TurboTest: Learning When Less Is Enough Through Early Termination of Internet Speed Tests

Authors: Haarika Manda, Manshi Sagar, Yogesh, Kartikay Singh, Cindy Zhao, Tarun Mangla, Phillipa Gill, Elizabeth Belding, Arpit Gupta

Presentation by: Haarika Manda

[MAPRG] Google UC SANTA BARBARA

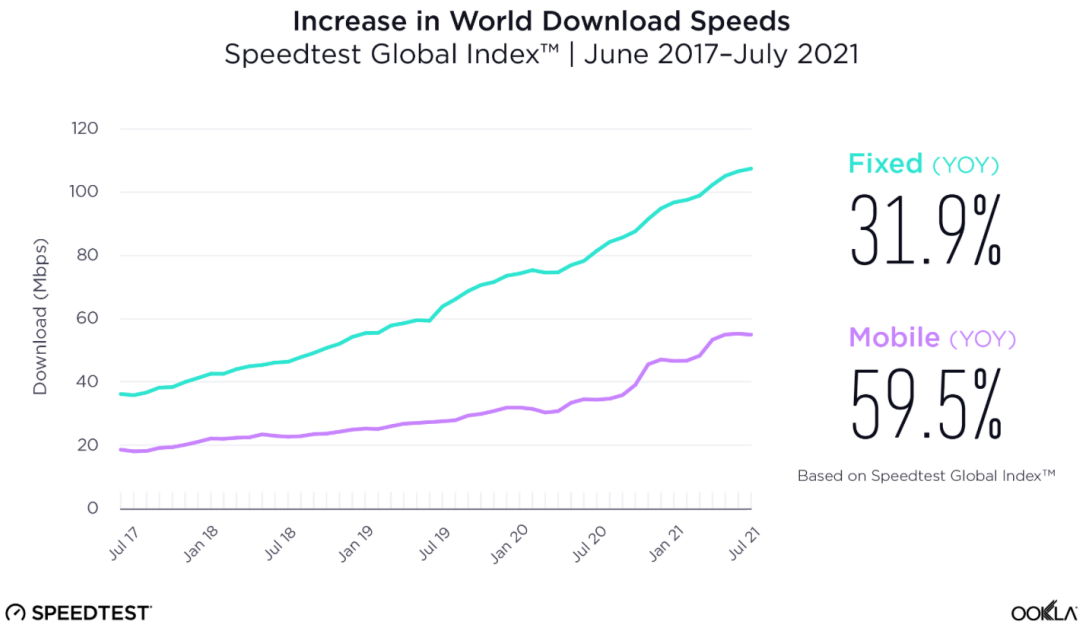
Speed Tests Impose Growing Costs at Scale

Speed Tests Impose Growing Costs at Scale

- **Expensive**
  - **As more users transfer to Gigabit connections, scaling infrastructure is problematic**

Speed Tests Impose Growing Costs at Scale

- **Expensive**
  - **As more users transfer to Gigabit connections, scaling infrastructure is problematic**

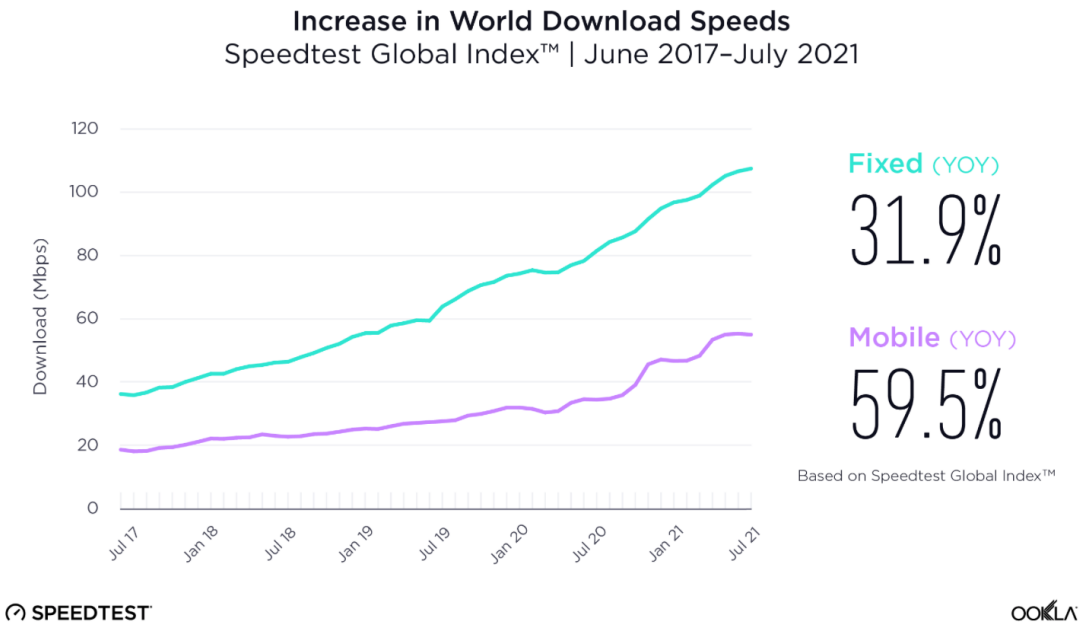


12 PB

per month at M-Lab
September 2024 (23x increase from 2023)

Speed Tests Impose Growing Costs at Scale

- **Expensive**
  - **As more users transfer to Gigabit connections, scaling infrastructure is problematic**
- **Redundant**
  - **Do we need to run 10+ second speed tests, for every connection ?**



12 PB

per month at M-Lab

September 2024 (23x increase from 2023)

Understanding the natural distribution of speed tests …

Understanding the natural distribution of speed tests …

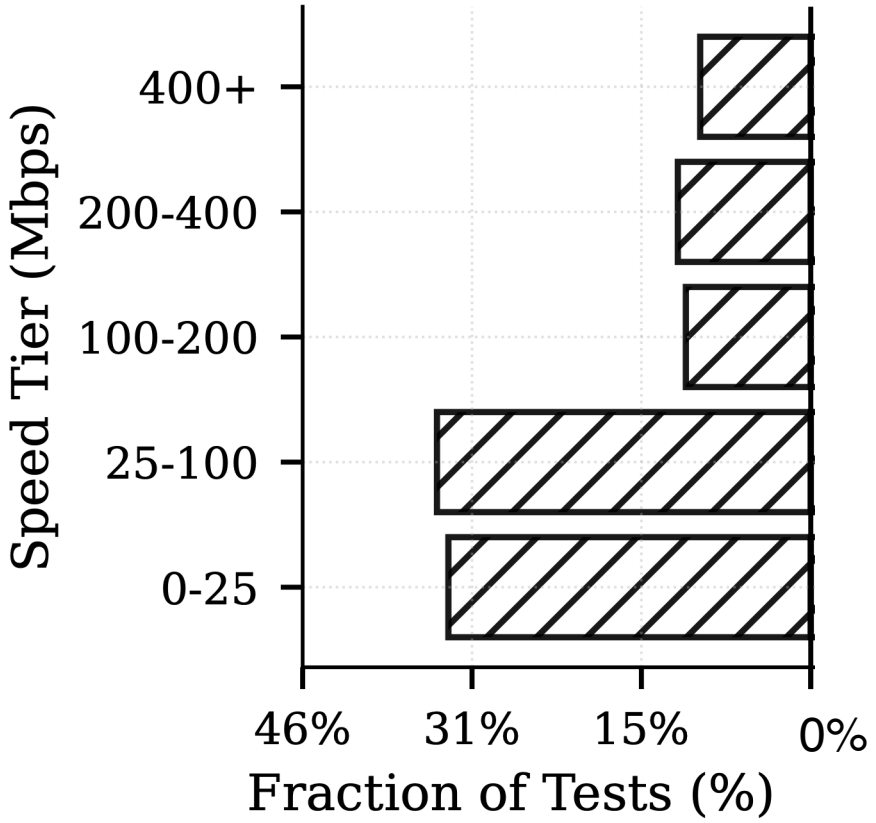
- Which types of tests contribute the most to data transfer costs ?

Understanding the natural distribution of speed tests …

- Which types of tests contribute the most to data transfer costs ?
- Let’s look at 1 million “Download” tests from (2024-2025)..

Understanding the natural distribution of speed tests …

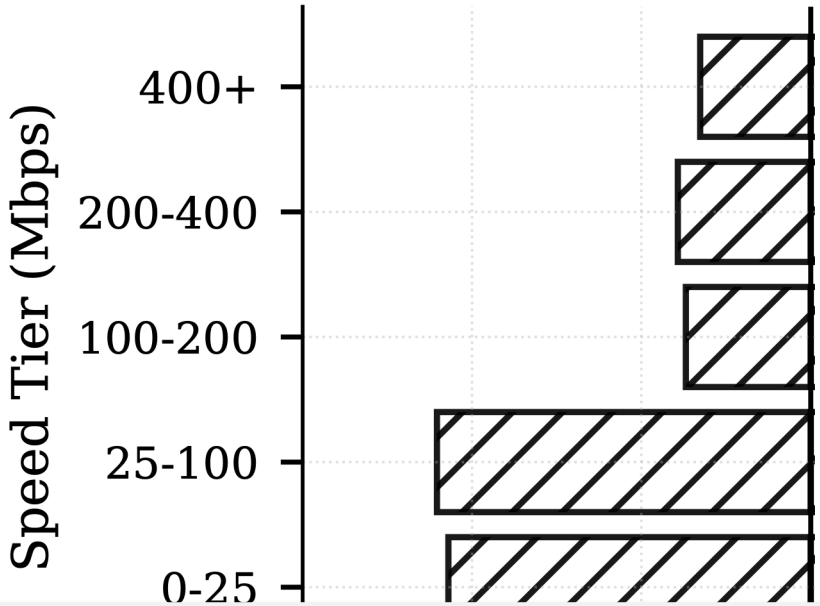
- Which types of tests contribute the most to data transfer costs ?
- Let's look at 1 million "Download" tests from (2024-2025)..



More than 30% of all speed tests are in low throughput categories…

Understanding the natural distribution of speed tests …

- Which types of tests contribute the most to data transfer costs ?
- Let's look at 1 million "Download" tests from (2024-2025)..

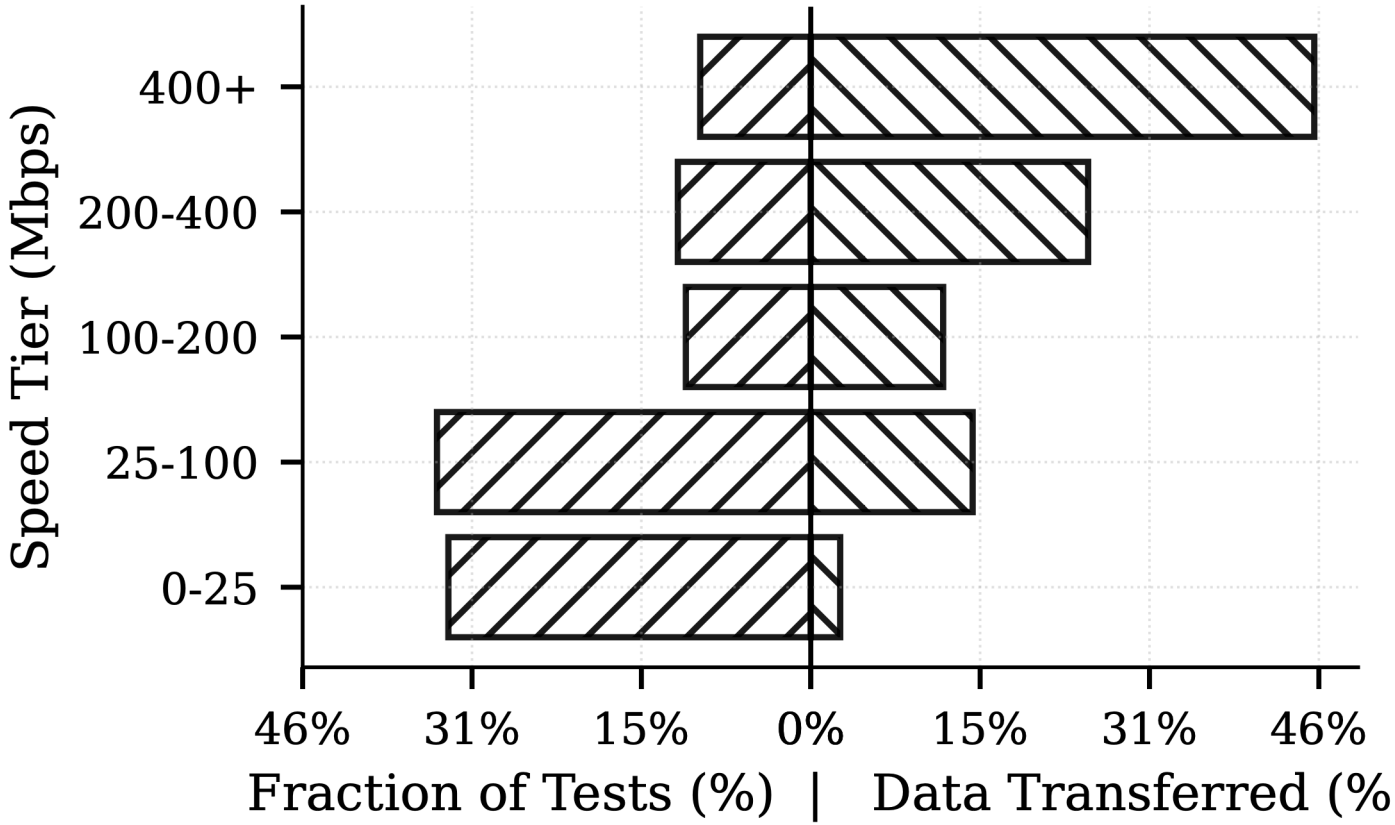


More than 30% of all speed tests are in low throughput categories…

Let's look at the raw data transferred …

High-Throughput Tests Drive Most Data Transfer Costs

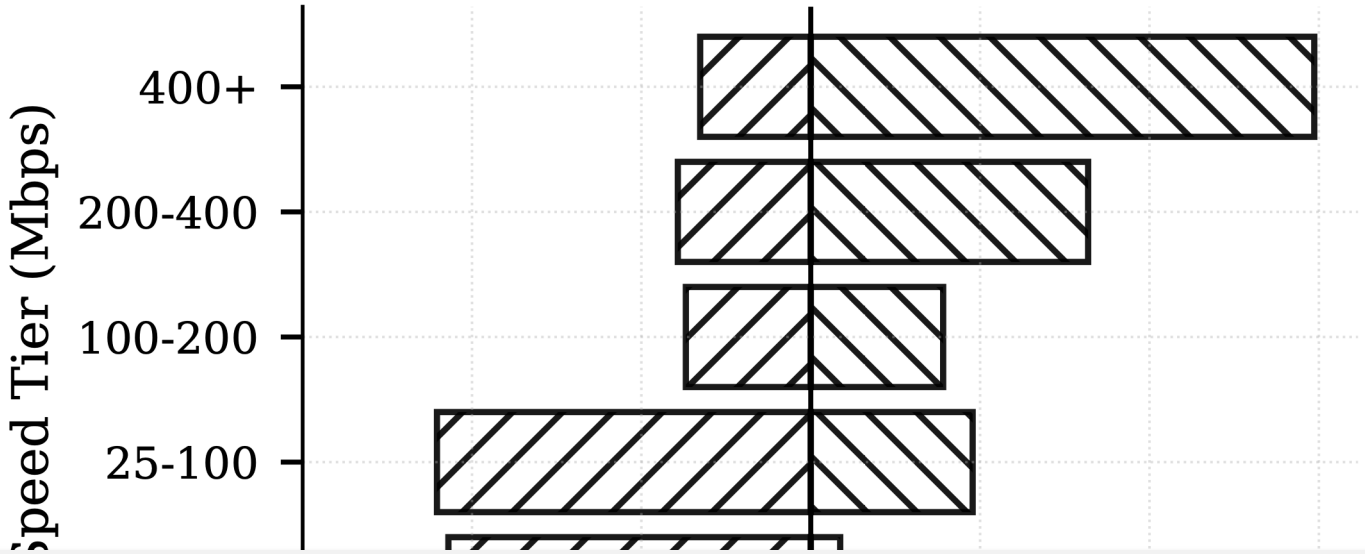
- Which types of tests contribute the most to data transfer costs ?
- Let's look at 1 million "Download" tests from (2024-2025)..



Gigabit connections are small in volume .. contribute to a lot of the data transferred

High-Throughput Tests Drive Most Data Transfer Costs

- Which types of tests contribute the most to data transfer costs ?
- Let's look at 1 million "Download" tests from (2024-2025)..

