Parameterized Concurrent Multi-Party Session Types

Minas Charalambides
charala1@illinois.edu

Peter Dinges
pdinges@acm.org

Gul Agha
agha@illinois.edu

Department of Computer Science
University of Illinois at Urbana–Champaign, USA
http://osl.cs.uiuc.edu/

Session types have been proposed as a means of statically verifying implementations of communication protocols. Although prior work has been successful in verifying some classes of protocols, it does not cope well with parameterized, multi-actor scenarios with inherent asynchrony. For example, the sliding window protocol is inexpressible in previously proposed session type systems. This paper describes System-A, a new typing language which overcomes many of the expressiveness limitations of prior work. System-A explicitly supports asynchrony and parallelism, as well as multiple forms of parameterization. We define System-A and show how it can be used for the static verification of a large class of asynchronous communication protocols.

1 Introduction

Session types [27] are a means of expressing the order of messages sent by actors [11] (or processes). In particular, session types can be used to statically check if a group of processes communicate according to a given specification. In these systems, a global type specifies the permissible sequences of messages that participants may exchange in a given session, as well as the types of these messages. The typing requires the programmer to provide the global type. A projection algorithm then generates the restrictions implied by the global type for each participant. Such restrictions are called end-point types or local types and describe the expected behavior of individual participants in the protocol. The actual program code implementing the participant behavior is checked for conformance against this localized behavior specification. We are interested in generalizing prior work on session types to typing coordination constraints on actors, which can then be enforced e.g. with Synchronizers [19, 20, 17] or other ways [30].

This requires addressing two limitations of previous work. First, session types do not (directly) support asynchronous events; asynchronous communication leads to delays which require considering arbitrary shuffles. Second, we wish to consider parameterized protocols. For example, consider two actors communicating through a sliding window protocol: the actors agree on the length of the window (i.e., the number of messages that may be buffered) and then proceed to an exchange of concurrent messages. Prior work on session types is not suitable for typing protocols like this: the reason for this deficiency is the fact that their respective type languages depend on other formalisms for type checking (such as typed $\lambda$-calculus [3] or System T [25]) and these formalisms do not support a parallel construct.

Contributions. We overcome many of these shortcomings by developing System-A, a new system for expressing types for multi-party interaction that does not depend on other formalisms for type checking. The main contributions are (a) parameterized constructs for expressing asynchrony, parallelism,
sequence and choice (Section 4), (b) a projection mechanism to provide type constraints on individual actors (Section 5), (c) the conditions under which conformance of the latter with the global type is assured (Sections 7, 8) and finally (d) we show that structural equivalence of types is decidable in System-A, by proving strong normalization of our local types (Section 6). Proofs of our theorems are included in the long version of this paper [13].

Limitations. Using the strong normalization result, we can decide whether the local behavior of an actor follows the protocol. However, this result relies on a type inference mechanism for the actor’s behavior (of the sort in Alur et al. [2]). We do not describe such a type inference mechanism in this article. Moreover, we omit support for session delegation. Finally, our realizability results rely on structural criteria and are hence conservative rather than precise [4].

2 Related Work

Session types [27, 39, 36, 26] originate from the context of $\pi$-calculi as statically derivable descriptions of process interaction behaviors. In two-party sessions, they allow us to statically verify that the participants have compatible behavior by requiring dual session types, that is, behaviors where each participant expects precisely the message sequence that the other participant sends and vice versa. Extensions to session types support asynchronous message passing [32] and introduce subtyping [21] for a looser notion of type compatibility. Session types have been integrated into functional [37, 35] and object-oriented [16, 29, 23] languages among others, with a wide range of applications including deadlock and livelock detection [24]. Other extensions deal with evolving system specifications using transformations [18]. Exception handling, which allows the participants of a protocol to escape the normal flow of control and coordinate on another, has been considered in [11, 10]. The present article combines three enhancements to session types that majorly extend their applicability: concurrent multi-party sessions, parameterized session types, and an enhanced syntax.

Asynchronous Multi-Party Sessions. Many real-world protocols involve more than two participants, which makes their description in terms of multiple two-party sessions unnatural. To overcome this limitation, Honda et al. [28] extend session types to support multiple participants: a global type specifies the interactions between all participants from a global perspective. A projection algorithm then mechanically derives the behavior specification of the individual participants, that is, the local type.

The notion of global type and the associated correctness requirements for projection were first studied by Carbone et al. [9]; Bonelli’s work on multi-point session types [7] treats multi-party protocols from the local perspective only. Bettini et al. [6] allow multi-party sessions to interleave and derive a type system guaranteeing global progress. Gay et al. [22] consider subtyping in presence of asynchrony.

The present article builds on the foundation of a global protocol specification and its projection onto local behaviors [28]. However, we do not address the question of local type safety and inference of actual programs, which is a major part of Honda et al.’s work. Furthermore, unlike their approach (but following Castagna et al. [12]), we simplify the notation for global types by replacing recursion with the Kleene star and limiting each pair of participants to use a single channel. We introduce an explicit shuffle operator to preserve the commutativity of message arrivals that can be achieved using multiple channels. Explicit shuffles also reduce the need for a special subtyping relationship that allows the permutation of (Lamport-style) concurrent asynchronous events for optimization [31].
Following Castagna’s global type syntax further, we support join operations. Joins cannot be expressed in Honda et al.’s global types because of the linearity requirement. However, as Denièlou remarks [15], join operations can only describe series–parallel graphs. Protocols such as the alternating bit protocol that require interleaved synchronization between two processes consequently cannot be expressed in our global type language. Our choice to not support generic graph structures as global types is founded on the desire to support parameterization and, at the same time, keep the language understandable; it remains unclear to us how to visualize parameterized graphs in an intuitive fashion.

