TATA CONSULTANCY SERVICES
Experience certainty.

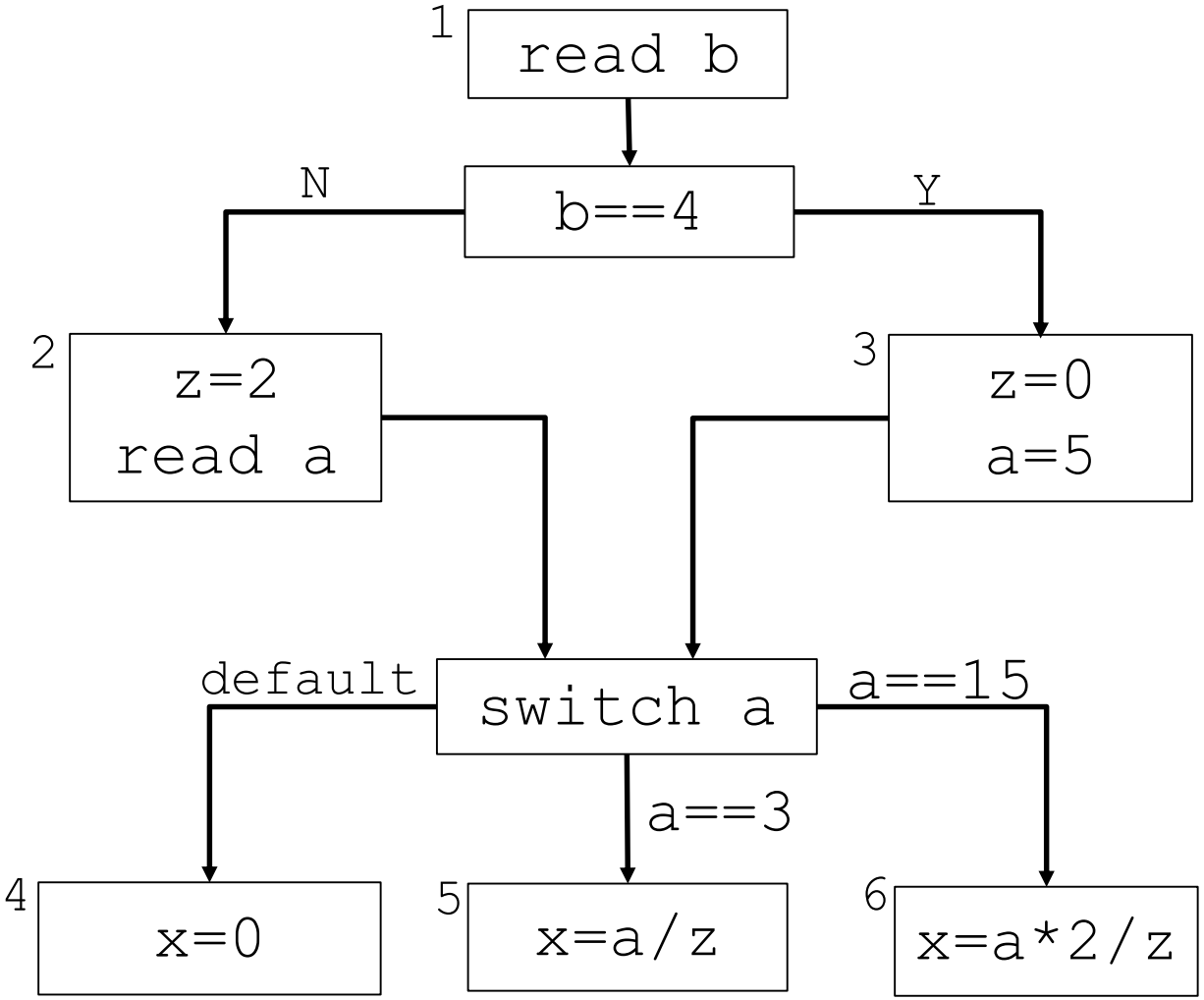
Path Sensitive MFP Solutions In Presence of Intersecting Infeasible Control Flow Path Segments

Komal Pathade (TCS Research, India)
and
Uday Khedker (IIT Bombay, India)

Outline

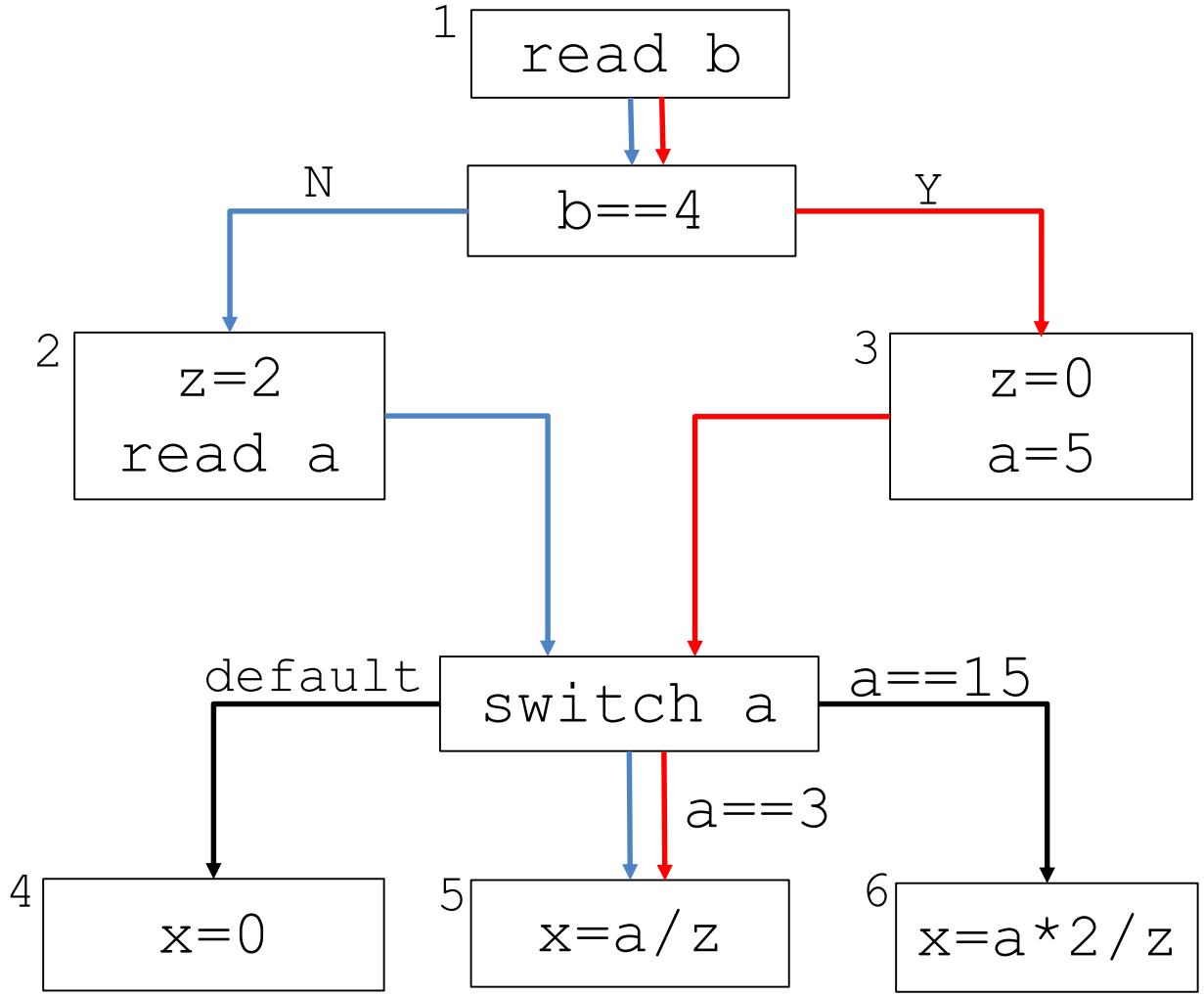
- Background : Solutions of Data Flow Analysis
- Motivation
- Our Solution
- Experiments and Results
- Direction of Future Work

