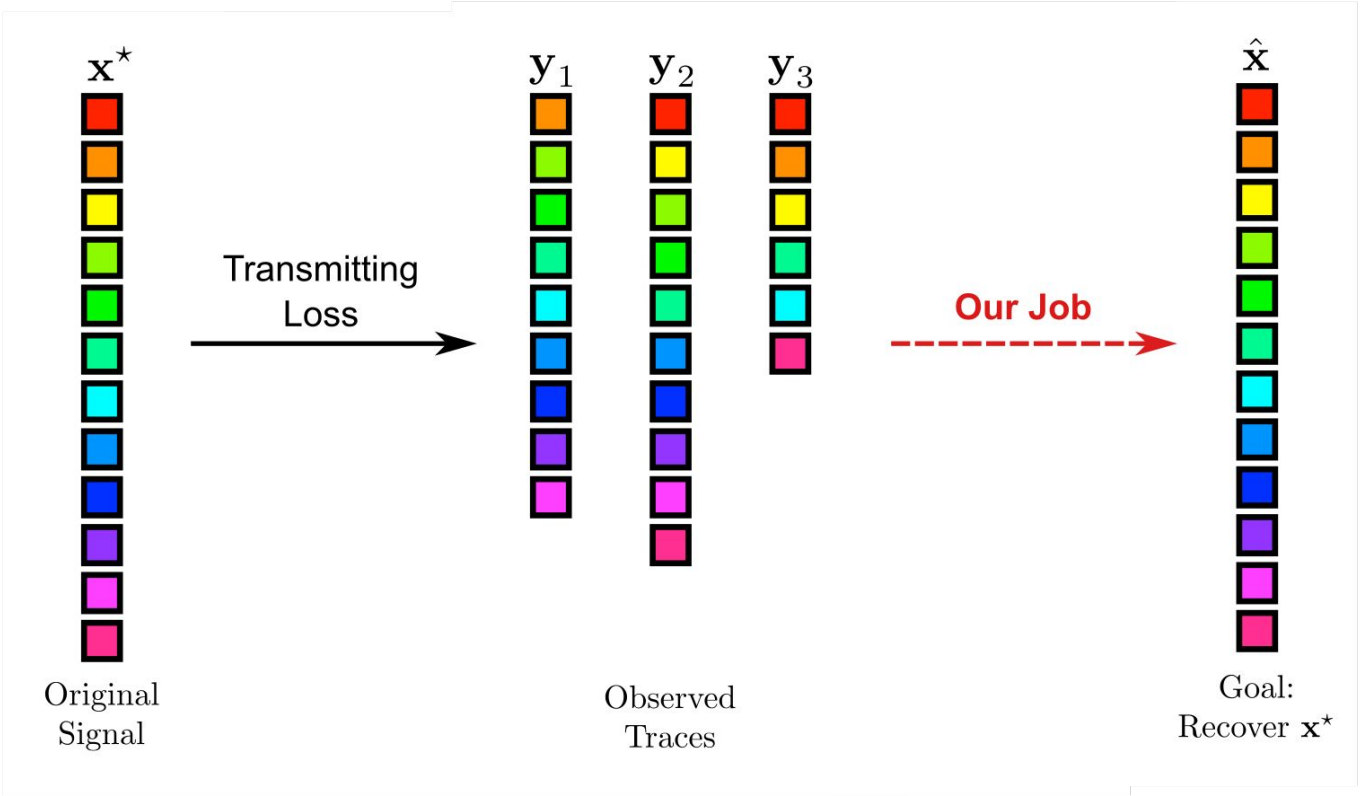
# Minimum-Length Trace Reconstruction

Via Integer Programming

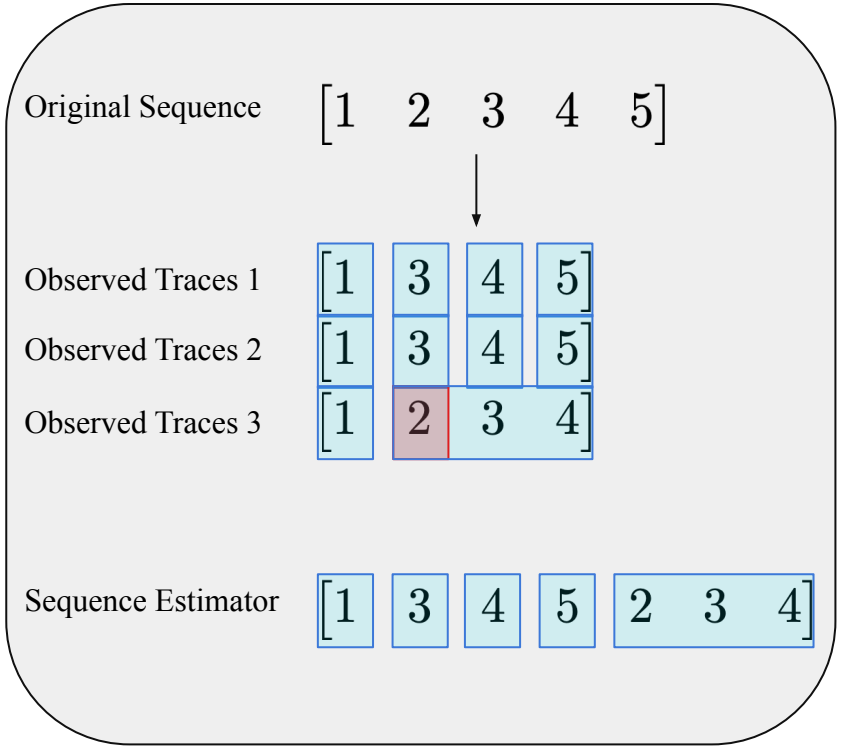
