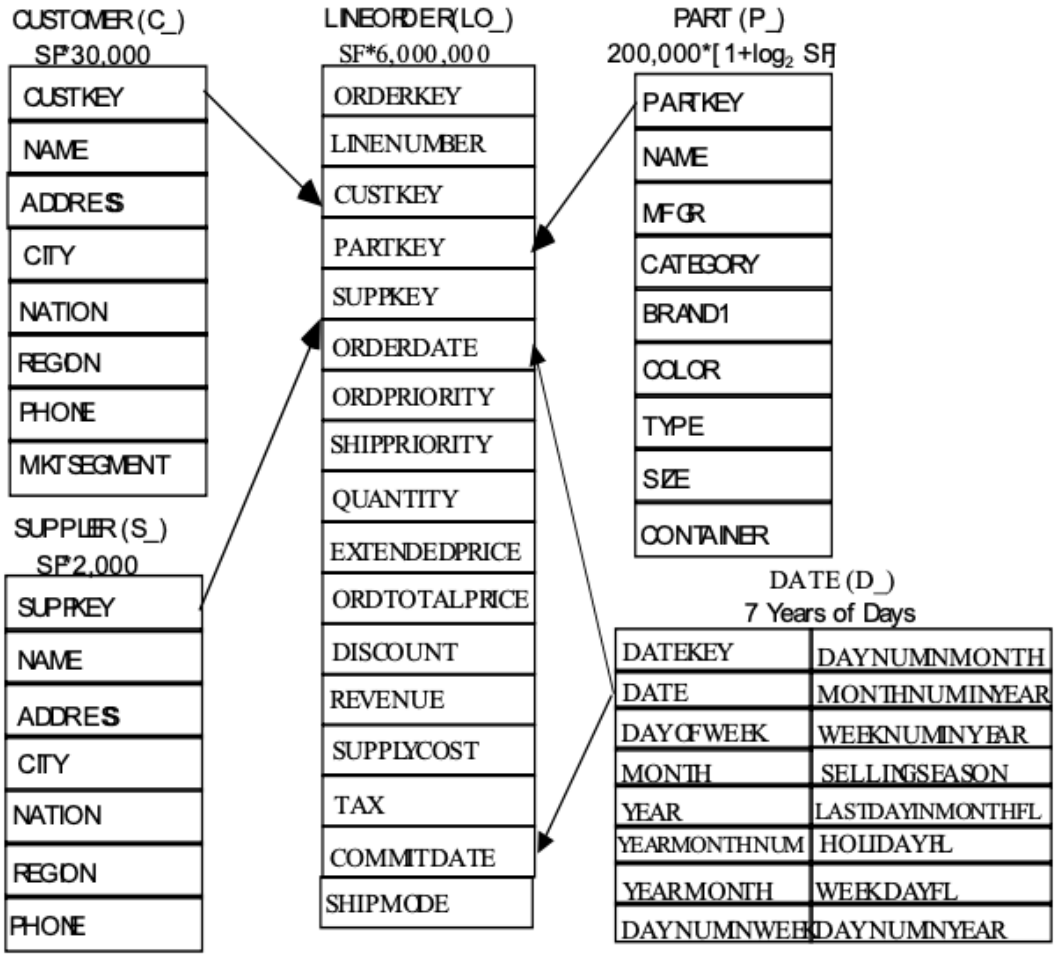
Accelerating Joins with Filters: Keeping a Limited Memory is Robust