Data Flow Analysis



Values of *z at node 5 ?*

Meet Over Paths (MOP) Solutions

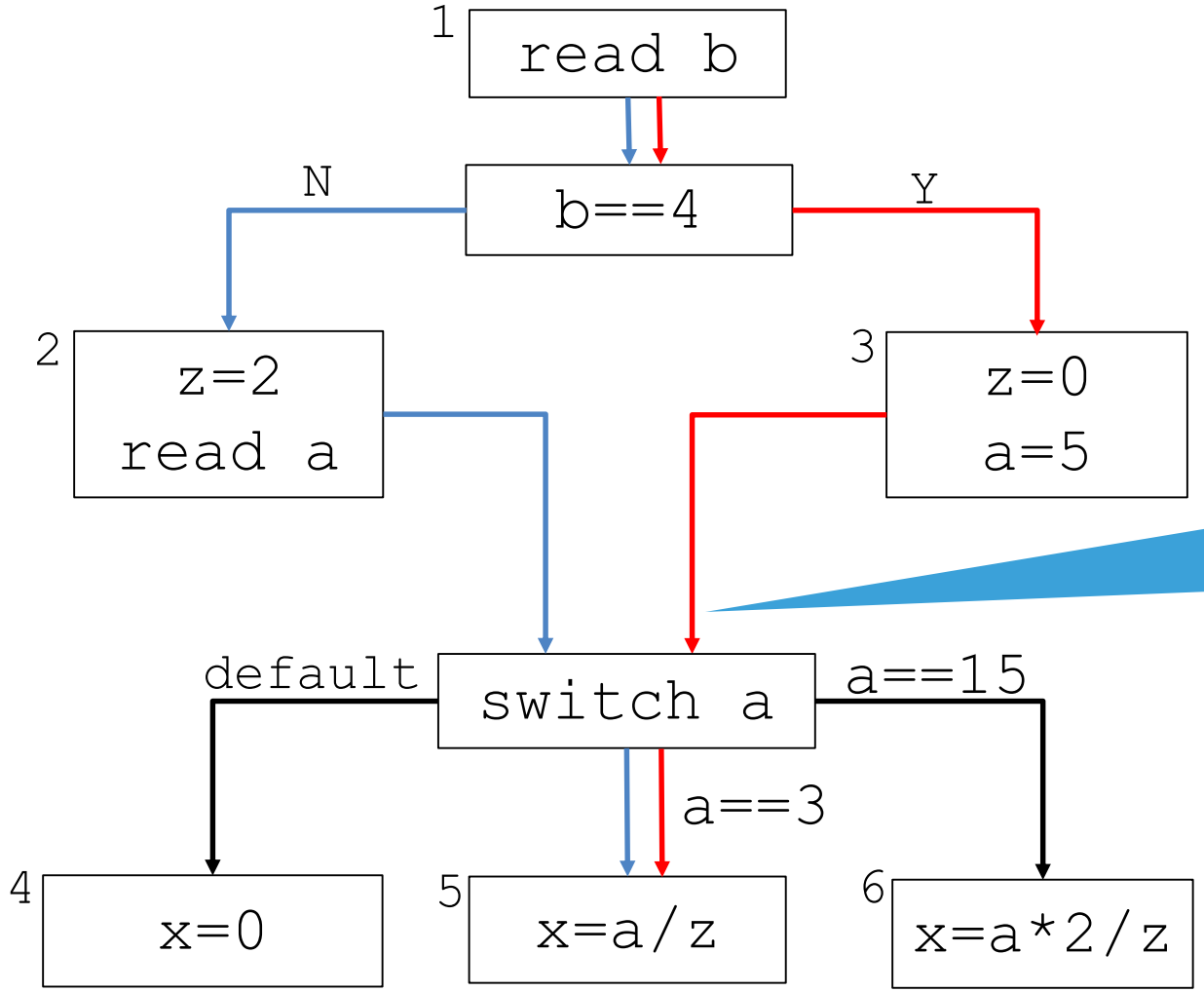


Values of *z at node 5 ?*

- Computes value along each path separately

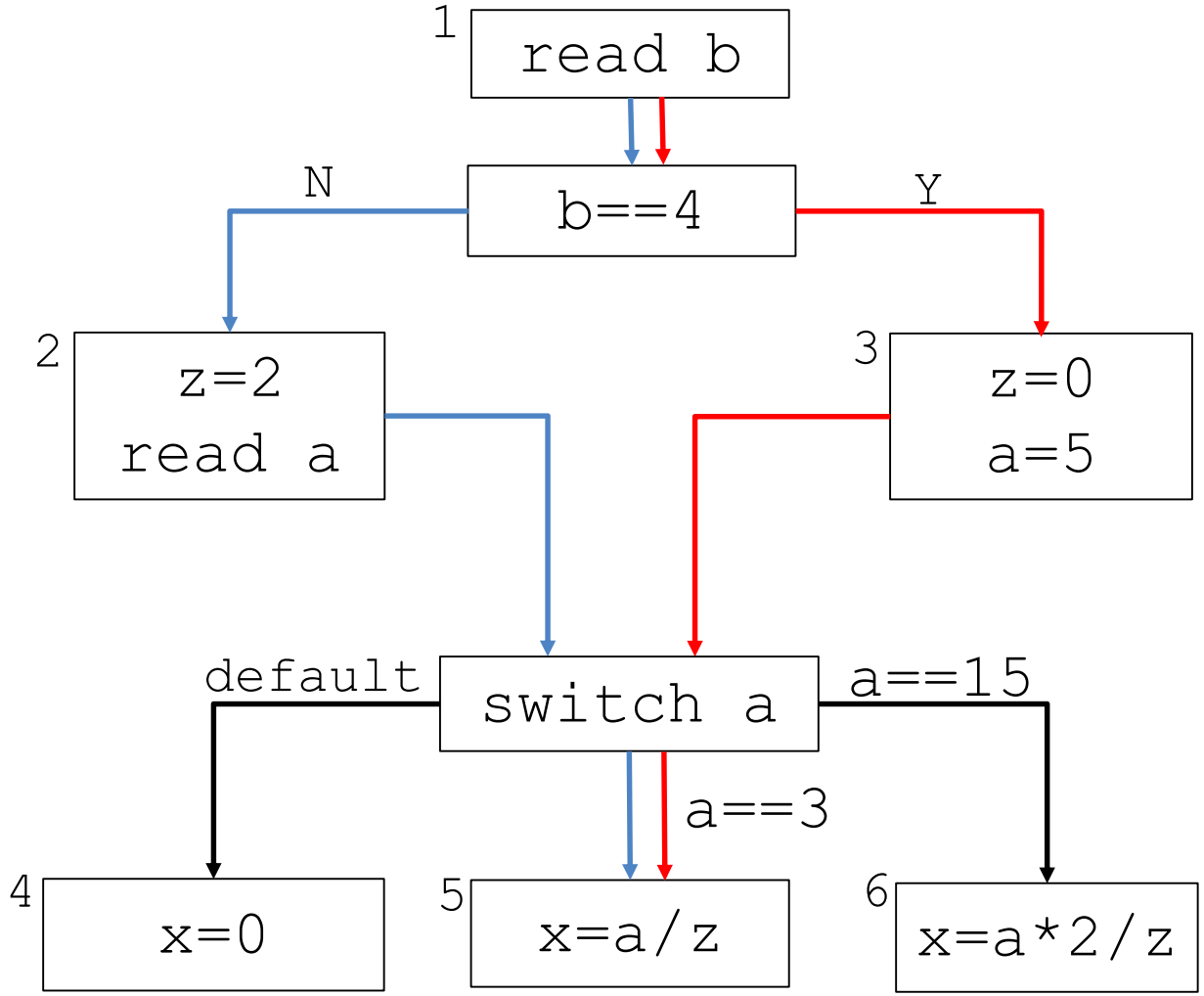
Node	Value of z from paths	
	Blue	Red
5	z=2	z=0

Meet Over Paths (MOP) Solutions



- An Infeasible Path
- Approaches exist to detect such paths

Meet Over Paths (MOP) Solutions

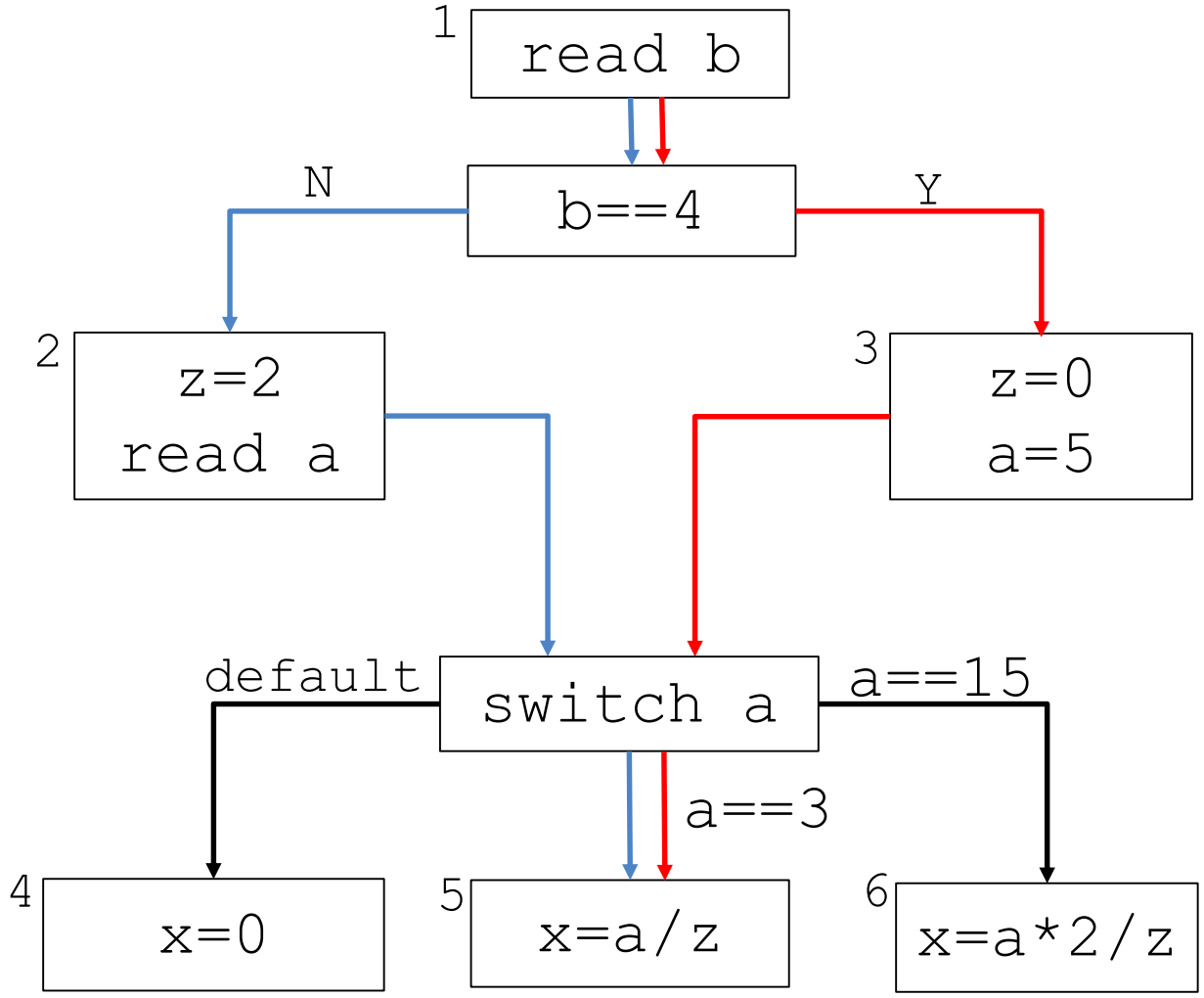


Values of *z at node 5 ?*

- Computes value along each path separately

Node	Value of z from paths	
	Blue	Red
5	z=2	z=0

Meet Over Paths (MOP) Solutions



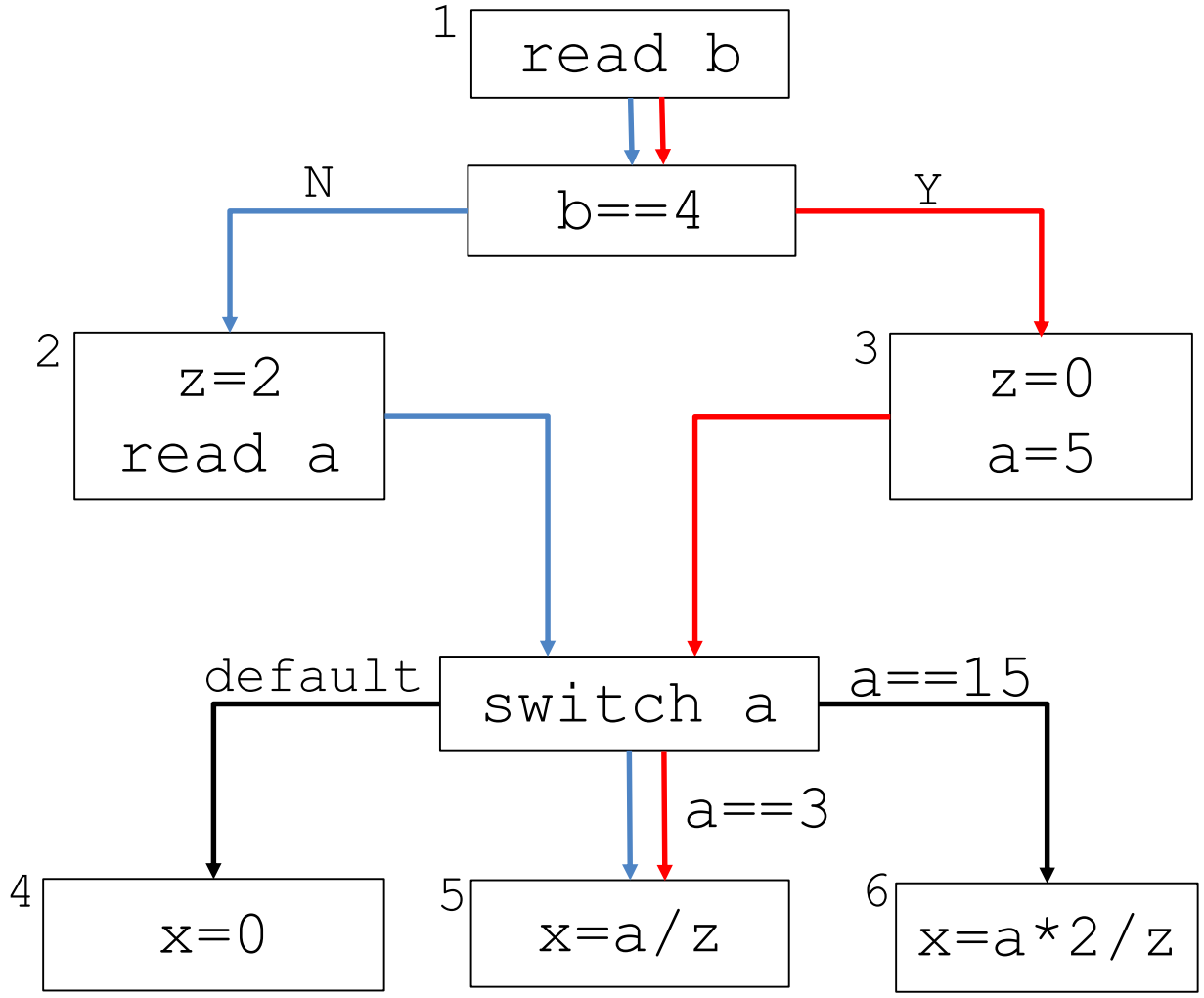
Values of *z at node 5 ?*

- Computes value along each path separately

Node	Value of z from paths	
	Blue	Red
5	z=2	z=0

- Red path is infeasible so z=0 is discarded, so z=2. (Precise)
- Not Scalable: maintains values of the order of O(number of paths * number of nodes)
- Number of paths is exponential.

