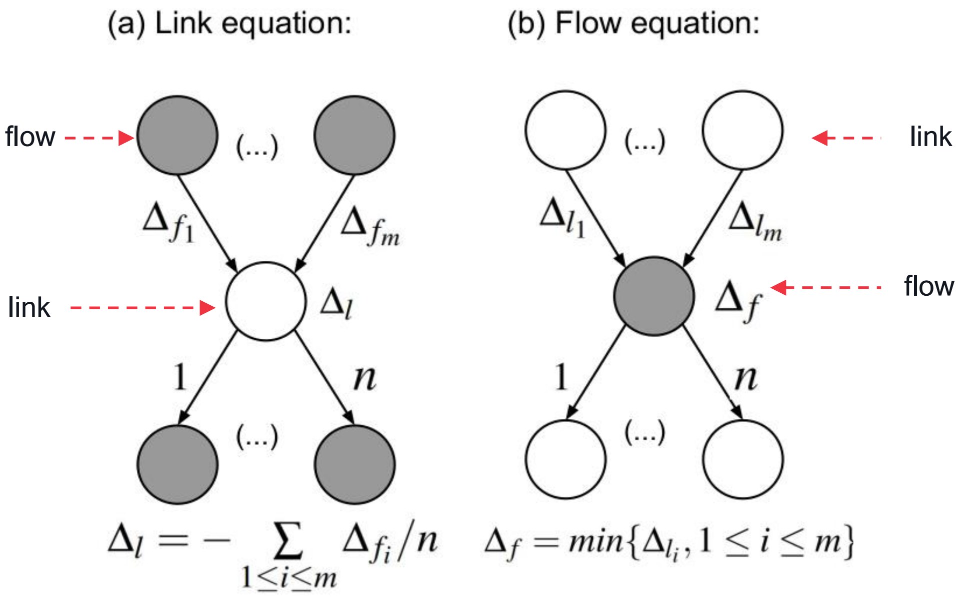
Propagation Lemma: A flow f can influence the performance of another flow f' iff the bottleneck structure has a directed path from f to f'. (Same lemma exists for bottleneck links.)



[*] SIGMETRICS 2020 and SIGCOMM 2021: https://bit.ly/3eGOPrb | [https://bit.ly/2UZCb1M]

Propagation Equations

Link and flow equations:

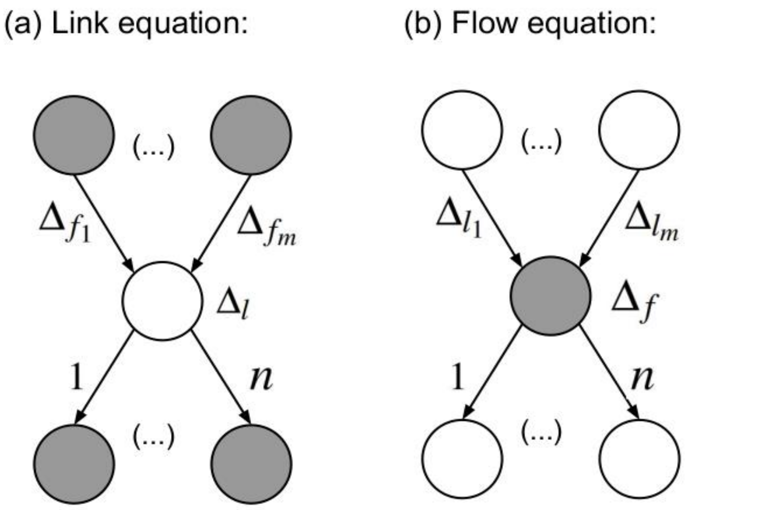


$$\Delta_l = -\sum_{1 \le i \le m} \Delta_{f_i}/n \qquad \Delta_f = min\{\Delta_{l_i}, 1 \le i \le m\}$$

[*] SIGMETRICS 2020 and SIGCOMM 2021: https://bit.ly/3eGOPrb | [https://bit.ly/2UZCb1M]

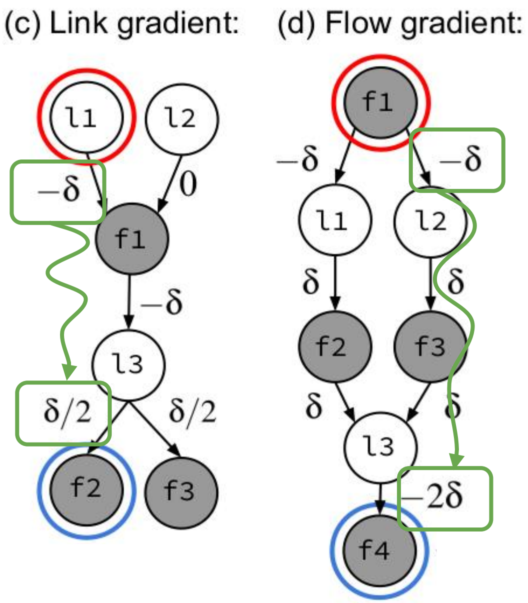
Propagation Equations

Link and flow equations:



$$\Delta_l = - \sum_{1 \le i \le m} \Delta_{f_i} / n \qquad \Delta_f = min\{\Delta_{l_i}, 1 \le i \le m\}$$

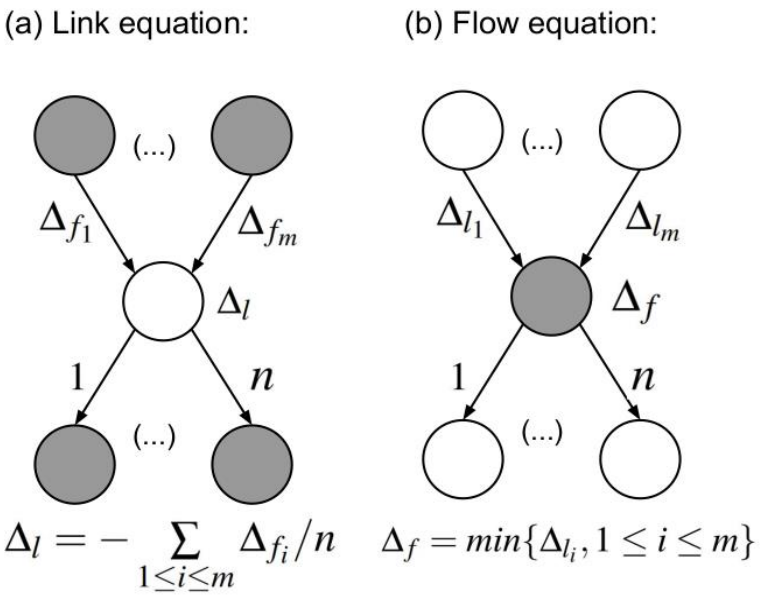
Example:



[*] SIGMETRICS 2020 and SIGCOMM 2021: https://bit.ly/3eGOPrb | [https://bit.ly/2UZCb1M]

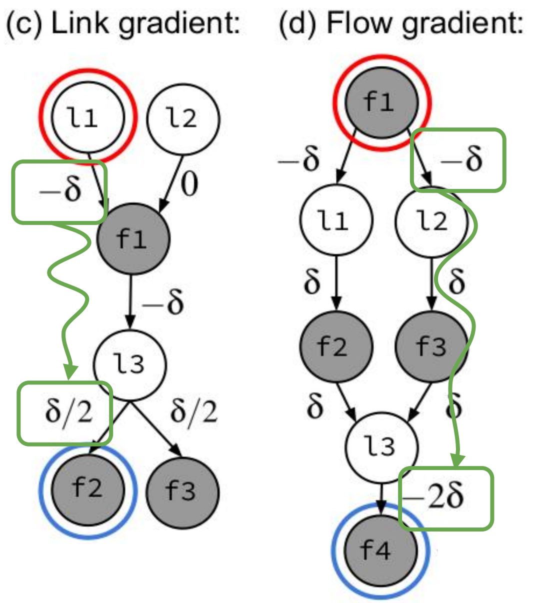
Propagation Equations

Link and flow equations:



$$\Delta_l = -\sum_{1\le i\le m} \Delta_{f_i}/n \qquad \Delta_f = min\{\Delta_{l_i}, 1 \le i \le m\}$$

Example:



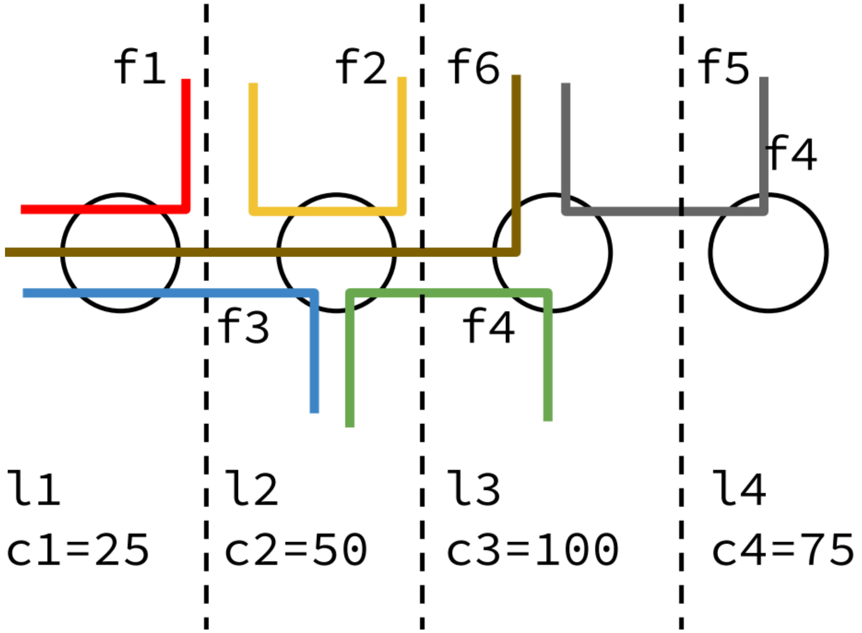
$$\nabla_{l_1}(f_2) = \partial r_{f_2}/\partial c_{l_1} = \Delta_{f_2}/(-\delta) = \frac{\delta/2}{-\delta} = -1/2$$

$$\nabla_{f_1}(f_4) = \partial r_{f_4}/\partial r_{f_1} = \Delta_{f_4}/(-\delta) = -2\delta/(-\delta) = 2$$

[*] SIGMETRICS 2020 and SIGCOMM 2021: https://bit.ly/3eGOPrb | [https://bit.ly/2UZCb1M]