Nicholas Corrado Xiating Ouyang

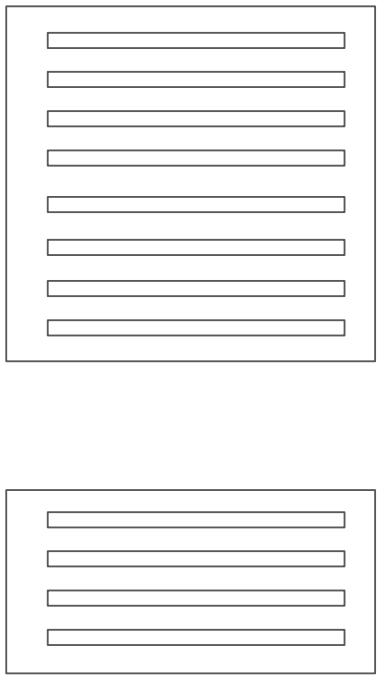
University of Wisconsin-Madison

Star Schema

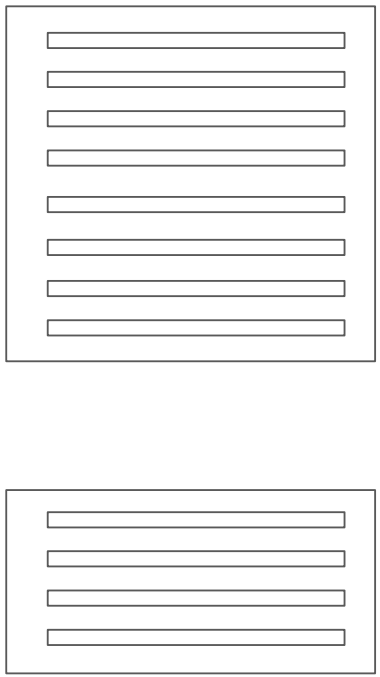


- Gigantic intermediate tables
- Filtering ahead of time before join

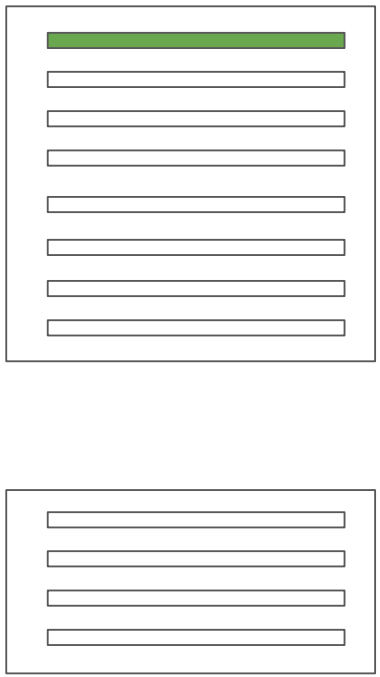
Lookahead Information Passing (LIP)



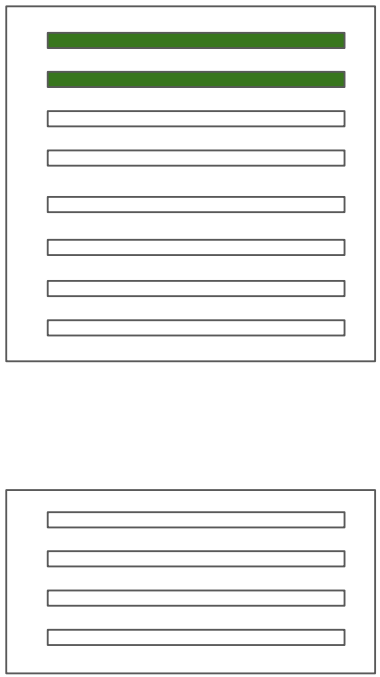
Lookahead Information Passing (LIP)



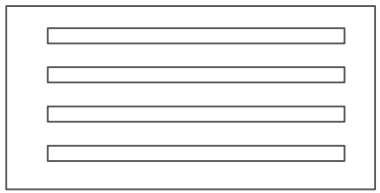
Lookahead Information Passing (LIP)



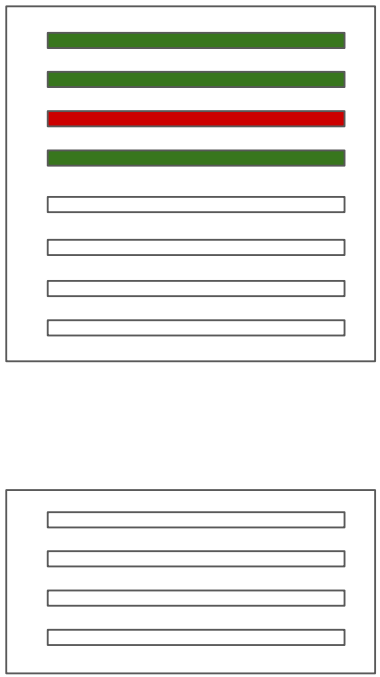
Lookahead Information Passing (LIP)



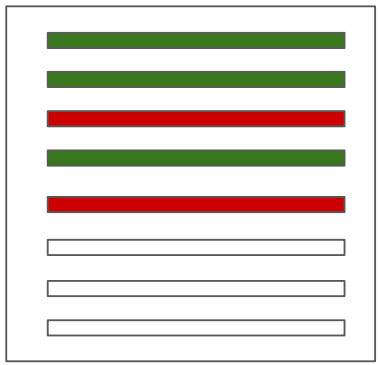
Lookahead Information Passing (LIP)

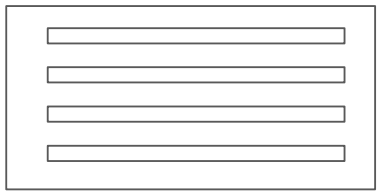


Lookahead Information Passing (LIP)

