nsdi'22

Elixir: A High-performance and Low-cost Approach to Managing Hardware/Software Hybrid Flow Tables Considering Flow Burstiness

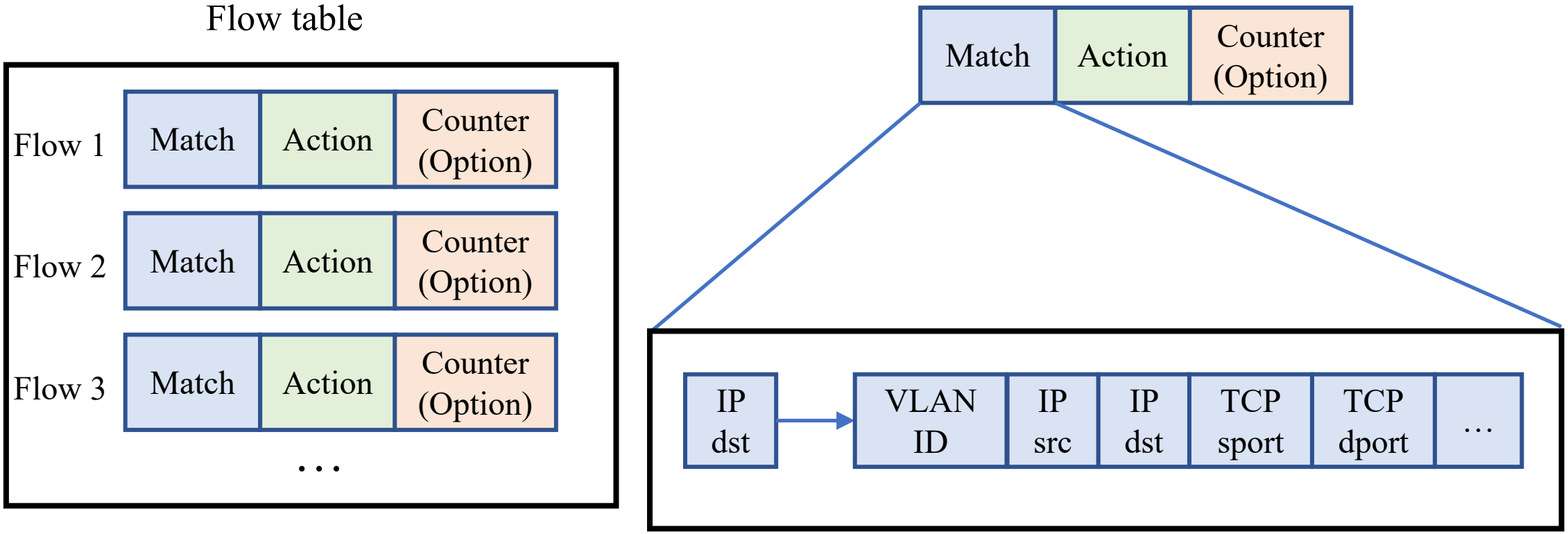
Yanshu Wang

Dan Li, Yuanwei Lu, Jianping Wu, Hua Shao, Yutian Wang

Tencent 腾讯

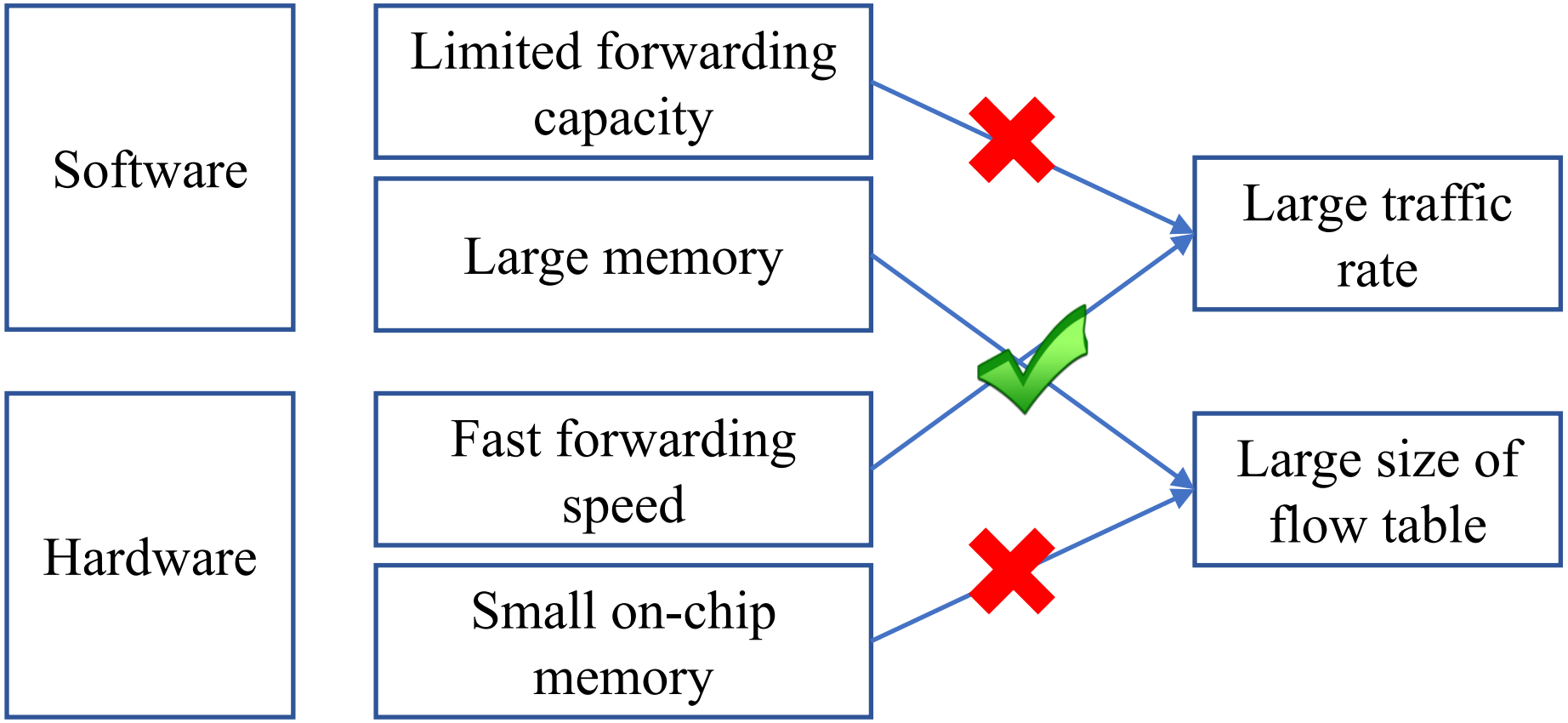
Flow Table

- More and more packet header fields taken as the input for forwarding rules, the size of flow table in modern network devices grows rapidly.