**Parameterized Session Types.** Our major extension of global types over Honda et al. and Castagna et al.’s work is the introduction of parameters. The starting point for our parameterization of session types is the work of Yoshida et al. [38] and Bejleri [5]. Yoshida et al. augment the global types of Bettini et al. [6] with primitive recursive combinators to obtain dependent types that support the parameterization of the repetition count and the connection topology. This allows, for example, using a single global type for a highly participant-count dependent butterfly network. Static verifiability—without instantiating the type parameters—is maintained by projecting onto parameterized local types that allow syntactic comparisons. In [14], Denièlou et al. achieve parameterization by means of quantification over behavior specifications they call roles. Like Bettini et al. and unlike System-A, neither Denièlou et al. nor Yoshida et al. support arbitrary, concurrency-induced shuffles in their global and local types. While Bettini et al. regain parallel composition through the interleaving of global types, it is unclear how the results transfer to the other two approaches.

**Modeling of Multi-Party Protocols.** Formalisms for describing multi-party communication protocols have been studied in the context of designing distributed systems and cryptographic protocols. As modeling tools, the formalisms provide ways to check a protocol for desired properties [40], or to synthesize such protocols [34]. In contrast to session types, the formalisms lack ways to statically verify the compliance of an actual protocol implementation against the specification. Denièlou and Yoshida [15] discuss session types and their relation to work on distributed systems or cryptographic protocols in greater depth.

### 3 Motivation

Formalisms introduced in previous work are not expressive enough to define the types of some interesting protocols such as the sliding window protocol, a locking–unlocking protocol, and a case of limited resource sharing. In this section, we demonstrate how the behavior of these protocols can be described in System-A.

**The Sliding Window Protocol.** Assume an actor \(a\) sends messages of type \(m\) to an actor \(b\), which acknowledges every received message with an \(ack\) message. The protocol determines that at most \(n\) messages can be unacknowledged at any given time, so that \(a\) ceases sending until it receives another \(ack\) message. In this example, the window size \(n\) is a parameter, which means we need a way to express the fact that \(n\) sending–acknowledging events can be in transit at any given instant in time. Following is the global type of the protocol.

\[
(a \xrightarrow{m} b ; b \xrightarrow{ack} a)^* \parallel (a \xrightarrow{m} b ; b \xrightarrow{ack} a)^* \parallel \ldots \parallel (a \xrightarrow{m} b ; b \xrightarrow{ack} a)^*
\]

\(n\) times
\( a \rightarrow m \rightarrow b \) denotes that \( a \) sends a message of type \( m \) to \( b \). Operator \( ; \) is used for sequencing interactions. \( \parallel \) is used for composing its left and right arguments in parallel. The Kleene star has the usual semantics and takes precedence over \( \parallel \).

The above type can be expressed using the notation of Castagna et al. [12], albeit with a fixed window size \( n \). In System-A on the other hand, we can parameterize the type in \( n \) and statically verify that participants follow the protocol without knowing its value at runtime. Using \( \parallel_i \rightarrow \) to denote the parallel composition of \( n \) processes, we obtain the following type in our notation:

\[
\left( a \rightarrow m \rightarrow b ; b \rightarrow \text{ack} \rightarrow a \right)^* \\
\]

### Locking / Unlocking

Consider a set of \( n \) processes, each of which needs to acquire exclusive access to a resource by sending it a lock message. The resource replies with ack, the process uses the resource and unlocks it by sending an unlock message, at which point the next process can do the same, and so on. The following type describes the locking–unlocking protocol for a fixed number of processes.

\[
(c_1 \rightarrow \text{lock} \rightarrow s ; s \rightarrow \text{ack} \rightarrow c_1 ; c_1 \rightarrow \text{unlock} \rightarrow s) \otimes \ldots \otimes (c_n \rightarrow \text{lock} \rightarrow s ; s \rightarrow \text{ack} \rightarrow c_n ; c_n \rightarrow \text{unlock} \rightarrow s)
\]

With \( \otimes \) denoting shuffling, this formula expresses that any ordering of the \((c_i \rightarrow \text{lock} \rightarrow s ; s \rightarrow \text{ack} \rightarrow c_i ; c_i \rightarrow \text{unlock} \rightarrow s)\) events is acceptable. To support a dynamic network topology, the number of participants should be a parameter. The following is the locking–unlocking example in System-A, where conformance to the protocol is statically verifiable without knowledge of the runtime value of \( n \).

\[
\left( c_1 \rightarrow \text{lock} \rightarrow s ; s \rightarrow \text{ack} \rightarrow c_1 ; c_1 \rightarrow \text{unlock} \rightarrow s \right) \|^n_l \left( c_2 \rightarrow \text{lock} \rightarrow s ; s \rightarrow \text{ack} \rightarrow c_2 ; c_2 \rightarrow \text{unlock} \rightarrow s \right)^* \\
\]

### Limited Resource Sharing

In this scenario, a server \( s \) grants two clients \( c_1 \) and \( c_2 \) exclusive access to a set of \( n \) resources. At any given point, a maximum of \( n \) resources can be locked, but the relevant lock–ack–unlock messages from both clients can be interleaved in any way. Following is the global type for this situation:

\[
\left( c_1 \rightarrow \text{lock} \rightarrow s ; s \rightarrow \text{ack} \rightarrow c_1 ; c_1 \rightarrow \text{unlock} \rightarrow s \right) \|^n_l \left( c_2 \rightarrow \text{lock} \rightarrow s ; s \rightarrow \text{ack} \rightarrow c_2 ; c_2 \rightarrow \text{unlock} \rightarrow s \right)^* \\
\]

The parallel composition is parameterized in \( n \), the number of resources. Each sequence of lock–ack–unlock messages is also parameterized in \( i \), which ranges from 1 to \( n \). This is necessary to ensure realizability of the protocol, as in the case of multiple outstanding requests, it allows the participants to disambiguate the responses they receive. Each parallel instance subsumed by the \( \|^n_l \) operator consists of a loop (Kleene Star) which entails a choice, indicated by \( \oplus \). Either \( c_1 \) gets access to a resource, or \( c_2 \) and this happens repeatedly.