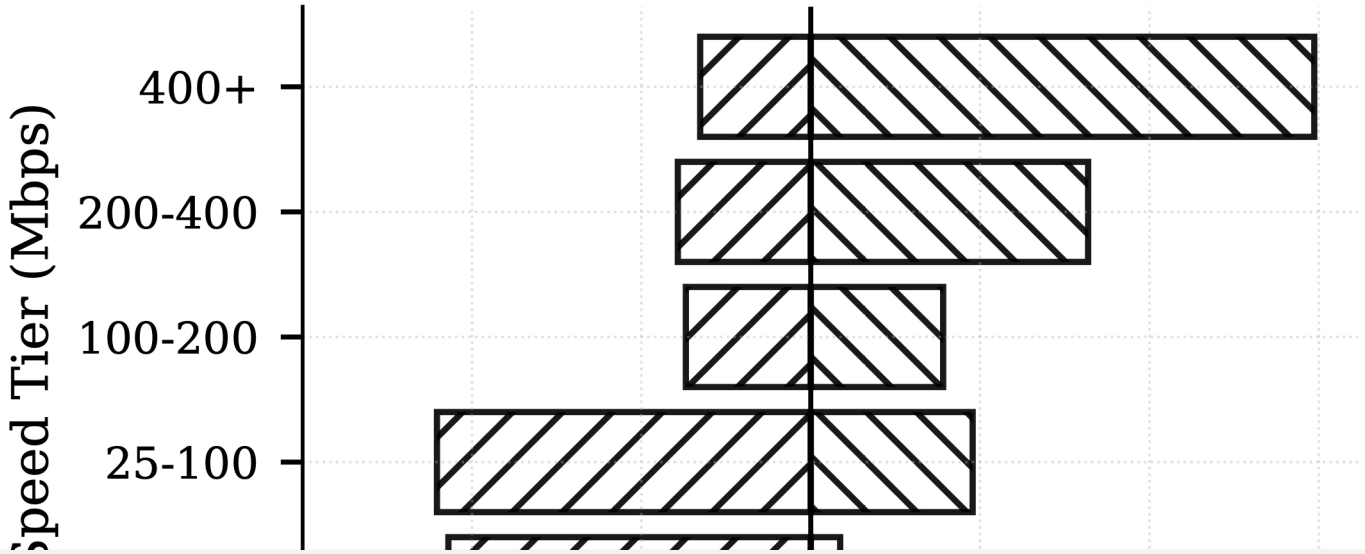


Gigabit connections are small in volume .. contribute to a lot of the data transferred

Can we send fewer packets/probes and still report an accurate throughput value ?

High-Throughput Tests Drive Most Data Transfer Costs

- Which types of tests contribute the most to data transfer costs ?
- Let's look at 1 million "Download" tests from (2024-2025)..

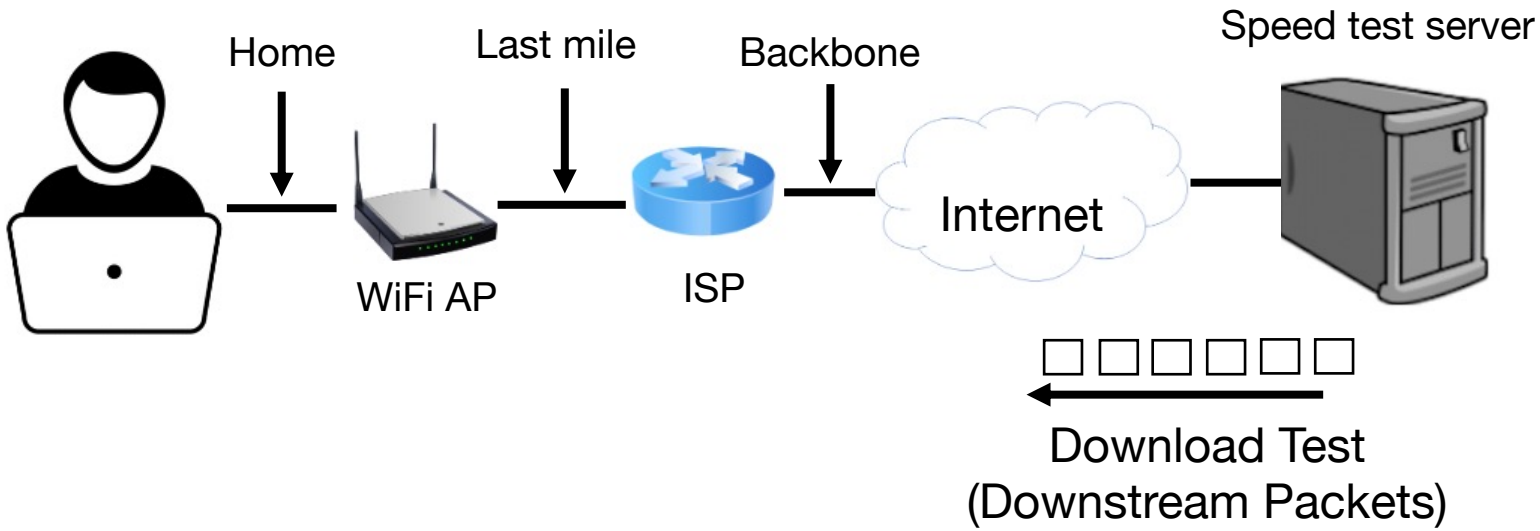


Gigabit connections are small in volume .. contribute to a lot of the data transferred

Can we send fewer packets/probes and still report an accurate throughput value ?

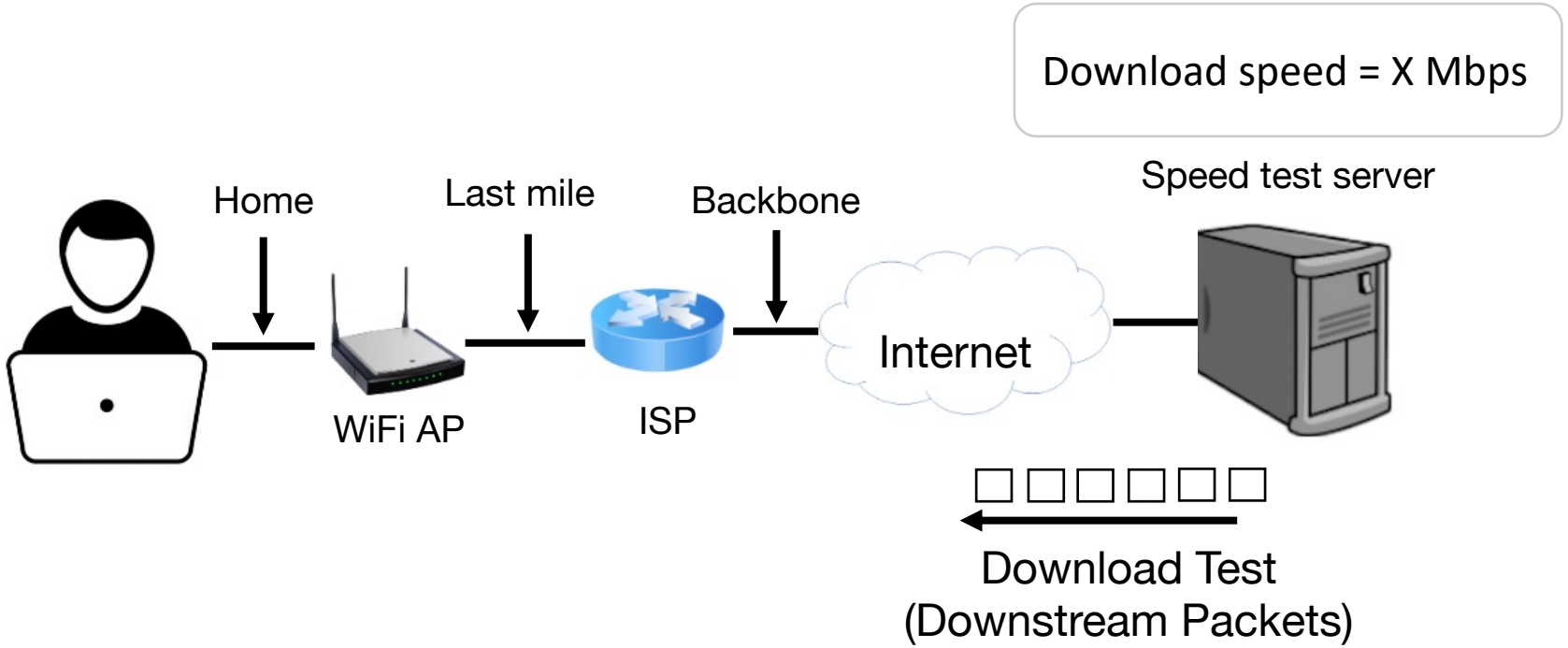
The Early Termination Problem

Let's look at an example …..



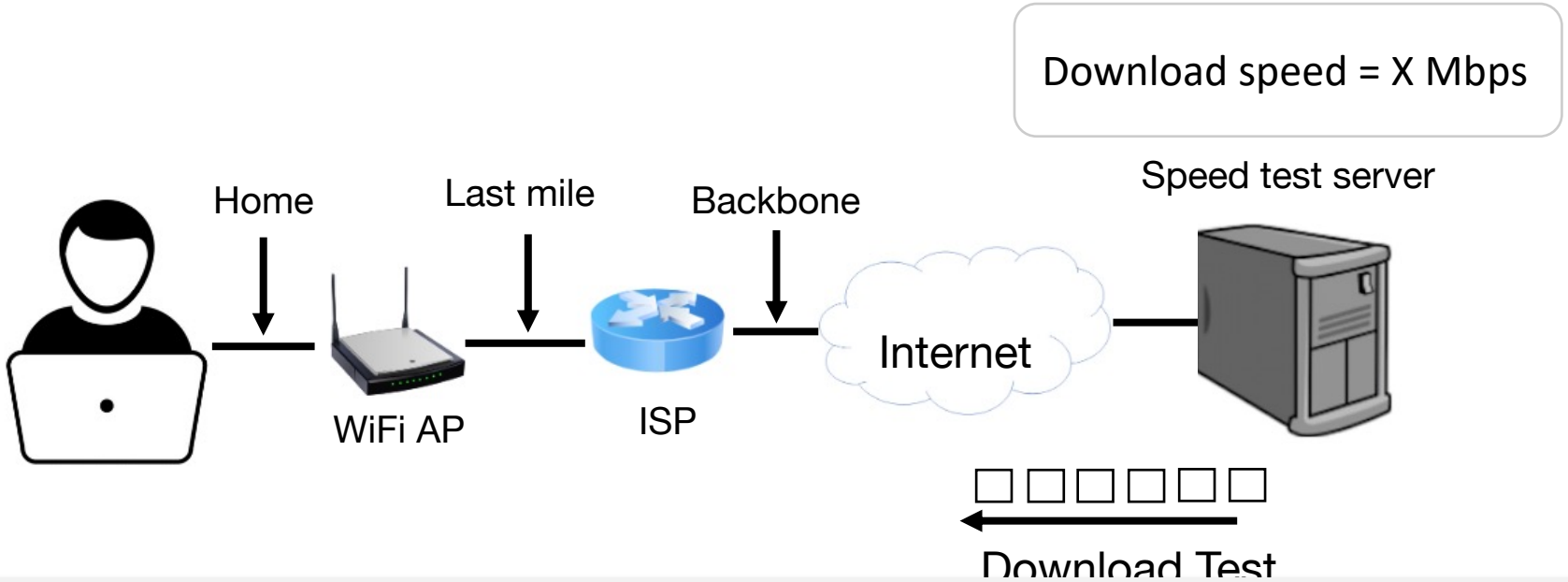
The Early Termination Problem

Let's look at an example …..



The Early Termination Problem

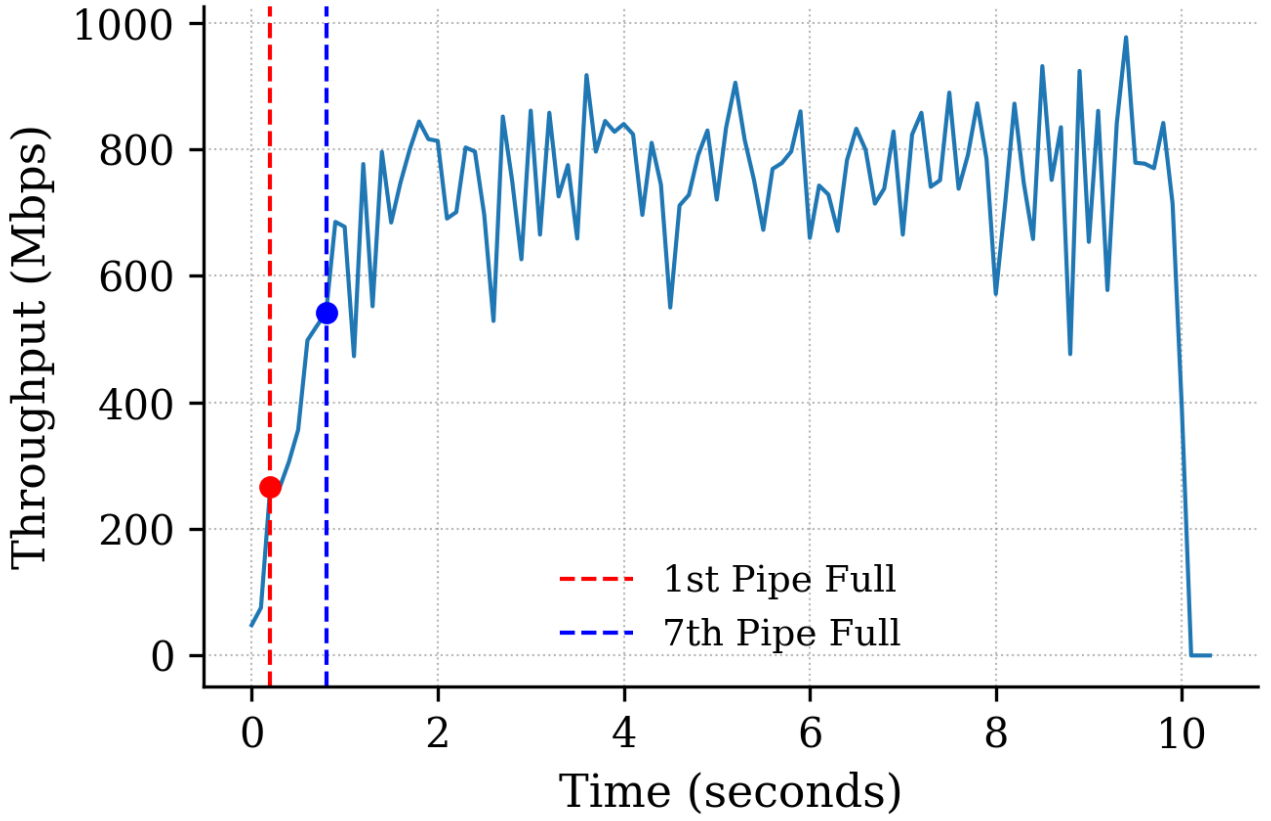
Let's look at an example …..



Can we stop a test early without sacrificing the accuracy of the reported throughput?

Can BBR Pipe-Full Signals Stop Tests Early?

- **Let's try leveraging transport layer signals from Congestion Control (ex: BBR[1])**

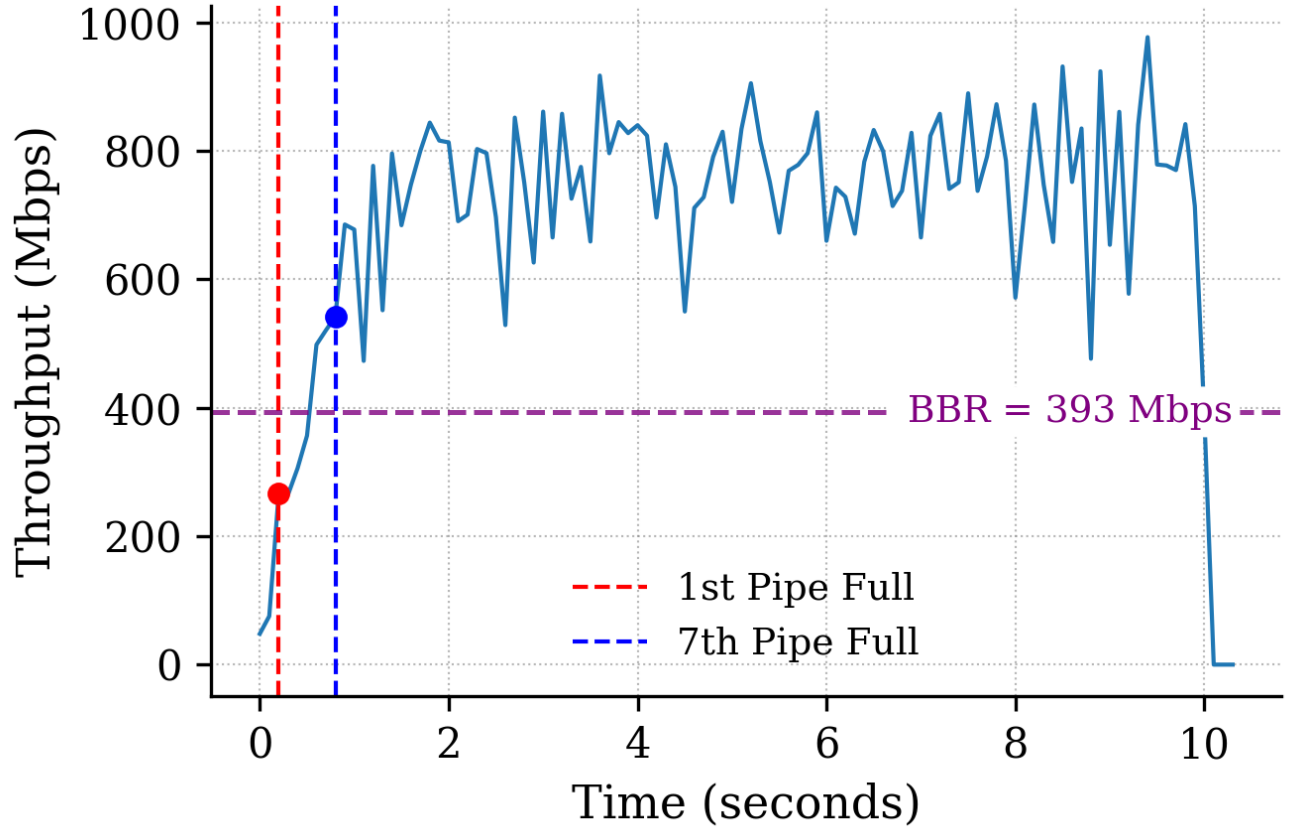


Ground Throughput (over 10 seconds) - **687** Mbps

1. Phillipa Gill, Cristina Leon, Neal Cardwell, and Christophe Diot. 2025. The Case for Leveraging Transport Signals to Improve Internet Speed Test Efficiency. SIGCOMM Comput. Commun. Rev. 55, 2 (April 2025), 23–28. https://doi.org/10.1145/3750832.3750836

Can BBR Pipe-Full Signals Stop Tests Early?