Maximum Fix Point (MFP) Solutions



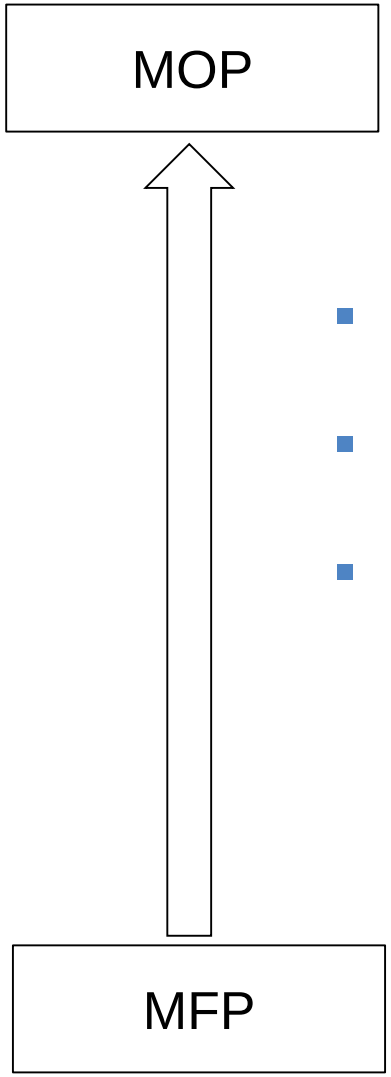
Values of *z at node 5 ?*

- Compute range of values reaching along all paths

Node	Value Range of z from all paths
5	z=[0,2]

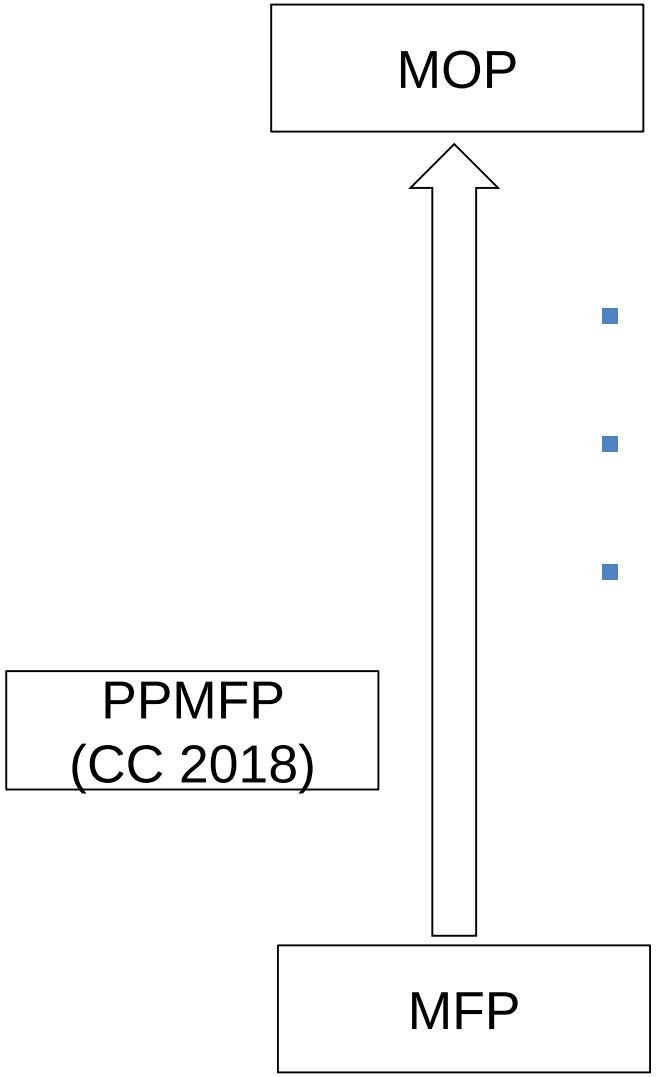
- Scalable: computes only one value per program point
- Imprecise: z=0 is included which reaches along infeasible path (marked in red)

Scalability and Precision Trade-off



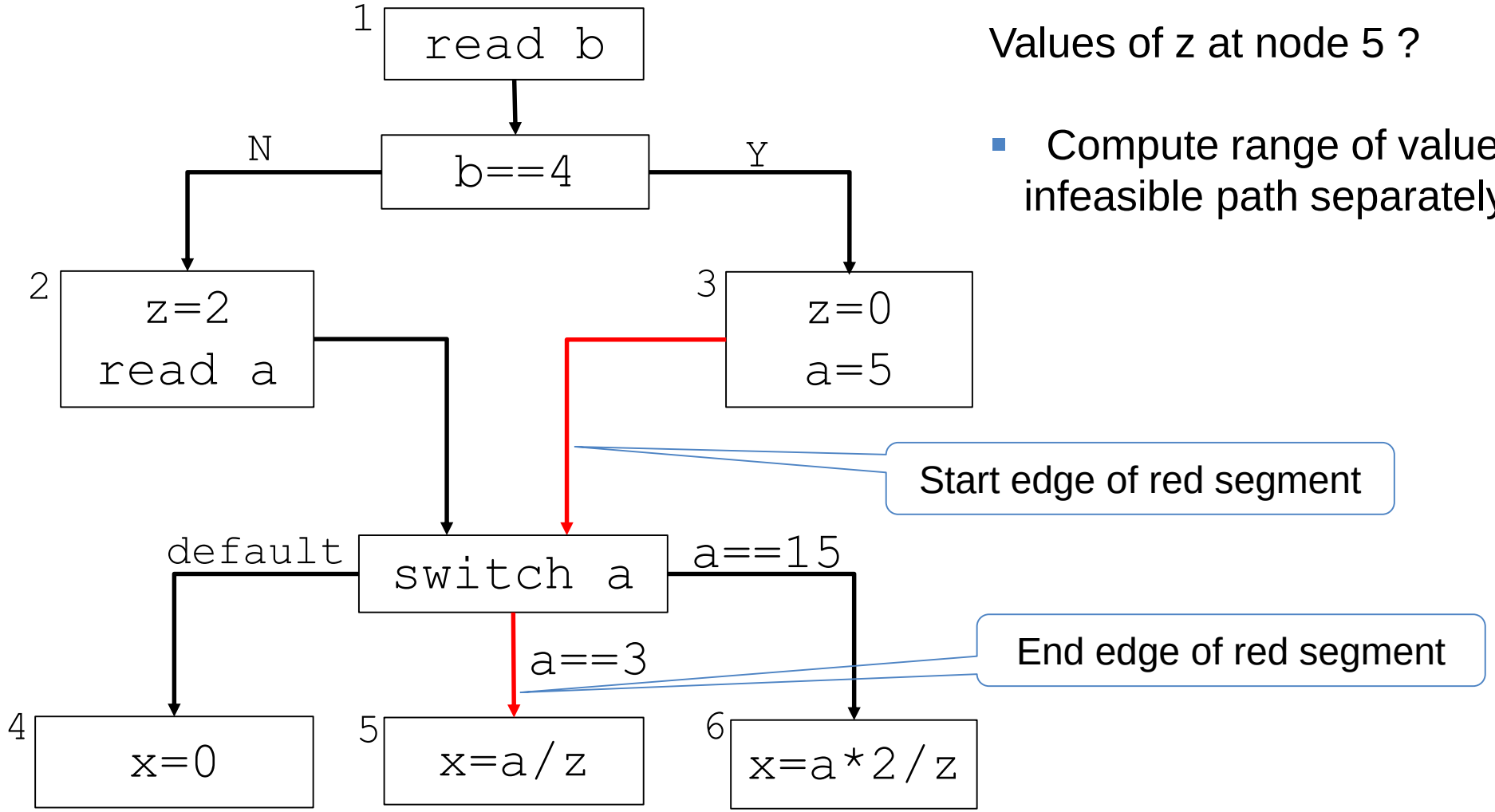
- Number of values maintained increases
- Precision Increases
- Scalability Decreases

Partially Path Sensitive Solutions



- Number of values maintained increases
- Precision Increases
- Scalability Decreases

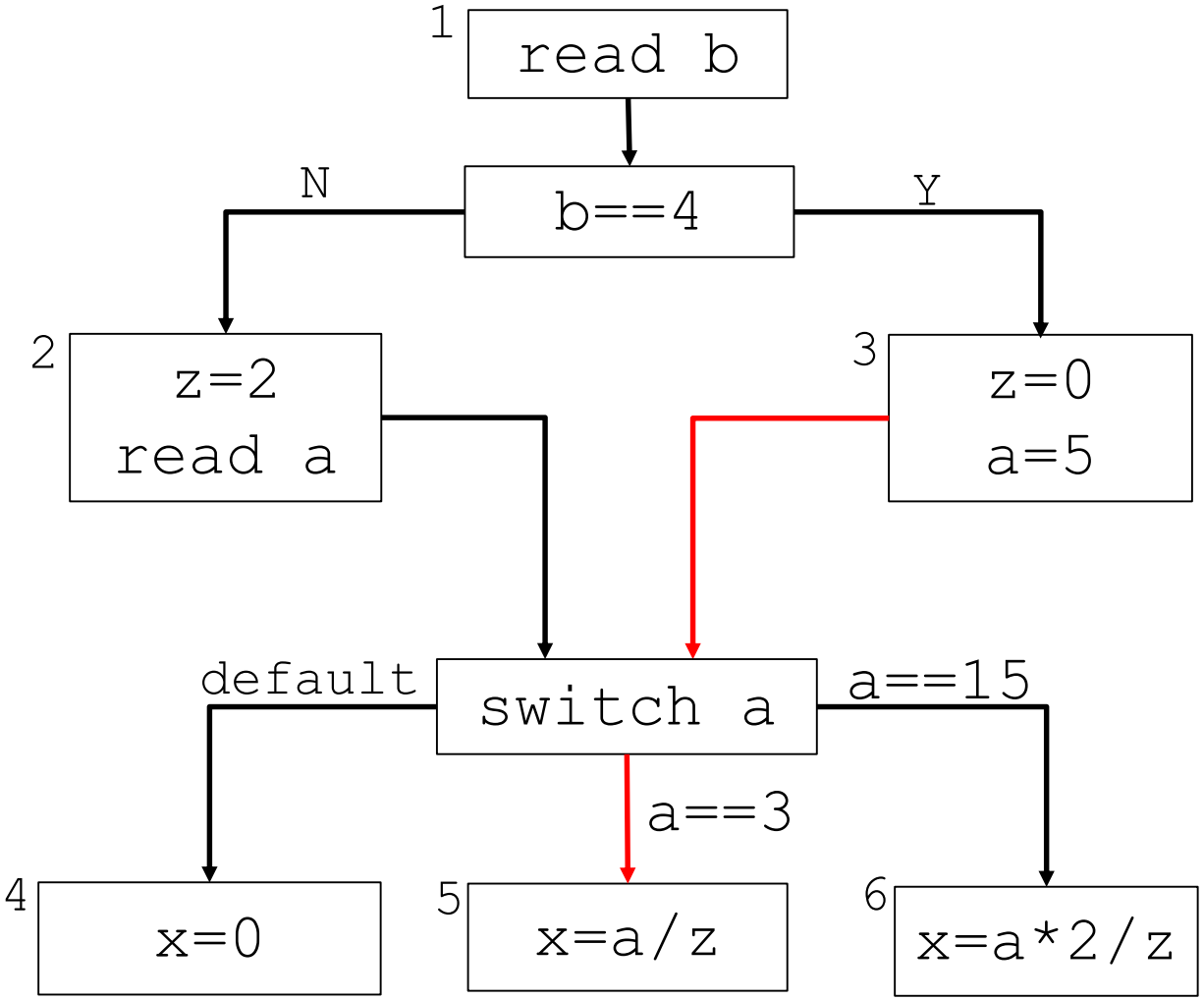
Prior Work: PPMFP (CC 2018)



Values of z at node 5 ?

- Compute range of values along feasible and infeasible path separately

Prior Work: PPMFP (CC 2018)

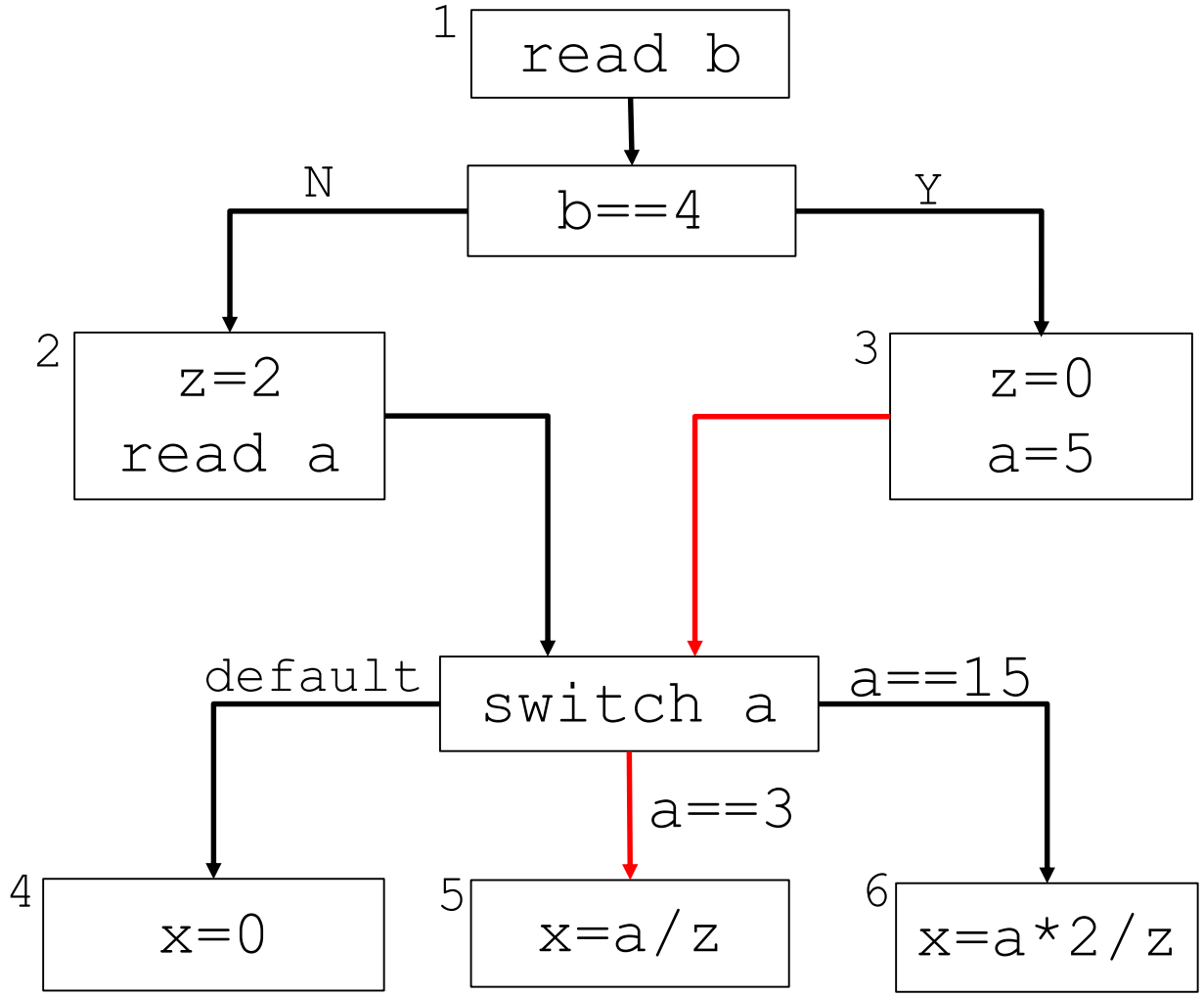


Values of z at node 5 ?