### 4 Type Syntax

#### 4.1 Global Types

A global type describes a protocol which the whole system must adhere to. The examples in Section 3 are all global types, since they describe the behavior of all participants. Global types in System-A can be constructed according to the grammar in Table 1 with operator descriptions following.
Parameterized Concurrent Multi-Party Session Types

Table 1: The grammar of global types

| G ::=          | (G)          | ε                  | ε                  |
|               | (G-Paren)    | (G-Empty)          | (G-Empty)          |
|               | G ; G        | (G-Seq)            | (G-Seq-N)          |
|               | G ⊕ G        | (G-Choice)         | (G-Choice-N)       |
|               | G ∥ G        | (G-Parallel)       | (G-Parallel-N)     |
|               | G ⊗ G        | (G-Shuffle)        | (G-Shuffle-N)      |
|               | p₁ type → p₂ | (G-Interaction)    | Gⁿ                 |
|               |              | (G-Exp)            | (G-KleeneStar)     |

(G-Seq) is used for the sequential composition of events.

(G-Choice) denotes exclusive choice between the arguments. For instance, G₁ ⊕ G₂ means that either G₁ or G₂ will be executed (but not both).

(G-Parallel) means that the arguments run in parallel; any interleaving of sequenced actions is possible. For instance, (a ↦₁ b ; a ↦₂ c) ∥ c ↦₃ b means that any of the interleavings ABC, ACB, CAB is possible, where A = (a ↦₁ b), B = (a ↦₂ c) and C = (c ↦₃ b). Notice that B is not allowed to precede A, as the ordering of actions as determined by operator ; is not allowed to change.

(G-Shuffle) means that both arguments are executed atomically, in an unspecified order. Formally, G₁ ⊗ G₂ ≡ (G₁ ; G₂) ⊕ (G₂ ; G₁).

(G-Interaction) denotes the sending and receiving of a message. For instance, p₁ ↦₁ → p₂ means that process p₁ sends a message of type t to process p₂.

(G-KleeneStar) has the usual semantics, of zero or more repetitions of the argument.

The n-ary versions of the operators express behaviors where the value of n is unknown at compile time. (G-Seq-N), (G-Choice-N), (G-Parallel-N), (G-Shuffle-N) apply the respective binary operator n−1 times to n global types, parameterized in i. (G-Exp) denotes n-fold repetition of the argument (in sequence). Note that for known values of n, we do not need the right column of Table 1 as the desired behavior can be produced by suitable repeated applications of the binary operators.

All of the operators are commutative, with the exception of sequencing. All operators are furthermore associative, with the exception of shuffling. In particular,

\[ \bigotimes_{i=1}^{n} G_i \neq \bigotimes_{i=1}^{n} (\cdots (G_1 \otimes G_2) \otimes G_3 \cdots) \otimes \cdots \otimes G_n \].

Instead, \( \bigotimes_{i=1}^{n} G_i \) means that all arguments Gᵢ are executed atomically, but in an unspecified order.

The distinction between the Kleene star and exponentiation is fundamental. The use of Gⁿ means that the protocol conformance checker will have to prove that the system is correct for any fixed value of the parameter n. G* on the other hand means an arbitrary number of repetitions of G. There is no parameter fixing this number and it may be different from instance to instance of the Kleene Star and/or among executions of the same program with the same run-time values for its parameters.
4.2 Local Types

A local type specifies the abstract behavior of a single protocol participant. The syntax of local types is given in Table 2 with descriptions following.

| L ::= (L) (L-Paren) | ε (L-Empty) |
| a!t (L-Send) | a?t (L-Recv) |
| L ; L (L-Seq) | \[ i=1 \to n \] L_i (L-Seq-N) |
| L ⊕ L (L-Choice) | \[ i=1 \to n \] (L-Choice-N) |
| L || L (L-Parallel) | \[ i=1 \to n \] L_i (L-Parallel-N) |
| L ⊗ L (L-Shuffle) | \[ i=1 \to n \] L_i (L-Shuffle-N) |
| L^n (L-Exp) | L^* (L-KleeneStar) |

(L-Send) denotes sending a message of type \( t \) to process \( a \).

(L-Recv) denotes receiving a message of type \( t \) from process \( a \).

In the sliding window example of Section 3, the behavior of the sender \( a \) is described by the local type \( n \parallel \{ (b!m ; b?ack) \}_i \). Leaving out the initial \( n \parallel \) symbol for the time being, what remains is \( (b!m ; b?ack)^* \). This means sending a message and then receiving an acknowledgment \( (b!m ; b?ack) \), an arbitrary number of times. Assuming that the window size \( n \) is a parameter, any interleaving of \( n \) of these sequences is possible, with the obvious constraint of not receiving more acknowledgments than the number of messages sent. This is ensured by composing sequences of the form \( (b!m ; b?ack) \), where ordering is forced by the \( ; \) operator.

