

## on Steroids with Learned Filters

Qitong Wang  
Harvard University

Ioana Ileana  
Université Paris Cité

Themis Palpanas  
Université de Paris & IUF

1

### Data Series Similarity Search

- Data series present in virtually every domain



- Similarity search (or pattern matching) identifies sequences similar to some query sequence

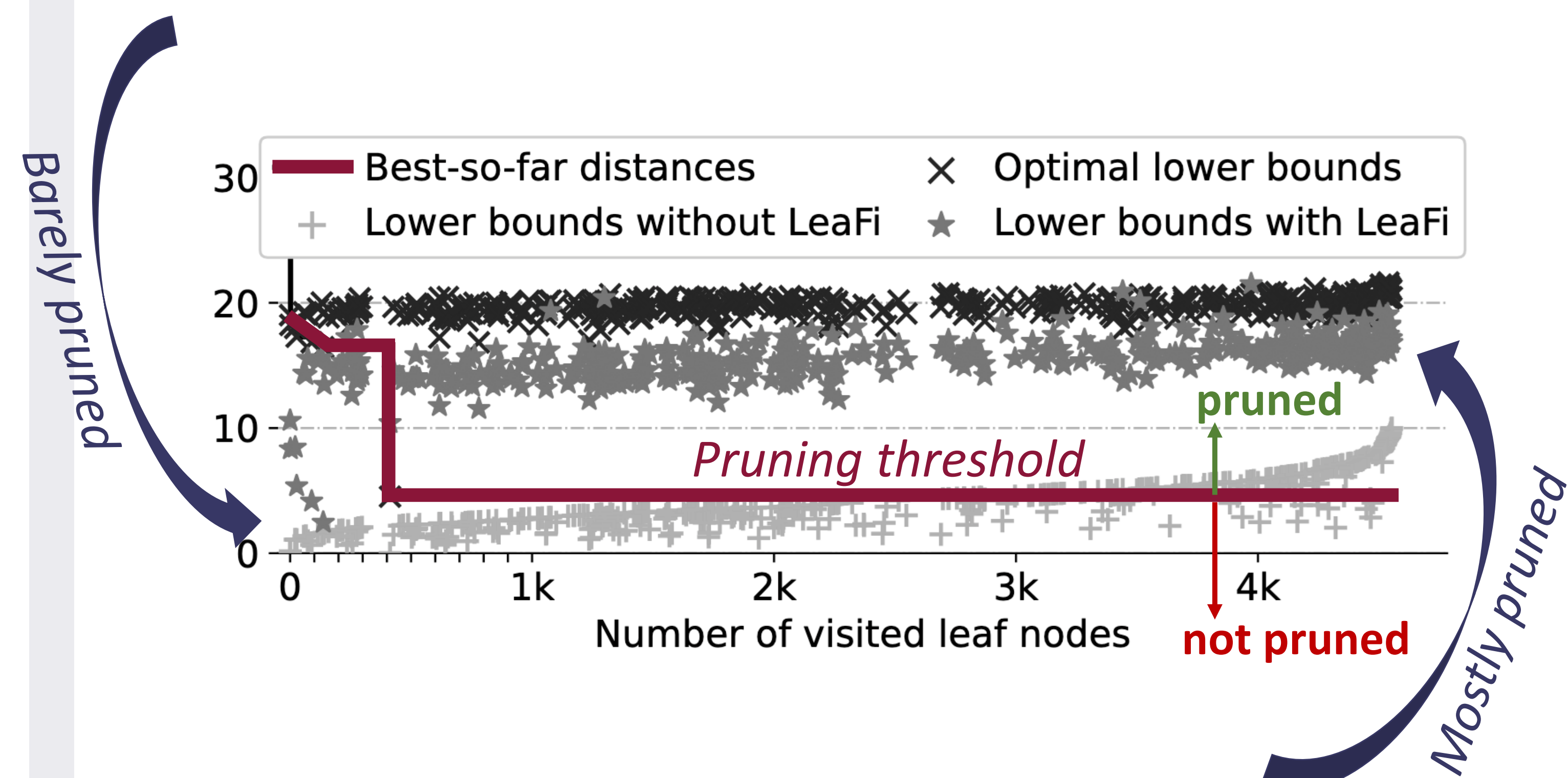


- Similarity search is a key operation for scalable data series analysis

3

### Accelerate Search using Learned Filters

- Lower bounds are used to prune nodes
- Loose lower bounds **without Learned Filters**