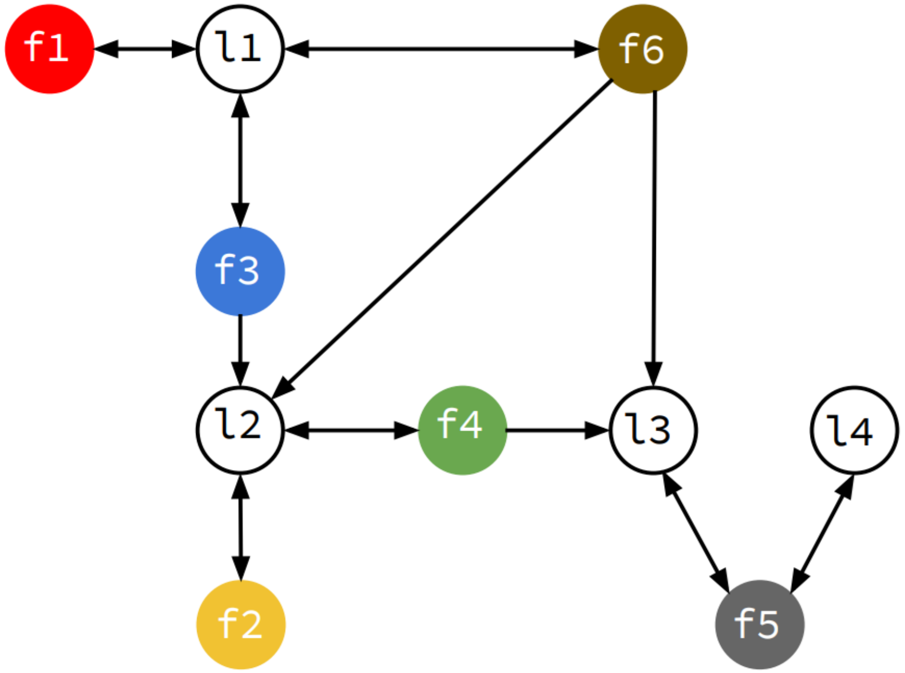
Optimal Flow Throughput Reduction

Communication Network:

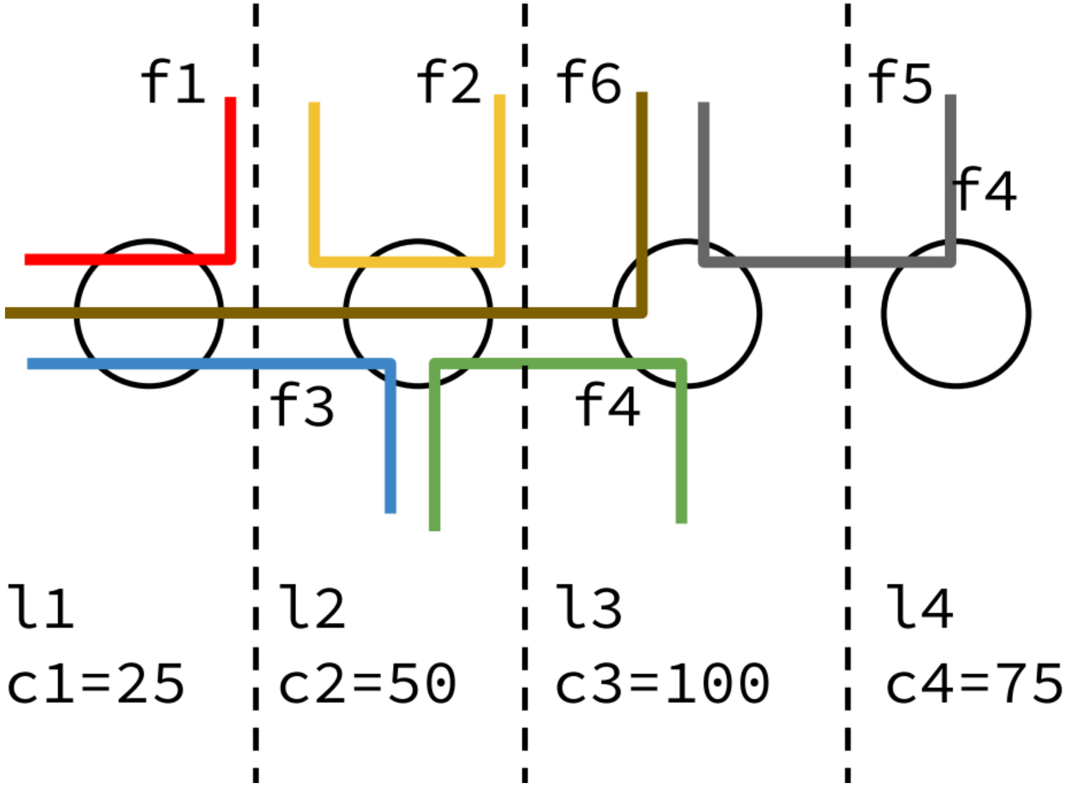


$$\mathbf{r} = [\underset{f_1}{8.3}, \underset{f_2}{16.6}, \underset{f_3}{8.3}, \underset{f_4}{16.6}, \underset{f_5}{75}, \underset{f_6}{8.3}]$$

Bottleneck Structure:



Optimal Flow Throughput Reduction



$$\begin{array}{cccccccc} & & f_1 & f_2 & f_3 & f_4 & f_5 & f_6 \\ \mathbf{r} & = & [8.3, & 16.6, & 8.3, & 16.6, & 75, & 8.3] \end{array}$$

F : Total network flow

$$\partial F/\partial r_1^- = 1$$

$$\partial F/\partial r_2^- = 1$$

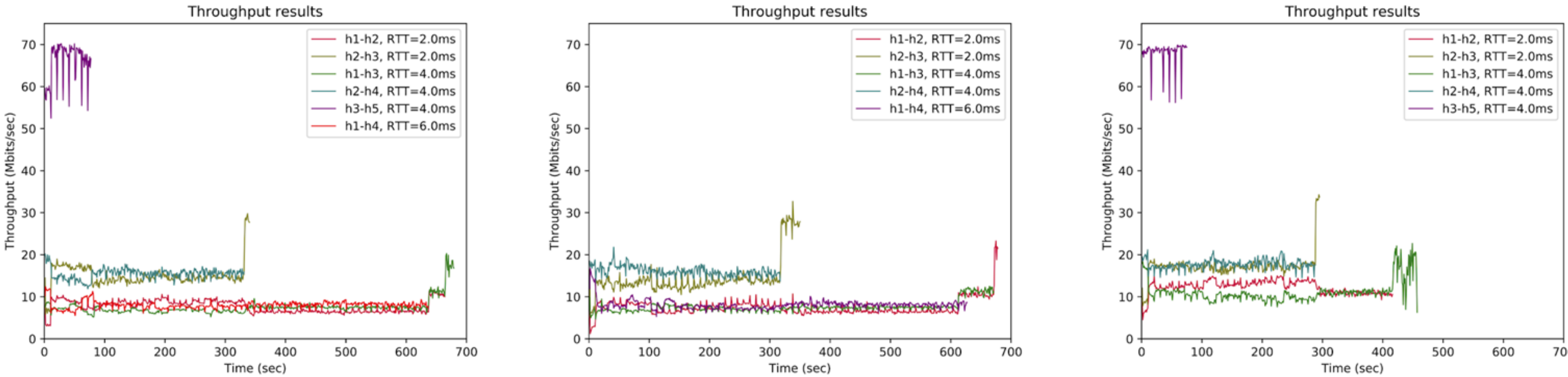
$$\partial F/\partial r_3^- = 1/4$$

$$\partial F/\partial r_4^- = 0$$

$$\partial F/\partial r_5^- = 1$$

$$\boxed{\partial F/\partial r_6^- = -1/2}$$

Optimal Flow Throughput Reduction



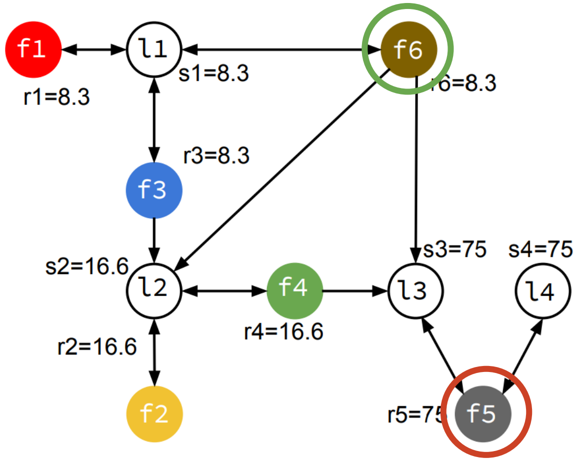
(a) Without removing any flow. **(b) Removing the heavy-hitter flow f_5.** **(c) Removing a low-hitter flow f_6.**

Table 3: As predicted by the theory of bottleneck ordering, flow f_6 is a significantly higher impact flow than flow f_5.

Comp. time (secs)	f_1	f_2	f_3	f_4	f_5	f_6	Slowest
With all flows	664	340	679	331	77	636	679
Without flow f_5	678	350	671	317	—	611	678
Without flow f_6	416	295	457	288	75	—	457

Avg rate (Mbps)	f_1	f_2	f_3	f_4	f_5	f_6	Total
With all flows	7.7	15.1	7.5	15.4	65.8	8.1	119.6
Without flow f_5	7.5	14.5	7.6	16.1	—	8.3	54
Without flow f_6	12.2	17.2	11.1	17.7	68.1	—	126.3

Bottleneck Structure:

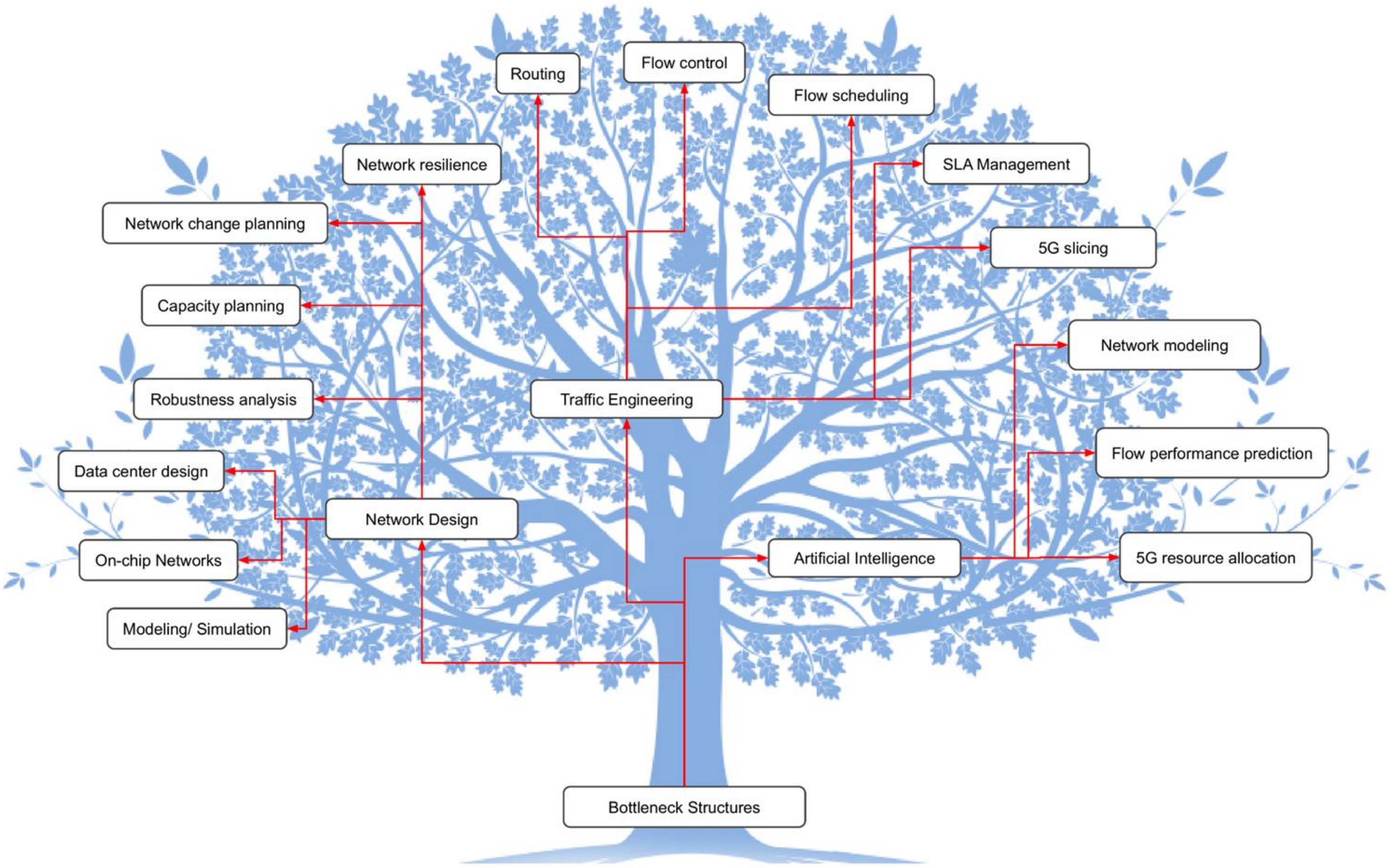


Types of Perturbations (derivatives) Supported by the Bottleneck Structure Graph

- Flow routing
- Traffic shaping (BW enforcement)
- Link capacity upgrades
- Link capacity fluctuations (e.g., SNR in a wireless channel)
- Path shortcuts
- Flow scheduling
- Flow completion
- Job mapping
- Multi-job scheduling

Bottleneck Structure Graph (BSG) Service: ALTO Use Cases

Bottleneck Structure Graphs (BSG): Use Cases



Potential WGs collaborations with ALTO
PANRG
PCE
TEAS
CDNI
COINRG
NETMOD
DETNET
NMRG / digital twins
CAN (BOF)
IAB / Path Signals
Others...

Bottleneck Structure Graphs (BSG): Use Cases Documented in the I-Draft

- Application Rate Limiting for CDN and Edge Cloud Applications
- Time-bound Constrained Flow Acceleration for Large Data Set Transfers
- Application Performance Optimization Through AI Modeling
- Optimized Joint Routing and Congestion Control
- Service Placement for Edge Computing
- Training Neural Networks and AI Inference for Edge Clouds, Data Centers and Planet-Scale Networks
- 5G Network Slicing

Bottleneck Structure Graphs (BSG): Use Cases Documented in the I-Draft

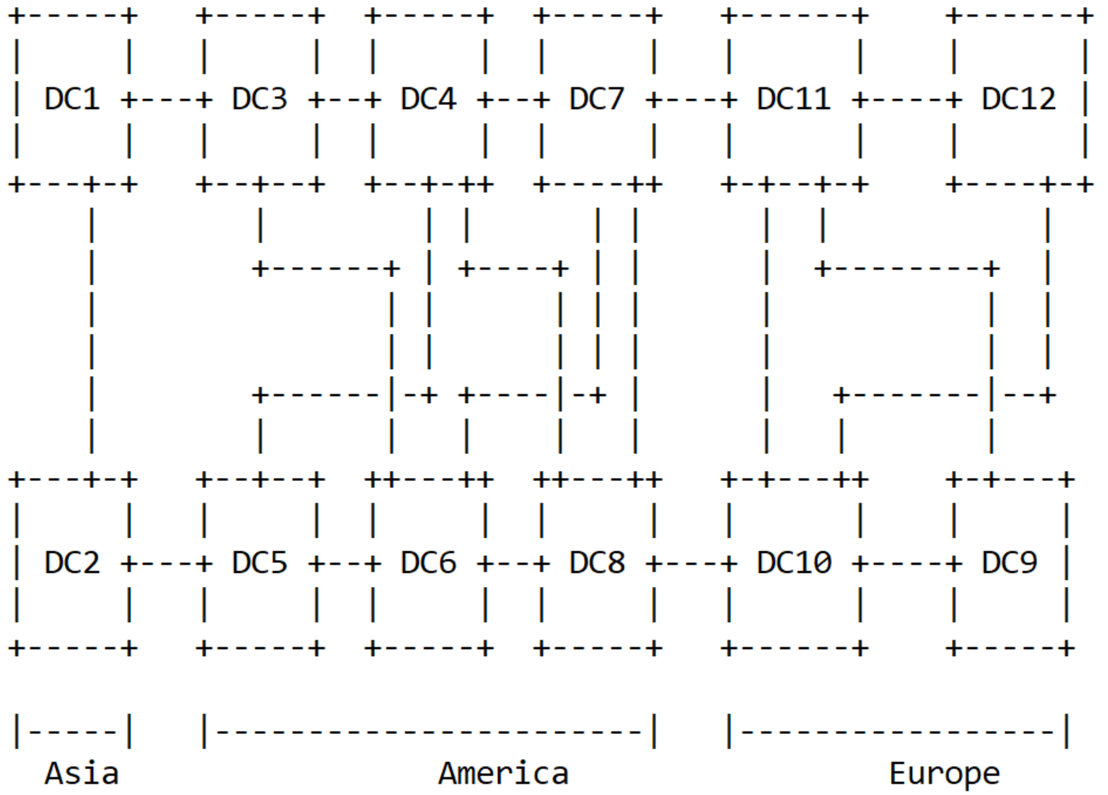
- Application Rate Limiting for CDN and Edge Cloud Applications
- Time-bound Constrained Flow Acceleration for Large Data Set Transfers
- Application Performance Optimization Through AI Modeling
- Optimized Joint Routing and Congestion Control
- Service Placement for Edge Computing
- Training Neural Networks and AI Inference for Edge Clouds, Data Centers and Planet-Scale Networks
- 5G Network Slicing

We will focus on “Optimized Joint Routing and Congestion Control”. For details on all other use cases, see the I-Draft:

https://datatracker.ietf.org/doc/html/draft-giraltyellamraju-alto-bsg-requirements-00

BSG Use Cases: Optimizing Joint Routing and Congestion Control

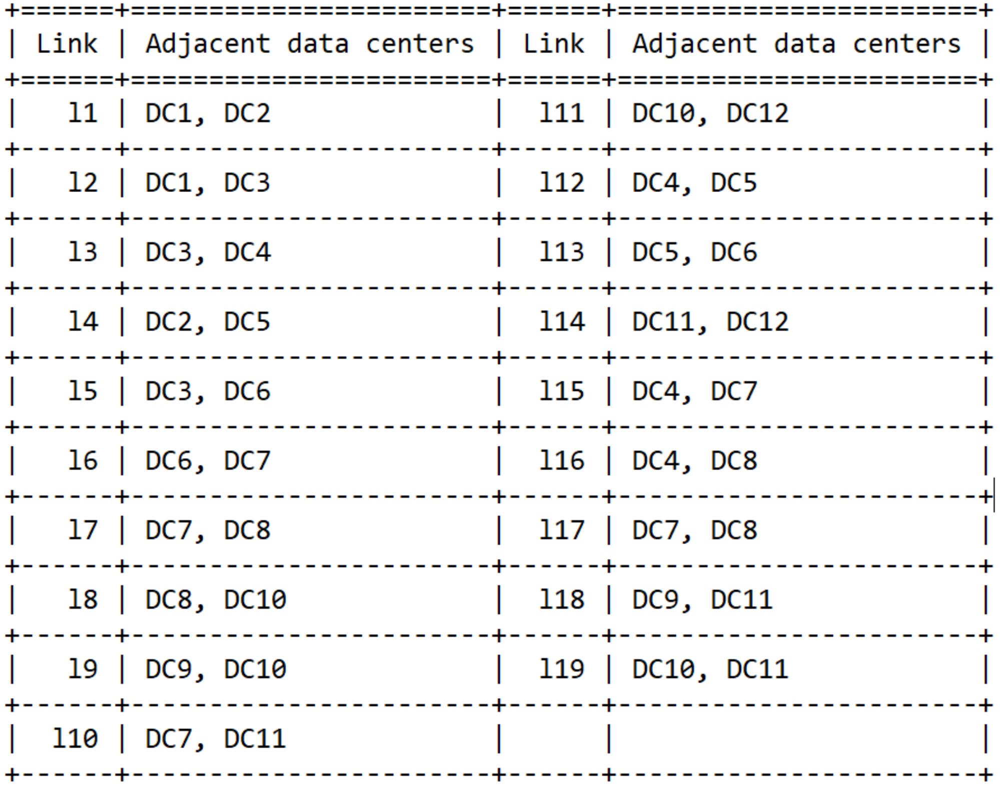
Assume Google's B4 Network from [B4-SIGCOM]:



```
+-----+   +-----+  +-----+  +-----+   +------+    +------+
|     |   |     |  |     |  |     |   |      |    |      |
| DC1 +---+ DC3 +--+ DC4 +--+ DC7 +---+ DC11 +----+ DC12 |
|     |   |     |  |     |  |     |   |      |    |      |
+---+-+   +--+--+  +--+-++  +----++   +-+--+-+    +----+-+
    |        |        | |      | |      |  |           |
    |        +------+ | +----+ | |      |  +--------+  |
    |               | |      | | |      |           |  |
    |               | |      | | |      |           |  |
    |        +------|-+ +----|-+ |      |   +-------|--+
    |        |      |   |    |   |      |   |       |
+---+-+   +--+--+  ++---++  ++---++   +-+---++    +-+---+
|     |   |     |  |     |  |     |   |      |    |     |
| DC2 +---+ DC5 +--+ DC6 +--+ DC8 +---+ DC10 +----+ DC9 |
|     |   |     |  |     |  |     |   |      |    |     |
+-----+   +-----+  +-----+  +-----+   +------+    +-----+

|-----|   |-----------------------|   |-----------------|
  Asia                 America                 Europe
```

Figure 4: Google's B4 network introduced in [B4-SIGCOMM].