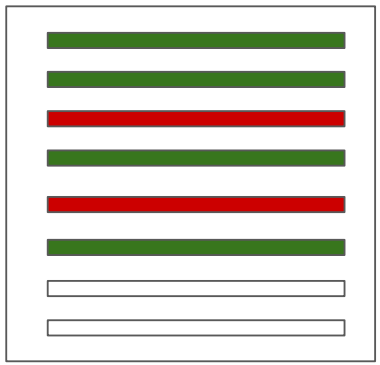


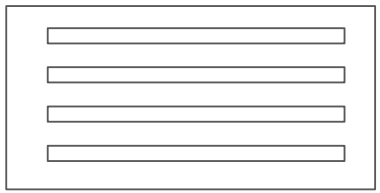
Lookahead Information Passing (LIP)



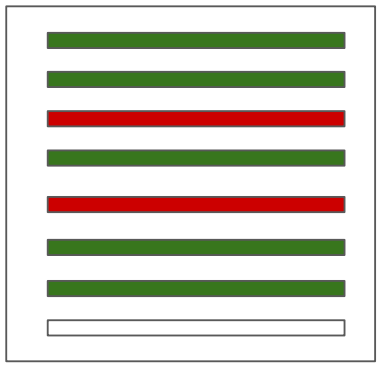


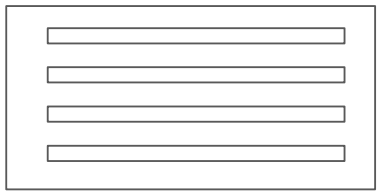
Lookahead Information Passing (LIP)



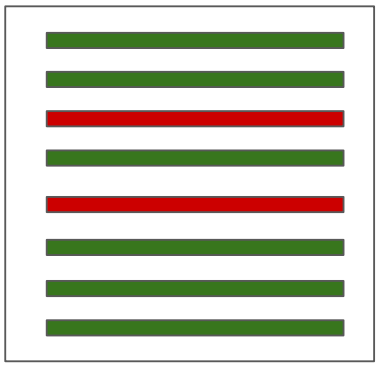


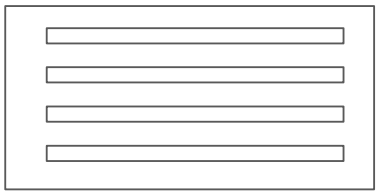
Lookahead Information Passing (LIP)



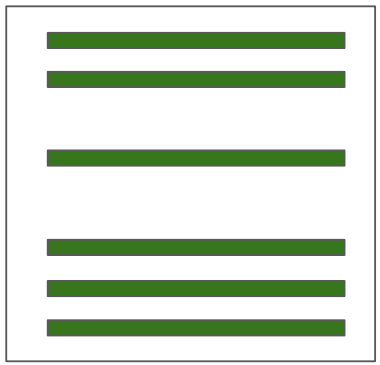


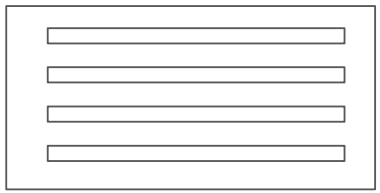
Lookahead Information Passing (LIP)



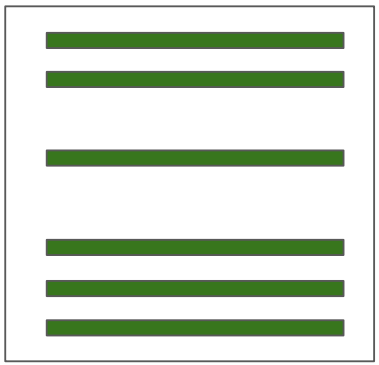


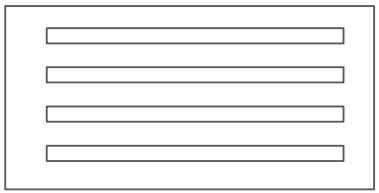
Lookahead Information Passing (LIP)



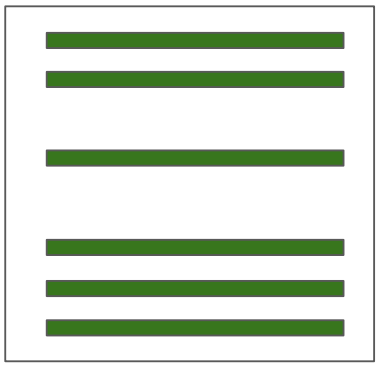


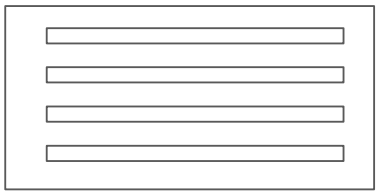
Lookahead Information Passing (LIP)