5 Projection

The local type of the sliding window protocol in Section 4.2 is a restriction of the respective global type in Section 3 onto the individual processes. In this section, we investigate a way of automating this process. \( \mathcal{G} \triangleright p \) is read “the projection of global type \( \mathcal{G} \) onto process \( p \)” and the result is a local type as defined in Table 2. The projection function \( \triangleright \) is formally defined in Table 3 and the result of applying it to all the processes in the system is an environment \( \Delta = \{ p_i : \mathcal{L}_i \}_{i=1} \) which maps processes to local types.
Table 3: The projection function

\[
(a \xrightarrow{m} b) \triangleright p \quad ::= \begin{cases} 
b!m & \text{if } p = a \\
a?m & \text{if } p = b \\
\varepsilon & \text{otherwise}
\end{cases}
\quad (P\text{-Interaction})
\]

\[
\mathcal{G}^n \triangleright p \quad ::= \quad (\mathcal{G} \triangleright p)^n
\quad (P\text{-Exp})
\]

\[
(\mathcal{G}_1 \oplus \mathcal{G}_2) \triangleright p \quad ::= \quad (\mathcal{G}_1 \triangleright p) \oplus (\mathcal{G}_2 \triangleright p)
\quad (P\text{-Choice})
\]

\[
(\mathcal{G}_1 \parallel \mathcal{G}_2) \triangleright p \quad ::= \quad (\mathcal{G}_1 \triangleright p) \parallel (\mathcal{G}_2 \triangleright p)
\quad (P\text{-Paral})
\]

\[
(\mathcal{G}_1;\mathcal{G}_2) \triangleright p \quad ::= \quad (\mathcal{G}_1 \triangleright p);(\mathcal{G}_2 \triangleright p)
\quad (P\text{-Seq})
\]

\[
(\mathcal{G}_1 \otimes \mathcal{G}_2) \triangleright p \quad ::= \quad (\mathcal{G}_1 \triangleright p) \otimes (\mathcal{G}_2 \triangleright p)
\quad (P\text{-Shuffle})
\]

\[
( n \bigodot_{i=1} \mathcal{G}_i ) \triangleright p \quad ::= \quad n \bigodot_{i=1} (\mathcal{G}_i \triangleright p)
\quad (P\text{-Seq-N})
\]

\[
( n \bigoplus_{i=1} \mathcal{G}_i ) \triangleright p \quad ::= \quad n \bigoplus_{i=1} (\mathcal{G}_i \triangleright p)
\quad (P\text{-Choice-N})
\]

\[
( n \bigotimes_{i=1} \mathcal{G}_i ) \triangleright p \quad ::= \quad n \bigotimes_{i=1} (\mathcal{G}_i \triangleright p)
\quad (P\text{-Shuffle-N})
\]

\[
( n \parallel_{i=1} \mathcal{G}_i ) \triangleright p \quad ::= \quad n \parallel_{i=1} (\mathcal{G}_i \triangleright p)
\quad (P\text{-Paral-N})
\]

For the lock/unlock example of Section 3, projecting onto a client \(c_k\) and the server \(s\) yields

\[
\mathcal{G} \triangleright c_k = n \bigotimes_{i=1} (c_i \xrightarrow{\text{lock}} s; s \xrightarrow{\text{ack}} c_i; c_i \xrightarrow{\text{unlock}} s \triangleright c_k)
\quad (P\text{-Shuffle-N})
\]

\[
= n \bigotimes_{i\neq k} \varepsilon \otimes (s!\text{lock}; s?\text{ack}; s!\text{unlock})
\quad (P\text{-Interaction}), (P\text{-Seq})
\]

\[
= s!\text{lock}; s?\text{ack}; s!\text{unlock},
\quad \text{(eliminating } \varepsilon \text{)}
\]

\[
\mathcal{G} \triangleright s = n \bigotimes_{i=1} (c_i \xrightarrow{\text{lock}} s; s \xrightarrow{\text{ack}} c_i; c_i \xrightarrow{\text{unlock}} s \triangleright s)
\quad (P\text{-Shuffle-N})
\]

\[
= n \bigotimes_{i=1} (c_i?\text{lock}; c_i?\text{ack}; c_i?\text{unlock}).
\quad (P\text{-Interaction}), (P\text{-Seq})
\]

Similarly, the projected local types for the resource sharing protocol of Section 3 are

\[
\mathcal{L}_s = n \parallel_{i=1} (c_i?\text{lock}_i; c_i?\text{ack}_i; c_i?\text{unlock}_i \oplus c_2?\text{lock}_i; c_2?\text{ack}_i; c_2?\text{unlock}_i)^*
\]

\[
\mathcal{L}_{c_1} = n \parallel_{i=1} (s!\text{lock}_i; s?\text{ack}_i; s!\text{unlock}_i)^*
\]

\[
\mathcal{L}_{c_2} = n \parallel_{i=1} (s!\text{lock}_i; s?\text{ack}_i; s!\text{unlock}_i)^*.
\]
6 Type Checking

Given a global type, we need to be able to check the respective projections against the local types inferred from the program itself. This is possible due to the following properties of our language of local types:

**Theorem 1** (Weak Normalization). For any local type \( L \) in System-A, there exists a finite sequence of reduction steps which brings the type to a normal form.

We prove this in the extended version of this paper [13], where we provide the reduction semantics and a normalization process.

**Corollary 1** (Strong Normalization). For every local type \( L \) in System-A, all sequences of reduction steps are finite and lead to the same normal form.