- Compute range of values along feasible and infeasible path separately

Node	Value range of z from	
	Feasible Paths	Infeasible segment start Red
4	z=[2,2]	z=[0,0]
5	z=[2,2]	z=[0,0]
6	z=[2,2]	z=[0,0]

Prior Work: PPMFP (CC 2018)

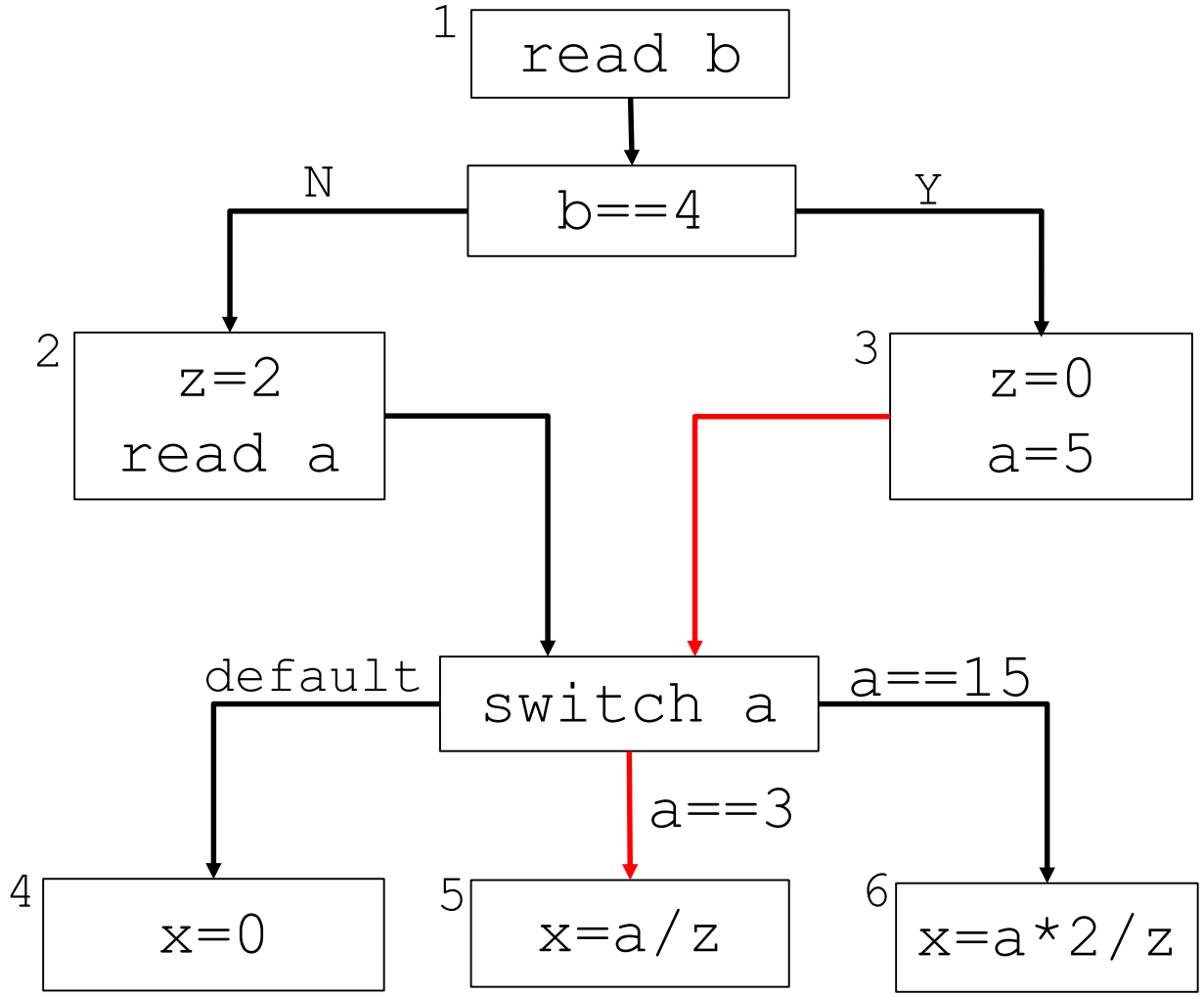


Values of z at node 5 ?

- Compute range of values along feasible and infeasible path separately

Node	Value range of z from	
	Feasible Paths	Infeasible segment start Red
4	z=[2,2]	z=[0,0]
5	z=[2,2]	z=[0,0]
6	z=[2,2]	z=[0,0]

Prior Work: PPMFP (CC 2018)



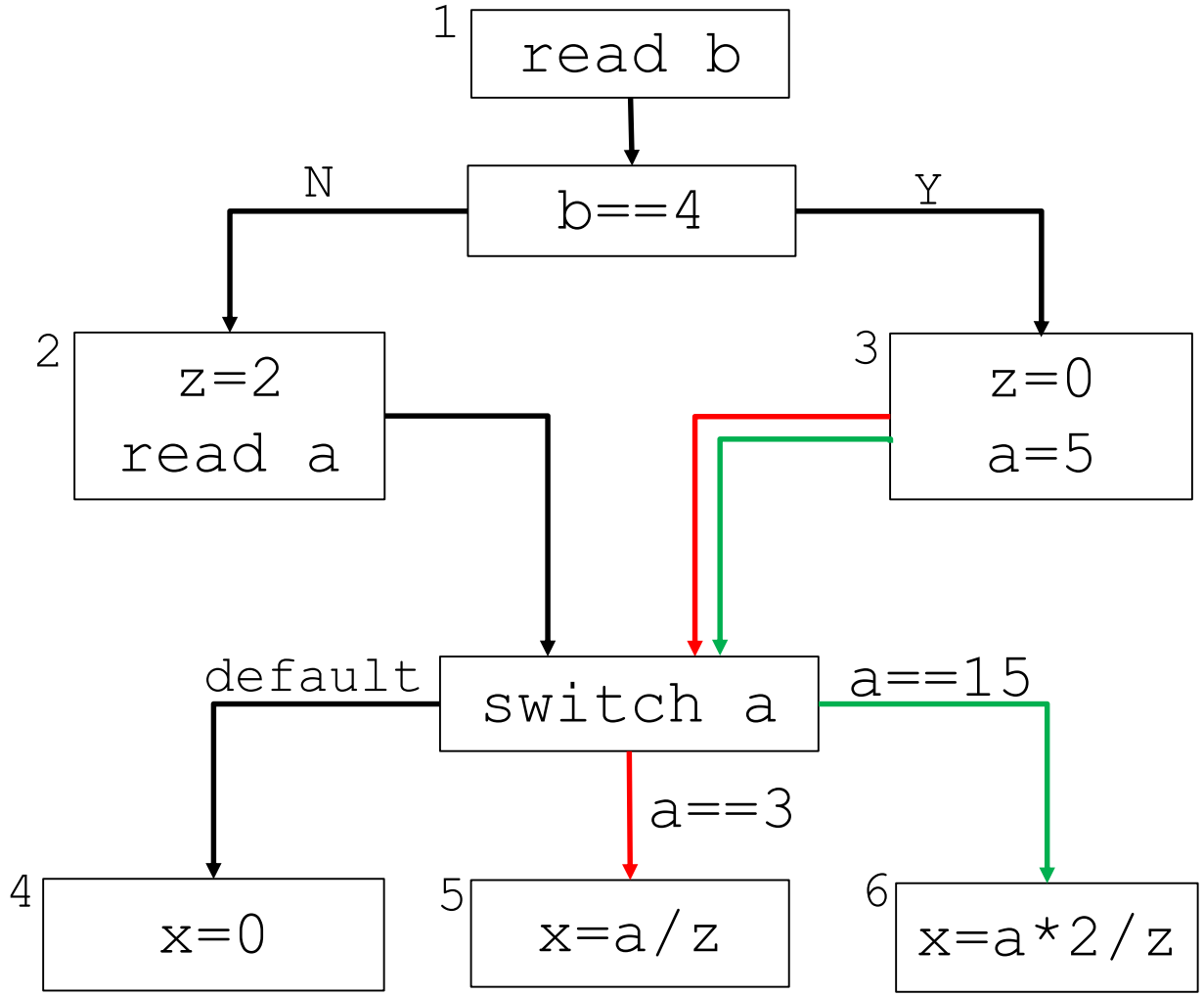
Values of z at node 5 ?

- Compute range of values along feasible and infeasible path separately

Node	Value range of z from	
	Feasible Paths	Infeasible segment start
		Red
4	z=[2,2]	z=[0,0]
5	z=[2,2]	z=[0]
6	z=[2,2]	z=[0,0]

- The final range is meet of ranges from all buckets (except discarded ranges)

Green Path Segment is also Infeasible



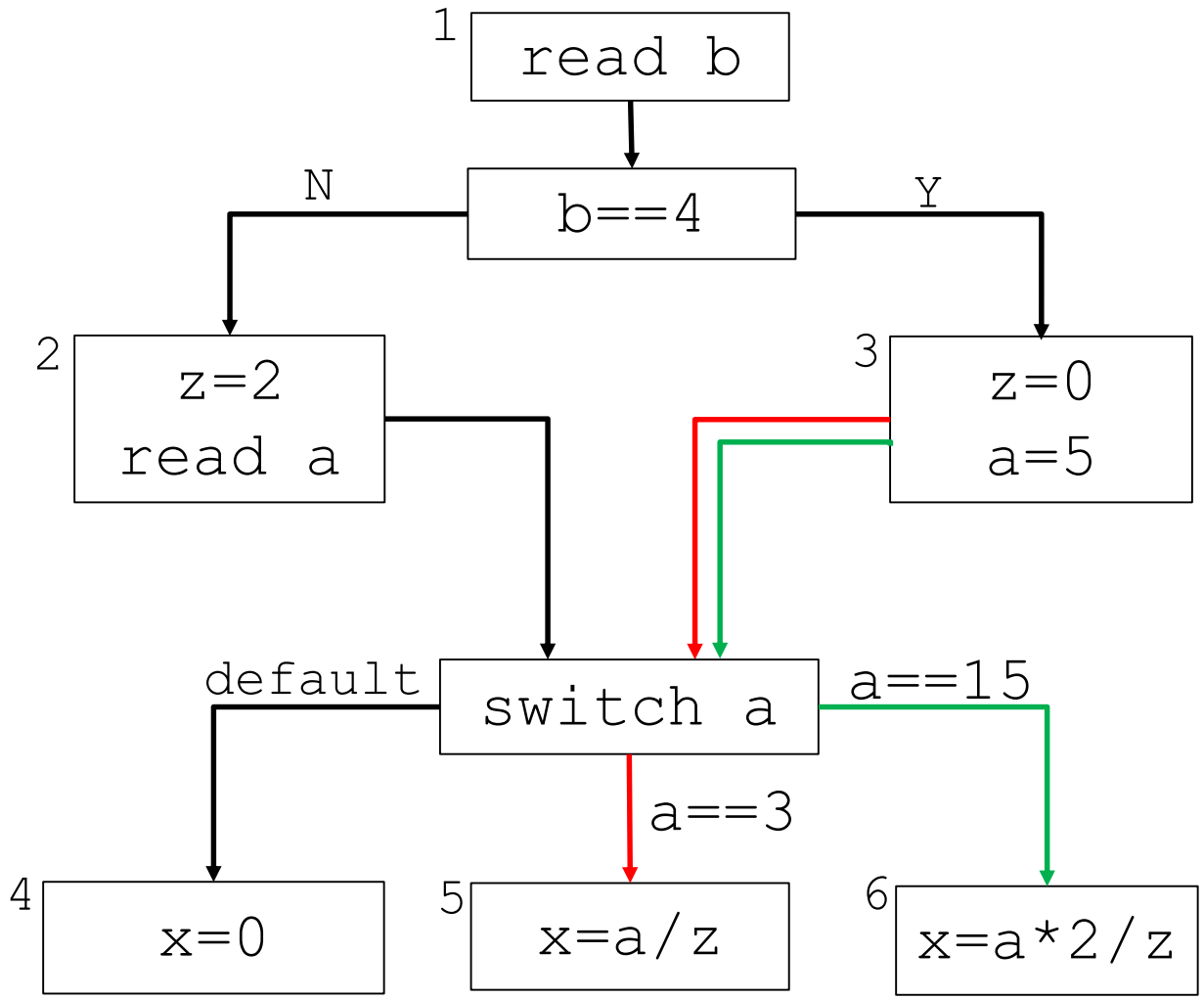
Values of z at node 5 ?