**Huanran Li, Daniel Pimentel-Alarcón**

**University of Wisconsin-Madison**

# Trace Reconstruction

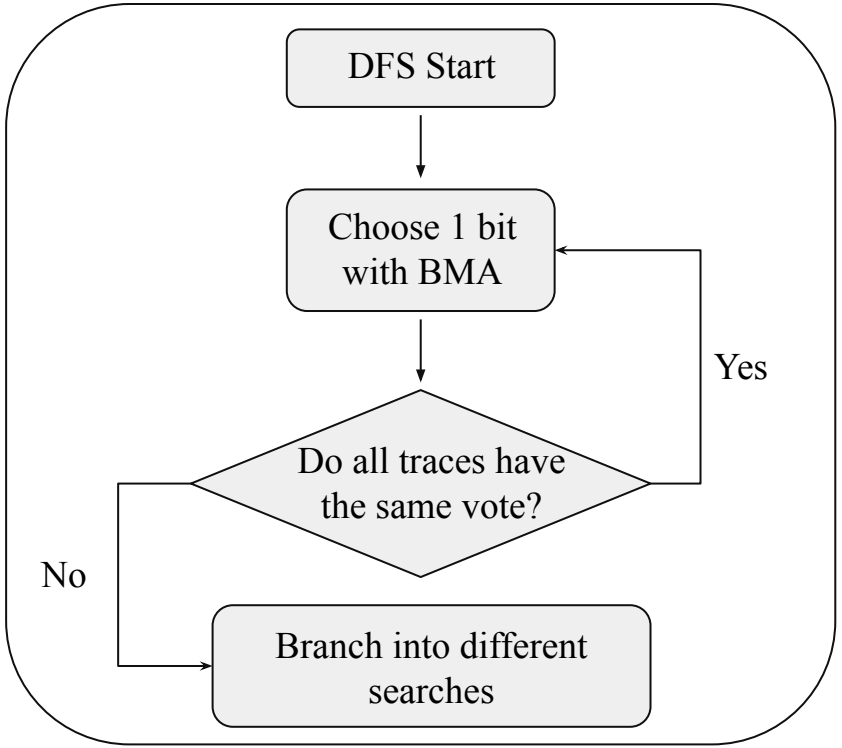


