On the Preciseness of Subtyping in Session Types: 10 Years Later

Tzu-Chun Chen  
Microsoft, Germany  
gina.chen@microsoft.com

Mariangiola Dezani-Ciancaglini  
University of Torino, Italy  
dezani@di.unito.it

Nobuko Yoshida  
University of Oxford, UK  
nobuko.yoshida@cs.ox.ac.uk

CCS CONCEPTS
• Theory of computation → Process calculi; Type theory.

KEYWORDS
π-calculus, Session types, Subtyping, Soundness and Completeness

ACM Reference Format:

The PPDP Most Influential Paper 10-Year Award for our work [11] was a wonderful surprise for us. We then looked at how the results in that paper have been used in the following literature. As outcome we wrote this short note in the hope of not missing too many crucial references.

Session types [23, 24, 26, 27, 36] are a successful formalism to structure interaction and to reason over communicating processes and their behaviour. The basic idea is to introduce a new form of polymorphism which allows the typing of channel names by structured sequences of types, abstractly representing the traces of channel usages. A crucial choice is if the processes communicate synchronously or asynchronously.

Subtyping enhances the expressiveness of session types. For synchronous processes, two distinct subtyping approaches have been proposed: one allowing the safe substitution of channels [18] and the other allowing the safe substitution of processes [13]. In our work [11], focused on replacing processes, we adopted the latter approach. We dub this subtyping synchronous subtyping. This subtyping has been extended to accommodate asynchronous communications [32], essentially capturing the freedom of outputs typical in such settings.

Preciseness of subtyping was first defined for the call-by-value λ-calculus with iso-recursive types in [30]. A subtyping is precise if it is both sound and complete. A subtyping relation is sound if no typable program is incorrect. It is complete if there is no strictly larger sound subtyping relation.

Our first major result in [11] was demonstrating the preciseness of synchronous subtyping for synchronous sessions. A key element in proving completeness was the construction of processes that characterise types.

Asynchronous communication is modelled using queues: output processes put messages in queues, while input processes read messages from these queues [26]. In such a scenario, the types should ensure not only deadlock-freedom, i.e., that input processes will always find messages, but also orphan message-freedom, i.e., that all messages in queues will eventually be consumed. This notion of soundness was first formulated in [11] and then widely adopted in the literature. The synchronous subtyping that is sound for synchronous sessions is also sound for asynchronous sessions, but it is not complete for asynchronous sessions. The subtyping proposed in [32] enjoys subject reduction, but it is unsound for asynchronous sessions because it does not ensure orphan message-freedom. It is also incomplete for asynchronous sessions. These incompleteness and unsoundness results are demonstrated in [11] through examples.

The main achievement of [11] is the definition of a new subtyping (dubbed asynchronous subtyping) together with the proof of its preciseness for asynchronous sessions. Again, a key aspect of proving completeness is the construction of processes that characterise types, even though this construction is much more complex than in the case of synchronous subtyping.

Preciseness can be defined operationally by means of processes or denotationally using type interpretations. Both forms of preciseness are proved for the synchronous and the asynchronous subtyping in [11]. Notably, [11] was the first paper discussing preciseness in the context of process calculi.

The most direct follow-up of [11] are papers discussing various aspects of preciseness. In [15], denotational and operational preciseness of subtyping for some λ-calculi and mobile processes are compared. While in [11] only binary sessions are considered, in [16] the operational and denotational preciseness of the synchronous subtyping for synchronous multiparty sessions types (MPST) is proved. The novelty of this paper is the introduction of characteristic global types to show the operational completeness.

In [10], new results about the uniqueness of precise subtyping relations are provided. In the same paper the approach of [11] is generalised to session initialisation and communication of expressions including shared channels. In [19] the preciseness of the synchronous subtyping for synchronous MPST is proved using a novel coinductive treatment of global type projections, based on global and local type trees.

For sure the most interesting development in this line of research are the papers [20, 21], where the first formalisation of the precise subtyping relation for asynchronous MPST is presented. The proof is based on a novel session decomposition technique, from full session types (including internal/external choices) into single input/output session trees (without choices). This session decomposition technique expresses the subtyping relation as a composition
of refinement relations between single input/output trees and provides a simple reasoning principle for optimising the order between asynchronous messages. A related denotational semantics of action permutations under different multiparty communication queues and buffers is studied in [14]. There the permutation defined in [21] is modelled as a valid semantic transformation which does not cause deadlock.