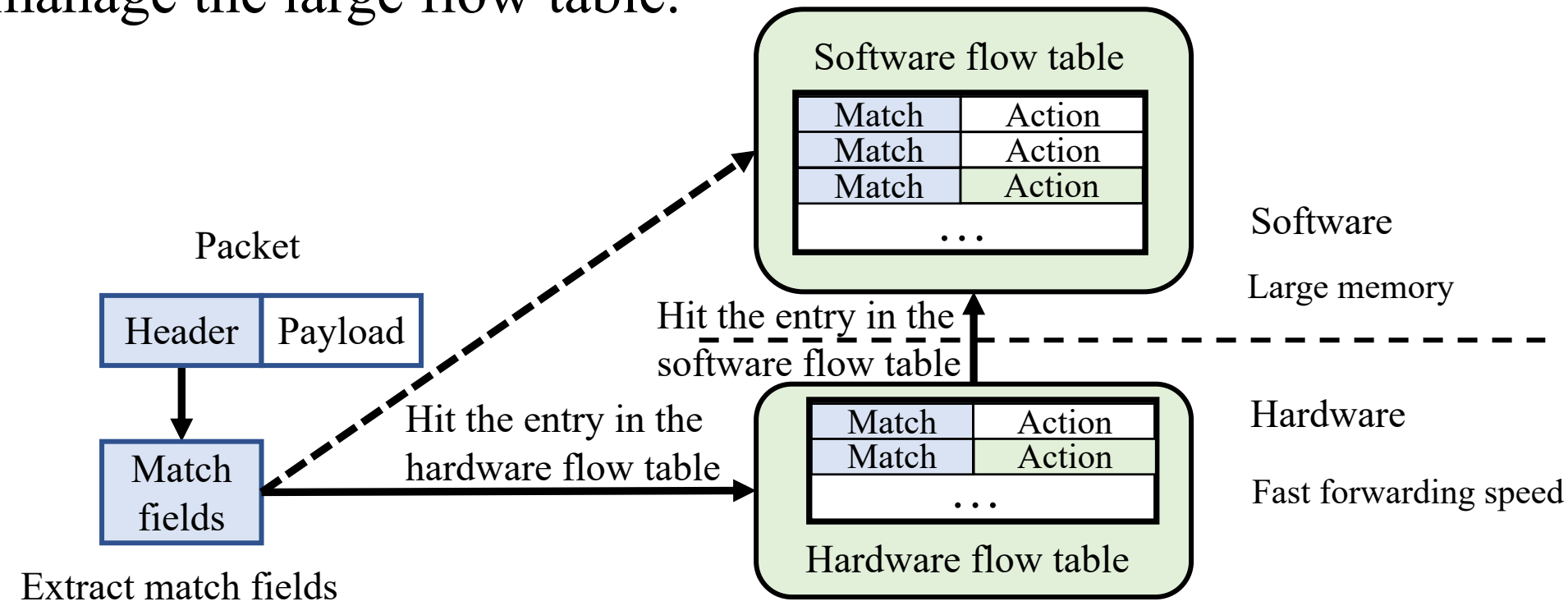
Hardware/Software characteristics

- The characteristics of hardware/software are complementary.



Hardware/Software Hybrid Flow Tables

- Many network devices take a hardware/software hybrid way to manage the large flow table.

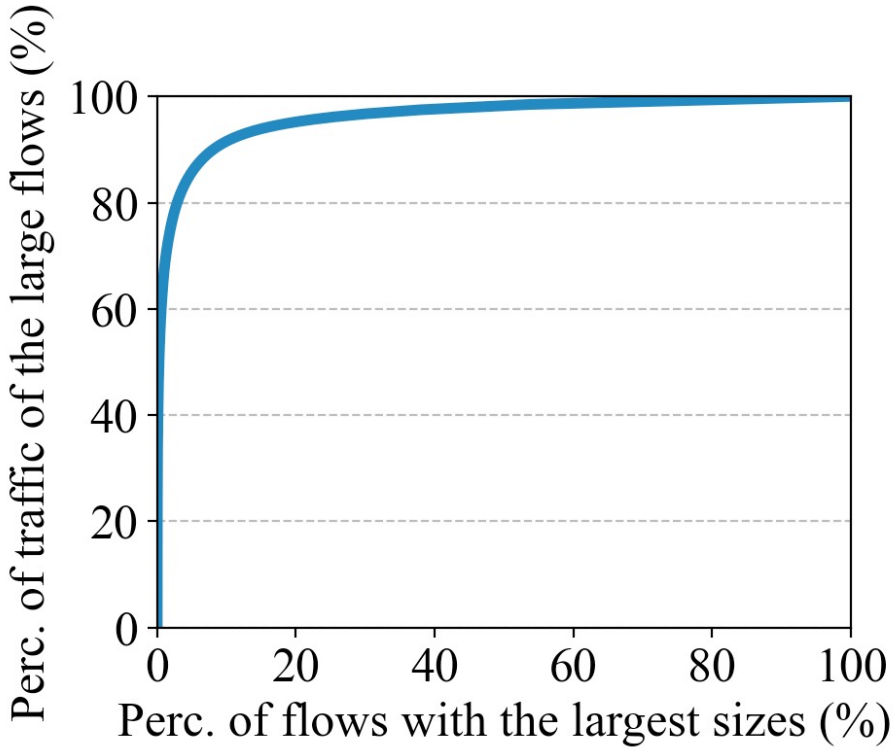


It is important to figure out how to split the flow table between hardware and software.

Limitations of Prior Work

- Previous works usually leverage the traffic skewness for flow table splitting.

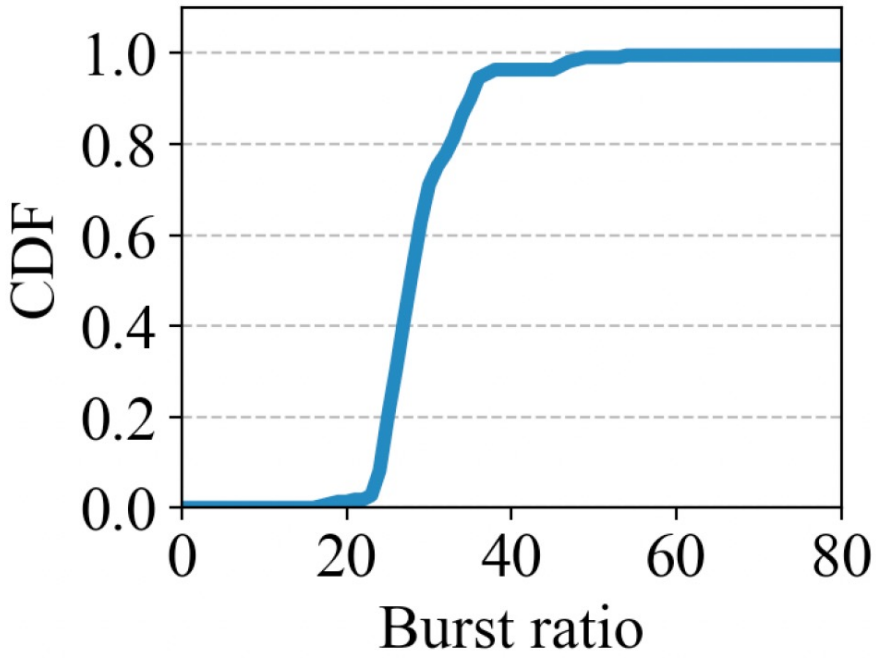
  [TFO SIGCOMM CCR'12, LFP TNSM'20]



- Offloading top 10% largest flows to the hardware can save up to 90% CPU resources?

Limitations of Prior Work

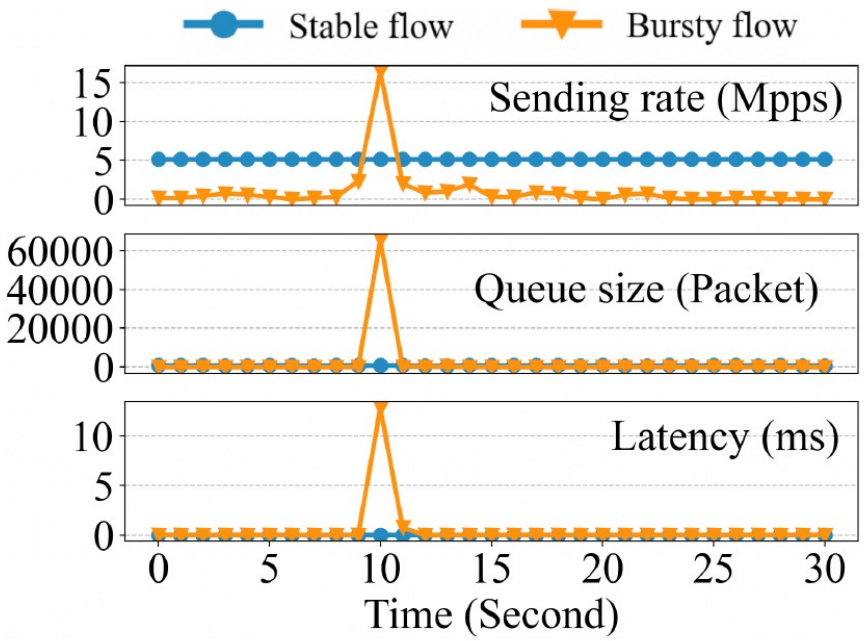
- However, the bursty flows are widely existent.