- Let's try leveraging transport layer signals from Congestion Control (ex: BBR[1])
- The BBR heuristic has a relative error of **42%** and is off by 200+ Mbps!

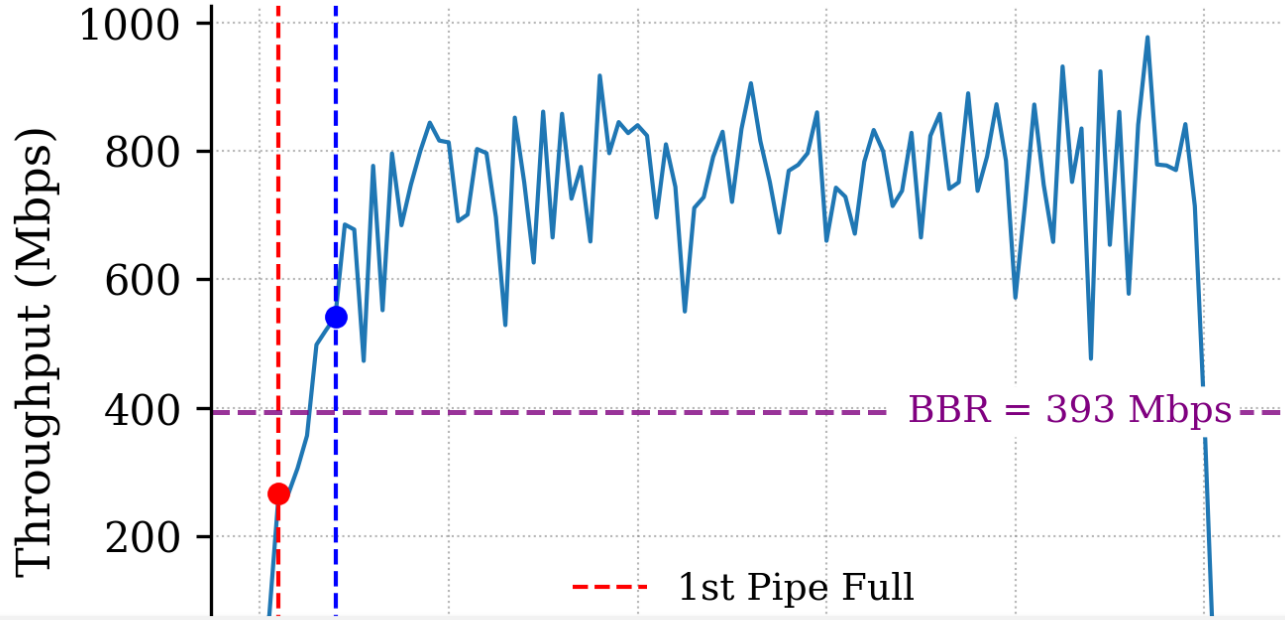


Ground Throughput (over 10 seconds) - **687** Mbps

1. Phillipa Gill, Cristina Leon, Neal Cardwell, and Christophe Diot. 2025. The Case for Leveraging Transport Signals to Improve Internet Speed Test Efficiency. SIGCOMM Comput. Commun. Rev. 55, 2 (April 2025), 23–28. https://doi.org/10.1145/3750832.3750836

Can BBR Pipe-Full Signals Stop Tests Early?

- Let's try leveraging transport layer signals from Congestion Control (ex: BBR[1])
- The BBR heuristic has a relative error of 42% and is off by 200+ Mbps!

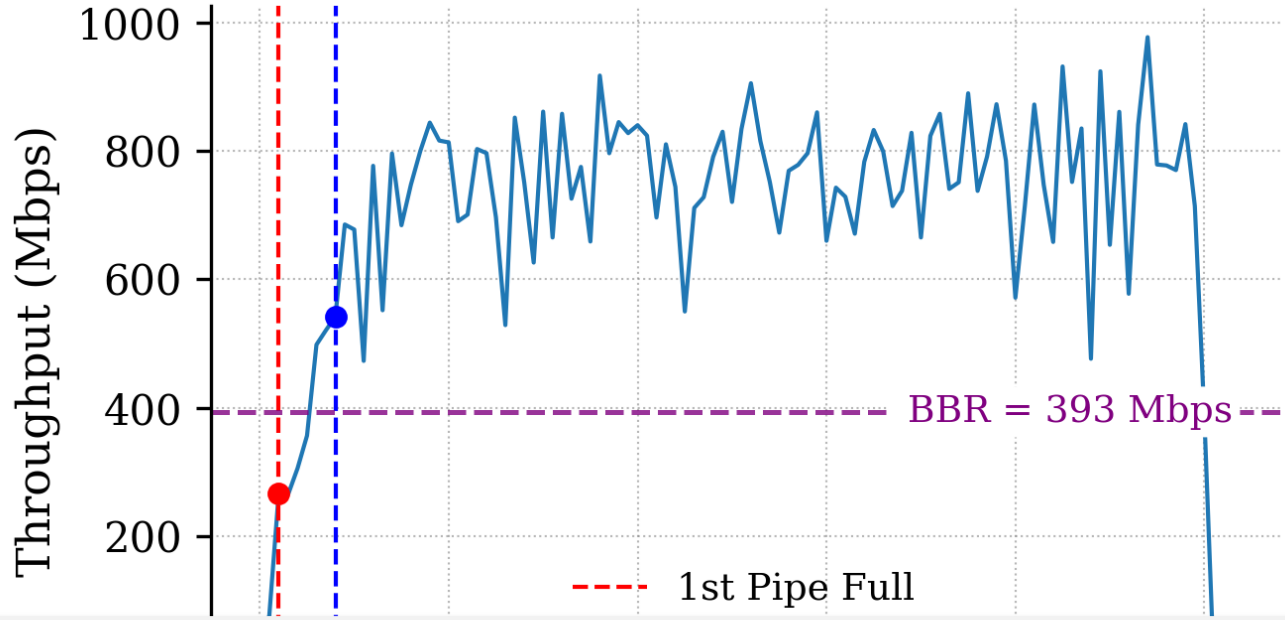


BBR fails to balance the tradeoff between data savings and accuracy

1. Phillipa Gill, Cristina Leon, Neal Cardwell, and Christophe Diot. 2025. The Case for Leveraging Transport Signals to Improve Internet Speed Test Efficiency. SIGCOMM Comput. Commun. Rev. 55, 2 (April 2025), 23–28. https://doi.org/10.1145/3750832.3750836

Can BBR Pipe-Full Signals Stop Tests Early?

- Let's try leveraging transport layer signals from Congestion Control (ex: BBR[1])
- The BBR heuristic has a relative error of **42%** and is off by 200+ Mbps!



Pipe-full ≠ throughput stable

BBR fails to balance the tradeoff between data savings and accuracy

1. Phillipa Gill, Cristina Leon, Neal Cardwell, and Christophe Diot. 2025. The Case for Leveraging Transport Signals to Improve Internet Speed Test Efficiency. SIGCOMM Comput. Commun. Rev. 55, 2 (April 2025), 23–28. https://doi.org/10.1145/3750832.3750836

Insights Into Early Termination

Goal: Predict a value close to the ground truth (Download speed for 10s test), while balancing ….

- **Low prediction error**
- **Low data transfer**

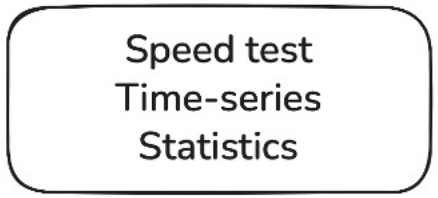
Can we decompose early termination into 2 learning strategies (using ML)?

Insights Into Early Termination

Goal: Predict a value close to the ground truth (Download speed for 10s test), while balancing ….

- **Low prediction error**
- **Low data transfer**

Can we decompose early termination into 2 learning strategies (using ML)?



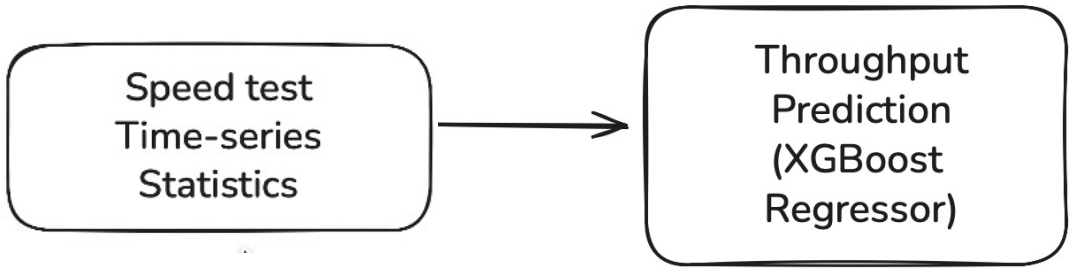
Insights Into Early Termination

Goal: Predict a value close to the ground truth (Download speed for 10s test), while balancing ….

- **Low prediction error**
- **Low data transfer**

Can we decompose early termination into 2 learning strategies (using ML)?

- **Prediction** (Given the time-series of statistics, accurately predict the throughput)



Minimize: $(\hat{y}_{\text{pred}} - y_{\text{final}})^2$

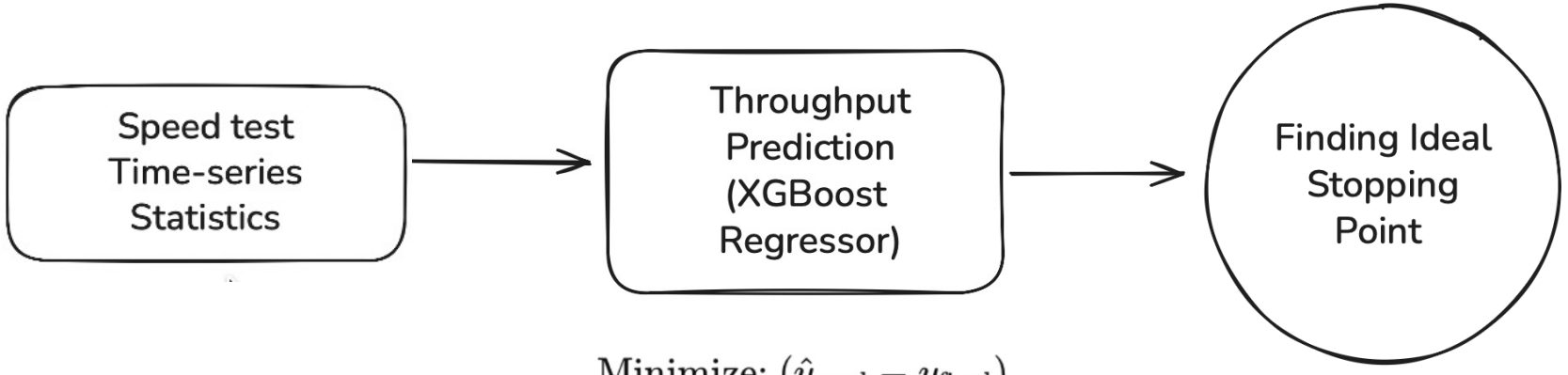
Insights Into Early Termination

Goal: Predict a value close to the ground truth (Download speed for 10s test), while balancing ….

- **Low prediction error**
- **Low data transfer**

Can we decompose early termination into 2 learning strategies (using ML)?

- **Prediction** (Given the time-series of statistics, accurately predict the throughput)
- **Termination** (Find an ideal stopping point, which saves data)



Minimize: $(\hat{y}_{\text{pred}} - y_{\text{final}})$

$$\tau_{\text{ideal}} = \text{timestamp when } \frac{|y_{\text{pred}} - y_{\text{final}}|}{y_{\text{final}}} < \epsilon$$

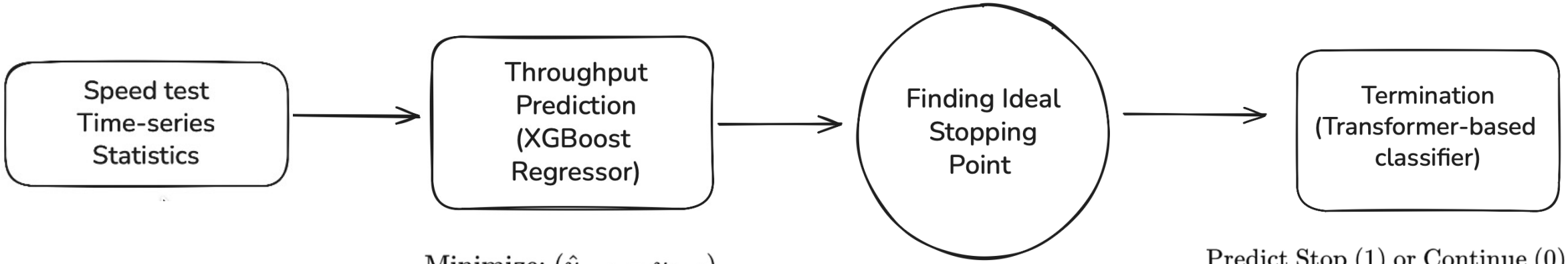
Insights Into Early Termination

Goal: Predict a value close to the ground truth (Download speed for 10s test), while balancing ….

- **Low prediction error**
- **Low data transfer**

Can we decompose early termination into 2 learning strategies (using ML)?

- **Prediction** (Given the time-series of statistics, accurately predict the throughput)
- **Termination** (Find an ideal stopping point, which saves data)



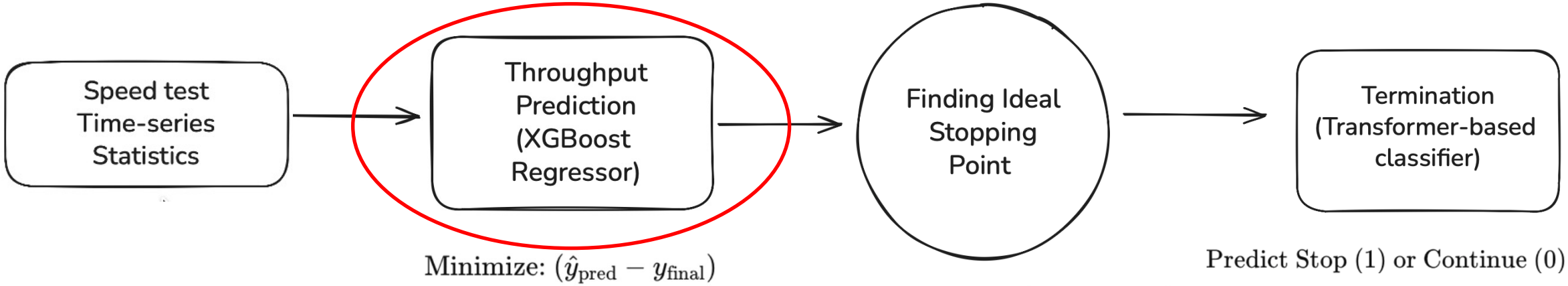
Insights Into Early Termination

Goal: Predict a value close to the ground truth (Download speed for 10s test), while balancing ….

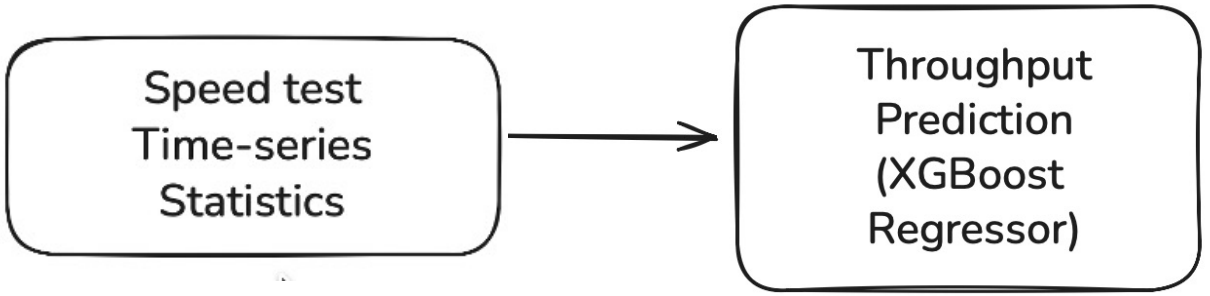
- **Low prediction error**
- **Low data transfer**

Can we decompose early termination into 2 learning strategies (using ML)?