A second line of research inspired by [11] addresses the undecidability of asynchronous subtyping. In [4] a core undecidable subtyping relation (obtained by imposing limitations on the structure of types) is devised. As a result of this initial undecidability finding, the asynchronous subtyping and the subtypings of [31, 32] are all shown to be undecidable. The undecidability proof for asynchronous subtyping in [29] relies on a new Turing-complete subclass of two-party communicating finite-state machines, demonstrating that asynchronous subtyping is equivalent to the halting problem for this class of machines. The undecidability result for asynchronous session subtyping is used to obtain an undecidability result for asynchronous contract refinement in [7] and for asynchronous communicating finite state machines in [8]. A novel variant of session subtyping that leverages the notion of controllability from service contract theory and that is a sound characterisation of fair refinement is proposed in [6]. Also, this subtyping and the fair refinement are undecidable.

A natural reaction to the undecidability of asynchronous subtyping is the search for either decidable restrictions or algorithms which can terminate without providing a definitive answer. The decidability of a fragment that does not impose any limitation on communication buffers and allows both the subtype and the supertype to include multiple choices is shown in [5]. The algorithm in [3] is based on a tree representation of the coinductive definition of asynchronous subtyping; this tree could be infinite, and the algorithm checks for the presence of finite witnesses of infinite successful subtrees. The proposal of [2] uses sets of traces instead of trees. The obtained algorithm applies abstract interpretation techniques.

The promotion of asynchronous subtyping incorporation in applications is also an interesting follow-up of [11]. The first work which informally introduces the idea of asynchronous subtyping in practice is [25], where asynchronous multiparty subtyping enables the programmer to permute the order of messages for performance gain without introducing deadlock. The asynchronous subtyping is used to model the double buffering protocol [28]. This approach was implemented and evaluated in C [35, 38] and MPI-C [33, 34] in the context of high-performance computing.

The tool presented in [1] integrates several algorithms for checking subtyping that can be invoked from an easy-to-use Python GUI. This interface allows users to input, using standard session type syntax, two types: the candidate subtype and supertype.

The recent work [9] proposes CAMP, which is a static performance analysis framework for message-passing concurrent and distributed systems based on MPST. CAMP augments MPST with annotations of communication latency and local computation cost, defined as estimated execution times, that is used to extract cost equations from protocol descriptions and to statically predict the communication cost. CAMP is also extended to analyse asynchronous communication optimised programs. The tool based on cost theory is applicable to different existing benchmarks and use cases in the literature with a wide range of communication protocols, including the implementations in [34, 35].

The Rust programming framework, Rumpsteak [12], incorporates multiparty asynchronous subtyping [21] to optimise asynchronous message-passing in the Rust programming language. Specifically, the authors propose an algorithm for asynchronous subtyping based on the session decomposition technique in [20, 21] that is bounded by a number of iterations and proved to be sound and decidable. They evaluate the performance and expressiveness of Rumpsteak against three previous Rust implementations. Rumpsteak is more efficient and can safely express many more examples by offering arbitrary ordering of messages. The authors also analyse the complexity of the new algorithm and benchmark it against the binary session subtyping algorithm in [3]. The algorithm in [3] turns out to be exponentially slower than Rumpsteak.

Hinrichsen’s PhD thesis [22] introduces Actris, a Coq tool that integrates separation logics and asynchronous binary session types with the asynchronous subtyping in [11].

The first formalisation of multiparty asynchronous subtyping within the Coq proof assistant is given in [17]. Session types are decomposed into session trees that do not involve choices, and then a coinductive refinement relation is established over them to govern subtyping. This approach allows for the proof of example subtyping schemas that appear in the literature. Notably, to the best of our knowledge, no other decidable sound algorithm is able to verify all these examples.

We conclude by observing that we were wrong in [11], since we wrote: “Algorithms for checking the synchronous and asynchronous subtypings of the present paper can be easily designed”. In fact, while there are algorithms for synchronous subtyping (see [37] and the references there), the asynchronous subtyping is undecidable as discussed above. The challenge of asynchronous subtyping remains intriguingly complex and theoretically rich. This underscores the evolving nature of the field and opens avenues for future exploration.

ACKNOWLEDGMENTS

We thank Alceste Scalas for his collaboration in writing the journal version [10] of [11]. We are grateful to Ilaria Castellani for her valuable suggestions that improved the presentation of this abstract. The last author is partially supported by EPSRC EP/T006544/2, EP/K011715/1, EP/K034413/1, EP/L00058X/1, EP/N027833/2, EP/N028201/1, EP/T014709/2, EP/V000462/1, EP/X015955/1 and Horizon EU TaRDIS 10109306.

REFERENCES