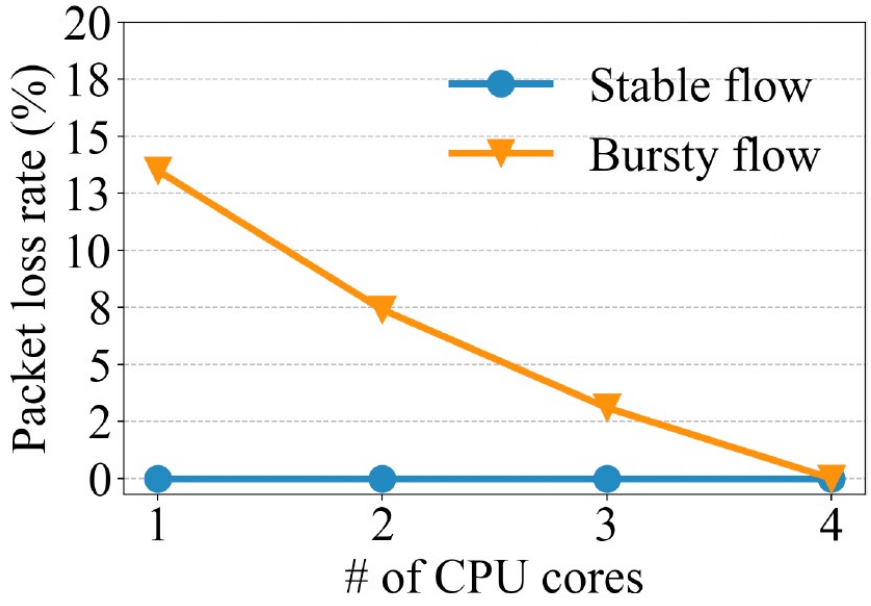


$$\text{burst ratio} = \frac{\text{peak rate}}{\text{average rate}}$$

Limitations of Prior Work

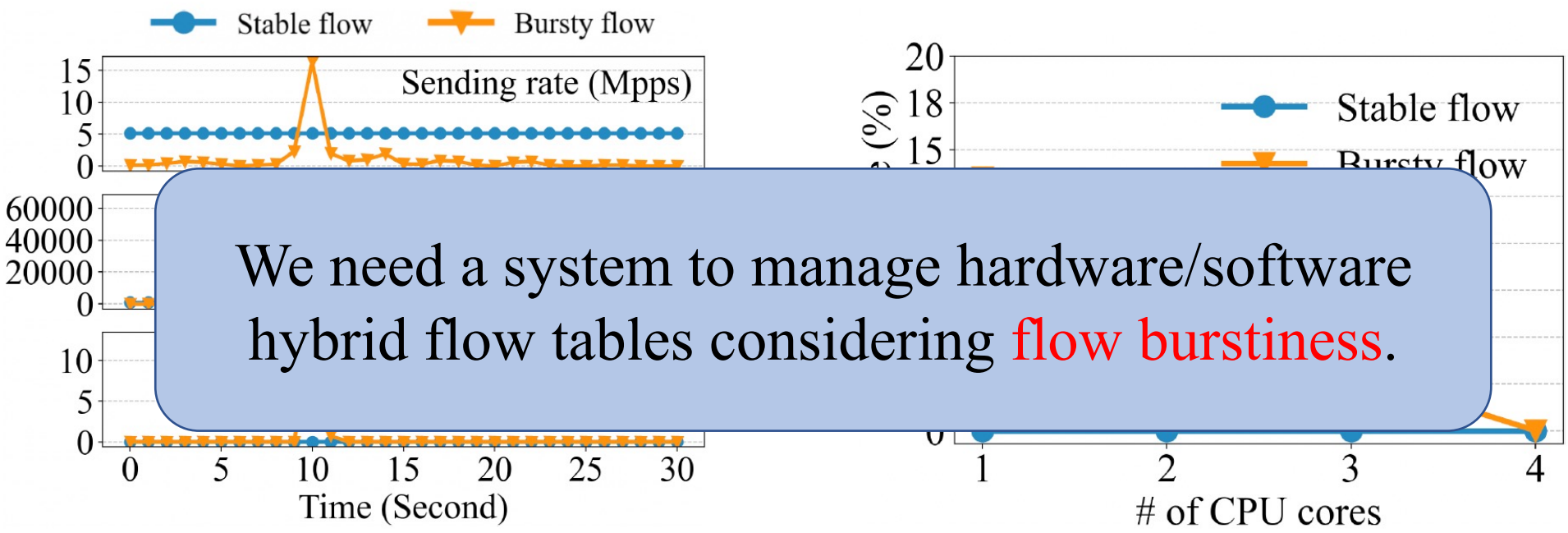
- Bursty flows lead to large queue size and high latency.
- Software need to reserve more resources for the bursty flow.





Limitations of Prior Work

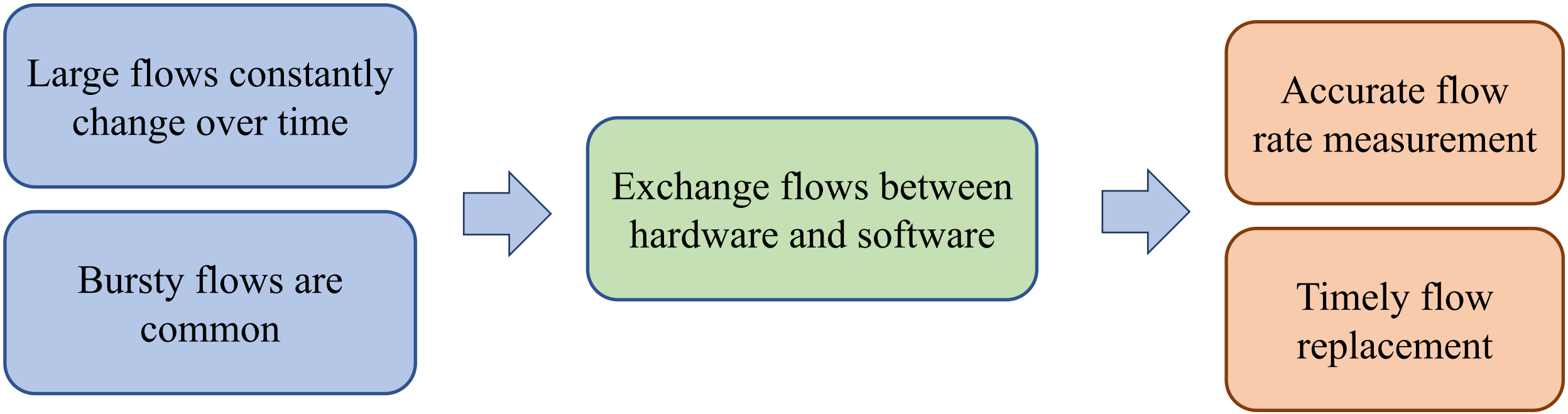
- Bursty flows lead to large queue size and high latency.
- Software need to reserve more resources for the bursty flow.



We need a system to manage hardware/software hybrid flow tables considering flow burstiness.

