Cross-Layer Pacing for Predictably Low Age of Information

Presentation of work originally published in the Proceedings of the 2019 Workshop on Ultra-Low Latency in Wireless Networks

Andreas Schmidt[1], Stefan Reif[2], Pablo Gil Pereira[1], Timo Hönig[2], Thorsten Herfet[1],
Wolfgang Schröder-Preikschat[2]

Abstract: For dynamic systems, it is mandatory that all components operate on predictably "fresh" data. This requirement constitutes a challenge, in particular, when Internet-based or wireless communication is involved. We propose X-Pace, an approach based on cross-layer pacing, to achieve predictably low communication latency in the presence of varying network channel properties and node performance.

Keywords: Cross-Layer Optimization; Pacing; Age of Information; Low Latency; Transport Protocols

Emerging application domains such as the Internet of Things (IoT), smart factories, and inter-connected cars have initiated a transition from deeply embedded cyber-physical systems to collaborating systems. Such networked cyber-physical systems no longer operate in isolation, but use shared communication media to cooperate.

One of the key challenges is to ensure that all system components operate on "fresh" data. The age of information depends on data dependencies, scheduling, function execution times, queueing delays, and network transmission times. In modern systems, information processing chains are so long and complex that various sources of non-determinism accumulate, and potentially cause system components to operate on outdated data. The main reason for non-determinism is the dynamic behavior of networked cyber-physical systems. Besides varying network channel conditions in the Internet or wireless links, node performance depends on available resources, for example due to power constraints in embedded nodes, or contention on edge-located computing nodes.

To achieve predictably low age of information, we propose cross-layer pacing. It eliminates queueing delays, which are a main source of latency and jitter in large-scale networked systems, by making sure that (a) all processing steps run at the same speed and (b) data is produced exactly at the moment it can be consumed by the following processsing step. We have implemented this approach in a prototypical run-time system, X-Pace[3] [Sc19]. One fundamental building block is BBR [Ca16], a congestion control algorithm that utilizes the available network data rate efficiently, while keeping network-related queues empty.

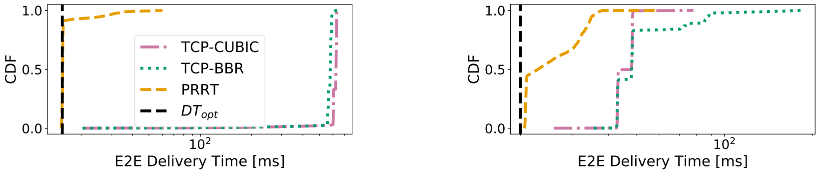
---

[1] Saarland Informatics Campus, `{andreas.schmidt,gilpereira,herfet}@cs.uni-saarland.de`

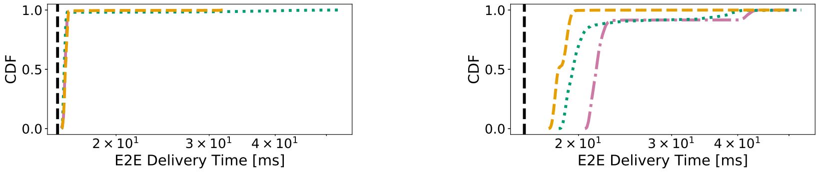
[2] Friedrich-Alexander-Universität Erlangen-Nürnberg, `{reif,thoenig,wosch}@cs.fau.de`

[3] X-Pace = short for cross-Layer pacing

doi:10.18420/inf2019_22



(a) Receiver System (b) Local Network Link



(c) Sender System (d) Internet Link

Fig. 1: Application-level packet delivery times depending location of the bottleneck (at system level for (a) and (c), at the network level for (b) and (d))

It simultaneously measures relevant channel properties needed for cross-layer pacing. In particular, these measurements allow our system to detect the current bottleneck component, to communicate the bottleneck pace through the entire system, and to enforce the correct speed at all layers in the system. Importantly, the pacing scheme includes the application layer. If an application constitutes the bottleneck, we detect this scenario by measuring its run-time. Otherwise, if the application is capable of running faster than appropriate, it is forced to slow down to enforce that all system components operate with the right timing. In summary, this cross-layer approach ensures that all buffers, network-based as well as node-based, remain empty.

We have implemented X-Pace in PRRT, a predictably reliable real-time transport protocol. It provides an ordered packet stream with partial reliability. The latter is based on expiry dates that are assigned to packets, incorporating the application's needs. PRRT also utilizes a hybrid error control scheme to achieve predictable reliability.

We compare X-Pace, implemented in PRRT, against TCP with the *CUBIC* and *BBR* congestion control algorithms. The results, visualized in Figure 1, demonstrate that X-Pace dynamically detects the bottleneck component and adapts appropriately. In consequence, PRRT achieves lower latency and jitter than both TCP variants.

Bibliography

[Ca16] Cardwell, Neal; Cheng, Yuchung; Gunn, C Stephen; Yeganeh, Soheil Hassas; Jacobson, Van: BBR: Congestion-based congestion control. ACM Queue, 14(5):50:20–50:53, Dec. 2016.

[Sc19] Schmidt, Andreas; Reif, Stefan; Gil Pereira, Pablo; Hönig, Timo; Herfet, Thorsten; Schröder-Preikschat, Wolfgang: Cross-Layer Pacing for Predictably Low Latency. In: Proc. 6th Intl. Worksh. on Ultra-Low Latency in Wireless Networks (Infocom ULLWN). IEEE, Apr. 2019.