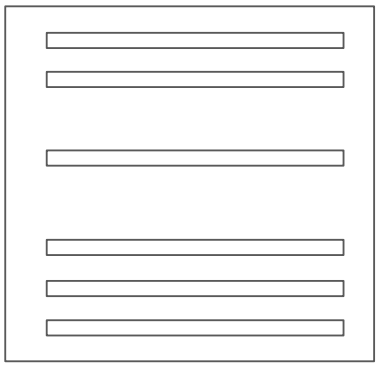
Lookahead Information Passing (LIP)

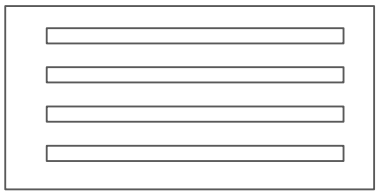




6/8

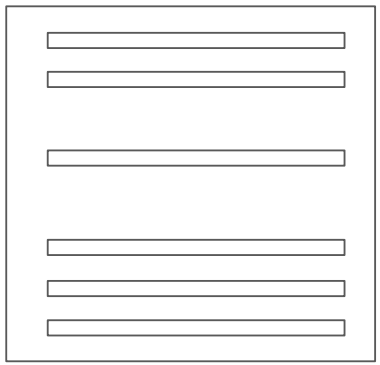
Lookahead Information Passing (LIP)

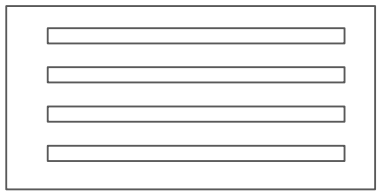




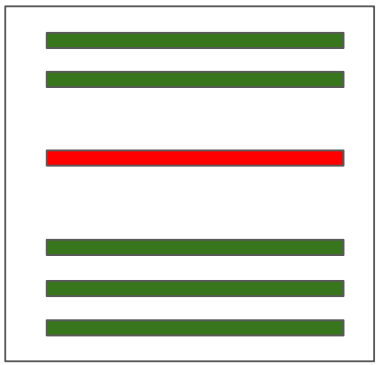
6/8

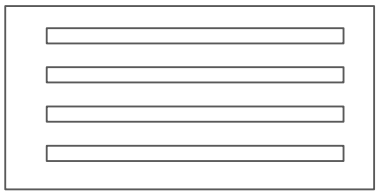
Lookahead Information Passing (LIP)

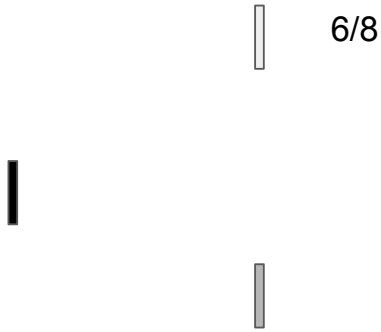




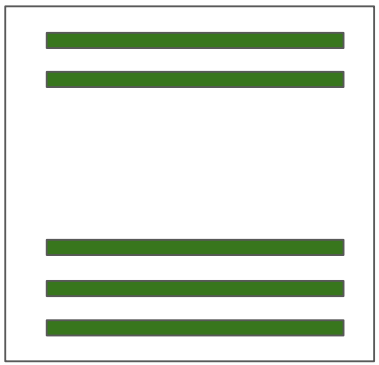
Lookahead Information Passing (LIP)

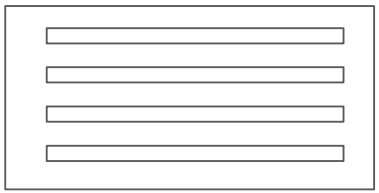


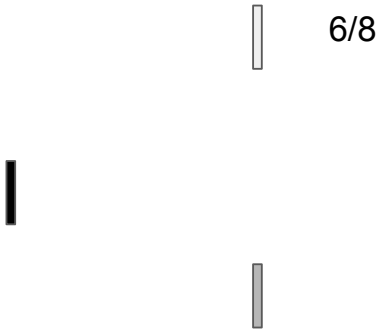




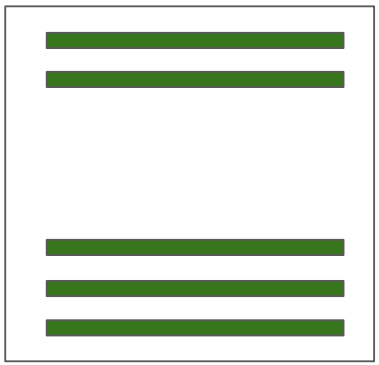
Lookahead Information Passing (LIP)