- **Prediction** (Given the time-series of statistics, accurately predict the throughput)
- **Termination** (Find an ideal stopping point, which saves data)



Stage 1: Predicting Final Throughput

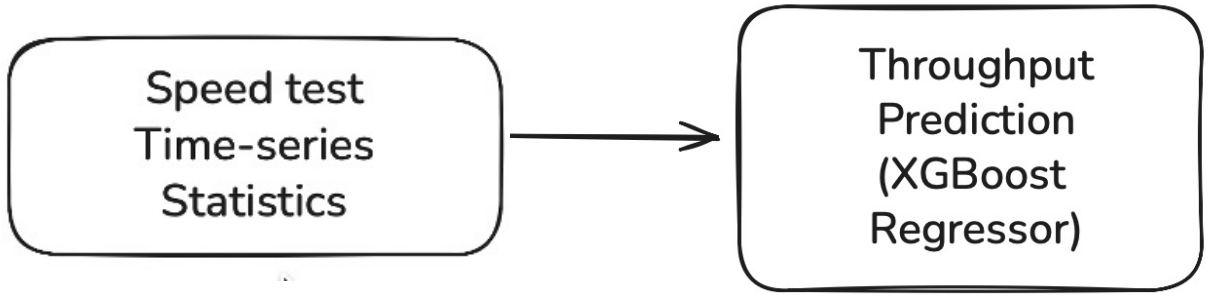


Minimize: $(\hat{y}_{\text{pred}} - y_{\text{final}})^2$

Stage 1: Predicting Final Throughput

Time-Series Input (TCP info)

- Throughput samples + RTT values
- Congestion window size
- Retransmissions + duplicate ACK
- Sampled at 100ms granularity

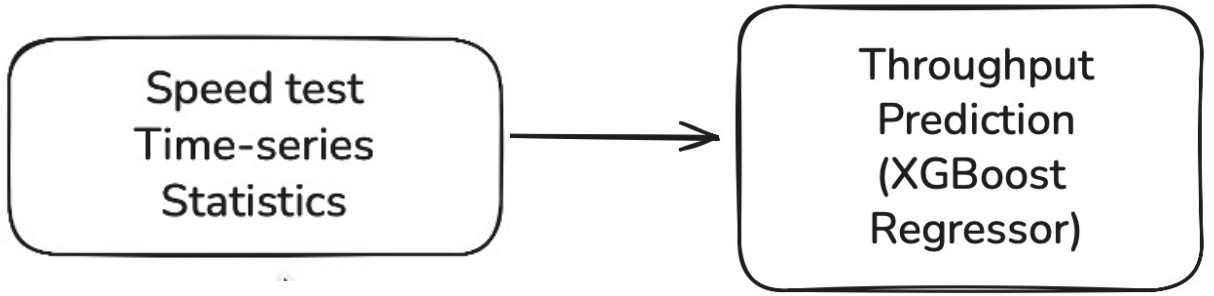


Minimize: $(\hat{y}_{\text{pred}} - y_{\text{final}})^2$

Stage 1: Predicting Final Throughput

Time-Series Input (TCP info)

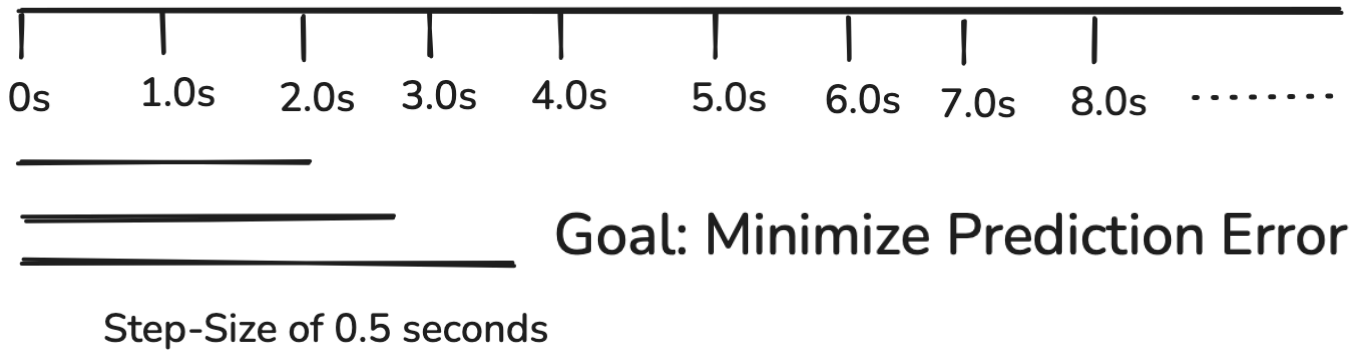
- Throughput samples + RTT values
- Congestion window size
- Retransmissions + duplicate ACK
- Sampled at 100ms granularity



Minimize: $(\hat{y}_{\text{pred}} - y_{\text{final}})^2$

(Example) Ground Truth: 50 Mbps

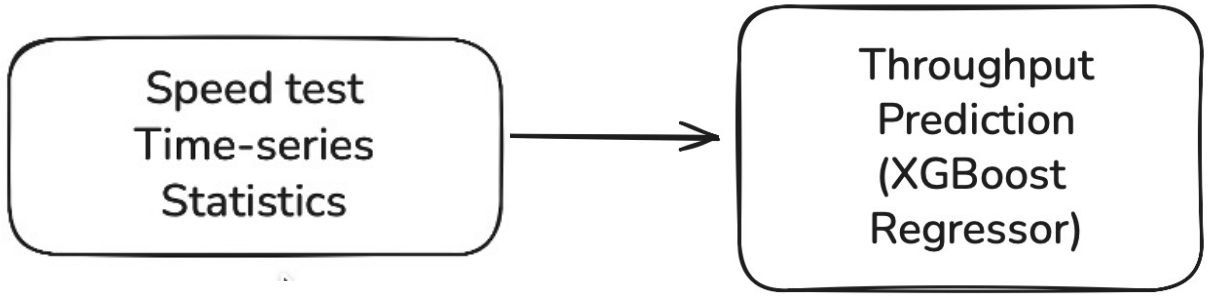
Input for Regressor: 13 features x 20 windows (for 2 seconds)



Stage 1: Predicting Final Throughput

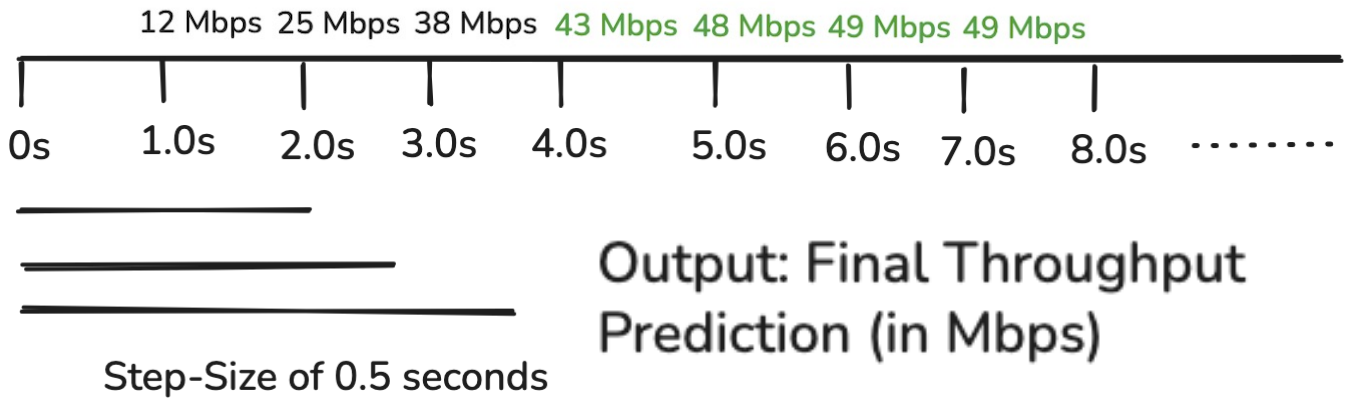
Time-Series Input (TCP info)

- Throughput samples + RTT values
- Congestion window size
- Retransmissions + duplicate ACK
- Sampled at 100ms granularity



Minimize: $(\hat{y}_{\text{pred}} - y_{\text{final}})^2$

(Example) Ground Truth: 50 Mbps

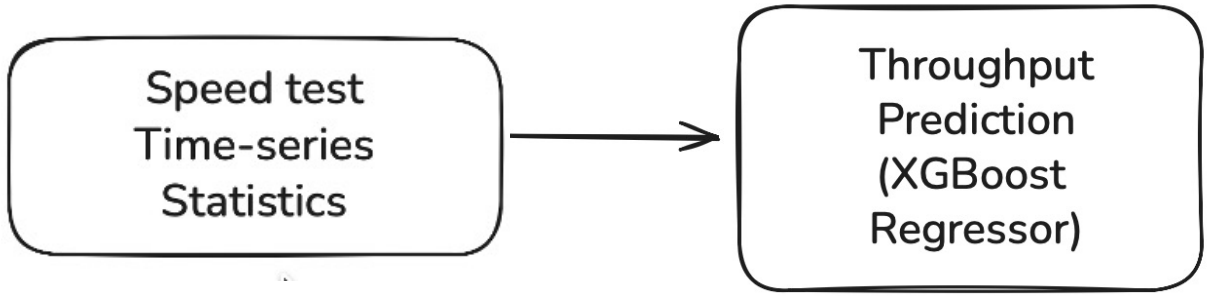


Output: Final Throughput Prediction (in Mbps)

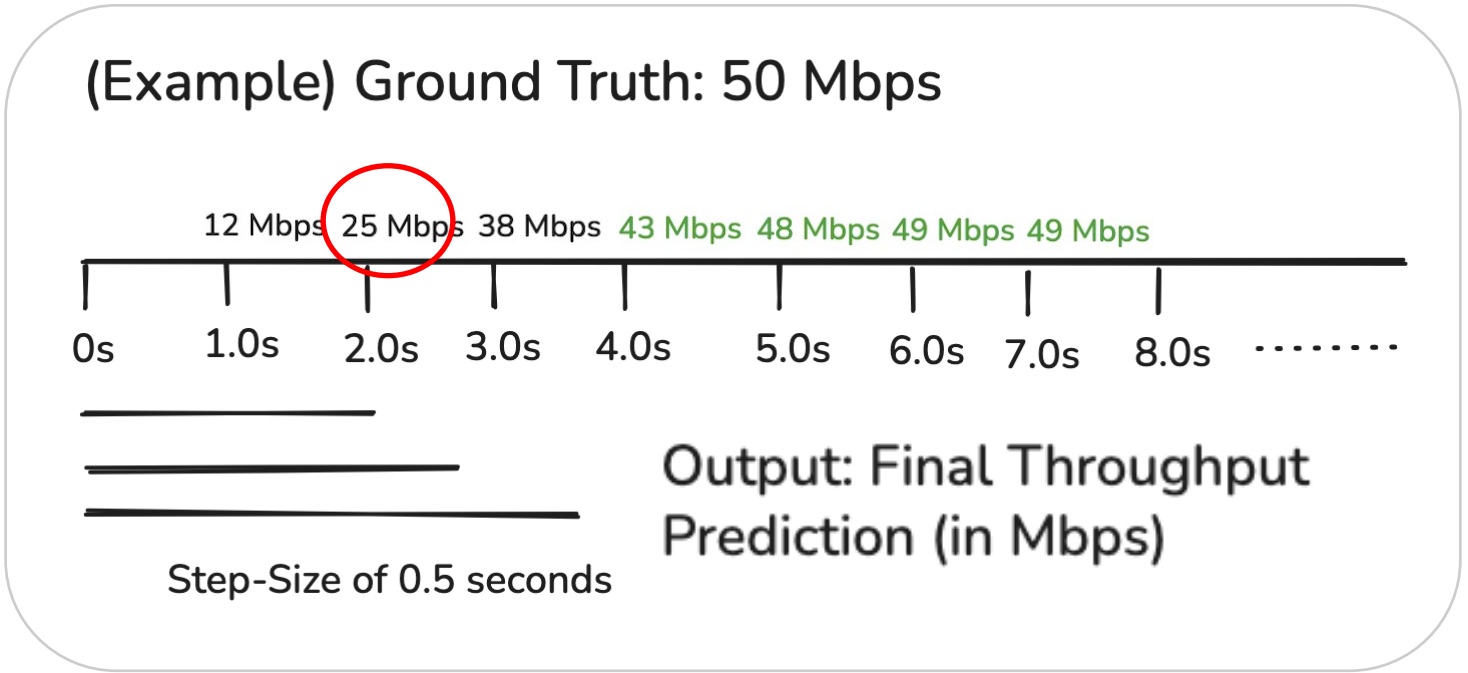
Stage 1: Predicting Final Throughput

Time-Series Input (TCP info)

- Throughput samples + RTT values
- Congestion window size
- Retransmissions + duplicate ACK
- Sampled at 100ms granularity



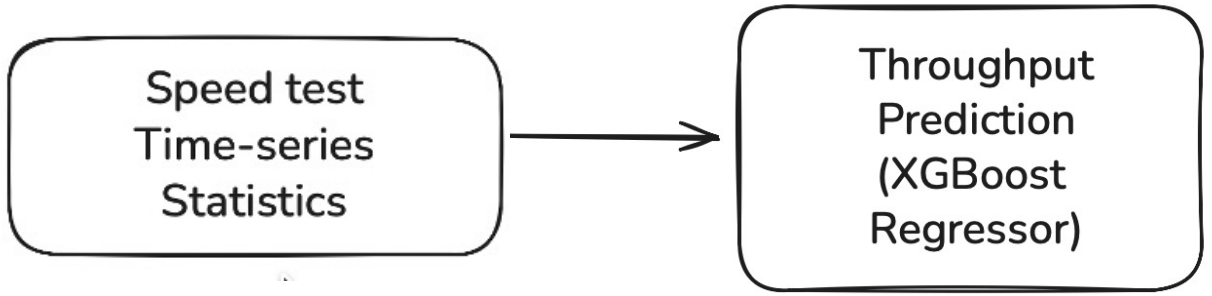
Minimize: $(\hat{y}_{\text{pred}} - y_{\text{final}})^2$



Stage 1: Predicting Final Throughput

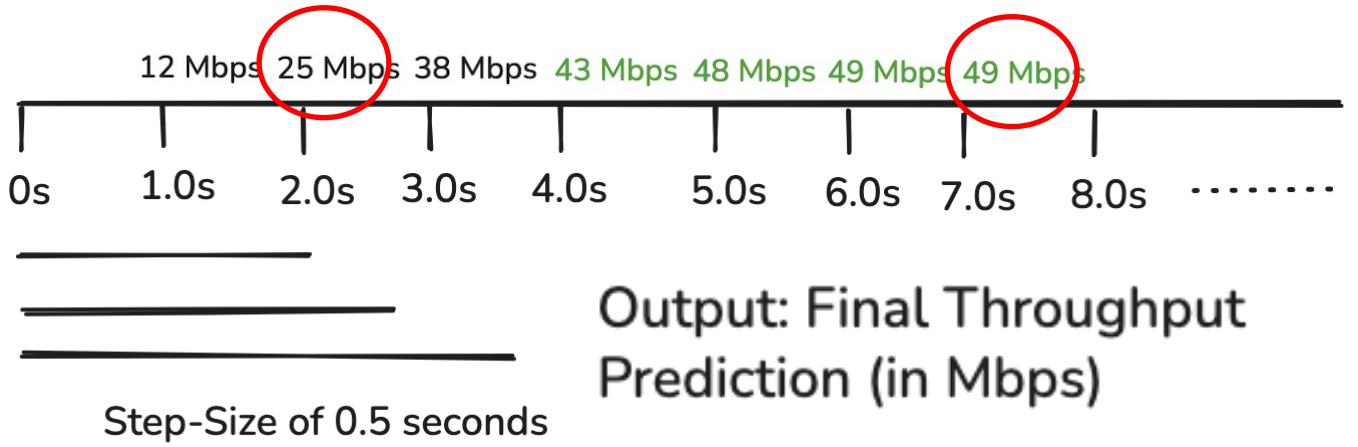
Time-Series Input (TCP info)

- Throughput samples + RTT values
- Congestion window size
- Retransmissions + duplicate ACK
- Sampled at 100ms granularity



Minimize: $(\hat{y}_{\text{pred}} - y_{\text{final}})^2$

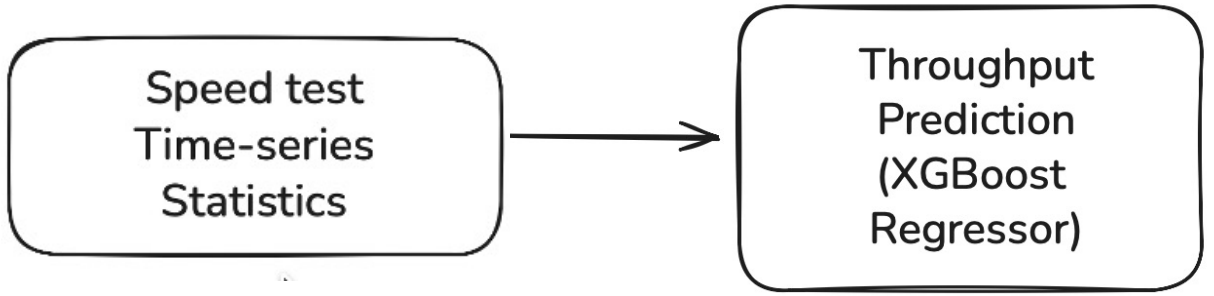
(Example) Ground Truth: 50 Mbps



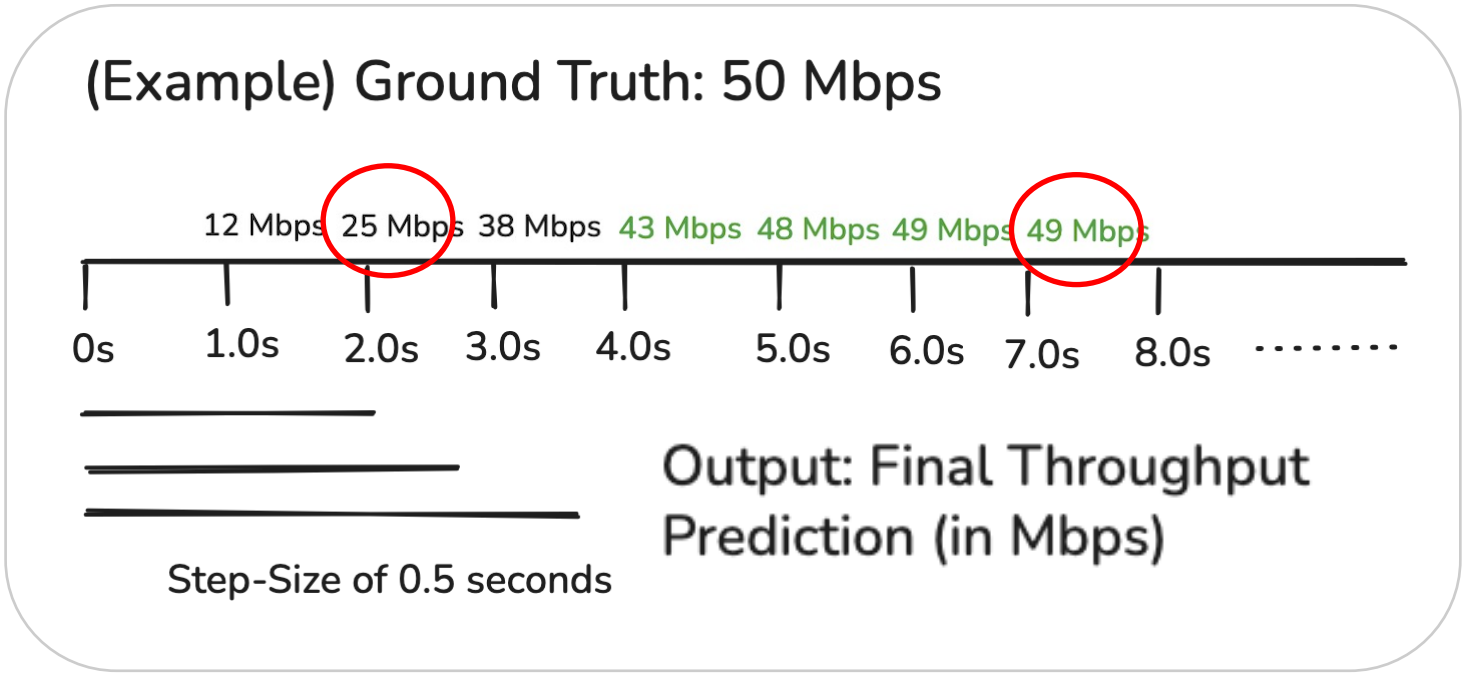
Stage 1: Predicting Final Throughput

Time-Series Input (TCP info)