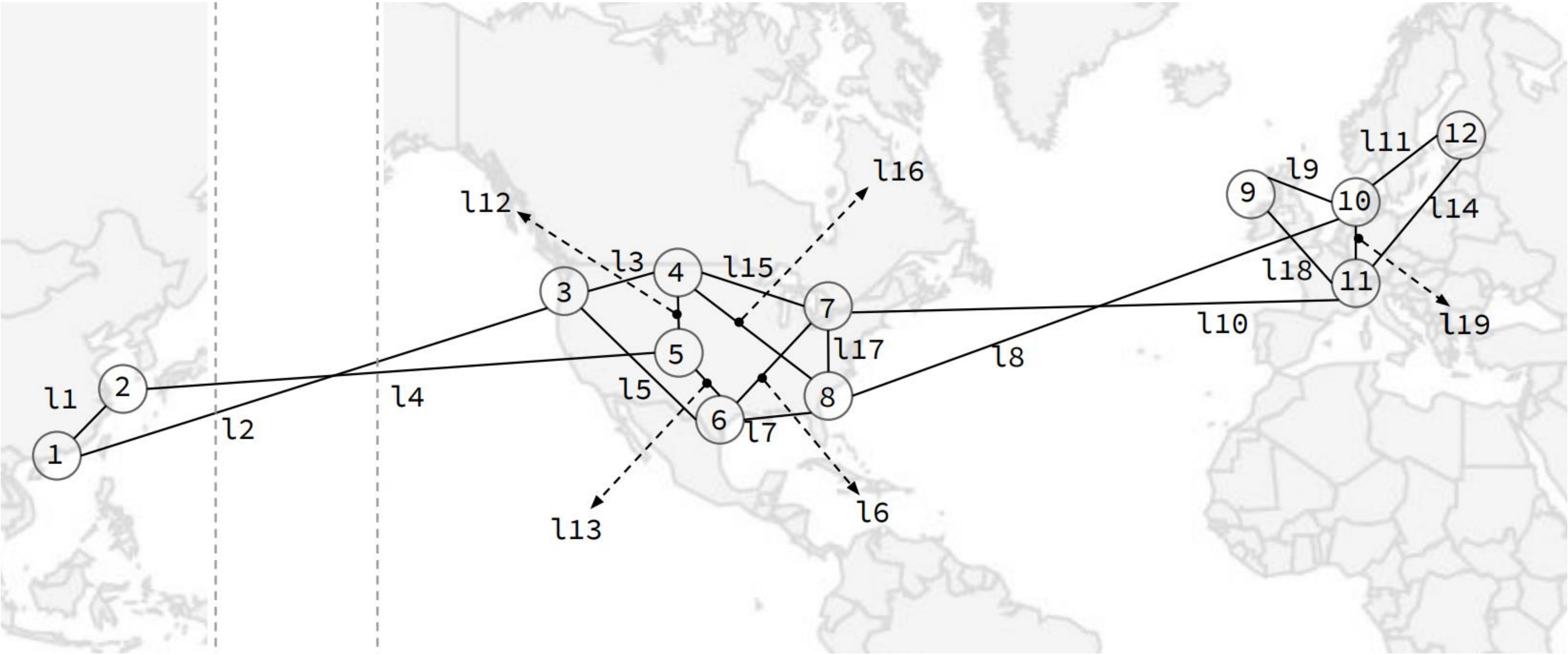


Link	Adjacent data centers	Link	Adjacent data centers
l1	DC1, DC2	l11	DC10, DC12
l2	DC1, DC3	l12	DC4, DC5
l3	DC3, DC4	l13	DC5, DC6
l4	DC2, DC5	l14	DC11, DC12
l5	DC3, DC6	l15	DC4, DC7
l6	DC6, DC7	l16	DC4, DC8
l7	DC7, DC8	l17	DC7, DC8
l8	DC8, DC10	l18	DC9, DC11
l9	DC9, DC10	l19	DC10, DC11
l10	DC7, DC11		

Table 1: Link connectivity (adjacency matrix) in the B4 network.

BSG Use Cases: Optimizing Joint Routing and Congestion Control

Assume Google's B4 Network from [B4-SIGCOM] (human friendly version):



BSG Use Cases: Optimizing Joint Routing and Congestion Control

- Assume a simple configuration with a pair of flows (one for each direction) connecting every data center in the US with every data center in Europe.
- All links are assumed to have a capacity of 10 Gbps except for the transatlantic links (DC7-DC11 and DC8-DC10), which are configured at 25 Gbps.
- Then the bottleneck structure is the graph shown on the right.

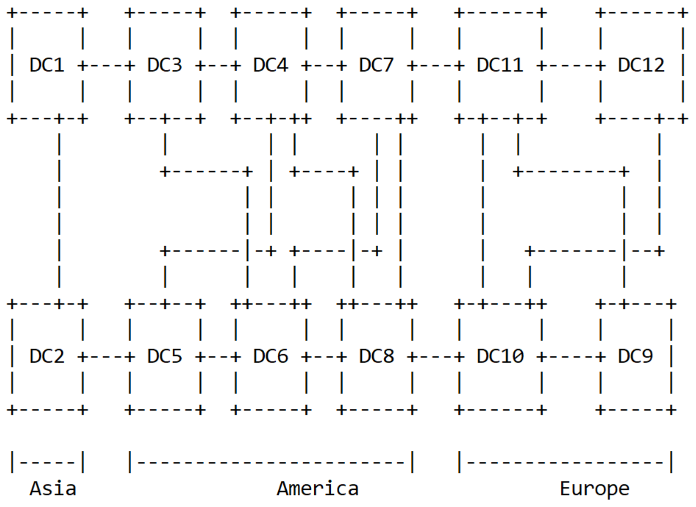


Figure 4: Google's B4 network introduced in [B4-SIGCOMM].

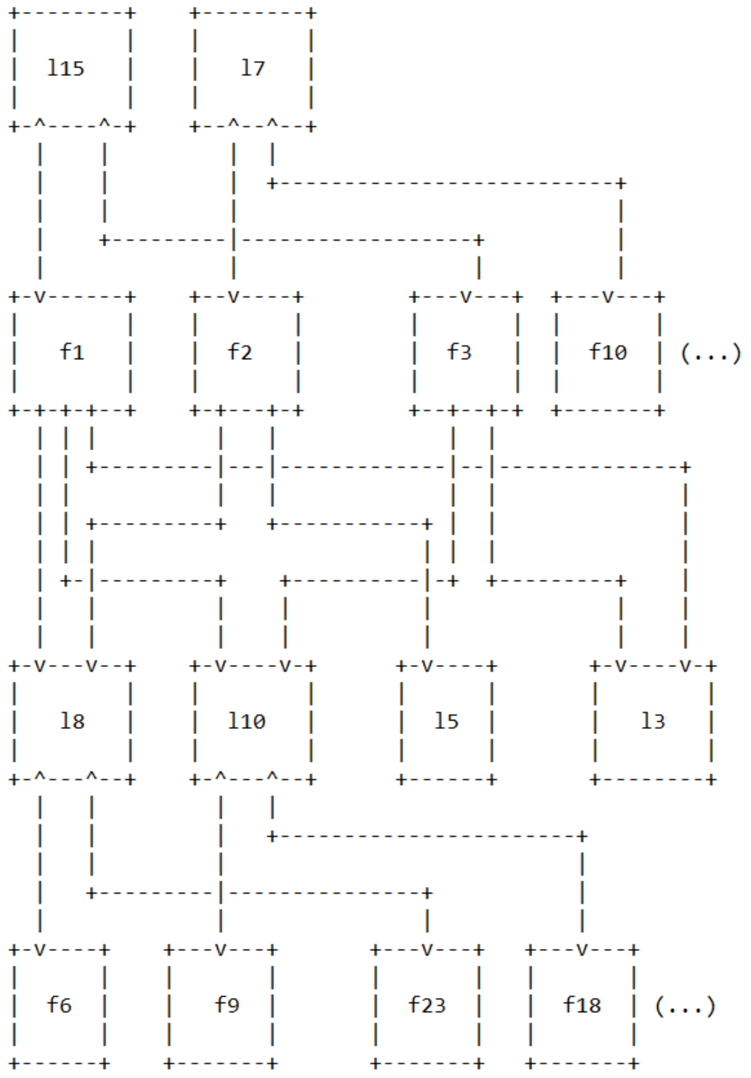
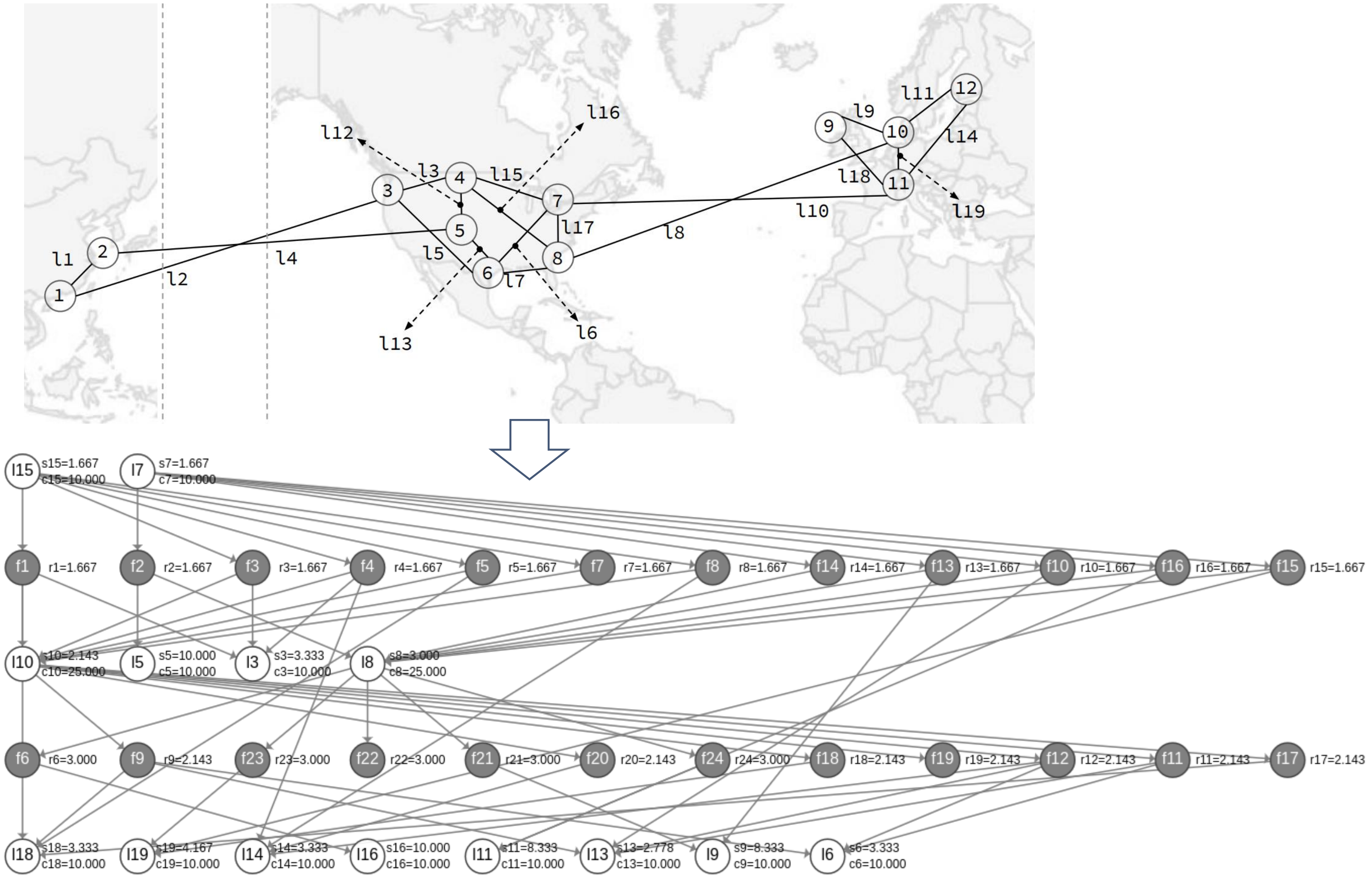


Figure 5: Bottleneck structure of Google's B4 network example.