## Bitwise Majority Alignment (BMA)




**Sequence is much longer!**

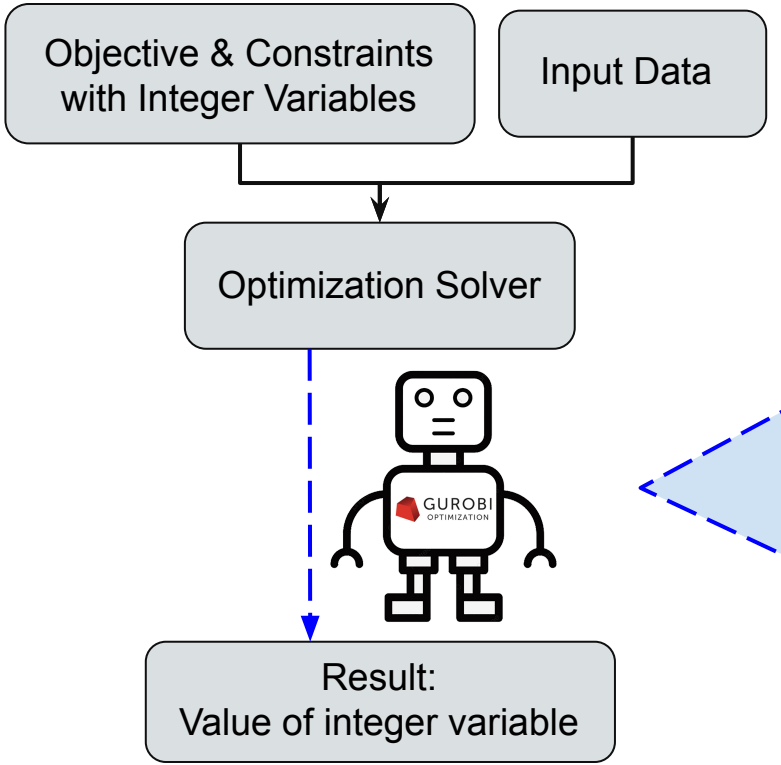
## Depth First Search (DFS)