- Throughput samples + RTT values
- Congestion window size
- Retransmissions + duplicate ACK
- Sampled at 100ms granularity



Minimize: $(\hat{y}_{\text{pred}} - y_{\text{final}})^2$

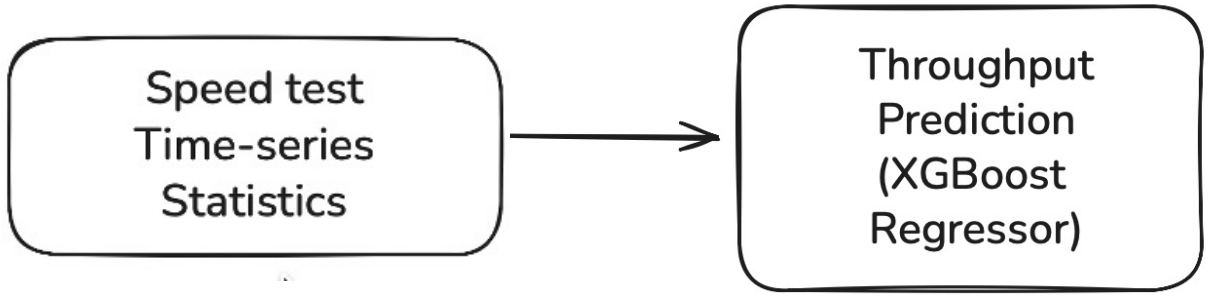


Accuracy vs Data Savings

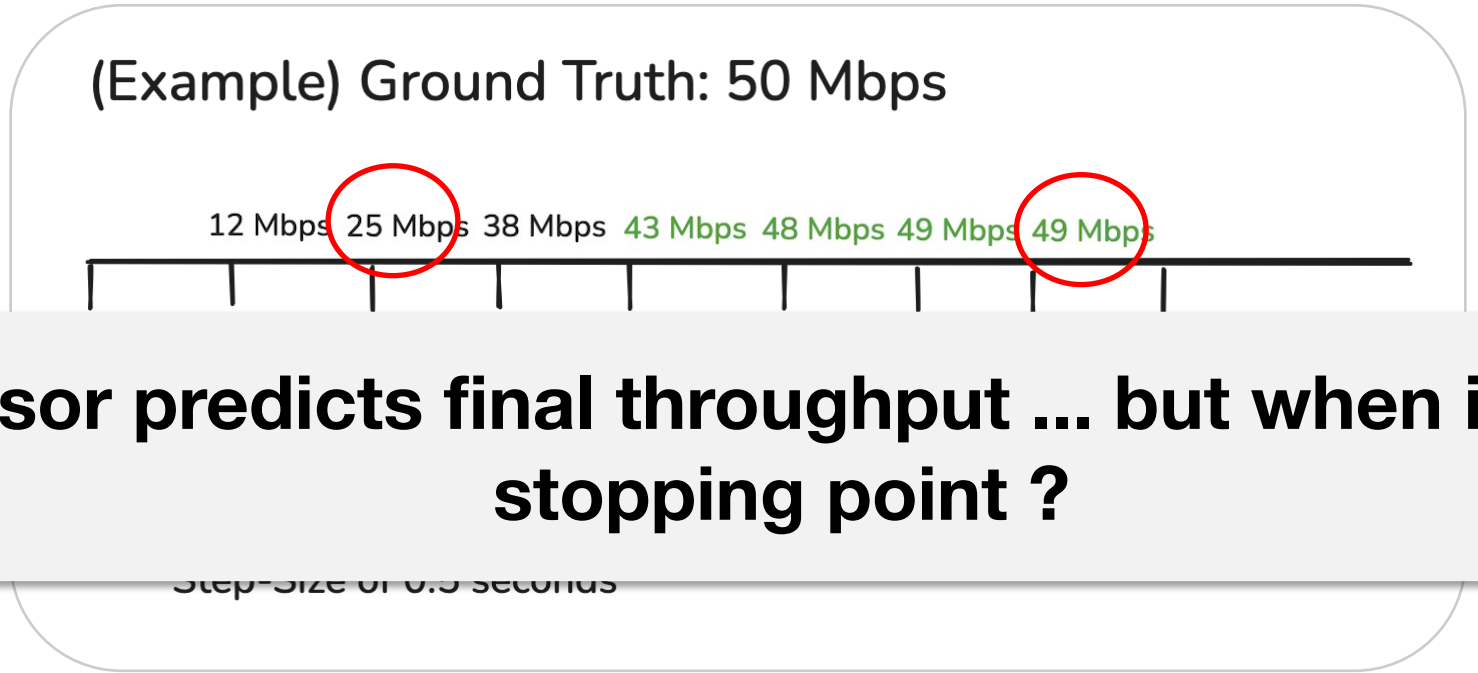
Stage 1: Predicting Final Throughput

Time-Series Input (TCP info)

- Throughput samples + RTT values
- Congestion window size
- Retransmissions + duplicate ACK
- Sampled at 100ms granularity



Minimize: $(\hat{y}_{\text{pred}} - y_{\text{final}})^2$



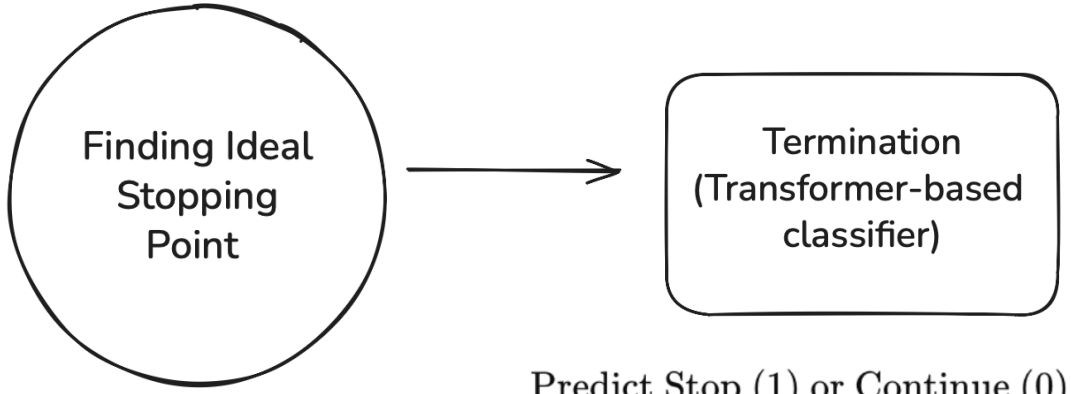
Accuracy vs Data Savings

Our regressor predicts final throughput ... but when is the optimal stopping point ?

Stage 2: Learning When to Stop

Labeling Ideal Stopping point

- For each test, find the earliest point (t*) where Stage 1 error falls within tolerance $\boldsymbol{\varepsilon}$ (%)
- ε ∈ {5%, 10%, 15% ..} - operator-tunable tradeoff



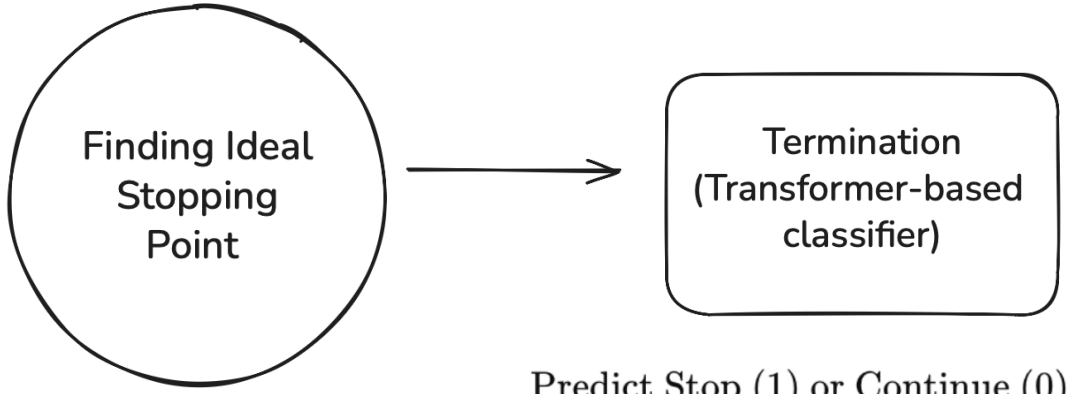
Predict Stop (1) or Continue (0) at each τ

$$\tau_{\text{ideal}} = \text{timestamp when } \frac{|y_{\text{pred}} - y_{\text{final}}|}{y_{\text{final}}} < \epsilon$$

Stage 2: Learning When to Stop

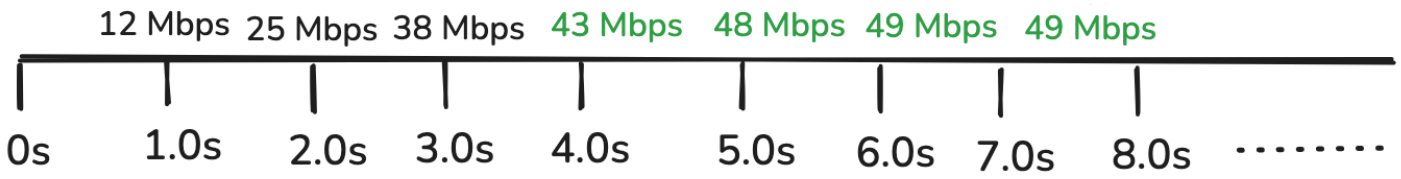
Labeling Ideal Stopping point

- For each test, find the earliest point (t*) where Stage 1 error falls within tolerance $\boldsymbol{\varepsilon}$ (%)
- ε ∈ {5%, 10%, 15% ..} - operator-tunable tradeoff



$$\tau_{\text{ideal}} = \text{timestamp when } \frac{|y_{\text{pred}} - y_{\text{final}}|}{y_{\text{final}}} < \epsilon$$

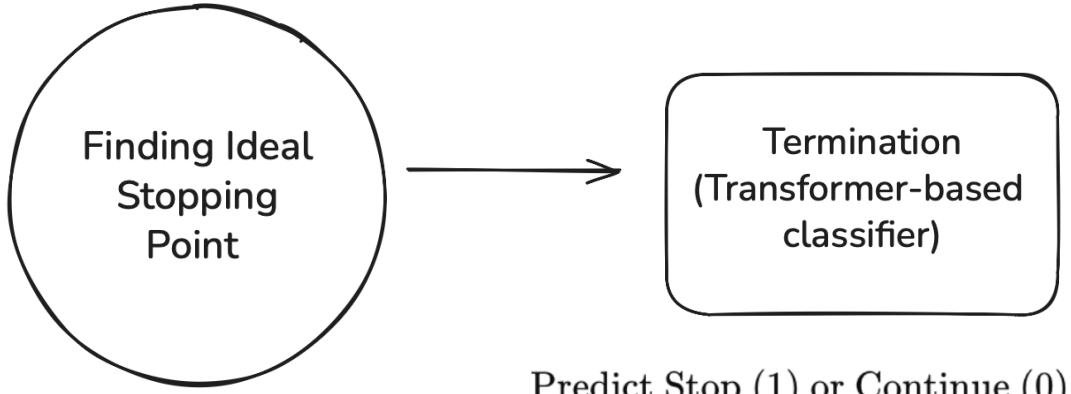
Example: Ground Truth: 50 Mbps,



Stage 2: Learning When to Stop

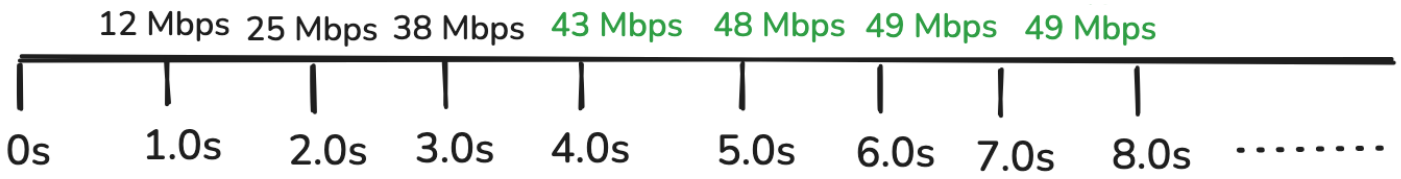
Labeling Ideal Stopping point

- For each test, find the earliest point (t*) where Stage 1 error falls within tolerance $\boldsymbol{\varepsilon}$ (%)
- ε ∈ {5%, 10%, 15% ..} - operator-tunable tradeoff



$$\tau_{\text{ideal}} = \text{timestamp when } \frac{|y_{\text{pred}} - y_{\text{final}}|}{y_{\text{final}}} < \epsilon$$

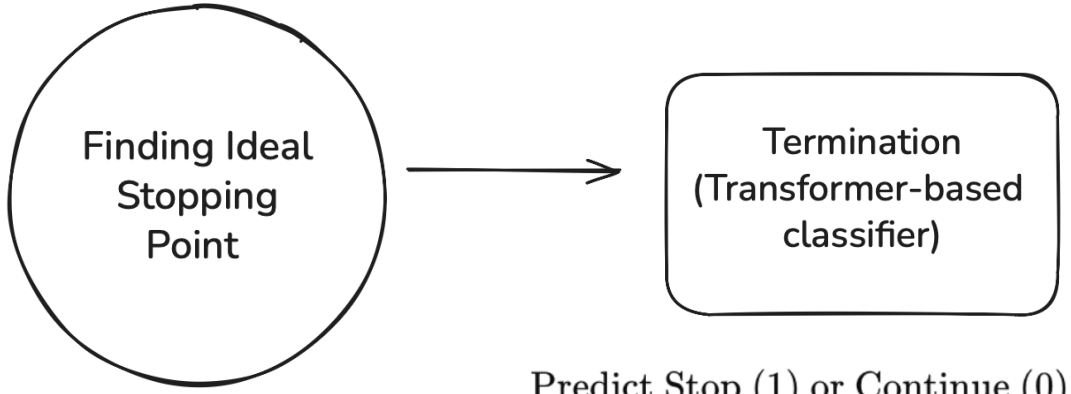
Example: Ground Truth: 50 Mbps, ϵ = 15 (%)



Stage 2: Learning When to Stop

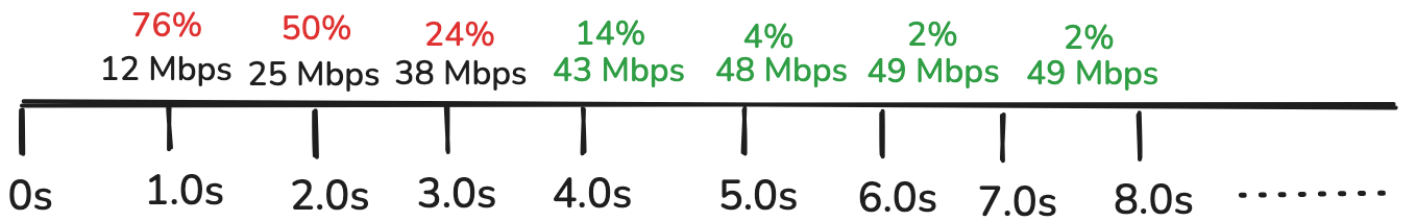
Labeling Ideal Stopping point

- For each test, find the earliest point (t*) where Stage 1 error falls within tolerance $\boldsymbol{\varepsilon}$ (%)
- ε ∈ {5%, 10%, 15% ..} - operator-tunable tradeoff



$$\tau_{\text{ideal}} = \text{timestamp when } \frac{|y_{\text{pred}} - y_{\text{final}}|}{y_{\text{final}}} < \epsilon$$