In the extended version of this paper, we show that the aforementioned normalization process uniquely determines the reduction semantics, implying the uniqueness of normal forms.

Checking structural equivalence of the types derived from the program against the projections is decidable up to \( \alpha \)-conversion. However, all that is required to overcome this issue is that names in the code are consistent with those in the supplied global type. In our opinion it is reasonable to expect programmers to adhere to such a naming convention.

7 Global Type Realization

In this section, we discuss the properties that a given global type must satisfy in order to be projectable. These properties are discussed while assuming actor semantics [1] for the messaging system; that is, asynchronous, unordered and eventual (guaranteed, albeit with arbitrary delay) delivery of messages. Applying the projection function to a projectable global type will result in local types for the participants, whose combined behavior is consistent with the global type—a fact we show in Section 8.

The subsequent discussion of projectability criteria uses the following definitions:

**Definition 1** (Event). An event is a single interaction \( p_1 \xrightarrow{m} p_2 \) in a global type.

We extend the projection function onto events and write \( e \triangleright p \) to denote the projection of event \( e \) onto process \( p \) using rule (P-Interaction).

**Definition 2** (Trace). A trace is a sequence of events producible by a global type \( \mathcal{G} \) and is of the form \( e_1 ; e_2 ; \ldots ; e_k \). The set of traces a global type \( \mathcal{G} \) can produce is denoted by \( \text{tr}(\mathcal{G}) \). The first and last events of a trace \( t \) are denoted \( \text{first}(t) \) and \( \text{last}(t) \) respectively. Abusing notation, the set of events that appear first in traces of \( \mathcal{G} \) is denoted \( \text{first}(\mathcal{G}) = \{ \text{first}(t) \mid t \in \text{tr}(\mathcal{G}) \} \). Similarly, the set of events that appear last in traces of \( \mathcal{G} \) is denoted \( \text{last}(\mathcal{G}) \).

Since a trace is simply a sequence of events of the form \( p \xrightarrow{m} q \), we extend the projection function onto traces in the natural way. We write \( t \triangleright p \) to denote the projection of trace \( t \) onto process \( p \) using rules (P-Seq) and (P-Interaction).

7.1 Sequentiality Criterion

The purpose of this criterion is to ensure that the sequential constructs of a global type retain sequential semantics after projection. As an example problematic case, consider \( \mathcal{G}_1 = a \xrightarrow{m_1} b ; c \xrightarrow{m_2} d \). Without the use of some covert coordination channel (for example by implementing a barrier mechanism), it is impossible for \( c \) to know when \( b \) has received the message. The two events \( a \xrightarrow{m_1} b \) and \( c \xrightarrow{m_2} d \) are impossible to order using our projection function, as the resulting environment would be \( \Delta_1 = \{ a :
b!m_1, b : a;m_1, c : d!m_2, d : c;m_2\}, which allows c to send m_2 to d before a sends m_1 to b. \mathcal{G}_1 does not satisfy the sequentiability criterion and thus is not projectable.

Another problematic case is \mathcal{G}_2 = a \xrightarrow{m_2} b ; a \xrightarrow{m_2} b, where a cannot know when m_1 has been received so as to start transmitting m_2, hence \mathcal{G}_2 is not projectable either. The following definition captures the conditions under which events are guaranteed to respect the sequencing restrictions imposed in a global type, when the latter is projected onto individual processes.

**Definition 3** (Sequentially Projectable Global Type). The set of *sequentially projectable* (SP) global types is defined inductively as follows:

\[
\begin{align*}
\mathcal{G} &\in \text{SP} \quad \forall \mathcal{G} \\
\mathcal{G}_1 \cdot \mathcal{G}_2 &\in \text{SP} \quad \forall \mathcal{G}_1, \mathcal{G}_2 \\
(\forall e_1 \in \text{last}(\mathcal{G}_1), e_2 \in \text{first}(\mathcal{G}_2) \Rightarrow (e_1 ; e_2) \in \text{SP}) &\Rightarrow \mathcal{G}_1 ; \mathcal{G}_2 \in \text{SP} \\
(\forall e_1 \in \text{last}(\mathcal{G}_1[i]), e_2 \in \text{first}(\mathcal{G}_1[j]) \Rightarrow (e_1 ; e_2) \in \text{SP}) &\Rightarrow \mathcal{G}_1[i] ; \mathcal{G}_1[j] \in \text{SP} \\
(\forall e_1 \in \text{last}(\mathcal{G}), e_2 \in \text{first}(\mathcal{G}) \Rightarrow (e_1 ; e_2) \in \text{SP}) &\Rightarrow \mathcal{G}^{n} \in \text{SP} \\
(\forall e_1 \in \text{last}(\mathcal{G}), e_2 \in \text{first}(\mathcal{G}) \Rightarrow (e_1 ; e_2) \in \text{SP}) &\Rightarrow \mathcal{G}^{*} \in \text{SP}
\end{align*}
\]

where \Pi denotes the set of processes.

Illustrating the third case of the definition above, the following global type is in SP:

\[
\mathcal{G} = (a \xrightarrow{m} b \parallel c \xrightarrow{m} b) ; (b \xrightarrow{m} l \parallel b \xrightarrow{m} k)
\]

It is easy to see that \text{last}\left( a \xrightarrow{m} b \parallel c \xrightarrow{m} b \right) = \{a \xrightarrow{m} b, c \xrightarrow{m} b\} and \text{first}\left( b \xrightarrow{m} l \parallel b \xrightarrow{m} k \right) = \{b \xrightarrow{m} l, b \xrightarrow{m} k\} so that all four sequences (e.g. \text{a} \xrightarrow{m} \text{b} ; \text{b} \xrightarrow{m} \text{k}) are in \text{SP} according to the first two lines of the definition above.