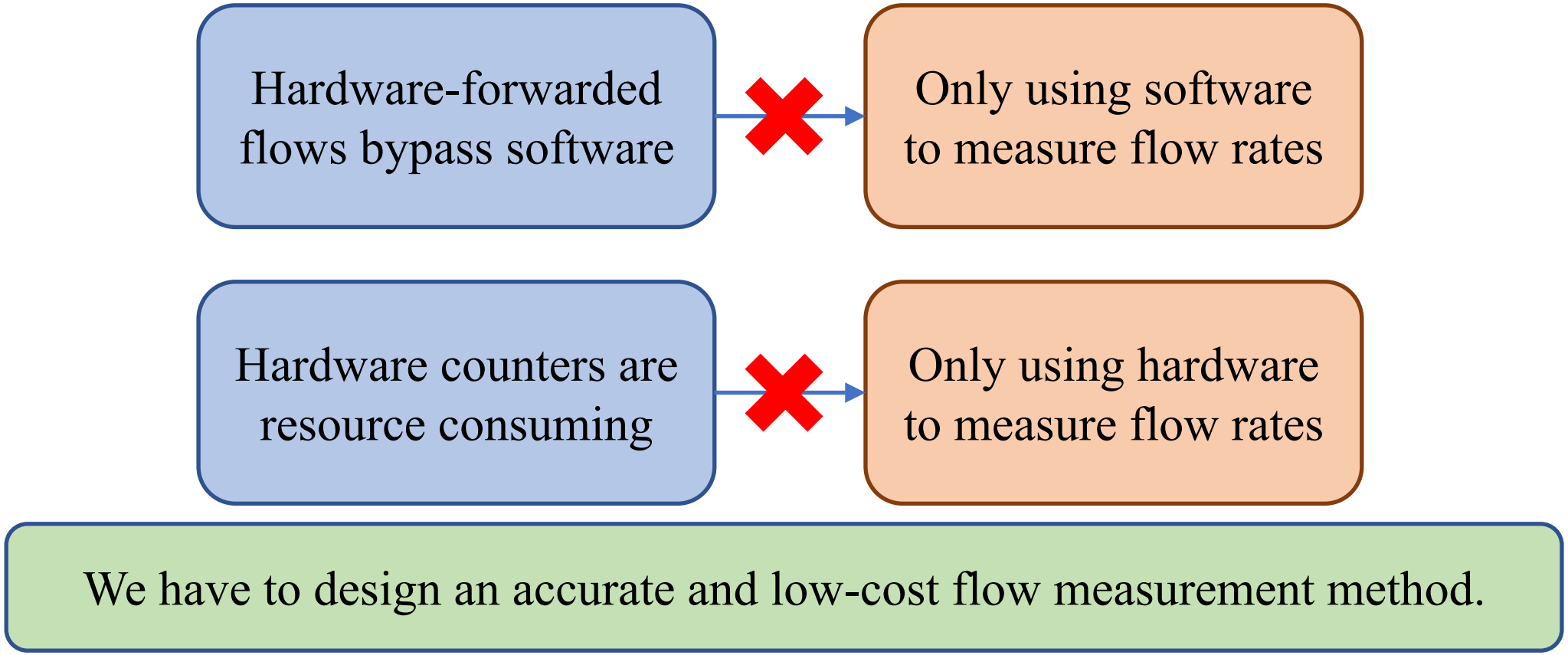
Design challenges

- Challenge # 1: Accurate flow rate measurement
- Challenge # 2: Timely flow replacement



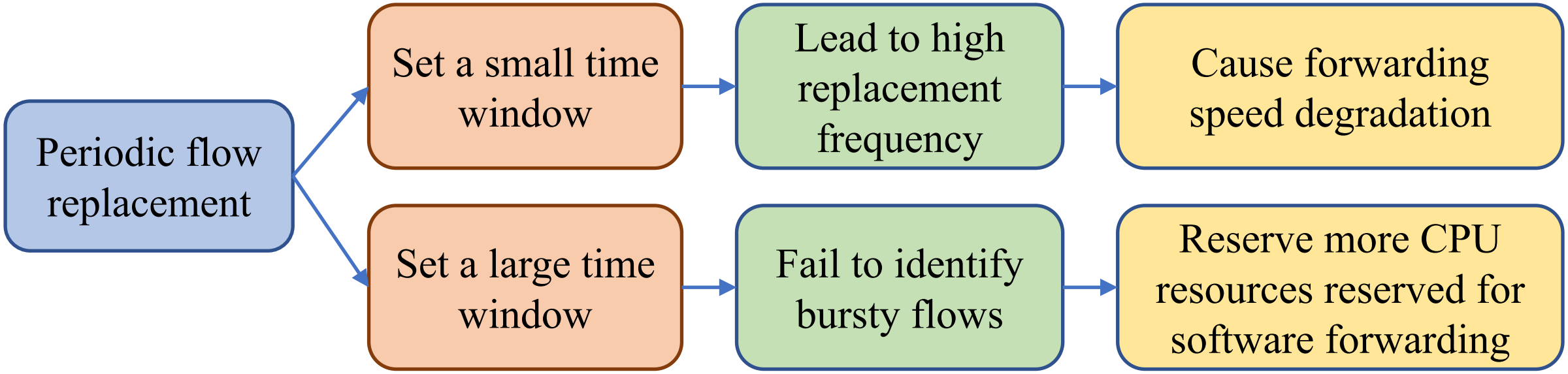
Accurate flow rate measurement

- Challenge # 1: How to accurately measure all the flow rates with low overhead on commodity devices?



Timely flow replacement

- Challenge # 2: How to set the proper timing for flow replacement between hardware and software?

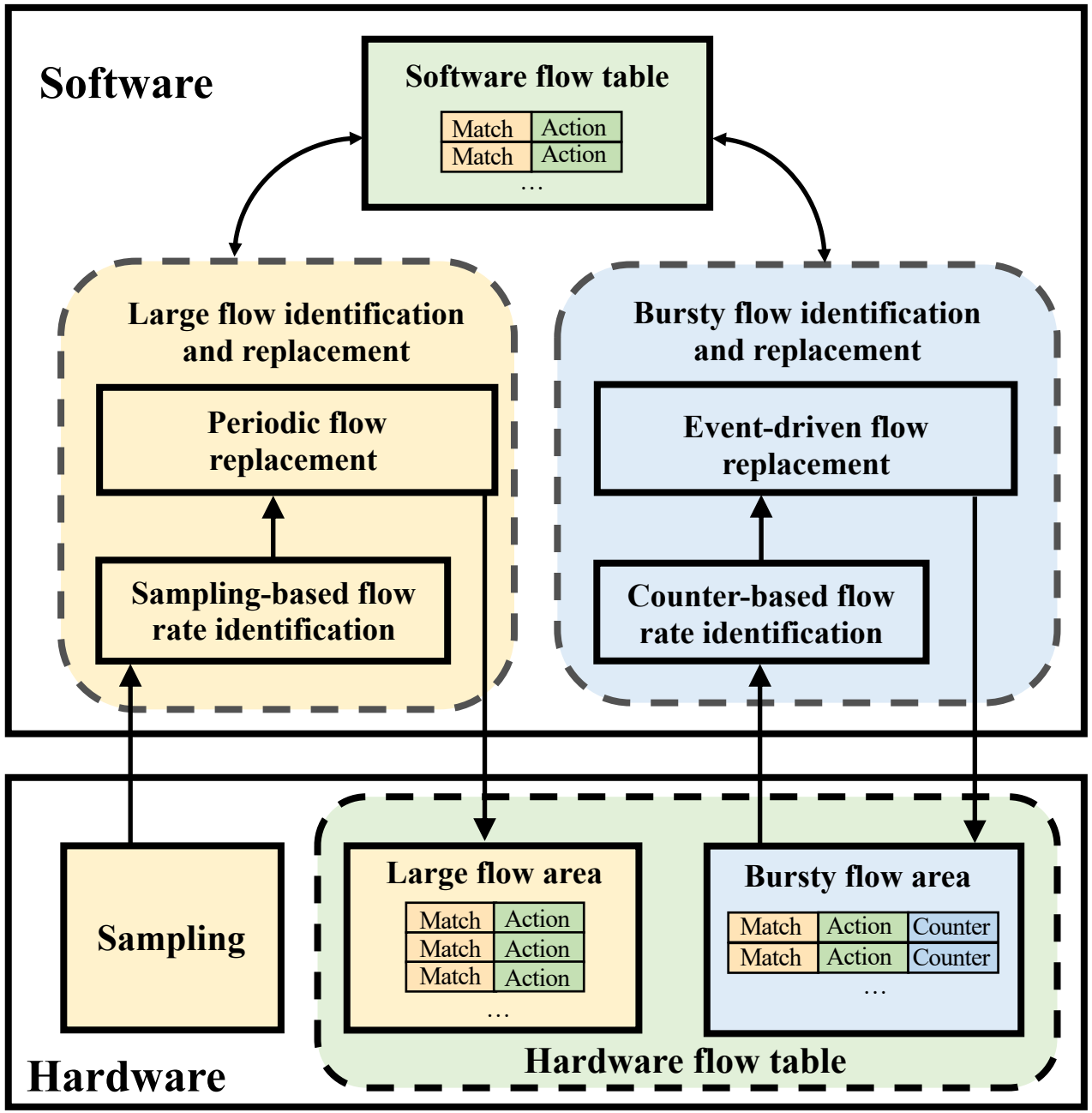


We have to design a method to identify bursty flows and keep replacement frequency low.