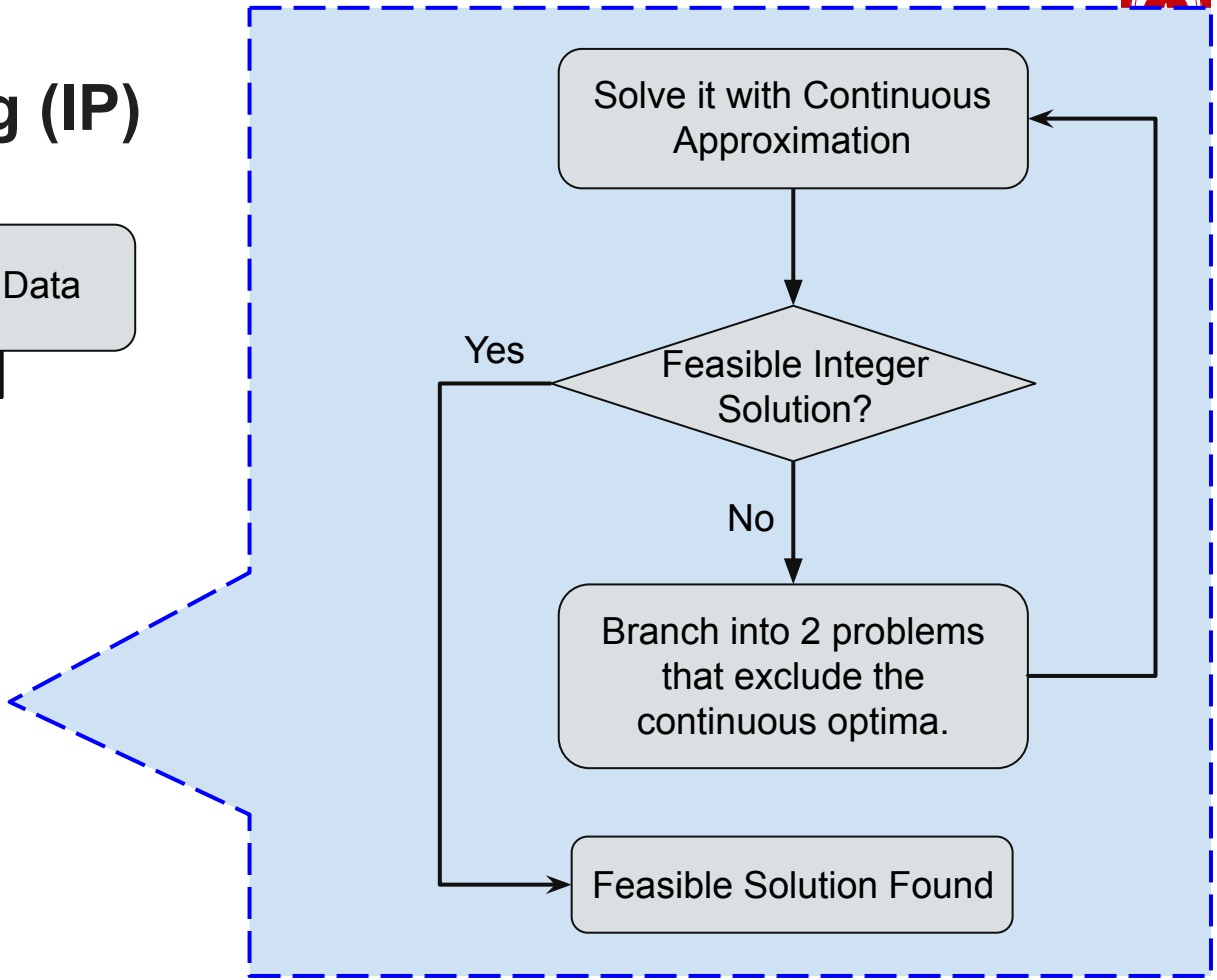



**Slow!**

How to **efficiently** reconstruct a sequence with the **shortest length possible**?

# Integer Programming (IP)



# Trace Reconstruction Model (Our Paper)

**Variable**

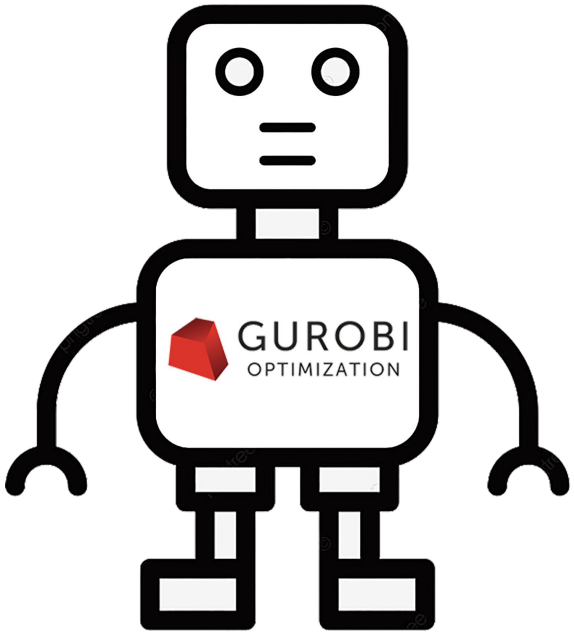
We created an **Estimator** $\mathbf{x} = [?, ?, ..., ?]$, and we need you to fill it in.

**Constraints**

We must observe all the **traces** from the **Estimator**.

**Objective**

We want you to use as less bits as possible in the **Estimator**.