7.2 Choice Criterion

The purpose of this criterion is to ensure that projecting \mathcal{G}_1 \oplus \mathcal{G}_2 maintains the choice semantics, meaning that all participants can recognize which branch of the choice operator they need to take during execution. As an example of a type that does not satisfy this criterion, consider

\[
\mathcal{G} = (a \xrightarrow{m} b ; b \xrightarrow{k} c ; c \xrightarrow{t_1} d) \oplus (a \xrightarrow{m} b ; b \xrightarrow{k} c ; c \xrightarrow{t_2} d).
\]

Here, a and b know which branch they are on, because on the left branch b receives a message of type m_1 from a, while on the branch on the right it receives a message of type m_2. However, from that point on, b behaves identically with respect to c, which has no way of telling whether the message to send to d should be of type t_1 or t_2. We call the first point at which two traces differ with respect to a given process the **distinctive point**, which can be \epsilon if no such point exists. This notion is formalized in the following definition:

**Definition 4** (Distinctive Point). The **distinctive point** of a process \( p \) with respect to a pair of traces \( t_1 = (e_1, \ldots, e_k) \in \text{tr}(\mathcal{G}) \) and \( t_2 = (f_1, \ldots, f_l) \in \text{tr}(\mathcal{G}) \) is an index \( i \) given by

\[
d_{t_1,t_2}(p) = \min\{i \mid (e_i \triangleright p) \neq (f_i \triangleright p)\}
\]

where \( e_i, f_i \) denote events. In the case where \( t_1 \triangleright p = \epsilon \), or \( t_2 \triangleright p = \epsilon \), or \( t_1 \triangleright p = t_2 \triangleright p \), no such \( i \) exists and the distinctive point is defined to be \( \epsilon \).
The definition that follows captures the conditions under which the choice semantics are maintained after projection. The first bullet deals with the non parameterized version of the choice operator $\oplus$. Item (i) captures the case where a process $p$ is the first process acting on the two branches, in which case it must inform the others of the branch they are on. It does so by either sending a different message, or by sending to a different process in each case. Note that the same process must inform the others on both branches. Item (ii) captures the case where $p$ is not the first process to act, in which case it must be informed of the branch it is on and the distinctive point should be a suitable receive event.

Notice how the second bullet deals with shuffling by means of choice. Clearly, if a process can tell whether it is on $G_1$ or $G_2$, it is also able to tell the order in which they appear.

The third bullet inductively uses the previous two to define choice-wise projectability in the parameterized cases of choice $\oplus$ and shuffle $\otimes$.

**Definition 5** (Choice-Wise Projectable Global Type). The set of Choice-Wise Projectable (CP) global types is defined inductively as follows:

- $G = G_1 \oplus G_2 \in CP \iff \forall p \in G$, either of the following is true:
  
  (i) $G = G_1 \oplus G_2$ and
  
  \[ \forall e_1 \in first(G_1), e_2 \in first(G_2) : e_1 = p \xrightarrow{m} q, e_2 = p \xrightarrow{m'} q' \quad \text{where } p \neq q \text{ and } p \neq q' \quad \text{and } (q \neq q' \text{ or } m \neq m') \]

  (ii) $\forall t_1 = (s_1, \ldots, s_{k_1}) \in tr(G_1), t_2 = (u_1, \ldots, u_{k_2}) \in tr(G_2)$,
  
  either $d_{t_1,t_2}(p) = e$, or
  
  \[ d_{t_1,t_2}(p) = i \text{ and } s_i = q \xrightarrow{m} p, u_i = q' \xrightarrow{m'} p \quad \text{where } p \neq q \text{ and } p \neq q' \quad \text{and } (q \neq q' \text{ or } m \neq m') \]

- $G = G_1 \otimes G_2 \in CP \iff G_1 \oplus G_2 \in CP$

- $G = \bigoplus_{i=1}^{n} G_i \in CP \iff (G_i\{1/i\} \oplus G_i\{2/i\}) \in CP$

- $G = \bigotimes_{i=1}^{n} G_i \in CP \iff (G_i\{1/i\} \otimes G_i\{2/i\}) \in CP$

The criterion for parameterized shuffling $\bigotimes$ is stricter than what one can derive if the value of $n$ is given. However, it is hard to loosen up the constraint when it is dealt with as a parameter.

### 7.3 Parallel Composability Criterion

As an example of what can go wrong when composing two global types using the $|$ operator, consider the example $G = \left( (a \xrightarrow{m_1} b ; b \cdot k_1 \cdot c) \oplus (a \xrightarrow{m_2} b ; b \cdot k_2 \cdot c) \right) \parallel a \xrightarrow{m_3} b$. The intended behavior of $G$ is that $a$ chooses whether to send a message of type $m_1$ or $m_2$ to $b$, which in turn decides whether to send $c$ a message of type $k_1$ or $k_2$. Concurrently with this, an additional $m_3$ is sent from $a$ to $b$. Assume that as far as $\oplus$ is concerned, $a$ decides to send $m_2$ to $b$. It is then obvious how the additional parallel event $a \xrightarrow{m_3} b$ might confuse $b$ to simultaneously take both branches of the choice operator.