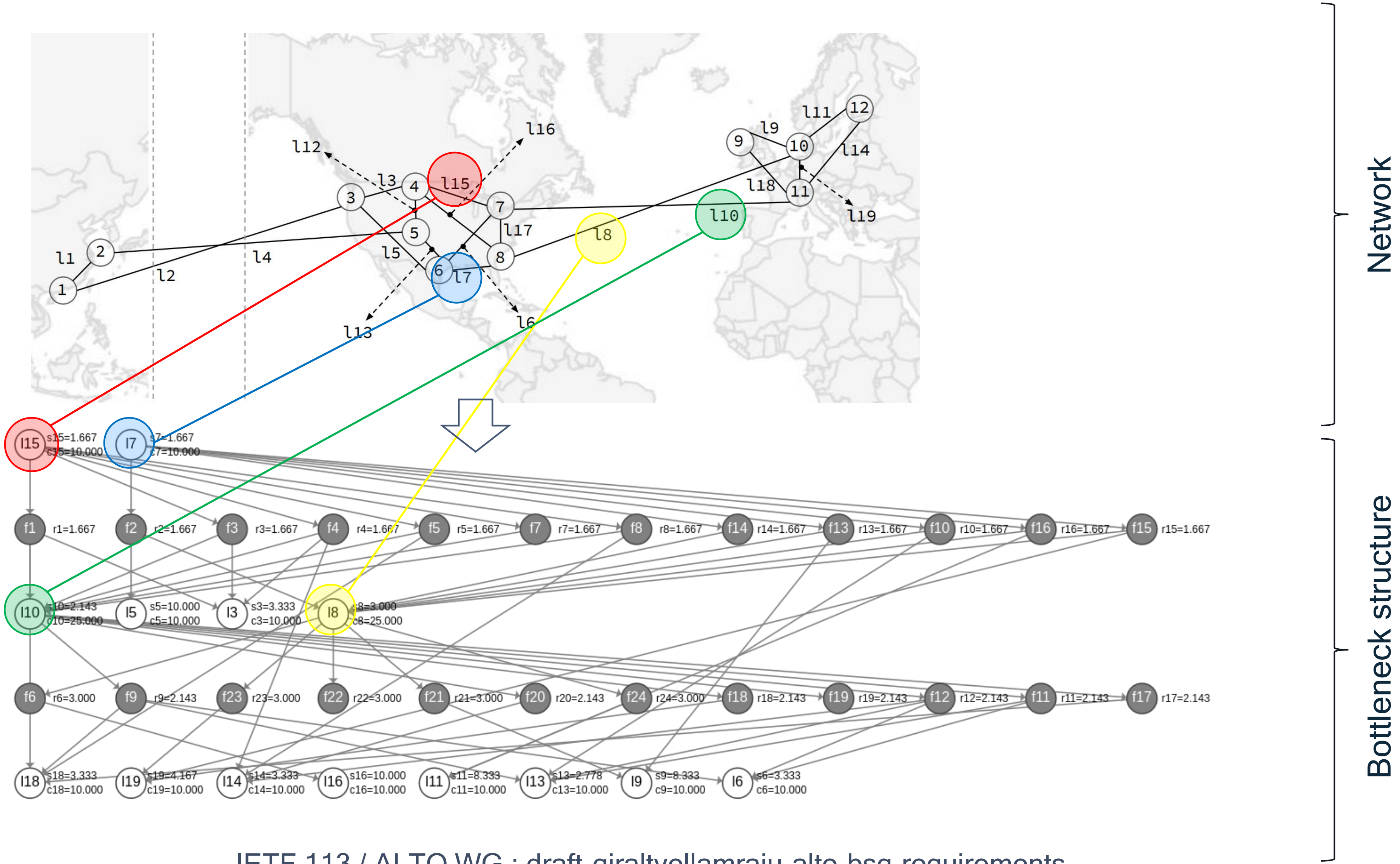
BSG Use Cases: Optimizing Joint Routing and Congestion Control



Network

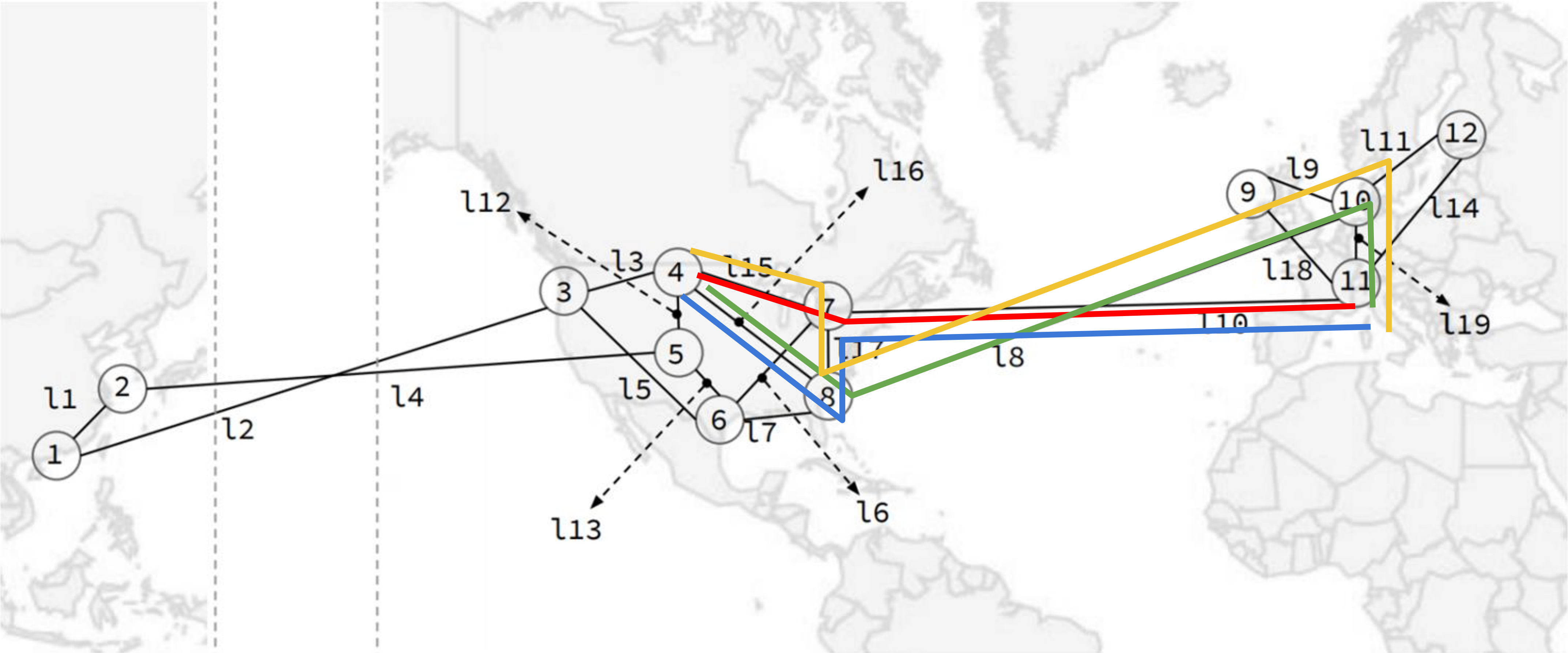
Bottleneck structure

BSG Use Cases: Optimizing Joint Routing and Congestion Control



BSG Use Cases: Optimizing Joint Routing and Congestion Control

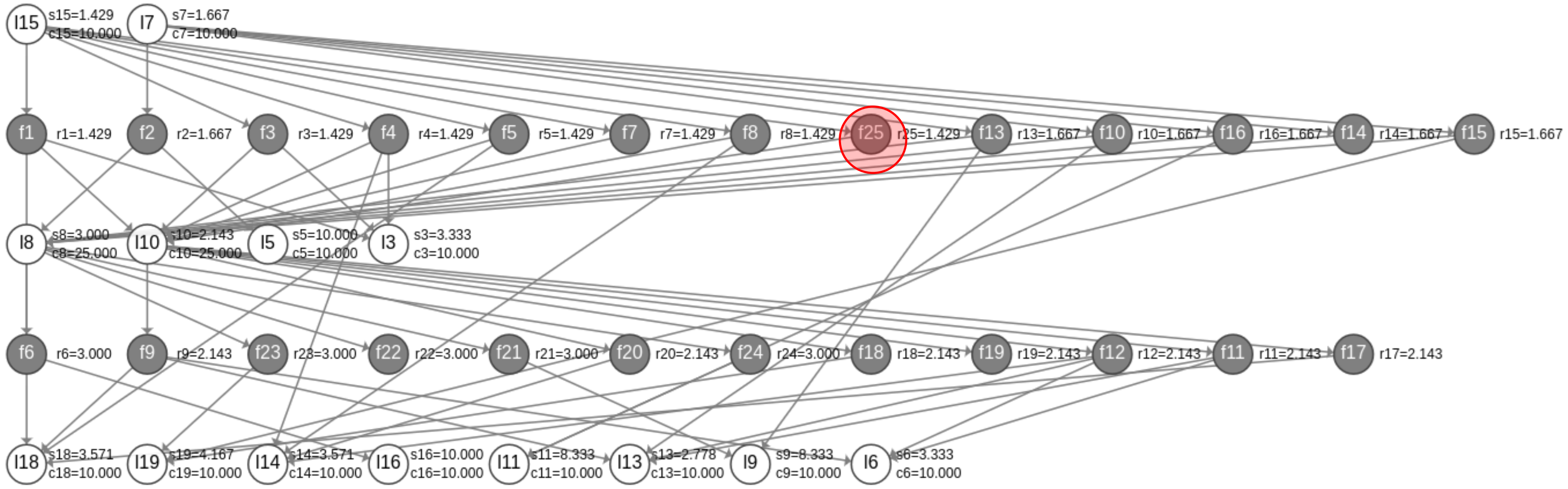
Suppose that an application needs to place a new flow on Google's B4 network to transfer a large data set from data center 11 (DC11) to data center 4 (DC4). There are multiple path choices:



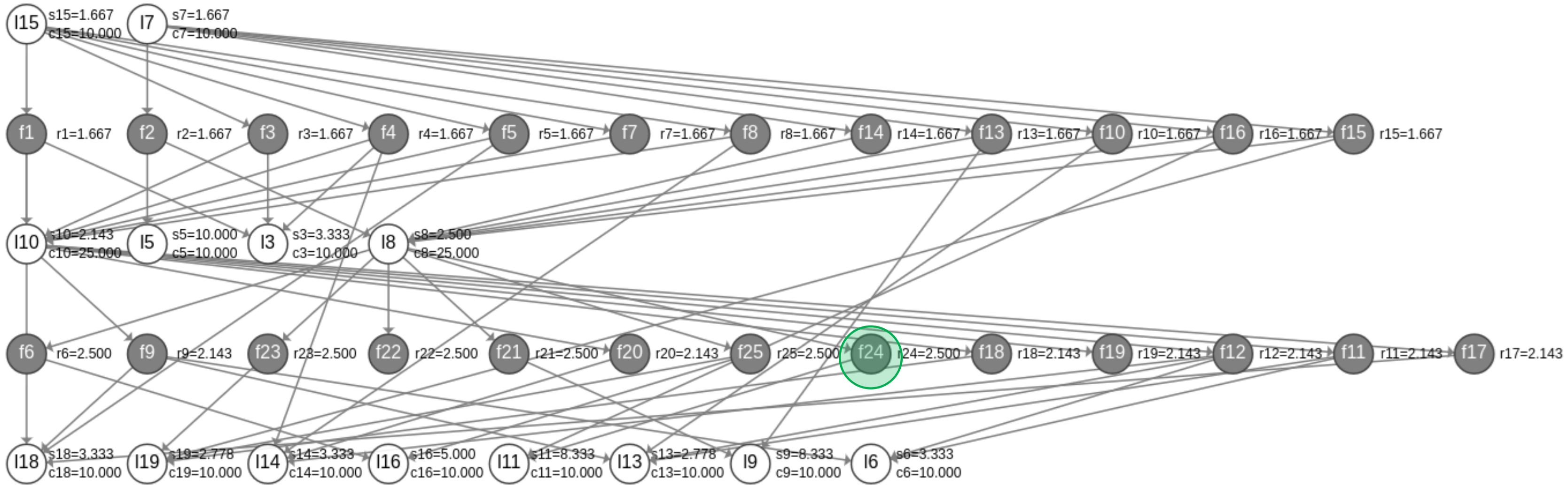
BSG Use Cases: Optimizing Joint Routing and Congestion Control

- Using bottleneck structures, we can compute in O(V+E*log(V)) the path that will yield maximal throughput while considering the reaction of the congestion control algorithm (See [G2-TREP]).
- The optimal path corresponds to DC11 → I19 → DC10 → I8 → DC8 → I16 → DC4 yielding a throughput of 2.5 Gbps.
- Note that this is higher than the shortest path DC11 -> I10 -> DC7 -> I15 -> DC4, which yields a throughput of 1.429 Gbps.
- SLA management: Bottleneck structures can also be used to qualify and quantify the ripple effects produced on all other flows when placing the new flow to ensure their SLAs are preserved. See next slide.

BSG Use Cases: Optimizing Joint Routing and Congestion Control

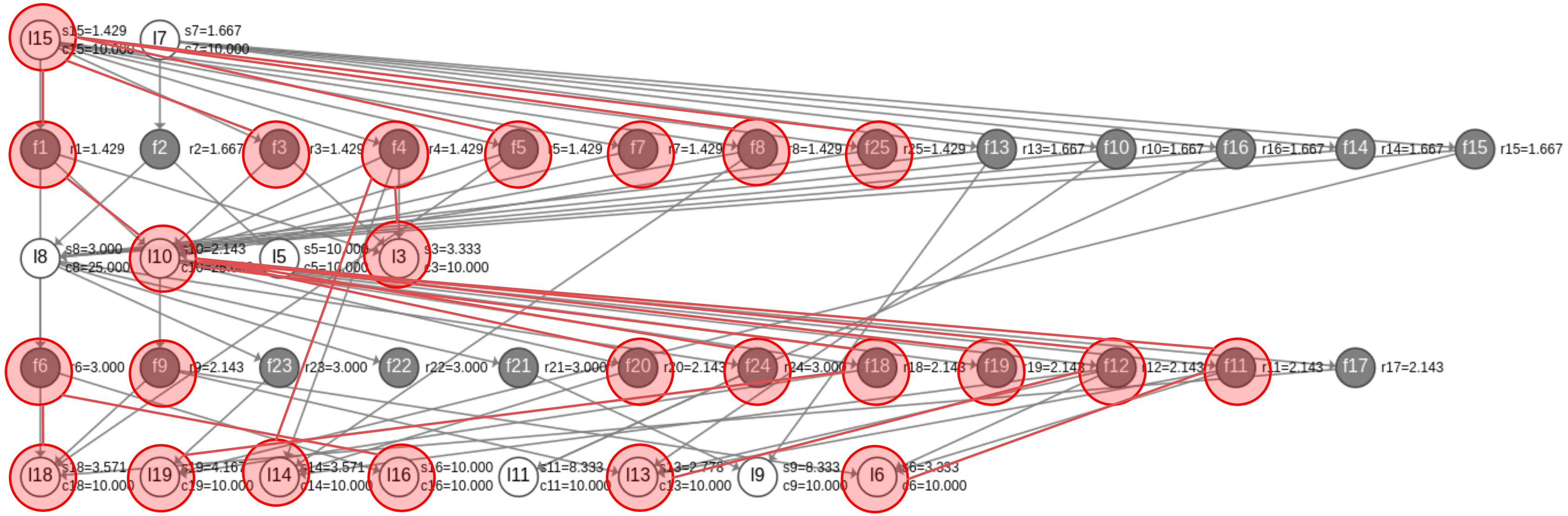


Routing on shortest path



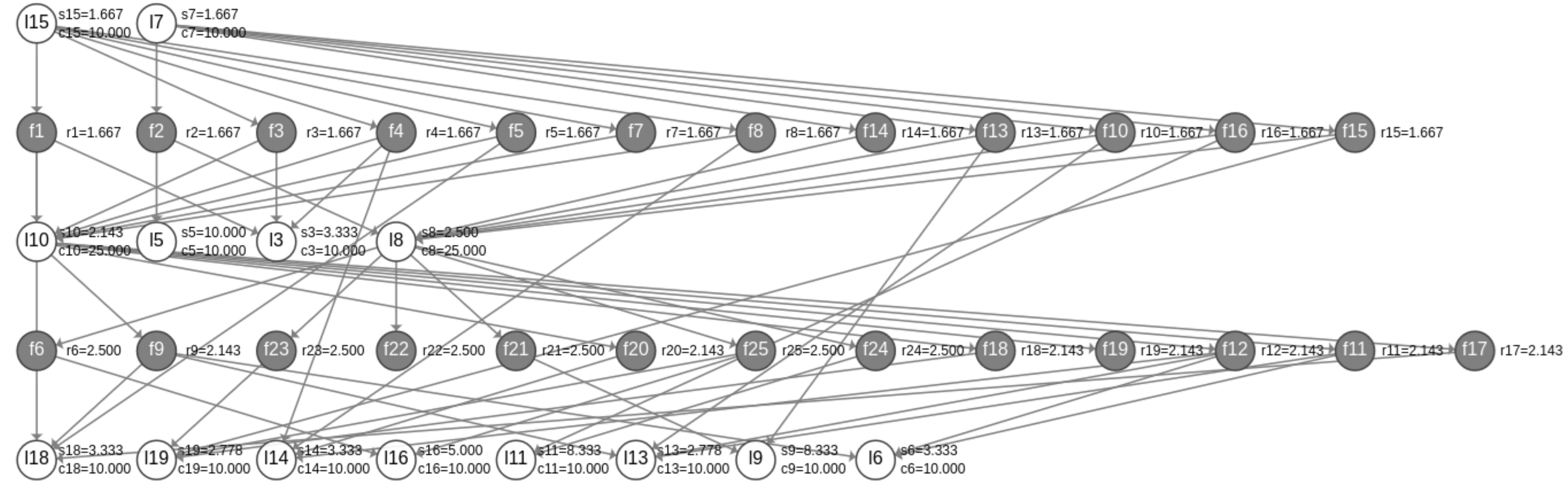
Routing on maximal throughput path

BSG Use Cases: Optimizing Joint Routing and Congestion Control



Routing on shortest path

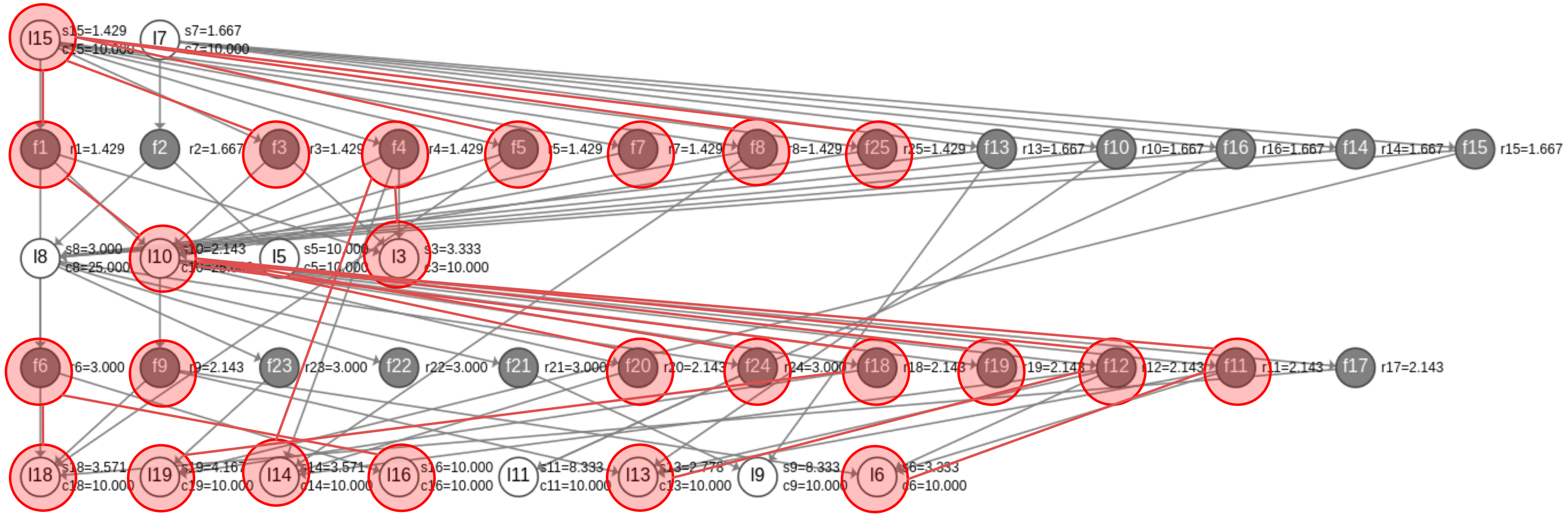
🔴 Flows affected when placing flow on shortest path