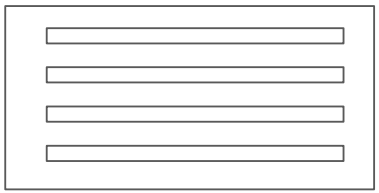






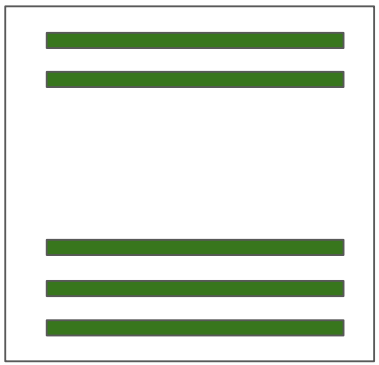
Lookahead Information Passing (LIP)

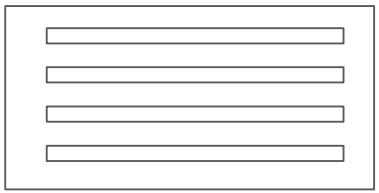




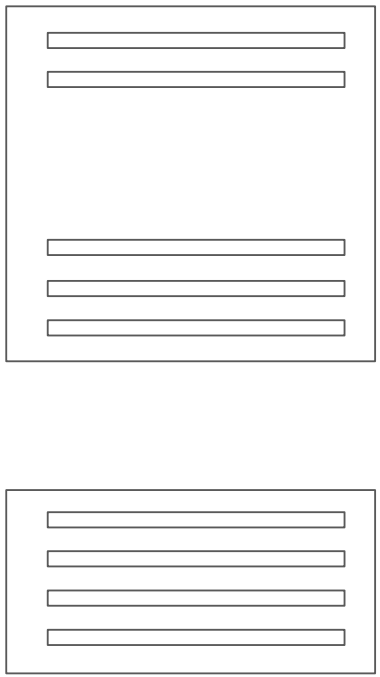
6/8

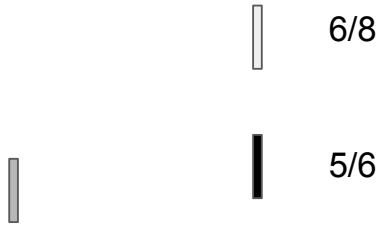
Lookahead Information Passing (LIP)



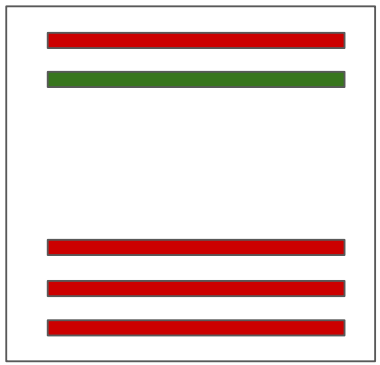


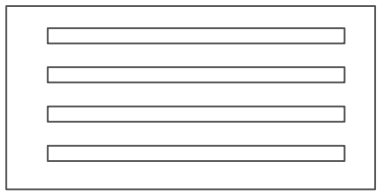
Lookahead Information Passing (LIP)

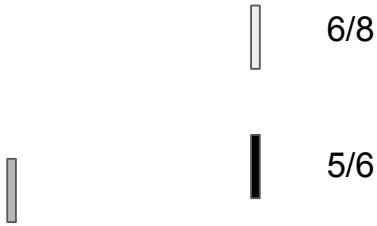




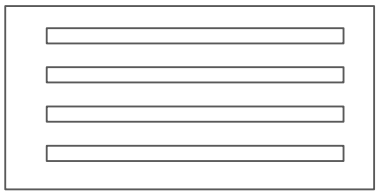
Lookahead Information Passing (LIP)

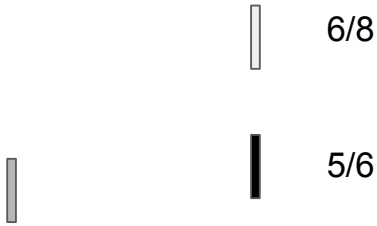




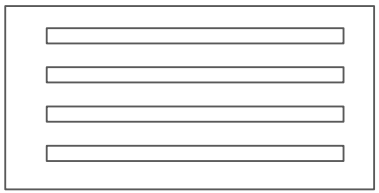


Lookahead Information Passing (LIP)