- Compute range of values along feasible and infeasible path separately

Node	Value range of z from	
	Feasible Paths	Infeasible segment start Red
4	z=[2,2]	z=[0,0]
5	z=[2,2]	z=[✖0]
6	z=[2,2]	z=[0,0]

- The final range is meet of ranges from all buckets (except discarded ranges)

One bucket for Green path segment

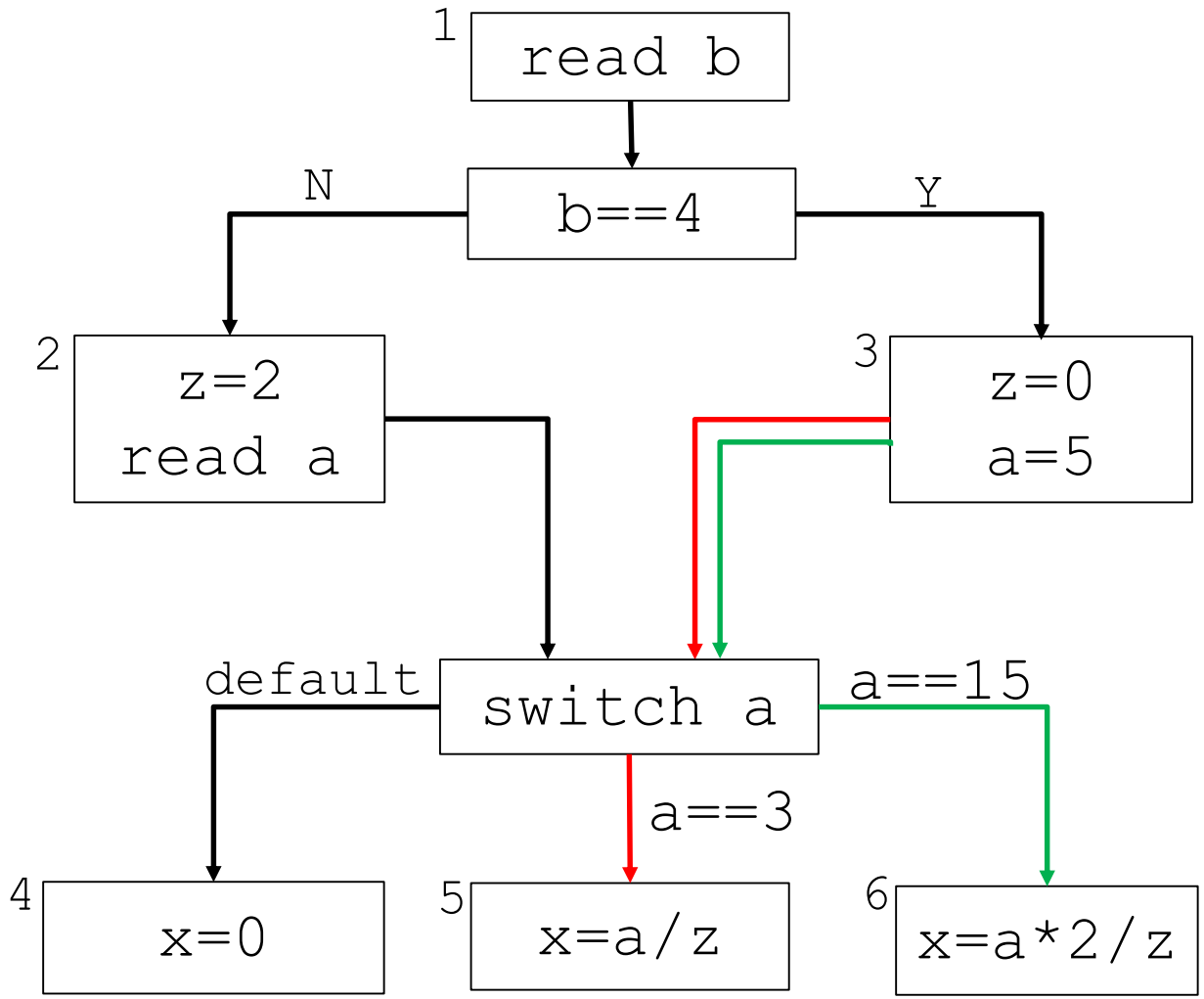


Values of z at node 5 ?

- Compute range of values along feasible and infeasible path separately

Node	Value range of z from		
	Feasible Paths	Infeasible segment start	
		Red	Green
4	z=[2,2]	z=[0,0]	z=[0,0]
5	z=[2,2]	z=[0,0]	z=[0,0]
6	z=[2,2]	z=[0,0]	z=[0,0]

Prior Work: PPMFP (CC 2018)

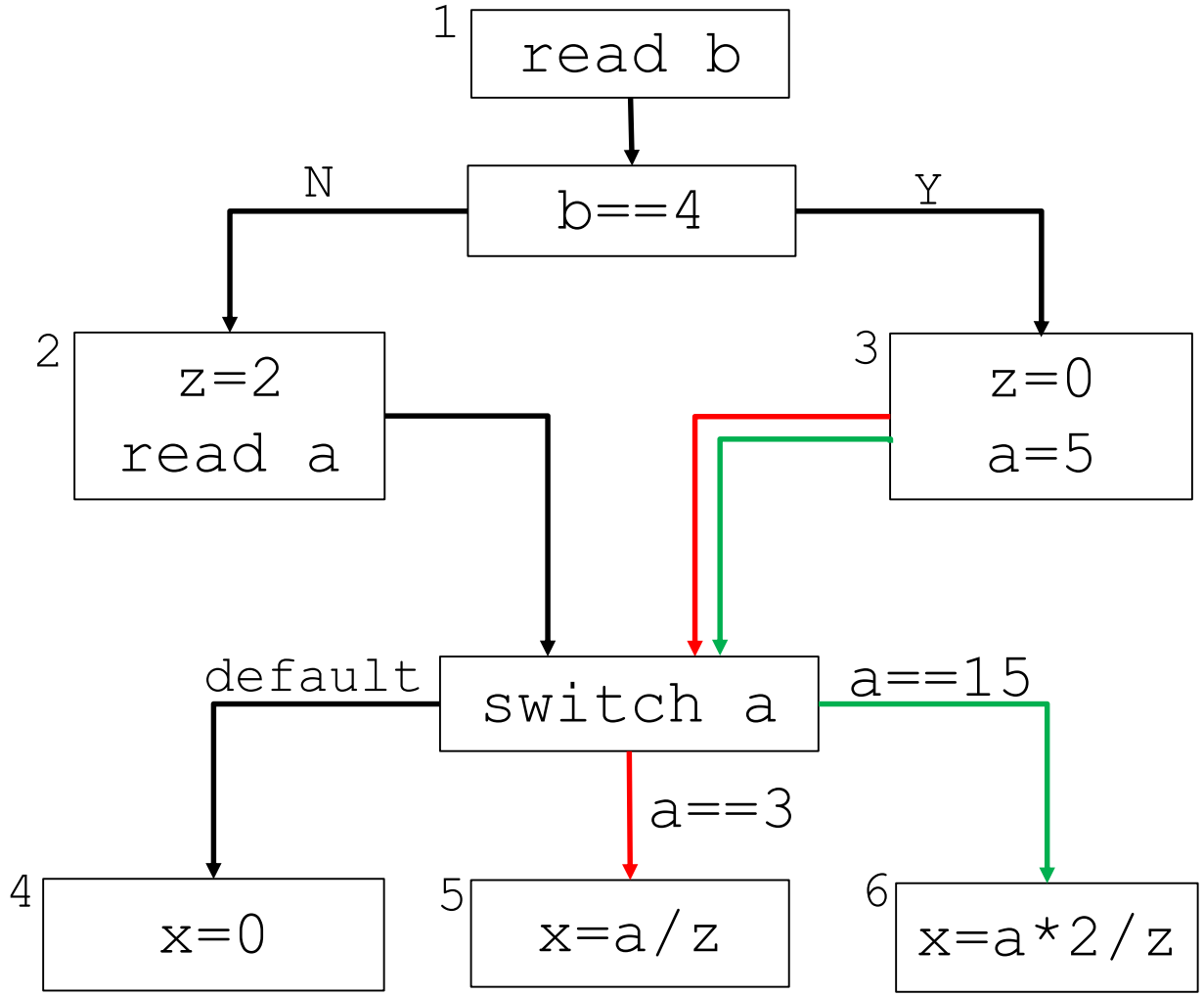


Values of z at node 5 ?

- Compute range of values along feasible and infeasible path separately

Node	Value range of z from		
	Feasible Paths	Infeasible segment start	
		Red	Green
4	z=[2,2]	z=[0,0]	z=[0,0]
5	z=[2,2]	z=[0,0]	z=[0,0]
6	z=[2,2]	z=[0,0]	z=[0,0]

Prior Work: PPMFP (CC 2018)

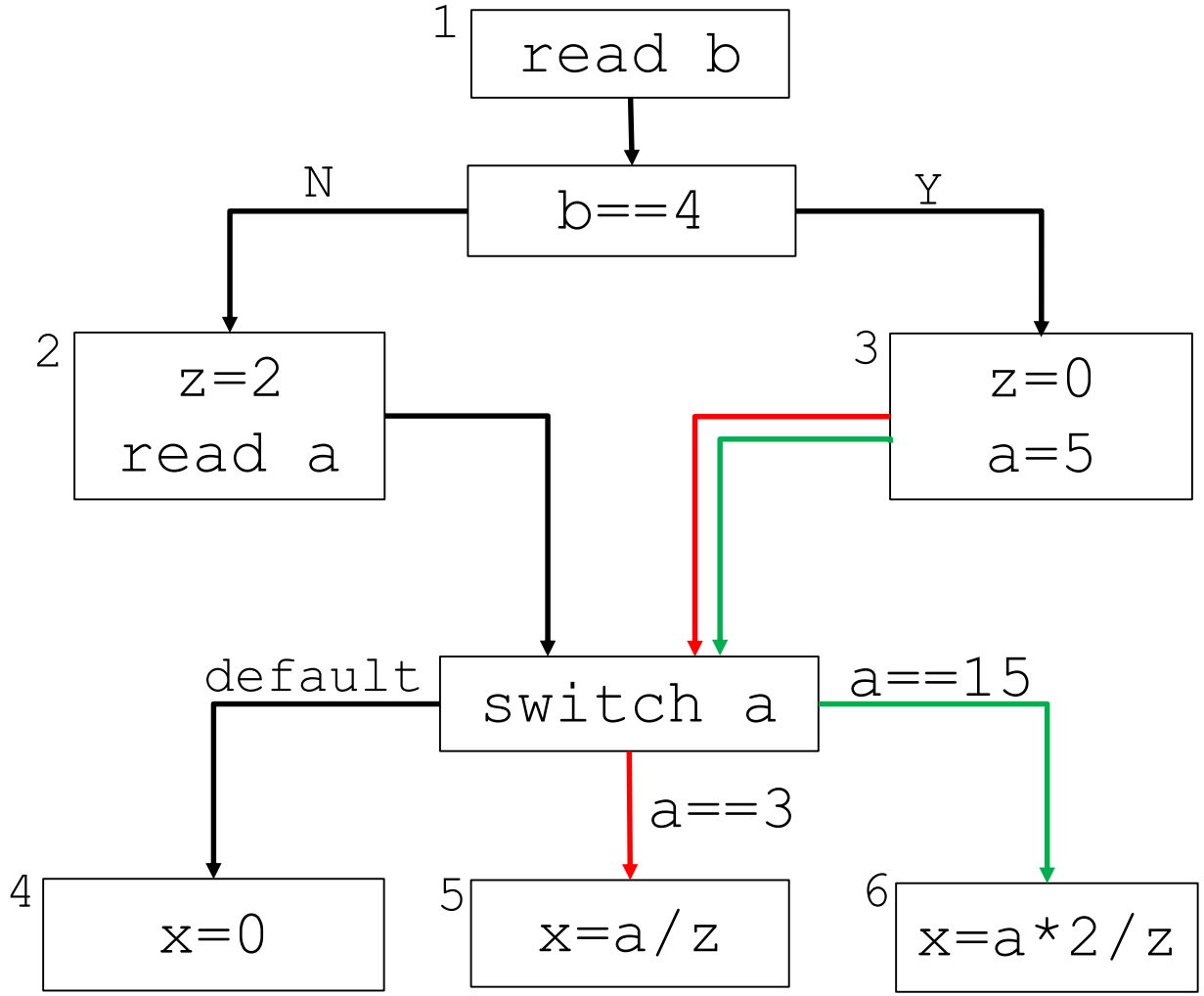


Values of z at node 5 ?