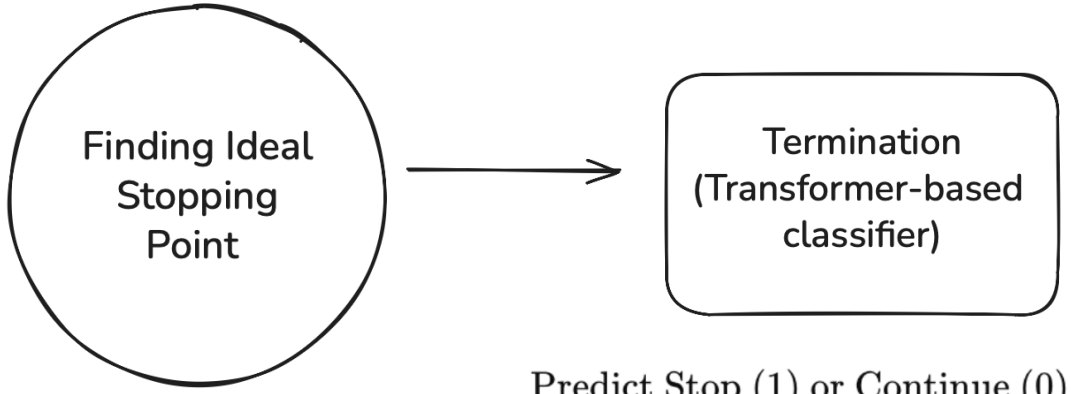
Example: Ground Truth: 50 Mbps, ϵ = 15 (%)



Stage 2: Learning When to Stop

Labeling Ideal Stopping point

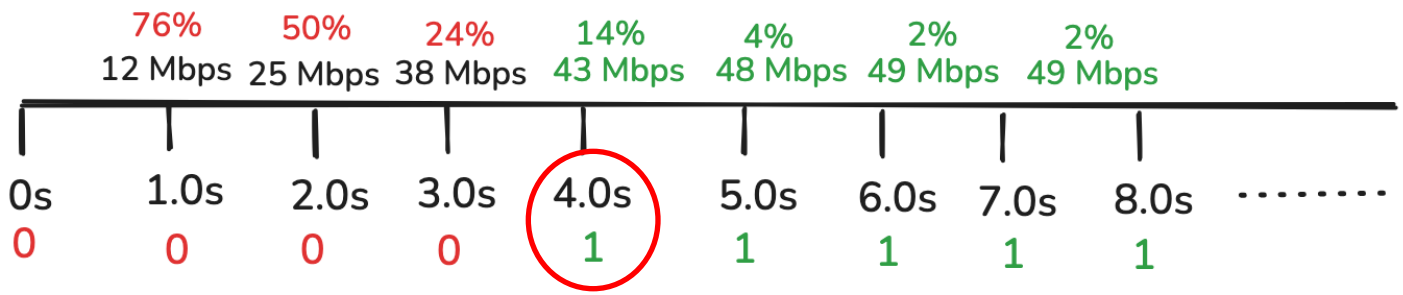
- For each test, find the earliest point (t*) where Stage 1 error falls within tolerance $\boldsymbol{\varepsilon}$ (%)
- ε ∈ {5%, 10%, 15% ..} - operator-tunable tradeoff



$$\tau_{\text{ideal}} = \text{timestamp when } \frac{|y_{\text{pred}} - y_{\text{final}}|}{y_{\text{final}}} < \epsilon$$

Example: Ground Truth: 50 Mbps, ϵ = 15 (%)

Label for transformer (classifier): STOP (1) /CONTINUE (0)

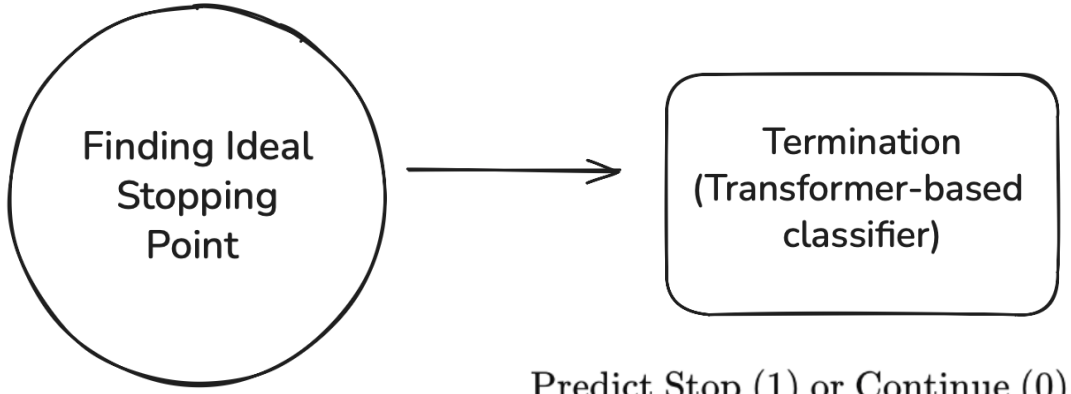


Ideal Stopping Point (t*) = 4 seconds

Stage 2: Learning When to Stop

Labeling Ideal Stopping point

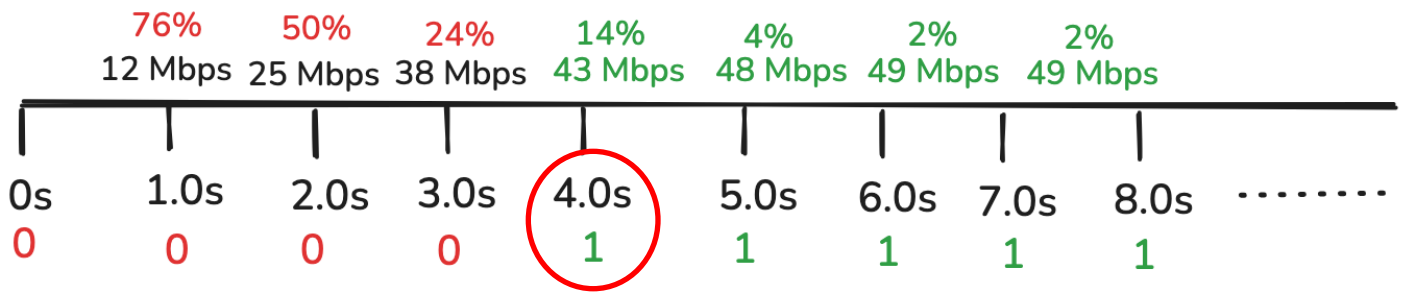
- For each test, find the earliest point (t*) where Stage 1 error falls within tolerance $\boldsymbol{\varepsilon}$ (%)
- ε ∈ {5%, 10%, 15% ..} - operator-tunable tradeoff



$$\tau_{\text{ideal}} = \text{timestamp when } \frac{|y_{\text{pred}} - y_{\text{final}}|}{y_{\text{final}}} < \epsilon$$

Example: Ground Truth: 50 Mbps, ϵ = 15 (%)

Label for transformer (classifier): STOP (1) /CONTINUE (0)



Ideal Stopping Point (t*) = 4 seconds

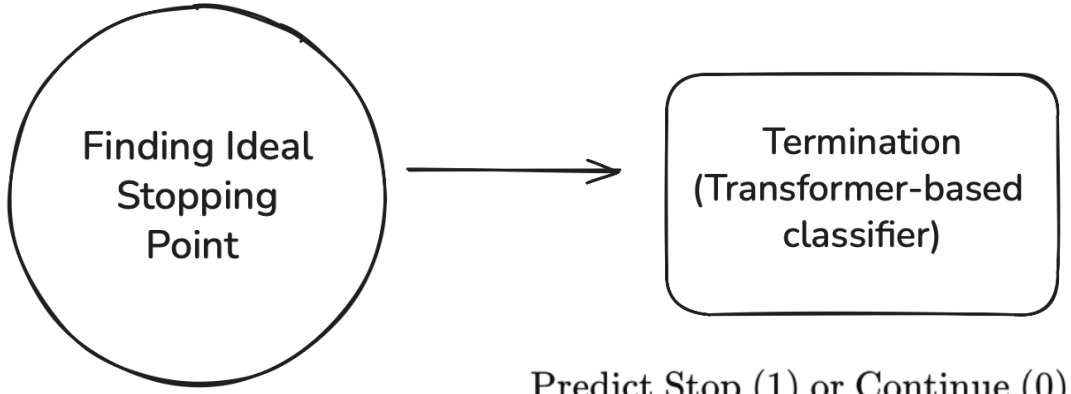
Low prediction error

Low data transfer

Stage 2: Learning When to Stop

Labeling Ideal Stopping point

- For each test, find the earliest point (t*) where Stage 1 error falls within tolerance ε (%)
- ε ∈ {5%, 10%, 15% ..} - operator-tunable tradeoff



$$\tau_{\text{ideal}} = \text{timestamp when } \frac{|y_{\text{pred}} - y_{\text{final}}|}{y_{\text{final}}} < \epsilon$$

Example: Ground Truth: 50 Mbps, ϵ = 15 (%)

Label for transformer (classifier): STOP (1) /CONTINUE (0)

76%	50%	24%	14%	4%	2%	2%
12 Mbps	25 Mbps	38 Mbps	43 Mbps	48 Mbps	49 Mbps	49 Mbps

Low prediction error

Ideal Stopping Point (t*) = 4 seconds

A transformer classifier, learns the earliest point where the Stage 1 error falls within an operator-specified tolerance, ε.

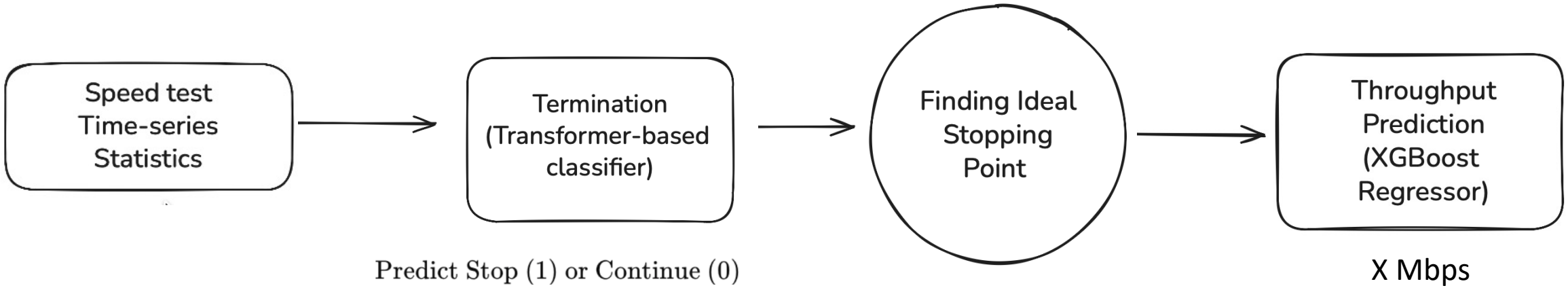
Putting the Two Stages Together – TurboTest

Putting the Two Stages Together – TurboTest

After training on 800k real-world speed tests from M-lab …

Putting the Two Stages Together – TurboTest

After training on 800k real-world speed tests from M-lab …

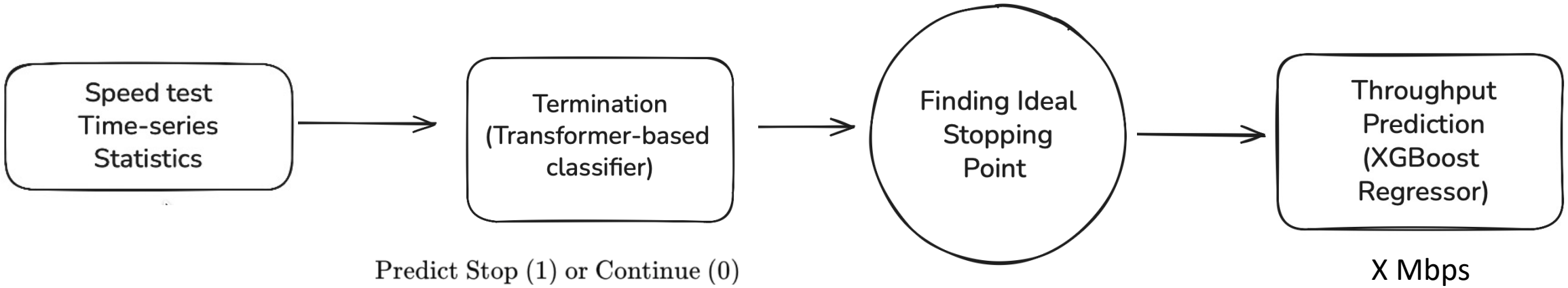


Inference Workflow

During inference/runtime, the classifier stage comes before the prediction stage whereas this was reversed during training time.

Putting the Two Stages Together – TurboTest

After training on 800k real-world speed tests from M-lab …



Inference Workflow

At inference, the classifier runs first. The regressor fires only once the classifier signals sufficient evidence.

Does TurboTest Extend the Accuracy-Savings Frontier?

Does TurboTest Extend the Accuracy-Savings Frontier?

Metrics for Success (calculated on dataset of 1 million speed tests)

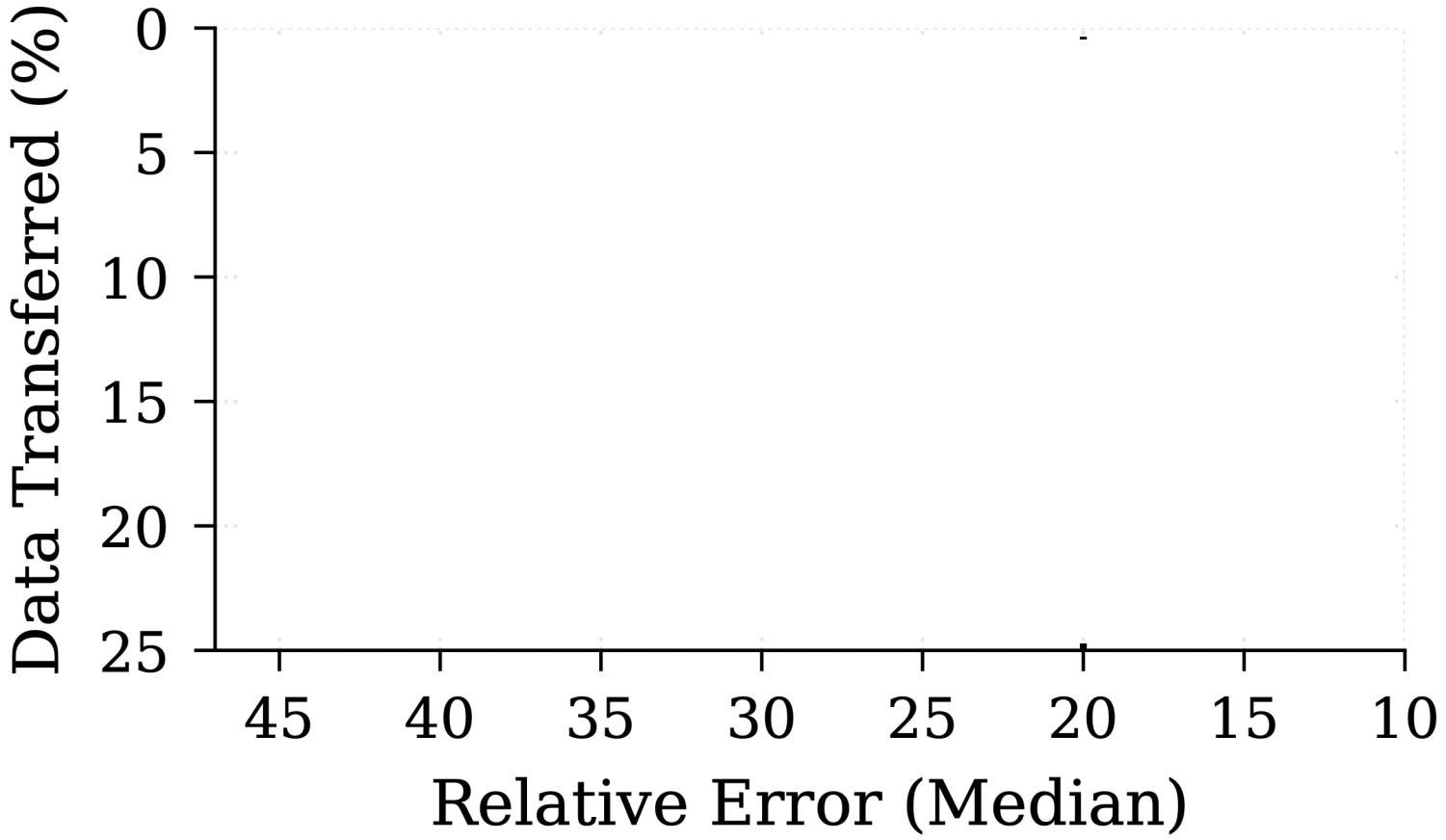
- **Low cumulative Data Transferred**

$$\frac{\sum_{i=1}^{n} B_i^{\text{early}}}{\sum_{i=1}^{n} B_i}$$

- **Low Median Relative Error**

$$\frac{\left|T - T^{\text{early}}\right|}{T} \times 100$$

Does TurboTest Extend the Accuracy-Savings Frontier?