Our solution: Elixir

C1: Accurate flow rate measurement

C2: Timely flow replacement

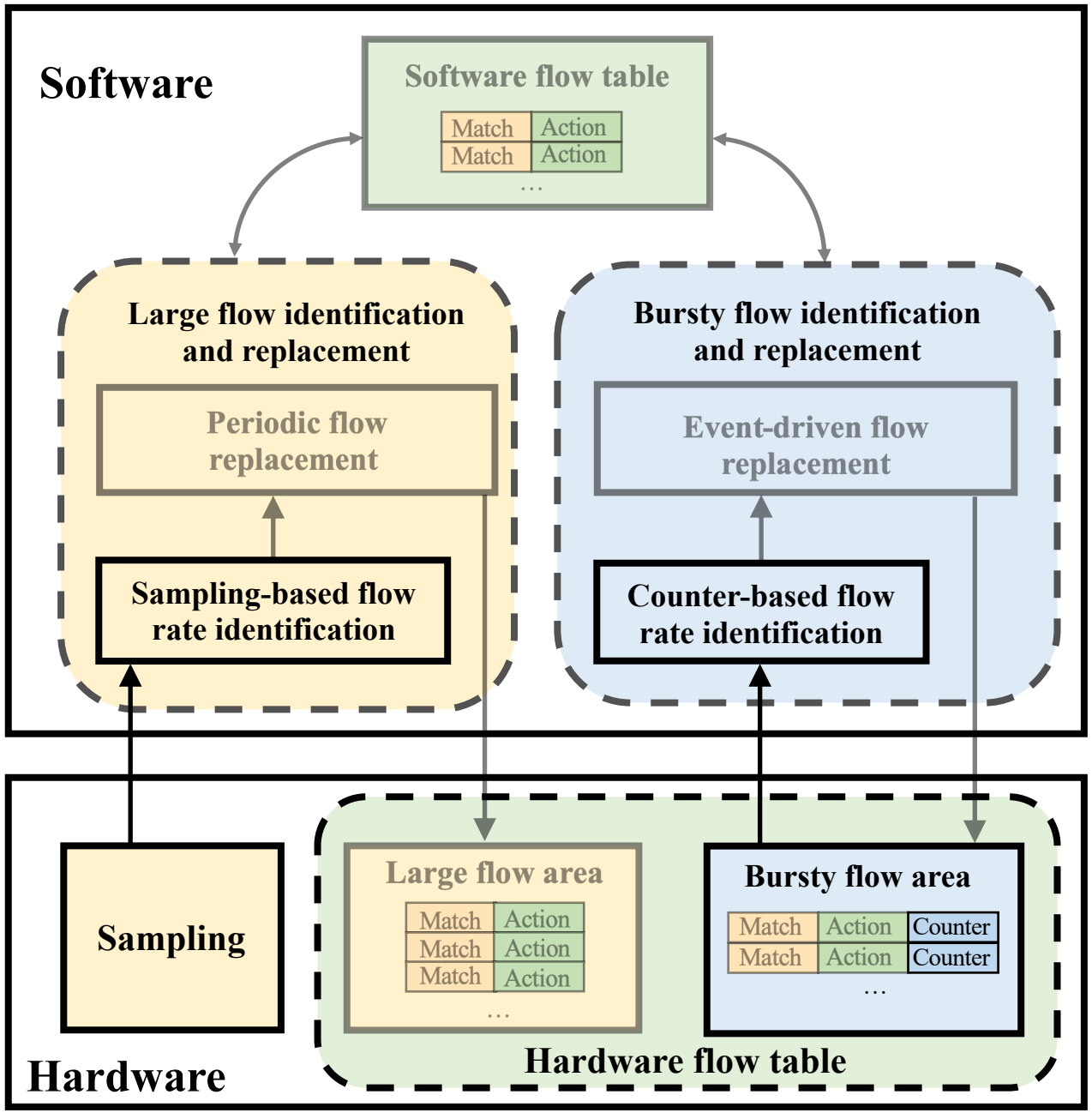


Our solution: Elixir

C1: Accurate flow rate measurement

C2: Timely flow replacement

S1: Separate the identification of large flows and bursty flow



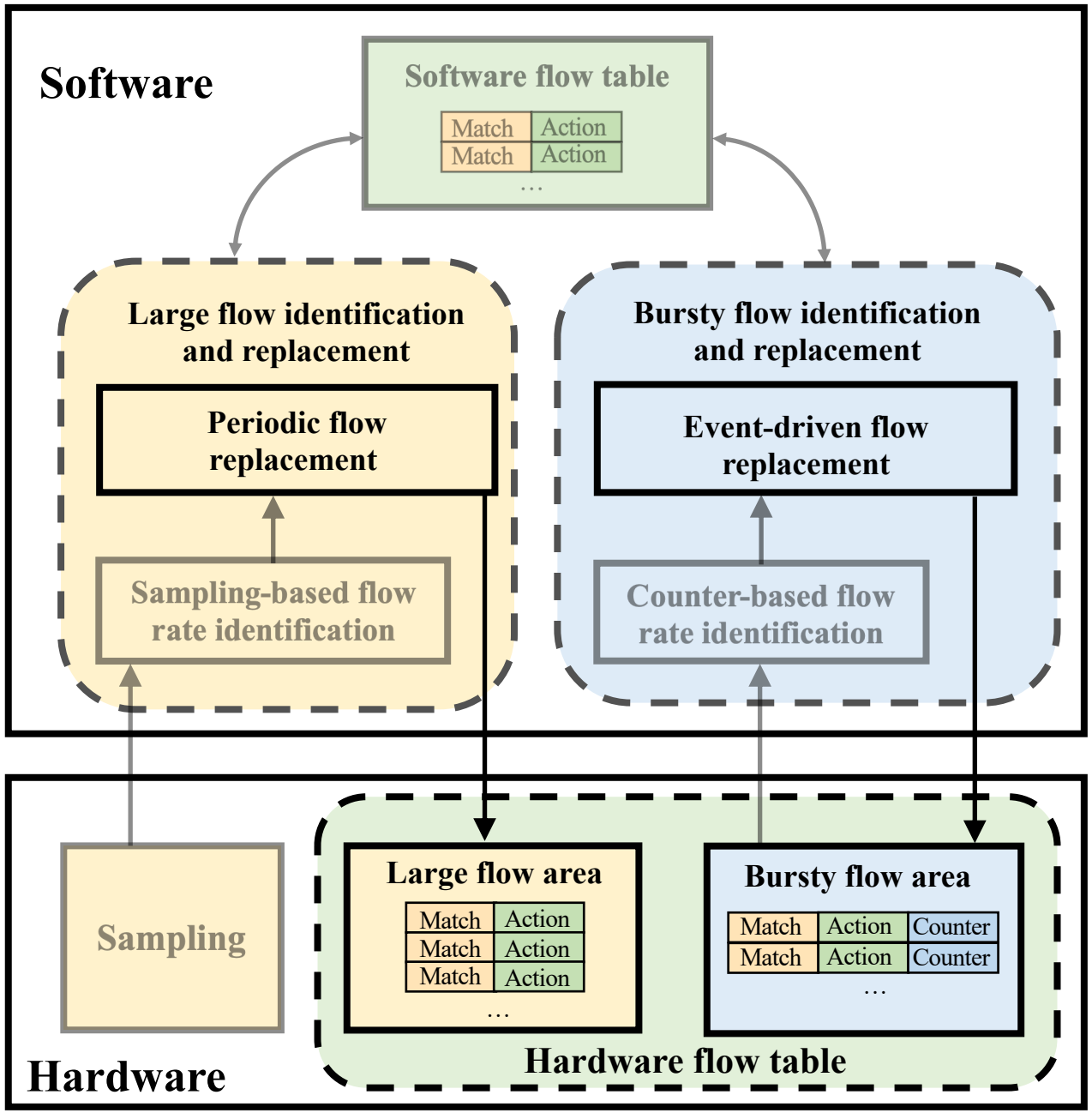
Our solution: Elixir

C1: Accurate flow rate measurement

S1: Separate the identification of large flows and bursty flow

C2: Timely flow replacement