- Compute range of values along feasible and infeasible path separately

Node	Value range of z from		
	Feasible Paths	Infeasible segment start	
		Red	Green
4	z=[2,2]	z=[0,0]	z=[0,0]
5	z=[2,2]	z=[0,0]	z=[0,0]
6	z=[2,2]	z=[0,0]	z=[0,0]

Prior Work: PPMFP (CC 2018)



Values of z at node 5 ?

- Compute range of values along feasible and infeasible path separately

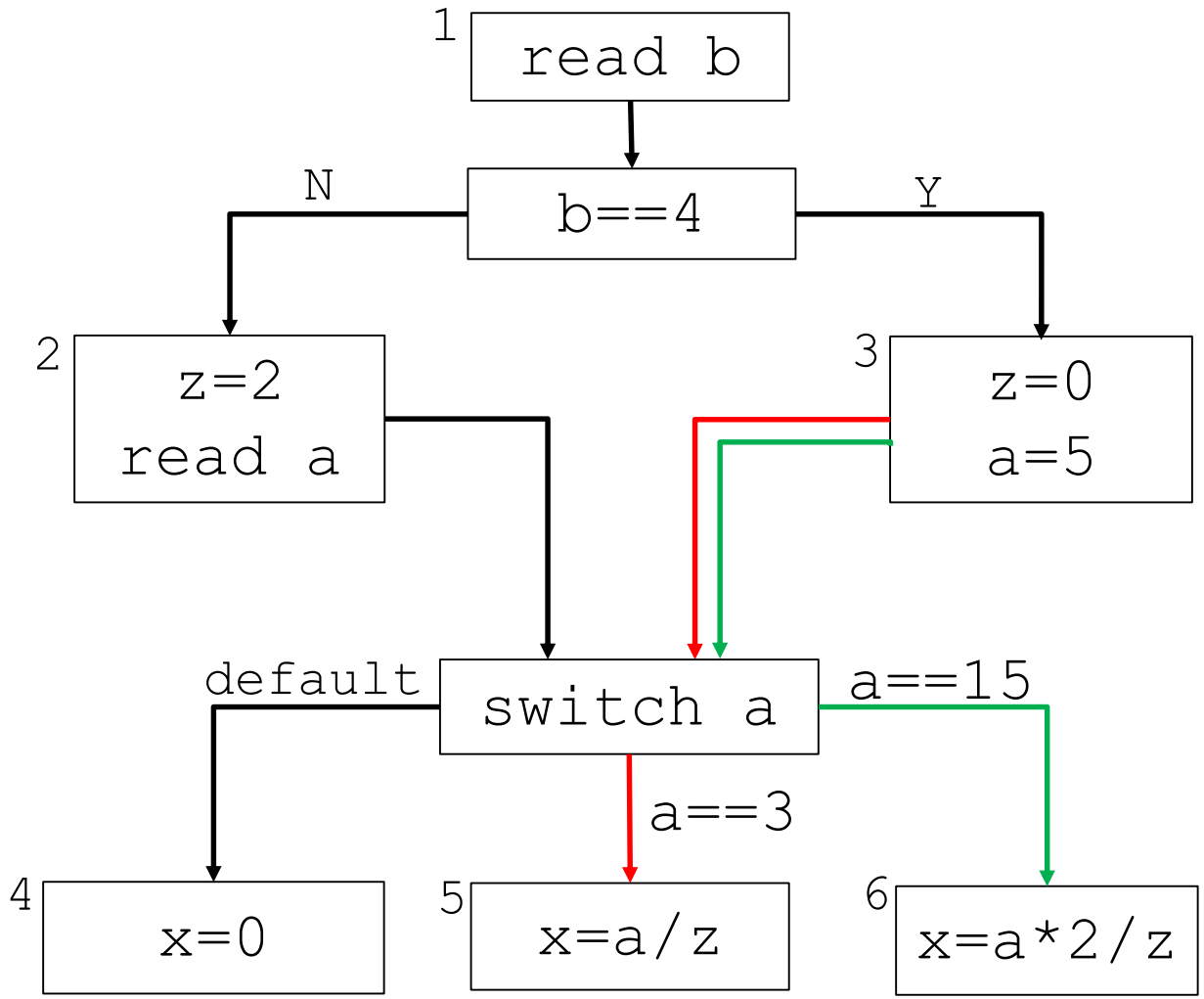
Node	Value range of z from		
	Feasible Paths	Infeasible segment start	
		Red	Green
4	z=[2,2]	z=[0,0]	z=[0,0]
5	z=[2,2]	z=[0,0]	z=[0,0]
6	z=[2,2]	z=[0,0]	z=[0,0]

- Imprecise: z=[0,0] is part of green bucket at node 5 and data flow value in green bucket is not discarded at node 5.

Intersecting Infeasible Path Segments

- On average 70% infeasible path segments were intersecting in our benchmarks
- Keeping one bucket for one path segment creates duplication, which brings imprecision
- We need to create buckets that represent group of path segments to achieve precision

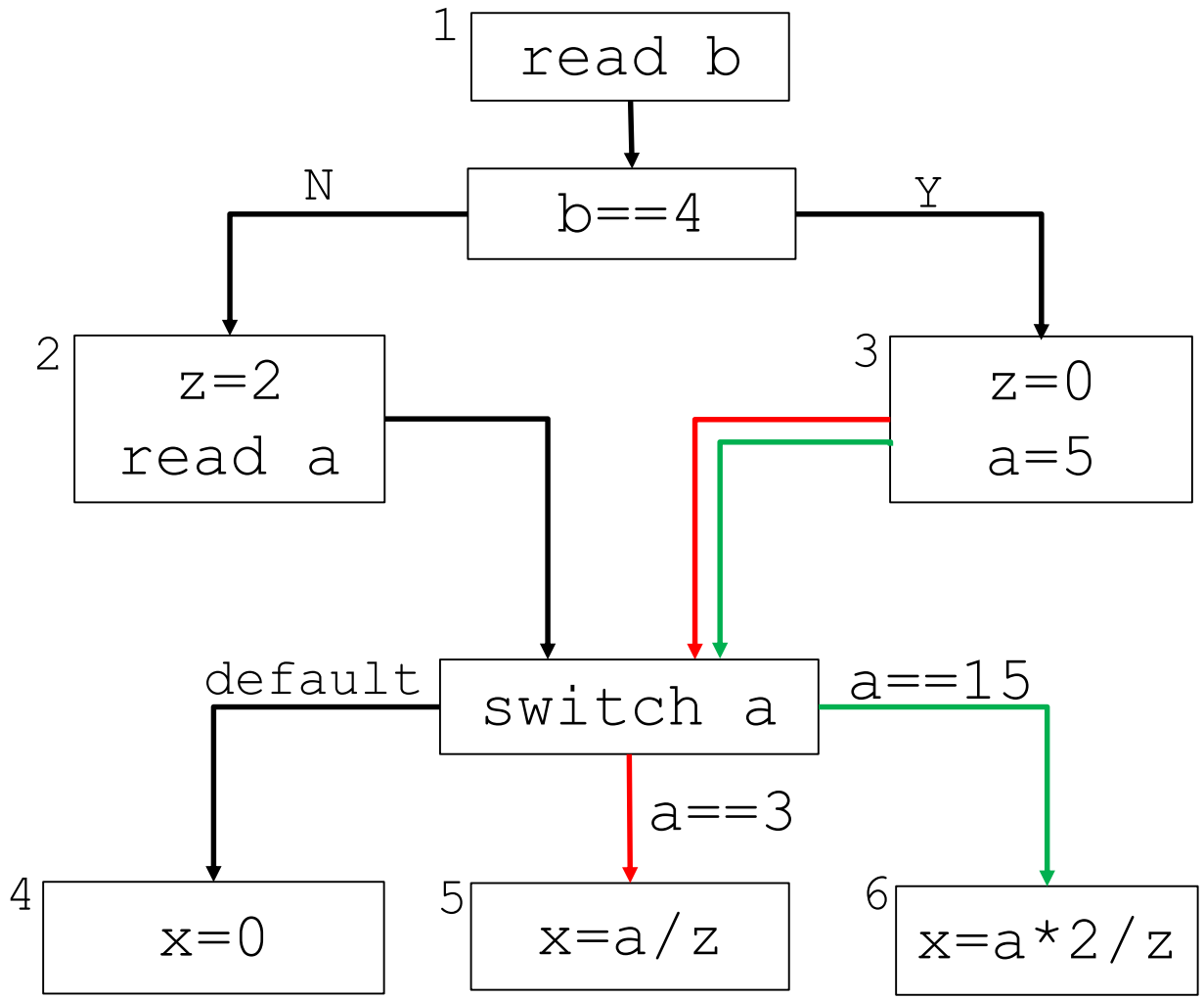
Our Idea



- Keep a bucket for each color combination (each color represents one infeasible path segment)

Node	Feasible Paths	Value range of z from Infeasible segment start		
		Red	Green	Red ∧ Green
5	z=[2,2]			z=[0,0]
6	z=[2,2]			z=[0,0]

Our Idea



- Keep a bucket for each color combination (each color represents one infeasible path segment)

Node	Feasible Paths	Value range of z from Infeasible segment start		
		Red	Green	Red ∧ Green
5	z=[2,2]			z=[[illegible]0]
6	z=[2,2]			z=[[illegible]0]

Why Prior Work (PPMFP) Did Not Do It

- Low assessed Impact: How many Infeasible Path Segments are intersecting?

- High cost - seems that exponential number of Buckets will be required: For K infeasible path segments:
  1. PPMFP Solution keeps K buckets
  2. Our Solution will require 2^K buckets?
  (Seems as bad as MOP solution in terms of complexity)

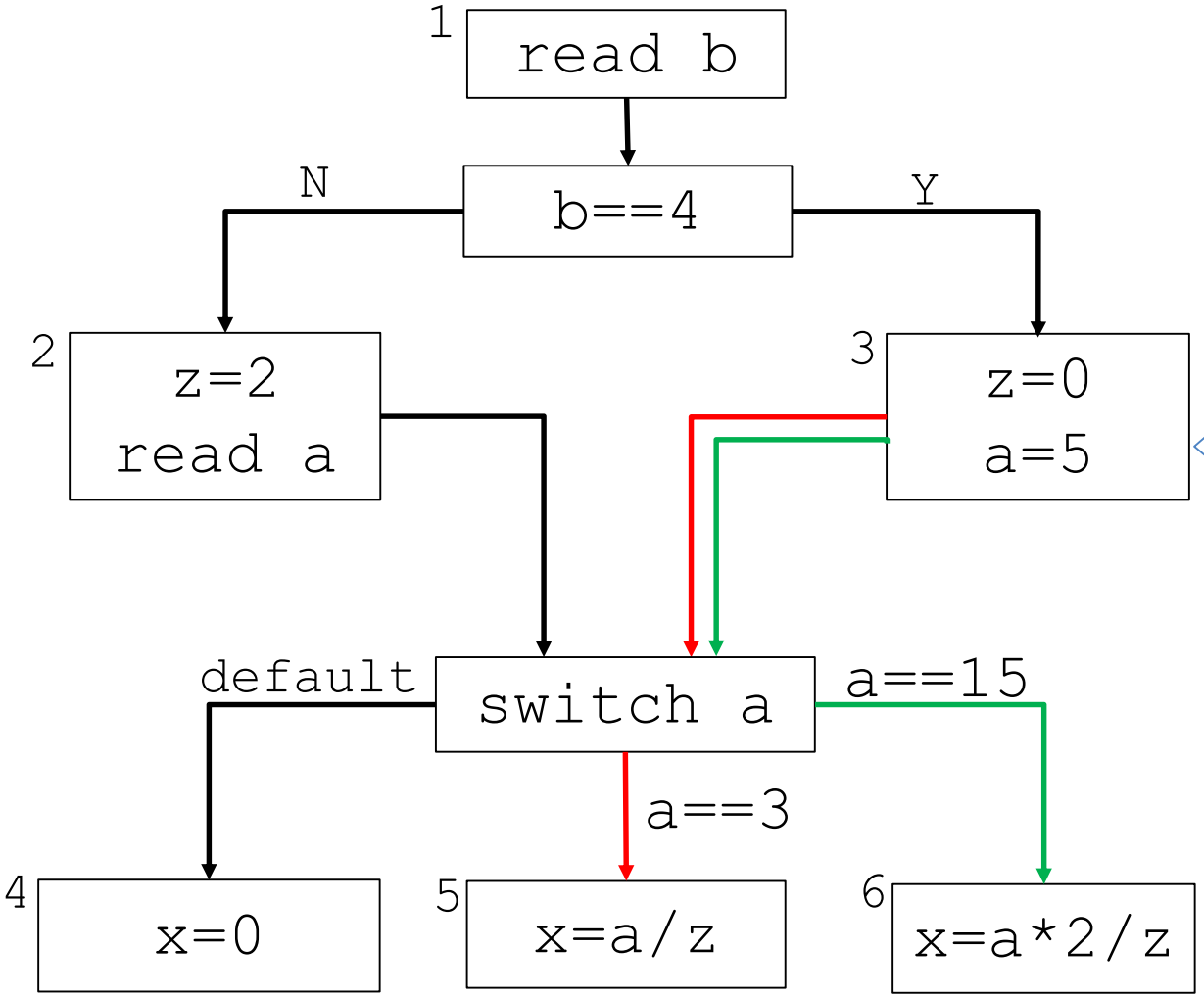
Our Contribution

- Low assessed Impact: How many Infeasible Path Segments are intersecting? **Average 70%**