Routing on maximal throughput path

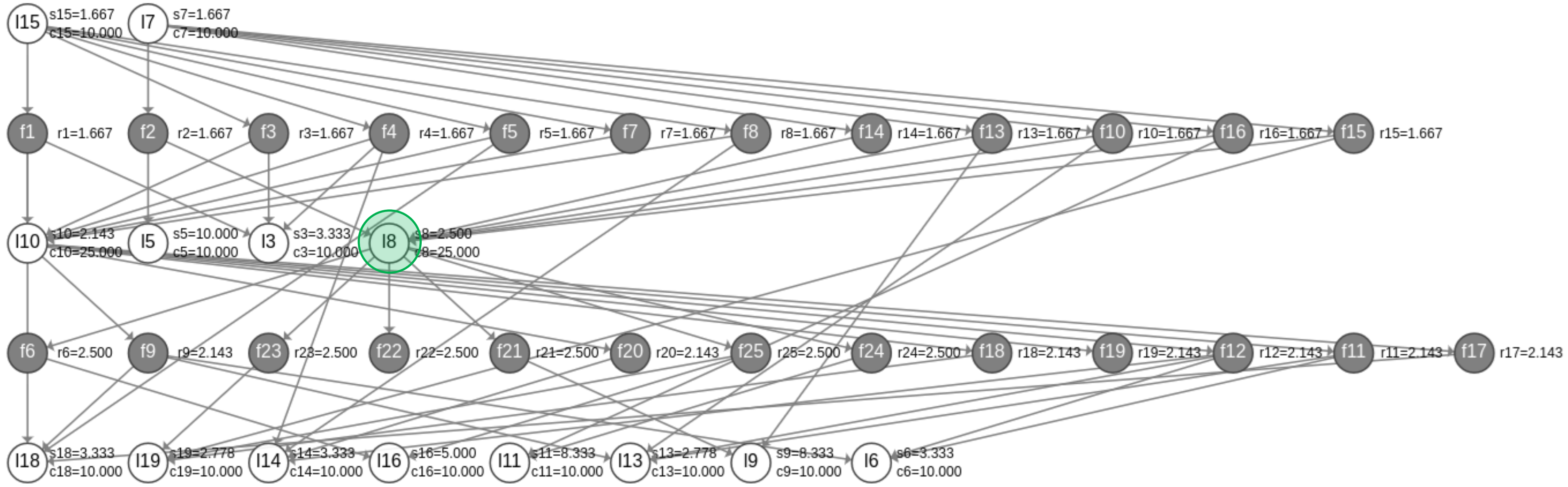
🟢 Flows affected when placing flow on maximal throughput path

BSG Use Cases: Optimizing Joint Routing and Congestion Control



Routing on shortest path

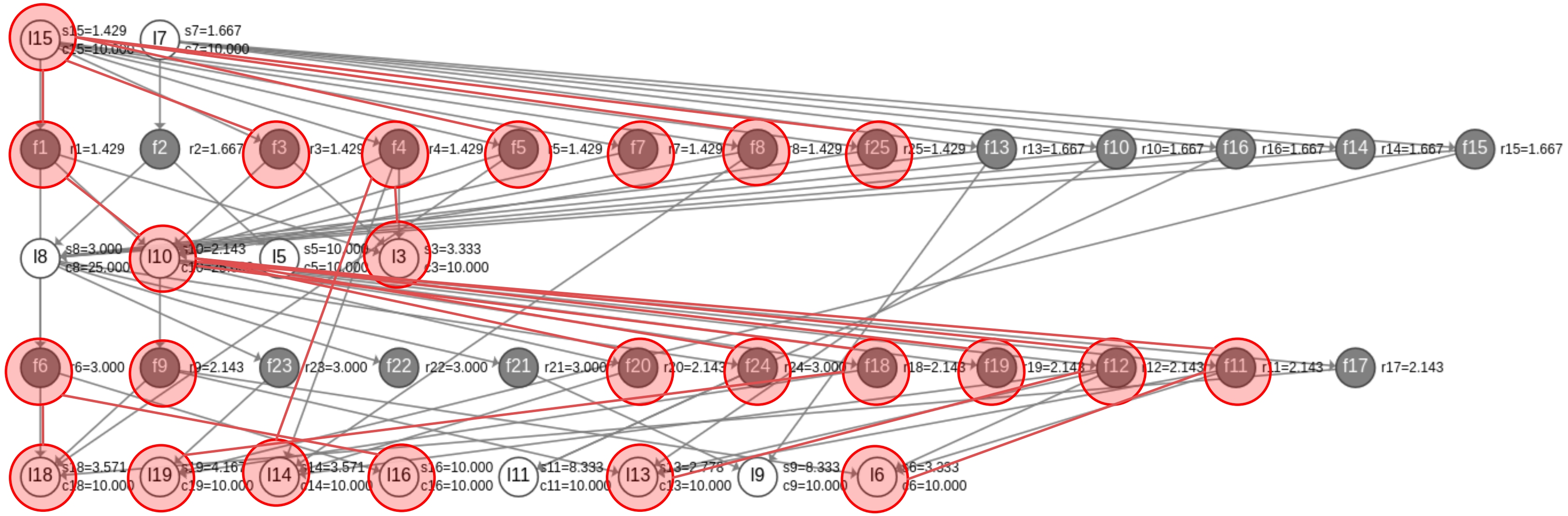
🔴 Flows affected when placing flow on shortest path



Routing on maximal throughput path

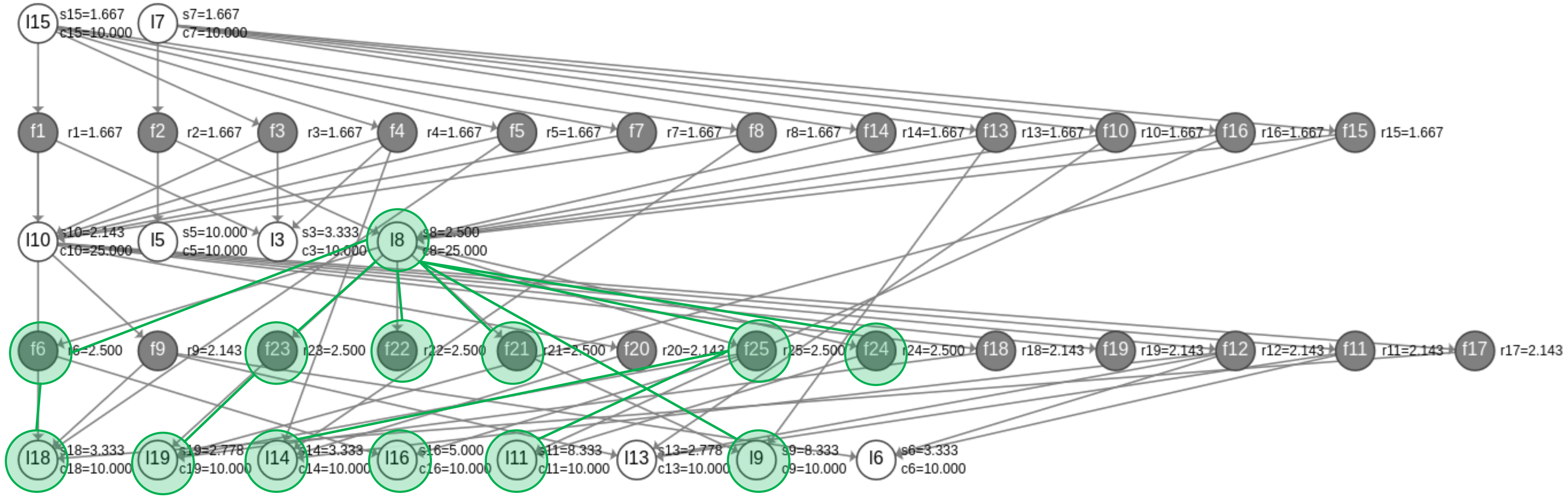
🟢 Flows affected when placing flow on maximal throughput path

BSG Use Cases: Optimizing Joint Routing and Congestion Control



Routing on shortest path

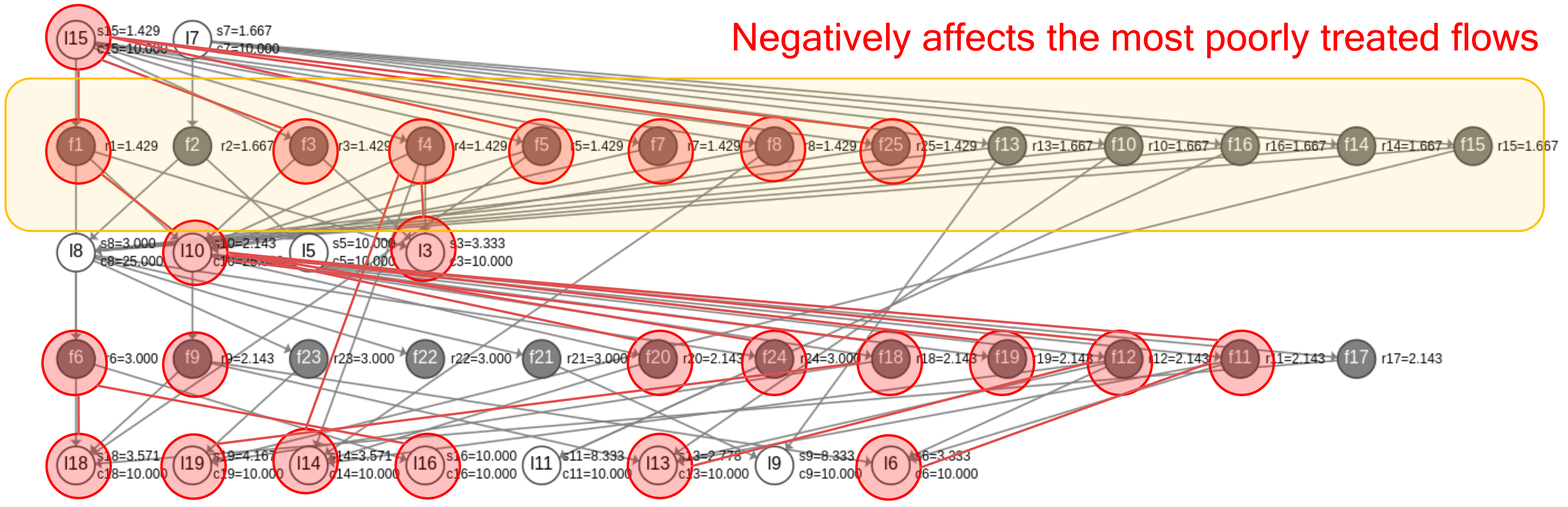
● Flows affected when placing flow on shortest path