S1: Separate the replacement of large flows and bursty flow



Our solution: Elixir

C1: Accurate flow rate measurement

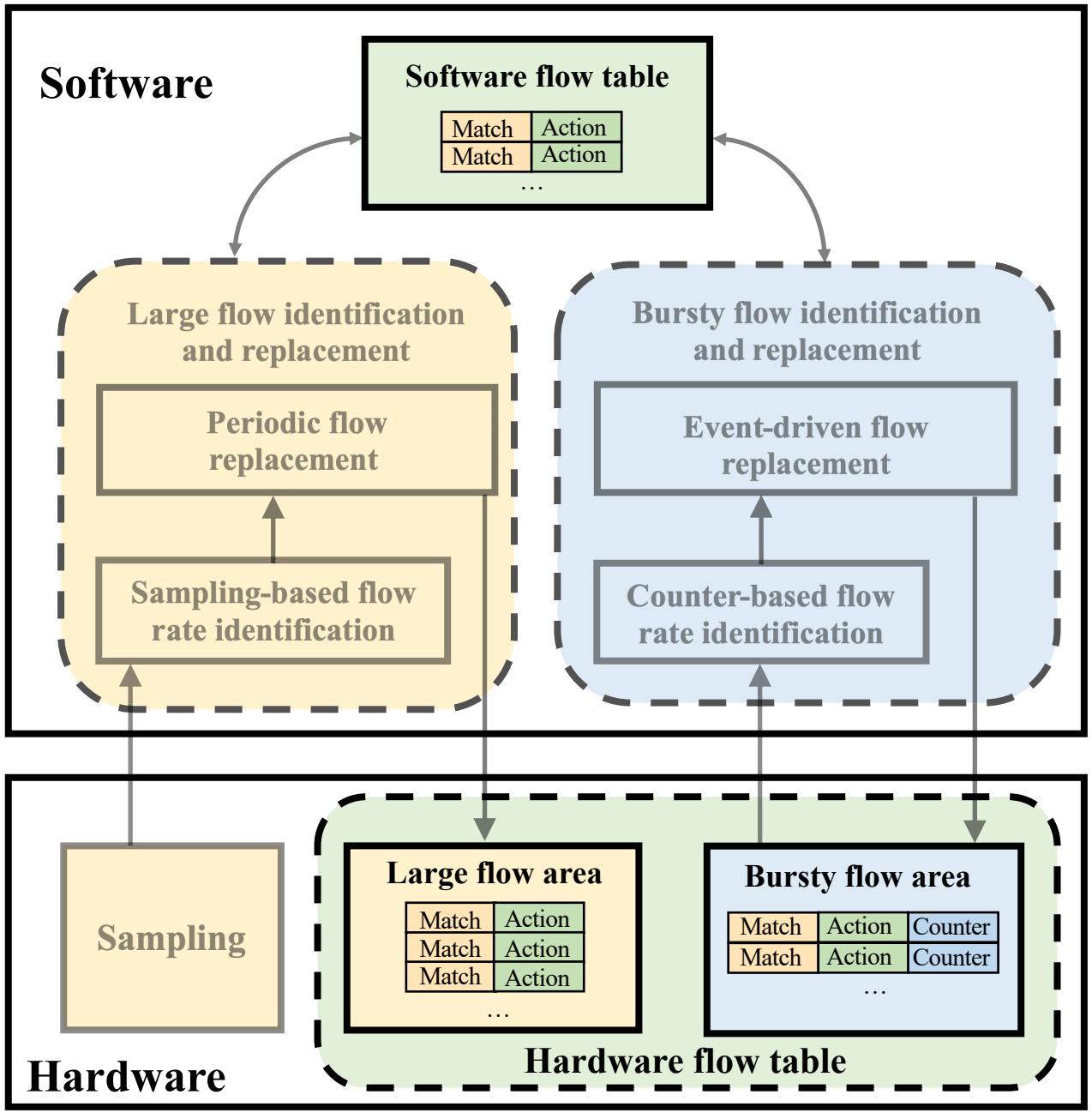
C2: Timely flow replacement

S1: Separate the identification of large flows and bursty flow

S1: Separate the replacement of large flows and bursty flow

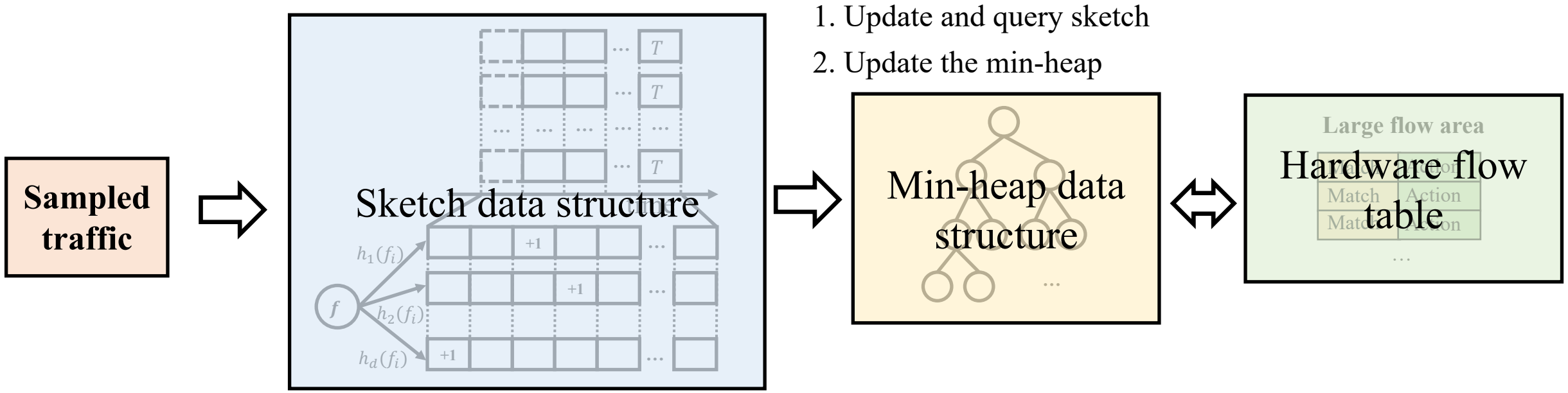
Shared software flow table

Separated hardware flow table



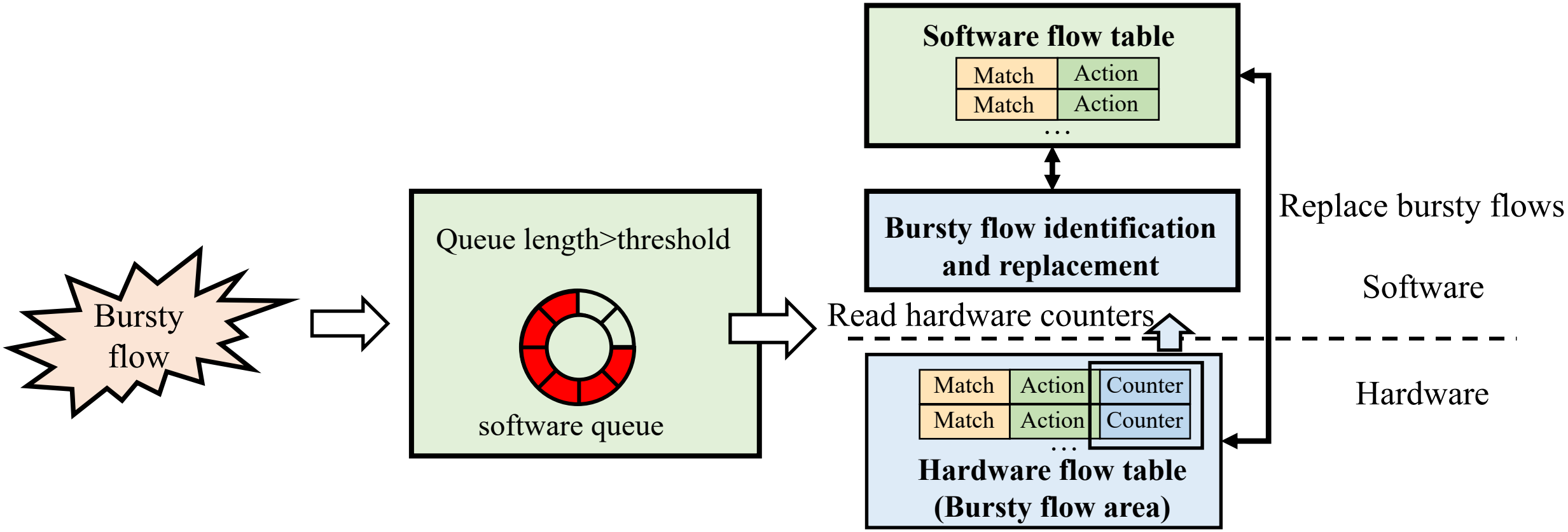
Periodic Large Flow Replacement