- High cost - seems that exponential number of Buckets will be required: For K infeasible path segments:
  1. PPMFP Solution keeps K buckets
  2. Our Solution will require 2^K buckets? **Requires number of buckets=O(Number of Edges)** (Seems as bad as MOP solution in terms of complexity)

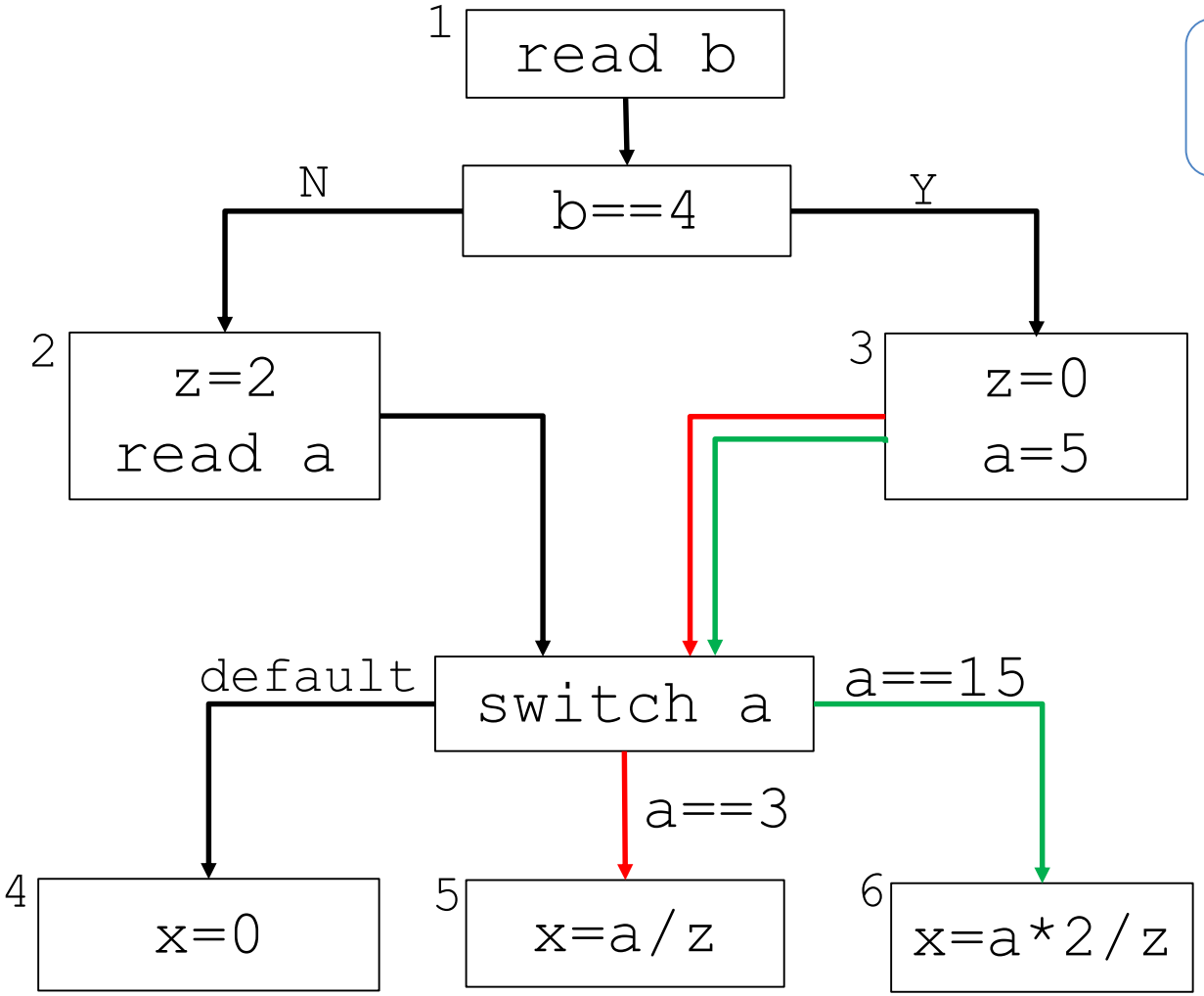
- We propose heuristics to reduce number of buckets even further

How Many Buckets



- A Data Flow Value at a node is assigned to buckets representing colors of outgoing edges
- Maximum buckets at a node = | out edges |
- Total buckets = N * E, E = | out edges|

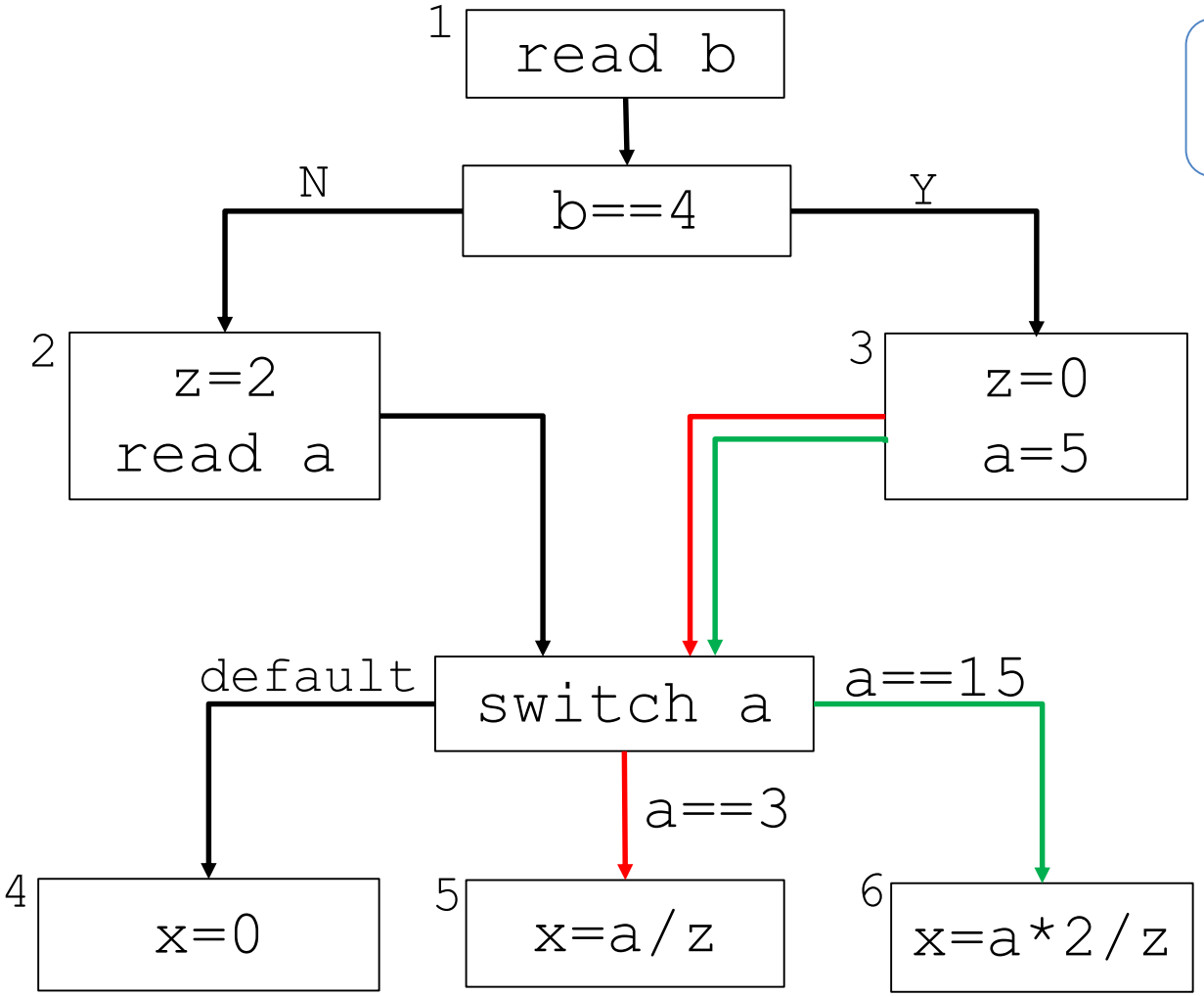
Reducing Number of Buckets



1. Eliminate Empty Buckets

Node	Feasible Paths	Value range of z from Infeasible segment start		
		Red	Green	Red ∧ Green
5	z=[2,2]			z=[0,0]
6	z=[2,2]			z=[0,0]

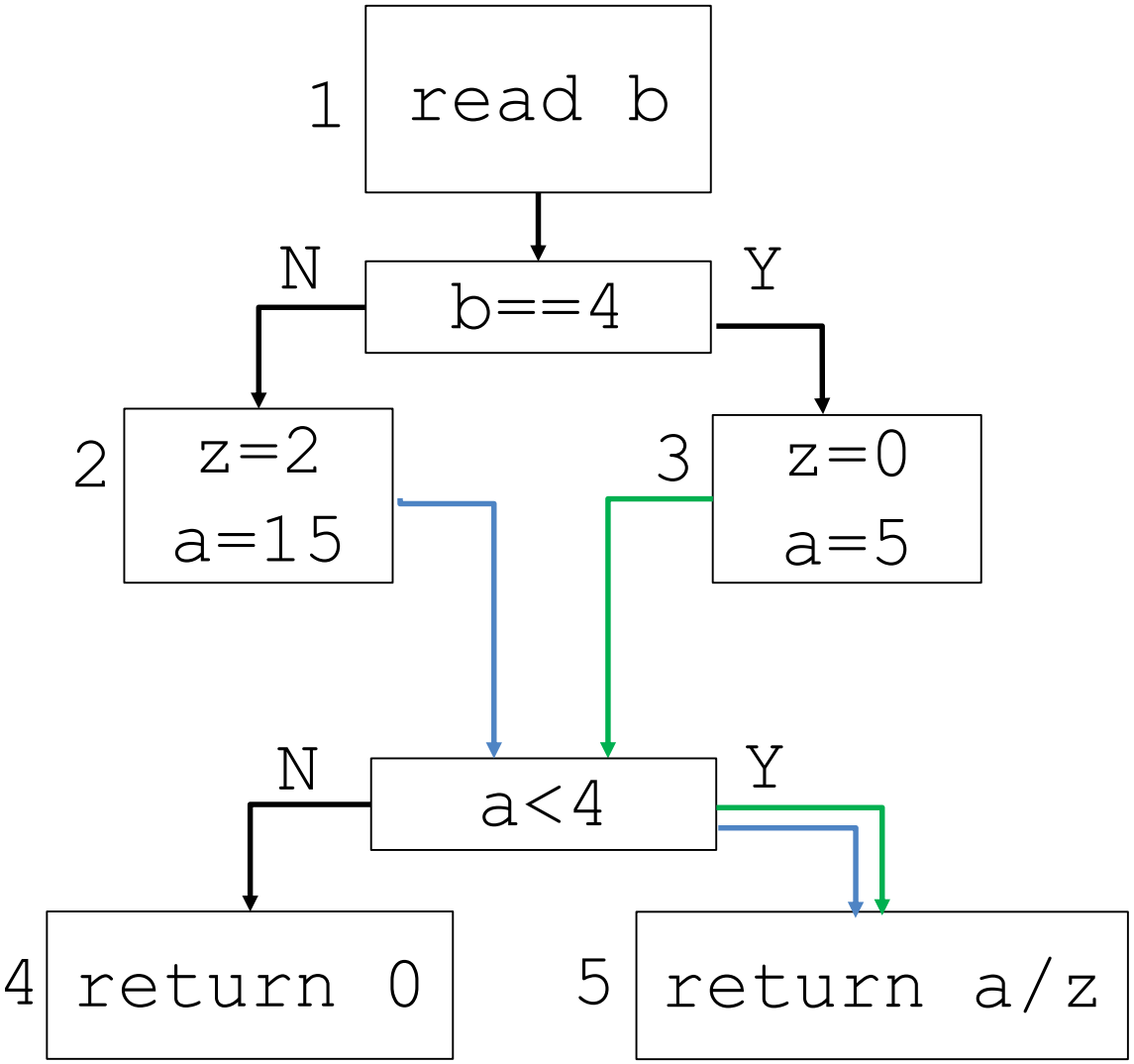
Reducing Number of Buckets



1. Eliminate Empty Buckets

Node	Value range of z from	
	Feasible Paths	Infeasible segment start Red ∧ Green
5	z=[2,2]	z=[0,0]
6	z=[2,2]	z=[0,0]