Metrics for Success (calculated on dataset of 1 million speed tests)

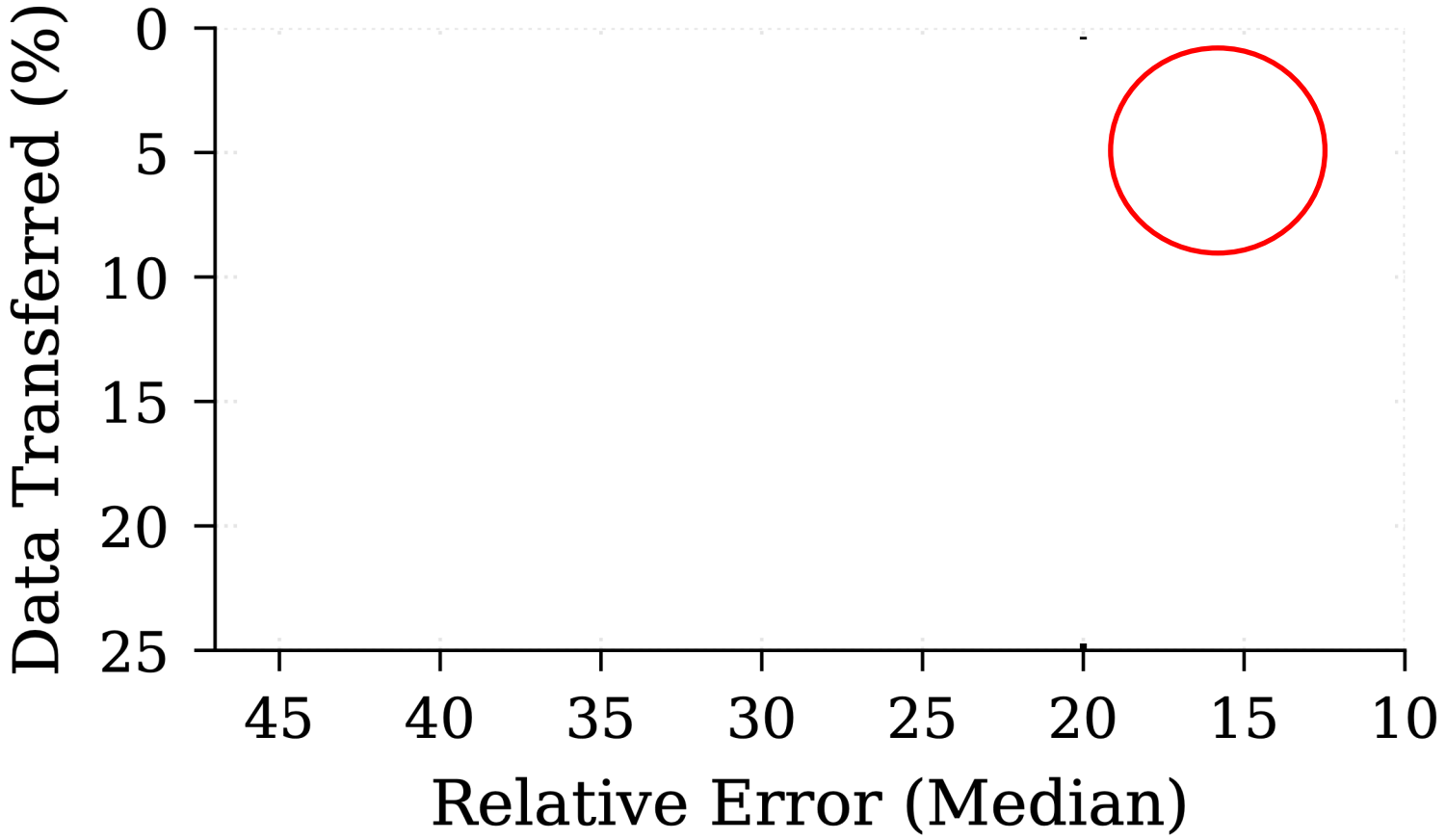
- **Low cumulative Data Transferred**

$$\frac{\sum_{i=1}^{n} B_i^{\text{early}}}{\sum_{i=1}^{n} B_i}$$

- **Low Median Relative Error**

$$\frac{\left|T - T^{\text{early}}\right|}{T} \times 100$$

Does TurboTest Extend the Accuracy-Savings Frontier?



Metrics for Success (calculated on dataset of 1 million speed tests)

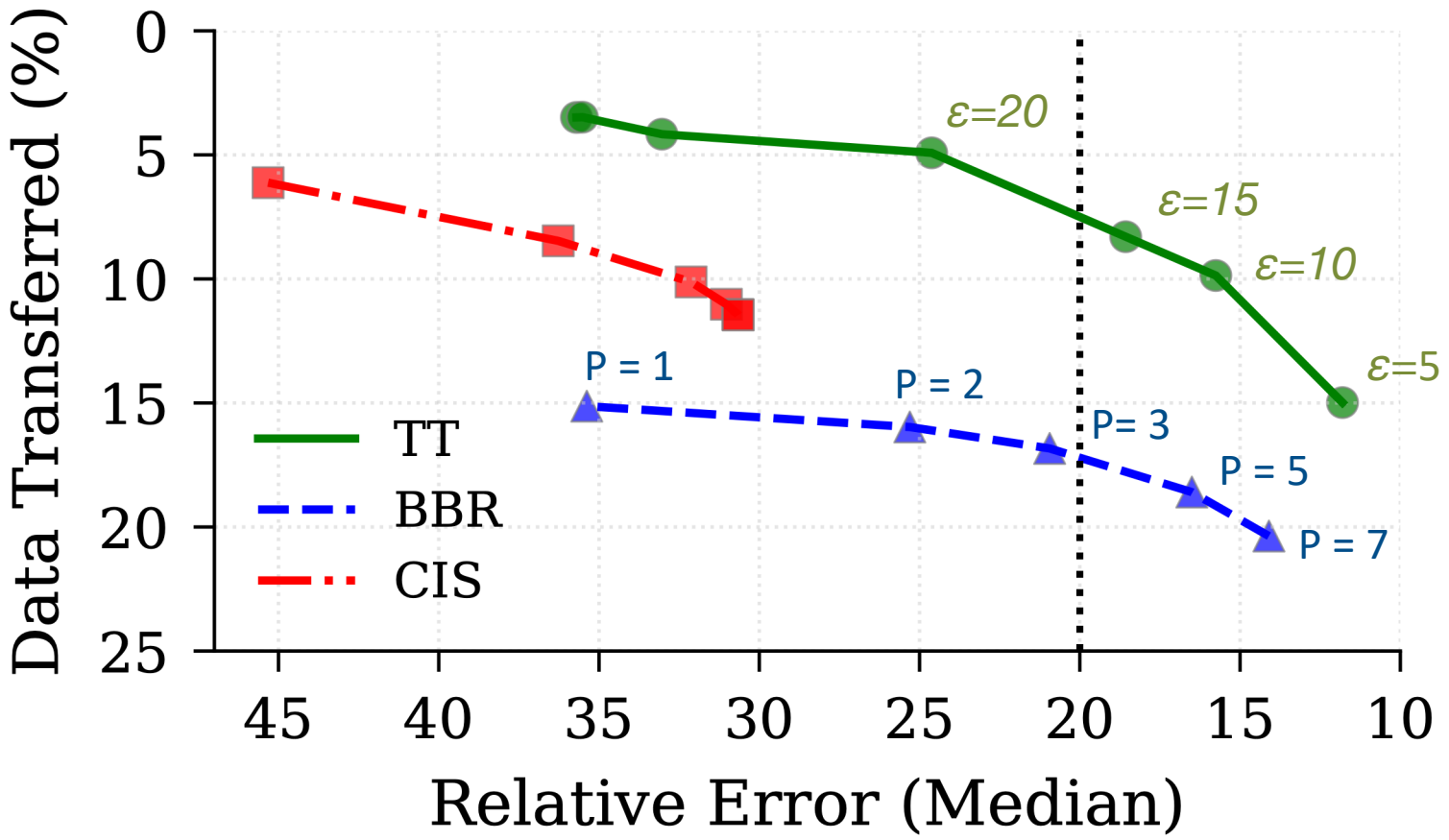
- **Low cumulative Data Transferred**

$$\frac{\sum_{i=1}^{n} B_i^{\text{early}}}{\sum_{i=1}^{n} B_i}$$

- **Low Median Relative Error**

$$\frac{\left|T - T^{\text{early}}\right|}{T} \times 100$$

Does TurboTest Extend the Accuracy-Savings Frontier?



Metrics for Success (calculated on dataset of 1 million speed tests)

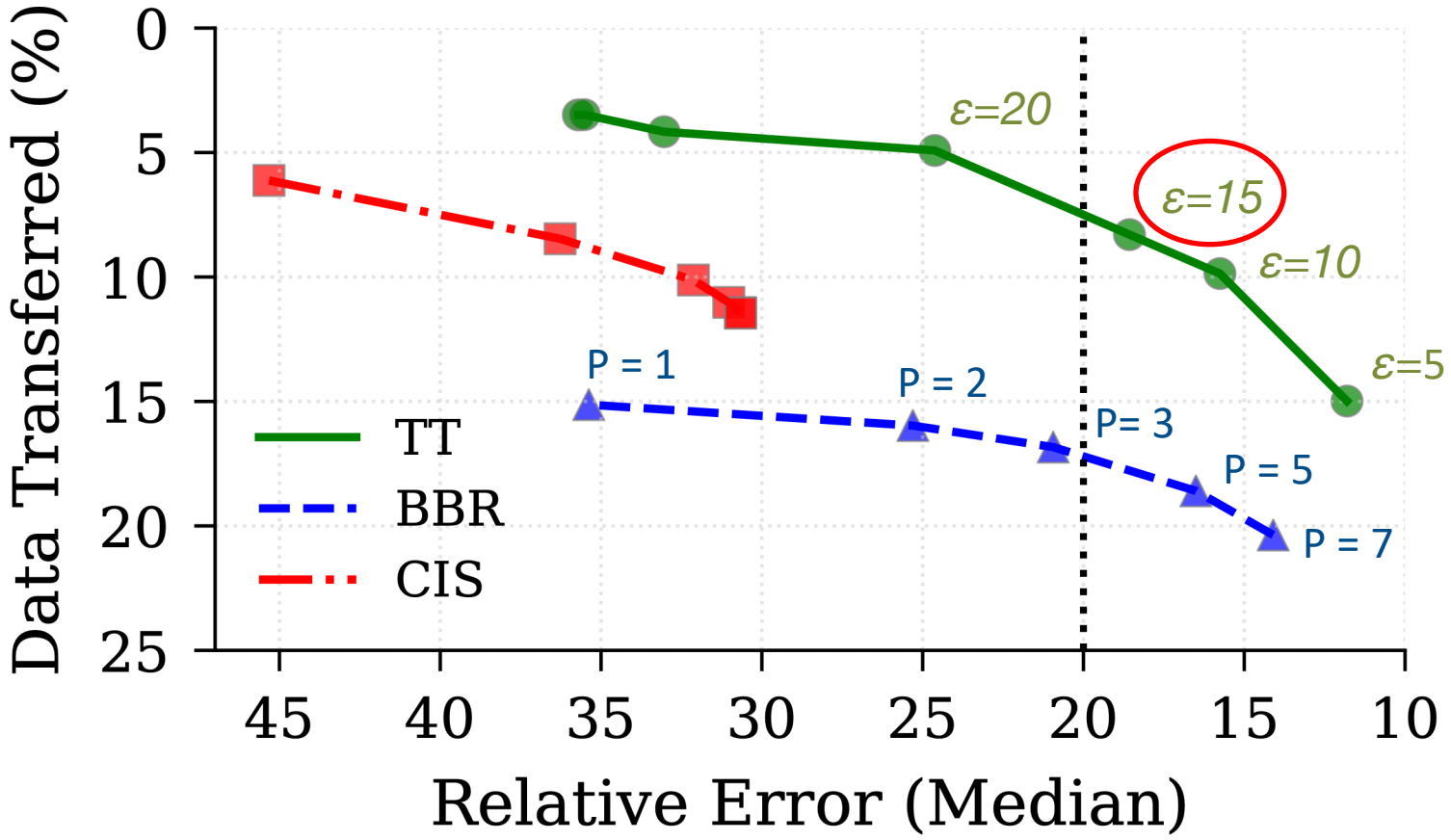
- **Low cumulative Data Transferred**

$$\frac{\sum_{i=1}^{n} B_i^{\text{early}}}{\sum_{i=1}^{n} B_i}$$

- **Low Median Relative Error**

$$\frac{|T - T^{\text{early}}|}{T} \times 100$$

Does TurboTest Extend the Accuracy-Savings Frontier?



Metrics for Success (calculated on dataset of 1 million speed tests)

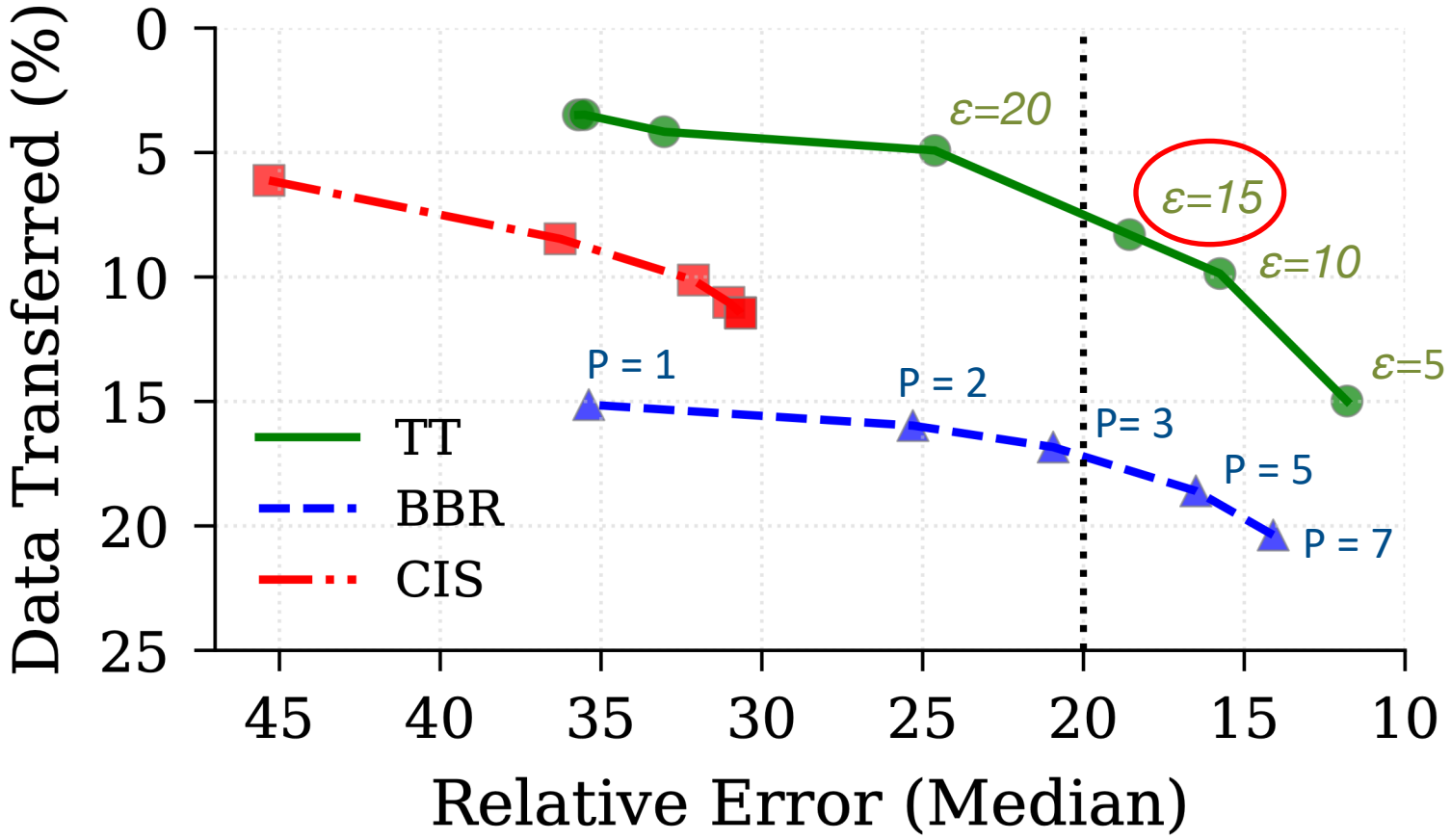
- **Low cumulative Data Transferred**

$$\frac{\sum_{i=1}^{n} B_i^{\text{early}}}{\sum_{i=1}^{n} B_i}$$

- **Low Median Relative Error**

$$\frac{|T - T^{\text{early}}|}{T} \times 100$$

Does TurboTest Extend the Accuracy-Savings Frontier?