Routing on maximal throughput path

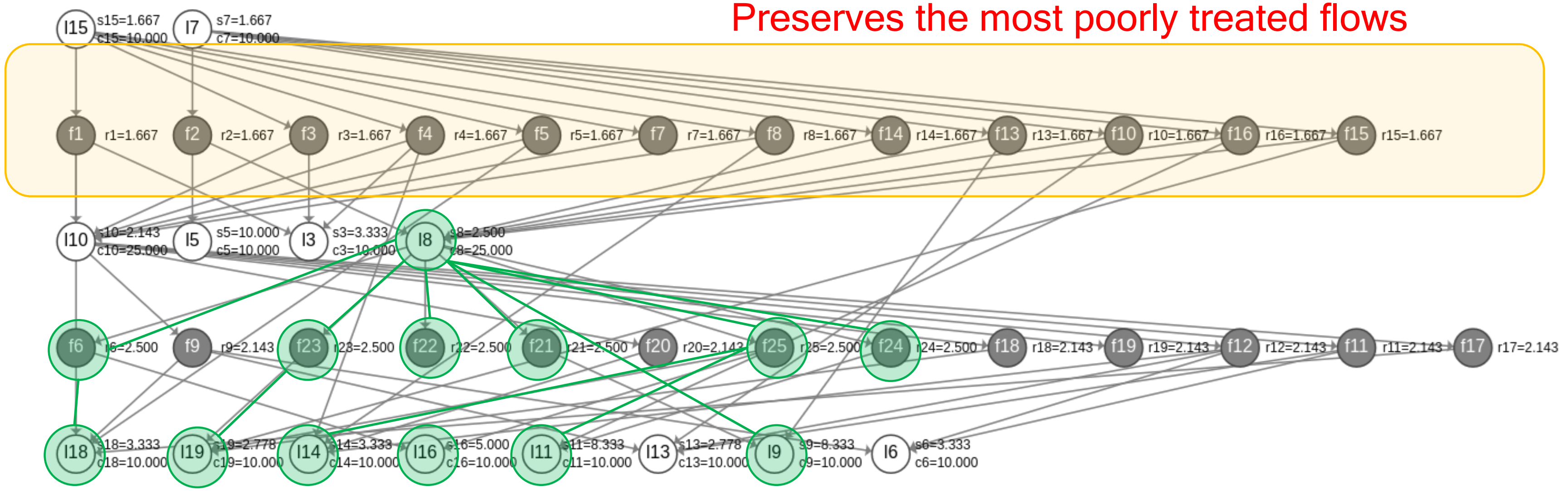
● Flows affected when placing flow on maximal throughput path

BSG Use Cases: Optimizing Joint Routing and Congestion Control


Negatively affects the most poorly treated flows

Routing on shortest path

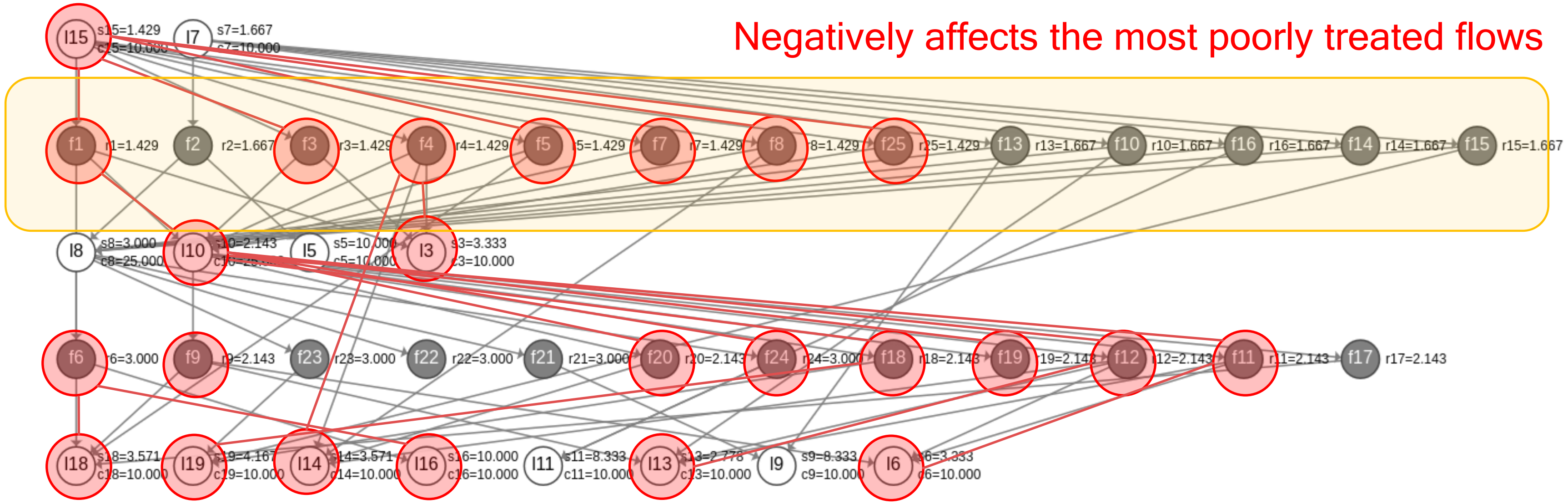
🔴 Flows affected when placing flow on shortest path


Preserves the most poorly treated flows

Routing on maximal throughput path

🟢 Flows affected when placing flow on maximal throughput path

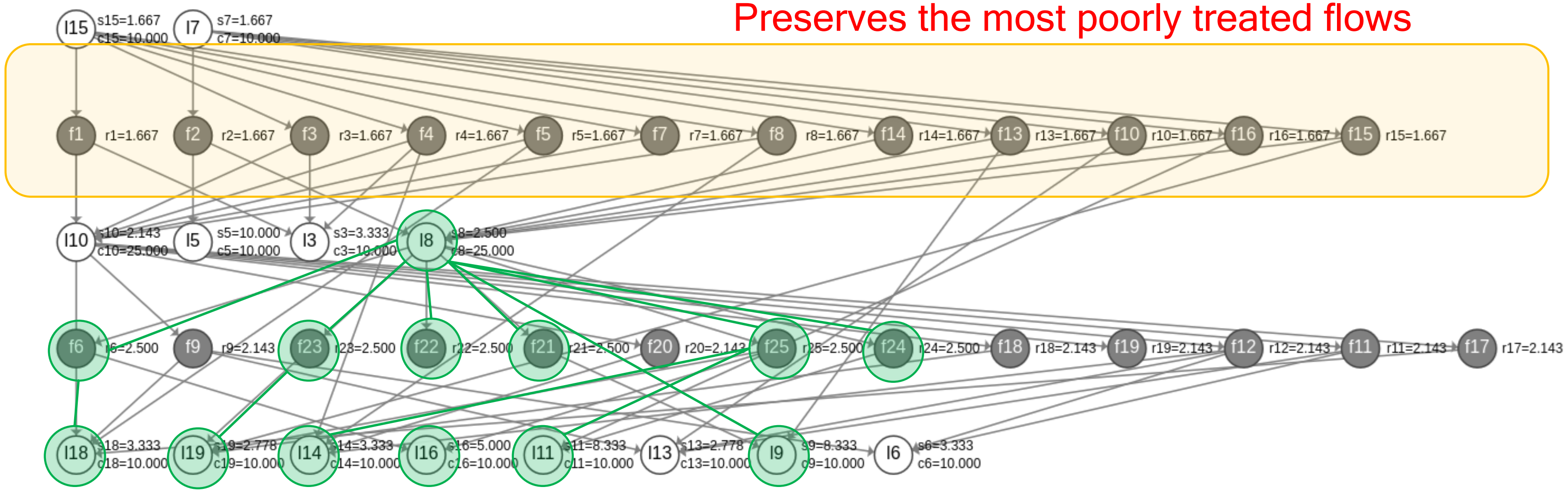
BSG Use Cases: Optimizing Joint Routing and Congestion Control



Negatively affects the most poorly treated flows

Routing on shortest path

● Flows affected when placing flow on shortest path



Preserves the most poorly treated flows

Routing on maximal throughput path

● Flows affected when placing flow on maximal throughput path

Introducing Bottleneck Structures in the IETF ALTO WG: Requirements

ALTO Requirements to Support Bottleneck Structures

Requirement 1: Bottleneck Structure Graph (BSG) Abstraction

- **Requirement 1A (R1A).** The ALTO server MUST compute the bottleneck structure graph to allow applications optimize their performance using the BSG service.
- **Requirement 1B (R1B).** The ALTO server MUST at least support the computation of one bottleneck structure type from Section 3.7. Depending on the network information available (e.g., presence of QoS class information), the ALTO server MAY support all the three bottleneck structure types, in which case the ALTO client MAY be able to choose the bottleneck structure type for retrieval. Also, it is RECOMMENDED that the ALTO server supports the computation of the path gradient graph (PGG) as the default bottleneck structure implementation for retrieval by the ALTO clients.

ALTO Requirements to Support Bottleneck Structures

Requirement 2: Information Received from the Network

- **Topology Object (T).** The T Object is a data structure that includes:
  (1) A Topology Graph (V, E), where V is the set of routers and E is the set of links connecting the routers in the network.
  (2) A Capacity Dictionary (C), a key-value table mapping each link with its capacity (in bps).

- **Flow Dictionary (F).** The F Dictionary is a key-value table mapping every flow with the set of links it traverses.

- **Requirement 2A (R2A).** The ALTO server MUST collect information about (1) the set of routers and links in a network, (2) the capacity of each link and (3) the set of links traversed by each flow.

ALTO Requirements to Support Bottleneck Structures

Requirement 3: Information Passed to the Application

- **Requirement 3A (R3A).** The ALTO client MUST be able to query the ALTO server to obtain the current bottleneck structure of the network, represented as a digraph.
- **Requirement 3B (R3B).** One or more ALTO services (the Network Map, the Cost Map, the Entity Property Map or the Endpoint Cost Map) MUST support reporting to ALTO clients additional network state information derived from the bottleneck structure to the ALTO client.

ALTO Requirements to Support Bottleneck Structures

Requirement 4: Features Needed to Support the Use Cases

- **Requirement 4A (R4A).** The ALTO BSG service MUST be able to compute the effect of network reconfigurations using bottleneck structure analysis and according to the types described in Section 3.9.

- **Requirement 4B (R4B).** The BSG service MUST be able to update the bottleneck structure graph in near-real time, at least once a minute or less.

References

- [G2-SIGCOMM] Ros-Giralt, J., Amsel, N., Yellamraju, S., Ezick, J., Lethin, R., Jiang, Y., Feng, A., Tassiulas, L., Wu, Z., and K. Bergman, "Designing data center networks using bottleneck structures", ACM SIGCOMM , 2021.
- [G2-TREP] Ros-Giralt, J., Amsel, N., Yellamraju, S., Ezick, J., Lethin, R., Jiang, Y., Feng, A., Tassiulas, L., Wu, Z., and K. Bergman, "A Quantitative Theory of Bottleneck Structures for Data Networks", Reservoir Labs (Qualcomm) Technical Report , 2021.
- [G2-SIGMETRICS] Ros-Giralt, J., Bohara, A., Yellamraju, S., Langston, H., Lethin, R., Jiang, Y., Tassiulas, L., Li, J., Tan, Y., and M. Veeraraghavan, "On the Bottleneck Structure of Congestion-Controlled Networks", ACM SIGMETRICS , 2020.
- [B4-SIGCOMM] Jain et al, S., "B4: Experience with a Globally-Deployed Software Defined WAN", ACM SIGCOMM , 2013.

Discussion Q&A

Thank you