- A periodic procedure to exchange large flows.
- Sampled traffic to identify the rates of large flows.



Event-Driven Bursty Flow Replacement

- Elixir leverages an event-driven process to detect and offload bursty flows.

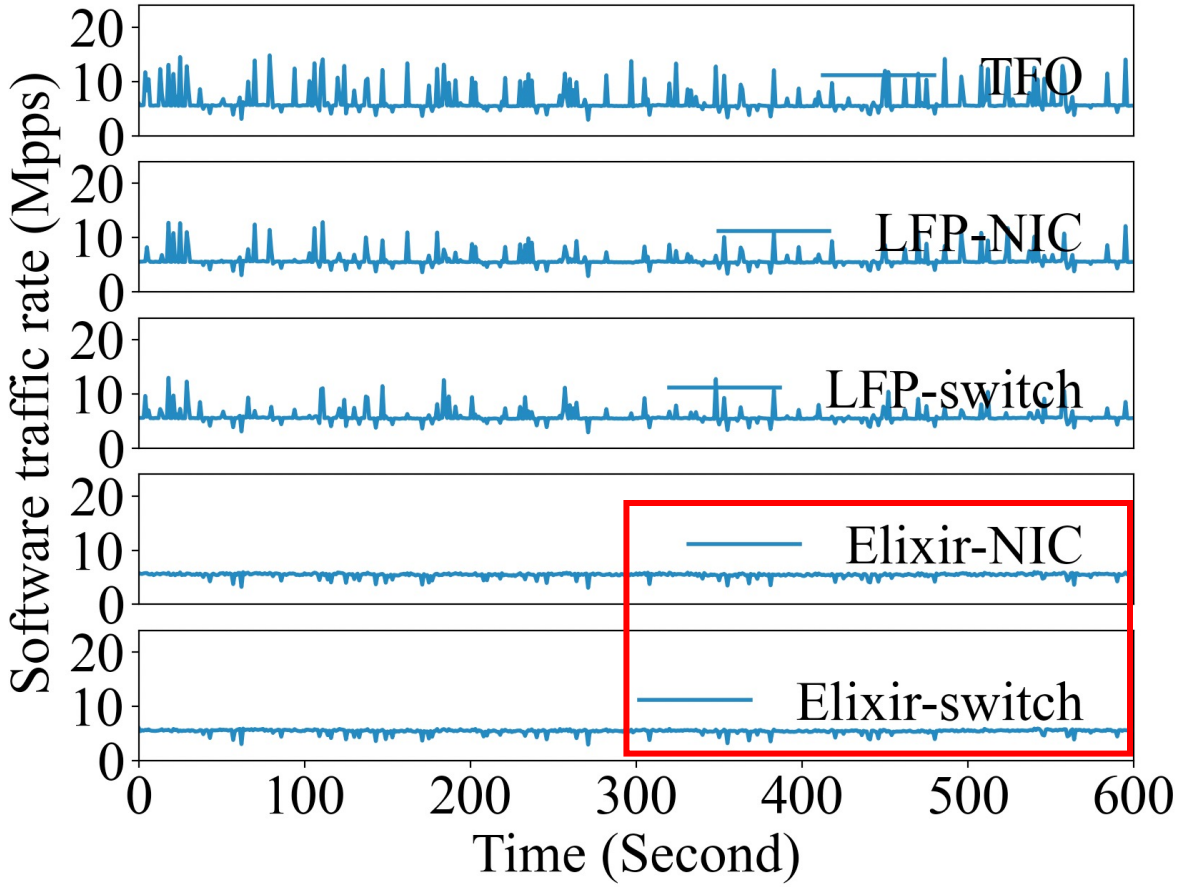


Evaluation

- Compared with:
  - The state-of-the-art solutions: TFO (Traffic-aware Flow Offloading) , LFP (Large Flow Predictor).
  - The baseline solution: PSF (Pure Software Forwarding).

