Formal Modelling and Analysis of Stream Processing Systems

Linh Thi Xuan Phan
Department of Computer and Information Sciences, University of Pennsylvania
Email: linhphan@seas.upenn.edu

Stream processing systems (SPS) encompass classes of prevalent systems such as media and graphics processing, network packet processing, and control systems. Diverse in range and features, these systems operate on a large amount of data in a stream-like fashion, thus availing themselves to modelling using *stream abstraction*.

Unlike traditional systems, SPS exhibit both the complexity of data-oriented and control-oriented applications. The event streams are highly bursty in nature and usually comprising multiple types, each with a different execution demand. The architecture resources are also heterogeneous, and they often vary in scheduling policies. Further, the processing of the events frequently depends on the state of the systems. For instance, the processor may provide different amount of resource to different streams depending on the fill-level of the buffers in the system. It may otherwise stall on a full output buffer. The processing of the events possibly also requires synchronisation among different streams.

Due to the above complex characteristics of SPS, current analysis methods are often inadequate due to their lack of modelling capability and accuracy. Here, we propose a spectrum of novel models and analysis techniques for tackling SPS, as depicted in Figure 1 and elaborated below.

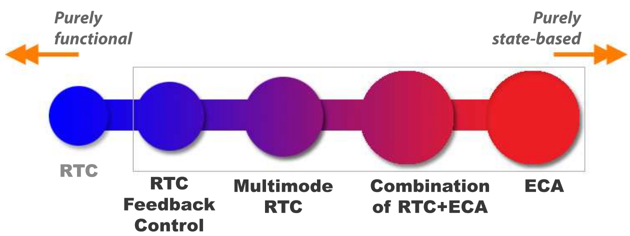

Fig. 1. Our models and analysis techniques (in rectangle box). The sizes of the circles represent their level of expressiveness. From left to right, the models advance in terms of expressiveness. From right to left, the models progress in terms of efficiency.

Event Count Automata (ECA) [2]. ECA is a state-based model that enables the modelling and analysis of SPS with state-information. The ECA model marries the count-based abstraction with the syntax of timed and hybrid automata, allowing for a natural and succinct representation of complex state-dependencies in SPS. We formulate the networks of ECAs communicating via buffers, which can be used to represent heterogeneous architectures and state-dependent processing and scheduling protocols. We apply ECA models to a range of scheduling policies and demonstrate how various performance properties can be analysed using existing verification techniques and tools.

Composition of ECA and Real Time Calculus (RTC) [4]. We introduce interface theories that enable the *composition of ECA models and RTC models* to achieve a good analysis trade-off between efficiency and accuracy. Using these interfaces, we develop a hybrid analysis framework for heterogeneous systems which comprise a mixture of components: with and without state-constraints. This compositional analysis framework allows one to leverage on the expressiveness of ECAs and the efficiency of RTC models. Analysis using this composed model is more efficient than using only ECAs and more accurate than using only RTC models.

Multi-Mode Real-Time Calculus (MMRTC) [3]. We introduce a second approach to integrate functional and state-based techniques in a mixed model termed MMRTC. The MMRTC model enables effective representation of multi-mode SPS, with streams (resources) specified as finite automata whose states are annotated with functions that define constraints on the arrival patterns of event streams (service patterns of the resources). We present a two-layered analysis technique that allows *RTC-based algebraic formulations to be combined with automata exploration techniques*, resulting in efficient analysis of MMRTC models. We develop technical results for computing various relevant properties of SPS using this technique.

RTC with Feedback Controls (RTC-FB) [1]. We introduce an algebraic technique to model and analyse a special class of state-dependencies in the existing RTC framework. Specifically, we show how back-pressure effects caused by the finite buffers such as blocking-write and state-based scheduling policies involving blocking-write can be captured using RTC-FB. We present methods to compute various performance properties of SPS that take into account of the capacity of the buffers. These properties can be computed efficiently in a purely functional and modular manner.

In addition, we present several case studies using variations of MPEG-2 decoders to demonstrate the applicability of our models and methods.

References

[1] A. Bouillard, L. T. X. Phan, and S. Chakraborty. Leight-weight modeling of complex state dependencies in stream processing systems. In *RTAS*, 2009.

[2] S. Chakraborty, L. T. X. Phan, and P. S. Thiagarajan. Event count automata: A state-based model for stream processing systems. In *RTSS*, 2005.

[3] L. T. X. Phan, S. Chakraborty, and P. S. Thiagarajan. A multi-mode real-time calculus. In *RTSS*, 2008.

[4] L. T. X. Phan, S. Chakraborty, P. S. Thiagarajan, and L. Thiele. Composing functional and state-based performance models for analyzing heterogeneous real-time systems. In *RTSS*, 2007.