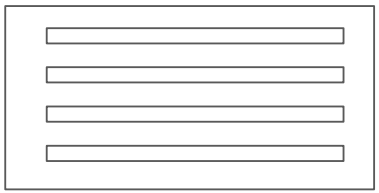




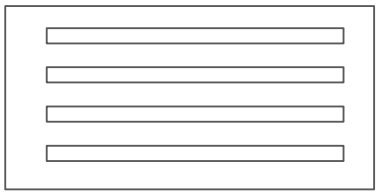
Lookahead Information Passing (LIP)



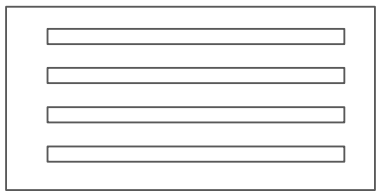
Lookahead Information Passing (LIP)

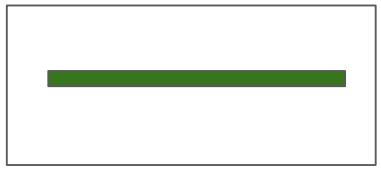


Lookahead Information Passing (LIP)

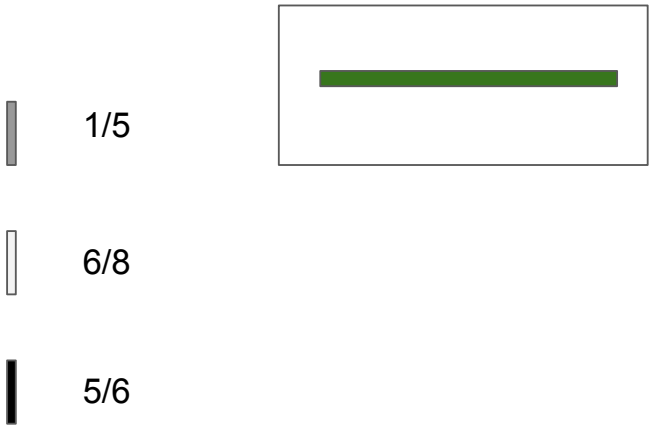


Lookahead Information Passing (LIP)



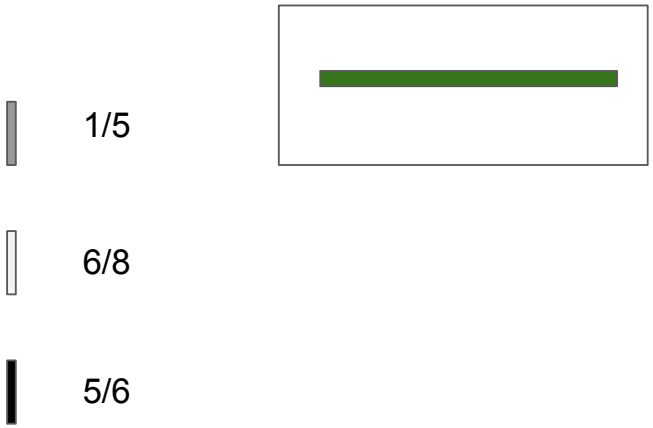


Lookahead Information Passing (LIP)



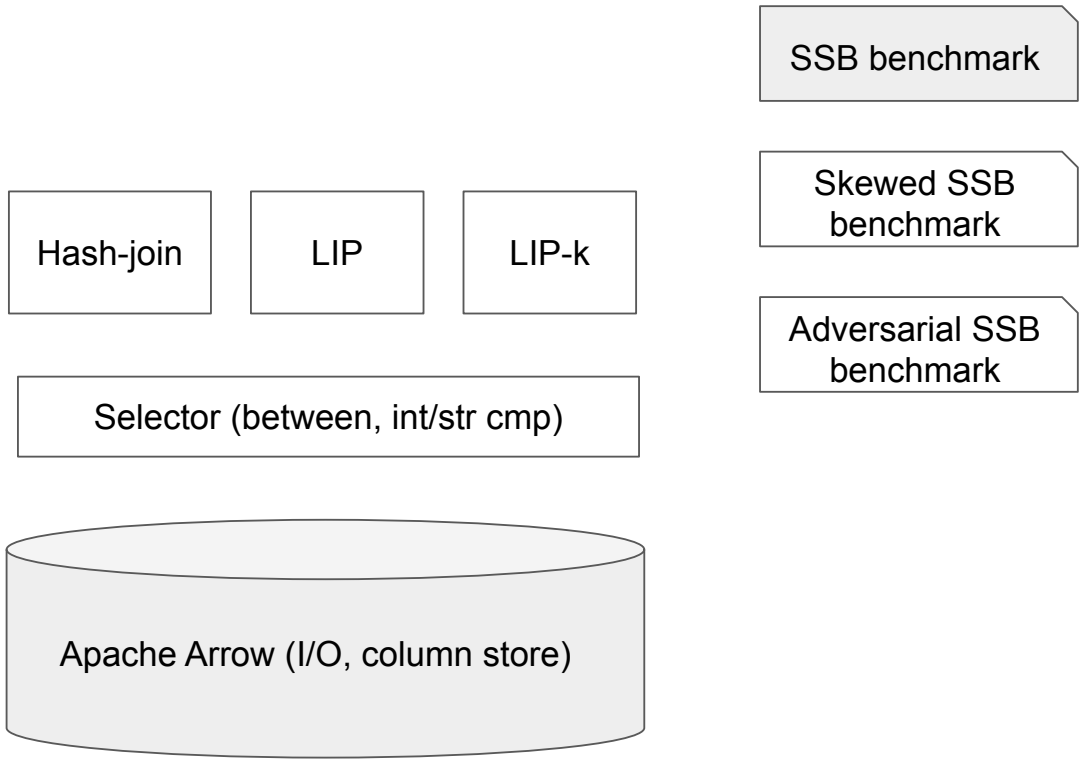
- Using statistics from all previous batches
- Just the previous k