Copy. Done. Easy. :)

# Trace Reconstruction Model (Our Paper)

**1.** The Estimator must observe all traces.

**Constraint 1: Selector i * Trace i + Filler i = Estimator**

Variable **Selector i** | Observed **Trace i** | Variable **Filler i** | Variable **Estimator**

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ 3 \\ 4 \end{bmatrix} + \begin{bmatrix} 0 \\ 2 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \\ 0 \\ 0 \end{bmatrix}$$

**a.** Each entry of the estimator cannot be selected from the **Trace** and filled by **Filler** at the same time.
**Constraint 2**: **Selector** and **Filler** cannot fill the same entry of the **Estimator**.

**b.** The entries selected from the **Trace** must follow the order in the **Trace**.
**Constraint 3**: **Selector** must be an identity matrix after deleting empty rows.

# Trace Reconstruction Model (Our Paper)

**2.** The optimal **Estimator** must have minimal length.

**Estimator:** $\mathbf{x} = \overbrace{[1, 2, 3, 4, 0, 0]}^{\textbf{Capacity}}$

$\underbrace{\qquad\qquad}_{\textbf{Length}}$

**Objective**: $\min ||\mathbf{z}||_1$

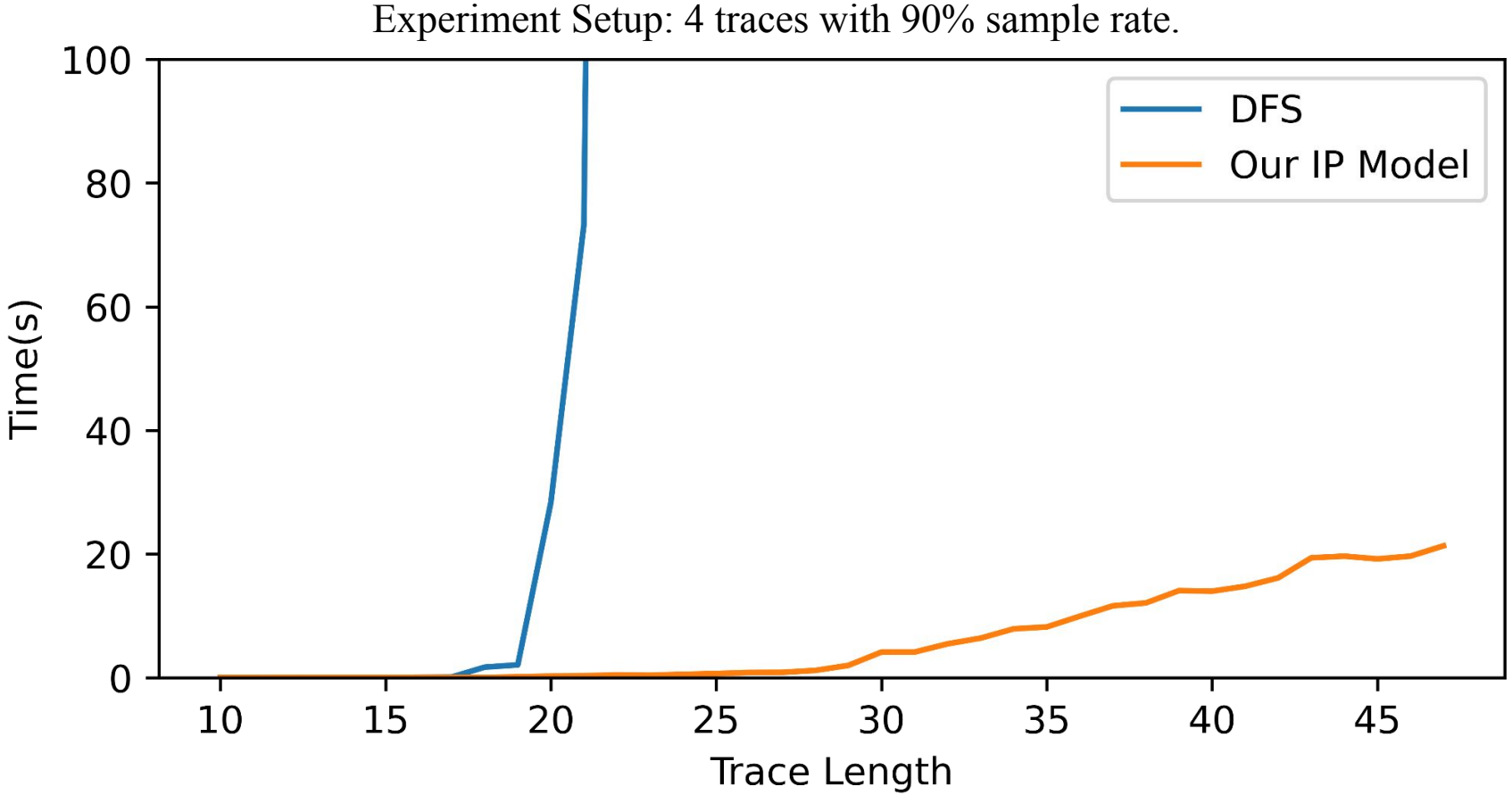
**Estimator:** $\mathbf{x} = [1, 2, 3, 4, 0, 0]$