Reducing Number of Buckets



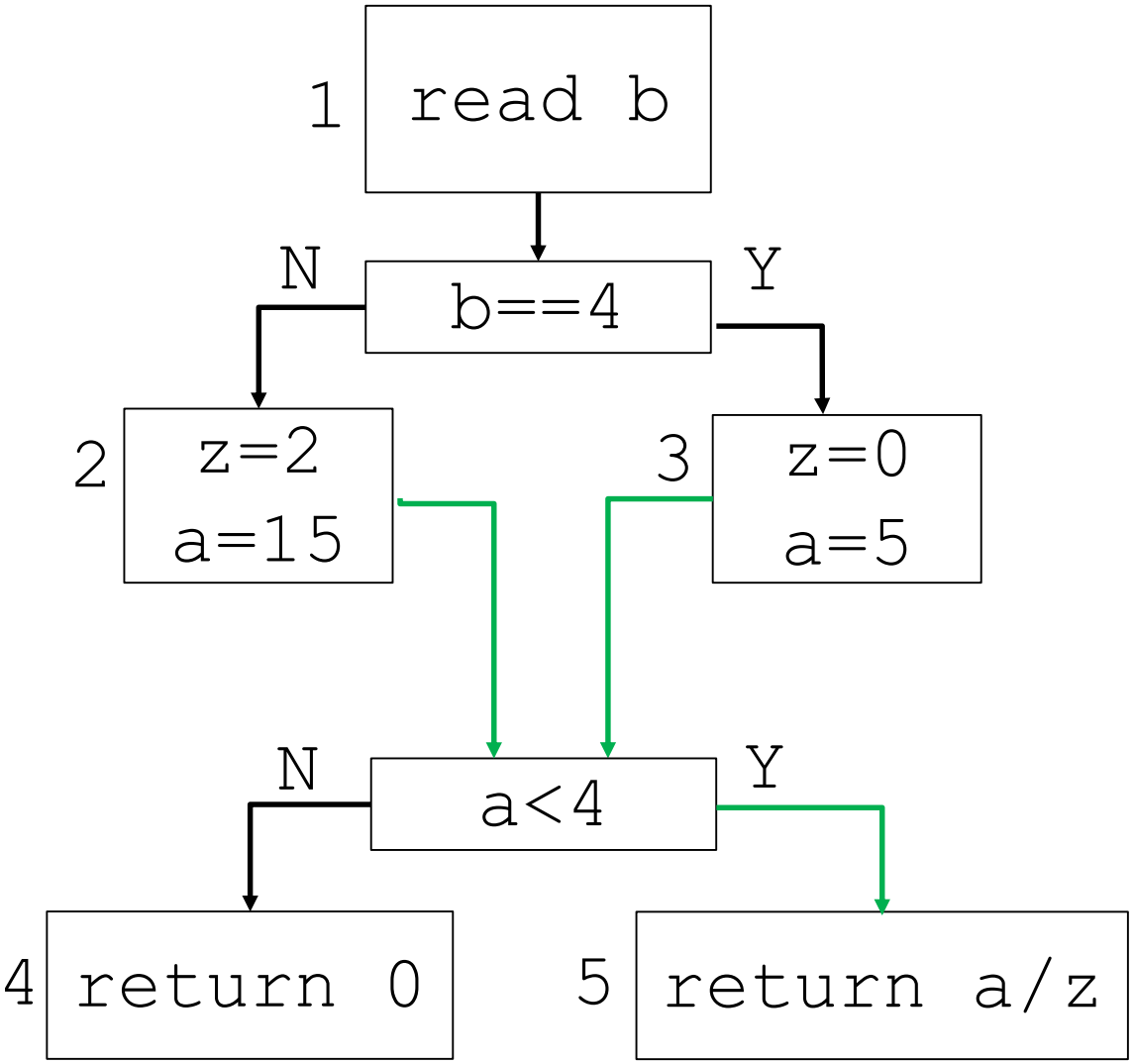
2. Merge bucket with same infeasibility Constraint

- Infeasibility condition: `a<4=false,` is same for blue and green path.

Node	Value range of z from		
	Feasible Paths	Infeasible segment start	
		Blue	Green
5		z=[2,2]	z=[0,0]

- So one bucket is sufficient for blue and green paths

Reducing Number of Buckets



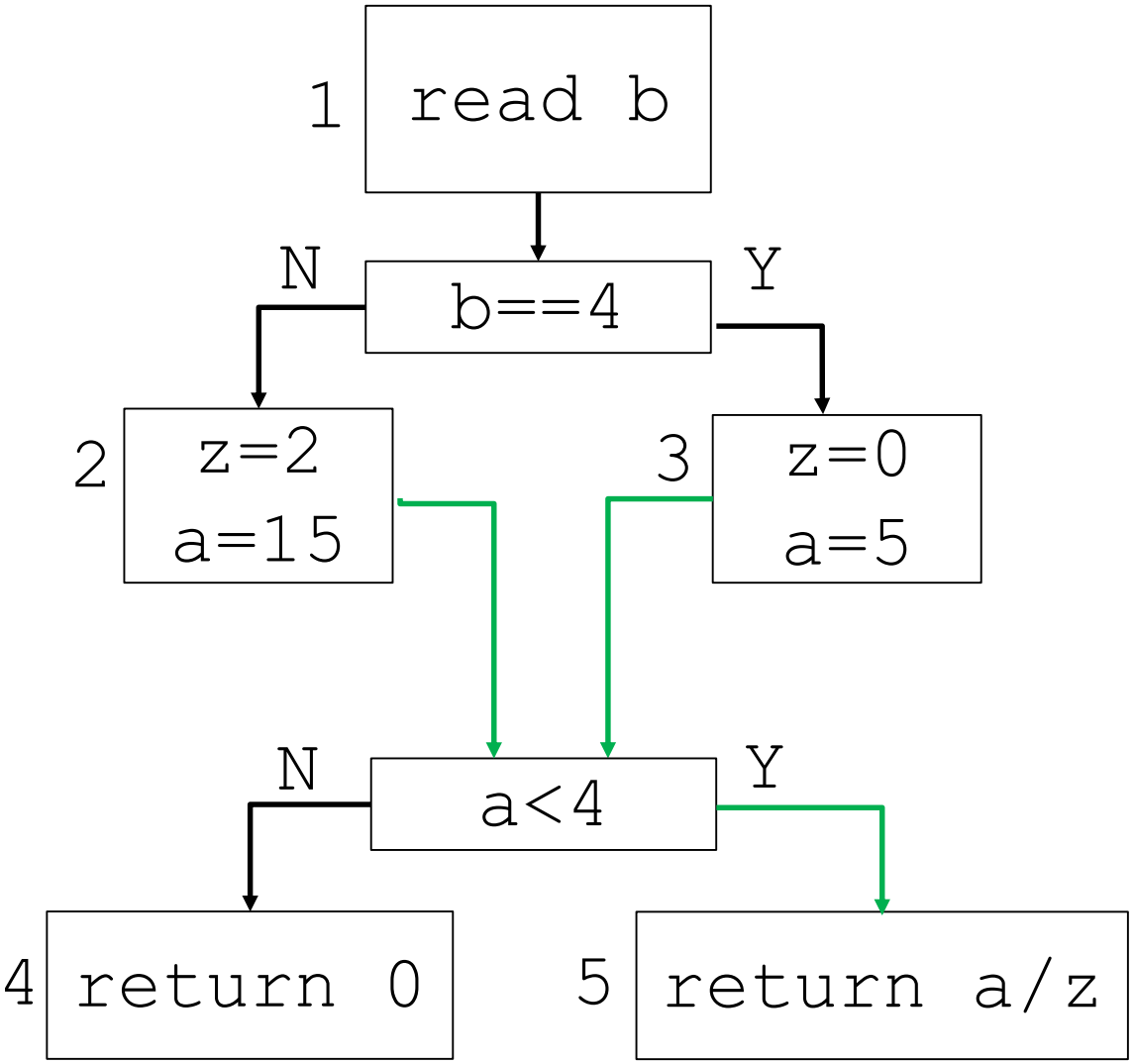
2. Merge bucket with same infeasibility Constraint

- Infeasibility condition: `a<4=false,` is same for blue and green path.

Node	Value range of z from	
	Feasible Paths	Infeasible segment start
		Green
5		z=[0,2]

- So one bucket is sufficient for blue and green paths

Reducing Number of Buckets



2. Merge bucket with same infeasibility Constraint

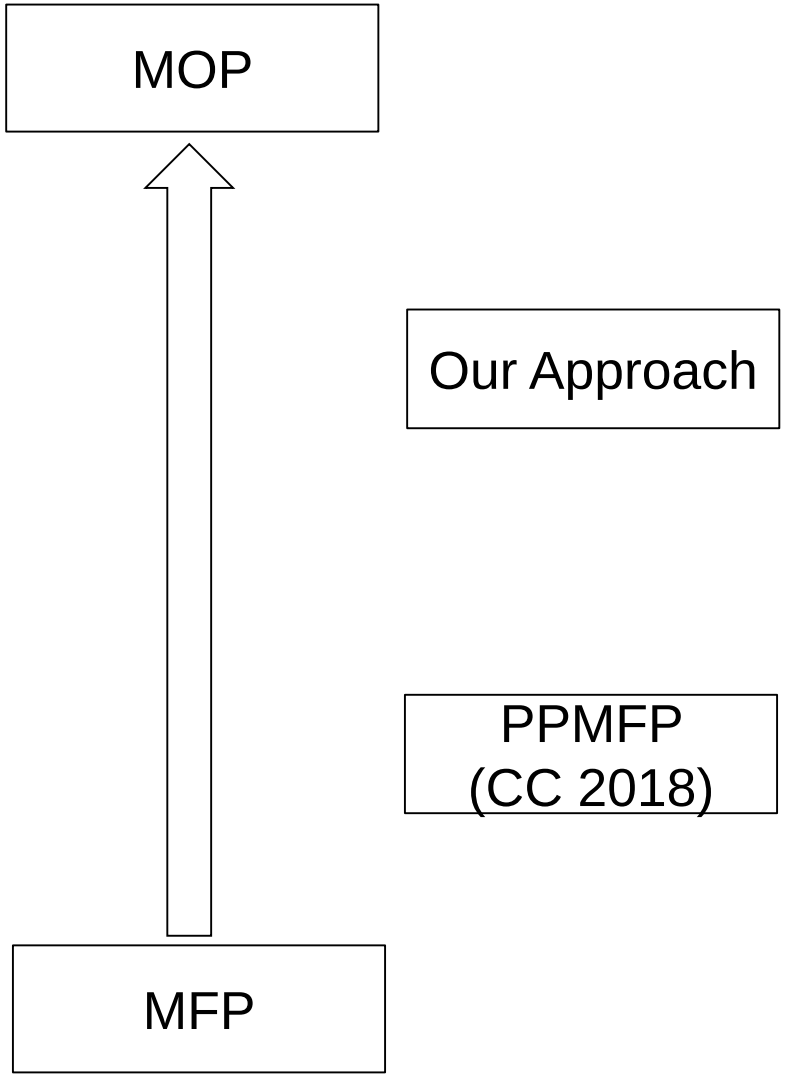
- Infeasibility condition: `a<4=false,` is same for blue and green path.

Node	Value range of z from	
	Feasible Paths	Infeasible segment start
		Green
5		z=[-∞,2]

- So one bucket is sufficient for blue and green paths

Related Work

- Number of values maintained increases
- Precision Increases
- Scalability Decreases



Benchmarks and Results

Benchmarks:

- Seven open source benchmarks: SPEC CPU 2006

- Three industry benchmarks

Results:

1. Improvement – up to 3 times improvement over PPMFP.

- Up to 20% reduction in reaching definitions

- Up to 10% reduction in uninitialized variables.

2. Analysis Time: Up to 300% more than PPMFP.

Future Work

- Detecting infeasible path segments in presence of Arrays and Pointers in condition expressions is an open problem.
- Improving scalability of the approach, currently scales to 75KLoc

Thank You