Lookahead Information Passing (LIP)



- Using statistics from all previous batches
- Just the previous k — LIP-k

Implementation and benchmarking



Experiments

LIP is solving an online problem

- Tuples arriving one at a time
- Upon arrival, decide a sequence of filters
- Minimize the total probes
- Deterministic!

LIP is solving an online problem

- Tuples arriving one at a time
- Upon arrival, decide a sequence of filters
- Minimize the total probes
- Deterministic!

Theorem

Let n be the number of filters in the LIP problem. There is no deterministic mechanism $\mathcal{M}$ achieving a competitive ratio less than n for the LIP *problem.*

LIP is solving an online problem

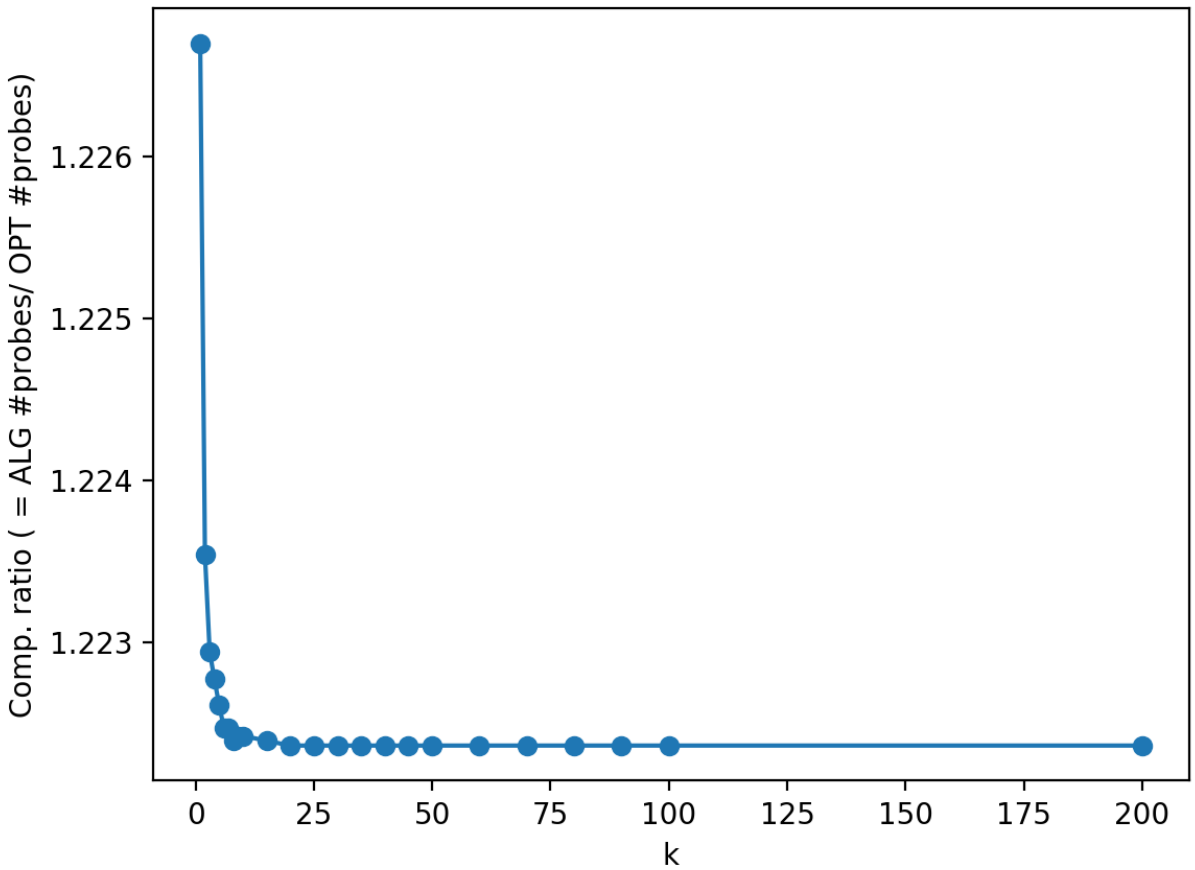
- Tuples arriving one at a time
- Upon arrival, decide a sequence of filters
- Minimize the total probes
- Deterministic!