**Mask:** $\mathbf{z} = [1, 1, 1, 1, 0, 0]$

**a.** Assign value 1 to the **Mask** entry prior to the last bit used in the **Estimator**.

**Constraint 4**: $[\mathbf{x}]_i \leq \beta[\mathbf{z}]_i$

**Constraint 5:** $[\mathbf{z}]_{i-1} \geq [\mathbf{z}]_i$


Experiment Setup: 4 traces with 90% sample rate.
DFS
Our IP Model
Time(s)
Trace Length
0
20
40
60
80
100
10
15
20
25
30
35
40
45

# More Useful Constraints!

**Estimator:** $\mathbf{x} = [x_1, x_2, x_3, x_4, x_5, x_6]$

**Trace i:** $\mathbf{y}_i = [1, 3, 4] \implies \begin{cases} x_1 + x_2 + x_3 + x_4 + x_5 + x_6 \geq 1 + 3 + 4 \\ x_1 + x_2 + x_3 + x_4 + x_5 \geq 1 + 3 \\ x_1 + x_2 + x_3 + x_4 \geq 1 \end{cases}$

**Fact: Actual length of the sequence =** $||\mathbf{z}||_1$ **!**

**Estimator:** $\mathbf{x} = [x_1, x_2, x_3, x_4, x_5, x_6]$

**Trace i:** $\begin{cases} \mathbf{y}_i = [1, 3, 4] \\ ||\mathbf{z}||_1 = 4 \end{cases} \implies \begin{cases} x_1 + x_2 + x_3 + x_4 \geq 1 + 3 + 4 \\ x_1 + x_2 + x_3 \geq 1 + 3 \\ x_1 + x_2 \geq 1 \end{cases}$

**1.** $\mathbf{y}_i = \begin{bmatrix} 1 \\ 3 \\ 4 \end{bmatrix} \implies \mathbf{U}_i = \begin{bmatrix} 1+3+4 & 1+3 & 1 & 0 & 0 & 0 \\ 0 & 1+3+4 & 1+3 & 1 & 0 & 0 \\ 0 & 0 & 1+3+4 & 1+3 & 1 & 0 \\ 0 & 0 & 0 & 1+3+4 & 1+3 & 1 \\ 0 & 0 & 0 & 0 & 1+3+4 & 1+3 \\ 0 & 0 & 0 & 0 & 0 & 1+3+4 \end{bmatrix}$

**2.** $[\mathbf{w}_i]_j = \begin{cases} [\mathbf{U}_i\mathbf{z}]_j - [\mathbf{U}_i\mathbf{z}]_{j+1} & j < m \\ [\mathbf{U}_i\mathbf{z}]_j & j = m \end{cases}$