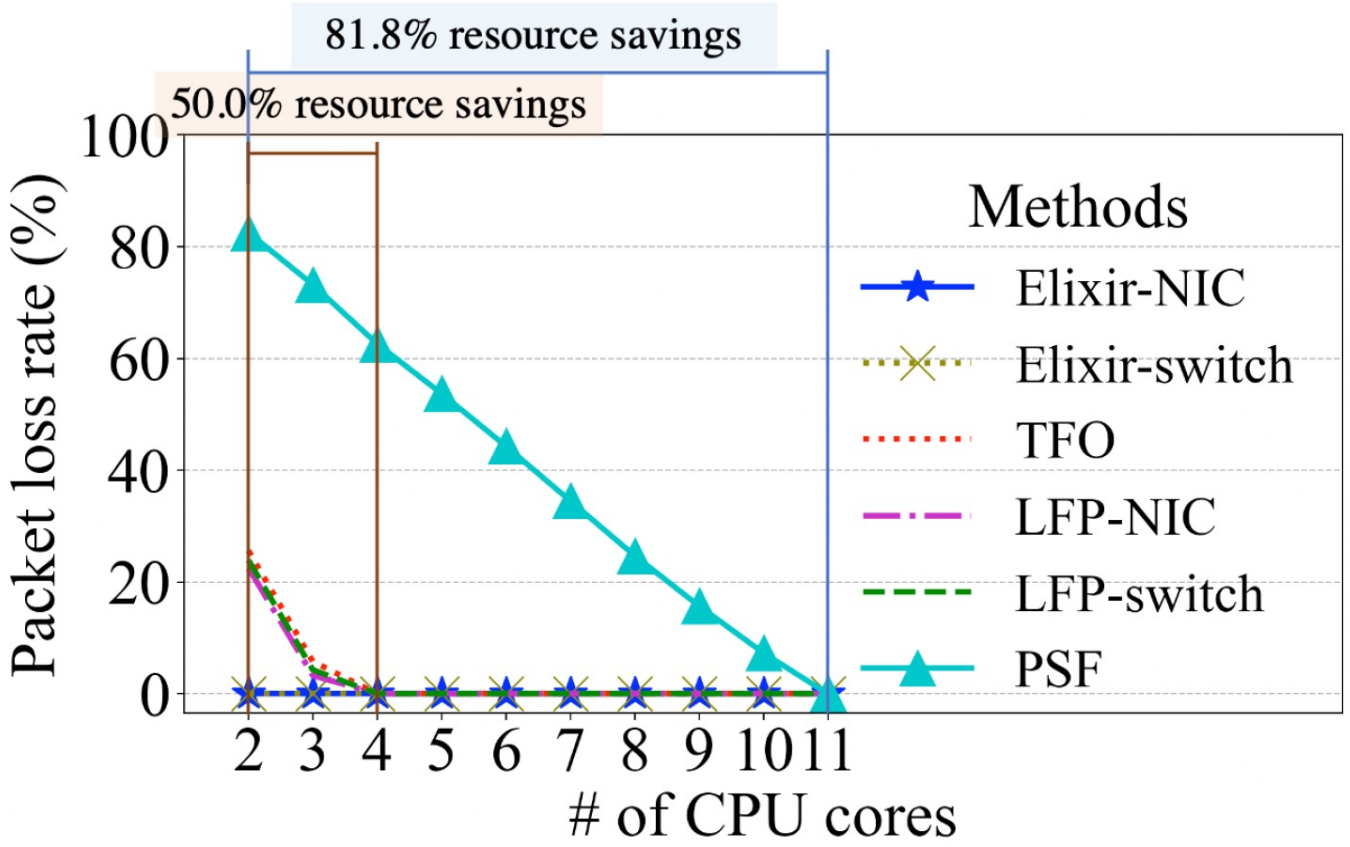
- Metric:
  - Traffic rate, CPU consumption, Forwarding capability, Latency, packet loss rate.

Evaluation:
Software-forwarded traffic rate



- Elixir can successfully keep the software forwarded traffic at a stable and low rate.

Evaluation:
CPU consumption with fixed forwarding capability



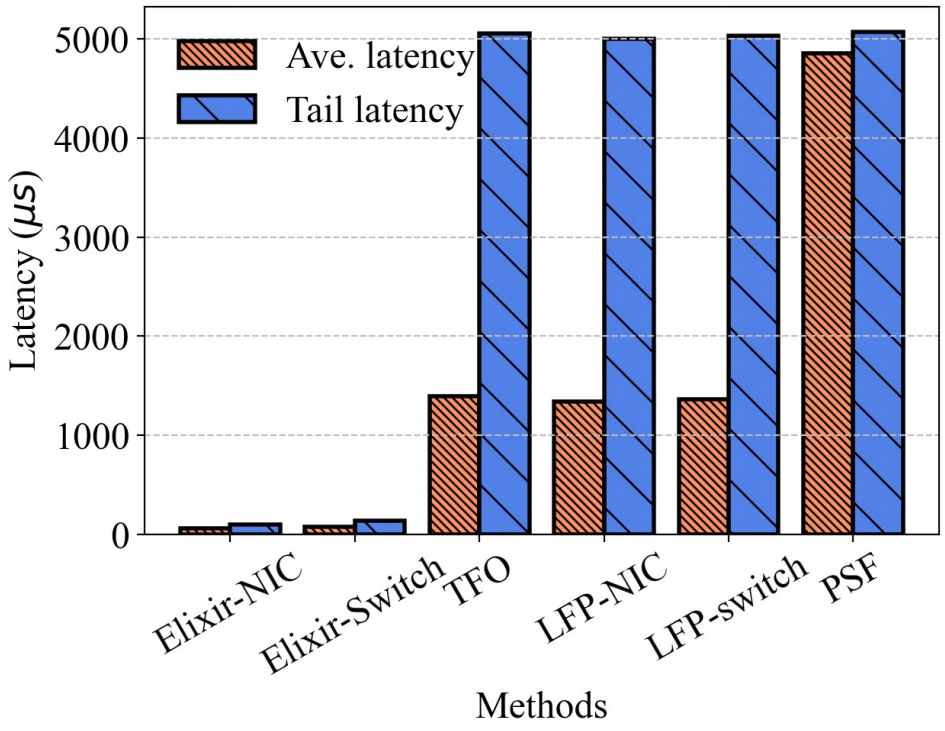
- Elixir can save ~81% CPU resources compared with PSF.
- Elixir can save ~50% CPU resources compared with TFO and LFP.

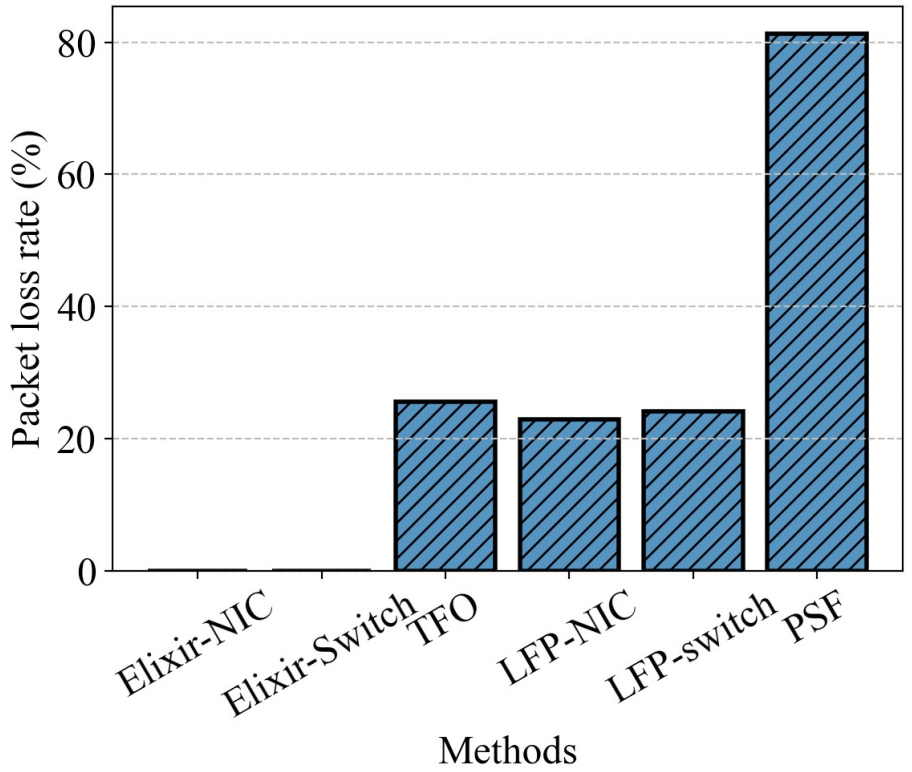
Evaluation:
Forwarding capability with fixed CPU consumption

- Maximum forwarding rate of Elixir is twice higher than that of TFO and LFP.
- PSF has the lowest maximum forwarding rate.

	Maximum forwarding rate (Mpps)
Elixir-NIC	56.2
Elixir-switch	56.1
TFO	26.4
LFP-NIC	28.3
LFP-switch	27.6
PSF	9.8

Evaluation:
Latency and packet loss rate with fixed traffic rate





- The tail latency of Elixir is ~97% lower compared with TFO and LFP.
- The packet loss rate of Elixir is lower than TFO and LFP.

Conclusion

- We present Elixir, the first system to manage hardware/ software hybrid flow tables considering flow burstiness.
  - Separate the identification and replacement processes of large flows and bursty flows.
- We prototype Elixir and compare it with the state-of-the-art solutions.
  - Less resource consumption.
  - Higher throughput.
  - Lower latency.

Thanks

Question?

wys17@mails.tsinghua.edu.cn