In general, the problem appears when actions in one parallel branch affect choices made on another. Global types that do not exhibit this problem are parallel projectable (PP).
Definition 6 (Parallel Projectable Global Types). For two global types $\mathcal{G}_1$ and $\mathcal{G}_2$, $\mathcal{G} = \mathcal{G}_1 \parallel \mathcal{G}_2$ is parallel projectable (PP) if there is no overlap between the distinctive points in $\mathcal{G}_1$ and events in $\mathcal{G}_2$. Formally, for two global types $G_1$ and $G_2$, $G = G_1 \parallel G_2$ is parallel projectable (PP) if there is no overlap between the distinctive points in $G_1$ and events in $G_2$. Formally, for two global types $G_1$ and $G_2$, $G = G_1 \parallel G_2$ is parallel projectable (PP) if there is no overlap between the distinctive points in $G_1$ and events in $G_2$. Formally,

- $\mathcal{G}_1 \parallel \mathcal{G}_2 \in \text{PP}$ iff $\forall t_1 = (e_1, \ldots, e_k), t_2 = (e'_1, \ldots, e'_k) \in \text{tr}(\mathcal{G}_1), t_2 \in \text{tr}(\mathcal{G}_2), p \in \Pi$ we have $d_{t_1, t'_1}(p) = i$ and one of the following is true:
  
  (a) $i = \varepsilon$

  (b) $e_i, e'_i$ both have $p$ as the sender

  (c) $e_i \notin t_2$ and $e'_i \notin t_2$

- $\bigoplus_{i=1}^{n} \mathcal{G}_i \in \text{PP}$ iff $(\mathcal{G}_i\{1/i\} \parallel \mathcal{G}_i\{2/i\}) \in \text{PP}$

where $\Pi$ denotes the set of processes. Notice how this definition incorporates parallel composability of two Kleene starred types (the Kleene Star entails a choice pertaining to loop entrance and exit).

7.4 Kleene Star Criterion

Use of the Kleene Star in global types can result in protocols whose projection is unsafe, that is, can result in execution traces that are not part of the original global type. To avoid this, a global type must be such that the entry and exit conditions to the starred type can be identified by all participants. Determining whether this is the case requires inspection of not only the starred type itself, but also of what comes after the starred section.

Definition 7 (Kleene Star Projectable Global Types). For global types $\mathcal{G}, \mathcal{G}'$, we say that $\mathcal{G}^*; \mathcal{G}'$ is Kleene Star Projectable (KP) iff $\mathcal{G} \oplus \mathcal{G}' \in \text{CP}$.

As an example of a type that is not in KP, consider $\mathcal{G} = (a \xrightarrow{m} b ; b \xrightarrow{m'} c)\cdot c \xrightarrow{m''} d$ where $c$ has no way of knowing whether it should wait for $m'$ from $b$, or proceed immediately with sending $d$ the message $m''$.

8 Correctness

The conditions discussed above are sufficient to ensure that the projection function generates local types which are functionally consistent with the global type. We call a global type that satisfies all of the above criteria projectable:

Definition 8 (Projectable Global Type). The set of projectable (PR) global types is inductively defined in Table 4.

Theorem 2 formalizes our intuition that under the constraints mentioned above, the projection function is correct; that is, the projected environment is consistent with the global type. In what follows, $\text{tr}(\Delta)$ with $\Delta = \{p_i : \mathcal{L}_i\}_{i \in I}$ denotes the set of traces producible by environment $\Delta$. Also, $\Delta_{\mathcal{G}}$ denotes the environment resulting from the projection of $\mathcal{G}$ onto the set of processes, i.e. $\Delta_{\mathcal{G}} = \{p : \mathcal{G} \triangleright p\}_{p \in \Pi}$.

Theorem 2. $\mathcal{G} \in \text{PR} \Rightarrow \text{tr}(\mathcal{G}) = \text{tr}(\Delta_{\mathcal{G}})$

We sketch the proof of this theorem in the extended version of this paper [13], where we inductively treat each of the cases in Table 4. What needs to be proved is essentially $\forall t \in \text{tr}(\mathcal{G}) \Leftrightarrow t \in \text{tr}(\Delta_{\mathcal{G}})$. While the forward direction is rather obvious, proving that the projected environment does not generate traces that are not part of the original global type is trickier and is why we need the criteria of Section 7.

Proving this theorem, we get a correctness proof of our projection function (given the premises discussed previously) for free.
Semantic comparison is unnecessary for the weak normalization proof, but would be useful in a practical implementation. An algorithm [13] already includes many cases of semantically equivalent, yet structurally differing types. Implementation of this algorithm along with the relevant type inference algorithm is the next step towards a practical system.

In System-A, we can statically verify—without instantiating the parameters—the compliance of implementations to protocols: we do this by first projecting the specification to parameterized types, and then comparing these projections against the types extracted from the program. An important result we obtain is that structural equivalence of types in System-A is decidable; we present this result in Section 6 by first showing weak and subsequently strong normalization of local types. Unlike other typing proposals, System-A does not depend on other theories (typed λ-calculus, system T, or system F) for type-checking. In Section 7 we discuss the conditions under which our projection function is correct and state their sufficiency in Section 8.

9 Conclusions and Future Work

We introduced System-A which allows for parameterized parallelism, where the number of participants, types of messages sent, as well as the number of such messages are controlled by type parameters. Choice among various execution paths can also be parameterized, so that the number and types of different paths to be taken is not known at compile time. System-A also introduces a shuffling operator, which expresses arbitrary reordering of its arguments, again in a parameterized fashion. A series of examples demonstrates the usefulness of these extensions, which allow us to specify and check previously inexpressible interactions such as the sliding window protocol and parallel resource locking/unlocking (Section 3). In System-A, we can statically verify—without instantiating the parameters—the compliance of implementations to protocols: we do this by first projecting the specification to parameterized types, and then comparing these projections against the types extracted from the program. An important result we obtain is that structural equivalence of types in System-A is decidable; we present this result in Section 6 by first showing weak and subsequently strong normalization of local types. Unlike other typing proposals, System-A does not depend on other theories (typed λ-calculus, system T, or system F) for type-checking. In Section 7 we discuss the conditions under which our projection function is correct and state their sufficiency in Section 8.