Theorem

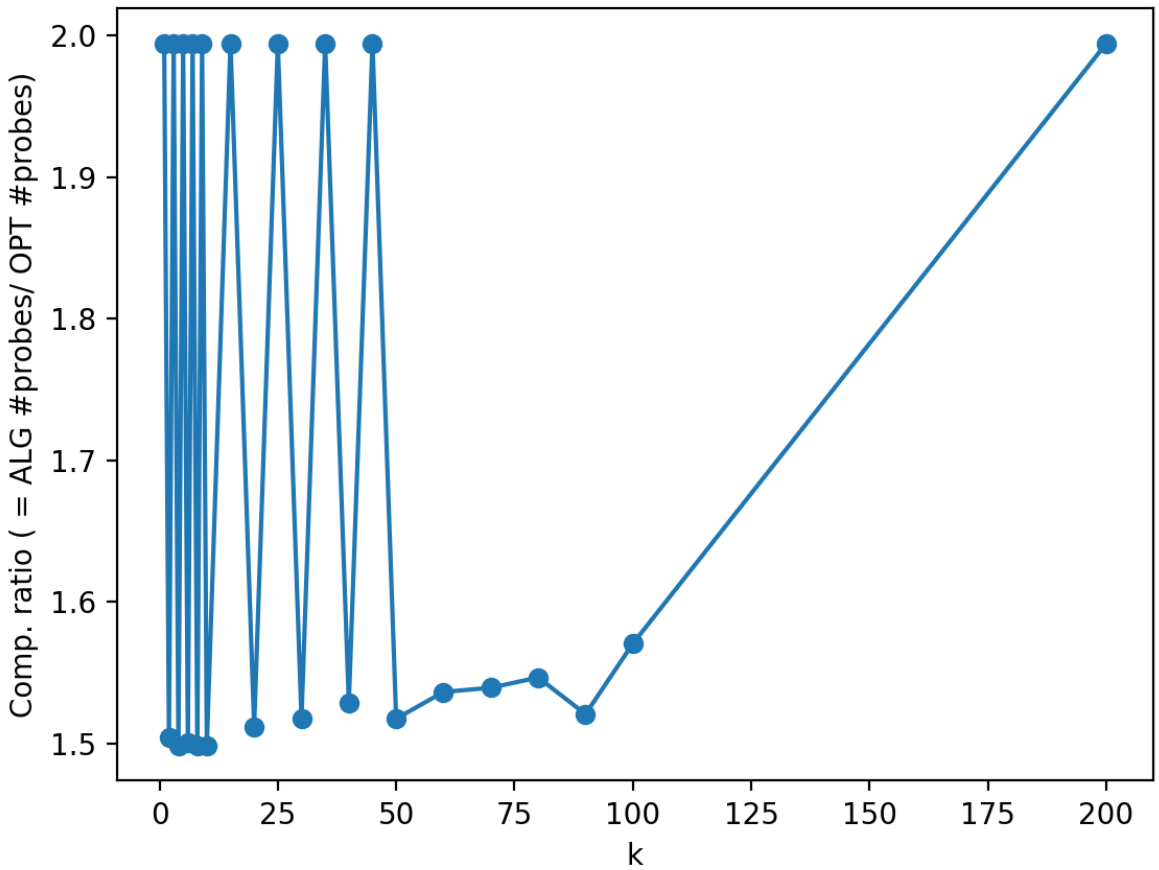
Let n be the number of filters in the LIP problem. There is no deterministic mechanism $\mathcal{M}$ achieving a competitive ratio less than n for the LIP *problem.*

- Not observed in practice, but a theoretical lower bound
- Randomness?

Competitive ratio vs. k on Uniform



Competitive ratio vs. k on skewed



Conclusion

- Implemented LIP and its variant LIP-k
- LIP-k is better than LIP in the adversarial/skewed settings
- Randomness to achieve better robustness guarantee

Thank you!

Questions?