**eg.** $\mathbf{z} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 0 \\ 0 \end{bmatrix} \implies \mathbf{w}_i = \begin{bmatrix} 0 \\ 1 \\ 1+3 \\ 1+3+4 \\ 0 \\ 0 \end{bmatrix}$

$$\begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \end{bmatrix} \geq \begin{bmatrix} 0 \\ 1 \\ 1+3 \\ 1+3+4 \\ 0 \\ 0 \end{bmatrix}$$

$$\begin{cases} \mathbf{y}_i = [1,3,4] \\ ||\mathbf{z}||_1 = 4 \end{cases} \implies \begin{cases} x_1 + x_2 + x_3 + x_4 \geq 1+3+4 \\ x_1 + x_2 + x_3 \geq 1+3 \\ x_1 + x_2 \geq 1 \end{cases}$$

| Method | | Original | Improved |
|---|---|---|---|
| **Sequence length = 50**<br>**Estimator Capacity = 60** | Mean (s) | **1126.7** | 1144.4 |
| | Std (s) | 2106.1 | **974.5** |
| | Min (s) | **40.7** | 58.0 |
| | Max (s) | 6989.4 | **2864.5** |
| **Sequence length = 50**<br>**Estimator Capacity = 200** | Mean (s) | 25907.4 | **15138.4** |
| | Std (s) | 11340.8 | **8812.3** |
| | Min (s) | 15444.8 | **8515.4** |
| | Max (s) | 41665.8 | **27592.5** |

# Conclusion

**Variable**

We created an **Estimator** $\mathbf{x} = [?, ?, ..., ?]$, and we need you to fill it in.

**Constraints**

We must observe all the **traces** from the **Estimator**.

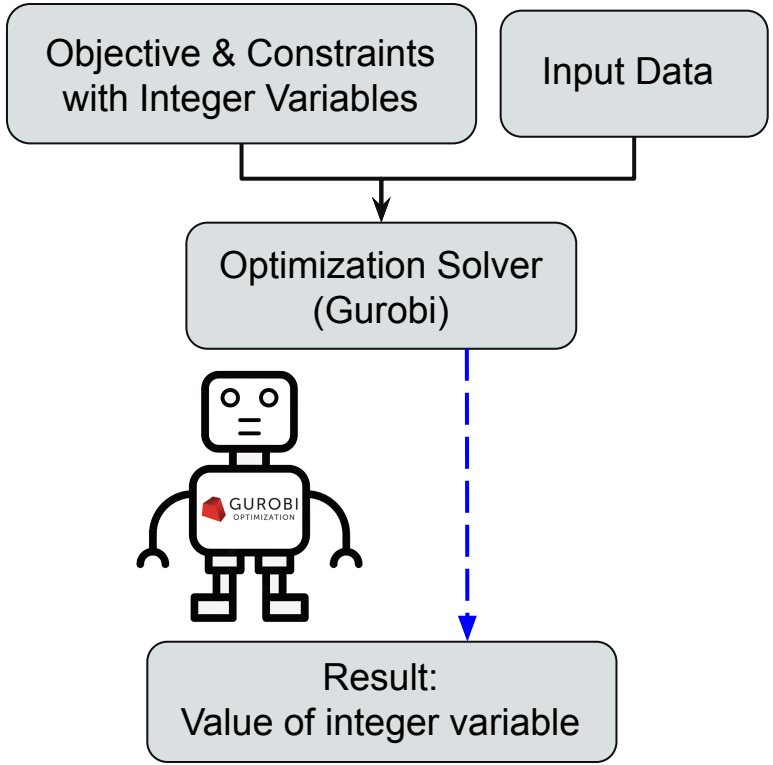
**Objective**

We want you to use as less bits as possible in the **Estimator**.

**Improvement**

**Estimator:** $\mathbf{x} = [x_1, x_2, x_3, x_4, x_5, x_6]$

**Trace i:** $\begin{cases} \mathbf{y}_i = [1, 3, 4] \\ ||\mathbf{z}||_1 = 4 \end{cases} \implies \begin{cases} x_1 + x_2 + x_3 + x_4 \geq 1 + 3 + 4 \\ x_1 + x_2 + x_3 \geq 1 + 3 \\ x_1 + x_2 \geq 1 \end{cases}$




WISCONSIN
UNIVERSITY OF WISCONSIN–MADISON