Future Work. Complete type checking with System-A is only decidable up to type inference; we do not provide an algorithm for inference of suitable types in an actor language. The design of a programming language along with the relevant type inference algorithm is the next step towards a practical implementation.

Another practical consideration includes semantic comparison of local types. Our normalization algorithm [13] already includes many cases of semantically equivalent, yet structurally differing types. Semantic comparison is unnecessary for the weak normalization proof, but would be useful in a prac-

---

Table 4: The set PR

<table>
<thead>
<tr>
<th>Table 4: The set PR</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \varepsilon \in PR )</td>
</tr>
<tr>
<td>( a \xrightarrow{m} b \in PR )</td>
</tr>
<tr>
<td>( \mathcal{G} \in PR ) and ( \mathcal{G}^n \in SP ) ( \Rightarrow ) ( \mathcal{G}^n \in PR )</td>
</tr>
<tr>
<td>( \mathcal{G}_1, \mathcal{G}_2 \in PR ) and ( (\mathcal{G}_1 ; \mathcal{G}_2) \in SP ) ( \Rightarrow ) ( (\mathcal{G}_1 ; \mathcal{G}_2) \in PR )</td>
</tr>
<tr>
<td>( \mathcal{G}_1, \mathcal{G}_2 \in PR ) and ( (\mathcal{G}_1 \oplus \mathcal{G}_2) \in CP ) ( \Rightarrow ) ( (\mathcal{G}_1 \oplus \mathcal{G}_2) \in PR )</td>
</tr>
<tr>
<td>( \mathcal{G}_1, \mathcal{G}_2 \in PR ) and ( (\mathcal{G}_1 \otimes \mathcal{G}_2) \in CP ) ( \Rightarrow ) ( (\mathcal{G}_1 \otimes \mathcal{G}_2) \in PR )</td>
</tr>
<tr>
<td>( \mathcal{G}_1, \mathcal{G}_2 \in PR ) and ( (\mathcal{G}_1 \parallel \mathcal{G}_2) \in PP ) ( \Rightarrow ) ( (\mathcal{G}_1 \parallel \mathcal{G}_2) \in PR )</td>
</tr>
<tr>
<td>( \mathcal{G}_1, \mathcal{G}_2 \in PR ) and ( \mathcal{G}^n_1 \in SP ) and ( (\mathcal{G}^n_1 ; \mathcal{G}_2) \in SP \cap KP ) ( \Rightarrow ) ( (\mathcal{G}^n_1 ; \mathcal{G}_2) \in PR )</td>
</tr>
<tr>
<td>( \mathcal{G}_i {1/i}, \mathcal{G}_i {2/i} \in PR ) and ( (\mathcal{G}_i {1/i} ; \mathcal{G}_i {2/i}) \in SP ) ( \Rightarrow ) ( \oplus \mathcal{G}_i \in PR )</td>
</tr>
<tr>
<td>( \mathcal{G}_i {1/i}, \mathcal{G}_i {2/i} \in PR ) and ( (\mathcal{G}_i {1/i} \oplus \mathcal{G}_i {2/i}) \in CP ) ( \Rightarrow ) ( \bigoplus \mathcal{G}_i \in PR )</td>
</tr>
<tr>
<td>( \mathcal{G}_i {1/i}, \mathcal{G}_i {2/i} \in PR ) and ( (\mathcal{G}_i {1/i} \parallel \mathcal{G}_i {2/i}) \in CP ) ( \Rightarrow ) ( \bigotimes \mathcal{G}_i \in PR )</td>
</tr>
<tr>
<td>( \mathcal{G}_i {1/i}, \mathcal{G}_i {2/i} \in PR ) and ( (\mathcal{G}_i {1/i} \parallel \mathcal{G}_i {2/i}) \in PP ) ( \Rightarrow ) ( \bigparallel \mathcal{G}_i \in PR )</td>
</tr>
</tbody>
</table>
tical setting where the user is interested in semantic adherence to a protocol. Specifically for the case where reordering of terms is possible as a result of operator commutativity, our suggested coding only serves as an existential proof. A more practical coding scheme could be developed, perhaps employing lexicographic ordering.

Denielou et al. [14] propose a system where parameterization is achieved by means of quantification over roles. Roles are behavior specifications that are taken up by processes while they participate in a protocol, and processes are allowed to join and leave protocols (respectively, adopt and drop roles) dynamically. Their notation’s expressiveness is limited when it comes to arbitrary, concurrency-induced interleavings of events. Nevertheless, incorporating their ideas in System-A would greatly expand the applicability of the ideas presented here, towards a different direction than what is addressed in the present paper.

Support for session delegation and exception handling (in the sense of Carbone et al. [10]) represents another opportunity for extension. Furthermore, it may be possible to transfer the recent, precise realizability results [4] for choreographies [33] to our parameterized specifications.

Acknowledgments

The authors would like to thank the anonymous reviewers for their insightful comments. This publication was made possible in part by sponsorships from the Air Force Research Laboratory and the Air Force Office of Scientific Research under agreement number FA8750-11-2-0084, as well as the Army Research Office under Award No. W911NF-09-1-0273. The U.S. Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright notation thereon.

References