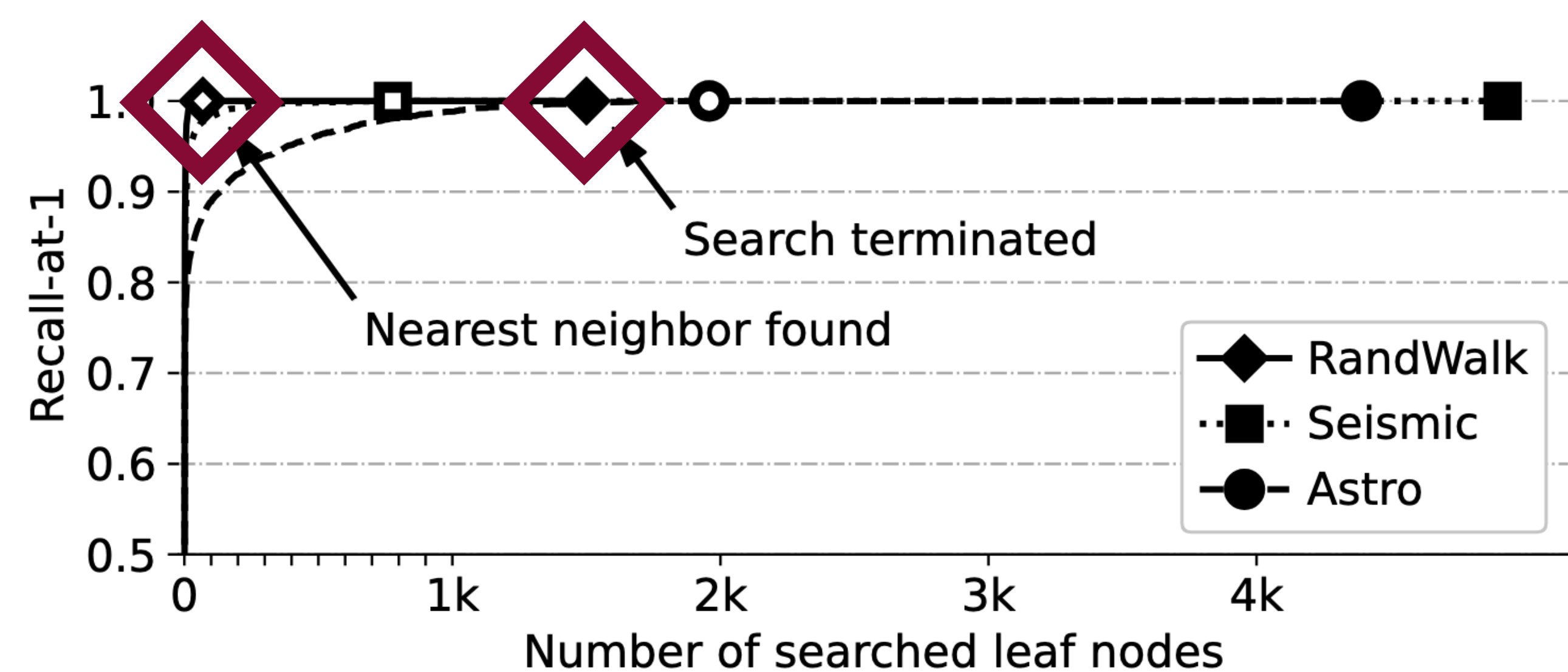


- Tight lower bounds **with Learned Filters**

2

### Search Is Slow! 😞

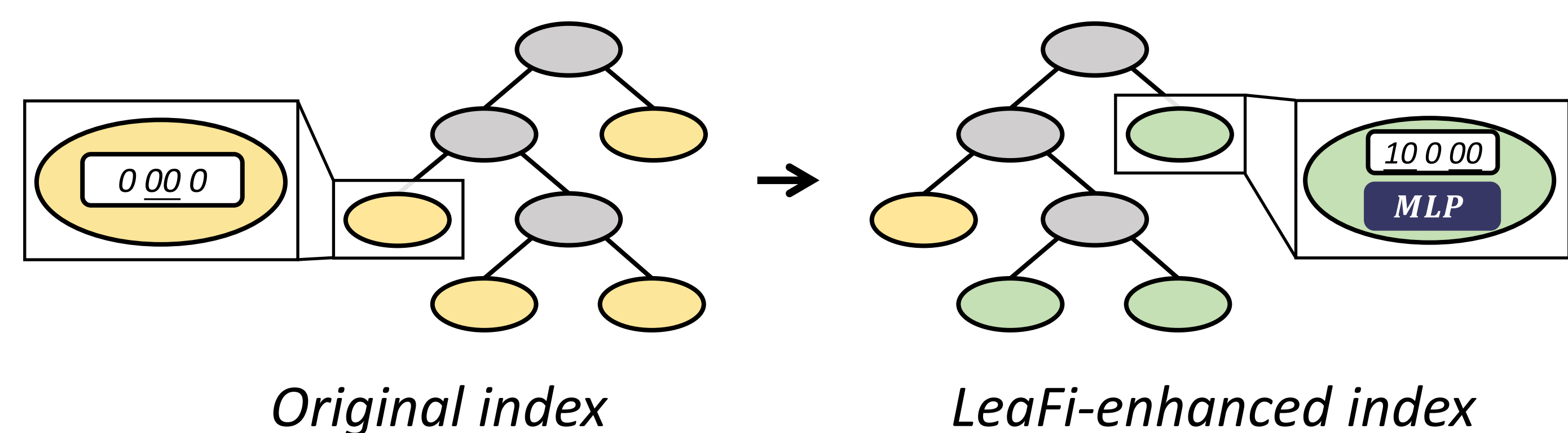
- Tree-based indexes are among SOTA solutions



- Search cannot be terminated, after results are found

4

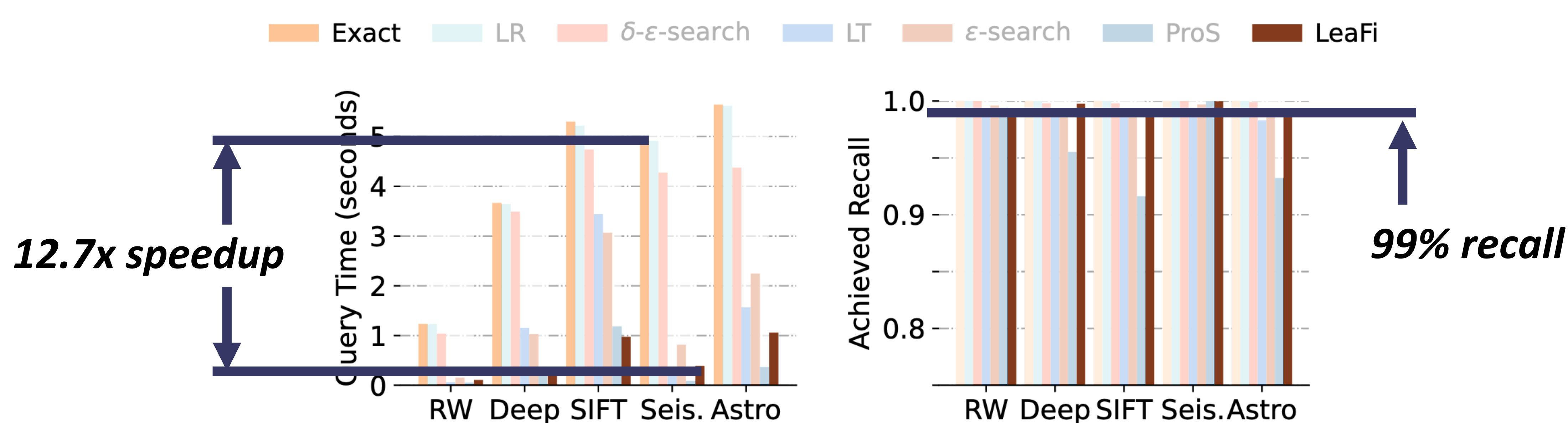
### How to Insert Learned Filters?



- Which nodes to insert learned filters?
  - ✓ Model it as a knapsack problem, then simplify
    - an item  $\rightarrow$  a filter, value  $\rightarrow$  search speedup, weight  $\rightarrow$  GPU memory
- How to control result quality?
  - ✓ Posterior statistical adjustment
    - inspired by conformal predictors

5

### Experiments: almost perfect recall, at a fraction of exact search time



Scan for more LeaFi!