Metrics for Success (calculated on dataset of 1 million speed tests)

- **Low cumulative Data Transferred**

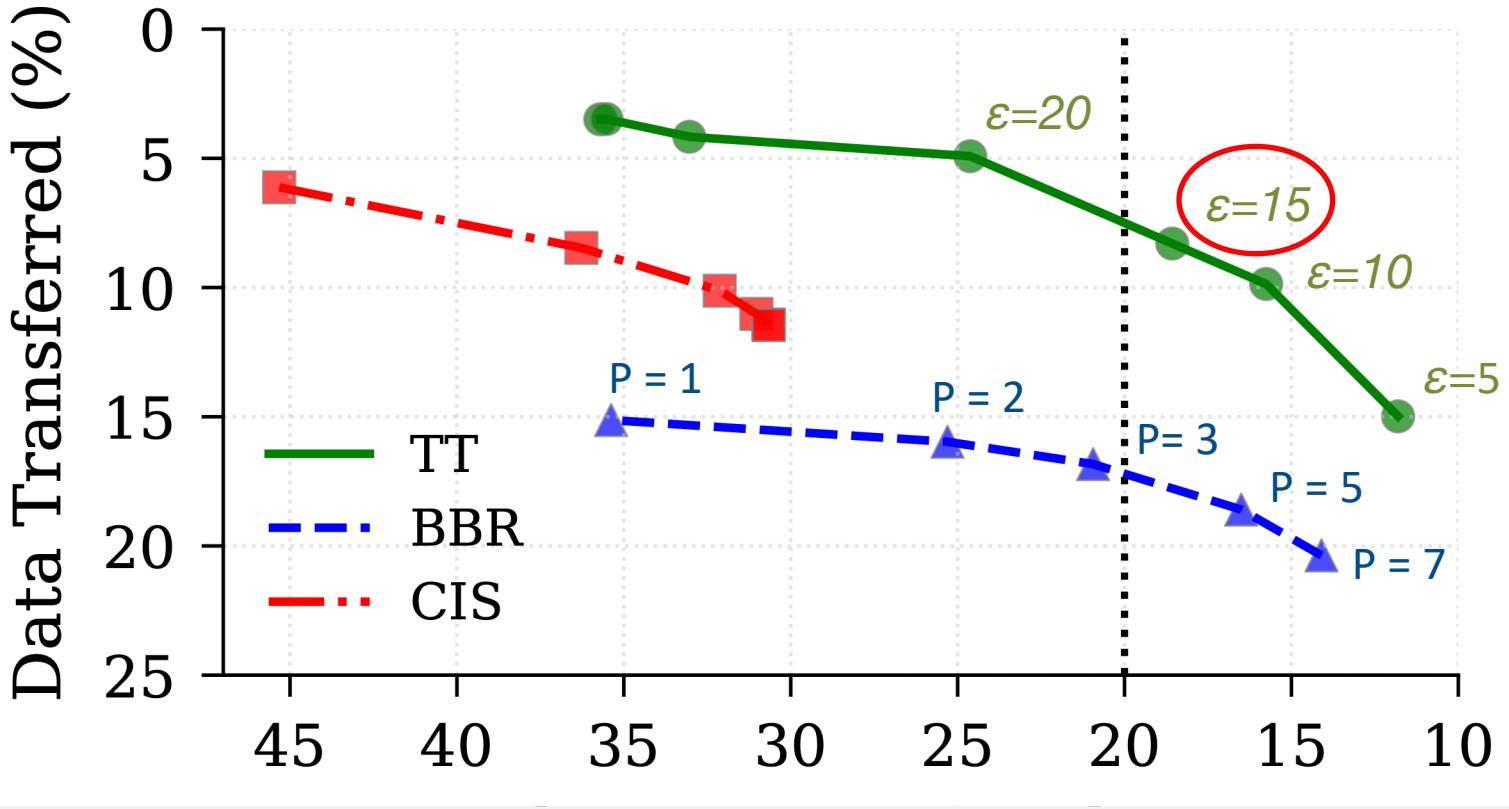
$$\frac{\sum_{i=1}^{n} B_i^{\text{early}}}{\sum_{i=1}^{n} B_i}$$

- **Low Median Relative Error**

$$\frac{|T - T^{\text{early}}|}{T} \times 100$$

At ε=15, TT achieves 8% data transfer -> Reducing M-lab's monthly overhead from 12 PB to under 1 PB

Does TurboTest Extend the Accuracy-Savings Frontier?



Metrics for Success (calculated on dataset of 1 million speed tests)

- **Low cumulative Data Transferred**

$$\frac{\sum_{i=1}^{n} B_i^{\text{early}}}{\sum_{i=1}^{n} B_i}$$

- **Low Median Relative Error**

$$\left|T - T^{\text{early}}\right|$$

How does TurboTest perform in the key high throughput and Gigabit speed tests ?

Which Tests Benefit Most from TurboTest?

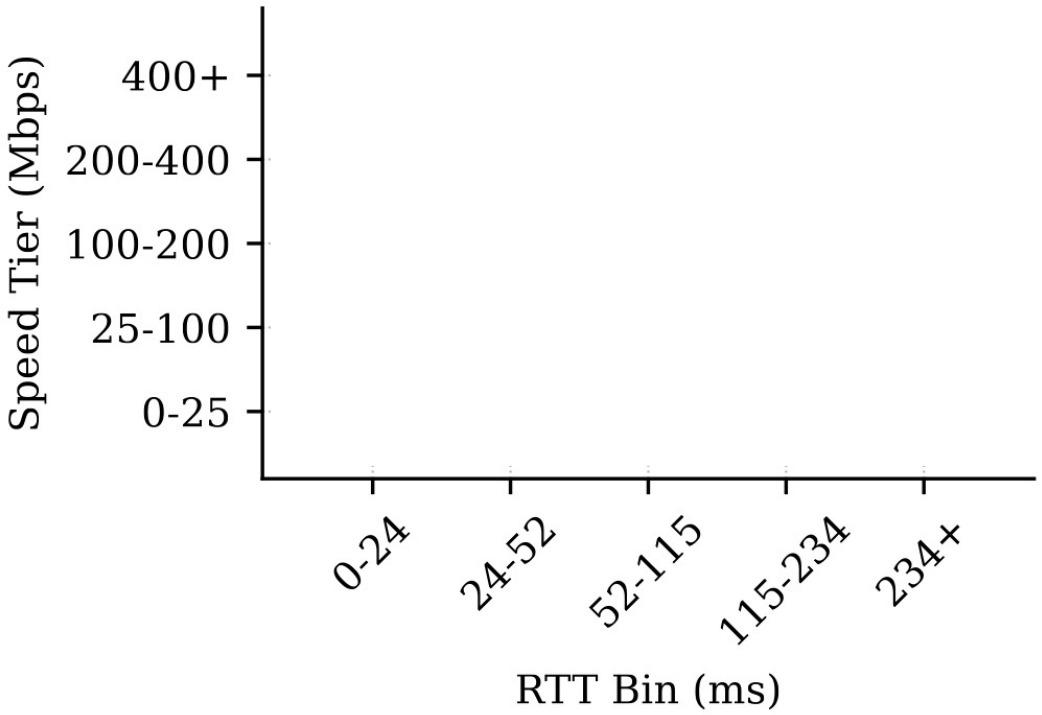
Which Tests Benefit Most from TurboTest?

- **Since, high throughput tests contribute the most to raw data volume ….**

Which Tests Benefit Most from TurboTest?

- Since, high throughput tests contribute the most to raw data volume ….
- Let’s breakdown our results based on throughput and RTT

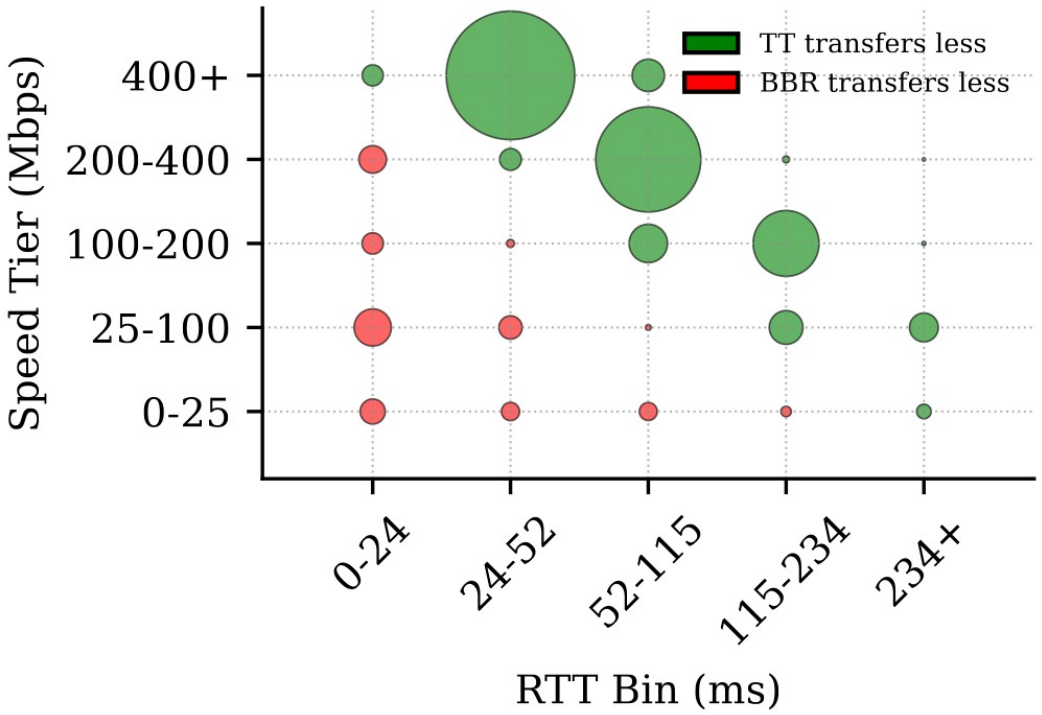
Which Tests Benefit Most from TurboTest?



- **Since, high throughput tests contribute the most to raw data volume ….**
- **Let's breakdown our results based on throughput and RTT**

Point size indicates the magnitude of the difference in data transferred

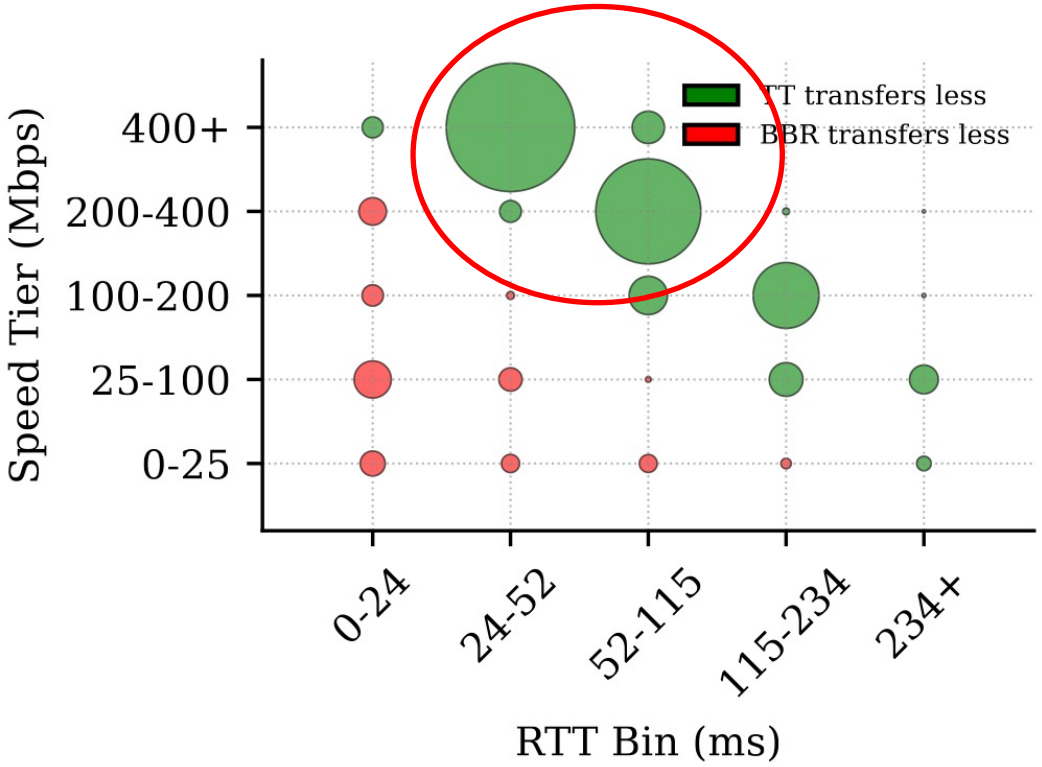
Which Tests Benefit Most from TurboTest?



- **Since, high throughput tests contribute the most to raw data volume ….**
- **Let's breakdown our results based on throughput and RTT**

Point size indicates the magnitude of the difference in data transferred.

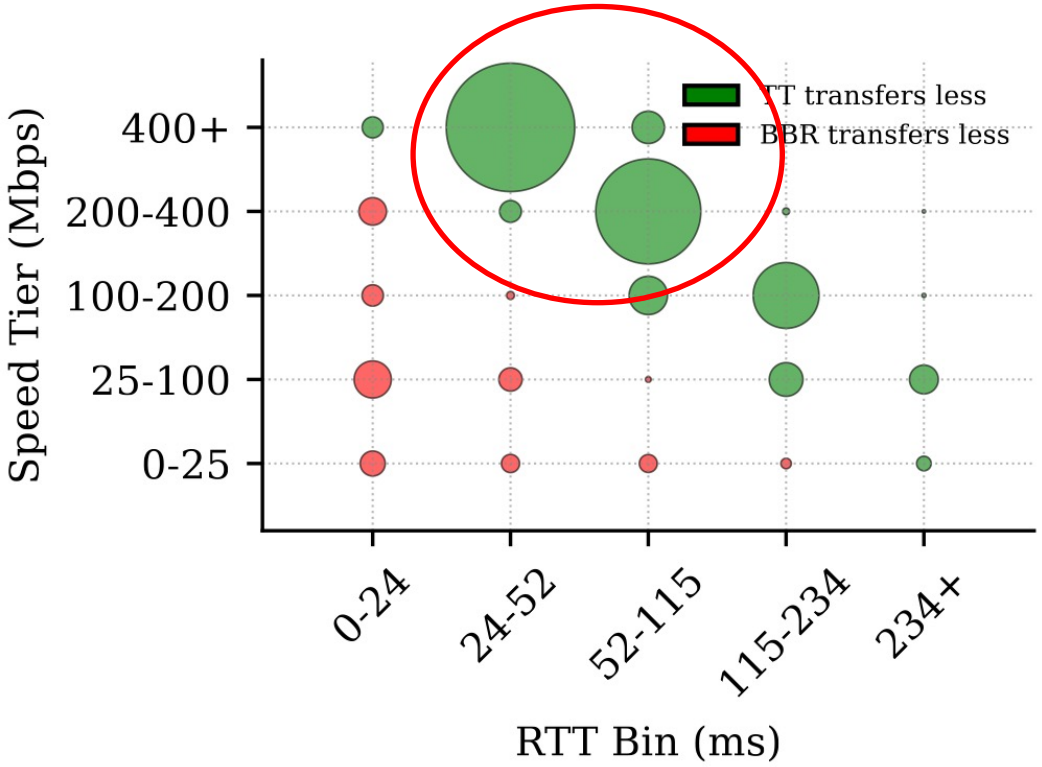
Which Tests Benefit Most from TurboTest?



- **Since, high throughput tests contribute the most to raw data volume ….**
- **Let's breakdown our results based on throughput and RTT**

Point size indicates the magnitude of the difference in data transferred.

Which Tests Benefit Most from TurboTest?

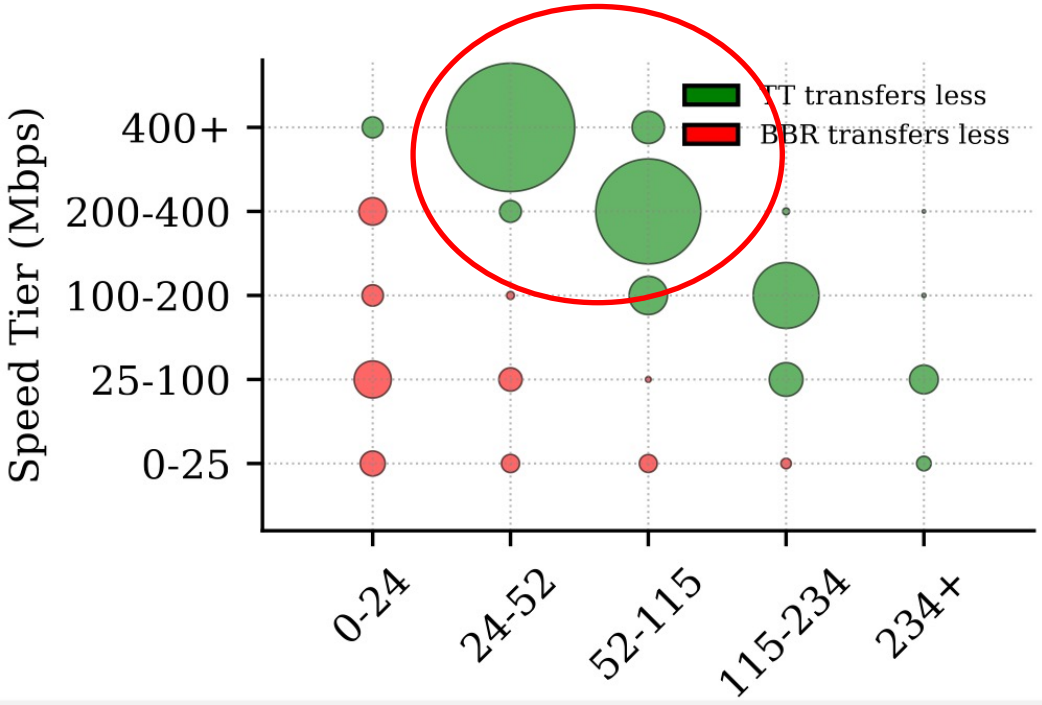


- **Since, high throughput tests contribute the most to raw data volume ….**
- **Let's breakdown our results based on throughput and RTT**

- ***TurboTest*** **outperforms BBR in high-speed tiers, which drives the majority of data savings**

Point size indicates the magnitude of the difference in data transferred.

Which Tests Benefit Most from TurboTest?



- Since, high throughput tests contribute the most to raw data volume ….
- Let's breakdown our results based on throughput and RTT

- *TurboTest* outperforms BBR in high-speed tiers, which drives the majority of data savings

TurboTest's largest gains are in high-throughput tests — the same conditions where BBR struggles most.

Takeaways

Takeaways

- **TurboTest** balances the trade-off between accuracy and data transferred, outperforms BBR (**1.8-4.4x**)

Takeaways

- **TurboTest** balances the trade-off between accuracy and data transferred, outperforms BBR (**1.8-4.4x**)
- It leverages temporal patterns to outperform prior methods on problematic high throughput tests (sample dataset of 1 million real-world speed tests)

Takeaways

- **TurboTest** balances the trade-off between accuracy and data transferred, outperforms BBR (**1.8-4.4x**)
- It leverages temporal patterns to outperform prior methods on problematic high throughput tests (sample dataset of 1 million real-world speed tests)
- By exposing a tunable parameter **ε, TurboTest** delivers efficient and robust termination policies, suitable for deployment at scale

Takeaways

- **TurboTest** balances the trade-off between accuracy and data transferred, outperforms BBR (**1.8-4.4x**)
- It leverages temporal patterns to outperform prior methods on problematic high throughput tests (sample dataset of 1 million real-world speed tests)
- By exposing a tunable parameter **ε, TurboTest** delivers efficient and robust termination policies, suitable for deployment at scale

Reach out to me:

Haarika Manda, hmanda@ucsb.edu

